

Keyifli Okumalar Dileriz

Binlerce kategoride milyonlarca PDF e-kitap, ders kitapları, dergiler ve romanlara platformumuz üzerinden ücretsiz ulaşabilirsiniz.

RESMİ WEB SİTEMİZ

<https://rantey.org>

Yedek ve Duyuru Kanallarımız

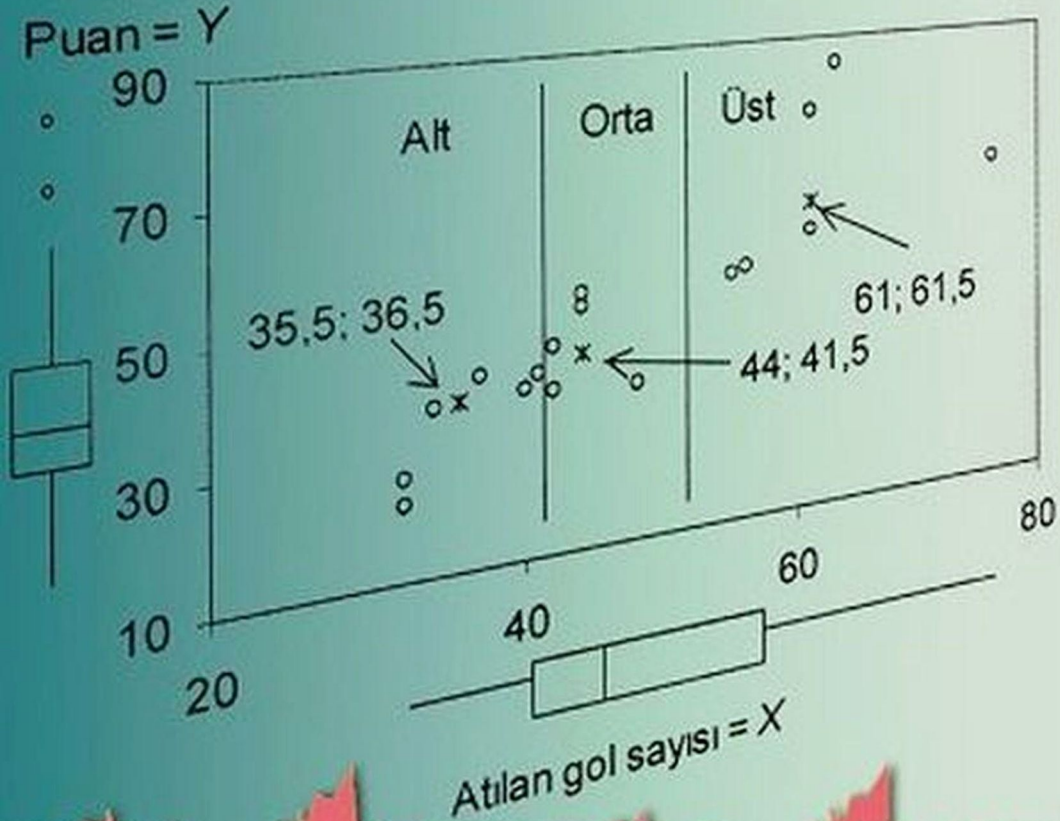
Sitemize erişim engeli gelmesi veya adres değişikliği durumunda aşağıdaki kanallardan güncel giriş adresine erişebilirsiniz:

- Telegram: <https://t.me/birkitapbir>
- WhatsApp Kanalımız : <https://whatsapp.com/channel/0029VbDHv8uE50Us4lvMoc0Y>

İSTATİSTİK

SAYILARIN ARKASINI ANLAMAK

ÜMİT ŞENESEN



İSTATİSTİK

Sayıların Arkasını Anlamak

İSTATİSTİK

Sayıların Arkasını Anlamak

Prof. Dr. Ümit Şenesen

İSTANBUL TEKNİK ÜNİVERSİTESİ
İŞLETME FAKÜLTESİ



Literatür Yayınları: 128

İSTATİSTİK

Sayıların Arkasını Anlamak

Ümit Şenesen

Birinci Basım: Ekim 2004, İstanbul

İkinci Basım: Ekim 2006, İstanbul

Üçüncü Basım: Ekim 2007, İstanbul

Dördüncü Basım: Ekim 2013, İstanbul

Genel Yayın Yönetmeni: Kenan Kocatürk

Uygulama: Aslı Akgün

Kapak Tasarım: Emel Atik

Baskı ve Cilt: Çevik Matbaacılık - Tel.: (0212) 501 30 19

ISBN: 978-975-04-0283-8

© Copyright 2004, Literatür: Yayıncılık

Bu kitabın yayın hakları Literatür: Yayıncılık, Dağıtım,

Pazarlama San. ve Tic. Ltd. Şti.'ne aittir. Kitabın tamamı veya bir bölümü

hiçbir biçimde fotokopi yoluyla çoğaltılamaz, dağıtılamaz,

yeniden elde edilmek üzere saklanamaz.

Sertifika No: 10843

LİTERATÜR: YAYINCILIK, DAĞITIM,

PAZARLAMA, SANAYİ VE TİCARET LTD. ŞTİ.

İstiklal Caddesi, Emgen Han, No: 47 Kat: 4

Beyoğlu 34433 İstanbul

T 0(212) 292 4120

F 0(212) 245 5987

E literatur@literatur.com.tr

www.literatur.com.tr

NEYİ BİLDİRİR SAYILAR?

*sayılar yaklaşan bir şeyleri bildirir
sayılar bildirir uzaklaşan bir şeyleri
nedir yaklaşan bize
bizden uzaklaşan nedir?*

...

*neyi bildirir sayılar
neyi bildirmeli
yaklaşan nedir bize
uzaklaşan nedir bizden?*

Nâzım Hikmet

Ümit Şenesen 1971'den beri görev yapmakta olduğu İstanbul Teknik Üniversitesi'nden 2011'de emekli olmuştur. Halen aynı yerde ders vermeyi sürdürmektedir. Gerek yerli, gerek yabancı bilimsel dergilerde çeşitli yazıları yayımlanmıştır. *Belirsizlik Koşullarında Karar Verme*(1981), *Türkiye Ekonomisi için Bir Ekonometrik Model Denemesi* (1985), *Betimleyici, Sorgulayıcı İstatistik* (1998) kitaplarının yazarıdır. İngilizce'den çevirdiği kitaplar şunlardır: C. H. Sharp'tan *Ulaştırma Ekonomisi* (1978), P. Sraffa'dan *Malların Mallarla Üretimi* (1981, 2011), A. Koutsoyiannis'ten (G. Günlük-Şenesen'le) *Ekonometri Kuramı* (1989), D. N. Gujarati'den (G. Günlük-Şenesen'le) *Temel Ekonometri* (1999, 2012), P. Newbold'dan *İşletme ve İktisat için İstatistik* (2000), J. E. Freund'dan *Matematiksel İstatistik* (2001).

senesen@itu.edu.tr

Önsöz	xiii
1 İSTATİSTİK NEDİR? NEYE YARAR?	1
1.1 Giriş	1
1.2 İstatistik nedir?	3
1.3 İstatistik neye yarar?	5
1.4 İstatistik ne gibi işler görür?	7
Veri derleme, deney tasarımı	7
Sorgulayıcı veri çözümlemesi	8
Kestirme	8
Sınama, karar verme	9
1.5 İstatistiğin kötüye kullanımı	10
1.6 Bilgisayarla istatistik	11
Okuma önerileri, Alıştırmalar	12
2 İSTATİSTİK NEREDEN ÇIKTI?	17
2.1 Veri derlemenin evrimi	17
2.2 İstatistik yöntemlerinin evrimi	20
Okuma önerileri, Alıştırmalar	23
3 NEYİ, NASIL ÖLÇELİM?	25
3.1 Anakütle, örneklem, değişken, değer	25
3.2 Ölçeklerin önemi	27
3.3 Ad ölçeği	27
3.4 Sıra ölçeği	31
3.5 Eşaralılık ölçeği	33
3.6 Oran ölçeği	36
3.7 Yöntem seçiminde ölçeklerin önemi	39
Okuma önerileri, Alıştırmalar	41

4	KESMEYELİM DE YUVARLATALIM MI?	51
4.1	Sayılar neden kesilir ya da yuvarlatılır?	51
4.2	Yuvarlatma kuralları	53
4.3	Yuvarlatma hataları	58
	Okuma önerileri, Alıştırmalar	60
5	HANGİ VERİYİ, NEREDE ARAYALIM?	65
5.1	Giriş	65
	Birinci el veri – ikinci el veri	66
5.2	Yerli kaynaklar	67
5.3	Yabancı kaynaklar	72
	Okuma önerileri, Alıştırmalar	73
6	ÖRNEKLEME RASSAL MI OLSUN?	75
6.1	Tamsayım – örnekleme	75
6.2	Rassal örnekleme – rassal olmayan örnekleme	77
6.3	Rassal örnekleme türleri	84
	Basit rassal örnekleme	84
	Atlamalı rassal örnekleme	87
	Katmanlı rassal örnekleme	89
	Aşamalı rassal örnekleme	94
	Kümeli rassal örnekleme	95
6.4	Deney tasarımı	96
	Okuma önerileri, Alıştırmalar	99
7	BU ÇİZELGE, BU ÇİZİM NE ANLATIYOR?	103
7.1	Verilerin sunulması	103
7.2	İyi bir çizelge nasıl olmalı?	107
7.3	İyi bir çizim nasıl olmalı?	116
7.4	Kötü örnekler	144
	Okuma önerileri, Alıştırmalar	147

8 AŞAĞI YUKARI NE KADAR?	167
8.1 Ortalama düzey	167
8.2 Berk ortalamalar	169
Ensıkdeğer (<i>ESD</i>)	170
Ortanca (<i>Ortc</i>)	175
Bölenler	179
Dördebölenler ortalaması	181
8.3 Duyarlı ortalamalar	182
Aritmetik ortalama (<i>AO</i>)	184
Geometrik ortalama (<i>GO</i>)	193
Ters (Harmonik) ortalama (<i>TO</i>)	204
Kareli ortalama (<i>KO</i>)	208
Yarı duyarlı bir ortalama: Kırpık ortalama (<i>KırO</i>)	206
Okuma önerileri, Alıştırmalar	211
9 NE KADAR OYNAK?	
BAKIŞIK MI, ÇARPIK MI? SIVRI Mİ, BASIK MI?	223
9.1 Oynaklık – yayıklık – değişkenlik	223
9.2 Nasıl ölçülür?	224
Aralık	225
Dördebölenler aralığı	226
Dördebölenler sapması	227
Ortalama mutlak sapma	228
Standart sapma – varyans	229
Değişkenlik katsayısı	237
Görelî fark	238
9.3 Çarpıklık ölçüleri	239
Ortalamalarla çarpıklık	239
Dördebölenlerle çarpıklık	240
Beklemlerle (momentlerle) çarpıklık	241
9.4 Basıklık ölçüleri	244
Beklemlerle (momentlerle) basıklık	244
Çarpıklık-basıklık ölçüsü	244
9.5 Gelir dağılımında çarpıklık ölçüleri	245
Lorenz eğrisi	246
Gini yoğunlaşma katsayısı	247
Toplanma oranı	253
Okuma önerileri, Alıştırmalar	255

10 SORGULAYALIM, BAKALIM NE ÇIKACAK?	269
10.1 Sorgulayıcı veri çözümlemesi	269
10.2 Tek değişkenli verilerin sorgulanması	271
Dal-yaprak	271
Beşli özet	280
Dışadüşenler	285
Kutu-çizimi	288
Standartlaştırma	294
Dönüştürme	296
Tek veri kümesini dönüştürme	299
Birden çok veri kümesini hep birlikte dönüştürme	306
Okuma önerileri, Alıştırmalar	313
11 ARALARINDA İLİŞKİ VAR MI?	323
11.1 İki boyutlu çizelgeler	323
Simgeli çizelgeler	324
Ortanca işleme	329
11.2 İki nicel değişken arasındaki ilişki	339
Direnci doğru bulma	343
11.3 İki değişkenli incelemede dönüştürme	357
Okuma önerileri, Alıştırmalar	366
KAYNAKLAR	375
TÜRKÇE-İNGİLİZCE DİZİN	385
İNGİLİZCE-TÜRKÇE DİZİN	393

Bu kitap 1998’de ilk basıldığında pek okunacağını sanmıyordum. Yine de okunup okunmadığını bilmiyorum ama baskısının tükenmesi beni biraz şaşırttı. Hele Bodrum’da Gümüşlük Akademisi’nde adı da verilerek bir bölümünün ele alınmasına inanmak epeyce güç oldu. Yaz sıcağında, deniz kıyısında sayılarla ilgili bir konuyu ele almayı kim kendiliğinden isteyebilir ki? Bu iş en azından “sıkıcı” gelmez mi?

“Demek ki ikinci bir baskı gerekiyor” diye düşündüm. “Ama iyice genişletilmiş, zenginleştirilmiş olarak.” Bu amaçla çalışmaya koyuldum. Bütün bölümleri baştan sona elden geçirip yeniden yazdım, alıştırıcıların hem sayısını çok artırdım, hem eskileri güncelledim. Her bölümün sonuna okuma önerileri ekledim, dileyen daha derinlemesine araştırabilsin diye!

İlk baskının önsözünde benzeri kitaplarda bulunmayan ya da pek az yer tutan konulardan söz etmiştim. İstatistiğin evrimi, veri kaynakları, çizelgeler, çizimler, ölçme düzeyleri, ortalamalar, çarpıklık, basıklık, sayıların yuvarlatılması, dönüştürme, sorgulayıcı veri çözümlemesi bunlar arasındaydı. Yeni baskıda daha da ayrıntılı ele alınan bu konulara, örneklem alma, iki değişkenli veri çözümlemesinde kullanılan simgeli çizelgeler, ortanca işleme, dirençli doğru, eğriden doğruya dönüştürme gibi yöntemler de eklendi.

İstatistik formüllerinin ezberlenmesinden çok kavramların anlaşılması amaçlandı, karmaşık gibi görünen sayı kümelerinin arkasında yatan gizli yapıların nasıl ortaya çıkarılacağı açıklanmaya çalışıldı.

Verilen örnekler, araştırma soruları ya Türkiye’deki yaşamın içinden ya da bu ülkede yaşayan insanları ilgilendiren konulardan seçildi. Bebek ölüm oranlarından vergi oranlarına, fiyatlardan futbol karşılaşmalarına, TV izlenme oranlarından boşanma oranlarına, gelir dağılımından not dağılımına. Irak savaşından ortalama ömüre, kadınların yönetimde görev almasından üretimin dışarıya bağımlılığına, savunma bütçesinden içki tüketimine kadar değişen bir yelpaze, sayılar kullanılarak incelendi.

İlk baskıda çok kişinin emeği geçmişti. Herkese yeniden teşekkür ederim. Ama bunlara ikinci baskıda yenileri eklendi. Araştırma görevlileri

Barış Turan ile Cem Çakmaklı metni baştan sona okuyup düzeltmeler önerdi. Onların titiz dikkatleri olmasa gülünç yanlışlar önlenemezdi. Cem'in katkıları iki harf arasının fazla açık olmasından başlayıp Kafka'dan yapılan alıntının doğruluğuna kadar uzandı. Ayrıca sondaki "dizin"in düzenlenmesine de katıldı. Kitabı öğrencilerine okutan meslektaşım Raziye Selim de güç anlaşılan yerleri belirterek yeniden yazılmasını sağladı. Öğrencilerim gördükleri yanlışları, kendini ele vermeyen kapalı söylemleri göstererek beni uyardılar. Hepsine gönül borcum var.

Bana okuma-yazmayı öğreten annemden başlayarak ilkokuldan bugüne kadarki bütün öğretmenlerimin, yalnızca bu kitabın değil, bütün yaşamımın biçimlenmesinde harcadıkları emeği saygıyla anıyorum.

Yazma aşamasında her başım sıkıştığında can yoldaşım Gülay'a danıştım. Beni hep yüreklendirdi. Kitabın bir an önce bitmesi için baskılar uyguladı. Ev sorumluluklarımı aksatmama göz yumdu. Bu arada kendi çalışmalarından nasıl vakit bulduysa, yazdıklarımı okuyup eleştirdi. Oğlum Cemre, ne zaman omzumun üstünden baksa bir yanlış gördü. Kimi verilerin derlenmesinde, kimi çizelgelerin düzenlenmesinde el verdi, bilgisayar donanım sorunlarımın çözülmesinde destek oldu. İkisi de sağ olsun.

Literatür Yayınevi genel müdürü Kenan Kocatürk, kitabın çok kısa sürede basılmasını sağladı. Hem ona, hem yayınevi çalışanlarına teşekkür ederim.

Bütün bu çabalara karşın yine de bulunabilecek yanlışların ya da eksikliklerin sorumluluğunu tek başıma üstlenmek zorundayım.

Ümit Şenesen
Ekim 2004

İstatistik Nedir? Neye Yarar?

Gün gelecek, istatistik, etkin vatandaşlık için okuma yazma kadar gerekli olacak.

Herbert G. Wells

1.1 GİRİŞ

H. G. Wells'in sözünü ettiği gün çoktan geldi. Çevrenize şöyle bir bakıp konuşulanlara bir kulak kabartsanıza. Gazetelerde, dergilerde, radyolarda, televizyonlarda durmadan istatistiklerden söz ediliyor. Her aybaşında enflasyon oranları, her gün İMKB indeksleri, borsa kâğıtlarının değerleri, yatırım fonları, zaman zaman ekonominin büyüme oranı, bütçe açıkları, dış ticaret, dış borçlar, dolar kuru, gelir dağılımı, gecekondulaşma hızı, başa güren gazetelerin günlük satış sayıları, televizyon kanallarının izlenme oranları, yayın kuruluşlarının reklam pastasından aldıkları paylar, seçim tahminleri, kamuoyu yoklamaları, araba fiyatları, özel okul fiyatları, asgari ücret, liselere ya da üniversitelere giriş puanları, suç oranları, doktor başına düşen kişi sayısı, kişi başına düşen gelir, hepimizin başına düşen zamlar sözlü-yazılı haberleşme araçlarının vazgeçemedikleri haberler arasında. Buna çeşitli spor karşılaşmalarına ilişkin puan cetvellerini, gol kralıklarını, her türlü sayısal veriyi ekleyin. Her an istatistiğin bombardımanı altındayız.

Bu durumun ilk ayırđına varan uluslardan biri Japonlar olmuş, 18 Ekim'i Ulusal İstatistik Günü olarak çoktan seçmişler bile (Moore, 1979:xi). Bu ülkede ilköğretimin ilk sınıflarındaki öğrenciler için açılan istatistik çizimleri yarışmasına her yıl on binlerce öğrenci katılıyor, yedi yaşındaki çocuklar ödüller kazanıyor (Tuft, 1983:84). Devlet İstatistik Enstitüsü (DİE) de belki bundan esinlenerek her 9 Mayıs'ta "Türkiye İstatistik Günü" kutlamalarını başlatmış bulunuyor.

İstatistik, sayılara en uzak görünebilecek alanlara bile el atmaktan çekinmiyor. Artık yalnızca basketbolde değil, sporun bir çok alanında istatistikten yararlanılıyor, giderek spor yazınında istatistik başlıklara bile çıkıyor (Bora, 2004; Eraybar, 2004). İstatistiğin girmedığı bir alan gösterebilir

misiniz? Güzel sanatlar mı dediniz? Alın size bir örnek. Sezgin (1993) istatistiği edebiyata uygulamış, “Türkçe metinlerde kelime uzunlukları”, “Türkçe metinlerin ölçülebilir özellikleri”, “Türk romanında batı kaynaklı kelimeler” konularında sayısal çalışmalar yapmıştır. Başka ülkelerde de Shakespeare’in kaç sözcük bildiği (Efron, Thisted, 1976), ABD’deki *Federalist Yazılar*’ı kimin yazdığı (Mosteller, Wallace, 1972) istatistikçilerce incelenmiştir. Son günlerde yine istatistiğe dayanarak Molière’in kimi oyunlarının asıl yazarının Corneille olduğu ileri sürülmektedir (Üster, 2003). Bu kitabın yazarı bile asıl çalışma alanı dışına çıkıp, içinde istatistik kullanılan müzikle ilgili bir yüksek lisans tezinde (Başar, 1993) jüri üyeliği yapmış, sporcu öğrencilerin yeteneklerinin incelendiği bir araştırmada (Aköz vd. 1994) görev almış, ayrıca kısa-orta mesafe koşularındaki olimpiyat derecelerinin zaman içindeki gelişmelerini incelemiştir (Şenesen, 1995).

Örnek olmak üzere Devore, Peck (2001), Mosteller vd. (1973), Fisher (1971), Hoaglin vd. (1983, 1985), Johnson, Kotz (1997), Marsh (1988), McNeill (1977), Salzburg (2001), Siegel (1997), Tanur vd. (1973, 1989), Tuft (1983, 1990, 1997), Tukey (1977) gibi bir kaç kaynağa göz atılırsa istatistiğin ne kadar değişik alanlarda kullanıldığına ilişkin daha iyi bir fikir edinilebilir: Adalet, araştırma-geliştirme, bankacılık, bilgisayar, borsacılık, cezaevleri, denizcilik, deprem, dış ticaret, din, dokumacılık, elektrik, el yazısı, emlakçılık, enerji, finansman, fiyatlar, gazetecilik, gelir dağılımı, hastabakıcılık, havacılık, hayvancılık, iklim, imalat, inşaat, istihdam, işsizlik, kalite denetimi, kentler, kumar, madencilik, mekanik, muhasebe, müzik, nüfus sayımı, otomotiv, para basma, pazarlama, polis, reklamcılık, resim, sağlık, savunma, seçim, sigortacılık, silahlar, siyaset, spor, şans oyunları, şarapçılık, tarım, teknoloji, terör, ticaret, trafik kazaları, ulaştırma, ulusal gelir, uzay mekiği kazaları, ücretler, üretim, vergiler, yerölçüm, yöneticilik. Daha sayalım mı?

Bugün dünya üniversitelerinde bitkibilim, budunbetim (kavmiyat, etnografya), budunbilim (ırkiyat, etnoloji), coğrafya, çevrebilim, davranış bilimleri, denizbilim, dilbilim, dirimbilim, edebiyat, eğitim, fizik, felsefe, genbilim, gökbilim, hayvanbilim, hukuk, iktisat, insanbilim, işbilim (ergonomi), işletmecilik, kimya, madenbilim, metalurji, meteoroloji, mühendislik, nüfusbilim, nükleer fizik, ruhbilim, tarih, tıp, toplumbilim, uzaybilim, yerbilim gibi alanlarda istatistik öğretilmektedir. Hatta bazı kitaplar bu öğretimi çekici kılmak için karikatür, şakacı yazım biçimi gibi ilginçliklere başvurmaktadır (Diamantopoulos, Schlegelmilch, 1997, Gick, Smith, 1993).

Üstelik burada sözü edilen alanlar yalnızca bazı örneklerdir, istatistik bunlar dışında başka alanlarda da uygulanmaktadır.

1.2 İSTATİSTİK NEDİR?

İstatistik, sağduyunun inceltilmiştir.

İstatistik sözcüğünün yukarıda değişik anlamlarda kullanıldığı dikkatli bir okuyucunun gözünden kaçmamıştır. Bunlardan biri sayısal verilerdir. “Ayn fiyat istatistikleri” ya da “istatistiklere göre son yirmi yılın en büyük dış ticaret açığı” dediğimizde bu anlamı vermek istiyoruz.

Bu verilerin derlenip toplanması, işlenip özetlenmesi, düzenlenip sunulması, çözümlenip sonuçlara varılması aşamalarında kullanılan kuramların, tekniklerin, yöntemlerin hepsine birden istatistik denir. “Bizim fakülte de istatistik okutuluyor” ya da “bu işyerinde istatistikle kalite denetimi yapılıyor” sözleri de bu ikinci anlamı taşımaktadır. Bu kitapta istatistik genellikle yöntem anlamında kullanılacaktır.

Aslında istatistiğin yalnız istatistik bilenlerce kullanılan üçüncü bir anlamı daha vardır ama buna ileride yeri geldiğinde değinilecektir.

İstatistiğin gerçekle ilişkisi olmadığı, zaman zaman da şakaya getirilerek bir tür yalan söyleme sanatı olduğu konusundaki kanılar kesinlikle haklıdır. İstatistiği, sayıların arkasında yatan gerçekleri anlamak, bunlardan işe yarar sonuçlar çıkarmakta kullanılan bir araç olarak düşünmek gerekir. Her araç iyiye de kullanılabilir, kötüye de. Elektrikle ister aydınlatma yaparsınız, ister işkence! Seçim sizin. İstatistikle doğruları gösterebileceğiniz gibi insanları aldatabilirsiniz de. Ama biraz istatistik öğrenirseniz başkalarının sizi aldatmaları pek kolay olmaz.

İş yaşamında pek çok bilgiye gerek duyarız, bunların çoğu sayısal veri biçimindedir. Bu sayılar denizinde boğulmadan yüzebilmek için bu verileri işleyip anlamlı olgulara, düşüncelere dönüştürmemiz, sonra da bunlardan kullanılabilir bilgiler türetmemiz gerekir. İşte istatistik bizi bu işleri yapacak yöntemlerle donatır.

İstatistik yöntemleri belirsiz durumlarda nasıl karar verilebileceğini gösterir. “İMKB bileşik indeksi önümüzdeki günlerde yükselecektir” ya da “bu fiyatı verirsek ihaleyi kazanamayız” gibi sözler aslında “İMKB indeksinin yükselmesi beklenmektedir”, “ihaleyi kazanma şansımız düşüktür” anlamlarına gelir. Bu beklenti ne kadar yüksek, bu şans ne kadar düşüktür? İstatistik bu belirsizliği sayılara dönüştürme yöntemlerini de içerir.

Bir çok iktisat yorumcusundan enflasyonun nedenleri konusunda çeşitli açıklamalar duymuş ya da okumuşsunuzdur. Faiz oranlarının yükselmesi ya da düşmesi yatırımlar üzerinde nasıl bir etki yaratır? Şirketlerin kâr payı dağıtımları pay senetlerinin piyasa değerini artırır mı? Parti başkanlarının TV tartışmasına katılmaları seçim sonuçlarına nasıl yansır? Bu sorular birden çok değişken arasındaki ilişkilerle ilgilidir. Bu ilişkinin gücü, yönü,

yarattığı etkinin büyüklüğü ne kadardır? İstatistik bu sorulara, veriler elverdiği ölçüde, yanıtlar bulmaya çalışır. •

Aldığımız kararların önemli bir bölümü geleceğe, bazıları uzak geleceğe ilişkin sonuçlar doğurur, dolayısıyla hayli erken alınmaları gerekir. Bu da bazı tahminlerin yapılmasını gerektirir. İstatistik burada da yardıma koşar, geçmiş olaylardaki, davranışlardaki düzenliliklere ya da değişkenler arasındaki ilişkilerin yapısına dayanarak gelecekteki gelişmelerin olası eğilimlerini kestirmede işe yarar.

Bütün bu işler sırasında kullanılan veriler çoğu zaman kesinlikten uzaktır. Yararlı bilgiler içerirler ama her şeyi açıklamazlar. Şu eşitlik hemen her zaman geçerlidir:

$$\text{Veri} = \text{Gerçeklik} + \text{rassallık}$$

Burada veri elimizde olandır, gördüğümüz, ölçtüğümüz, saydığımız şeydir. Gerçeklik ise öğrenmek, bilmek istediğimizdir. Rassallığa gelince o da gerçekliğe tam olarak ulaşmamızı engelleyen, araya giren bozukluklar, bilinmeyenler, eksiklikler, hatalar, yanlışlıklardır. İstatistik rassallığın boyutlarıyla ilgilenir, rassallığı olabildiğince azaltıp gerçekliği kavramaya çalışır.

Bir örnek olarak depremi ele alalım. Yerbilimciler ne zaman, nerede, ne büyüklükte deprem olacağını bilmeye çalışırlar. Bu çok karmaşık bir sorundur. Depremlere ilişkin sürekli veri toplanmasını gerektirir. Bu verilerin içerdiği rassallıktan kurtulmak zor bir iştir. Verilerin çözümlemesi istatistik bilgisinin yanı sıra yerbilimi uzmanlığı da ister. Bu iki bilim dalı işbirliği yaparak rassallığı azaltmaya, bütün insanlığı ilgilendiren yaşamsal sorulara güvenilir yanıtlar bulmaya çalışır.

Aslında her bilim dalı istatistikle benzer yardımlaşmalara girer, birbirlerinden karşılıklı yararlanırlar. Bu da bilimdeki bilinmezlerin sayısını azaltmada önemli bir işlev görür.

İstatistik bir yandan resmin bütününe görmeye, göstermeye uğraşır, öte yandan tekil gözlemleri de göz ardı etmez (Siegel, 1997:4). Tek tek gözlemler birleştirilerek istatistikle bütüncül bir bilgiye ulaşılabilir. Ama veriler arasında bütüne aykırı duran, özel konumda olanlar varsa bunların ne anlama geldiği de çok önemlidir.

Kalabalık bir veri kümesine siz baktığınızda ne görürsünüz, bir uzman baktığında ne görür? Size şaşırtıcı gelebilir ama ikinizin gördükleri arasında çoğu zaman pek fark yoktur. Çünkü fazla bir şey görünmez. Bu sayılardan anlam çıkarabilmek için biraz uğraşmak gerekir. Dolayısıyla sayılardan oluşan bir çizelgeye baktığınızda bir şey anlamıyorsanız kendinizi bozguna uğramış gibi görmeyin. Biraz çaba gösterip istatistik öğrendiğinizde siz de bu verilerden anlamlar çıkarabileceksiniz.

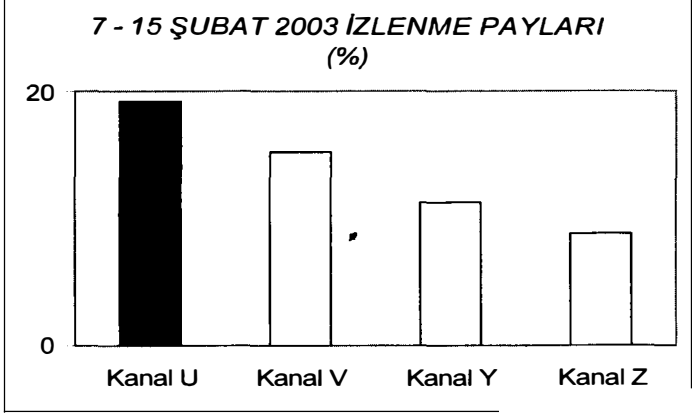
1.3 İSTATİSTİK NEYE YARAR?

*İstatistik dersi grip aşısından daha sev.mli
değildir ama en az o kadar yararlıdır.*

Paul Newbold

Sıradan vatandaş olarak istatistik öğrenmenin bir yararı, çevreden gelen veri bombardımanına karşı hazırlıklı olmaktır. Sayıların desteğinde ileri sürülen düşüncelerin doğru olup olmadığını başka türlü nasıl değerlendirebiliriz ki? İşte size gazetelerden bazı örnekler:

- Ekonomiden sorumlu bir devlet bakanı diyor ki: *“Türkiye’de aylık enflasyon Mart’ta % 7 oranına çıktı. Şimdi bakın son aylara, Nisan % 4,8, Mayıs % 2,1, Haziran % 2,6, Temmuz % 2,2. Bu dönemlere bakarsanız enflasyon % 7’lerden % 2-2,5 seviyelerine inmiştir.”* (Cumhuriyet, 14 Ağustos 1988) Ülkemizde enflasyon yaz aylarında genellikle yavaşlar. Bakanın vurguladığı yavaşlamanın ne kadarı mevsim etkilerine, ne kadarı hükümetin önlemlerine bağlanabilir? Aslında enflasyonda bir yavaşlama olmadığı halde bazı düşüşler sırf mevsim nedeniyle mi ortaya çıkmıştır? Bu soruları belli güven düzeylerinde yanıtlayabilirsek daha bilinçli bir vatandaş olabiliriz.
- *“Mevlevilerin dönme hareketi hem ruhu, hem bedeni arındırıyor. (Bir araştırmaya göre) hiçbir semazende yüksek tansiyon rahatsızlığına rastlanmadı. Ayrıca semazenlerin kanındaki yağ oranlarının ve kolesterollerinin de hep normal düzeyde olduğu belirlendi. ... Bale ve buz pateni (ile) uğraşan insanların sağlığında da benzer olumlu etkiler (saptandı).”* (Radikal, 18 Şubat 2003) Bu hastalıklardan kurtulmak için semazen (balerin, buz patencisi) mi olmalı? Yoksa semazen (balerin, buz patencisi) olabilmek için zaten sağlıklı bir yapıya sahip olmak mı gerekir? İstatistikle araştırma yöntemlerini öğrenirsek bu sorulara daha “sağlıklı” yanıtlar verebiliriz.
- 24 Ocak 1980 kararlarını değerlendiren şu iki demece bakalım. (1990’a girerken Türk Ekonomisi, Sabah gazetesi eki, 24 Ocak 1990) *“24 Ocak istikrar sağladı”* (A. Tigrel, DPT Müsteşarı). *“Hükümet 10 yıl işçinin ezilmesini tercih etti”* (Ş. Yılmaz, Türk-İş Genel Başkanı). Verileri incelemesini bilirsek, bu demeçlerin etkisinde kalmadan kendi görüşümüzü daha kolay oluşturabiliriz.
- Bir reklamda kullanılan Çizim 1.1’e bakarak dört TV kanalının izlenme payları konusunda doğru bilgi edinebiliyor muyuz? Yoksa çizimler bize gerçeği biraz da olsa çarpıtarak mı yansıtıyor?



ÇİZİM 1.1 TV kanallarının izlenme payları (Radikal, 17 Şubat 2003)

Bunu anlayabilmek için her kanalın izlenme paylarını bilmek gerekiyor. Bunlar sırasıyla % 18,9, 18,1, 13,1, 12,4. Gazetede verilen değerlerle çizimdeki sütunların boyları birbirini tutmuyor. Ayrıca gazetedeki çizimde sol eksene hiçbir değer konmamış. Çizim burada yeniden düzenlenirken hem % 20 değeri eklendi, hem kanal adları gizlendi. Sizce gazetedeki bu reklamı U, V, Y, Z kanallarından hangisi vermiş olabilir?

İşte istatistik kullanan bu gibi yayınlardan doğru bilgiler alabilmek için istatistik bilmek gerekiyor. İstatistik bilgisi bizi böyle durumlarda yanıltmaktan korumakla kalmaz, sunulan çizelgelerle çizimlerden doğru sonuçlara ulaşmamızı, başka bir deyişle sayıların arkasında yatanları da görmemizi sağlar.

İstatistik **karar vermede** de çok işe yarar. İstatistik öğrenmenin bir başka nedeni de budur. Hangi alanda çalışıyor olursak olalım, bize sunulan istatistik verileri anlamamız, belirsizliklerle dolu yaşamda kararlarımızı onlara dayandırmamız gerekir. İstatistik bu kararların doğruya en yakın biçimde verilmesinin yollarını gösterir.

1.4 İSTATİSTİK NE GİBİ İŞLER GÖRÜR?

*İstatistikçinin bir işi de sayılara bakıp
neden öyle göründüklerini sorgulamaktır.*

Stella Cunliffe

İstatistik yöntemlerin kullanılışını anlayabilmek için önce “istatistik” olarak adlandırılan çeşitli işler şöyle sıralanabilir:

- Veri derlemek, gerekirse deney tasarlamak,
- Verileri sorgulamak,
- Bir büyüklüğü kestirmek (tahmin etmek),
- Bir önsavı sınamak (hipotezi test etmek), karar vermek.

Veri derleme, deney tasarımı

İstatistik işlemler için her şeyden önce veri bulmak gereklidir. Veriler ilgili konuya bağlı olarak ya onları yayımlayan Devlet İstatistik Enstitüsü (DİE) gibi kurum ya da kişilerden hazır olarak alınır, ya kamuoyu yoklamalarında olduğu gibi alandan derlenir, yahut da deneyler düzenlenerek elde edilir.

Aradığımız veriler geçmişle ilişkiliyse ilk yola başvurmaktan başka çare yoktur.

Araştırma konusu, birimlerin gözlenmesiyle elde edilecek verilere bağlıysa kullanılacak yöntem **örneklemedir**.

Sözelimi yeni bir gübrenin ürün verimliliğini artırıp artırmadığını anlamak istiyorsak deney yapılmalıdır. Deney tasarımı, toplanacak verilerin işe yarar olabilmesi için nasıl derlenmeleri gerektiğini gösterir. Deneye girişmeden önce tasarımı üzerinde kafa yormak genellikle meyvelerini verir. İyi tasarlanmış bir deneyin bulgularını çözümlemek kolaylaşır, çoğu zaman sonuçlar hemen göze çarpar. Buna karşılık deney kötü tasarlanmışsa elde edilen veriler başlangıçta sorulan sorulara yanıt veremez.

Bir deneyde araştırmacı, bir ya da bir kaç girdinin, öteki adıyla **bağımsız değişkenin** çeşitli değerlerini seçerek çıktının ya da **bağımlı değişkenin** değerlerine ne olduğunu gözler. Amaç, girdi ile çıktı arasındaki ilişkiyi anlamak, bunun gerisinde yatan süreci belirginleştirmeye çalışmaktır. Öyleyse bir deney tasarımının özü şu üç soruya yanıt bulmaktır: Hangi etmenleri ele almalıyız? Bunların düzeyleri nasıl değişmelidir? Bu düzeyler bir arada nasıl kullanılmalıdır? Bu konuya ileride yeniden dönecektir.

Sorgulayıcı veri çözümlemesi

Sorgulayıcı veri çözümlemesi derlenen verilere alıcı gözle bakmak gibidir. Tukey kurucusu olduğu bu alanı “sayısal dedektiflik işi” diye yorumlar. “Bir suçu araştıran dedektif hem bazı araçlara gerek duyar, hem olayı kavramak zorundadır. Yanında parmak izi tozu yoksa çoğu zaman parmak izlerini bulamaz. Ama suçlunun en çok nerelerde parmak izi bırakabileceğini bilmiyorsa bu kez de doğru yerlere bakmayı beceremez. Veri çözümleyen bir kimse de tıpkı bunun gibi hem araçlara gerek duyar, hem anlamak ister” (Tukey, 1977:1).

Sorgulayıcı veri çözümlemesi, çıkarımcı istatistikten farklı olarak verileri sorguya çeker, onlarla konuşur, ama daha önemlisi verileri dinler, ondan sonra ne yapılacağına karar verir (Marsh, 1988:xviii).

Bir veri kümesinin başlıca özellikleri şunlardır: Ortalama düzey, ortalama-
madan uzaklaşma derecesi, tepe sayısı, çarpıklık, sivrilik-basıklık, sıradışı gözlemler. Bu özellikleri çok yalın, çok kolay yollarla ama kaba olarak görmek için sorgulayıcı veri çözümlemesi kullanılır. Bu yöntemlerle birden çok veri kümesinin benzer özellikleri hemen karşılaştırılabilir. Ayrıca iki değişken arasındaki ilişkinin özellikleri de sorgulayıcı yöntemlerle kolayca ortaya çıkarılabilir. Böylece verilerin gerisinde yatan yapı anlaşılabilir. Veriler gerçekliği açıklamada ilginç bir yapıya sahipse ama bu yapı, pek de o kadar ilginç olmayan rassallıklarla perdeleniyorsa, sorgulayıcı yöntemler, görüntüyü bulanıklaştıran rassallık öğelerini ayıklayarak verilerin asıl önemli olan bileşenini gözler önüne serebilir.

Bu yöntemlerin çabuk anlaşılır, kolay uygulanabilir olmalarının bir nedeni, çoğu zaman dört işlemden öteye bir matematik bilgisi gerektirmemeleridir. Ayrıca yalın bazı çizimlerden de yararlanırlar. Bugün giriş düzeyindeki ders kitaplarında bile yer alan **dal-yaprak**, **kutu çizimi**, **dönüştürme merdiveni** gibi kavramlar çok işlevseldir.

Bir araştırmacı, toplanmış olan verileri kullanmadan önce iyice tanımak zorundadır (Hartwig, Dearing, 1979:5). Özellikle neyi, nerede arayacağını tam olarak bilemiyorsa sorgulayıcı veri çözümlemesi daha da önem kazanır, araştırmacıya çok değerli ipuçları sunabilir. Bu nedenle adı “verilerin ön çözümlemesi” olarak da anılır (Chatfields, 1986). İleride sırası geldikçe sorgulayıcı veri çözümlemesi yöntemleri tanıtılacaktır.

Kestirme

İstatistikte **kestirim**, elimizde bilinmeyen bir büyüklüğe ilişkin gözlenmiş veriler varken o büyüklüğü tahmin etmektir. Sözelimi LES (lisanüstü eğitim sınavı) sonuçları öğrencilerin yeteneklerini, fiyat indeksleri enflasyonu,

seçim araştırma sonuçları oy dağılımını, fosil kayıtları dünyanın yaşını (Siegel, 1988:7) kestirmeye çalışır.

Kimi kestirimler tepki toplar, kimileri üzerinde uzlaşılır. Aslında, aradaki fark hata diye tanımlandıktan sonra her şey her şeyin kestirimi olabilir. Önemli olan bu hatanın olabildiğince küçük kalmasıdır. Bunu sağlamak için çeşitli kestirim yolları, başka bir deyişle **kestirimciler** kullanılır. Bu kestirimcilerin yansızlık, etkinlik gibi özellikleri olması beklenir. Bu konuya ileride değinilecektir.

Kuşkusuz bütün kestirimler az ya da çok hata içerir. İstatistiğin buradaki katkısı, yapılan bir kestirimde bizi kör karanlıktan çıkarıp hatanın yaklaşık büyüklüğü konusunda bir fikir vermesidir (Siegel, 1997:7).

Sinama, karar verme

Keynes'in 1930'larda yaptığı gibi "toplam tüketim, toplam gelire bağlıdır" diye kuramlar ortaya atılabilir, "geliştirdiğimiz yeni yöntem üretimde verimliliği artırır" türünden savlar ileri sürülebilir. Bunların geçerliliğinin sınanması gerekir. Yeni bir ilaca üretim izni verilirken hastalığa iyi gelmesi, yan etkilerinin olmaması gibi bir dizi sınamayı geçmesi beklenir. Aksi halde "ilaç" sayılmaz. İşte bu nedenle 1980'lerin sonlarında bir hekimimizin zakkumdan elde ettiği bir özütle kanseri yendiği yolundaki savı tıp çevrelerini birbirine kattı. Tabipler Odası bu hekimin bir süre mesleğini yapmasını yasakladı. İş mahkemeye yansıdı. Bu hekim Tabipler Odası'nın haksızlığına mı uğramıştı, yoksa bir şarlatan mıydı (Batuhan, 1993:373-416)? Tabipler Odası'nı harekete geçiren, gerekli sınamalar yapılmadan bir özüte "ilaç" denmesi, bunun insanlar üzerinde denenmesiydi.

İstatistikte **önsav sınaması** (hipotez testi) denilen bu işlemde önce bir **sıfır önsavı** tanımlanır. Buna göre ortada ilginç bir durum yoktur, gözlenen fark rassallıktan kaynaklanmaktadır. **Karşı önsav** ise görünen ilginçliğin rassallığın ötesinde bir nedene dayandığını ileri sürer. Sınama sonucunda bu önsavlardan hangisinin ret, hangisinin kabul edileceğine karar verilir. Elbette bu karar hatalı olabilir. Ama kullanılan istatistik yöntemleri, bu hatayı işleme olasılığının büyüklüğü konusunda bize bir fikir verir.

Önsav sınaması yapılırken bazı önyargıların, yanlış anlamaların işe karışmamasına özen gösterilmelidir. Sınamayı yapanlar tarafsız mıdır, yoksa sonucun o yönde çıkmasında bir çıkarları var mıdır? Sigara ile kanser arasında bir ilişki olup olmadığı sınanırken sigara şirketlerinin desteklediği araştırmaları mı, yoksa tarafsız kamu kuruluşlarının yaptırdıklarına mı güvenirsiniz? Ya bir siyasi partinin yapılmasına para ödediği bir kamuoyu araştırmasının sonuçlarına ne dersiniz?

1.5 İSTATİSTİĞİN KÖTÜYE KULLANIMI

*Üç türlü yalan vardır: Sıradan yalan,
kuyruklu yalan, bir de istatistik.*

Benjamin Disraeli

Her bilim dalı gibi istatistik de, bilerek ya da bilmeyerek, kötüye kullanılabilir. Verileri başkalarını yanıltmak amacıyla kullanmak başışlanamaz bir davranış biçimidir. Hiçbir zaman böyle bir yola başvurulmamalıdır. Ne yazık ki bu görüş bir dilek olmaktan öteye gidememektedir. Serdaoğlu'na (2001) göre "PKK teröründen bor işletmelerinin değerine, SSK verilerinden hayali ihracata kadar her alanda sayısız gayrı ciddi rakam ortaya atılıyor."

İstatistiğe güvenme konusunda insanlar şu iki görüş arasında bir yerdedir: Bir ucta sayılara değişmez doğruları yansıtan bilgiler olarak yürekte inanmak, öbür ucta istatistikle her şeyin kanıtlanabileceğini, dolayısıyla da hiçbir şeyin kanıtlanamayacağını düşünmek. Her iki görüş de bizi gerçekten uzaklaştırır.

Başlıca kötüye kullanma biçimleri

- konuyu bilmemek,
- hatalı ham verilerle iş görmek,
- araştırmayı yanlış tasarlamak,
- sonuçları yanlış ya da yanıltıcı yorumlamak,
- yanlış ya da yetersiz çözümleme yöntemleri kullanmak,
- bulguları çarpıtmak olarak sayılabilir (Jaffe, Spirer, 1987:12).

Bunlardan biri bile geçerliyse kötüye kullanmadan, birden çoğu geçerliye katmerli kötüye kullanmadan söz edilebilir.

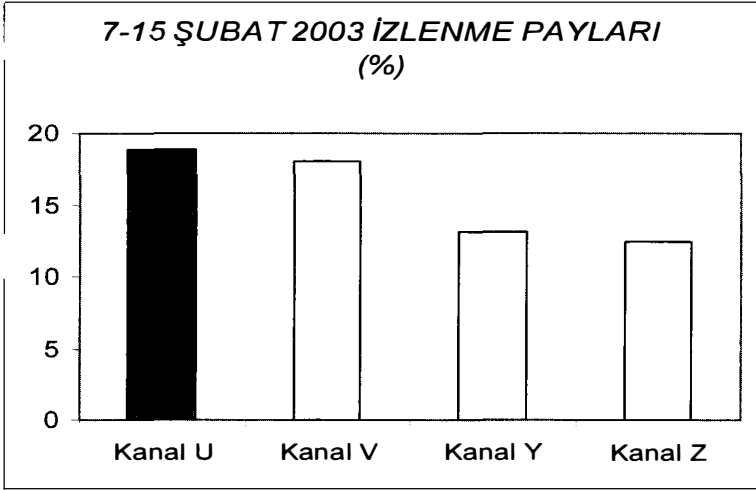
DİE'ye (1993:287) göre 1988'de kesilen hayvan sayısı toplamı 15 milyonun, et üretimi de 550 bin tonun üzerindeyken, 1989'da bu sayılar hayvan sayısında 14 milyonun, et üretiminde 510 bin tonun altına düşmüştür. Konuyu bilmeyen biri bu azalma neden olmuş diye düşünebilir. Oysa bir azalma yoktur. DİE bu arada tanım değiştirmiş, "mezbahalarla kombinalarda" kesilen hayvan sayısına 1988'de % 10 kaçak kesim eklerken 1989'da bunu kapsam dışı bırakmıştır.

Bir elektrik mühendisinin yaptığı yanlış tasarım uygulamada çok çabuk ortaya çıkar, ama bir istatistikçinin yanlış tasarımının sonuçları hemen görülemeyebilir. İstatistiğin kötüye kullanımı öyle yanlış kararlara yol açabilir ki bunun anlaşılması hem zaman alabilir, hem de hiç düzeltilemeyecek haksızlıklar doğabilir.

Bunun en çarpıcı örneği 1970'te ABD'de Vietnam savaşına gidecek askerlerin seçimi sırasında yaşanmıştır (Moore, 1979:33-34, Sandy, 1990:223-224). TV'de yayımlanan bir seçme işlemiyle 1952 doğumlu gençler arasından savaşa gidecek olanlar **örnekleme** yoluyla belirlenmiştir. Ocak ayının her günü için birer top bir kutuya konup karıştırılmış, bunlara Şubat'ın günleri eklenerek bir daha karıştırılmış, yıl sonuna kadar her ay için bu işlem yeniden yapılmıştır. Daha sonra çekilen toplar hangi günleri gösteriyorsa 1952'nin o günlerinde doğanlar askere alınmıştır. Hemen anlaşılacağı gibi Aralık topları öbür ayların toplarıyla bir, Kasım topları iki, ... Şubat, Ocak topları on birer kez karıştırılmıştır. Dolayısıyla son aylara ilişkin toplar daha üstlerde kalmış, seçimler de üste yakın toplardan yapıldığı için yılın son aylarında doğanlardan çok fazla kişi seçilirken ilk aylardan pek azı askere alınmıştır. yanlışlık anlaşıldığında iş isten geçmiş, seçilenler Vietnam'a gönderilmiştir bile. Bu örnekte amaçlanmayan bir kötüye kullanma söz konusudur. Ertesi yıl örneklem seçme işi istatistikçilere bırakarak seçimin rassal olması sağlanmıştır.

Huff (1973:14) şöyle bir örnekten söz eder. Bir gazete haberine göre *“son on yılda dört milyon dolayında Amerikalı katolik, protestan olmuştur”*. Bilginin kaynağı olan protestan bir din adamı, 25 bin protestan papazına gönderdiği bir soru kâğıdıyla kendi bölgelerinde protestanlığa geçen katolik sayısını sormuş, aldığı 2.219 yanıtın elde ettiği 51.361 sayısını ülke çapında 181 bin dolayındaki katolik din adamına genelleyerek 4.144.366 sayısına ulaşmıştır. Oysa burada bilinmesi gereken % 90'dan fazla din adamının neden yanıt vermediğidir. Bunların büyük bir bölümü kendi bölgelerinde böyle bir din değiştirme olayına rastlamadığı için yanıt vermemiş olamaz mı? Akla yakın görünen bu varsayım kullanılır da, bu soru formunun 25 bin yerine 181 bin din adamına gönderildiği düşünülürse toplam din değiştirme sayısının 370 bin dolayında olacağı kestirilebilir. Yetersiz veri çözümleme yöntemi kullanılarak bulguların yanlış yorumlanması böylesine büyük bir hataya yol açmıştır.

Kötüye kullanım her zaman istenmeden yapılmaz; bazen bile bile, isteyerek de yapılır. Çizim 1.1'e dönecek olursak burada özellikle yapılan bir çarpıtmadan söz edilebilir. Aynı verilerle düzenlenen Çizim 1.2 bu bulguyu daha doğru yansıtmaktadır.



ÇİZİM 1.2 Çizim 1.1'in verilere göre düzeltilmiş biçimi

Kanal U'nun izlenme payının öbür kanallardan yüksek olduğu gerçektir ama bu farklar ilk çizimde bir hayli abartılmıştır. Belki de bu hile açığa çıkmasın diye sol ekseninde değerler yazılmamıştır. Çizimlerdeki çubukların boylarını ölçersek, V ile U'nun izlenme payları arasındaki fark aslında % 4'ten azken, Çizim 1.1'de % 20'den büyükmüş gibi çizildiğini görürüz. U ile Z kanalları için bile bu fark çok abartılmıştır. Gerçekte Z, U'nun yaklaşık 2/3'ü kadarken ilk çizimde yarıdan az gibi gösterilmiştir. Bu bağışlanamaz bir davranıştır. İstatistik kötüye kullanılarak bir kanalın öbürlerine göre ezici bir üstünlük taşıdığı izlenimi uyandırılmaya çalışılmıştır.

Bazen gereken yerde istatistik kullanılmaması da kötüye kullanım gibi istenmeyen sonuçlara yol açabilir. Bunun tüyler ürperten bir örneği 1986'da Challenger uzay mekiğinin fırlatıldıktan 73 saniye sonra havada patlamasıdır. Kazayı araştıran kurul, “eldeki veriler bir istatistikçi tarafından incelenmiş olsaydı, kalkışa izin verilmeyeceği” sonucuna varmıştır. (Dalal vö. 1989; Tufte, 1997:18-53).

1.6 BİLGİSAYARLA İSTATİSTİK

Alet işler, el övünür.

Atasözü

Bilgisayarların belleklerinin büyümesi çok büyük boyutlardaki veri kümelerinin saklanması, çalışmalarının hızlanması ise bu verilerin çözülmesini kolaylaştırmıştır. Bu sayede hayli karmaşık modeller uygulamaya girebilmiştir. Ayrıca veri derleyen resmi kurumlar veri tabanlarının hiç olmazsa bir bölümünü kamunun ücretsiz kullanımına açmıştır. Bu kurumlardan satın alınabilen verilerin eskiden olduğu gibi basılı biçimde değil de elektronik ortamda elde edilebilmesi, kullanıcıyı verileri yeniden bilgisayara girme yükünden kurtarmıştır. Her gün kullanışlı bir sürü yazılım geliştirilip hem istatistikçilerin, hem istatistik öğrencilerinin yararlanmalarına sunulmaktadır.

Kimi yabancı ders kitapları bu yazılımları kullanarak istatistik öğretmeye girişmektedir. Sözgelimi Becker (1997) ile Neufeld (1997) Excel yazılımını, Monk (1991) MINITAB'ı, Lee (2000) ikisini birden, Hamilton (2002) STATA'yı temel almıştır. Bunun yanı sıra istatistik kitaplarıyla birlikte okuyucuya en azından alıştırma verilerini içeren bir "disket" ya da "CD" verilmektedir (sözgelimi Lind vd. (2002)). Bu kervana, yabancı dilde kitap yayımlayan yerli yayınevlerinden de katılanlar olmuştur (Özdemir, 2001). Hatta bazı yayınevleri buna yazılım bile eklemektedir. Bu yazılımlar "Phstat" (McClave vd. (2001)), "PHStat2", "TreePlan" (Newbold vd. (2003)) gibi adlar almakta, Excel'in olanaklarını geliştirerek ders kitabında geçen istatistik tekniklerinin bilgisayarda uygulanmasını öngörmektedir. İstatistiği alıştırmalarda bilgisayar kullanarak öğretmeyi amaçlayan bir örnek de VISUAL STATISTICS'tir (Doane vd. 2001). Burada veri kümelerini istediğiniz özelliklere uygun olarak değiştirip çeşitli yöntemleri yeni verilerle uygulama olanağı sağlanmaktadır. Bir başka örnekte (Aczel, Sounderpandan, 2002) yine Excel'den yararlanılmakta, çeşitli istatistik teknikleri için özel kutucuklar hazırlanıp kullanıcının verilerini buralara yazması istenmekte, bu işlem tamamlanınca hesaplama sonucu ortaya çıkmaktadır. Bu yazılımları kullanmak çok kolaydır, Excel'le tanışık olan herkes bunu hemen yapabilir.

MINITAB, SPSS gibi yazılımların öğrenci sürümleri vardır. Bu yazılımlardan en az birini kullanmayı öğrenmek öğrenciye büyük kolaylıklar sağlayacaktır. Bu konuda yardımcı olacak Türkçe kitaplar da vardır (sözgelimi Özdamar, 1999). Yukarıda adı geçenler dışında EVIEWS, LIMDEP, MICRO TSP, PcGive, SAS, SHAZAM, SYSTAT adlı yazılımlar da kullanılmaktadır.

Bilgisayarlar, çözümleri kolaylaştırmanın yanı sıra verilerin ön inceleme-
mesi, çizelgelerin, çizimlerin düzenlenmesi aşamalarında da ilerlemelere yol
açmıştır. Ancak çizimler konusunda dikkatli olmak gerekir. İleride yeri gel-
diğinde gösterileceği gibi bilgisayarların hazır çizim yazılımları, bilgi iletme
bakımından çok yetersiz, hatta yanıltıcı çizimler bile üretebilmektedir.

Okuma önerileri

Batuhan, H. (1993) *Bilim ve Şarlatanlık*, Yapı Kredi.

Huff, D. (1973) *How to Lie with Statistics*, Penguin (Türkçesi:

Koparan, E. (1995) *İstatistik ile Nasıl Yalan Söylenir?*, Sarmal).

Jaffe, A. J., Spierer, H. F. (1987) *Misused Statistics*, Marcel Dekker.

Salzburg, D. (2001) *The Lady Tasting Tea – How Statistics Revolutionized Science
in the Twentieth Century*, Freeman.

Sezgin, F. (1993) *Dil ve Edebiyatta İstatistik ve Bilgisayar Uygulamaları*, Dergâh.

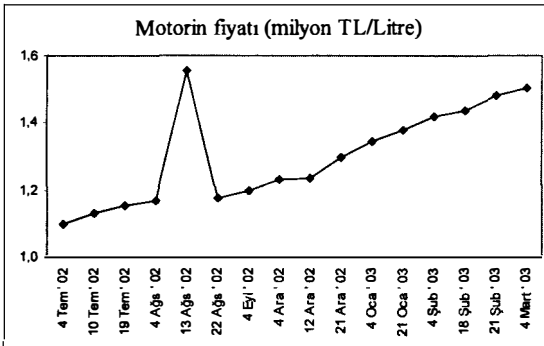
Tanur, J. M., Mosteller, F., Kruskal, W. H., Lehman, E. L., Link, R. F., Pieters, R.
S., Rising, G. R. (1973, 3rd ed., 1989) *Statistics: A Guide to the Unknown*, Hol-
den Day.

ALIŞTIRMALAR

1.1 Yayın organlarında istatistik içeren haberleri inceleyiniz.

- Bunlar hangi alanlarda yoğunlaşıyor?
- Bilginize yeni bir ekleme yapıyor mu?
- Herhangi bir kötüye kullanım içeriyor mu?
- Yalnızca derlenmiş verileri mi sunuyor, yoksa yapılmış bir sınamanın,
kestirimin, çözümlemenin sonuçlarını mı yansıtıyor?

1.2 Bir gazetede çıkan aşağıdaki çizimi inceleyin.

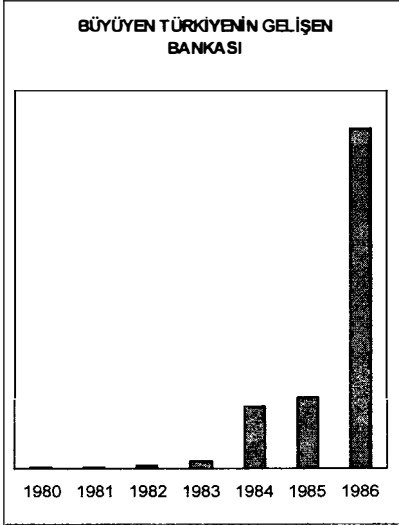


Cumhuriyet, 6 Mart 2003

- a) Bu çizim size neyi anlatıyor?
- b) Size garip gelen bir yönü var mı?
- c) 13 Ağustos 2002 tarihindeki fiyat ötekilere göre çok farklı değil mi? Bunun nedeni ne olabilir? O tarihte yapılan çok yüksek oranlı bir zam on gün sonra geri mi alınmıştır? Yoksa verilerde ya da çizimde bir yanlışlık mı var?
- d) Yatay eksenlerdeki tarihlerin aralarındaki gün sayıları eşit mi?
- e) Bunlar eşit olmadığı halde çizimde eşit gösterilmiş. Bu ne gibi bir yanlış algılamaya yol açabilir? Bu çizime bakan bir kimsede zamların hep aynı aralıklarla yapıldığı izlenimi uyanabilir mi?

1.3 Aşağıdaki çizim bir gazetede tam sayfa, renkli yayımlanan bir banka reklamından çubuk boyları arasındaki oran aynı kalmak üzere aktarılmıştır. Çubuklar bankanın yıllık kârlarını göstermektedir. Reklamda belirtildiğine göre her yılın kârı sırasıyla 1,65, 2,254, 15,33, 127,84 milyon TL, 2,324, 2,485, 13,5 milyar TL olmuştur.

- a) Bu çizimde herhangi bir kötüye kullanma söz konusu mu? Açıklayınız.
- b) Bu verilerle bir çizim düzenleyip yukardakiyle karşılaştırınız.



Cumhuriyet, 5 Şubat 1987

1.4 Bir gazetede "Akşamcılar rakıya iyiden iyiye küstü" başlığıyla şu haber çıkmıştır: "TEKEL'in resmi rakamlarına göre 1999'da 79 milyon litre olan üretim 2001'de 67 milyon litreye düştü. İşçi sayısının artmasına karşın tüketim aynı dönemde 72 milyon litreden 63 milyona indi. Üretim düştü ama TEKEL'in gelirleri arttı. 1999'da rakı satışından 185,1 trilyon lira gelir elde

~~edilirken~~ **bu rakam 2001'de 497 trilyon liraya çıktı. Yani satışlar 3 yılda yaklaşık yüzde 15'den fazla düşerken gelir yüzde 170 oranında arttı.**” (Milliyet, 15 Şubat 2003).

- a) Gazetenin verdiği değişim oranları doğru mu? Bu oranları bir kez de siz hesaplayıp karşılaştırınız.
- b) Haberde işçi sayısı arttı deniliyor. Verilen sayılardan bu sonucu çıkarabilir misiniz? İşçi sayısı artmamış olabilir mi?
- c) Üretim litre ile ölçülürken satış rakamları TL ile verilmiş. Farklı iki ölçü birimiyle ölçülen bu iki değişkendeki değişimler karşılaştırılabilir mi? Satış gelirlerinin arttığını söyleyebilmek için rakı satış fiyatındaki değişimleri bilmemiz gerekmez mi?

İstatistik Nereden Çıktı?

*Daha uzağı görebildiysem devlerin
omuzları üstünde olduğumdandır.*

Isaac Newton

Toplumsal gereksinimleri bilimin gelişmesinde itici bir güç olur, bunun karşılığında bilim de toplumu geliştirir (Bernal, 1965:54-57). İstatistik için de benzeri bir etkileşim söz konusudur. Bu etkileşimin zaman içinde geçirdiği evrelere iki açıdan bakılabilir. Birincisi sayısal veri derlemenin, ikincisi istatistik yöntemlerinin gelişme tarihi.

2.1 VERİ DERLEMENİN EVRİMİ

Sayısal veri derlemenin tarihçesini, ilk insan topluluklarına kadar geriye götürebiliriz. Yazının henüz bilinmediği dönemlerde bile avcılıkla, çobanlıkla ya da ilkel tarımla geçinen topluluklar kendilerini (kişileri, aileleri, silah taşıyabilen erkekleri) ya da mallarını (hayvanlarını, tarım ürünlerini) bazı işaretler kullanarak saymış, kayıtlara geçirmişlerdir. İlk devletlerle birlikte bu sayma işlemi dönemsel olmaya başlamıştır. Eski Mısır'da, Asur'da, İsrailoğulları'nda, Çin'de, Eski Yunan'da, Roma'da düzenli sayım yapıldığı bilinmektedir (Çömlekçi, 1985; Göçmençelebi, 1976:9-10; Gürtan, 1979:8; Kutsal, Toktamış, 1981).

Bugün Batı dillerinde nüfus sayımı anlamında kullanılan *census* sözcüğü Romalılardan kalmadır. VII. yüzyılda Charlemagne dönemi Kıta Avrupası'nda yapılan, XI. yüzyılda İngiltere'de gerçekleştirilip Doomsday Book adlı kütüklere geçirilen sayımlar, Orta Çağdaki ender sayımlardır. Bu örneklerin aslında çağdaş anlamda nüfus sayımları olmadığı, genellikle askeri amaçla yetişkin erkeklerin ya da vergi toplamaya yönelik olarak arazinin, servetin sayıma konu olduğu görülmektedir. O dönemlerin toplumsal gereksinimleri için sayımların bu biçimlerde yapılması yeterli bu-

lunmuştur. Buna aykırı ilk örnek XIII. yüzyılda Venedik'te yapılan bir sayımda yaşayan herkesin sayıma katılmasında görülür (Kendall, 1960:46).

Orta Çağdaki sayımların durgun bir toplumda yapıldığını unutmamak gerekir. O dönemde üretim bile sabit bir büyüklük olarak düşünülmektedir. Dolayısıyla da küçük değişimlerin kaydını tutmaya gerek yoktur. Bu durgunluğun ilk kırıldığı devletler ticarete, ilkel sanayide gelişmeye başlayan İtalyan kent-devletleridir. Muhasebede, matematikte çarpıcı ilerlemelerin görüldüğü XV. yüzyıl İtalya'sında ticari gereksinimleri karşılamak amacıyla büyük ölçeklerde veri derlendiği bilinmektedir. Ancak bu verilerden yalnızca o andaki durumun saptanması için yararlanılmakta, genişleyen ekonomi için kestirimlere temel olabilecekleri düşünülmemektedir.

İstatistik sözcüğünün ilk kez, Girolama Ghilini adlı bir İtalyan tarihçi tarafından kullanıldığı ileri sürülmektedir (Kendall, 1960:45; Göçmençelesi, 1962:29). XVI.-XVII. yüzyıllarda pek çok yazar bu sözcüğü kullanmışsa da amaçları yalnız siyasal devletleri tanıtmak olduğundan yapıtlarında pek az sayısal bilgiye rastlanmaktadır. Bu yazarların yaptıkları işin istatistikle ilgisi, Almanların *Staatskunde* (siyasal coğrafya) kavramının istatistikle ilgisi kadardır. Sömürgeleri olmayan, Kıta Avrupa'sında kendini sıkışmış görüp genişlemek isteyen Almanya'da, gerek bu ülke, gerek komşu ülkeler hakkında arazi, nüfus, mali-askeri-yönetmelik örgütlenme biçimleri, iktisadi-askeri güç konularında çoğu sayısal olmayan bilgiler derlenmesi, daha sonra da bunların bir ders olarak okutulması sonucunda doğan siyasal coğrafya kavramı zamanla sayısal verilerle çok daha fazla ilgilenmeye başlamış, böylelikle betimleyici istatistiğin temeli olma yoluna girmiştir. Ancak eğer siyasal coğrafya için istatistik sözcüğüyle aynı kökten gelen bir sözcük kullanılmamış olsaydı, bugünkü istatistikçiler bu çalışmalara hiçbir ilgi göstermeyeceklerdi (Kendall, 1960:45).

Yine Kendall'a göre günümüz istatistiğine az da olsa bir benzerlik gösteren çalışmalara 1660'lardan önce pek rastlanmamaktadır. Siyasal çalkantıların durulduğu bu dönemde, devlete ilişkin konularda sayılarla akıl yürütme sanatı olarak tanımlanan siyasal aritmetik doğmuş, bugünkü istatistiğin temellerini atmıştır. John Graunt 1662'de çıkan yapıtında insan ömrüyle meslekler, insan ömrüyle cinsiyet ilişkilerini incelemiştir. Ardından gelen William Petty'nin *Political Arithmetick* (1690) çalışması da dönemin çığır açan yapıtlarından sayılmaktadır. Bu dönemin ürünü olan sigorta matematikçileri, sayısal verileri çözümleyip yorumlayarak kurallar çıkarmayı, ilerisi için kestirimler yapmayı amaçlamaktaydılar. Adını bir kuyruklu yıldız veren Edmund Halley, 1693'te her yaş için hem ölme, hem yaşama olasılıklarını hesaplamış (Gülçür, 1946:4), doğum, ölüm, evlenme gibi olayların aralarındaki ilişkileri inceleyerek saptadığı düzenliliklerle çağımız

nüfusbiliminin kuruluşuna öncülük etmiştir. Yeni çağın ilk nüfus sayımı 1703'te İzlanda'da yapılmıştır.

Çağdaş anlamda ilk nüfus sayımının 1790'ı beklemesi gerekmiştir. O yıllarda yönetim örgütünü yeni oluşturan Amerika Birleşik Devletleri'nin Temsilciler Meclisi'nde her eyaletin kaç kişiyle temsil edileceğini saptamak bu sayımın ana amacı olmuştur. Avrupa ülkeleri de, başta Fransa olmak üzere 1801'de sayımlara başlamıştır.

XIX. yüzyılın sonlarına kadar sayımlarla derlenen verilerin düzenlenmesi, yayıma hazır duruma getirilmesi yıllarca sürmekteydi. 1890 ABD sayımını yöneten Herman Hollerith, daha sonra kendi adıyla anılacak olan delikli bilgisayar kartlarının ilk örneklerini geliştirerek bunları sayım sonuçlarının değerlendirilmesinde kullanmış, böylece bu süreyi üçte bir oranına indirmiştir. 1911'de görevinden ayrılarak, daha sonra IBM adıyla devleşecek bir bilgisayar ortaklığının kurucuları arasına katılmıştır. Yerine gelen James Powers da benzeri bir yol izlemiş, geliştirdiği makineler sayesinde Sperry Rand Co. adlı başka bir bilgisayar ortaklığını kurmuştur.

Osmanlılarda yapılan ilk sayım 1831 tarihlidir. Ancak hem ülkenin bütününe kapsamamış, hem de yalnızca erkekler sayılmıştır. Aslında Orhan Gazi ya da I. Murat zamanından beri arazi yazımının yapıldığı Osmanlı devletinde zaman zaman bütüncül olmayan nüfus sayımına gidildiği de bilinmektedir. Bunlar yalnızca vergi toplama amacına yöneliktir (Cillov, 1976:16; DİE, 2002a:38, Göçmençelebi, 1962:18). 1912 yılında sonuçlarına pek güven duyulmayan bir tarım sayımı, 1913-15 yıllarında da, savaş nedeniyle bitirilemeyen bir sanayi sayımı yapılmaya çalışılmıştır. Her iki sayım da, iktidara yeni gelen İttihad ve Terakki Cemiyeti'nin bazı verilere gereksinme duymasından kaynaklanmıştır (DİE, 1973:83).

Cumhuriyet kurulduktan sonra bu konuya gereken önemin verildiği, Başbakanlığa bağlı bir İstatistik Genel Müdürlüğü kurulmasından (1927) anlaşılmaktadır. O yıl yapılan ilk genel nüfus sayımı 1935 yılından sonra 1990'a kadar her beş yılda bir yinelenmiştir. 23.2.1990 günlü bir kanun hükmünde kararname gereği, bundan böyle nüfus sayımlarının her on yılda bir sonu sıfır ile biten yıllarda yapılacağı duyurulmuştur. 22 Ekim 2000'de on dördüncü genel nüfus sayımı yapılmıştır.

1927'den 1980'e kadar beş kez yinelenen tarım sayımları, 1989 tarihli bir kanun hükmünde kararnameyle her on yılda bir sonu birle biten yıllarda yapılmaktadır (DİE, 2002a:271). 1927'deki ilk sanayi sayımından sonra sanayi istatistikleri genellikle yıllık anketlerle derlenmiştir. 1966'dan sonra hem aylık, hem üçer aylık anketler uygulanmaya başlamıştır (DİE, 2002a:311-312).

Günümüzde DİE pek çok alanda veri derlemektedir. Bu konuya ileride daha ayrıntılı olarak değinilecektir.

2.2 İSTATİSTİK YÖNTEMLERİNİN EVRİMİ

İstatistik yöntemlerinin tarihçesine gelince, olasılığın atası sayılan şans oyunlarının 5-6 bin yıl önce "icat" edildiği ileri sürülmektedir (David, 1955:1). "Aşık atmak", "aşığı cuk oturtmak" deyimlerinde hâlâ yaşayan aşık oyununda kullanılan aşık kemiği, bugünkü zarın o dönemdeki ilkel karşılığıdır (David, 1962:3). (Bu aşık sözcüğünün başındaki "a"nın kısa okunduğunu anımsatmaya gerek var mı?)

Bu tür şans oyunları çağlar boyunca yalnız boş zamanlarda eğlence olsun diye oynanmakla kalmamış, günümüz Tibet’inde hâlâ izleri görülebilecek biçimde, Eski Mısır’da, Mezopotamya’da, Eski Hindistan’da, Eski Yunan’da dinsel amaçlarla da kullanılmışlardır (David,1955:8-10). İskambil kâğıtlarının 1350 yılı dolaylarında ortaya çıkışına kadar şans oyunlarında zar kullanılmıştır (Kendall 1956:19). Hristiyan din adamları Orta Çağ boyunca aşık kemiği ya da zarla oynanan oyunlara, dolayısıyla kumara, falcılığa sürekli karşı çıkmışlardır. 2004 yılı başında Diyanet İşleri Başkanlığı’nın piyangoya karşı çıkışı da buna benzetilebilir mi?

Bugün İngilizce’de şans, risk gibi anlamlar yüklenen *hazard* sözcüğü, Üçüncü Haçlı Seferi’ne katılanların dönüşte Avrupa’ya getirdikleri bir zar oyununun adı olup Arapça *al zar* sözcüğünden türemiştir (Kendall, 1956:21).

Bu dönemlerin hem parası, hem zamanı bol bazı "soylu"larının kumara düşkünlükleri nedeniyle, kendilerine kazanma yolları bulmaları için matematikçilere iş verdikleri bilinmektedir. XVII. yüzyılda bunlar arasında Galileo Galilei, Blaise Pascal gibi ünlüler de vardı (Akalin, 1977:3-4). Ancak onlardan önce "her yüzünün üste gelme şansı aynı olan bir zar" kavramından olasılık hesaplamayı İtalyan matematikçiler bilmekteydiler. Bu soyutlamayı ilk kez, kumarbaz bir matematikçinin yaptığı sanılmaktadır (David, 1955:14).

XVII. yüzyılla birlikte olasılık kuramının gelişmesi büyük bir ivme kazanmıştır. Blaise Pascal (1623-1662) **matematiksel beklenti** kavramından yola çıkarak **karar kuramının** temellerini atmış, **olasılık** terimini kullanmamış olmakla birlikte eşit şanslı olaylar, rassal yürüyüş, olasılığa sıklık yaklaşımı kavramlarını geliştirmiş (Johnson, Kotz, 1997:61-63), **ikiterimli** (binom), **eksi ikiterimli** (negatif binom) **dağılımlarının** olasılıklarını bulmakta kullanılan katsayıları incelemiş, olasılık hesaplarının yolunu açmıştır. Jakob Bernoulli (1654-1705) olasılık kuramını ele alan ilk yapıtı yazmış, burada **büyük sayılar yasasını** geliştirip (Sheynin, 1968:232) **kombinasyon**, **permutasyon**¹ hesaplarını ortaya atmıştır. Pierre-Simon Laplace

¹ Bunlara Türkçe karşılık olarak "birleşim, sıradüzen" (İnal vö. 1983), "katışım, devşirim" (Çoker, Karaçay, 1983) sözcükleri önerilmektedir.

(1749-1827), 1812 tarihli yapıtıyla olasılık kuramını geliştirmiş, bu kuramı yalnız sıradan şans problemlerine uygulamakla kalmamış, olayların nedenlerini, yaşam istatistiklerini, gelecek olayların kestirimini de incelemesine katmış, Thomas Bayes'tan (1701-1761) habersiz olarak Bayes teoremini bulmuştur. Carl Frederich Gauss (1777-1855), pek çok olasılık hesabının temelini oluşturan çan biçimindeki **normal dağılım** ya da **Gauss dağılımı** eğrisini bulmuş (1820'ler), bugün **en küçük kareler** yöntemini bir gökcisminin yörüngesini hesaplamakta çok başarılı bir biçimde kullanmıştır (1801). Olasılık kuramıyla ilgilenen bu kişilerden hiçbirinin tek konularının bu olmadığı, hepsinin matematik, fizik, gökbilim, vb. dallarda isim yapmış olmaları o dönemin çok yönlü bilim adamı geleneğine uymaktadır.

Çağdaş istatistiğin kurucusu sayılan Adolphe Quetelet (1796-1874) üç ayrı akımın (betimleyici istatistikçiler diyebileceğimiz Alman okulunun, sigorta matematikçilerinin, olasılık kuramının) tekniklerini bir araya toplayıp istatistiği bütün toplumsal olaylara, özellikle de insanbilime uygulamıştır. Yalnız istatistik araçlarıyla gözlenebilecek, tanımlanabilecek davranış türlerine ilk kez dikkat çeken bu bilim adamı, binlerce kişi üzerinde yaptığı ölçümlerin Gauss dağılımına uyduğunu göstererek "ortalama insan" kavramını ortaya atmıştır (Seal, 1967:207).

Bütün bu gelişmelere karşın XIX. yüzyıl felsefi belirlenimciliğin doruğa çıktığı bir dönemdir. Bu inanışa göre, evrenin başlangıç koşullarından yola çıkıp hareketleri betimleyen matematik formülleri yardımıyla her şey önceden kestirilebilirdi. Isaac Newton'un (1642-1727) hesaplamaları kullanılarak o zamana kadar bilinmeyen bir gezegen olan Neptün'ün elle konmuş gibi bulunması, 1840'larda "saat gibi işleyen" evren düşüncesini adanmakıllı pekiştirmişti. Ancak gerek Newton'un, gerek gezegenlerin, kuyruklu yıldızların gelecekteki konumlarının yerden yapılan gözlemlerle nasıl kestirilebileceğini gösteren Pierre Simon Laplace'ın (1749-1827) hesaplamalarında "hata fonksiyonu" denilen bir kavrama gerek vardı. Kesitirimleri kusursuz değildi, küçük hatalar içeriyordu. Bu hataların dünya atmosferinin yarattığı yanılgılardan, gözlemlerden, insan yanılgılarından doğduğu düşünülüyordu. Hata fonksiyonu bütün bunları emiyor, geriye saf hareket yasaları kalıyordu. Daha duyarlı ölçme araçları geliştikçe bu hataların azalacağı umuluyordu. Ama öyle olmadı. XIX. yüzyılın son dönemlerinde hatalar azalacağına arttı. Ölçmeler hassaslaştıkça hatalar daha da belirginleşti. Saat gibi işleyen evren düşüncesi çöktü. Bilim yavaş yavaş gerçekliğin yeni bir modeliyle, istatistikle çalışmaya başladı (Salzburg, 2001.vii-viii).

Bilimler "**gözlem değerleri = istatistik modeli + hata**" eşitliğine önem vermeye yöneldiler. Kurulan model ne kadar başarılıysa hata teriminin gözlem değerine göre büyüklüğü de o kadar azdır. Hata terimleri belki sıfır-

lanamaz ama çok küçük boyutlara indirilebilir. Böylece istatistik modellerle gözlemlerin içerdiği gerçeklik yakalanmaya çalışılır.

Francis Galton (1822-1911) bu dönemde **korelasyon (ilişki)** sözcüğünü istatistikteki anlamıyla ilk kez kullanmış, 1877’de ilk **regresyon (bağlanım) katsayısı** hesabını yapmıştır (Johnson, Kotz, 1997:109). Çağdaş istatistiğin en temel direklerinden biri olan Karl Pearson (1857-1936) (Brookes, Dick, 1967:2), istatistiği biyolojiye, toplumsal bilimlere uygularken **ki-kare sınaması** gibi kuramsal katkılarda bulunmuştur. Biometri terimini ilk kullanan Pearson, *Biometrika* dergisini de çıkarmıştır. Ronald Aylmer Fisher (1890-1962) **deney tasarımı, varyans çözümlemesi** gibi konulardaki katkıları dolayısıyla büyük istatistikçiler arasına girmiş, soyadının baş harfi, kendi bulduğu **F dağılımına** verilmiştir (Johnson, Kotz, 1997:99-108).

XX. yüzyıl istatistiğin parlak isimleriyle doludur. William Sealy Gosset “Student” (1876-1937) **t dağılımını**, Andrey Andreviç Markov (1856-1922) **Monte Carlo benzetimlerinin** yolunu açan **Markov zincirini**, Corrado Gini (1884-1965) gelir dağılımındaki eşitsizliklerin ölçülmesinde kullanılan **Gini katsayısını** bulmuş, Irving Fisher (1867-1947) **indeks sayılarını** kuramlaştırmış, Charles Edward Spearman (1863-1945) **etmen çözümlemesinin** (faktör analizinin) temellerini atmış, Jerzy Neyman (1894-1981) ile Egon Sharpe Pearson (1895-1980) **önsav sınamasını** geliştirmiş, yine Neyman istatistiğe **güven aralığı** kavramını kazandırmış, Andrey Nikolayeviç Kolmogorov (1903-1987) **olasılığın belitlerini** (aksiyomlarını) yazmış, Pefnuti Livoviç Çebişev (1821-1894) bazı temel limit teoremlerini, Paul-Pierre Lévy (1886-1971) **merkezi limit teoremini** kanıtlamış, Frank Wilcoxon (1892-1965) **katsayısal olmayan sınamalar** alanının kapılarını açmış, Leonard Jimmie Savage (1917-1971) ile Bruno de Finetti (1906-1985) **kişisel olasılık** kavramını ortaya atmış, William Edwards Deming (1900-1993) **kalite denetiminde** istatistiği kullanmış, John W. Tukey’in (1915-2000) başını çektiği Princeton takımı **berk (robust) kestirim yöntemlerini** tanıtmış, yine Tukey **sorgulayıcı veri çözümleme yöntemlerini** yaymıştır. Bu yüzyılda istatistiğin kullanılmadığı bilimsel alan kalmamış gibidir.

İstatistiğin, özellikle iktisatçılar için gittikçe daha çok önem kazanmaya başlayan bir türevi olan **ekonometrinin** doğumu da bu yüzyılın ilk çeyreğine rastlar. Bu terim çağdaş anlamıyla ilk kez Norveçli iktisatçı-istatistikçi Ragnar Frisch tarafından 1926 yılında önerilmiştir (Lange, 1978:15; Ertek, 1996:14; Genceli, 2001:1) ama iktisadi olaylara ekonometri tekniklerinin uygulanmasına 1900’lerden bu yana rastlanmaktadır. Hatta Pawel Ciompa adlı bir Alman yazar daha dar bir anlamda da olsa “ökonometrie” terimini bile kullanmıştır (Berndt, 1991:iii). Ekonometrinin bu dönemde ortaya

çıkması bir rastlantı değildir. Bu dönem batı ülkelerinde hem tekel gücünün, hem devlet kapitalizminin gelişme yıllarıdır. Lange'ye göre ekonometri bu iki olgunun yarattığı gereksinimleri karşılamak amacıyla gelişmiştir (Lange, 1978:18-23).

Biometri, sosyometri gibi bilim dalları da istatistikteki bu hızlı gelişmenin ürünleridir.

Bütün bu gelişmelerin sonucunda bugün hemen her bilim dalında öğrenim gören gençler istatistik dersi okumakta, hemen her alandaki araştırmacılar istatistik tekniklerinden yararlanmaktadırlar.

Okuma Önerileri

- Bernal, J. D. (1965) *Science in History*, 3rd ed., Penguin,
 Çömlekçi, N. (1985) *İstatistiğin Tarihçesi*; Anadolu Üniversitesi.
 David, F. N. (1962) *Games, Gods and Gambling*, Charles Griffin.
 DİE (1973) *Toplumsal ve Ekonomik Gelişmenin 50 Yılı*, DİE Matbaası.
 DİE (1993) *Osmanlı Devleti'nde Bilgi ve İstatistik*, DİE Matbaası.
 Johnson, N. L., Kotz, S. (1997) (eds.) *Leading Personalities in Statistical Sciences*, Wiley.
 Kutsal, A., Toktamış, Ö. (1981) *İstatistik (Terimler, Tarihçe, Örgütler, Yayınlar, Öğretim)*, Hacettepe Üniversitesi.
 Pearson, E. S., Kendall, M. G. (1970) (eds.) *Studies in the History of Statistics and Probability*, Griffith.
 Salzburg, D. (2001) *The Lady Tasting Tea – How Statistics Revolutionized Science in the Twentieth Century*, Freeman.

ALIŞTIRMALAR

- 2.1 Osmanlıların *ihsaiyat* adını verdikleri istatistik biliminin Türk tarihindeki yerini inceleyiniz.
- 2.2 Aşağıda adları yazılı bilim insanlarından istediklerinizi seçerek istatistiğe yaptıkları katkıları inceleyiniz. Bugün kullanılan hangi kavramları, yöntemleri, modelleri onlara borçluyuz?

T. Bayes (1701-1761)	Bernoulli ailesi (17.-18. yüzyıl)
A. L. Bowley (1869-1957)	A. Cauchy (1789-1857)
W. G. Cochran (1909-1980)	G. M. Cox (1900-1978)
P. L. Çebişev (1821-1894)	F. N. David (1909-1993)
B. De Finetti (1906-1985)	A. De Moivre (1667-1754)
W. E. Demming (1900-1993)	F. Y. Edgeworth (1845-1926)
C. L. E. Engel (1821-1896)	R. A. Fisher (1890-1962)

I. Fisher (1867-1947)	F. Galton (1822-1911)
C. F. Gauss (1777-1855)	C. Gini (1834-1965)
W. S. Gosset (1876-1937)	L. Guttman (1916-1987)
J. Hayek (1926-1974)	W. Hoeffding (1914-1991)
H. Hotelling (1895-1973)	P. Laplace (11749-1827)
P. Levy (1886-1971)	M. G. Kendall (1907-1983)
A. N. Kolmogorov (1903-1987)	P. C. Mahalanobis (1893-1972)
A. A. Markov (1856-1922)	J. C. Maxwell (1831-1879)
S. F. Nightengale (1820-1910)	I. Newton (1642-1727)
J. Neyman (1894-1981)	B. Pascal (11623-1662)
E. S. Pearson (1895-1980)	K. Pearson (1857-1936)
W. Petty (1623-1687)	L. A. J. Quetelet (1796-1874)
F. P. Ramsey (1903-1950)	E. E. Slutsky (1880-1948)
N. V. Smirnov (1900-1966)	G. W. Snedecor (1881-1974)
C. E. Spearman (1863-1945)	A. Wald (1902-1950)
N. Wiener (1894-1964)	F. Wilcoxon (1892-1965)
S. S. Wilks (1906-1964)	H. O. A. Wold (1908-1992)
J. Wolfowitz (1910-1981)	F. Yates (1902-1994)
G. U. Yule (1871-1951)	

Neyi, Nasıl Ölçelim?

Bir gün Nasreddin Hoca'ya sormuşlar:
 "Efendi! Sen bilirsin, şu dünyanın tam ortası neresidir?"
 Hoca sol ayağının bastığı yeri göstermiş: "İşte tam burası."
 Gülüşmüşler: "Eğlenme! Merak ettik de soruyoruz!"
 "Eğlenmiyorum," demiş Hoca. "İnanmıyorsanız ölçün."
Memet Fuat'ın söyleyişiyle

İstatistik çalışmalarında kullanılan veriler ya araştırmacının kendisince derlenir ya da yayımlanmış kaynaklardan hazır alınır. Birinci yolu kullanmak, veri derleme yöntemlerini doğru uygulamayı, ikinci yolla veri toplayabilmek de hangi yayınlarda, ne gibi verilerin bulunduğunu bilmeyi gerektirir. 5. Bölümde her iki yola ilişkin kısa bilgiler verilecektir. Ancak verileri nasıl bulacağımıza geçmeden önce verinin ne olduğu, nasıl ölçüleceği, nasıl kullanılacağı konuları üzerinde kısaca durmakta yarar var. Bunun için de bazı kavramları tanımamız gerekiyor.

3.1 ANAKÜTLE, ÖRNEKLEM, DEĞİŞKEN, DEĞER

Herhangi bir istatistik çalışmasında ele alınan konunun sayısal bakımdan incelenebilmesi için önce bu olaya ilişkin birimlerin belli bir ya da bir kaç özelliğinin saptanarak sayılması/ölçülmesi gerekir.

Sözgelimi bir ülkedeki ailelerin gelir düzeylerine ilişkin veri derlemek istediğimizi düşünelim. Burada ülkede yaşayan bütün ailelerin oluşturduğu topluluğa **anakütle** denir. Zaman darlığı, pahalıya çıkma gibi nedenlerle bütün ailelere ilişkin gözlem yapmak elverişli olmuyorsa, o zaman özel yöntemlerle seçilmiş bazı ailelerle ilgili bilgi derlemek gerekir. Bu daha küçük aile kümesine **örneklem**, örneklem giren ailelerin her birine **gözlem birimi** ya da kısaca **birim** denir. Bu örnekte ilgilendiğimiz nitelik olan gelir düzeyine **değişken** adını verebiliriz. Her ailenin gelir düzeyi, bu değişkenin aldığı çeşitli **değerler**dir. Bunlara **gözlem değeri** de denir. Ailelerin gelir düzeyleri, bir başka deyişle gelir düzeyi değişkeninin aldığı değerler bizim toplamak istediğimiz verilerdir.

Bu örnekte anakütle somut bir kavramdır, ama soyut bir anakütle de inceleme konusu olabilir. Sözelimi, yeni geliştirilen bir ürün türünün ileride üretilecek bütün birimleri anakütleyi oluşturabilir. Bu ürünün belli bir özelliğinin, diyelim dayanıklılığının ölçülmesi gerekiyorsa, örnek olarak üretilmiş az sayıda birimin dayanıklılığının incelenmesiyle anakütleyle ilişkin bilgi sahibi olunabilir.

İncelemeye konu olan birimin ille de elle tutulabilir, sürekli bir varlık olması gerekmez (Gürtan, 1987:23). Doğum, ölüm, trafik kazası, fiyat gibi birimler de pekâlâ kullanılabilir. Dikkat edilirse insan, konut, mal stoğu, seçmen gibi belli zaman dilimlerinde varolan verilerle, dış ticaret, fiyat, evlenme, boşanma gibi anlık birimlere ilişkin verilerin toplanma biçimlerinin de farklı olacağı açıktır. İlkinde belli aralıklarla gözlemler yapmak yeterliyken ikinci türde sürekli gözlem zorunludur. Bu nedenle nüfus sayımları, stok sayımları belli aralıklarla yapılır, ama trafik kazası, dış ticaret, fiyat verileri sürekli olarak derlenir.

Hakkında veri derlenecek birimin çok iyi tanımlanması büyük önem taşır. Sözelimi nüfus sayımlarının birimi olan insanı nasıl tanımlayabiliriz? Sayım günü ülkede bulunan herkes, bu arada yabancı uyruklular da, sayılacak mıdır? Sayım görevlisi bir eve girdiğinde az önce ölmüş biri ya da henüz doğmuş bir bebek de sayıma katılacak mıdır? Herkes sayım sırasında bulunduğu yerde mi, yoksa yasal olarak oturduğu yerde mi sayılacaktır? Bu son soruya verilecek yanıtı göre fiili ya da yasal sayım gibi iki değişik nüfus sayımı biçiminden bile söz edilebilir.

Öyleyse bir anakütle insanlardan oluşuyorsa bunların meslekleri, yaşları, işsiz olup olmadıkları, gelir düzeyleri ilgilendiğimiz özellikleri arasında sayılabilir; birim trafik kazasıysa ele alınan değişken kaza yeri, ölü-yaralı sayısı, maddi hasar tutarı olarak belirlenebilir; birim dışsatım olayı ise malın cinsi, miktarı, fiyatı, tutarı, gittiği ülke, CIF ya da FOB gibi satış biçimleri değişkenleri oluşturulabilir. Üniversite öğrencisi birim olarak alınmışsa değişkenlerin neler olabileceğini saymak okuyucuya bırakılmıştır.

Bu değişkenlerden meslek, işsiz olup olmamak, kaza yeri, malın cinsi, gittiği ülke, satış biçimi gibi bazılarının değerleri sayılarla gösterilemezken; yaş, gelir düzeyi, ölü-yaralı sayısı, maddi hasar tutarı, malın miktarı, fiyatı, tutarı gibi değişkenlerin sayısal değerler aldığı söylenebilir. Birinci türden değişkenlere **nitel** (kategorik), ikincilere **nicel** (nümerik) **değişken** denmektedir.

Görüldüğü gibi hem nitel, hem nicel değişkenlere ilişkin veri derlemek, bunları istatistik araştırmalarında kullanmak olanaklıdır. Böyle bir ayırım yapmamızın nedeni iki tür değişkenin değerlerini gerek **ölçme** sırasında, gerek istatistikle inceleme sırasında farklı yöntemlerin kullanılması gerektiğidir. İleride bu konuya ilişkin açıklamalar yapılacaktır.

3.2 ÖLÇEKLERİN ÖNEMİ

*Ölçün, ölçemiyorsanız ölçmeye çalışın.
Galilei Galileo*

Buradaki ölçme sözcüğünden ne anladığımız çok önemlidir. Eğer uzunluk, zaman gibi ölçmelerden söz ediyorsak kullanabileceğimiz matematik işlemleri konusunda fazla bir sıkıntımız yoktur. Ama ölçme kavramını biraz genişletip toplum bilimlerinde kullanılan zekâ, yetenek, yoksulluk, toplumsal konum gibi ölçmeleri de bu kavrama katıyorsak işlemler sırasında bazı sorunlar çıkabilir. Bu durumda çeşitli **ölçme düzeyleri** arasında bir ayırım yapıp her biri için kendine uygun farklı istatistik modeller bulmamız gerekir.

Burada uygulamacılara çok kolaylık sağlayan bir ayırıma uyularak dört değişik ölçme düzeyi ele alınacaktır (Siegel, Castellan, 1988: 23): Ad ölçeği, sıra ölçeği, aralık ölçeği, oran ölçeği.

3.3 AD ÖLÇEĞİ

Her bilimde en temel işlem kümelere ayırma, bölümlendirmedir. Birimleri kümelendirmek, belli bir niteliğe göre hangi birimlerin birbirine benzediğini, hangilerinin farklı olduğunu belirlemek demektir. Başka bir deyişle gözlem birimlerini, değişkenin aldığı değerlere göre değişik kümelerde toplamaktır. Bunu yapmaktaki amaç, aynı kümeye giren birimlerin birbirinden farkının, kümeler arasındaki farklarla karşılaştırıldığında, olabildiğince az olmasıdır, Yani her kümenin kendi içinde türdeşlik sağlanmasıdır.

Sözgelimi mesleklere göre kümelendirme yaparken, bilgisayar üretimi yapan bir işletmenin ortaklarını, burada çalışan mühendisleri, ustaları, işçileri, muhasebecileri, satıcıları, bilgisayar dışalımçıları, dışsatımcılarını, tamircilerini, yazılımcılarını aynı kümede mi, yoksa farklı kümelerde mi toplayacağız? Bu sorunun yanıtı amacımızın ne olduğuna bağlıdır. Eğer bize bilgisayarın üretimi, dağıtımı, bakımı alanında çalışanların bir bütün olarak sayısı gerekliyse yukarıda sayılan kimseleri aynı kümeye sokabiliriz. Farklı amaçlar için başka kümelendirmeler uygun olabilir.

Kümelere ayırmanın bir işe yarayabilmesi için her kümenin, başka değişkenler için yapılmış kümelendirmelere göre de türdeş olması sağlanmalıdır (Blalock, 1972:12). Sözgelimi mesleklere göre bir kümelendirme yapıp meslek türü ile ömür uzunluğu ya da gelir düzeyi arasında bir ilişki var mı, yok mu diye bakılsa, diyelim maden işçilerinin ortalama ömrünün he-

kimlere göre daha kısa, ortalama gelir düzeylerinin sanayicilere göre daha düşük olduğu bulunabilir. Ama benzeri bir karşılaştırmayı saç rengiyle gelir düzeyi arasında yapsak anlamlı bir fark çıkmayabilir. Başka bir deyişle, değişik saç renklerine sahip **kümeler arasındaki** gelir düzeyi farkları, aynı saç rengine sahip **küme içindeki** gelir düzeyi farkları yanında hafif kalabilir. Başka bir bağlamda, sözgelimi ana-babaların saç rengiyle çocuklarının saç rengi arasındaki ilişkinin araştırılması konusunda pekâlâ işe yarayabilecek bir değişken olan saç rengi, ilk durumda işe yaramamaktadır.

Ad ölçeklerinde iki temel özellik bulunur. Biri **bakışım**, öteki **geçişlilik**. Bakışım, eğer $A = B$ ise $B = A$ anlamına gelir: Yani A gözlemi B 'nin kümesindeyse, B de A 'nın kümesindedir. Geçişlilik ise şöyle gösterilebilir. Eğer A ile B 'nin aynı kümede olduğunu biliyorsak, B ile C 'nin kümeleri de ortaksa, A ile C de aynı kümededir.

Kümelendirmeyi, ölçmenin en düşük düzeyi olarak alabiliriz. Kümelere ayırma hem **kapsayıcılık**, hem **örtüşmezlik koşullarını** sağlamalıdır. Eğer kurduğumuz kümeler, bütün gözlemleri kapsıyorsa (yani kümelerden hiçbirine girmeyip dışarıda kalan gözlem yoksa), aynı zamanda bu kümeler birbirleriyle örtüşmüyorsa (yani her gözlem yalnız bir tek kümeye giriyorsa) en azından istatistik işlemlerin uygulanması için gerekli koşullar sağlanmış demektir.

Bu kümelerden her birine uygun adlar takabiliriz ama bunlardan biri ötekinden “daha iyi, daha üstün” gibi nitelemelerde bulunamayız. Kümelerin adlarını harfler, sözcükler ya da sayılardan oluşturabiliriz ama kullandığımız sayıların bize bunlar arasında aritmetik işlemler yapma olanağı vermediğini unutmamalıyız.

Öğrencilere verilen okul numaralarının en sol basamakları okula kaydoldukları yılları gösterecek biçimde düzenlenirse, öğrenciler bu basamaklardaki sayılara bakılarak okula girdikleri yıllara göre kümelendirilebilirler. Benzeri bir uygulama taşıt plaka numaralarında illere göre yapılmaktadır. Burada da taşıtlar plakalarının ilk iki sayısına bakılarak illere göre kümelendirilebilir. Ancak akli başında hiç kimse bu okul numaralarını ya da plaka numaralarını toplamaya ya da başka bir aritmetik işlemde kullanmaya kalkmaz. Bunun gibi evlenmemişlere 0, evlilere 1, boşanmışlara 2 vb. sayılar versek, bunlarla da işlem yapılamayacağı açıktır.

İşte böyle en alt ölçme düzeyinde, sayısal olmayan nitel değişkenlerin kümelendirilmesinde kullanılan ölçeğe **ad ölçeği** denir. Bunu Arıcı (1988:13) *sınıflama-kategorileme-gruplama ölçeği*, Orhunbilge vö. (1997) *sınıflayıcı ölçek*, Sencer (1981) *adlayıcı ölçek* olarak adlandırmaktadır. Her ne kadar küme adı olarak kullanılan sayılar arasında işlem yapamasak da bu kümelerin her birine giren gözlemlerin sayısı bulunabilir. Bu sayılara **sıklık** (frekans) denir. Yukarıda belirtilen **kapsayıcılık-örtüşmezlik** koşul-

ları yerine geliyorsa, her kümedeki gözlem sayısı (yani her kümenin sıklığı) toplam gözlem sayısına bölünerek **paylar**, **yüzdele** hesaplanabilir, her kümedeki sıklık sayısı, payı, yüzdesi başka kümedekilerle karşılaştırılabilir ya da onlara **oranlanabilir**. Bunlardan en yüksek sıklığa sahip olanı belirlenebilir. Şimdi Çizelge 3.1'e bir göz atalım. Bu çizelgedeki değişkenimizin adı "medeni durum"dur, aldığı değerler de "hiç evlenmedi", "evli", "eşi öldü", "boşandı" kümelerine ayrılmıştır. Kapsayıcılık özelliğinin sağlanması için bunlara bir de medeni durumunun ne olduğu belli olmayan (sayım yapılırken medeni durumu sağlıklı biçimde saptanamamış) kimseler için "bilinmeyen" kümesi eklenmiştir. Görüldüğü gibi bu kümelerden hiçbirinin ötekilere göre daha üstün olduğu söylenemez. Bekârlık herkes için sultanlık mıdır acaba? Medeni durum değişkenin aldığı değerler bir büyüklük sırasına konulamayacak ölçüde niteldir. Çizelgedeki küme adlarının sıralarını değiştirsek verilerin açıklama gücü değişmez.

Kümelere giren kişi sayıları çizelgede bin kişiye yuvarlatılarak verilmiştir. Her kümedeki kişi sayısının toplam içindeki payının yüzde karşılığı da çizelgede gösterilmiştir. Bu hesaplamaları yapabilmemiz için 11 yaşından büyük *herkesin* bu kümelerden birine ama *yalnız* birine girdiğine, yani hem kapsayıcılık, hem örtüşmezlik koşullarının sağlandığına güvenmemiz gerekir. Bu örnekte bilinmeyen kümesinin varlığı kapsayıcılık özelliğinin yerine geldiğini gösterir; tek tek küme sıklıkları toplamının, toplam sıklığa eşit olması da örtüşmezlik özelliğinin sağlandığı anlamına gelir. Böylece

ÇİZELGE 3.1 2000 Genel Nüfus Sayımında İstanbul'da yaşayan 11 yaşından büyüklerin medeni durumları

Medeni durum	Kişi sayısı (bin)	Pay (%)
Hiç evlenmedi	2.710	34,22
Evli	4.746	59,93
Eşi öldü	325	4,11
Boşandı	136	1,72
Bilinmeyen	1	0,02
Toplam	7.919 ¹	100.00

DİE (2003:135)

¹ Yuvarlatma nedeniyle toplam 7.919'u vermemektedir. Bu konu, ileride yuvarlatma anlatılırken ele alınacaktır.

yüzde payları ya da **görelî sıklıkları** hesaplamamız için bir engel kalmamaktadır. Hiç evlenmeyenlerin toplam içindeki payının yaklaşık % 34, evlilerinse yaklaşık % 60 kadar olduğunu görmekteyiz. Eşi ölenler % 4'ün üstünde, boşanmışlar da % 2'ye yakındır. Her onbin kişiden ikisinin medeni durumunun saptanamadığı anlaşılmaktadır.

Ayrıca küme sıklıkları arasındaki oranlar da hesaplanabilmektedir. Sözelimi evlilerin sayısı, hiç evlenmeyenlerin yaklaşık 1,9 katıdır ya da hiç evlenmeyenler evlilerin 0,57'si kadardır. Öteki kümeler arasında da benzer oranlar hesaplanabilir.

Kümelendirmenin bize sağladığı bilgilerden biri de en yüksek sıklığın hangi kümede olduğudur. **Ensıkdeğer** (mod, tepedeğer, egemen değer) diye adlandırılan bu küme örneğimizde “evli” kümesidir. Başka bir deyişle evli olma, 2000 nüfus sayımında İstanbul'da yaşayan 11 yaşından büyük kimseler arasında *en sık rastlanan* medeni durumdur.

Eğer incelediğimiz değişken, örneğimizdeki medeni durum değişkeni gibi nitelse, derlenen bilgilerle burada yaptıklarımızdan daha fazla istatistik işlemi yapılamaz. Bu söze şöyle karşı çıkılabilir. Toplam kişi sayısı çizelgedeki beş kümeye bölünerek **aritmetik ortalama** bulunamaz mı? Ya da bilinmeyen kümesindeki bin kişi dışındakiler, kalan dört kümeye bölünerek her kümede *ortalama kaç kişi* olduğu görülemez mi? Bu işlemler elbette yapılabilir, bulunan sonuçlara belli anlamlar da yüklenebilir. Ancak elde edilen ortalama, bizim ilgilendiğimiz değişken olan medeni durum değişkeninin değil, “her medeni durumdaki kişi sayısı” diyebileceğimiz başka bir nicel değişkenin ortalamasıdır. Dolayısıyla “medeni durum denilen nitel değişken için aritmetik ortalama hesaplanamaz” savı yine geçerliliğini korur.

Pay, bir kümenin bütün içinde tuttuğu yer olarak tanımlanabilir. Evlilerin 11 yaşından büyükler içindeki payı yaklaşık 3/5 kadardır dediğimizde bu kolaylıkla anlaşılır. Bunu **yüzde** olarak % 60 ile de gösterebiliriz. Bu değer aynı zamanda **görelî sıklık** olarak da adlandırılabilir. **Oran** ise daha çok iki kümedeki sayıların birbirine oranı anlamında kullanılmaktadır. Boşanmışların, eşi ölenlere oranı 5:2, ya da tersi eşi ölenlerin boşanmışlara oranı 2:5 kadardır. Bu kavramlar ileride daha ayrıntılı olarak incelenecektir.

Diyelim yukarıdaki medeni durum değişkeninin çeşitli kümelere dağılımını, 1990 nüfus sayımındaki dağılımla karşılaştırmak, on yıl içinde bir değişme olup olmadığını, yahut 2000 nüfus sayımında İstanbul'un yanı sıra bir kaç ilimizdeki benzer medeni durum dağılımlarını karşılaştırıp, bunların birbirine benzeşmelerini görmek istiyoruz. Ad ölçeğiyle kümelendirilmiş verilerden bu tür çıkarsamalar yapmak için **ikiterimli** (binom) *sınaması*, *ki-kare sınaması* gibi bazı **katsayısal olmayan yöntemlere** başvurulabilir.

3.4 SIRA ÖLÇEĞİ

Ad ölçeğinde ölçülmüş gözlemlerin toplandığı kümelerin kendi aralarında bir sıralama yapamayız demiştik. Eğer nitel bir değişkenin gözlemlerinin dağıtıldığı kümeler arasında genel kabul gören bir sıra varsa kullanılan ölçeğe **sıra ölçeği** denebilir. Buna *sıralama ölçeği* (Arıcı, 1988:14) ya da *sırasal ölçek* (Kurtuluş, 1985:313; Orhunbilge vö. 1997:44), *sıralayıcı ölçek* (Sencer, 1981) adları da verilmektedir. Burada gözlemlerin belli bir sıraya göre dizildiklerini, ancak bunlardan bazılarının sıralandıkları eksen üzerinde aynı noktayı paylaştıklarını düşünebiliriz. Aynı noktaya düşen gözlemler arasında bir ayırım yapamıyorsa bunları aynı kümenin içine koyabiliriz. Yine de belli bir kümeye giren gözlemlerin, başka kümelere girenlere göre daha yüksek ya da daha düşük değer aldıklarını söyleyebiliriz.

Sözelimi bir kentteki aileler toplumsal-ekonomik konumlarına göre “üst”, “üst-orta”, “alt-orta”, “alt” kümelerine ayrılabilir. Ya da örnekleme seçilmiş bazı kimselerden bir ürünün çeşitli markaları arasında bir beğenme sıralaması yapmaları istenebilir. Böylece “en çok beğenilen”, “daha az beğenilen”, “en az beğenilen” markalar sırası saptanabilir. Bu ilişkiler $>$ simgesiyle gösterilebilir. $>$ simgesinin solundaki gözlem sağındakine göre daha üstündür, daha beğenilir, daha zordur anlamlarını taşır.

Sıra ölçeğinde de ad ölçeğindeki gibi **bakışım** özelliği bulunur, ancak $A > B$ ise $B > A$ olamaz (Siegel, Castellan, 1988: 23). Sıra ölçeği bu anlamda bakışık değildir. Yani A gözlem değeri B ile aynı kümedeyse, B de A ile aynı kümededir, ama A, B 'nin bulunduğu kümenin daha üstünde bir kümedeyse, B, A 'nın kümesinden yukarı bir kümede olamaz. Sıra ölçeği **geçişlilik** özelliğini de taşır. Eğer $A > B$ ise, aynı zamanda $B > C$ ise, $A > C$ 'dir. Öyleyse A gözlemi B gözleminden, B de C 'den daha yukarı bir kümede yer alıyorsa, A, C 'den daha üst bir kümede bulunmalıdır.

Sıra ölçeğiyle yapılan ölçmelerin, ad ölçeğiyle yapılan ölçmelere göre daha üst düzeyde oldukları söylenebilirse de, sıra ölçeğiyle yapılan ölçmelerin, gözlem değerleri arasındaki farkların büyüklükleri konusunda hiçbir bilgi vermedikleri unutulmamalıdır. Sözelimi en çok beğenilen markayla ikinci sırada beğenilen marka arasındaki fark kıl payı kadarken, ikinci sırada beğenilen markayla üçüncü arasında dağlar kadar fark bulunabilir. Bir başka deyişle K markasını L 'ye yeğleyen bir kimse bu seçimini epey zorlanarak yaparken, L ile M markaları arasındaki sıralamada hiç duraksamadan L markasını daha çok beğendiğini dile getirebilir. Elbette bunun tam tersi de söz konusu olabilir. Önemli olan bu sıralamaya bakıp farkların büyüklüklerine ilişkin bir fikir edinemeyişimizdir. Bunu şöyle açıklayabiliriz. K ile M arasındaki uzaklığı $KM = KL + LM$ ilişkisiyle gösterebiliriz ama KL

ÇİZELGE 3.2 İTÜ öğrencilerinin başlangıçtaki beklentileriyle bugünkü izlenimlerinin karşılaştırılması

Beklentiye göre İTÜ'yü beğenme düzeyi	Kişi sayısı	Pay (%)
Beklediğimden daha iyi	583	10,0
Beklediğim düzeyde	2.604	44,7
Beklediğimden kötü	2.639	45,3
Toplam	5.826	100,0

İTÜ Vakfı (1991:15)

ile *LM* uzaklıklarını birbirleriyle karşılaştıramayız. Bu sıra ilişkilerini aritmetik işlemlere aktaramayız. Yani burada toplama, çıkarma, çarpma, bölme işlemlerini genellikle kullanamayız.

Sıra ölçeğine bir örnek vermek amacıyla 1991 yılı başlarında yapılan bir araştırmanın sonuçlarından uyarlanmış Çizelge 3.2'ye göz atarsak İTÜ öğrencilerine İTÜ'yü başlangıçtaki beklentilerine göre nasıl buldukları sorusuna alınan yanıtların özetlendiğini görürüz.

Çizelgede yanıtlar üç kümede toplanmıştır: “Beklediğimden daha iyi”, “aynı düzeyde”, “daha kötü”. Bu örnekte değişken “beklentiye göre İTÜ'yü beğenme düzeyi” olarak adlandırılabilir. Ad ölçeği için kullanılan örnekteki “medeni durum” değişkeni gibi bu değişken de niteldir ama beğenme düzeyleri bir büyüklük sırasına konulabilir. Her küme bu sıralamadaki yerini almıştır.

Bu örnekte de ad ölçeğinde kullanılan medeni durum değişkeni için yapılan bütün hesaplamaların benzerleri yapılabilir. Sözgelimi düş kırıklığına uğrayanların en kalabalık kümeyi oluşturduğu, dolayısıyla ensıkdeğerin bu küme olduğu söylenebilir. Burada da her kümenin görelî sıklıkları (toplam içindeki payları) hesaplanmıştır. Ayrıca çeşitli kümelerin sıklıklarının birbirleriyle oranları da bulunabilir. Kısacası ad ölçeğiyle ölçülen değişkenler için yapılabilen hesaplamaların hepsi sıra ölçeğiyle ölçülen değişkenler için de geçerlidir.

Bunlara ek olarak, kümelerin sırayla bir araya getirilmesiyle **birikimli sıklık**, **birikimli görelî sıklıklar** da hesaplanabilir. Sözgelimi ilk iki küme birleştirilerek, “İTÜ'yü beklediğimden kötü bulmayanlar” 3187 kişidir, bunların toplam içindeki payı % 54,7 kadardır diyebiliriz. İlk sayı birikimli sıklık, ikincisi birikimli görelî sıklık olarak adlandırılır.

Ayrıca büyüklük sırasına dizilmiş verilerin tam ortasına rastlayan gözlem değeri anlamında **ortanca** (medyan) adıyla anılan bir başka değer daha bulunabilir. Buna **ortalayan** da denmektedir (İpek 1988:139; Sınıksaran, 2001:26). Örneğimizdeki 5826 öğrenci, başlangıçtaki beklentilerine göre

İTÜ'yü beğenme düzeyleri dikkate alınarak sıralanırlarsa tam ortaya denk gelen 2913'üncü - 2914'üncü öğrencilerin İTÜ'yü beklentileriyle aynı düzeyde buldukları anlaşılır. Öyleyse bu dağılımın ortancası “beklediğim düzeyde” kümesidir.

Bunun gibi, sıraya dizilmiş öğrencilerin en üst ya da en alt dörtte bir noktasına rastlayanların beğenme düzeyleri de saptanabilir. Bu değerlere de **dördebölen** (kartil) adı verilebilir. Dörtte bir yerine herhangi bir başka oran geçirilerek benzer saptamalar yapılabilir. Bu konu ileride yeniden ele alınacaktır.

Bu ölçekle ölçülmüş verilerle dört işlem yapılamadığı için, **aritmetik ortalama** hesaplanamaz. Ad ölçeğinde verilen örnekte olduğu gibi burada da toplam öğrenci sayısının küme sayısına bölünmesiyle bulunacak olan değer, beğenme düzeyi değişkeninin değil, her beğenme düzeyindeki öğrenci sayısı değişkeninin ortalamasıdır.

Yukarıdaki dağılımı daha önce ya da daha sonra (eğer varsa) yapılmış benzeri bir araştırmanın sonuçlarıyla karşılaştırıp İTÜ öğrencilerinin beklentilerine göre İTÜ'yü beğenme düzeylerinin zaman içinde değişip değişmediğini görmek isteyebiliriz. Bu amaçla **sıra örneklem değeri** kullanan çeşitli **katsayısal olmayan sınamalar** kullanılabilir. Ayrıca bu iki dağılımdaki sıralanmaların benzerliğini anlayabilmek için **sıra ilişki** (korelasyon) **katsayısı** hesaplanabilir (Arıcı, 1988:6). Ancak bütün bu işlemler sırasında bazı ek varsayımların sağlanması gerektiği de unutulmamalıdır (Siegel, Castellan, 1988:27).

3.5 ARALIK ÖLÇEĞİ

Aralık ölçeği, ad ölçeğiyle sıra ölçeğinden daha üst düzeyde ölçüm yapabilen bir ölçektir. *Aralıklı ölçek* (Arıcı, 1988:16; Sencer, 1981) diye de anılır. Daha önce gördüğümüz iki ölçeğin bakışım, geçişlilik özelliklerini taşır. Bunlara ek olarak bu ölçek üzerindeki iki nokta arasındaki fark, anlamlı bir büyüklüktür. Eğer gözlem değerlerimizi büyüklük sırasına koyabiliyor, üstelik aralarındaki uzaklığın da belli bir birim cinsinden ne kadar olduğunu belirleyebiliyorsak aralık ölçeğiyle ölçüm yapıyoruz demektir.

Bu iş için herkesin üzerinde anlaşacağı ortak bir ölçü birimi gerektiği açıktır. Sıcaklık santigrat (C) ya da fahrenheit (F) ile ölçülür, ama fayda, toplumsal-ekonomik konum, zekâ ya da gönenc (refah) için böyle genel kabul görmüş ölçüler yoktur. Gerçi bir çok yazar (Babbie, 1992:142; Black, 1999:52; Bernard, 2000:41; Frankfort-Nachimas, Nachimas, 1992:155; Graziano, Raulin, 1993:79) zekâ ölçümlerini aralık ölçeğiyle ölçülmüş gibi kullansa da buna karşı çıkanlar şu örneği verir “Sözgelimi bir karıkocanın

gelirleri toplanarak anlamlı bir değer elde edilebilir ama bu iki kişinin zekâ düzeylerini gösteren IQ değerlerinin toplamı bir anlam ifade etmez” (Blalock, 1972:14-15). Dolayısıyla IQ’ların aralık ölçeğinde ölçülmüş olmaları en azından tartışmalıdır denebilir.

Burada önemli olan ölçülen değişkenin çeşitli değerleri arasındaki farkların eşit aralıklara anlamlı biçimde bölünüp bölünmediğidir. Eğer değişik toplumsal-ekonomik konumlara verilen değerlerin farklarını eşit aralıklı bir ölçü birimiyle ölçemiyorsak bu kavramı da ancak sıra ölçeği düzeyinde ölçebildiğimiz ortaya çıkar. Bu konuya az ileride yeniden döneceğiz.

Aralık ölçeğiyle ölçüm yapabilmek için bu ölçek üzerinde **isteğe bağlı** olarak belirlenmiş bir **sıfır noktasının** varlığı da gereklidir. Sözelimi sıcaklığı C ile ölçüyorsak sıfırı suyun donduğu noktaya koyarız. Suyun kaynadığı noktayla sıfır noktası arası da yüz eşit parçaya bölünür. Yok eğer ölçü birimiz F ise bu kez suyun donma noktasına 32, kaynama noktasına 212 değerleri verilerek bu ikisinin arası 180 eşit parçaya bölünür. 32’den geriye doğru gidilerek sıfır noktasının yeri de belirlenir. Sıfır noktası isteğe bağlı olarak seçilmiş bu gibi ölçü birimleriyle ölçülen iki değer arasındaki uzaklık, başka iki değer arasındaki uzaklıkla karşılaştırılabilir, ama bu uzaklıkların kendileri birbirleriyle oranlanamazlar.

Örnek olarak 40° C ile 30° C arasındaki sıcaklık farkı, 20° C ile 10° C arasındaki farka eşittir diyebiliriz, ancak 40° C sıcaklık, 20° C sıcaklığın iki katıdır demek doğru olmaz. Bunu anlayabilmek için 40° ile 20° C sıcaklıklarının sırasıyla 104°, 68° F ile de gösterilebileceğini anımsamak yeter. 40, 20’nin iki katıdır ama 104, 68’in iki katı değildir. Gerek C, gerek F ile yapılan ölçümlerin geçerli olduğunu bildiğimize göre buradan şu sonucu çıkarabiliriz. Böyle bir oranlama yapmak doğru olamaz.

Yalnız bir noktayı açıklıkla belirtmek gerekir. Gözlem değerleri birbiriyle oranlanamaz ama iki değer arasındaki fark, başka bir farka oranlanabilir. $26 - 20 = 6^{\circ} \text{ C}$ farkı, $20 - 18 = 2^{\circ} \text{ C}$ farkına bölünebilir. Böylece 6° C sıcaklık düşüşü, 2° C sıcaklık düşüşünün 3 katıdır denebilir. Aynı işlemi F ile yinelemek okuyucuya bırakılmıştır.

C ile yapılan ölçüme X , F ile yapılan da Y desek, X ile Y arasında şöyle bir **doğrusal dönüştürme** yapılabildiğini görürüz:

$$Y = 32 + 1,8X$$

Bu eşitliği, aralık ölçeğiyle, fakat hem sıfır noktaları, hem birimleri farklı olan iki ayrı ölçü ile ölçülmüş bütün değişkenlere genellersek:

$$Y = a + bX$$

yazabiliriz. Buradaki “ a ” sabit sayısının varlığı, herhangi iki X değeri arasındaki oranın, bu X değerlerine karşılık iki Y değeri arasındaki orana

eşit olmasını engeller. X 'e X_1, X_2 gibi iki değer verelim. X_1, X_2 'nin iki katı, yani $X_1 = 2X_2$ olsun. Yukarıdaki eşitlikte bunları yerine koyarak Y değerlerini sırasıyla:

$$Y_1 = a + bX_1 \qquad Y_2 = a + bX_2$$

buluruz. Bunlardan ilkinde X_1 yerine $2X_2$ konursa

$$Y_1 = a + 2bX_2$$

elde edilir. Bunun da Y_2 'nin iki katına, yani $2a + 2bX_2$ 'ye eşit olmadığı açıktır.

Şimdi de bu ölçümlerin farklarını alalım.

$$\begin{aligned} \Delta X &= X_1 - X_2 = 2X_2 - X_2 = X_2 \\ \Delta Y &= Y_1 - Y_2 = a + 2bX_2 - a - bX_2 = bX_2 \end{aligned}$$

Bu iki fark birbirine oranlandığında $\Delta Y / \Delta X = b$ bulunur. Bu da anlamlı bir orandır. “ a ”nın varlığı bu bölme işlemini engellemez.

Benzeri durumlar miladi, rumi, hicri, vb. takvimler ya da “standart puan”lar gibi ölçekler için de geçerlidir. Sıfır noktası takvimlerde belli olaylara göre, standart puanlarda da ortalamaya bağlı olarak belirlenir.

Ad ya da sıra ölçeğiyle ölçülmüş değişkenler **nitel** değerler aldıkları halde aralık ölçeğiyle ölçülenlerin değerleri **nicel** olur. Dolayısıyla da istatistik işlemlere daha elverişlidirler. Bu nedenle çeşitli toplumsal ya da davranışsal bilimlerde geliştirilmiş sıra ölçeklerini aralık ölçeğine çevirebilmek için çalışmalar yapılmaktadır (Bernard, 2000: 285-312, Sekaran, 2000:187-203)

Aralık ölçeğiyle ölçülmüş değişken değerleriyle **aritmetik ortalama**, **standart sapma** gibi değerler hesaplanabilir, çünkü bu hesaplamalar sırasında değerler arasında bir oranlama yapılmaz. Ancak standart sapmanın aritmetik ortalamaya oranını gösteren **değişkenlik katsayısı** hesaplanamaz, çünkü burada aralık ölçeği üzerindeki aritmetik ortalama bölme işleminin paydasında yer almaktadır. Bu ölçek üzerindeki değerlere (kök, üs, logaritma gibi) türlü **dönüştürmeler** uygulanabilir, dolayısıyla **geometrik**, **ters** (harmonik), **kareli ortalama** hesaplanabilir. Böyle iki değişken arasında **ilişki** (korelasyon) **katsayısı** bulunabilir, gerekli varsayımlar sağlanıyorsa **t sınaması** ya da **F sınaması** yapılabilir. Şurası unutulmamalıdır ki, normal dağılım varsayımı gerektiren bu gibi sınamalarda, aralık ölçeğiyle ölçme **gerekli** bir koşuldur ama **yeterli değildir** (Siegel, Castellan, 1988:30).

Aralık ölçeğine bir örnek olmak üzere iki bölgemizde beşer merkezde ölçülen hava sıcaklıklarının yıllık ortalamaları Çizelge 3.3'te verilmiştir.

ÇİZELGE 3.3 Bazı merkezlerin yıllık ortalama hava sıcaklıkları (C)

Merkez	Marmara	Merkez	Akdeniz
Balıkesir	14,5	Adana	18,8
Bursa	14,6	Anamur	19,2
Çanakkale	14,8	Antalya	18,1
Edirne	13,4	Islahiye	16,6
Göztepe (İst.)	14,0	Muğla	14,9

DİE (2002a:16)

Burada gözlem birimlerimiz meteoroloji merkezleri, değişkenimiz yıllık ortalama hava sıcaklıkları; gözlem değerlerimiz ise C ile ölçülmüş hava sıcaklığı değerleridir. Çizelgedeki sayılar, daha önceki çizelgelerdekinden farklı olarak sıklıkları değil ölçümlerin değerlerini göstermektedir. Daha önceki değişkenler nitel olduklarından aldıkları değerler sayılarla gösterilememiş, ancak aynı kümeye düşenlerin kaçar tane oldukları belirtilmekle yetinilmişti. Bu örnekte ise değişken nicel olduğundan sayısal değerler alabilmektedir.

Kullanılan ölçek aralık ölçeği olduğundan gözlem değerlerini büyüklük sırasına koymakla yetinmeyebilir, aralarındaki farkların ne kadar olduğunu da söyleyebiliriz. Daha ileri giderek iki bölgemizin ayrı ayrı ortalamalarını da hesaplayabiliriz. Marmara ortalaması yaklaşık 14,3°, Akdeniz ortalaması da 17,6° C'dir. Demek ki Akdeniz bölgesinin yıllık hava sıcaklığı ortalaması, Marmara Bölgesi'nden 3,3° C daha yüksektir.

Adana ile Akdeniz ortalaması farkı 1,2° C iken, Anamur ile aynı ortalama arasında 1,6° C fark vardır. $1,2 / 1,6 = 0,75$ işlemi anlamlı bir sonuç verir. Böylece Adana'nın Akdeniz ortalamasından sıcaklık farkının, Anamur'ununkinin dörtte üçü kadar olduğunu anlarız. Dikkatli bir okuyucu son bölme işleminin *farklar* arasında olduğunu gözden kaçırmamıştır.

Bu verilerle daha pek çok işlem yapılabilirse de henüz yöntemleri öğrenmemiş olduğumuzdan bu işleri daha sonraya bırakıyoruz.

3.6 ORAN ÖLÇEĞİ

Son olarak **oran ölçeği**ni ele alacağız. *Oranlı ölçek* Arıcı, 1988:17), *oransal ölçek* (Orhunbilge vö., 1997; Sencer, 1981) gibi adlarla da anılan bu ölçek yukarıda gördüğümüz ölçeklerin bütün özelliklerini taşır, üstelik hepsinden daha üst düzeydedir. Çünkü ölçtüğümüz değişkenin aldığı değerlerin aralarındaki farkı belirlemekle yetinmez, bu değerlerin birbirlerine göre

oranlarını da bulmamıza olanak tanır. Bu nedenle yalnız nicel değişkenlerin değerlerini ölçer. Çok yaygın olarak kullanılan kg, m, YTL gibi ölçüler hep bu ölçeğin örnekleridir. Aralık ölçeğinden farkı, sıfır noktasının, **isteğe bağlı olmadan** tanımlanmasıdır. Sözelimi hiç satış yapılmamışsa “satış miktarı” değişkeni doğal olarak sıfırdır. Kelvin ölçeğiyle sıcaklık ölçüldüğünde de mutlak sıfır vardır.

Aralık ölçeğinde olduğu gibi oran ölçeğinde de değişik ölçü birimleri arasında doğrusal dönüştürmeler yapılabilir. Sözelimi kg-lb gibi iki ağırlık ölçüsü ya da TL-€ gibi iki para birimi arasında dönüştürmelere sık sık rastlarız. Sıfır noktası her iki ölçü birimi için de aynı ($0 \text{ kg} = 0 \text{ lb}$ ya da $0 \text{ TL} = 0 \text{ €}$) olduğundan ölçülen değerler arasında oranlama da yapılabilir. $1 \text{ €} = 1,8 \text{ milyon TL}$ iken (siz bu satırları okurken 1 € kaç YTL olmuştur bilemem), 4 € , 1 € ’nin dört katı olduğu gibi, bunların TL karşılıkları olan $7,2 \text{ milyon TL}$ de $1,8 \text{ milyon TL}$ ’nin 4 katıdır. Öyleyse iki ölçü biriminden birine X , ötekine Y dersek, iki ölçü birimi arasında

$$Y = c X$$

biçiminde bir dönüştürme uygulandığı açıktır. Bu ilişkideki “ c ”, kg-lb için $0,453$; TL-€ için (şu sıralar) $1,8$ milyondur. Buradaki dönüştürme eşitliğinin aralık ölçeğindeki farkı “ a ” gibi bir sabit terim içermemesidir. Aralık ölçeğinde bu “ a ” yüzünden oranlama yapamadığımızı söylemiştik, oysa oran ölçeğindeki dönüştürme eşitliğinde böyle bir engel olmadığından değişken değerleri arasında rahatlıkla oranlama yapabiliriz. Orada yaptığımız gibi X ’e X_1, X_2 gibi iki değer verelim. X_1, X_2 ’nin iki katı, yani $X_1 = 2X_2$ olsun. Yukarıdaki eşitlikte bunları yerine koyarsak Y değerlerini sırasıyla

$$Y_1 = cX_1 \quad Y_2 = cX_2$$

elde ederiz. Bunların ilkinde X_1 yerine $2X_2$ konursa

$$Y_1 = 2cX_2$$

bulunur. Görüldüğü gibi $Y_1 = 2Y_2$ çıkmıştır. Buradan Y ’ler arasındaki oranın X ’ler arasındaki orana eşit olduğunu görebilir, dolayısıyla değişken değerleri arasında bölme işlemi uygulayabiliriz. Öyleyse oran ölçeğinde ölçülmüş değişkenlerle (daha önceki ölçeklerle ölçülen değişkenlerle hesaplayabildiklerimize ek olarak) **değişkenlik katsayısı, çarpıklık, basıklık katsayıları** gibi değerler hesaplanabilir, gerekli varsayımların sağlanması koşuluyla her türlü katsayısal sına uygulanabilir. Birden çok değişken için **bağlanım** (regresyon), **ilişki** (korelasyon), **varyans çözümlemesi** gibi yöntemler kullanılabilir.

Oran ölçeğine örnek vermek üzere iki çizelge düzenlenmiştir. Çizelge 3.4, cam işkolunda çalışan, Kristal-İş sendikasına üye işçilerin 1990’da

“yaşadıkları haneyi paylaştıkları kişi sayıları”nı, Çizelge 3.5, Kafkasya’daki, Orta Asya’daki bazı ülkelerle Türkiye’nin iki ayrı yılda “kişi başına gelir düzeyleri”ni göstermektedir.

Çizelge 3.4’te gözlem birimlerinin ilgilendiğimiz niteliği, bir başka deyişle ele alınan değişken, “hanedeki kişi sayısı”dır. Kişilerin tamsayı olmaları gereği açıktır. Bir hanede yaşayan 1,7 kişiden söz edilemez. Yalnız tamsayı değerler alabilen bu tür değişkenlere **kesikli değişken** denmektedir.

Buna karşılık Çizelge 3.5’teki kişi başına gelir değişkeni **sürekli değişkendir**, duyarlı bir biçimde ölçülebilirse ille de tamsayı değer alması gerekmez. Ama genellikle ölçüm çok duyarlı olmayabilir. Sözelimi bu örnekteki eski Sovyetler Birliği ülkelerinin kişi başına gelir değerlerine güvenilirliğin öteki ülkelerinkine göre daha az olduğu verilerin alındığı kaynağa da belirtilmektedir.

Sürekli değişkenler çoğunlukla yuvarlatılarak kullanılırlar. Bunun bir nedeni ölçme sırasında çok duyarlı davranılmamış olması, başka bir nedeni de algılamayı kolaylaştırmak için, önemsiz ayrıntı sayılan bölümlerin yuvarlatılarak duyarlılığı daha az sayıların kullanılmasıdır. Sözelimi Çizelge 4.3’teki hava sıcaklığı değişkeni sürekli bir değişkendir, onun değerleri de genellikle yuvarlatılmıştır.

ÇİZELGE 3.4 Kristal-İş Sendikası’na üye işçilerin yaşadıkları haneyi paylaştıkları kişi sayısı

Kişi sayısı	Hane sayısı
1	104
2	660
3	1.464
4	2.669
5	2.145
6	1.244
7	536
8	354
≥ 9	402
Toplam	9.560

Kristal-İş (1990:17)

ÇİZELGE 3.5 Kafkasya’daki, Orta Asya’daki bazı ülkelerle Türkiye’nin kişi başına gelirleri

Ülke	Kişi başına gelir (\$)	
	1991	2001
Tacikistan	1.050	170
Özbekistan	1.350	550
Kırgızistan	1.550	280
Gürcistan	1.640	620
Azarbeycan	1.670	650
Türkmenistan	1.700	950
Türkiye	1.780	2.540
Ermenistan	2.150	560
Kazakistan	2.470	1.360

World Bank (1993:238-39, 2003:234-35)

Öyleyse aralık ya da oran ölçeğiyle ölçülmüş değişkenler **nicel değişkenler**dir, **sürekli-kesikli** olmak üzere ikiye ayrılabilirler. Sürekli değişkenlerin değerleri çoğu zaman yuvarlatılır. Yuvarlatma kuralları aşağıda ele alınacaktır. İstatistik incelemelerimizi yaparken yalnızca sürekli değişkenleri yuvarlatmakla yetinmeyiz. Kesikli değişkenler de yuvarlatılabilir. Sözgelimi ülke nüfusu genellikle yuvarlatılır. Ayrıca sıklık sayılarını da yuvarlatabiliriz. Gerçekten de Çizelge 3.1'deki sıklık değerleri kesikli olmalarına karşın bin kişi olarak yuvarlatılmışlardır.

3.7 YÖNTEM SEÇİMİNDE ÖLÇEKLERİN ÖNEMİ

Ölçemediğin şeyi yönetemezsin
Peter Drucker

Görüldüğü gibi ölçekler arasında bir düzey sıralaması vardır. Ad ölçeği en alt düzeyde ölçüm yapar. Gözlem değerlerini çeşitli kümelere ayırmakta kullanılır. Öteki bütün ölçekler ad ölçeğinin içerdiği bakışım, geçişlilik özelliklerine ek olarak bir de **sıra** özelliği taşırlar, yani kümeleri kendi aralarında sıraya dizerler. Aralık ölçeği bunların yanı sıra kümelerin ya da tekil gözlem değerlerinin arasındaki **farkları** karşılaştırabilir. Oran ölçeğiyle gözlem değerleri arasında **oran** uygulayabilme özelliği de taşır. Bundan dolayıdır ki araştırmalarda elden geldiğince oran ölçeği kullanılmak istenir, ancak bu istek her zaman yerine gelmeyebilir. Böyle durumlarda aralık ölçeği, o da olmazsa geriye doğru gidilerek sıra ölçeği kullanılır, en olanaksız durumda ad ölçeğiyle iş yapmak zorunda kalınır.

Şunu unutmamak gerekir ki eğer elimizdeki veriler oran ölçeğiyle ölçülmüşlerse, bu verileri daha düşük ölçme düzeyindeki verilermiş gibi kullanabiliriz. Diyelim ki A ürününün son bir haftadaki satışı 3 ton iken B ürünününki 2 ton olsun. Bunu A, B'den çok satmıştır biçiminde de dile getirebiliriz. Ama o zaman aradaki satış farkının 1 ton olduğu, A'nın satışının B'ninkinin 1.5 katı olduğu bilgilerini hiç değerlendirmemiş oluruz.

Peki, bunun tersini yapamaz mıyız? Elimizdeki veriler daha düşük bir ölçme düzeyiyle elde edilmiş olmalarına karşın, sanki daha yüksek düzeyde bir ölçek kullanılmış gibi davranıp daha gelişkin istatistik yöntemlerinden yararlanamaz mıyız? Gerçek yaşamda, ölçek kavramını iyi değerlendiremeyen bazı araştırmacıların zaman zaman yanlışlıkla bu yola başvurdıkları görülebiliyor. Sözgelimi sıra ölçeğiyle ölçülerek sıralı kümelere ayrılmış gözlemlere “ad” olarak verilmiş “sayı”ları aritmetik işlemlerde kullanıp ortalama, standart sapma hesaplayan, bunlara *t* sınaması uygulayanlara rastlanabiliyor. Ancak belli bir istatistik tekniğinin kullanılabilmesi için

verilerin de belli bir ölçüm düzeyinin altındaki bir ölçekle ölçülmemeleri gerekir. Verilerin ölçüm düzeyiyle kullandığı istatistik yönteminin ölçüm düzeyinin tutarlı olup olmadığını saptama sorumluluğu araştırmacıya düşer (Blalock, 1972:17). Bunu yapmazsa vardığı sonuçların güvenilirliği de tartışmalı olur.

Sözgelimi gözlem birimlerimizi toplum bilimlerinde sık kullanılan bir kavram olan “toplumsal-ekonomik konum”larına göre kümelere ayırmak; bu amaçla da gelir, meslek, eğitim düzeyi, oturulan semt gibi bazı değişkenlerden yararlanmak istediğimizi düşünelim. Dikkat edilirse gelir değişkeni (sayısal bilgi varsa) oran ölçeğiyle, eğitim düzeyi değişkeni genellikle sıra ölçeğiyle, meslek, semt değişkenleri de ad ölçeğiyle ölçülebirlirler. Ancak toplumsal-ekonomik konum belirlenirken son iki değişkenin kendi aralarında sıralanabileceğinin varsayıldığı görülüyor. Belli meslekler, belli semtler bir kümeye alınırken, başka bazı meslekler, başka bazı semtler daha üst ya da daha alt düzeydeki başka bir kümeye atanabiliyor. Bu sıralamalar genel kabul gördüğü sürece böyle bir varsayıma karşı çıkılamaz. Yine de bu tür sıralamalarda özen gösterilmesi gereği ortadadır.

Bir an için mesleklerin de, semtlerin de sıra ölçeğine uyabileceğini düşünelim. Gözlem birimlerimizden A, bu değişkenlerin ikisinde de B'den üstün olsun. Bu durumda A'nın toplumsal konumunun B'ninkinden üstün olduğu açıktır. Ama işler her zaman bu kadar yolunda gitmez. A bir değişkene göre B'den üstün iken, başka bir değişkene göre de durum tam tersi olabilir. Bu durumda A ile B'den hangisinin konumunu daha üstün sayacağız. Bazı kimselerin yaptıkları gibi, sözgelimi çeşitli mesleklere ya da semtlere puan vererek bu engeli aşabilir miyiz? Bu puanları vermek demek, sıra ölçeğiyle ölçüldüğünü az önce zoraki biçimde benimsediğimiz bu değişkenlerin şimdi de aralık ölçeğiyle ölçülmüş olmalarını kabul etmemiz anlamına gelir. Buna hakkımız var mı?

Her şeye göğüs germek üzere gözümüzü kapayıp derin bir soluk alarak bu varsayımı da yaptık diyelim. Toplumsal-ekonomik konumu belirleyebilmek için bu kez de meslek-semt değişkenlerinin eşit ağırlık taşıyıp taşımadıklarını düşünmemiz gerekecektir. Mesleği mi yoksa semti mi bu bağlamda daha önemli saymalıyız. Birini ötekinden daha önemli saysak bile bunu sayısal olarak dile getirmemiz söz konusu olacak, o zaman da bu önem farkının ne oranda olduğuna karar vereceğiz demektir. Kişisel yargılarımızı bir yana bırakabilirsek bile, hem çeşitli mesleklere, semtlere kendi aralarında puan verirken, hem de meslek-semt arasındaki önem oranını sayısallaştırırken, başka bir deyişle bu iki değişkeni tartılandırırken, herkesçe kabul edilebilecek yargılara varmamız zorunlu görünüyor. Bundan kesinlikle emin olmadıkça **tartılı aritmetik ortalama** ya da benzeri bir

yöntem kullanarak bir kimsenin toplumsal-ekonomik konumunu belirlemeye kalkışmak pek çok sakıncalar içerir.

Bu örneği istediğiniz gibi değiştirebilir, siz de başka örnekler türetebilirsiniz. Bir ürünün çeşitli markalarını, fiyat, renk, çekicilik, dayanıklılık, satış sonrası hizmet gibi değişik özelliklere göre bir beğenme sıralamasına dizerken; bir ülkedeki bankaları öz sermaye / aktifler, kâr / öz sermaye, mevduat / aktifler gibi ölçütlere göre başarı sırasına koyarken hep böyle yanıışlara düşme tehlikesiyle karşı karşıya kalırız.

İstatistikte gerek bu tür, gerek başka tür yanıışlardan kaçınmanın yolu her zaman **sağduyumuza** başvurmak olmalıdır. Kullandığımız yöntemi ya da vardığımız sonucu hep sağduyumuzun süzgecinden geçirmekte yarar vardır.

Ölçeğin türü çoğu zaman kullanılacak istatistik yöntemini belirler. Ad ya da sıra ölçeğiyle ölçülmüş nitel değişkenlerle işlem yaparken genellikle **katsayısal olmayan** (nonparametrik) yöntemler uygun olur. Aralık ya da oran ölçeğine uygun değişkenler için ise çoğu zaman **katsayısal (parametrik)** yöntemler elverişlidir (Cooper, 2003:221-229).

Okuma önerileri

- Babbie, E. (1992) *The Practice of Social Research*, 6th ed. Wadsworth
 Bernard, H. R. (2000) *Social Research Methods*, Sage (285-316).
 Blalock, Jr., H.M. (1972) *Social Statistics*, 2.b., McGraw-Hill (11-20).
 Cooper, D. R., Schindler, P. S. (2003) *Business Research Methods*, McGraw-Hill (221-229).
 Frankfort-Nachimas, C., Nachimas, D. (1992) *Research Methods in the Social Sciences*, 4th ed. St. Martin's.
 Moore, D. S. (1979) *Statistics: Concepts and Controversies*, Freeman (100-112).
 Sekaran, U. (2000) *Research Methods for Business*, 3rd. ed. Wiley (187-203).
 Siegel, S., Castellan, Jr., N. J. (1988) *Nonparametric Statistics for the Behavioral Sciences*, 2. b. McGraw-Hill (23-33).

ALIŞTIRMALAR

- 3.1** Piyasaya sürülmesi düşünülen yeni bir oto lastiği markası için pazar araştırması yapılmış, bu amaçla seçilen bir örnekleme giren oto sürücülerine çeşitli sorular arasında
- hangi marka lastiği,
 - kaç aydır kullandığı,
 - bu lastiklerle kaç km yaptığı,

- çok memnunum, memnunum, idare eder, memnun değilim, hiç memnun değilim seçeneklerinden hangisini seçtiği,
- lastik yenilerken yine aynı markayı mı alacağı sorulmuştur.
- a) Bu sorularla hakkında bilgi derlenmek istenen değişkenlerin adlarını açıkça yazın.
- b) Bu değişkenlerden her birinin alabileceği değerler nelerdir?
- c) Her bir değişken hangi ölçekle ölçülebilir?
- d) Hangi değişken(ler) için en sık değeri bulabilirsiniz?
- e) Hangi değişken(ler) için ortancayı hesaplayabilirsiniz?
- f) Hangi değişken(ler) için aritmetik ortalamayı bulabilirsiniz?
- g) Hangi değişken(ler)in kendi değerleri birbirine oranlanabilir?

3.2 Aşağıdaki değişkenlerden hangileri sürekli, hangileri kesiklidir? Her biri hangi ölçekle ölçülmelidir?

- a) Çeşitli binek otolarının silindir hacimleri,
- b) Bir fabrikadaki işçilerin çıplak saat ücretleri,
- c) Bir fabrikadaki işçilerin net aylık ücretleri,
- d) Son yılın aylık fiyat artış (enflasyon) oranları,
- e) Aynı ürünü satan firmaların pazar payları,
- f) Aynı ürünü satan firmaların pazar payları bakımından sıraları,
- g) Türkiye'deki hanehalklarının yıllık gelir düzeyleri,
- h) Kastamonu'da hanehalklarındaki kişi sayıları,
- i) Bir belediye otobüsünün günlük sefer sayısı,
- j) Belediye otobüs sürücülerinin ehliyet belgelerinin yaşı,
- k) Turistik otellerin yıldız sayıları,
- l) Bir "puan cetveli"ndeki takımların sıra numaraları,
- m) Son 20 yılda gol krallarının maç başına attıkları gol sayıları,
- n) Öğrencilerin sınav notları,
- o) Öğrencilerin sınav notlarının ortalaması.

3.3 İstanbul Kültür ve Sanat Vakfı'nca düzenlenen Uluslararası İstanbul Film Festivali'nin ortalama izleyicisi,

- 26-27 yaşlarında,
- bekâr,
- yüksekokul mezunu ya da öğrencisi,
- (1990 yılı başı fiyatlarıyla) 550 bin lira aylık gelirli,
- İngilizce'yi iyi derecede, ikinci bir yabancı dili (Fransızca ya da Almanca'yı) orta derecede bilen bir kişi olarak tanımlanmıştır. (*Cumhuriyet gazetesi*, 3.2.1990),
- a) Bu haberde hangi değişkenlere ilişkin bilgi veriliyor? Bunların her biri hangi ölçekle ölçülebilir?
- b) Değişkenlerin her biri hangi hangi değişik değerleri alabilir?

- c) Hangi değişken(ler) için ensıkdeğer, ortanca, aritmetik ortalama hesaplanabilir?
- d) Hangi değişken(ler)in kendi değişik değerleri arasında oranlama yapılabilir?
- e) Medeni durum değişkeninin “bekâr” değeri, sizce ensıkdeğer, ortanca, aritmetik ortalamadan hangisidir? Neden?

3.4 TÜBİTAKın çıkardığı Bilim ve Teknik Dergisi’nin okuyucu anketi sonuçları açıklamasında şöyle denmektedir. (TÜBİTAK, 1995:81): “Anketimizi yanıtlayan 1.727 kişinin % 67,3’ünü erkekler, % 32,5’ini ise kadınlar oluşturuyor. Yaş ortalaması 22 olan okuyucularımızın % 83,7’si bir yabancı dil biliyor. En genci 10, en yaşlısı 80 yaşında olan okurlarımızın % 42,7’si henüz öğrenci.”

- a) Burada hangi değişkenler ölçülmeye çalışılmıştır?
- b) Her bir değişken için hangi ölçek kullanılabilir?
- c) Değişkenlerin her biri hangi değişik değerleri alabilir?
- d) Hangi değişken(ler) için ensıkdeğer, ortanca, aritmetik ortalama hesaplanabilir?
- e) Hangi değişken(ler)in herhangi iki değeri arasında oranlama yapılabilir?
- f) Haberde geçen % değerlerine genel olarak ne ad verebiliriz?
- g) Kadınlarla erkeklerin toplamı % 100,0 etmiyor. Aradaki farkı nasıl açıklarsınız? Bu kapsayıcılık özelliğinin olmadığı anlamına gelir mi?

3.5 2000 Genel Nüfus Sayımı sonuçlarına göre bölgelerimizin nüfusları aşağıda gösterilmiştir.

- a) Buradaki değişkenin adını açıkça yazınız. Bu değişken hangi ölçekte ölçülmüştür?
- b) Bu verilerle görelî sıklıklar, birikimli görelî sıklıklar, ensıkdeğer, ortanca, aritmetik ortalama bulunabilir mi? Açıklayınız.

Bölgeler	Nüfus (bin kişi)
Marmara	17.354
Ege	8.951
Akdeniz	8.717
İç Anadolu	11.636
Karadeniz	8.439
Doğu Anadolu	6.142
Güneydoğu Anadolu	6.613
Toplam	67.853

Yuvarlatma nedeniyle toplam tutmamaktadır.
DİE (2002a:46)

- 3.6 Aşağıdaki çizelgede en az 72 yıldır gözlem yapılan meteoroloji istasyonlarında yıllık ortalama yağışlı gün sayısı gösterilmiştir.

İstasyon	Yağışlı gün sayısı	İstasyon	Yağışlı gün sayısı
Adana	76,5	Eskişehir	107,8
Afyon	106,5	Göztepe (İst.)	123,8
Ankara	103,6	Isparta	101,6
Balıkesir	95,5	Kırşehir	92,2
Bolu	137,7	Konya	81,5
Bursa	113,5	Kütahya	117,1
Edirne	107,1	Muğla	98,2
Erzurum	120,8	Sivas	110,4

DİE (2002a:26)

- a) Buradaki değişkenin adını açıkça yazınız.
b) Bu değişken hangi ölçekte ölçülmüştür?
c) Bu verilerle göreceli sıklıklar, birikimli göreceli sıklıklar, ensıkdeğer, ortanca, aritmetik ortalama bulunabilir mi? Açıklayınız.
- 3.7 Bir gazetede *Sayılarla kadınlar* başlığıyla şu çizelge verilmiştir:

Hayvan sahibi olarak kadın	Yüzde
2 + büyük baş hayvan	16,3
2 + küçük baş hayvan	9,6
2 + kanatlı hayvan	22,6
2 + arı kovanı	2,9
2 + İkiden fazla anlamındadır.	

Cumhuriyet Pazar eki, 9 Mart 2003

- a) Burada ölçülen değişkenler nelerdir?
b) Bu değişkenler hangi ölçeklerle ölçülmüştür?
c) “Yüzde” başlığı altında gösterilen değerler toplanabilir mi? Neden?
d) Sizce bunlar neyin yüzdesidir? Yani gözlemlerin tamamını içeren topluluk sizce nasıl tanımlanır?
- 3.8 Çizelge 3.3'teki verilerle her iki bölge için ayrı ayrı ortanca değerleri bulunup karşılaştırılabilir mi?
- 3.9 Çizelge 3.4'teki veriler için yüzde sıklık değerleri, ensıkdeğer, ortanca hesaplanabilir mi?
- 3.10 Çizelge 3.5'e dönelim.
- a) Buradaki veriler için ortanca bulunabilir mi?
b) Türkiye dışındaki ülkelerin gelirlerinde on yılda düşüşler olmuştur. Bunun nedenleri (her iki yıldaki dolar kurunun belirlenme yöntemleri, on

yılda ortaya çıkan enflasyon, gerçek yoksullaşma, vb.) üzerinde düşünün.

3.11 Bir gazetede *Vatandaşlar cezadan kaçtı* başlığıyla verilen bir haberde yer alan aşağıdaki çizelgeyi inceleyin.

- Buradaki değişkenler nelerdir? Her biri kesikli mi yoksa sürekli midir? Hangi ölçekte ölçülmüşlerdir? Yuvarlatma yapılmış mıdır?
- Hangi yılda en çok ceza kesilmiştir (ya da ödenmiştir)? Bu soruya daha anlamlı bir yanıt vermek için bu yıllarda enflasyonun ne kadar arttığını bilmemiz gerektiğini düşünüyor musunuz? Neden?

Yıllar	Kesilen ceza (milyar TL)	Ödenen ceza (milyar TL)
1993	8.998	3.428
1994	22.232	10.172
1995	39.111	18.755
1996	61.730	28.040
1997	134.971	71.570
1998	283.828	158.175
1999	502.536	274.611
2000	827.774	447.114
2001	1.757.307	927.945
2002	2.795.862	1.391.076

Radikal gazetesi, 6 Mart 2003

- Kesilen cezaların yüzde kaçının ödendiğini gösteren bir değişken tanımlayalım. Hangi yılda bu değişken en yüksek değerini alır? Bu soruyu yanıtlarken enflasyona ilişkin bilgi gerekli olabilir mi? Neden?
- Bu çizelgeye ödenen cezanın kesilen ceza içindeki payını gösteren bir sütun eklemek işe yarar mı?

3.12 Aşağıdaki çizelgede AB ülkeleriyle Türkiye’de 2000 yılında beş yaşından küçüklerin binde ölüm oranları verilmiştir:

Ülke	%	Ülke	%	Ülke	%
Almanya	6	Hollanda	7	Lüksemburg	6
Avusturya	6	İngiltere	7	Portekiz	8
Belçika	7	İrlanda	7	Yunanistan	8
Danimarka	6	İspanya	6	Türkiye	43
Finlandiya	5	İsveç	4		
Fransa	6	İtalya	7		

World Bank, 2003:234-235.

- Buradaki değişken hangi ölçekte ölçülmüştür?
- AB ülkeleri için en sık rastlanan değer (ensıkdeğer) kaçtır?
- İsveç'in en düşük değeriyle ensıkdeğer arasındaki fark kaçtır?
- Portekiz ya da Yunanistan'ın en yüksek değeriyle ensıkdeğer arasındaki fark kaçtır?
- Türkiye'nin değeriyle ensıkdeğer arasındaki fark kaçtır?
- AB ülkeleri için ortanca kaçtır?
- (c), (d), (e)'deki soruları ortanca için de yanıtlayınız.
- Bulgularınızı nasıl yorumluyorsunuz?
- Türkiye'deki bebek ölüm oranı sizce aşırı uçta bir değer midir?

3.13 1987 Genel Seçimleri öncesi siyasi parti önderlerine çeşitli bakımlardan verilen notları gösteren aşağıdaki çizelge bir gazetede yayımlanmıştır.

- Bu özelliklerin ölçülebileceği en yüksek ölçek hangisidir?
- Önderlere verilen bu notlar sizce hangi ölçekte ölçülmüştür?
- Çizelgenin son satırındaki ortalamanın hesaplanabilmesi için söz konusu özellikler hangi ölçekte ölçülmüş olmalıdır?
- Yukarıda (a)'da verdiğiniz yanıtla daha sonraki yanıtlarınız birbirini tutuyor mu? Tutmuyorsa burada ne gibi bir yanlışlık yapılmış olabilir?

	Özal	İnönü	Demirel	Ecevit	Erbakan	Türkeş	Edibali
Sevimsilik	7	7,5	5	6	5,5	2	5
Güvenilirlik	5,5	7	6	6	5	3	4,5
Konuşma	7,5	7	6	7	7	4	6
İnandırma	6	7	5,5	5,5	4,5	4	5
Liderlik	7	7	6,5	6	5,5	4,5	4
Ortalama	6,6	7,1	5,8	6,1	5,5	3,5	4,9

Milliyet gazetesi, 16 Ekim 1987.

3.14 İstanbul Büyükşehir Belediyesi APK Daire Başkanlığı'nın 1201 kişiyle yaptığı "İstanbul'da üst sosyoekonomik gruplar ve kentsel dayanışma" araştırmasına göre (*Radikal*, 12 Şubat 2004)

- Katılanların %1'i kendini çok zengin, %26'sı zengin, %72'si orta halli, %0,4'ü fakir olarak görüyor.
- %38'inin aylık geliri 3,5 – 5, %10'unun 5 – 7, %42'sinin 7 – 10, %7'sinin 11 – 20, %2,4'ünün 21 – 40 milyar TL arasında, %0,5'inin 40 milyar TL'den çok.
- %62,5'i sosyal statülerini kadere, %36'sı kendi gayretlerine bağlıyor.
- %22,4'ü İstanbul, %5,3'ü Sivas, %5,2'si Malatya, %4,1'i Mardin, %3,8'i Trabzon, %3,3'ü yurtdışı doğumlu.

- %77,3'ü İstanbul'a geldikten sonra gelir düzeylerinin arttığını, %14,2'si aynı kaldığını, %6,4'ü düştüğünü söylüyor.
 - %85,3'ü İstanbul dışında yaşamayı düşünmüyor.
 - Yurtdışına çıkanların oranı %42. Yurtdışına çıkış nedenlerinin %19'u gezi, %17'si iş. %21'i Almanya'ya, %9'u Britanya'ya, %8,7'si İtalya'ya, %8,5'i Fransa'ya, %6,4'ü ABD'ye, %4'ü Rusya'ya, %3,8'i Suudi Arabistan'a gitmiş.
 - %28'i kendini laik, %24'ü demokrat, %22'si milliyetçi, %14'ü Kemalist, %12'si sosyal demokrat, %12'si muhafazakâr, %8'i dindar, %7'si İslamcı olarak tanımlıyor.
 - %88'inin en az bir dairesi, %7,5'inin villası, %2,7'sinin yatı ya da teknesi var.
 - %84'ü malvarlığını kendi edinmiş.
 - %60'ı parasını faize, repoya, altına, borsaya, hazine bonosuna, koleksiyonculuğa yatırıyor.
 - %85'i herhangi bir sosyal örgüte üye değil. Üye olanların %40'ı spor kulüplerine, %30'u hemşire derneğine üye.
 - %61,7'si herhangi bir yardım kurulunun adını bilmiyor. Bildiğini söyleyenlerin %40'ı büyükşehir, %27'si Darülaceze, %14,4'ü Deniz Feneri, %10,8'i Çocuk Esirgeme Kurumu, %10'u ilçe belediyeleri, %6,6'sı aşevleri, %6'sı Kızılay adını veriyor.
 - %22,4'ü televizyonda haberleri, %13,7'si yerli dizileri, %8,9'u haber programları izliyor.
 - %21,4'ü Türk halk müziği, %20'si Türk sanat müziği dinliyor.
 - %60'ı en az bir yabancı dil bildiğini söylüyor. Bunların %84'ü İngilizce, %15'i Almanca, %8'i Rusça, %7,6'sı Fransızca, %7'si Arapça biliyor.
 - %46'sının hiç çocuğu yok. Ötekilerin %28'i bir, %41'i iki, %20'si üç, %6'sı dört çocuklu.
 - %54'ünün aylık mutfak masrafları 1 milyar TL'yi aşmazken, %25'inin 1 – 2 milyar TL arasında.
 - %64'ü yılda en az bir, % 16'sıysa yılda en az iki kez tatil yapıyor.
- a) Burada hangi değişkenler ölçülmeye çalışılmıştır?
 - b) Her bir değişken için hangi ölçek kullanılabilir?
 - c) Değişkenlerin her biri değişik değerleri alabilir?
 - d) Hangi değişken(ler) için *ESD*, *Ortc*, *AO* hesaplanabilir?
 - e) Hangi değişken(ler)'in herhangi iki değeri arasında oranlama yapılabilir?
 - f) Haberde geçen % değerlerine genel olarak ne ad verebiliriz?

3.15 Bir gazetenin kendi okurları arasında yaptığı bir araştırmaya göre (*Cumhuriyet*, 15 Şubat 2004) okurların özellikleri şöyledir:

- % 75'i erkek, % 25'i kadın,
- %12'si 15-20, %57'si 21-44 yaşları arasında, %31'i daha yaşlı,
- %66'sı Marmara-Ege, % 16'sı İç Anadolu, %13'ü Akdeniz, %5'i öbür bölgelerde yaşıyor,
- %15'i ilkokul, %17'si ortaokul, %37'si lise, %31'i üniversite bitirmiş,
- %54'ü A-B, %19'u C1, %12'si C2, % 15'i D-E sosyo-ekonomik düzeyde,
- %69'u evli, %27'si bekâr, %4'ü dul/ boşanmış,
- %55'i çalışmakta, (%20'si memur, %15'i işveren/yönetici, %15'i işçi) %21'i emekli, %13'ü öğrenci, %6'sı ev kadını, %5'i işsiz,
- %48'i son bir yılda tatile çıkmış (bunların %88'i yurt içinde),
- %4'ü gazeteyi 5 dakikadan kısa sürede, %10'u 15-20, %37'si 21-30, %32'si 31-60 dakika arasında, %17'si daha uzun sürede okuyor,
- %22'si yalnız *Cumhuriyet*, %36'sı *Cumhuriyet*'le birlikte *Hürriyet*, %29'u *Milliyet*, %16'sı *Sabah*, %14 *Posta*, %12 *Radikal* gazetelerini de okuyor,
- %80'i cep telefonu, %72'si ev, %68'i fotoğraf makinesi, %64'ü buzdolabı, %57'si bulaşık makinesi sahibi.

- a) Burada hangi değişkenler ölçülmeye çalışılmıştır?
- b) Her bir değişken için hangi ölçek kullanılabilir?
- c) Değişkenlerin her biri hangi değişik değerleri alabilir?
- d) Hangi değişken(ler) için *ESD*, *Ortc*, *AO* hesaplanabilir?
- e) Hangi değişken(ler)'in herhangi iki değeri arasında oranlama yapılabilir?
- f) Haberde geçen % değerlerine genel olarak ne ad verebiliriz?

3.16 İstanbul A.Ş. 10.000 metro yolcusuyla yaptığı sormacada şu bulguları elde etmiştir (*Gazete Metro*, 23-24 Şubat 2004).

- İstanbul metrosu yolcularının %74'ü erkek, %26'sı kadındır.
 - Bunların %57,6'sının arabası vardır, %42,4'ünün yoktur.
 - 15-25 yaş arasında olanlar %37,78; 26-35 arasındakiler %25,23; 36-45 arasındakiler %17,60; 46-55 arasındakiler %9,99; daha yaşlılar %9,40'tır.
 - %10'u ilkokulu, %12'si ortaokulu, % 33'ü liseyi, %38'i üniversiteyi bitirmiş, % 7'si lisansüstü eğitim görmüştür.
 - %50,16'sı metroyu işe, %21,07'si gezintiye, %11,64'ü okula, %9,60'ı alışverişe gitmek, %7,53'ü başka amaçlarla kullanmaktadır.
- a) Burada hangi değişkenler ölçülmeye çalışılmıştır?
 - b) Her bir değişken için hangi ölçek kullanılabilir?

- c) Değişkenlerin her biri hangi değişik değerleri alabilir?
- d) Hangi değişken(ler) için *ESD*, *Ortc*, *AO* hesaplanabilir?
- e) Hangi değişken(ler)in herhangi iki değeri arasında oranlama yapılabilir?
- f) Metroya binen 15 yaşından küçüklerin oranı kaçtır? Bunu nasıl açıklayabilirsiniz?
- g) Öbür oranlar virgülden sonra iki basamağa kadar belirtilirken cinsiyeti, eğitimi gösteren oranların tam sayıyla gösterilme nedeni sizce ne olabilir?

3.17 Bir habere göre (*Cumhuriyet*, 3 Ocak 2004) 741 erkek hükümlü ile yapılan bir araştırmaya göre bunların %52'si evli, %44'ü bekâr, %28'i okur-yazar. Yüksekokul bitirenlerin oranı %5,5. Düşük eğitimliler genellikle cinayet, yükseköğrenimliler sahtecilik suçları ndan hükümlü. %25'i esnaf, %18'i işçi. %24'ü birden çok kez cezaevine girmiş.

- a) Burada hangi değişkenler ölçülmeye çalışılmıştır?
- b) Her bir değişken için hangi ölçek kullanılabilir?
- c) Hangi değişken(ler) için *ESD*, *Ortc*, *AO* hesaplanabilir?
- d) Herhangi bir değişkenin iki değeri arasında oranlama yapılabilir mi?

Kesmeyelim de Yuvarlatalım mı?

Asmayalım da besleyelim mi?

Kenan Evren

4.1 SAYILAR NEDEN KESİLİR YA DA YUVARLATILIR?

Herhangi bir kişiye yaşını sorduğunuzda çoğunlukla yuvarlatılmış bir yanıt alırsınız. Kimse size yıldan daha ayrıntılı bir bilgi vermez. Belki küçükler üç buçuk yaşında olduklarını söyleyebilirler ama onların pek çoğu da kendini 4 yaşında göstermek ister. Hele yaşlılar “altmışıma merdiven dayadım” gibi yanıtlar verebilirler. Günlük yaşamımızda çoğu sayıyı yuvarlatarak kullanırız. Bir kilo kıyma aldığımızda bunun 997 ya da 1004 gr olması bizi pek rahatsız etmez.

Ad ölçeğini açıklarken kullandığımız Çizelge 3.1’de her bir medeni durumun sıklığının bin kişiye yuvarlatılarak kullanıldığını gördük. Bu çizelgede 2000 sayımında İstanbul’da yaşayan 11 yaşından büyüklerin toplam nüfusu 7.919 bin kişiydi. Aslında bu yaşlardaki kişi sayısı 7.918.500 ile 7.919.500 arasındadır. İstatistikte kullanılan verilerin çok önemli bir bölümüne böyle yuvarlatmalar uygulanmaktadır. Çünkü insan beyni bir sürü basamak içeren sayıları algılamakta zorlanır. Bir-iki basamaklı, en çok üç basamaklı sayıları daha kolay algılayabilir, aklımızda tutabiliriz. O yüzden Türkiye’nin 2000 yılı nüfusunu 68 ya da 70 milyon kişi, 2002 dış borç stoğunu 130 milyar \$ diye yuvarlatıp söylersek bu sayıların hem anlaşılması, hem bellekte saklanması daha kolay olur.

İkiden çok basamaklı sayıları karşılaştırmak bile sorun yaratabilir. Sözelimi 17,9 ile 35,2 sayısını karşılaştırmak istediğimizi düşünelim. Bırakın bunları birbirine bölmeyi, aralarındaki farkı bile kafadan hesaplamak zor olabilir. Ama ikişer basamak kullanılırsa 35 ile 18’in farkı 17’dir, ya da 35 bölü 18 yaklaşık 2’dir demek kolaydır.

Bu yuvarlatma ya ölçme-sayma yapılırken kendiliğinden olur ya da daha sonra bu kadar ayrıntının gereksiz olduğunu düşünen kullanıcı tarafın-

dan yapılır. Sözelimi elimizdeki cetvelle bir milimetreden daha duyarlı ölçümler yapamıyorsa ölçme sonuçlarımızı milimetreye yuvarlatarak dile getirebiliriz. Ama daha duyarlı bir ölçme aracıyla ölçümler daha az yuvarlatılabilir. Öte yandan ölçümlerimiz için santimetre düzeyinde bir duyarlılığın yeterli olduğu kanınsındaysak ölçme sonuçlarımızı buna göre kendimiz yuvarlatabiliriz.

Eğer veriler birbirlerine oranlanarak yeni veriler türetiliyorsa, o zaman yuvarlatma bir zorunluluk olarak karşımıza çıkabilir. Oran ölçeğini açıklarken kullandığımız “kişi başına gelir”¹ değişkeni toplam gelirin nüfusa bölünmesiyle bulunmuş bir değişkendir. Diyelim elimizde bir dizi yıl için Türkiye'nin milli gelir-nüfus verileri var. Bunları birbirine bölerek kişi başına gelir verileri üretmek istiyoruz. 2000 yılı “gayrisafi milli hasılası (GSMH)” 125.596.129 milyon TL (DİE, 2002a:643), 2000 yıl ortası nüfus tahmini de 67.469 bin kişi² ise (DİE, 2002a:44), kişi başına GSMH 1.861.538.321 TL bulunur. Burada yuvarlatma bir zorunluluk olarak karşımıza çıkar. (Siz bu satırları okurken paradan altı sıfır atılmışsa bu yazılanları anlamakta çekeceğiniz güçlük için bağışlanmamı dilerim.)

Bugün artık hesaplama işlemleri gitgide daha çok elektronik hesap makinası ya da bilgisayar kullanılarak yapılmakta, bunlarla bulunan işlem sonuçları da genellikle gereğinden daha duyarlı olarak elde edilmektedir. Dolayısıyla makinanın ya da bilgisayar programının özelliğine bağlı olarak sonuçlardaki basamak sayıları bir hayli yüksek olmaktadır. Ne yazık ki bu araçları kullanan bazı kimseler bulgularını başkalarına yazılı olarak aktarıırken aynı basamak sayısını korumaktadırlar. Oysa aşağıda göreceğimiz kurallara göre bu basamakların önemli bir bölümü yanıltıcı bir duyarlılık izlenimi yaratır. Bu nedenle, özellikle elektronik hesap makinası/bilgisayar kullananların yuvarlatma hatalarına karşı çok uyanık olmaları gerekmektedir.

Yukarıda bulduğumuz 1.861.538.321 TL kişi başına gelirin de böyle yanıltıcı bir duyarlılığı vardır. DİE (2002a:644)'ye göre bu sayı 1.851.412.612 TL'dir. İki hesap arasındaki fark 10 milyon TL kadardır. DİE hesabı yuvarlatmadan önceki sayılarla yaptığı için, milli gelir-nüfus sayılarının doğru olması koşuluyla, onun bulgusu daha güvenilir. Biz yuvarlatılmış sayılarla işlem yaptığımızdan bölmenin sonucu farklıdır. Öyleyse 2000 yılı kişi başına gelirini 1.9 milyar TL diye yuvarlarsak her iki hesap sonucu birbiriyle uyuşur.

¹ Bu örnekte “kişi başına gelir” ile “kişi başına GSMH” kavramları aynı anlamda kullanılacaktır.

² 2000 yılı nüfus sayımı sonucu 67.853 bin kişidir, ama bu sayı *sayım günündeki* nüfusu göstermektedir. Her nüfus sayımı arasındaki büyüme oranları kullanılarak *yıl ortası* nüfusları tahmin edilir.

Bu iki sayıyı yuvarlatmasak da fazlalıklarını budasak ne olur? Her iki bulguyu 1.8 milyar TL olarak elde ederiz. Burada ilk iki basamak dışındaki sayıların büyüklüğüne bakılmadan hepsi atılmıştır. Bu tür **kesme** işlemleri bazen işe yarar. İleride bu konuya yeniden döneceğiz.

Şimdi yuvarlatma kurallarıyla yuvarlatmadan doğan hataları ele alıp inceleyelim.

4.2 YUVARLATMA KURALLARI

- Planlama teşkilatının saygıdeğer hesap uzmanları, Sayın Başbakanımızın isteğiyle yine ülkemizin kalkınma ve büyüme hızını hesaplamak için toplanmış bulunuyoruz...
- Ben deminki hesabımın sonucunda bir adet virgül buldum.
 - Nasıl yani? Virgülün sağında ve solunda hiç rakam yok mu? Sıfır virgül sekiz olur, yedi virgül beş olur. Ama sadece virgül olması biraz garip değil mi Hikmet Bey?
 - Maalesef sadece virgül çıktı efendim.
 - Baari bir rakam bulsaydınız yuvarlak hesap yükseltir bir şeyler uydururduk. Virgülü en fazla yuvarlatsak nokta haline dönüştürebiliriz ki bunun da konumuzla bir ilgisi yok.
 - Naapabilirim yüz kere kontrol ettim virgül çıkıyor....
 - Zamanımız yok. Birazdan Başbakan cevap isteyecek. Şöyle yüksekçe bir şey söylemek lazım yoksa feci bozulur Hikmet Bey bulduğunuz virgülü verir misiniz? Evet gördüğünüz gibi hesap neticesinde bulunan virgül dikkatle bakılıp daha büyütülmüş gibi tasavvur edildiğinde aynen dokuza benzemektedir. Zira virgül minyatür bir dokuz rakamı değil midir? Bütün bu bilgilerin ışığı altında memleketimizin kalkınma ve iktisaden büyüme hızı yüzde dokuzdur. Hadi şimdi daalın.

Atilla Atalay, Gırgır

Yuvarlatma yapılırken genellikle iki yoldan biri uygulanır: Veriler ya cm, bin kişi, milyon TL, ton gibi belli bir birime yuvarlatılır, ya da belli sayıda basamak kalacak biçimde yuvarlatılır (Allen, 1964:68).

Sözelimi Türkiye'nin 2001 yılı dış satımı 31.339.991 bin \$'dır (DİE, 2002a:509). Bu sayı yuvarlatılarak yayımlanmıştır. Biz bunu birinci yola göre bir kez daha yuvarlatırken milyon \$ ya da milyar \$ birimlerinden birini seçebiliriz. Bu durumda 2000 dışsatımı 31.340 milyon \$ ya da 31 milyar

\$ olur. İkinci yola göre, üç basamak yeterlidir düşüncesiyle yuvarlatma yapsak 31,3 milyar \$ yazmamız gerekirdi.

Burada bir kaç noktaya dikkat etmemiz yerinde olur. Bir sayıyı yuvarlatırken, yuvarlatmadan sonra en sağda kalan basamak, ya yuvarlatma öncesi değerini aynen korumakta ya da bir büyümektedir. 31.339.991 bin \$, milyar \$'a yuvarlatılırken 31, üç anlamlı basamağa yuvarlatılırken 31.3 aynı kalmış, milyon \$'a yuvarlatılırken ise 31.339 bir büyüyerek 31.340 olmuştur.

Bu da, yuvarlatma sırasında gözden kaybolan basamaktaki değer 5'ten büyük olup olmamasına bağlıdır. Örneğimizde 31'den sonra gelen 3 < 5 olduğundan yuvarlatma yapılırken kesilip atılmış, ama 31.339'dan sonra gelen 9 > 5 olduğundan yuvarlatma yukarı doğru yapılmıştır.

Bu değer tam 5'e eşitse yuvarlatma hangi yöne doğru yapılacaktır? Uygulamada iki yol izlenmektedir. Birincisi 4/5 kuralıdır (Stytely, 2003:64). Buna göre silinen basamak 5'e eşitse hep yukarı doğru yuvarlatılır. İkinci yola göre yuvarlatmadan sonra en sağda kalan basamak çift sayı olmalıdır. Örnek olarak 101,5 ile 102,5 sayılarını alalım. 4/5 kuralına göre ilki 102'ye, ikincisi 103'e yuvarlatılır. Oysa çift sayıya yuvarlama kuralına göre her iki sayı da 102'ye yuvarlatılır. Böylelikle, bu durumdaki bütün sayıların hepsinin yukarıya doğru yuvarlatılması önlenmiş; bazılarının yukarıya, bazılarının aşağıya yuvarlatılması sağlanarak, yuvarlatmadan doğacak hata azaltılmış olur. Bu iki sayının yuvarlatmadan önceki toplamlarının 204 ettiğini biliyoruz. 4/5 kuralına göre yuvarlatılmış toplam 205 olur ki bu yuvarlatmadan doğan bir hatadır. Çift sayıya yuvarlatma sonuçlarının toplamı ise 204 eder, yani yuvarlatma bir hata doğurmamış olur. Uzun dizilerin toplamlarında bu yuvarlatma hatası çok daha önem kazanır. İkinci yöntem kullanılarak bu hata küçük tutulabilir.

Dikkat edilecek ikinci nokta, yuvarlatılmış bir sayıda en sağdaki basamağın sıfır olmasının ne anlama geldiğidir. Sözelimi 101,90 ton ile 101,9 ton aynı anlamı taşımaz mı? Bunun yanıtı kesinlikle "hayır"dır. Çünkü 101,90 ton 101.895 – 101.905 kg arasında bir sayıyı gösterirken; 101,9 ton, 101.850 – 101.950 kg arasında bir sayıyı temsil etmektedir. Eğer ilk sayımız 101.895 – 101.905 kg arasında bir değerden yuvarlatılarak elde edilmişse ondalık işaretinin sağında kalan, aynı zamanda elimizdeki sayının da en sağdaki basamağında bulunan "sıfır" kesinlikle belirtilmeli, sayı 101,90 diye yazılmalıdır. Yok eğer ilk sayımız 101.850 – 101.950 kg arasında bir sayının yuvarlatılmasıyla bulunmuşsa o zaman da en sağdaki sıfırın kesinlikle yazılmaması, sayının 101,9 diye gösterilmesi doğru olur.

Ondalık basamağı olmayan sayıların en sağ basamağındaki sıfırın anlamı nedir? Bu basamak gerçekten sıfır olarak mı ölçülmüştür, yoksa sonradan yuvarlatma sırasında mı ortaya çıkmıştır? Bu sorunun yanıtı biliniyorsa sorun çıkmaz ama ya bu konuda kesin bir bilgi yoksa! O zaman, yanıltıcı

bir duyarlılıktan kaçınmak isteniyorsa son basamağın yuvarlatma sonucu olduğu varsayılır. Böylece eldeki sayının basamakları bir azaltılmış olur. Sözelimi 2000 yılı sonunda 120 milyar \$ dış borçtan söz edildiğinde bunun 119,5 –120,5 milyar \$ arasında değil de, 115 – 125 milyar \$ arasında olduğunu düşünmemiz daha doğru olur. Gerçekten de o tarihteki borcumuz 118,615 milyar \$'dır (DİE, 2002a:636). Aslında bu borç 119,5 –120,5 milyar \$ arasında olsaydı yuvarlatma yapıldığında 120,0 milyar \$ diye yazılması gerekirdi.

Çok acele işlem yapma gerektiğinde ya da böyle acele durumlar için kullanılan **dal-yaprak** gösterimlerinin hazırlanması sırasında yuvarlatma yerine **kesme** diyebileceğimiz başka bir yol da izlenebilir (Erickson, Nosanchuk, 1992:21). Bu yol kullanıldığında, elde kalacak basamaklardan sonra gelen sağdaki her sayı, beşten büyük olup olmadığına bakılmaksızın kesilip atılır. Böylece Türkiye'nin 2001 dışsatımı 31.339 milyon \$ ya da 31 milyar \$ olarak gösterilir. İlk bakışta oldukça sakıncalı gibi görünen bu yolun, özellikle dal-yaprak gösteriminde sağladığı kolaylıklar, o konu işlenirken açıklanacaktır.

Yuvarlatma yapılırken kaç basamak kullanılması gerektiğine nasıl karar verilecektir? Burada sağduyunun rolü çok önemlidir. Ancak yuvarlatılacak veriler bazı hesaplama işlemlerinde kullanılacaksa, yuvarlatma işlemlerinde fazla ileri gitmemek, hesap sonucunda bulacağımız değerın kaç basamak içermesini istiyorsak, genellikle bundan bir fazla basamak kalacak biçimde yuvarlatma yapmak uygun olur.

Yeniden kişi başına gelir hesapladığımız yukarıdaki örneğe dönelim. Gelir 125.596.129 milyon TL, nüfus da 67.469 bin kişi olarak yuvarlatılmıştı. Buradan kişi başına gelir (= gelir / nüfus) 1.861.538.321 TL bulunmuştu. Eğer kişi başına geliri milyon TL cinsinden bulmak istiyorsak gelirdeki 9, nüfustaki 5 basamağın hepsi gerekli midir? Bu iki sayı arasında çarpma ya da bölme yapılacaksa (biz bölme yapmaya niyetliyiz) bulunacak sonucun en çok 5 anlamlı basamağı olabilir. Çünkü işlemde kullanılan iki sayıdan, basamakları daha az olan nüfusta 5 basamak bulunmaktadır. Sonucun basamak sayısı da bundan çok olamaz. Belki de bir eksik olduğu bile söylenebilir.

Kişi başına geliri milyon TL'ye yuvarlatırsak 1.862 milyon TL yani bir milyar sekiz yüz altmış iki milyon TL buluruz. Buradaki basamak sayısı dördüttür. Öyleyse bölmede kullanılan basamak sayıları (9 ile 5) gereğinden çoktur. Geliri 5 basamak içerecek biçimde 125,60 trilyon TL, nüfusu da yine 5 basamak içerecek biçimde 67,469 milyon kişi diye yazıp bölme işlemini bu yuvarlatılmış sayılarla yapsak 1.862 milyon TL elde edilir. Görüldüğü gibi dört basamak içerecek bu bölme sonucu için bölen, bölünen terimlerinin beşer basamak içermesi yeterlidir.

Bölenle bölünen de üç basamaklıysa sonuçtaki üç basamak da anlamlı değer taşır mı? Çoğu zaman hayır. Bu örnekte 126 trilyon TL, 67,5 milyon kişiye bölünse (1,86 değil) 1,87 milyar TL çıkar. Bu da bize en sağdaki basamağın yanıltıcı olduğunu gösterir.

Peki kişi başına geliri milyon yerine bin TL duyarlılığında bulmak istersek ne olacaktır? Yanıt 1.861.538 bin TL yedi basamak içermektedir. Acaba bu sayının en sağdaki iki basamağının değerine (38'e) güvenebilir miyiz? Başka bir deyişle kişi başına gelir 1.861.537 ya da 1.861.539 bin TL olabilir mi?

9 basamaklı gelir değeri (daha önce yuvarlatılmış olduğundan) 125.596.128,5 ile 125.596.129,5 milyon TL arasında bir değerdir. Nüfusun da benzer biçimde 67.468,5 ile 67.469,5 bin kişi arasında olduğu söylenebilir. Bu yuvarlatmalar gelirin üst sınırına, nüfusun alt sınırına yakın değerlerden yapılmışsa kişi başına gelir aslında daha yüksek; yok eğer tersine gelirin alt, nüfusun üst sınırına yakın yerden yapılmışsa bu kez de kişi başına gelir aslında daha düşüktür. En uç çapraz değerler kullanılırsa kişi başına gelir kaç çıkar?

Gelirin üst sınırı olan 125.596.129,5 milyon TL, nüfusun alt sınırı 67.468,5 bin kişi değerine bölünürse kişi başına gelirin üst sınırı

$$125.596.129,5 / 67.468,5 = 1.861.552.124 \text{ TL}$$

bulunur. Şimdi de gelirin en düşük, nüfusun en büyük değerlerini birbirine bölelim. Kişi başına gelirin alt sınırı olan

$$125.596.128,5 / 67.469,5 = 1.861.524.518 \text{ TL}$$

çıkarak. Bu sınır değerlerden üstteki, bin TL'na yuvarlatılırsa 1.861.552 bin TL eder. Alttakine benzer bir yuvarlatma uygulansa bu kez de 1.861.524 bin TL bulunur. Öyleyse kişi başına gelir 1.861.524 bin TL ile 1.861.552 bin TL arasında olabilir. Demek ki en sağdaki iki basamağa yüzde yüz güvenmemiz doğru değilmiş. Bin TL'ye yuvarlatılmış, tam güvenilebilir bir kişi başına gelir hesabı yapabilmek için gelir ve nüfus değişkenlerinin ikisinin de en az yedi basamağının bulunması gerekmektedir. Ama nüfus değişkenimizin değeri bize beş basamak içeren yuvarlatılmış bir sayı olarak verildiğinden bu işlemleri bu bilgiyle yapamayız.

Bu örnekteki bulgumuzu genelleştirerek şu kuralı yazabiliriz:

Çarpma-bölme işlemlerinde sonuç, işlemde kullanılan sayılardan basamağı en az olanunki kadar, bazen de ondan bir eksik doğru basamak içerir.

Aynı örneğe tersten de yaklaşabiliriz. 9 ve 5 basamaklı iki sayıyı birbirine böldüğümüzde 1.861.538.321 TL olarak bulunacak sonucu en fazla beş doğru basamakla 1.861,5 milyon TL'ye yuvarlatmalıyız. Sonuçtan iyice

emin olmak istersek dört basamaklı 1.862 milyon TL yuvarlatması daha uygundur. 1.861.540 ya da 1.861.538 bin TL gibi bir yuvarlatma yanıltıcı bir duyarlılık taşır.

Benzer kural kök almalarda da geçerlidir. Sözelimi, yuvarlatılarak elde edilmiş 90 sayısının karekökü 9,48683298'dir. Acaba bunun kaç basamağı doğrudur? 9,5 mi yoksa 9,49 mu yazmalıyız? Yuvarlatma öncesinde 89,5 – 90,5 arasında bir sayımız olduğunu biliyoruz. Bu iki sınır değer kareköklerini alırsak, sırasıyla 9,46 ile 9,51 elde ederiz. Bu sayıları birer basamak yuvarlatırsak aynı değere ulaşırız. Her iki sınır değeri için de ortak karekök 9,5 olduğuna göre, yuvarlatılmış 90 sayısının karekökünü en çok iki basamaklı yazabiliriz. Bunu üç basamağa çıkarıp 9,49 yazmak yanıltır. Yuvarlatma öncesi sayımız 90,155'ten büyükse karekökü 9,50'ye, 89,965'ten küçükse 9,48'e yuvarlanır. Demek ki 9,49 sayısı, 89,5 – 90,5 arasındaki bütün sayıların değil, yalnızca 89,965 – 90,155 arasındaki sayıların karekökü için doğru bir gösterim olabilir. Başka bir deyişle yuvarlatılmış 90 sayısının karekökünü en çok iki basamaklı bir sayıyla gösterebiliriz.

Çizelge 3.1 ile 3.2'deki yüzde hesaplamalarını bu açıklamaların ışığında inceleyecek olursak bu kurallara uyulduğunu görürüz. Bütün göreceli sıklıklardaki basamak sayısının, bölme işleminde kullanılan sayılardan basamağı az olanın basamak sayısına eşit olduğu görülür. Bu sözün doğruluğunun sınanması okuyucuya bırakılmıştır.

Uygulamalarda kaç basamaklı sayılar kullanmalıyız? Bazı savlara göre sayıları iki **etkin** basamağa yuvarlamak yeter, ileri düzeyde istatistik çözümlmeleri bile üçüncü ya da dördüncü basamaktaki sayılara bağlı değildir (Ehrenberg, 1982:219). Etkin basamak sayısı şudur: Diyelim bir öbek insanın boyları 156, 177, 172, 169, 183, ... cm olarak ölçülmüş olsun. Baştaki ilk basamak bütün gözlemler için hep 1 olduğundan etkin sayılmaz. Bu sayılarda etkin basamak sayısı ikidir. Son üç ayda satışları 24.352.128, 24.569.265, 24.943.729 kg olan bir işletme, bunları 24.350, 24.570, 24.940 ton olarak gösterirse yine iki etkin basamak kullanmış olur.

Bu dönüşürmede bilgi kaybı çok azdır. Sözelimi 24.352.128 kg 24.350 tona geçerken (2.128 / 24.352.128) onbinde 1'den küçük bir değer değişmesi olmuştur. Bu bilgi kayıpları öbür satışlar için sırasıyla (735 / 24.569.265) yüzbinde 3, (3.729 / 24.943.729) onbinde 1,5 kadardır. İki etkin basamak kullanmadan doğan bilgi kaybı aynı sayıları bir çizimde işaretlerken uğranan kayıptan daha küçüktür. Ama ikiden az etkin basamak kullanmak bu kaybı yükseltir. İlk ayın satışı 24,4 bin ton gösterilse, bilgi kaybı %2'ye yaklaşır.

Söz konusu bilgi kaybı oranı, etkin basamaklardan önceki basamak sayısına bağlıdır. Aylık satışların her biri 20 bin ton daha eksik, yani

4.352.128, 4.569.265, 4.943.729 kg olsaydı, yuvarlatma sonucunda elde edilen 4.350, 4.570, 4.940 ton değerlerindeki bilgi kayıpları sırasıyla onbinde 4,9; 1,6; 7,5 düzeylerine fırlardı.

İk etkin basamaklı sayı kullanımı sunuş ya da yorum yaparken çok işe yarayabilir. Ama, diyelim DİE gibi bir kuruluş ya da bir işletme bütün verilerini iki etkin basamağa indirgeyemez.

4.3 YUVARLATMA HATALARI

Yanlışlardan kaçınmak için yuvarlatma yollarından hangisine uyulduğunun bir kenara yazılması uygun olabilir. İleride bu yuvarlatılmış sayılara başvurma gereği doğduğunda bu not işe yarar.

Yuvarlattığımız sayılar başka işlemlerde kullanılmayacaklarsa yuvarlatma hataları önemli sorunlara yol açmaz. Çünkü zaten bu işi gereksiz ayrıntı olarak gördüğümüz bölümlerden kurtulmak için yaparız. Ama yuvarlatılmış sayılarla işlemler yapacaksak o zaman yuvarlatmanın doğuracağı hatalar da önem kazanır.

Önce yuvarlatma hatasıyla ne demek istediğimizi açıklayalım: Bir sayı yuvarlatılırken ya biraz küçülür ya da biraz büyür. Aradaki fark yuvarlatma hatasıdır:

$$\text{Yuvarlatma hatası} = \text{Asıl sayı} - \text{yuvarlatılmış sayı.}$$

Sözgelimi 14,6 ile 15,9 sayıları yuvarlatılmış sayılarsa, ilki 14,55-14,65, öteki de 15,85-15,95 arasındaki sayılardan yuvarlatılmıştır. Her birindeki yuvarlatma hatası her iki yönde en çok 0,05 kadardır. Bu iki sayıyı toplayalım: $14,6 + 15,9 = 30,5$. Doğru toplam

$$14,55 + 15,85 = 30,40 \text{ ile } 14,65 + 15,95 = 30,60$$

arasındadır. Bunu $30,50 \pm 0,10$ diye yazabiliriz. Öyleyse her bir sayıdaki yuvarlatma hatalarının toplamı ($0,05 + 0,05$), toplamdaki yuvarlatma hatasını ($0,10$) verir. $15,9 - 14,6 = 1,3$ çıkarma sonucunun da ne kadar hata içerdiği benzer biçimde bulunabilir. Doğru fark

$$15,85 - 14,65 = 1,20 \text{ ile } 15,95 - 14,55 = 1,40$$

arasındadır, demek ki $1,30 \pm 0,10$ kadardır. Burada da farktaki hatanın büyüklüğü, tek tek sayılardaki yuvarlatma hatalarının toplamına eşittir. Öyleyse kuralı şöyle genelleayebiliriz:

Yuvarlatılmış sayılarla yapılacak toplama-çıkarmalarda sonucun yuvarlatma hatası, işleme giren her bir sayıdaki yuvarlatmadan doğan hataların toplamına eşittir (Allen, 1964:70).

Bir dizi yuvarlatılmış sayıyı toplarken farklı duyarlılık düzeyleri kullanmak doğru olmaz. Her bir sayı aynı ölçü birimine göre yuvarlatılmalıdır. 14,6 ton ile 15,92 ton değerlerini oldukları gibi toplamamalıyız, çünkü biri yüz kg, öteki on kg düzeyinde yuvarlatılmıştır. Çizelge 3.1'deki bütün nüfus değerleri bin kişi olarak yuvarlatılmıştı. Böyle toplamalarda her sayıdaki hata, toplama giren değer sayısı ile çarpılarak toplamdaki yuvarlatma hatası elde edilebilir. Bu çizelgede, her birindeki yuvarlatma hatası 500 kişi olan 5 sayı toplanmış, dolayısıyla toplamda en çok $5 \times (\pm 500) = \pm 2.500$ kadar bir hata yapılmıştır. Çoğu zaman tek tek sayılardaki hataların işaretleri hep aynı olmadığından bunlar kısmen de olsa birbirlerini götürürler, dolayısıyla toplamdaki hata daha küçük olur. Gerçekten de Çizelge 3.1'deki yüzdelerin toplamı, tam 100,00 etmektedir. Demek ki her bir sayıdaki yuvarlatma hataları birbirini götürmüştür. Ancak aynı şey bu çizelgedeki kişi sayılarının toplamı için söylenemez. Kişi sayıları toplanırsa 7.918 bin kişi elde edilmektedir. Oysa çizelgede gerçek toplam 7.919 bin kişi olarak verilmiştir. Toplamın yuvarlatma hatası bin kişi kadardır. Toplamın tam tutmadığı durumlarda bunu açıkça yazmakta yarar vardır. Çizelge 3.1'de bu açıklama bir dipnot ile yapılmıştır.

Toplama-çıkarmalarda her sayının aynı ölçü birimine göre yuvarlatılması gerektiğini gördük. Çarpma-bölmelerde ise asıl önemli olan basamak sayısıdır. Sözgelimi $1,2 \times 0,74 = 0,888$ çarpımının duyarlılığı, çarpılan sayılarda virgülden sonra kaç basamak olduğuna değil, bu sayılarda kaç basamak bulunduğuyla bağlıdır. Çarpım sonucu $12 \times 74 = 888$ olarak bulunup ondalık basamaklar sonradan ayrılabilir.

Burada ilgili kavram **görelî hatadır**. Eğer 1,2 ile 0,74 yuvarlatılmış sayılarsa, bunların her birinin görelî hatası, “yuvarlatma hatası / sayı”, yani

$$0,05 / 1,2 = \% 4,17 \text{ ile } 0,005 / 0,74 = \% 0,68$$

kadardır. Bu durumda çarpım sonucunun en küçük ve en büyük olası değerleri

$$1,15 \times 0,735 = 0,845 \text{ ile } 1,25 \times 0,745 = 0,931$$

eder. Aradaki fark $0,931 - 0,845 = 0,086$ 'dır. Bunun yarısı 0,043 her iki yöndeki yuvarlatma hatasıdır. Öyleyse çarpım yaklaşık olarak $0,888 \pm 0,043$ sınırları arasındadır. Buradan da çarpımın görelî hatası $0,043 / 0,888 = \% 4,84$ bulunur. Dikkat edilirse çarpımın görelî hatası $\% 4,84$, tek tek çarpanların görelî hataları toplamı $4,17 + 0,68 = 4,85$ 'e çok yakındır. Buna yaklaşık eşittir bile diyebiliriz. Bu bulgu şöyle genellenebilir:

İki yuvarlatılmış sayının çarpımlarının ya da bölümlerinin görelî hatası, bu sayıların ayrı ayrı görelî hatalarının toplamına yaklaşık olarak eşittir (Allen, 1964:72)

Görüldüğü gibi yuvarlatma hataları çarpımda da, bölümde de birbirine eklenerek artmaktadır. Öyleyse sonuçtaki basamak sayısı, işleme giren terimlerin basamak sayısından çok olamaz, tam tersine genellikle daha az olur. Örneğimizde çarpımın alt sınırı 0,845, üst sınırı 0,931 idi. Çarpanlarda ikiye basamak vardı. Sonucu iki basamakla bile yazamayız, çünkü iki sınır arasındaki sayıların hepsinin aynı iki basamaklı sayıya yuvarlatılması olanaksızdır. Giderek sonucu 0,9 biçiminde tek basamakla yazmanın bile içimizi tam anlamıyla rahatlatamayacağını söyleyebiliriz. Az bir olasılıkla da olsa sonucun 0,845 – 0,850 arasında çıkması durumunda 0,9 yuvarlatması bile yanıltıcı olacaktır.

Az önce verdiğimiz kişi başına gelir örneğimizde sonuçta 5 basamağa sahip olabilmek için gelir ve nüfus değişkenlerinin her birinin neden en az 6 basamak içermeleri gerektiği umarız şimdi daha iyi anlaşılmıştır.

Yuvarlatmalarla ilgili olarak bilmemiz gereken son bir nokta daha var. Doğal tamsayılar, sonsuz sayıda basamakları varmış gibi işlem görürler. Dolayısıyla yuvarlatılmış bir sayıyı doğal bir tamsayıya bölersek, sonucun basamak sayısı yalnızca yuvarlatılmış sayının basamak sayısına bağlıdır. Sözgelimi boyları cm cinsinden ölçülmüş on basketbol oyuncusunun boylarını toplayıp ona bölerek bulduğumuz boy ortalamasında üç basamak olmasıdır. Bunu kişi başına gelir hesaplarken de anımsamak da yarar var. Eğer DİE'nin bilip de bize söylemediği yuvarlatılmamış gelir değerini yuvarlatılmamış kişi sayısına bölersek, sonucun basamak sayısı sonsuza kadar bile uzayabilir. Ancak bu kadar duyarlı bir kişi başına gelir değerini bilmek ne işe yarar?

Okuma Önerileri

Allen, R. G. D. (1964) *Statistics for Economists*, Hutchinson (64-75).

Ehrenberg, A.S.C. (1975) *Data Reduction*, Wiley.

Ehrenberg, A.S.C. (1982) *A Primer in Data Reduction*, Wiley (219-222).

Stutely, R. (2003) *The Definitive Guide to Managing the Numbers*, Prentice Hall (63-66).

ALIŞTIRMALAR

- 4.1** Çizelge 3.1’de gösterilen “pay”ların her birini inceleyiniz, basamak sayıları yanıltıcı bir duyarlık veriyor mu? Paylardan herhangi biri daha çok doğru basamak içerebilir mi? Neden? Bunlardan kimileri için daha az basamak kullanılmasını önerir misiniz? Neden?
- 4.2** Çizelge 3.2’deki “pay”ları ele alalım. Bunlar hesaplanırken kullanılan kişi sayıları doğal tamsayı olduklarından % değerler çok daha fazla basamak içerebilirler. Bu pay değerlerini daha çok basamakla göstermenin sağlayacağı herhangi bir yarar var mı? Neden? Peki basamak sayısını azaltmak işe yarar mı? Gerekçenizi açıklayınız?
- 4.3** Çizelge 3.4’e dönelim. Buradaki verilerin doğru basamak sayıları konusunda ne düşünüyorsunuz? Kişi sayıları yanlarındaki hane sayılarıyla çarpılsa elde edilen çarpımlar en çok kaç basamaklı sayılarla gösterilebilir? Sizce okuyucuya iletiyi en uygun biçimde vermek için kaç basamak kullanılmalı? Neden?
- 4.4** Zonguldak ilinin tüketici fiyatları indeks sayıları 1994:100 olmak üzere 1995 yılında 184,2 iken 2001 yılında 4909,7’ye çıkmıştır. (DİE, 2002a:561).
- a) 2001 yılı indeks sayısı 1995 yılının kaç katıdır?
- b) İndeks sayılarının yuvarlatılmış olduğunu gözönüne alarak bulgunuzun alt ve üst sınırlarını hesaplayınız. Buna göre yanıtınızın basamak sayısı en çok kaç olabilir?
- 4.5** Çizelge 3.5’e dönelim. Buradaki verilen hepsinin en sağ basamağı sıfırdır. Bunu nasıl yorumluyorsunuz? Her bir verinin doğru basamak sayısını bulunuz. Her bir ülkede iki yıl arasında yüzde kaçlık bir değişme olduğunu görmek için oranlar hesaplasanız sonuçları en çok kaç basamakla gösterebilirsiniz? Sizce bu değişim oranları için en uygun basamak sayısı kaçtır? Neden?
- 4.6** Bir üreticinin bir ayda yaptığı satışların bölgelere göre dağılımı yuvarlatılmış olarak aşağıda verilmiştir. Ay boyunca fiyat hiç değişmemiştir.

Bölge	Satış miktarı (ton)	Satış geliri (YTL)
Marmara	3.350	184.424
Ege	2.763	151.991
Akdeniz	938	51.570
Ötekiler	2.122	116.730
Toplam	9.173	504.515

- a) Toplam satışların YTL değerini, ton değerine oranlayarak ton fiyatını bulunuz. Bu fiyatın alt-üst sınırları kaçtır? Bu durumda yanıtınızda en çok kaç basamak olmalıdır?
- b) Aynı işlemi her bölge için yeniden yapınız. Her bölge için bulduğunuz fiyatın basamak sayıları birbirine eşit mi? Neden? Her bölge fiyatının basamak sayısı ile o bölgenin satış miktarı ile satış tutarının basamak sayıları arasında bir ilişki var mı? Belirtiniz.
- c) Marmara bölgesine yapılan satışların toplam satışlar içindeki % payını hem miktar, hem tutar cinsinden ayrı ayrı bulunuz. Yanıtlarınızı en çok doğru basamak içerecek biçimde düzenleyiniz.
- d) (c)'de bulduğunuz iki % değerini karşılaştırınız. İkisindeki doğru basamak sayısı birbirine eşit çıkıyor mu? Bunlardan birini kullanmak isterseniz hangisini seçersiniz? Neden? Yoksa her bir pay değerini ayrı ayrı kendi alanlarında kullanmayı mı daha doğru bulursunuz? Neden?

4.7 Türkiye'nin yedi bölgesinin 2000 sayımına göre nüfuslarıyla yüzölçümleri aşağıdaki çizelgede gösterilmiştir:

- a) Her bölgenin nüfus yoğunluğunu bulunuz. Her bir değer için basamak sayısını gerekçenizle birlikte belirtiniz.

Bölge	Nüfus (bin kişi)	Yüzölçümü (km ²)
Marmara	17.354	72.578
Ege	8.951	90.442
Akdeniz	8.717	115.730
İç Anadolu	11.636	89.247
Karadeniz	8.439	186.100
Doğu Anadolu	6.142	75.358
Güneydoğu Anadolu	6.613	145.360
Toplam	67.853	774.815

Yuvarlatma nedeniyle toplam tutmamaktadır.

DİE, 2002a:46, DİE, 1995:6

- b) Her bölgenin gerek toplam nüfus, gerek yüzölçümü içindeki paylarını ayrı ayrı bulunuz. Bunların basamak sayılarını irdileyiniz.

- 4.8** 1990'ların başlarında bazı ülkelerdeki meslek yüksek okulu öğrencilerinin yüksek öğrenimdeki toplam öğrenciler içindeki yüzde payları aşağıdadır. Bu verilerin doğru basamak sayıları konusunda ne söyleyebilirsiniz?

Ülke	%	Ülke	%
Pakistan	82	Yunanistan	39
Avusturalya	66	İrlanda	38
Belçika	60	Almanya	35
Brezilya	53	Fransa	27
Hollanda	53	Türkiye	10
Şili	47		

Gürüz vö. 1994:73

- 4.9** Alıştırma 1.3'te verilen kâr rakamlarının doğru basamak sayılarını bulunuz. Düşüncenizi yazınız.
- 4.10** Alıştırma 1.4'te aktarılan gazete haberinde geçen her verinin basamak sayısı konusunda yorumunuzu yazınız.
- 4.11** Alıştırma 3.4'te geçen yüzde ile gösterilmiş verilere uygulanan yuvarlatma konusunda ne düşünüyorsunuz? İçlerinde, basamak sayıları gereğinden fazla yazıldığı için yanlış duyarlılık izlenimi verenler olabilir mi?
- 4.12** Alıştırma 3.6'da gösterilen yağışlı gün sayılarına uygulanan yuvarlatma konusunda ne düşünüyorsunuz? Basamak sayıları sizce uygun mudur?
- 4.13** Alıştırma 3.11'e dönelim.
- Burada yer alan ceza tutarlarının basamak sayıları konusunda ne düşünüyorsunuz?
 - Yıllar geçtikçe, enflasyon nedeniyle, paranın satın alma gücü azalmaktadır. Bunu gözönüne alarak basamak sayılarını yeniden düzenlemeyi düşünür müydünüz? Bunu nasıl yaparsınız?
 - Bütün tutarları iki etkin basamağa yuvarlatırsanız her işlemde yuvarlatma hatası ne kadar olur?
- 4.14** Alıştırma 3.12'de verilen çocuk ölüm oranlarını inceleyiniz. Çok sayıda ülkede aynı oranı görebilirsiniz. Bu size yuvarlatmada biraz ileri gidildiği, ülkeler arası farkların kaybolduğu izlenimini veriyor mu? Yoksa bu basamak sayıları uygun bir duyarlılığı mı yansıtıyor?

Hangi Veriyi, Nerede Arayalım?

"Veri! Veri! Veri!" diye sabrı taşarak bağırdı.

"Çamur olmadan tuğla yapamam ki."

A. C. Doyle, Sherlock Holmes,

5.1 GİRİŞ

Veri derleme sürecinin tasarlanması, akıllı veri derlemenin can alıcı bir adımıdır. Bu sürecin bilinçli, uyumlu olması gereken aşamaları şöyle sıralanabilir (Ott, Longnecker, 2001:18):

- İncelemenin, araştırmanın ya da deneyin amacı belirlenmeli,
- İlgilenilen değişken(ler) saptanmalı,
- Alan araştırması ya da bilimsel deney için uygun bir tasarlama düşünülmeli, veriler ondan sonra toplanmalıdır.

Çalışmanın amacının belirlenmesi için önce ele alınan sorun iyice anlaşılmış olmalıdır. Diyelim ki bir belediyenin toplu taşımacılık birimi, yeni işletmeye açılan bir raylı sistem konusunda halkın ne düşündüğünü bilmek istiyor. Bu sistemin hangi öğeleri, insanları bu yolu kullanmaya yöneltir? Bu çalışmanın amacı bu etmenleri bulup ortaya çıkarmak olabilir. Daha sonra bu değişkenler üzerinde oynanıp değişiklik yapılarak halkın sistemi kullanması özendirilebilir.

İlgili değişkenleri saptamak için çalışmanın amacı iyice incelenmelidir. Bazı önemli etmenler, başka belediyelere ilişkin araştırmaları okuyarak ya da taşımacılık birimindeki görevlilerin kafa patlatıp düşünmeleriyle öne çıkabilir. Bunlar arasında güvenlik, maliyet, temizlik, rahatlık, durakların kişilerin evine, işyerlerine, alışveriş ya da eğlence merkezlerine yakın olması, arıza sıklığı gibi konular bulunabilir. Yolculara ilişkin yaş, cinsiyet, gelir, oturulan / çalışılan semt gibi veriler de gerekebilir. Ayrıca kişinin raylı sistemi ne sıklıkta kullandığını bilmek de önem taşıyabilir.

Bütün bunlar saptandıktan sonra verilerin nasıl derleneceğini düşünmeye sıra gelir. Gerekli veriler alan araştırmasıyla, deneylerle, daha önceki çalışmaların bulgularının incelenmesiyle, resmi ya da özel kuruluşların konuyla ilgili yayımlanmış verilerinin taranmasıyla toplanabilir.

Alan araştırması ya da deneyler aracılığıyla veri toplayabilmek için veri derleme yöntemlerini doğru uygulamayı bilmek, hazır verileri kullanmak için de hangi yayınlarda, ne gibi verilerin bulunduğunu öğrenmek gerekir.

Birinci el veri – ikinci el veri

Çeşitli bakanlıklar, Devlet İstatistik Enstitüsü (DİE) gibi resmi kurumların topladığı verilerle araştırmacının kendisi tarafından toplanan verilere **birinci el veri** denir. Bu kurumlar ya da kişiler tarafından toplanan verileri düzenleyerek yayımlayan başka kurumların ya da kişilerin verilerine ise **ikinci el veri** adı verilir (Sandy, 1990:10). Dünya Bankası'nın yayımladığı veriler ikinci el verilere bir örnek olarak gösterilebilir. Bu veriler çeşitli ülkelerde derlenir, bir merkezde toplanarak özetlenip açıklanır. DİE de bazı verileri ikinci el olarak yayımlamaktadır. Sözgelimi sağlık istatistikleri Sağlık Bakanlığı tarafından derlenmekte, DİE'de işlenerek yayıma hazırlanmaktadır.

Araştırmalarda olabildiğince birinci el verilerden yararlanılmalıdır. İkinci el veriye ancak birinci el verinin bulunamadığı zamanlarda başvurulmalıdır. Bu bölümde gerek yerli, gerek yabancı veri kaynakları incelenecek, örnekleme yoluyla veri derleme işi gelecek bölümde ele alınacaktır.

Bu konuda ilk bilinmesi gereken hangi tür verilerin hangi kuruluşlarca derlenip yayımlandığıdır. Ülkemizde çok çeşitli verileri derleyip yayımlayan kuruluşların başında Devlet İstatistik Enstitüsü (DİE) gelmektedir. Ayrıca iktisat konularında çalışanların işlerine yarayacak bazı veriler Devlet Planlama Teşkilatı (DPT), Başbakanlık Hazine Müsteşarlığı, Merkez Bankası, Maliye Bakanlığı, Çalışma ve Sosyal Güvenlik Bakanlığı, Sermaye Piyasası Kurulu (SPK), İstanbul Menkul Kıymetler Borsası (İMKB) gibi resmi kuruluşlar; İstanbul Ticaret Odası (İTO), İstanbul Sanayi Odası (İSO), Türkiye Bankalar Birliği (TBB), Türkiye Sanayiciler ve İşadamları Derneği (TÜSİAD), Türkiye Ticaret, Sanayi, Deniz Ticaret Odaları ve Ticaret Borsaları Birliği (TTOB) gibi meslek kuruluşları; çeşitli bankalar, bazı sendikalar tarafından yayımlanmaktadır. Öbür konularda da çeşitli bakanlıklar ya da resmi kurumlar veri toplamaktadır.

Uluslararası verileri yayımlayan kuruluşların başında da Ekonomik İşbirliği ve Kalkınma Örgütü (OECD), Uluslararası Para Fonu (IMF), Birleşmiş Milletler (UN), Dünya Bankası (WB) gibi kuruluşlar gelmektedir. Aşağıda bu kuruluşların her birinin yayımladıkları verilere ilişkin özet bilgiler bulunacaktır.

5.2 YERLİ KAYNAKLAR

Bir Hint yargıcı bir gün genç bir İngiliz görevlisine der ki: "Biraz yaşlanınca Hindistan istatistiklerini bu kadar rahat kullanamayacaksın. Devlet bu verileri derliyor topluyor, çıkarıyor, n'inci kuvvetini, küpkökünü alıyor, göze hoş görünen çizimler hazırlıyor. Ama unutma ki bu sayılar ilk önce köy muhtarının ağzından çıkıyor, o da canının istediği rakamı söylüyor."

Darrell Huff

DİE yayınları

DİE'nin güncel yayınları için ana başvuru kaynağı *Yayınlar ve Elektronik Hizmetler Kataloğu*'dur (DİE, 2002b). Geçmiş yayınları incelemek isteyenler 1927'den sonraki dönemi kapsayan Yayın Kataloğu'ndan yararlanabilirler (DİE, 2001b). Ayrıca her iki yayın da <http://www.die.gov.tr> adresinde yer almaktadır. DİE kuruluşundan bu yana kitap biçiminde yaptığı yayınlarına son yıllarda elektronik ortamlarda sunduğu verileri de ekleyerek yeni bir olanak yaratmıştır. DİE yalnız kendisi veri toplamakla kalmaz, hemen her bakanlık, Anayasa Mahkemesi, Yüksek Öğretim Kurumu, Merkez Bankası, Hazine Müsteşarlığı, Devlet Meteoroloji İşleri, Emniyet, Karayolları, Devlet Demiryolları, Devlet Tiyatroları, PTT, Tapu ve Kadastro Genel Müdürlükleri, Tekel, Milli Kütüphane Başkanlığı, belediyeler, İstanbul Ticaret Odası, kütüphane, sinema, tiyatro müdürlükleri, Devlet Malzeme Ofisi, Emekli Sandığı, SSK, Bağ-Kur, Çocuk Esirgeme Kurumu, Kızılay, çeşitli fabrikalar, gümrük örgütü, liman başkanlıkları gibi sayılamayacak kadar çok kuruluş aracılığıyla elde ettiği verilere de yayınlarında yer verir. Aşağıda DİE'nin bazı yayınları kısaca tanıtılmaktadır.

Türkiye İstatistik Yıllığı En yaygın kullanılan DİE yayınıdır. Yıllık olarak yayımlanmasına karşın kullanıcıya genellikle sekiz yıllık bir karşılaştırma yapma olanağı sağlar. Bu yayında şu konularda veri bulunmaktadır:

Arazi	Çalışma	Fiyatlar
İklim	Tarım	İndeksler
Çevre	Madencilik	Para
Nüfus	Enerji	Banka
Demografi	İmalat sanayii	Sigorta
Sağlık	Bina inşaatı	Maliye
Eğitim	Ulaştırma	Ulusal hesaplar
Kültür	Haberleşme	Gelir dağılımı
Adalet	Turizm	Tüketim
Seçimler	Ticaret	Araştırma-geliştirme
Sosyal güvenlik	Hizmetler	Satınalma gücü paritesi
Sosyal yardım	Dış Ticaret	Uluslararası istatistikler

Türkiye Ekonomisi İstatistik ve Yorumlar Temel olarak ulusal hesaplar, sanayi, inşaat, beklentiler, işgücü, dış ticaret, fiyatlar, bütçe, dış borç, para konularını kapsamaktadır. En güncel verilerle düzenlenen çizimlerle çizelgelerin yanı sıra çözümlemeleri de içeren bu yayın ayda bir çıkmaktadır.

Aylık İstatistik Bülteni Ulusal hesaplar, sanayi, bina inşaatı, ulaştırma, çevre, dış ticaret, fiyatlar, indeksler, maliye, para, banka, konularında veriler içeren bu yayın 1994 yılında önemli bir değişim geçirerek, o yıla kadar ayrı ayrı yayımlanmakta olan *Aylık Dış Ticaret Özeti*, *Aylık Ekonomik Göstergeler*, *Çiftçinin Eline Geçen Fiyatlar İndeksi Aylık Bülteni*, *Toptan Eşya ve Tüketici Fiyatları Aylık İndeks Bülteni*, *Aylık Sanayi Üretim İndeksi ve İmalat Sanayii Aylık Eğilim Anketi Sonuçları*, *Sanayi Üretim İndeksi*, *İmalat Sanayinde Çalışanlar ve Üretimde Çalışılan Saat İndeksi*, *Dönemlere Göre Bina İnşaatı Maliyet İndeksi* adlı yayınların yerine geçmiştir.

Bunlar dışında DİE düzenli aralıklarla şu yayınları da yapmaktadır:

- üç aylık aralıklarla *İmalat Sanayii İstihdam-Ödemeler-Üretim Eğilim*, *Bina İnşaatı Maliyet İndeksi*, *İşgücü Piyasası Analizi*,
- yıllık aralıklarla *İstatistiklerle Türkiye* (son beş yılın verileri, içeriği *Türkiye İstatistik Yıllığı* gibidir), *İstatistik Göstergeler 1923-***** (bu dizi 1923-1990 ile başlamıştır, içeriği *Türkiye İstatistik Yıllığı*'na benzer), *Adalet İstatistikleri*, *Bina İnşaatı İstatistikleri*, *İnşaat ve Tesisat İşyerleri İstatistikleri*, *Ölüm*, *İntihar*, *Evlenme*, *Boşanma İstatistikleri*, *Dış Ticaret İstatistikleri*, *Dış Ticaret Fiyat ve Miktar İndeksleri*, *Perakende Fiyat İstatistikleri*, *Toptan Fiyat İstatistikleri*, *Yıllık*

İmalat Sanayii İstatistikleri, Elektrik, Gaz ve Su İstatistikleri, Madencilik ve Taşocakçılığı İstatistikleri, İmalat Sanayiinde Enerji Tüketimi, Çalışma İstatistikleri, Bütçeler: Belediyeler ve İl Özel İdareleri, Kesin Hesaplar: Belediyeler ve İl Özel İdareleri, Şirket, Kooperatif ve Firma İstatistikleri, Milli Eğitim İstatistikleri: Örgün Eğitim, Milli Eğitim İstatistikleri: Yaygın Eğitim, Kültür İstatistikleri, Spor Faaliyetleri ve Tesisleri, Sporcu İstatistikleri, Spor Kulüpleri İstatistikleri, Tarımsal Yapı: Üretim, Fiyat, Değer, Tarım İstatistikleri Özeti, Su Ürünleri İstatistikleri, Turizm İstatistikleri, Ulaştırma İstatistikleri Özeti, Motorlu Kara Taşıt İstatistikleri, Karayolu Trafik Kaza İstatistikleri, Deniz Taşıtları ve Denizyolu Taşıma İstatistikleri, Hava yolu Taşımacılığı İstatistikleri, Gayrisafi Milli Hasıla, İllere Göre Gayrisafi Yurtiçi Hasıla, Dönemlere Göre Gayrisafi Milli Hasıla, Sayılarla Türkiye ve Avrupa Birliği, Seçilmiş Ülkeler İçin Ekonomik ve Sosyal Göstergeler

- sayım, anket ya da seçimin yapıldığı yıllarda yayımlanan *Genel Nüfus Sayımı: İdari Bölünüş, Genel Nüfus Sayımı: Nüfusun Sosyal ve Ekonomik Nitelikleri, Genel Tarım Sayımı, Genel Sanayi ve İşyerleri Sayım, Araştırma-Geliştirme İstatistikleri, Bina Sayımı, Katı Atık İstatistikleri, Hava Kirliliği, Hanehalkı İşgücü Anketi Sonuçları, Çocuk İşgücü, Örgün ve Yaygın Eğitim Anketi Sonuçları, Milletvekili Genel Seçim Sonuçları, Mahalli İdareler Seçimi Sonuçları, Ticaret İstatistikleri, Hizmet İstatistikleri, Otel ve Lokanta İstatistikleri, Mali Aracı Kuruluşlar İstatistikleri, Radyo ve Televizyon Kurumları İstatistikleri.*

Ayrıca dönemselsel olarak yayımlanmayan *Gayri Safi Milli Hasıla: Kavram, Yöntem ve Kaynaklar* (1994), *Gelir Yöntemiyle Gayri Safi Yurtiçi Hasıla*, yalnızca 1968, 1973, 1979, 1985, 1990, 1996 yılları için yayımlanan *Türkiye Ekonomisinin Input-Output Yapısı*, zaman zaman basılmış *Hanehalkı Gelir Dağılımı Anketi Sonuçları, İl ve Bölge İstatistikleri, GAP İl İstatistikleri, Ekonomik ve Sosyal Göstergeler, Türkiye İmalat Sanayiinde Yoğunlaşma, İmalat Sanayiinde Teknik İlerleme ve Etkinlik, İmalat Sanayiinde İstihdam, Türk İşgücü Piyasası İle İlgili Temel Gelişmeler, Türkiye’de İstihdam, İşsizlik ve Ücretler, Türkiye’de Çalışan Çocuklar, Türkiye’de Tarımsal Yapı ve İstihdam, Türkiye Nüfusu: Demografi Yapısı ve Gelişimi 21. Yüzyıl Ortalarına Kadar Projeksiyonlar, İstatistiklerle Kadın, Türkiye ve Dünya Dış Ticareti, Türkiye-Avrupa Birliği Dış Ticareti, Tarımsal Göstergeler, Tarihi İstatistikler Dizisi’nde çıkan çeşitli yayınlar da toplumsal konularla ilgilenenler için çok yararlıdır.*

Bunlara ek olarak *Sorularla İstatistik Dizisi*'nde çıkan *Satınalma Gücü ile Milli Gelir, Toptan Eşya ve Tüketici Fiyatları İndeksi, Tarım İstatistikleri, Dış Ticaret İstatistikleri, Sanayi ve Enerji İstatistikleri, İnşaat İstatistikleri, Fiyat İndeksleri ve Enflasyon, Örneklem ve Anket, İstatistiksel Kalite Kontrol* adlı küçük kitaplar yöntem sorunlarını anlamak için iyi birer kaynaktır.

Bütün bunlardan başka DİE çıkardığı aylık haber bültenleriyle en son verileri sıcaklığına kamuoyuna sunmaktadır.

DPT yayınları

DPT Yayınları arasında en kapsamlılarının başında, genellikle beş yılda bir yayımlanan *Beş Yıllık Kalkınma Planları* gelir. Planın temel amaçlarını, politikalarının, üretim hedeflerinin, uygulanması planlanan çeşitli iktisadi politikaların anlatıldığı bu yayında yer alan en önemli veriler ekonominin genel dengesi, yatırım-tasarruf dengesine ilişkin sayısal bilgilerdir. Bu yayının kardeşi olarak nitelendirilebilecek *Beş Yıllık Kalkınma Planı Öncesinde Gelişmeler* adlı yayın da bir başvuru kitabı niteliğindedir. Yıllık aralıklarla yayımlanan **** Yılı Programı, *** Yılı Programı Destek Çalışmaları: Ekonomik ve Sosyal Sektörlerdeki Gelişmeler, *** Yılı Yatırım Programı, *** Yılı İcra Planı* gibi veri bakımından zengin yayınların yanı sıra *Aylık Temel Ekonomik Göstergeler* de iyi birer kaynak kitabıdır.

Hazine Müsteşarlığı yayınları

Bu kamu kuruluşunun yayınlarını da yayımlanma aralıklarına göre şöyle sıralayabiliriz:

- aylık yayınlar: *Başlıca Ekonomik Göstergeler, Aylık Dış Ticaret Bülteni, Hazine Aylık Göstergeleri,*
- üç aylık yayınlar: *Dış Borçlar Bülteni, Türk Mali Sisteminin Temel Göstergeleri,*
- yıllık yayın: *Hazine İstatistikleri,*
- *Ekonomik Kavramlar ve Göstergeler Kılavuzu.*

Merkez Bankası yayınları

Merkez Bankasının “vaziyet”leri, “analitik bilançosu”, para, kredi verilerini içeren *Haftalık Para ve Kredi İstatistikleri Bülteni, Aylık İstatistik Bülteni, Üç Aylık Bülten, Yıllık Rapor* başlıklı yayınları bulunmaktadır. Bu yayınlarda Türkiye ekonomisinin geneline ilişkin verilerin yanı sıra, dış ilişkiler,

bütçe, faiz, döviz kurları, para piyasaları, banka konularında veriler kapsar.

Maliye Bakanlığı yayınları

Genellikle kamu maliyesine, harcama türlerine, gelir türlerine ilişkin ayrıntılı verilerin yer aldığı bu yayınların adları *Devlet Bütçe Gelirleri Aylık Bülteni*, *Kamu Hesapları Aylık Bülteni*, *Yıllık Ekonomik Rapor*, *** *Mali Yılı Bütçe Gerekçesi*'dir. Örnek olmak üzere bir yıllık ekonomik raporu ele alacak olursak şu başlıklara rastlarız: Genel ekonomik hedefler, milli gelir, sabit sermaye yatırımları, üretim, işgücü, istihdam, kamu maliyesi, devlet borçları, para, banka, mali piyasalar, fiyatlar, ödemeler dengesi, dünyadaki ekonomik gelişmeler.

Öteki yurtiçi yayınlar

Çalışma ve Sosyal Güvenlik Bakanlığı'nca yayımlanan *Çalışma Hayatı İstatistikleri* adlı yıllık yayın, toplu iş sözleşmeleri, grev, lokavt uygulamaları, sendikalar, konfederasyonlar, asgari ücret, işyeri sayıları, işçi sayıları konularında veriler içerir.

Sermaye Piyasası Kurulu haftalık, aylık bültenlerin yanı sıra yıllık raporlar, İstanbul Menkul Kıymetler Borsası da günlük, haftalık, aylık, dönemsel bültenlere ek olarak yıllık raporları, *Şirketler Yıllığı* adlı çalışmalarını yayımlamaktadırlar.

İstanbul Ticaret Odası (İTO) yayınları arasında kurumun kendi derlediği ücretliler geçinme indeksleriyle toptan eşya fiyatları indekslerinin gösterildiği *İTO Fiyat İndeksleri* adlı aylık yayın özellikle önemlidir. Ayrıca aylık olarak yayımlanan *Aylık Ekonomik Veriler*, *Dış Ticaret Bülteni*, iki ayda bir çıkan *Ticari Bilgi Bülteni*, yıllık olarak çıkan *Ekonomik Rapor*'dan da söz etmek gerekir.

İstanbul Sanayi Odası (İSO) yayınlarından aylık *İstanbul Sanayi Odası Dergisi*'nin genellikle Eylül (Ekim) aylarında çıkan özel sayılarında Türkiye'nin 500 (onlardan sonra gelen 250) büyük sanayi kuruluşunun satış hasılatı, brüt katma değer, öz sermaye, net aktifler, bilanço kârı, ihracat tutarı, ücretle çalışanlar ortalaması verileri zengin bir kaynaktır. Derginin aynı sayılarında bu veriler çeşitli yazarlarca değerlendirilip yorumlanmaktadır. Son yıllarda söz konusu veriler elektronik yollarla da alınabilmektedir.

Türkiye Bankalar Birliği'nin Merkez Bankası, Ticaret Bankaları, Kalınma ve Yatırım Bankaları'na ilişkin verilerin yer aldığı *Aylık Göstergeler* ile yıllık olarak yayımlanan *Bankalarımız* adlı yayınları da önemli veri kaynaklarıdır.

TÜSİAD'ın zaman zaman yayımlanan özel konulara ilişkin çalışma raporlarının yanı sıra Türkiye ekonomisine ilişkin belli başlı verilerin toplu olarak bulunabileceği *** *Yılına Girerken Türk Ekonomisi, Türkiye Ekonomisi* yayınlarından söz edilebilir.

TOBB'nin *Üç Aylık Ekonomik Rapor'u* ile yıllık *İktisadi Rapor'u* da genel veriler içermektedir.

Bunların yanı sıra çeşitli bankalar, işçi, işveren sendikaları da değişik aralıklarla bültenler yayımlamaktadır. Bunlardan Türk-İş yıllıkları ile Türkiye Petrol, Kimya, Lastik İşçileri Sendikası'nın yıllıkları toplu veriler için iyi birer kaynaktır.

Eğitimle ilgili konularda DİE'den başka veri yayımlayan bir kuruluş da Yüksek Öğrenim Kurumu'na bağlı Öğrenci Seçme Yerleştirme Merkezi'dir. Bu merkez ÖSS puan dağılımları, yüksek öğrenim istatistikleri gibi konularda yıllık yayınlar yapmaktadır.

5.3 YABANCI KAYNAKLAR

Uluslararası kuruluşların sunduğu veri kaynaklarına gelince, bunlar başta İngilizce olmak üzere Fransızca, Almanca, İspanyolca gibi dillerde yayımlanmaktadır. Bunların çoğu elektronik ortamda elde edilebilir.

OECD yayınlarından *OECD Economic Surveys* tek tek üye ülkelere ilişkin yıllık raporlardır. *OECD Economic Outlook* yılda iki kez yayımlanır. Ulusal gelir, dış ticaret, ödemeler dengesi, istihdam ve işgücü istatistikleriyle öncül gösterge bulgularını içerir. *Main Economic Indicators* adlı aylık yayın da benzer bilgiler sunar. Konuları adlarından anlaşılacak *Quarterly National Accounts* (Üçer Aylık Ulusal Hesaplar), *Monthly Statistics of Foreign Trade* (Aylık Dış Ticaret İstatistikleri), *Financial Statistics Monthly* (Aylık Mali İstatistikler), *Financial Market Trends* (Mali Piyasaların Eğilimleri), *Quarterly Labour Force Statistics* (Üçer Aylık İşgücü İstatistikleri) gibi yayınlar da OECD tarafından çıkarılır.

Uluslararası Para Fonu (IMF) tarafından yayımlanan başlıca veri kaynakları *International Financial Statistics* (Uluslararası Mali İstatistikler), *Balance of Payments Statistics Yearbook* (Ödemeler Dengesi İstatistikleri Yıllığı), *Direction of Trade Statistics* (Dış Ticaret İstatistiklerinin Yönü), *Government Finance Statistics Yearbook* (Kamu Maliyesi İstatistikleri Yıllığı), *Annual Report* (Yıllık Rapor), *World Economic Outlook* (Dünyaya İktisadi Bakış) olarak sayılabilir.

Birleşmiş Milletler Örgütü'nün veri içeren yayınlarından başlıcaları şunlardır: *Monthly Bulletin of Statistics* (Aylık İstatistik Bülteni), *Statistical Yearbook* (İstatistik Yıllığı), *World Statistics in Brief* (Özet Dünya İstatis-

tikleri), *International Trade Statistics Yearbook* (Uluslararası Dış Ticaret İstatistikleri Yıllığı), *Handbook of International Trade and Development Statistics* (Uluslararası Ticaret ve Kalkınma İstatistikleri Yıllığı), *National Account Statistics* (Ulusal Hesaplar İstatistikleri).

Dünya Bankası'nın *Financial Flows to Developing Countries* (Kalkınmakta Olan Ülkelere Mali Akımlar), *The World Bank Annual Report* (Dünya Bankası Yıllık Raporu), *World Development Report* (Dünya Kalkınma Raporu), *World Development Indicators* (Dünya Kalkınma Göstergeleri), *World Tables* (Dünya Çizelgeleri), *World Debt Tables* (Dünya Borç Çizelgeleri), *The World Bank Atlas* (Dünya Bankası Atlası), *Global Economic Prospects* (Küresel İktisadi Gelecek) adlı yayınları da istatistik veri bakımından zengindir.

UNDP (Birleşmiş Milletler Kalkınma Programı) örgütünün yıllık ara-
lıklarla çıkardığı önemli bir yayın da *Human Development Report* (İnsani Gelişme Raporu) adlı yayındır.

Yukarıda özet olarak belirtilen veri kaynaklarına ilişkin en ayrıntılı bilgi, bu kaynakların kendilerine bakılarak edinilebilir. Buradaki özetin bir yol göstericilikten öteye gidemeyeceği doğaldır. Bu amaçla aşağıdaki alıştırma-
lar arasında çeşitli kuruluşların bilgisayarla ulaşılabilen ana sayfa adresleri verilmiştir.

Okuma Önerileri

Barzun, J., Graff, H. F. (1993) *Modern Araştırmacı*, 7. b., (Çeviren: F. Dilber) TÜBİTAK (45-86).

DİE (2001b) *Yayın Kataloğu, 1927-2000*, DİE Matbaası.

DİE (2002b) *Yayınlar ve Hizmetler Kataloğu 2002*, DİE Matbaası.

Sel, C. (1994) "Ekonomik bilgilerin elde edilebileceği basılı ve on-line kaynaklar", *Ekonomiyi İzleme Sempozyumu 5-6 Mayıs 1994* içinde (sayfa numaraları yok) Hacettepe Üniversitesi.

Tütengil, C. O. (1975) *Sosyal Bilimlerde Araştırma ve Metod*, 3. b., İÜ İktisat Fakültesi (17-41).

ALİŞTIRMALAR

5.1 Yurt içinde yayın yapan bazı kuruluşların bilgisayarla erişilebilen ana sayfa adresleri aşağıdadır. Bu adreslere girerek hangi verileri bulabileceğinizi araştırın.

<http://www.die.gov.tr>

<http://www.hazine.gov.tr>

<http://www.tcmb.gov.tr>

Devlet İstatistik Enstitüsü

Hazine Müsteşarlığı

Merkez Bankası

<http://www.imkb.gov.tr>
<http://www.tobb.org.tr>
<http://www.ito.org.tr>
<http://www.iso.org.tr>
<http://www.tbb.org.tr>
<http://www.tüsiad.org.tr>
<http://www.musiad.org.tr>
<http://www.turkis.org.tr>
<http://www.disk.org.tr>
<http://www.hakis.org.tr>
<http://www.tisk.org.tr>
<http://www.mess.org.tr>

- 5.2 Yurt dışında yayın yapan bazı kuruluşların bilgisayarla erişilebilen ana sayfa adresleri aşağıdadır. Bu adreslere girerek hangi verileri bulabileceğinizi araştırın.

<http://www.worldbank.org>
<http://www.imf.org>
<http://www.undp.org>
<http://www.unesco.org>
<http://www.unctad.org>
<http://www.oecd.org>
<http://www.opec.org>
<http://www.ilo.org>
<http://www.nyse.com>
<http://www.economist.com>
<http://www.ft.com>
<http://www.wsj.com>
<http://www.gpn.org>
<http://www.iceg.org>
<http://www.nber.org>

<http://www.imd.cu>

İstanbul Borsası
 Odalar ve Borsalar Birliği
 İstanbul Ticaret Odası
 İstanbul Sanayi Odası
 Türkiye Bankalar Birliği
 TÜSİAD
 MÜSİAD
 Türk-İş
 DİSK
 Hak-İş
 TİSK
 MESS

Dünya Bankası
 Uluslararası Para Fonu
 BM Kalkınma Programı
 BM Eğitim, Bilim, Kültür Örgütü
 BM Ticaret, Kalkınma Konferansı
 İktisadi İşbirliği, Kalkınma Örgütü
 Petrol Satan Ülkeler Örgütü
 Uluslararası Çalışma Örgütü
 New York Borsası
The Economist
Financial Times
Wall Street Journal
 Küresel Politika Kurumu
 Uluslararası İktisadi Büyüme Merkezi
 Ulusal İktisadi Araştırma Merkezi
 (ABD)
 Uluslararası Yönetim Geliştirme
 Enstitüsü

Örnekleme Rassal mı Olsun?

Az olsun, öz olsun.

Atasözü

Bir önceki bölümde yayımlanmış kaynaklardan ne gibi veriler bulabileceğimizi gördük. Şimdi de kullanacağı verileri kendi derlemek isteyen bir araştırmacının nasıl bir yol izlemesi gerektiğini ele alacağız. Toplumsal bilimlerde de temel bilimler ya da mühendislikte olduğu gibi deneyler yapılabilir, ancak bu yol bir kaç alan dışında seyrek kullanılır. Onun yerine “alana çıkmak” diye bilinen yöntem daha sık başvurulur, yani **örnekleme** yapılır. Örnekleme bir üretim sürecinden çıkan ürünler arasında yapılabileceği gibi insanlar arasında da uygulanır. Aşağıda bu kavram ayrıntılı olarak açıklanacaktır.

6.1 TAMSAYIM – ÖRNEKLEME

Bir istatistik çalışmasında, hakkında sayısal bilgiler derlenecek gözlem birimlerinin hepsini içeren topluluğa **anakütle** dendiğini biliyoruz. Belli bir günde bir ülkede yaşayan insanların hepsi, bir ilde kayıtlı motorlu taşıtların hepsi, bir üniversitede kayıtlı öğrencilerin hepsi, bir malı gelecek yıl satın alabilecekleri düşünülen insanların hepsi, bir bölgede bir yıl içinde gerçekleşen kendini öldürmelerin hepsi, bir ülkede bir ayda alım-satıma konu olan fiyatların hepsi, bir işyerinde belli bir sürede üretilen malların hepsi, gelecek seçimde oy kullanabilecek seçmenlerin hepsi birer anakütle örneği olarak düşünülebilirler. Anakütleyi oluşturan gözlem birimlerinin hepsinin birden gözlenmesine **tamsayım** denir. Nüfus sayımları bu tür tamsayımların *ani* (kısa bir sürede) yapılanlarına örnek olarak gösterilebilir. Bir ilde motorlu taşıtların kaydı sürekli olarak tutulur. Bu da anakütlenin *sürekli* sayımının bir örneğidir.

Tamsayımlar, zorunlu durumlar dışında pek kullanılmazlar. Bir anakütleyi oluşturan gözlem birimlerinin toplam sayılarının bulunmasında tamsayımlar işe yarayabilmektedir. Ancak anakütle içindeki gözlem birim-

lerinin toplam sayılarının ortaya çıkarılmasında bile tamsayımlar eski saygının yerlerini koruyamamakta, onların yerine, bilgisayar olanaklarının da gelişmesinin etkisiyle, **örnekleme** (eski deyimle numune alma, sonday) geçmektedir.

Gözlem birimlerinin toplam sayılarından çok, belli değişkenlere göre aldıkları değerler ya da özellikleri incelenmek istendiğinde tamsayım yapmak çoğu zaman gereksiz olabilir. Öyle ya, denizin suyunun tuzlu olduğunu anlamak için hepsini içmek gerekmez.

Gelişen örnekleme teknikleri kullanılarak tamsayımlara göre daha az hatalı sonuçlar elde edilebilir. Bu inanılması güç bir sav gibi görünebilir ama nüfus sayımlarında tamsayımından vazgeçmenin önemli gerekçelerinden biri budur. Her nüfus sayımından sonra sayım memurlarının uğramadığı evler, hatta semtler ortaya çıkabilmektedir. 2000 nüfus sayımındaki aksaklıkların giderilebilmesi için sonradan epey çalışılması gerekmiş, sonuçların yayımlanması bu nedenle bir hayli gecikmiştir.

Nüfus sayımının bir günde tamamlanabilmesi için çok fazla sayım memurunun görev alması gerekmektedir. Bu sayının yüksekliği, giderleri arttırmanın ötesinde, sayım memurlarının yeterince yetiştirilememesi gibi sakıncaları da beraberinde getirmektedir. Dolayısıyla soru kâğıtları doldurulurken yanlışlar yapılabilmektedir. Daha az sayım görevlisiyle, daha iyi düzenlenmiş bir örnekleme, tamsayımına göre daha güvenilir sonuçlar alınmasını sağlayabilir. Örnekleme yoluna gidildiğinde sonuçların ne ölçüde yanlışlık içerebileceğinin hesaplanabilir olması tamsayımlardan vazgeçmenin önemli bir nedenidir.

Örnekleme yapmanın tamsayımına yeğlenmesinin öteki nedenleri arasında, örneklemenin tamsayımına göre hem daha ucuz olması, hem daha çabuk bitirilmesi gösterilebilir. Bazı durumlarda tamsayımın olanaksız olduğu da unutulmamalıdır. Bir işyerinde üretilen metal bir parçanın basınca karşı direncinin ölçüldüğü, belli bir düzeyin altında kalan parçaların fire sayıldığı bir durumu düşünelim. Üretilen parçaların her birinin ne kadar basınç altında kırıldığını saptayan bir düzeneğimiz olsun. Üretilen bütün parçaları bu düzenden geçirerek ne kadarlık bir basınç altında kırıldığını gözlemlemeye kalkmak, elde satılacak parça kalmaması anlamına gelir. Böyle bir durumda tamsayım yapılamayacağı açıktır.

Gerek tamsayımlarda, gerek örneklemelemlerde gözlem birimlerine ilişkin çeşitli değişkenlerin aldıkları değerlerin yazılması için özel olarak düzenlenmiş çizelgeler, soru kâğıtları kullanılır. Bunların hazırlanma-kullanma yolları, ayrı bir uzmanlık gerektirmesi nedeniyle burada ele alınmayacaktır (bkz. Baş, 2001, Duverger, 1980:195-229, Sencer, Irmak 1984:115-217, Bernard, 2000:227-276).

6.2 RASSAL ÖRNEKLEME – RASSAL OLMAYAN ÖRNEKLEME

Bir anaküleyi oluşturan birimlerin hepsinin değil de küçük bir bölümünün incelemeye alınmasına karar verildikten sonra sıra, hangi birimlerin gözleneceğini belirlemeye gelir. Yalnızca incelemeye alınan birimlere bakarak anakütleyle ilişkin sonuçlar çıkarmak istediğimize göre, örneklemeye alınan birimlerin anakütlenin incelenen özelliği konusunda iyi bilgi vermeleri gerekir. İyi bilgi, yansız, etkin bilgidir. **Yansızlık, etkinlik** kavramları örneklemeye konusunun en önemli kavramlarından ikisi oldukları için özenle açıklanmaları yararlı olur.

Bir anakütlenin herhangi bir özelliğiyle, diyelim bir fabrikada üretilen belli bir elektronik aygıtın ortalama ömrüyle ilgileniyor olalım.

Ortalama ömrün anakütledeki değerini μ (mü) ile gösterelim. Bu fabrikanın ürettiği bütün aygıtların ömürlerini ölçmemiz olanaksız olabileceği için bu işi bir örneklem üzerinde yapmamız uygun olur. Bu örneklem seçilen aygıtların ömürlerini ölçerek ortalamasını hesaplırsak, bulduğumuz değeri anakütlenin ortalamasını kestirmek için kullanabilir miyiz?

Örneklemeye seçilen aygıtların ömürlerini ölçüp toplayalım, aygıt sayısına bölelim. Bulduğumuz değere örneklem ortalaması diyebiliriz. Bu ortalama değerini $\bar{X}$ (X çizgi) ile gösterirsek, aynı gözlem sayısını içeren bu tür çeşitli örneklemelerden bulunacak $\bar{X}$ değerlerinin, anakütle ortalamasının genellikle altında ya da genellikle üstünde olması istenmeyen bir durumdur. Örneklem ortalamasının ($\bar{X}$), anakütle ortalamasını (μ) genellikle düşük ya da genellikle yüksek kestirdiği durumlarda örneklemde elde edilen bilginin **yanlı** ya da **sapkın** olduğu söylenebilir. Örneklem bilgisinin yansız olabilmesi için örneklem ortalamasının, ana kütle ortalamasını genellikle düşük ya da genellikle yüksek kestirmemesi, tam tersine bazen düşük, bazen yüksek ama ortalama olarak doğru kestirmesi, bir başka deyişle örneklem ortalamalarının beklenen değerinin ana kütle ortalamasına eşit olması gerekir.

Beklenen değer kavramı şu anlama gelir: Eğer bu aygıtın bütün üretiminden aynı büyüklükte ne kadar farklı örneklem çekebileceksek hepsini çektiğimizi, bunların her birinin ortalamasını teker teker bulduğumuzu düşünelim. Bu $\bar{X}$ 'lerin bir kez daha ortalamasını alalım. Bulduğumuz ortalamaların ortalamasını $\bar{\bar{X}}$ (X çift çizgi) ile gösterelim. İşte bu değer $\bar{\bar{X}}$ 'nin beklenen değeridir, $E(\bar{\bar{X}})$ ile gösterilir. Bu durumda $\bar{\bar{X}}$, μ 'nün **kestirimcisi** olarak adlandırılabilir. Bu kavramlarla yansızlık şöyle gösterilebilir:

$$\text{Yansız durum: } E(\bar{\bar{X}}) = \mu \quad \text{Yanlı (sapkın) durum: } E(\bar{\bar{X}}) \neq \mu$$

Stuart (1984:5)'tan uyarladığımız küçük bir örneklem yansızlığın ne olduğunu açıklayalım. Gerçek yaşamda pek rastlanmayacak kadar küçük bir

anakütle alıp işlemlerimizi bunun üzerinde gösterelim. Anakütleyi küçük almamızın nedeni, içinden seçilebilecek bütün değişik örneklemelerin sayısını düşük tutabilmektir. Aksi halde örneklem sayısı o kadar büyür ki, söylemek istediğimizi anlatamaz duruma düşeriz. Öte yandan anakütledeki bütün birimler genellikle gerçek yaşamda gözlenemez. Biz burada bunların gözlenebileceğini varsayıyoruz. Bunu yapmazsak yansızlığın ne olduğunu böyle bir örneklemle anlatamayız.

Anakütlemizi oluşturan A, B, C, D, E, F adları verilmiş altı aygıtın tek tek ömürleri sırasıyla 102, 106, 108, 110, 110, 112 saat olarak ölçülmüş olsun. Her birinden 100 çıkarıp (etkin basamak kavramını anımsayalım) işlemleri 2, 6, 8, 10, 10, 12 üzerinden yapmak kolaylık sağlar. Bulunan değerlere en sonunda 100 ekleyip özgün ölçümlere dönmek de kolaydır. Altı sayıyı toplayıp altıya bölersek anakütle ortalamasının $\mu = 8$, yani 108 saat olduğu açıktır.

Biz altı aygıt arasından dörder gözlemlilik bütün örneklemeleri seçip her birinin ortalamasını bulacak, sonra bu ortalamaların da ortalamasını alıp kaç olduğuna bakacağız. Eğer 8 bulursak “örneklem ortalaması, anakütle ortalamasının yansız bir kestirimcisidir”, yani $E(\bar{X}) = \mu$ diyebileceğiz. Yok eğer 8’den farklı bir sayı çıkarsa örneklem ortalamasının, anakütle ortalamasının yanlı (sapkın) bir kestirimcisi, yani $E(\bar{X}) \neq \mu$ olduğu sonucuna varacağız.

Altı gözlemlilik bir anakütleden çekilebilecek dört gözlem içeren 15 farklı örneklem çekilebilir. Bunların her birine giren dörder gözlemin değerleri aşağıdadır. Son satırdaki sayılar örneklem ortalamalarıdır.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	A	A	A	A	A	A	A	A	A	B	B	B	B	C
B	B	B	B	B	B	C	C	C	D	C	C	C	D	D
C	C	C	D	D	E	D	D	E	E	D	D	E	E	E
D	E	F	E	F	F	E	F	F	F	E	F	F	F	F
2	2	2	2	2	2	2	2	2	2	6	6	6	6	8
6	6	6	6	6	6	8	8	8	10	8	8	8	10	10
8	8	8	10	10	10	10	10	10	10	10	10	10	10	10
10	10	12	10	12	12	10	12	12	12	10	12	12	12	12
6,5	6,5	7	7	7,5	7,5	7,5	8	8	8,5	8,5	9	9	9,5	10

En alt satırdaki 15 örneklem ortalamasının toplamı 120 eder, bunun 15’e bölümü de 8’i verir. Yani $E(\bar{X}) = \mu$ olur. Demek ki örneklem ortalaması, anakütle ortalamasının yansız bir kestirimcisiymiş.

Gerçek yaşamda anakütle ortalamasının 8 (ya da 108) saat olduğunu bilemeyiz. Bu 15 örneklemde de çoğu zaman yalnızca birini çekebiliriz.

Ama görüldüğü gibi, hangi örneklemi çekersek çekelim bunun ortalaması 6,5 – 10 arasında sıralanan $\bar{X}$ 'lerden biri çıkacaktır. O halde bu anakütleden bir örneklem çekmekle, bu 15 örneklem ortalamasından birini çekmek aynı anlama gelir.

Yukarıda örneklem bilgisinin yansız olabilmesi için örneklem ortalamasının, ana kütle ortalamasını genellikle düşük ya da genellikle yüksek kestirmemesi, tam tersine bazen düşük, bazen yüksek ama ortalama olarak doğru kestirmesi gereğinden söz etmiştik. İşte bu örneğimizde örneklem ortalaması, anakütle ortalamasını bazen 6,5, 7, 7,5 gibi 8'den küçük değerlerle, bazen 8'e eşit değerlerle, bazen da 8,5, 9, 9,5 10 gibi 8'den büyük değerlerle, ama *ortalama olarak* 8'e eşit kestirmektedir. Başka bir deyişle örneklem ortalamamızın beklenen değeri ile anakütle ortalamamız arasında fark yoktur. Demek ki bu örnekte örneklem ortalamasının, anakütle ortalamasının yansız bir kestirimcisi olduğunu gösterebildik.

Örneklemden elde edilen bilginin yansız olmasının yanısıra etkin olması gerektiğinden de söz etmiştik. Yukarıdaki örneğimize dönecek olursak bu tür örneklemelerden elde edilecek $\bar{X}$ değerlerinin μ değerinin yakınlığında gezinmesi, fazla uzaklaşmaması durumunda örneklemden elde edilen bilginin etkin olduğu söylenebilir. Anakütle ortalaması $\mu = 108$ saat iken $\bar{X}$ değerleri 106,5 ile 110,0 saat arasındadır. Hatta bunlardan % 80'i 107 ile 109,5 arasında değerler almaktadır. Bu yeterli bir etkinlik midir?

Buna yanıt verebilmek için etkinliğin görece bir kavram olduğunu söyleyelim. Yukarıda anakütle ortalamasını kestirmeye çalışırken aritmetik ortalamayı bir kestirimci olarak kullandık. Aritmetik ortalama yerine başka bir ortalama hesaplama yöntemi kullanıp daha dar bir aralıkta kestirimler yapabilirsek daha etkin bir kestirimci bulduğumuz söylenebilir.

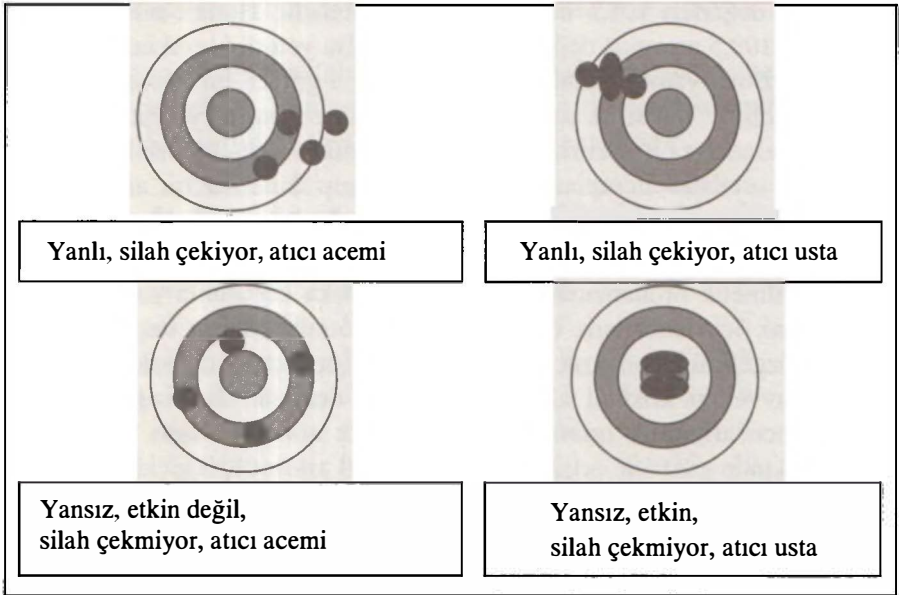
Şunu belirtelim ki anakütle ortalamasının yansız kestirimcileri arasında, aritmetik ortalamadan daha etkin bir başka kestirimci yoktur. Bunu anlamak için isterseniz bu 15 örneklemin ortalamalarını başka bir yöntemle, sözgelimi geometrik ortalamayla, yani örneklem değerleri çarpımının 4. kuvvetten kökünü alarak hesaplamayı deneyin. Bulacağınız aralık daha önce aritmetik ortalamayla bulduğumuz aralıktan daha geniş çıkacaktır. İlkinde anakütle ortalaması 106,5 ile 110 arasında, ikincide 105,6 ile 109,9 arasında *bir yerde tahmin edilecektir* (Geometrik ortalama konusu anlatıldıktan sonra bunun doğrulanması okuyucudan istenecektir). Yani aritmetik ortalama, geometrik ortalamaya göre daha etkindir.

Aslında buradaki etkinlik standart sapma ya da varyansla ölçülürse daha duyarlı sonuçlar elde edilir. Ama bu kavramları henüz görmediğimiz için bu işi de daha sonraya bırakıyoruz.

Etkinliği artırmanın bir yolu örneklemi büyütmektir. Bu örneğimizde dörtlü değil de beşli örneklem çekseydik elde edeceğimiz örneklem orta-

lamalarının en küçüğü 7,2, en büyüğü de 9,2 olurdu. Bunun doğrulanmasını da okuyucuya bırakıyoruz. Öte yandan, üçer gözlemlili örneklerle etkinliğin azlacağını göstermesi yine okuyucudan istenecektir.

Yansızlık-etkinlik kavramlarını atıcılıktan örnekler vererek de açıklayabiliriz. Bir silahla atış yapıldığını düşünelim. Burada her atış bir örneklem alma ya da bir örneklem ortalaması bulmaya karşılık gelsin. Bu atışlarda silahtan ya da atıcıdan ileri gelen hatalar söz konusu olabilir. Eğer silah tam hedefe değil de biraz yana (ya da aşağıya, yukarıya) doğru ateş ediyorsa “silah çekiyor” denir. İşte bu durum **yanlı** (sapkın) kestirimciye karşılıktır. Eğer silahın böyle bir bozukluğu yoksa, bu durum da **yansız** kestirimciye benzetilebilir. Atıcı usta değilse, birbirine pek yakın olmayan atışlar yapacaktır. Bu da **etkin olmayan** duruma karşılıktır. Usta atıcının atışlarıysa birbirine çok yakın olacaktır. Bu da (silahın çekmemesi koşuluyla) **etkinlik** gibi düşünülebilir. Çizim 6.1 silahtan ya da atıcıdan doğabilecek aksaklıklarla ilgili dört ayrı durumu göstermektedir.



ÇİZİM 6.1 Yansızlık-etkinlik özellikleri

Bunlardan üsttekilerde silah çekiyor. Atıcı usta da olsa pek işe yaramıyor. Böyle bir durum yanlı (sapkın) kestirimcilerle çalışmaya benzer. Son çizimde ise her şey düzgün. Hem silah hiçbir yana çekmiyor, tam hedef alınan yere ateş ediyor; hem atıcı usta, vuruşları birbirine çok yakın. Öyleyse bu da hem yansız, hem etkin bir kestirimciyle iş görmek gibidir.

Buraya kadar yansızlık, etkinlik özellikleri bir kestirimcinin marifetleri gibi sunuldu. Örneklem almanın bu işle ne ilgisi var diye sorulabilir. Eğer örneklem çekme yönteminizde yanlışlık yaparsanız, kestirimciniz ne olursa olsun, onun yansızlığı ya da etkinliği konusunda bilgi edinemeyebilirsiniz.

Bir örnekleme kestirimcinin hem yansız hem de etkin bilgiler vermesi çok iyi olur. Öyleyse örnekleme işini öyle becermeliyiz ki hem yansızlığı hem de etkinliği sağlayabilelim. Her iki özelliği birden sağlayabilmek için **rassal örnekleme** denilen yollardan yararlanmak gerekir. Rassal olmayan (keyfi) örnekleme yollarıyla örneklem seçilirse yansızlık-etkinlik koşullarının yerine gelmesi bir yana bu kavramlara ilişkin hiçbir bilğimiz olmayabilir. Bu nedenle istatistik araştırmalarında örneklemin rassal seçilmesine büyük önem verilir.

Rassallık kavramının istatistikte özel bir anlamı vardır. Günlük yaşamda kullandığımız *rast gele* sözcüğünün anlamı bunu karşılamaz. Rast gele demek, kafadan atarak, içten geldiği gibi demektir. Oysa bu, rassallığı sağlamayabilir. Bir kimseden, fazla düşünmeye fırsat vermeden sıfır ile dokuz arasında bir sayı söylemesi istendiğinde çoğu kimsenin yedi sayısını verdiği görülmüştür. Bunun nedenleri ruhbilim araştırmalarına konu olmuştur (Miller, 1956).

Bizi hemen yedi sayısını söylemeye iten neden, haftanın yedi günü, dünyanın yedi harikası, yerin yedi kat dibi, göğün yedi katı, güneşin yedi rengi, yediveren gülü, müzikteki yedi nota, yedi uyur (eshab-ı kehf) söylencesi, eski Yunan söylencesindeki Atlas'ın yedi kızı, yine aynı söylence-deki Niobe'nin yedi oğluyla yedi kızı, Anadolu'daki en eski yedi kilise, yedi ünlü Arap şairinin yedi kasidesinin duvarlara asılması demek olan Yedi Askı, Simbadname'deki yedi bilge öyküsü, İstanbul'un yedi tepesi, yine İstanbul'daki Yedikule semti, edebiyatımızdaki yedi meşaleciler, halk kültürümüzdeki yedi kardeşe yedi gelin, Pamuk Prenses masalındaki yedi cüce, yerli müzikli oyun Yedi Kocalı Hürmüz, Bertold Brecht ile Kurt Weill'in müzikli oyunu Yedi Ölümcül Günah (filmi de çevrildi), Kurosawa'nın filmi Yedi Samuray, Ian Fleming'in yarattığı 007 James Bond, bir zamanlar ABD petrol piyasasını elinde tutan Yedi Kız Kardeş denilen yedi şirket, dünyanın en güçlü ekonomilerine sahip G-7 adlı yedi ülke gibi örneklerin yediden yetmiş kadar hepimizin bilinç altında yer etmesi midir acaba?

Nedeni ne olursa olsun içimizden geldiği gibi sayı söylerken yedi sayısını öteki sayılardan çok daha fazla kullanıyorsak bu yolla rassallık sağlamıyoruz demektir. Çünkü 0 – 9 arasındaki her sayıya eşit şans verdiğimizizi düşünüyor ama bunu uygulamıyoruz. Bir başka deyişle on birimlik bir anakütleden bir birimlik bir örneklem seçmek istediğimizde çoğunlukla yedinci birimi seçersek bütün birimlere eşit şans tanımadığımız halde tanıdığımızı sanarak yanılıyoruz.

Milli piyangoyu kazanan numaraların nasıl çekildiğini hemen herkes televizyonda görmüştür. Büyük cam kürelerin içinde, üzerinde 0 – 9 arasında sayılar yazılı toplar sürekli olarak oradana oraya çarpıp dolaşmaktadır. Bir düğmeye basmakla bu toplardan biri bir delikten çıkmakta, kazanan numaranın bir basamağını oluşturmaktadır. Böylelikle çekilen bir sayının rassal olduğu söylenebilir. Ancak bunun en iyi yolu **rassal** (tesadüfi) **sayılar** kullanmaktır. Rassal sayılar kafadan atılarak değil, bilgisayarlar kullanılarak türetilir. Bu sayılar herhangi bir kurala bağlı olmadan dizilen, her birinin bu dizilişteki yeri kendinden öncekilerden **bağımsız** olarak belirlenen sayılardır. Art arda çekilen rassal sayıların birbirinden bağımsızlığını anlayamazsanız Demirkan'ın (2003) haberine göre milli piyango'nun 2004 yılbaşı çekilişinde bazı numaraların daha şanslı olduğunu ileri süren uzman (!) gibi komik (ya da acıklı) duruma düşersiniz

Örneklemin rassal olmasını sağlayan bu sayılar cebe sığabilen elektronik hesap makinelerinin pek çoğunda bir düğmeye dokunmakla elde edilebilmektedirler. Ayrıca hemen her istatistik yazılımında rassal sayı türeten alt programlar bulunmaktadır.

Rassal sayı kullanmanın önemini, Vietnam'a gönderilmek üzere seçilen ABD'li gençlerin öyküsünü anlatırken görmüştük. ABD'den bir başka örnek daha verelim (Ott, Longnecker 2001:169). Korkunç bir işsizliğin kol gezdiği 1936 başkanlık seçimleri öncesi *Literary Gazette* adlı bir yayın organı, 10 milyon soru kâğıdı dağıtıp kime oy verileceğini saptamaya çalışmıştı. Geri gelen 2,4 milyon yanıtta Demokrat Parti adayı Roosevelt % 43, bugün adını kimsenin anımsamadığı öteki aday (Landon) % 57 oy alacak gibi görünüyordu. Ama seçim tam tersi bir sonuç verdi. Roosevelt açık farkla kazandı. Gazetenin yanlış rassal olmayan örneklemeye başvurmasydı. Soru yönelttiği 10 milyon kişiyi telefon rehberleriyle "sosyal klup"lerin üye listelerinden bulmuştu. Üstelik bunların da dörtte birinden azı yanıt vermişti. Yanıt verenler genellikle daha varlıklı kesimlerden idi. Oysa Roosevelt'in oylarının çoğu yoksul kesimlerden gelmişti. Aynı seçimin sonuçlarını George Gallup 3.000 kişilik rassal bir örneklemle doğruya çok yakın kestirmişti. Bu başarısıyla Gallup büyük ün kazandı.

Rassal örnekleme alma yollarına geçmeden rassal olmayan örneklemlerden de kısaca söz edelim.

Sırf uygulanması kolay olsun diye en rahat ulaşılabilecek birimleri örnekleme almak rassal olmayan bir yöntem uygulandığı anlamına gelir. Burada temsil özelliğinin zedeleneyeceği çok açıktır. Şu örneklere bir göz atalım.

- Bir fabrikadan dağıtım kamyonlarına mal yüklenirken en üstten birkaç parça alıp bunların belli özelliklerini ölçmek,
- Futbol maçlarındaki kavgalarla ilgili bir araştırma yaparken yalnız kendi tuttuğumuz takımın yandaşlarıyla görüşmek,
- Önümüzdeki seçim sonuçlarını kestirmek için öğrencilere ödev olarak yakın çevrelerinin seçimde nasıl oy kullanacaklarını saptamalarını istemek,
- Canlı yayımlanan bir TV programında izleyicilerin belli konudaki görüşlerini bildirmek için telefonla bağlantı kurmalarını istemek.

Başka bir çok örnek sayılabilir, ama hepsinde örneklemenin rassal yapılmadığı çok bellidir. Başka örneklerde durum bu kadar açık olmayabilir ama doğacak sonuçlar daha az zararlı değildir.

Rassal olmayan başka bir yol da araştırmacının kendi kişisel değerlendirmesiyle temsil özelliği taşıdığına inandığı birimleri örnekleme seçmesidir. Eğer değerlendirmeleri gerçekçiye bu yolla iyi bir örneklem seçilebilir. Ama bundan nasıl emin olabiliriz ki? Bilgi eksikliği, önyargı, beğeni farklılığı örneklem seçiminde sapmalara yol açabilir.

Zaman zaman keyfi seçimle birimlere ulaşılmadaki kolaylık bir arada kullanılır. Sokakta sizinle görüşüp bazı sorular sormak isteyen “anketçi”lerle karşılaşmışsınızdır. Genellikle bu tür örnekleme anketi yapacaklara yaşa, medeni duruma, mesleğe, vb özelliklere göre kaçır kişiyle görüşmeleri gerektiği söylenir. Onlar da bu özelliklere uyduğunu düşündükleri kimselere yanaşıp sorularını sorarlar. Bunu yaparken ters bir tutum takınmayacak kimseleri seçmeye çalışırlar. Her yaştan ya da her meslekten kaçır kişiyle görüşülmesi gerektiği saptanırken uygun bir tasarımdan yararlanmış olunabilir ama bu, örnekleme giren kişilerin rassal seçilmesini sağlamaya yetmez.

Bu türden bütün örnekleme seçme yöntemlerini rassal olmayanlar kümesine katmak gerekir. Bu tür örneklemlerden elde edilen bilgilerin anakütleye genellenmesi sakıncalıdır. Çünkü bu genellenmenin nasıl bir hesaplama dayanağını anlatan örnekleme kuramı, örneklemenin hep rassal alındığı varsayımına dayanır. Örneklem rassal değilse örnekleme kuramı orada işe yaramaz. Ancak bazı özel durumlarda rassal olmayan örnekleme yollarından yararlanılması kaçınılmaz olabilir (Bernard, 2000:174).

6.3 RASSAL ÖRNEKLEME TÜRLERİ

Yukarıda rassal sayı kullanılarak çekilen örneklemelerle yansız, etkin kestirimcilerle ulaşılabileceğinden söz edilmişti. Rassal örneklemenin temel özelliği anakütledeki her birimin örneklem seçilme şansının bilinmesidir. Bunların birbirlerine eşit olmaları gerekmez, yalnızca biliniyor olmaları yeterlidir. (Stuart, 1984:3). Bunun sağlandığı bütün örneklemelere rassal örneklem denilebilir.

Böyle rassal sayı kullanılarak çekilen örneklemeler de çekiliş biçimlerine göre değişik adlar alırlar, ama hepsinin ortak özelliği anakütledeki her birimin örneklemde yer alma şansının önceden belli olmasıdır. Bunlardan en çok bilineni *basit rassal örneklem*dir. Öteki yaygın kullanılanlar *atlamalı rassal örneklem*, *katmanlı rassal örneklem*, *kümeli rassal örneklem*, *aşamalı rassal örneklem* türleridir. Aşağıda bunlar teker teker ele alınıp açıklanacaktır.

Basit rassal örneklem

Her gözlemin yalnızca şansa bağlı olarak çekildiği bir örneklemidir. Böylelikle

- hem anakütledeki her gözlem biriminin örneklem girme şansını birbirine eşit olmakta,
- hem de bu anakütleden seçilebilecek aynı büyüklükteki her örneklem seçilme şansını birbirine eşit olmaktadır.

Ayrıca hiçbir birim örneklem birden çok kez giremez. Yukarıda kullandığımız örnekte altı birimlik anakütleden dörder gözlemli örneklem seçilmişti. Aslında burada yapılan basit rassal örneklemekten başka bir şey değildi. Gerçekten de A, B, C, D, E, F gözlemlerinin her biri 15 örneklem onar tanesinde yer aldıklarına göre her gözlemin örneklem girme şansını $10/15 = 2/3$ olmaktadır. Bu anakütleden dörder gözlemli 15 farklı örneklem çekilebildiğine göre, çekilebilecek her örneklem, sözgelimi ACEF örnekleminin ya da başka herhangi bir örneklemin çıkma şansını $1/15$ 'tir. Az önce sözü edilen iki özellik de sağlandığına göre yaptığımız işlem basit rassal örneklem çekmektir.

Basit rassal örneklem yapılabilmesi için anakütledeki bütün birimlerin bir dökümünün oluşturulabilmesi gerekir. Bu dökümdeki her birime bir sıra numarası verilmeli, daha sonra rassal sayılar çekilerek bu sayılara karşılık gelen sıradaki birimler örneklem alınmalıdır. Diyelim ki 332 öğrencisi olan bir bölümün öğrencileri arasından 20 öğrenciyi basit rassal örneklem yoluyla seçeceğiz. İlk olarak öğrenci işleri görevlisinden bütün öğrencilerin

dökümünü alıp her öğrenciye 1'den 332'ye kadar bir sıra numarası veririz. Daha sonra hesap makinesinden, bilgisayardan ya da rassal sayılar çizelgelerinden yararlanarak rassal sayı çekme işlemine geçeriz.

Anakütle büyüklüğü 332 sayısı üç basamaklı olduğundan, hangi araç kullanılırsa kullanılsın bulunan rassal sayının da üç basamaklı olması gerekmektedir.

Eğer çekilen üç basamaklı rassal sayı 1 – 332 arasındaysa bir sorun çıkmaz. Çekilen sayıya karşılık gelen öğrenci örnekleme alınır. Ama 332'den büyük rassal sayılara karşılık gelen hiçbir sıra numarası olmadığından ya bu rassal sayı kullanılmaz, ki o zaman bu örnekte çekilen rassal sayıların üçte ikisine yakını kullanılamayacak demektir (bunun nedenini bulmak okuyucuya bırakılmıştır), ya da şu yola başvurulur. Çıkan rassal sayı

333 – 664 ($= 2 \times 332$) arasındaysa 332 çıkarılarak kalan alınır,
665 – 996 ($= 3 \times 332$) arasındaysa 664 çıkarılarak kalan alınır,
996'dan büyükse o sayı kullanılmaz.

Son yolun *modulus*¹ hesabı olduğunu bazı okuyucular hemen anlamış olmalıdır. Yani çıkan rassal sayı 332'ye bölündüğünde kalan, rassal sayı gibi işlem görmektedir. Bu yollardan biri kullanılarak bulunan rassal sayılar şunlar olsun:

100	850	353	396	231	695	243	782	375	168
904	055	450	054	489	084	584	665	148	321

Bu durumda örnekleme seçilecek öğrencilerin sıra numaraları aşağıdaki gibi olur. Bunların doğrulanması okuyucuya bırakılmıştır.

100	186	21	64	231	31	243	118	43	168
240	55	118	54	157	84	252	1	148	321

Dikkatli bir okuyucu 118 numaralı öğrencinin örnekleme iki kez seçildiğini görmüştür. Bu durumda yapılacak iş yeni bir rassal sayı çekmek olmalıdır. Diyelim 516 çıktı. O zaman örnekleme ($516 - 332$) 184 numaralı öğrenci girecek demektir.

Bu örnekte kullanılan basit rassal örnekleme yolunda, anakütleyi oluşturan 332 öğrenciden her birinin örnekleme seçilme şansı eşit olduğu gibi, bu anakütleden çekilebilecek 20 birimli bütün örneklemelerin her birinin seçilme şansı da birbirine eşittir. Acaba böyle kaç örneklem vardır? 332

¹ Çoker, Karaçay (1983:481) “modulus” karşılığı “ölçke” terimini önermektedir.

sayıdan 20'şerli kaç örneklem oluşturabilir. Lisede öğrendiği kombinasyon hesabını anımsayan okuyucu bunu kolayca yanıtlayabilir: 332'nin 20'li kombinasyon sayısı yani $332! / (20! \times 312!)$. *Faktöryel*² adı verilen ! simgesi, peşine takıldığı sayıdan geriye doğru 1'e kadar sayıların art arda çarpılacağı anlamına gelir. Öyleyse 1'den 332'ye kadar bütün sayılar çarpılacak, bulunan çarpım önce 1'den 20'ye kadar olan sayıların çarpımına, sonra da 1'den 312'ye kadar olan sayıların çarpımına bölünecektir. Bu işlemin sonucunu el hesap makinesiyle bulamayabilirsiniz. Bu işlemin sonucunun ne kadar büyük bir sayı olduğunu bulmak okuyucuya bırakılmıştır.

Örneğimizdeki bölümün öğrenci sayısı 332 değil de 1332 olsaydı, bulunacak rassal sayının da dört basamaklı olması gerekecekti. Ancak hesap makinelerinden bazıları ancak üç basamaklı rassal sayı türetmektedir. Bu üç basamağın, kullanılacak rassal sayının soldaki (ya da sağdaki) üç basamağı olacağına önceden karar verilip, çekilecek ikinci üç basamaklı sayının yine önceden karar verilmiş bir basamağı, rassal sayının dördüncü basamağı olarak kabul edilebilir. Böylece oluşturulan dört basamaklı rassal sayıya karşılık gelen sıra numarasına sahip birim örneklem seçilmiş sayılır.

Basit rassal örneklem kolay anlaşılabilen, görece olarak kolay uygulanabilen bir örneklem seçme biçimidir. Örneklemde elde edilecek aritmetik ortalama, anakütle ortalamasının yansız bir kestirimcisidir. Bu özellikleriyle anakütleyle ilişkin kestirim yapılmasında elverişli bir araçtır. Daha da ileri giderek diyebiliriz ki (bu kitapta ayrıntısına girilmeyecek olan) örneklem kuramının öğretilmesinde temel taş işlevi görür.

Ancak çok gözlemlili büyük bir örneklem alırken ya da anakütlerdeki birimlere birer sıra numarası vermek olanaksızken bu yöntemi kullanmak uygun düşmez. Bu durumda atlamalı rassal örneklem işimize yarayabilir.

Atlamalı Rassal Örneklem

SistematiK (DİE, 1998; Orhunbilge vö., 1997), **dizgesel** (İnal vö. 1983), **dizgeli** (Sencer, 1981) örneklem adlarıyla da anılan bu yöntemin kimi uygulamalarında anakütle birimlerinin hepsini içeren bir döküm gerekir. Ancak böyle bir döküm olmadan da atlamalı rassal örneklem kullanılabilir. Giderek denebilir ki anajütlenin birim sayısı bilinmese bile atlamalı rassal örneklemekten yararlanılabilir. Anakütle birimlerinin özelliğine bağlı olarak

- anakütleyi oluşturan dökümdeki gözlem birimleri belli aralıklarla atlanarak,

² Çoker, Karaçay (1983), İnal vö. (1983) "factorial" karşılığı "çarpım" terimini önermektedir.

- anakütleyi oluşturan gözlem birimleri ortaya çıktıkça bunlar arasından belli birim sayısı kadar aralıklarla atlanarak,
- anakütleyi oluşturan gözlem birimleri bir zaman sırası izliyorsa, bunlar arasından belli zaman aralıklarıyla atlanarak,
- anakütleyi oluşturan gözlem birimleri arasından belli mesafelerle atlanarak örneklem seçilebilir.

Anakütledeki birim sayısı N , örnekleme seçilecek birim sayısı n , atlama aralığı k olsun. $N / k = n$ ise işler kolaydır. 1 ile k arasında bir sayı rassal olarak seçilir. Bu sayıya karşılık gelen birim ilk gözlem olur. Daha sonra bu sayıya k eklenerek ikinci gözlem, buna k eklenerek üçüncü gözlem, ... bulunur. İlk çekilen sayıya $(n - 1)$ kez k eklendiğinde sonuncu gözlem elde edilmiş olur. Böylece örnekleme seçilen gözlem sayısı n 'ye eşitlenir. Bölümümüzde 340 öğrenci olsaydı, biz de 20 öğrencilik bir örneklem seçmek isteseydik $340 / 17 = 20$ olduğundan ilk gözlemi 1 ile 17 arasından çekilmiş bir rassal sayıyla elde ederdik. Diyelim ki 5 çıktı. Bundan sonra örnekleme girecek öğrencileri gösteren sayılar $5 + 17 = 22$, $5 + 2 \times 17 = 39$, ..., $5 + 19 \times 17 = 328$ diye belirlenirdi.

Dikkat edilirse ilk öğrenci rassal olarak seçilmiş, ondan sonrakiler hiçbir rassal süreçten geçmemiştir. İlk çekilen sayı belli olduktan sonra ötekiler, buna k 'nin katları eklenerek bulunmuştur. Dolayısıyla 5 sayısı, örnekleme girecek bütün birimlerin hangileri olacağını belirlemektedir. 5'in yer aldığı bir örnekleme 22, 39, ..., 328 dışındaki birimlerin seçilmesi olanaksızdır. İşte bu özellik nedeniyle her örneklemin seçilme şansı birbirine eşit değildir. Sözelimi 5, 19, 33, ..., gibi bir örneklemin bu yöntemle seçilmesi olanaksızdır. Bir başka deyişle aralarındaki fark k olmayan herhangi iki sayı aynı örnekleme yer alamaz.

Peki $N / k = n$ eşitliğinin geçerli değilse ne yapılabilir? Yukarıda verilen örnekte olduğu gibi, 332 öğrencinin içinden 20 öğrencilik bir seçme yapmayı atlamalı rassal örnekleme yoluyla yapmak istesek, 332'yi 20'nin bir eksiği 19'a bölüp kalanı buluruz. Bölüm 17, kalan 9 olduğuna göre ilk 9 gözlem arasından rassal seçilen birini örnekleme alırız. 4 çıkmış olsun. Sıra numarası 4 olan öğrenci örnekleme seçilmiş olur. Geriye kalan $n - 1$ gözlem az önce gördüğümüz yolla belirlenir. $4 + 17 = 21$, $4 + 2 \times 17 = 38$, ..., $4 + 19 \times 17 = 327$ sıra numaralı öğrenciler örnekleme girer.

Özetlersek, $N = (n - 1)k + t$ olsun. Burada $0 < t \leq k$ eşitsizliği geçerlidir. $t < k$ ise ilk t gözlem arasından biri rassal seçilir. Daha sonra her k 'ninci gözlem örnekleme alınır, n büyüklüğünde bir örneklem elde edilir. Burada n , N / k 'ye eşit ya da ondan büyük ilk tamsayıdır (Barnett, 1991:35). Eğer $t = k$ ise ilk k sayı arasından biri rassal seçilir, sonra her k 'ninci gözlem örnekleme alınır.

Bu yöntem özellikle sık örneklem alınan durumlarda ya da örneklemin iş sürerken çekilmesinin zorunlu olduğu ortamlarda yaygın olarak kullanılır.

Bu örnekleme yolu kullanılırken özellikle uyanık olunması gereken nokta, örnekleme dizgesel (sistematik) bir hatanın sokulmaması olmalıdır. Bu **dizgesel hata** kavramını şöyle bir örnekle açıklayalım. İki yanındaki evlerin doğalgaz ile ısıtıldığı bir caddedeki 200 ev arasından atlamalı rassal örneklemeyle 20 ev seçtiğimizi düşünelim. Bu evlerin ya hepsi tek numaralı ya da hepsi çift numaralı olmak zorundadır. (Neden?) Ev numaraları da caddenin bir yanındaki evlerde tek, öteki yanındaki evlerde çift olduğundan örnekleme giren evler caddenin hep aynı yanından olacaktır. Eğer örnekleme seçilen evlerin hepsi soğuk kuzey rüzgârlarına açık yandaysa bu evlerde yapılan soruşturma ısınma konusunda pek çok yakınma olduğunu ortaya çıkarabilir. Yok eğer öteki yanda kuzey esintilerinden pek etkilenmeyen evlerden oluşan bir örnekleme soruşturma yapılırsa, bu yakınmalara pek rastlanmayabilir. Bu demektir ki her iki örneklem de **yanlı** (sapkın) sonuçlar doğurur.

Başka bir örneği kalite denetimi alanından seçelim. Bir meyve suyu dolmuş yerinde, dolu şişelerin kasalara doğru ilerlediği şerit üzerindeki her yüz şişeden bir tanesinin laboratuarda incelenmek üzere örnekleme seçildiği bir düzen kurulduğunu varsayalım. Bu yürüyen şeride şişe veren A, B, C, D gibi dört doldurma makinesinin de aynı hızda çalıştığını düşünelim. Böylece art arda gelen her dört şişe, dört ayrı makineden, sözgelimi B, D, C, A sırasıyla gelmiş olur. Bu durumda da örnekleme alınan her yüzüncü şişe bu dört makinenin yalnızca birinden gelmiş olur. Rassal seçilen sayı 4'ün tam katıysa şişelerin hepsi A makinesinde, 4'e bölününce 1 artıyorsa hepsi B'de, 2 artıyorsa hepsi D'de, 3 artıyorsa hepsi C'de doldurulmuş şişeler elde edilir. Öyleyse kalite denetimi yalnızca bir makine için yapılmış, öteki üç makinenin şişeleri kalite denetimi dışında kalmış olur. Burada da alınan örneklem yanlış kararlara yol açar.

Bu tehlike dikkate alınarak doğru biçimde uygulanan bir atlamalı rassal örnekleme, böylesi dizgesel hatalardan arınmış olabilir. Tek bir rassal sayı çekmek gibi bir kolaylığa sahip olması nedeniyle de yaygın olarak kullanılır.

Katmanlı Rassal Örnekleme

Bu seçme yöntemine **tabakalı örnekleme** (DiE, 1998; Orhunbilge vö., 1997) adı da verilir. Basit rassal örnekleme için verdiğimiz öğrenci seçme örneğine geri dönelim. Burada elde ettiğimiz örneklem anakütleyi temsil

etmede yetersiz kalabilir mi? Sözelimi dört sınıfın her birinden yeterince öğrenci örnekleme girmiş midir?

Diyelim ki 332 öğrenci dört sınıf arasında eşit dağılmış olsun. Yani sıra numaralarına göre 1 – 83 arasındakiler birinci sınıf, 84 – 166 arasındakiler ikinci sınıf, 167 – 249 arasındakiler üçüncü sınıf, 250 – 332 arasındakiler dördüncü sınıf öğrencisi olsun. Örnekleme seçilen öğrencilerin sınıflara göre dağılımlarının da buna benzer olması, sınıfların örneklemede iyi temsil edilmesini sağlar. Yani 20 kişilik örneklemede her sınıftan beşer kişinin örnekleme girmesi sınıfların daha iyi temsil edilmesini sağlayabilirdi. Aca-ba örneğimizde durum nedir?

Birinci sınıftan 7 öğrenci (1, 21, 31, 43, 54, 55, 64), ikinci sınıftan 6 öğrenci (84, 100, 118, 148, 157, 168), üçüncü sınıftan 5 öğrenci (184, 186, 231, 240, 243), dördüncü sınıftan 2 öğrenci (252, 321) örnekleme girmiştir. Görüldüğü gibi sınıflar örneklemede eşit temsil edilmemektedirler. Bunun ne gibi bir sakıncası olabilir? Eğer inceleme konumuz öğrencilerin hangi sınıfta olduklarından bağımsız bir olguysa hiçbir sorun yoktur. Ama ya durum böyle değilse, öğrencilerin ilgilendiğimiz özelliği okudukları sınıfla yakından ilişkiliyse durum tatsızdır. Sözelimi öğrencilerin not ortalamalarıyla, barındıkları yerle, üniversiteye uyum sağlama düzeyleriyle ilgileniyorsak, bunların öğrencinin bulunduğu sınıftan bağımsız olduğunu söyleyebilir miyiz? Bu durumda bize her sınıftan aynı oranda öğrenci seçmeyi sağlayan bir örnekleme yöntemi gerekmektedir. İşte bu yöntem de katmanlı rassal örneklemedir.

Eğer anakütledeki katmanların örneklemede de benzer oranlarda temsil edilmesini istiyorsak katmanlı rassal örnekleme bu konuda bize yardımcı olabilir. Özellikle anakütle içindeki oranı düşük olan, dolayısıyla da örneklemede az ya da hiç temsil edilememe olasılığı yüksek olan katmanlara ilişkin bilgilerin, araştırmamız için özel bir önemi varsa bu yol kullanılmalıdır.

Bir tencere pilav çok iyi karıştırılmışsa bundan bir kaşık almakla tadının ya da tuzunun yerinde olup olmadığını anlayabiliriz. Bu işlemi basit rassal örneklemeye benzetebiliriz. Türdeş anakütlelerde basit rassal örneklemeler, temsil yeteneği yüksek örneklemelerdir. Ama pilav iyi karıştırılmamışsa tencerenin çeşitli yerlerinden birer kaşık almamız daha doğru sonuç verir. Bunun gibi anakütle türdeş değilse, anakütleyi oluşturan katmanların araştırma konumuz bakımından farklılıklar gösterdiği düşünülüyorsa, her katmanın anakütle içindeki payı saptanarak örneklem içinde de uygun oranlarda yer almaları sağlanır. Böylelikle hem örneğin temsil yeteneği yükseltilmiş, hem de örneklem büyüklüğünü daha düşük düzeylerde tutma olanağı sağlanmış olur (Günlük-Şenese, 1988:133).

Bir pazar araştırmasında hedef alıcı kitlesinin, cinsiyet, yaş, sosyo-ekonomik konum, vb. bakımlardan türdeş katmanlara ayrılarak her katmana

örneklemin içinde yeteri kadar yer vermek uygun görülebilir. Seçimlerle ilgili bir kamuoyu araştırmasında seçmen kitlesini katmanlara ayırmada, cinsiyet, yaş, sosyo-ekonomik konum özelliklerinin yanısıra bölgenin siyasal eğilimi gibi bazı özellikler de kullanılabilir. Süper marketlerle ilgili bir araştırma yapılırken bunların çoğunun küçük ya da orta ölçekli, bir kaç tanesinin büyük olduğu, büyüklerin satışının toplam içinde önemli bir pay oluşturduğu da biliniyorsa, örnekleme girecek süper marketler seçilirken bu olgular dikkate alınarak katmanlara ayırma yoluna gidilmesi doğru olabilir.

Burada gerçek yaşamdan bir örnek alalım (Neter, 1989). ABD'de kendi hatlarında birbirlerinin yükünü de taşıyan iki demiryolu şirketi, aralarındaki borç-alacak hesaplaşmasını katmanlı örnekleme yoluyla yapmaktadırlar. Bu örnekte, hesaplaşma süresi içinde yaklaşık 23.000 yük taşıma belgesi düzenlenmiştir. Eğer örnekleme yapılmazsa bunların her birinin teker teker incelenmesi gerekir. Toplam borç-alacak tutarının belgelerin tek tek tutarlarına bağlı olduğu açıktır. Eğer belgelerin tutarları birbirine yakın değerlerdeyse basit rassal örnekleme çok uygun bir yol olabilir. Oysa belge tutarlarının 2 \$ ile 200 \$ arasında değiştiği bilinmektedir. Basit rassal örnekleme yoluna gidilirse sayıları çok az olan büyük tutarlı belgelerin örnekleme girememesi tehlikesi ortaya çıkacaktır. Bunun üzerine anakütleyi oluşturan 22.984 belge aşağıdaki gibi katmanlarına ayrılmış, her katmandan ne oranda örneklem seçileceği ayrı ayrı saptanmıştır:

Katmanlar (\$)	2 -<5	5 -<10	10 -<20	20 -<40	40 +
Örnekleme girme oranı	% 1	%10	%20	%50	%100

İlk katmandaki belge tutarları dar bir aralıkta (3 \$) yer aldığından bunlardan % 1'lik çok küçük bir oranın örnekleme girmesi yeterli görülmüş, buna karşılık dördüncü katmandaki belgelerin yarısının, son katmandakilerin ise hepsinin örneklemede yer alması uygun bulunmuştur. Böylece sonucun gerçeğe yakın çıkması sağlanmıştır.

Bu örnek biraz daha uzatılıp yukarıda açıklanan atlamalı rassal örnekleme yolunun burada nasıl kullanıldığı da gösterilebilir. İlk katmandaki yaklaşık 9000 belge arasından basit rassal seçim yapmak oldukça zaman alıcıdır. Bunun yerine atlamalı rassal örnekleme kullanılarak iş çok kolaylaştırılabilir. 5 \$'dan az tutarlı belgelerin yalnızca % 1'i örnekleme alınacağına göre 00 ile 99 arasındaki sayılardan biri rassal olarak seçilse, belge numaralarının son iki basamağı buna uyan belgeler örnekleme alınarak sorun çözülebilir. Rassal sayı 02 seçildiğinden sonu 02 ile biten belgelerin tutarları örneklemede yer almıştır. Benzeri işlemler öteki katmanlarda da yapılmıştır. Böylece derlenen 2.072 belge tutarı (toplamın yaklaşık % 9'u) incelenerek şirketlerin birbirlerine olan borç-alacakları saptanmış, her kat-

man için toplam borç-alacak bulunmuş, bunlar da toplanarak genel tutar hesaplanmıştır. Bu uygulama yalnızca bir deneme olduğundan sonucun doğruluk derecesinin anlaşılması amacıyla belgelerin bir de tamsayım yoluyla incelenmesi yapılmıştır. Örneklemden elde edilen toplam borç-alacak tutarı 64.568 \$ iken tamsayım ile bulunan sonuç 64.651 \$ çıkmıştır. Aradaki fark 83 \$'dır. Öte yandan, örneklem giderleri 1.000 \$'dan azken, tamsayım için harcanan para 5.000 \$ tutmuştur.

Yamane (2001:121)'ye göre bu örneklem yönteminin kullanılması belli koşullarda duyarlılığı artırmakta, her katmana ilişkin bilgi elde edilebilmekte, verilerin toplanması kolaylaşabilmektedir. Artan duyarlılık, anakütleyle ilişkin kestirimlerin etkinliğinin artması anlamına gelir. Bu kavramı daha önce kullandığımız mini anakütleli örneğimizle açıklayalım.

Anakütlemizi oluşturan A, B, C, D, E, F adları verilmiş altı aygıtın ömürlerinin sırasıyla 102, 106, 108, 110, 110, 112 saat bulunduğunu anımsayalım. Bunlardan basit rassal örneklem yoluyla çektiğimiz dördü örneklemelerin ortalamalarının en az 106,5, en çok 110 çıktığını, ortalama olarak da 108 saat olduğunu bulmuştuk. Acaba katmanlı örneklem seçerek gene yansız ama daha dar bir aralıkta örneklem ortalamaları elde edebilir miyiz? Bunu başarabilirsek, basit yerine katmanlı örneklem almakla kestirimcimizi daha **etkin** kılabiliriz, bir başka deyişle anakütle ortalamasını kestirirken birbirine daha yakın iki sınır değeri elde edebiliriz demektir.

Bir an için katmanların şöyle olduğunu düşünelim.

Katman I	A, B, D	2, 6, 10
Katman II	C, E, F	8, 10, 12

Amacımız yine anakütle ortalamasının beklenen değerini bulmaktır. Bunu iki katmandan ayrı ayrı örneklemeler olarak yapacağız. Önce her iki katmandan eşit oranda birimi örneklem seçelim, daha sonra bunu değiştireceğiz. Basit yöntemde dördü örneklem aldığımız için şimdi her iki katmandan da ikişer olmak üzere yine dördü örneklem çekebiliriz. Bunların sayısı 9'dur. Aşağıdaki çizelgede bu 9 örneklem giren birimlerle bunların değerleri gösterilmiştir. Örneklemelerin ortalamaları yine son satırdadır.

	1	2	3	4	5	6	7	8	9
	A	A	A	A	A	A	B	B	B
I	B	B	B	D	D	D	D	D	D
	C	C	E	C	C	E	C	C	E
II	E	F	F	E	F	F	E	F	F
	2	2	2	2	2	2	6	6	6
I	6	6	6	10	10	10	10	10	10
	8	8	10	8	8	10	8	8	10
II	10	12	12	10	12	12	10	12	12
	6,5	7	7,5	7,5	8	8,5	8,5	9	9,5

Görüldüğü gibi 9 örneklemin ortalamalarının toplamı 72'dir, bunun da örneklem sayısı 9'a bölünmesiyle örneklem ortalamalarının beklenen değeri 8 (yani 108 saat) çıkar. Demek ki kestirimcimiz bu yöntemde de yansızdır. Acaba etkinliği biraz olsun artmış mıdır? En küçük örneklem ortalamasıyla en büyüğü arasındaki farka bakarak bu konuda bir fikir edinebiliriz. Basit rassal örnekleimde bu aralık 106,5 ile 110 arasındaydı, burada ise 106,5 ile 109,5 arasındadır. Öyleyse bu aralık az da olsa küçülmüştür. Demek ki katmanlı örneklem almakla kestirimcimizi daha etkin kılabilirdik.

Bunun nedenini şöyle açıklayabiliriz. Basit rassal örnekleimde küçük birimlerin hepsinin örnekleme girmesi çok küçük bir örneklem ortalamasına yol açar. Tam tersine örneklemin büyük birimlerden oluşması da örneklem ortalamasını büyütür. Ama katmanlamaya başvurduğunuzda birinci katmanda görece küçük değerler, ikinci katmanda da görece büyük değerler yer aldığından böyle bir durumla karşılaşmayız. Demek ki örneklem ortalamalarının uç değerleri basit rassal örnekleimde, katmanlı örneklemeye göre daha dışta olur.

Doğruyu söylemek gerekirse etkinlik burada biraz kabaca ölçülmektedir. Bu işi daha duyarlı yapabiliriz ama henüz ortalamadan sapmaların nasıl ölçüldüğünü görmediğimiz için şimdilik bununla yetiniyoruz. Ama ileride varyans ya da standart sapma hesaplamasını öğrendiğimizde bu etkinlikleri yeniden ölçme yoluna gideceğiz.

Bir soruya daha değinelim. Acaba katmanlar buradaki gibi değil de başka türlü ayrılıyorsa etkinlik yine artar mıydı? Bunun sınanması da yine okuyucuya bırakılmıştır.

Son örneğimizde her iki katmanda da aynı oranda birimi örnekleme seçtik. Acaba katmanlardan farklı oranlarda birimi örnekleme seçseydik sonuç nasıl çıkardı? Bu durumda her birimin örnekleme girme şansı aynı olmaz. Bir katmandakilerin şansı artarken ötekinde olanlarınki azalır. Bunun hiç bir sakıncası yoktur. Yeter ki ne yaptığımızı unutmayalım, her

katmanın örneklem ortalamasını uygun biçimde hesaba katalım. Diyelim ilk katmandan üç, ikinci katmandan bir birimi örneklem almaya karar verdik. Aşağıdaki çizelgede yer alan üç örneklemle karşılaşırız. Burada önce her katmanın ortalaması alınmış, sonra bunların da ortalaması alınarak örneklem ortalamasına ulaşılmıştır. Son satır örneklem ortalamalarını göstermektedir.

Örneklem	1	2	3
Katman I	A 2	A 2	A 2
	B 6	B 6	B 6
	D 10	D 10	D 10
Katman II	C 8	E 10	F 12
Katman I ortalaması	6	6	6
Katman II ortalaması	8	10	12
Örneklem ortalaması	7	8	9

Bu örneklem biçiminde de kestirimcimiz yansızdır. Çünkü örneklem ortalamalarının ortalaması ya da beklenen değeri yine 8 (yani 108 saat) çıkmıştır. Ya etkinlik diyecek olursanız çizelge bize etkinliğin bir hayli arttığının ip uçlarını vermektedir. Artık örneklem ortalamaları 107 – 109 gibi çok daha dar bir aralıkta yer almaktadır.

Dikkat ettiyseniz ilk katmanımızın kendi içindeki değişme aralığı (102 – 110), ötekinkine (108 – 112) göre epeyce geniştir. Öyleyse dar aralıklı olandan daha az, geniş aralıklı olandan daha çok birimi örneklem alarak bu özellikten yararlandık, böylece örneklem ortalamalarının değişim aralığını iyice daralttık. Kısacası, değişkenliği çok olan katmandan daha çok birimi örneklem katmak her zaman yararlıdır (Stuart, 1984:40).

Aşamalı rassal örneklem

Bu yöntem **çok-aşamalı** (DİE, 1998; İnal vö., 1983; Sencer, 1981), **çok kademeli** (Orhunbilge vö., 1997) gibi adlarla da anılır. Uygulamalarda yukarıda açıklanan örneklem teknikleri aşamalı olarak birlikte kullanılabilirler. Örneğin yurt çapında yapılacak bir gelir dağılımı araştırmasında bütün hane halklarının bir listesi çıkarılıp bunlar arasından rassal seçim yapılamaz. Ya ne yapılır?

Bunu uygulamadan bir örnekle açıklayalım (Esmer, vö. 1986:III-IV). İller beş bölgeye ve seçmen sayısı bakımından üç katmana (>400 bin, 200-400 bin, <200bin) ayrılmış, her katmandan birer il (toplam 15 il) rassal olarak seçilmiş, ayrıca bunlara İstanbul, Ankara, İzmir eklenmiştir. İkinci

aşamada her ilden merkez ilçeye ek olarak iki ilçe nüfusları gözönüne alınarak rassal olarak seçilmiştir. Daha sonra örnekleme giren bütün yerleşim birimleri önce kır-kent olarak ikiye ayrılmış, kentlerde bütün mahalleler gönenç (refah) düzeylerine, siyasal eğilimlerine göre katmanlandırılarak, her katmandan mahalleler, sokaklar, haneler rassal olarak seçilmiştir. Kırsal yerlerde ise seçmen nüfusuna ve siyasal eğilime göre katmanlar oluşturulmuş, bu katmanlardan rassal olarak köyler seçilmiştir. Daha sonra da hanelerle görüşülerek örnekleme yaşama geçirilmiştir.

Bir başka örnek daha verip iki çalışmada kullanılan yöntemlerin benzer yönlerine dikkat çekelim. *"...(B)irinci aşamada seçmen kitlesi coğrafi temele göre yedi bölgeye ayrılmış, her bölge içinde kırsal ve kentsel kesim ayrımına gidilmiştir. İstanbul, Ankara ve İzmir büyükşehir belediyeleri sınırları içinde kalan yerleşim birimleri ise ... seçim bölgelerine ayrılmıştır. ... (Örneklemin) bölgelere ve kırsal-kentsel, metropoliten alanlarına göre dağılımının anakütleyle orantılı olması ilkesi benimsenmiştir (tir). ... Metropollerde "mahalle", geri kalan hem kentsel, hem de kırsal kesimlerde "ilçe" birimi kullanılmıştır. ... Mahallelerde sokak seçimi, sokaklarda hane seçimi, hanelerde kişi seçimi (uygulanmıştır). ... İlçelerde de köy ve bucaklar arasından rassal seçim yapılmış, ...hanelere ve kişilere ulaşılmıştır."* (TÜSİAD, 2001:II, 13-17).

Bir örnek de iş dünyasından verelim. Araba lastiği satıcılarıyla yüzyüze görüşme yapacak bir araştırmacı örneklemini şöyle seçmektedir (ATM, 1988:5-14). Çok aşamalı, katmanlı rassal örnekleme yöntemine başvuran araştırmacı, ilk aşamada satıcıları, dağıtımını yapmakta oldukları markalara göre katmanlara ayırmıştır. Her katmandan anakütle dağılımına orantılı olarak örnek seçmiştir. İkinci aşamada satıcıları bölgelere göre yeniden katmanlara ayırmış, her bölgede hangi firmadan kaç satıcıyla görüşüleceğini yine anakütle içindeki oranlara göre saptamıştır. Üçüncü aşamada illerin seçimi gelmektedir. İstanbul, Ankara, Adana, Antalya illerini doğrudan örnekleme alan araştırmacı, öteki illeri, o ildeki satıcı sayısının bölge satıcı sayısına oranını temel alarak rassal seçmiştir. Son aşamada her ilde görüşülecek satıcıları da rassal seçimle saptamıştır.

Son örneği de DİE'nin çalışmalarından alalım (DİE, 1997a:ix-x). Gelir dağılımı araştırması için tasarlanan katmanlı örnekleme de nüfusu 20.000'e kadar olan yerler kır, daha kalabalık yerleşimler kent sayılmıştır. Hem kır, hem kent yerleşimleri yine nüfuslarına göre kendi içlerinde üçer tabakaya daha ayrılmıştır. Ayrıca 7 coğrafi bölge de birer katman olarak alınmıştır. Daha sonra bu katmanlarda hep rassal örnekleme uygulanmıştır.

Kümeli Rassal Örneklem

Bu yönteme **küme örneklemesi** (DİE, 1998; İnal vö., 1983; Sencer, 1981), **kümelere göre örneklem** (Orhunbilge vö., 1997) gibi adlar da verilir. Anakütleyi oluşturan birimlerin bir listesi yoksa ya da tek tek rassal olarak seçilecek birimlerin çok yaygın bir alana dağılmış olmaları nedeniyle ulaşıl-maları pahalıya çıkıyorsa, kümeli rassal örneklemeye başvurulabilir. Burada örneklem anakütle birimleri tek tek girmez, bu birimlerden oluşan kümeler girer. Rassal olarak seçilen kümelerdeki bütün birimlere tamsayım uygulanır.

Bir bölgedeki iş koşullarını araştırmak için, o bölgede çalışan işçilerin tam bir dökümü elde edilebilse bile, bunlar arasından seçilecek rassal bir örneklem girecekler, çok çeşitli yerlerdeki değişik işyerlerine dağılmış olacaklarından, bunlara tek tek ulaşmak çok güç olabilir. Bu durumda iş-yerleri arasında rassal bir seçimle bir örneklem oluşturmak, daha sonra bu işyerlerine gidilerek oralaradaki bütün işçilerle görüşmek uygun bir yol ola-bilir. Bu da kümeli rassal örneklem kullanıldığı anlamına gelir.

Bu örneklem tekniğinin kullanılmasıyla zamandan, paradan kazanıldı-ğı, uygulamada önemli kolaylıklar sağlandığı açıktır. Ancak önemli bir sakıncaya kapı açıldığı da unutulmamalıdır. Basit rassal örneklemde tek tek birimlere tanınan eşit seçilme şansı, burada yalnızca kümelere tanınmış-tır. Eğer küme içi birimler kendi aralarında ilişkili iseler, ki genellikle öyle-dirler, işin içine dizgesel (sistematik) hatanın girmesi söz konusudur. Bu nedenle kümeli örneklemeler, aynı büyüklükteki basit rassal örneklemelere göre genellikle daha yüksek bir örneklem hatası içerirler. Öyleyse bunu bile bile neden böyle bir yola başvurulur? Bunun nedeni paradan, zaman-dan, emekten edilen tasarrufun bu sakıncadan daha önemli olmasıdır. Eğer durum böyleyse katlanılan maliyete göre daha çok etkinlik kazanılmış ola-bilir (Stuart, 1984:63).

Küçük anakütle örneğimize bir de bu yöntemle yaklaşalım. Üç birimli örneklem seçeceğimizi varsayalım. Nedenlerinin tartışmasına girmeden anakütlemizi şöyle iki kümeye ayırdığımızı düşünelim.

Küme I	2, 6, 8	Küme II	10, 10, 12

Bu iki kümeden biri rassal olarak seçilecekse iki değişik örneklem ortaya çıkabilir. İlkini örneklem ortalaması 5,33, ikincinin 10,67'dir. Bunların ortalamasının 8 (yani 108 saat) ettiği açıktır. Öyleyse yansız bir kestirimci kullanıyoruz demektir. Ama etkinliğe ne oldu diye soracak olursanız, azaldı-ğını kendiniz de görebilirsiniz. Çünkü örneklem ortalamalarının aralığı 105,33 – 110,67 saat arasına genişlemiştir. Gerçi burada dört yerine üç birim-li örneklem seçmek de değişim aralığını genişletmiştir. Alıştırmalarda bu konuyu açıklığa kavuşturması okuyucudan istenecektir.

6.4 DENEY TASARIMI

Bilimsel araştırmalarda ortaya atılan önsavların doğruluğunu sınamak için örneklemelerden elde edilen bilgiler kullanılabileceği gibi düzenlenen deneylerden sağlanan veriler de yararlı olur. Bu doğrulama işlemi gözlemlerle yapılacaktır. Deneyin tasarlanması, bu gözlemlerin nasıl elde edileceği konusunda bir model kurmak demektir (Çömlekçi, 1988:6). Bu tasarlama iyi yapılmazsa elde edilen veriler işe yaramaz, yanlış (sapkın) olabilirler.

1. Bölümden alıntılarla deney tasarımı açıkla mak yerinde olur. Deney tasarımı, toplanacak verilerin işe yarar olabilmesi için nasıl derlenmeleri gerektiğini gösterir. Deneye girişmeden önce tasarımı üzerinde kafa yormak genellikle meyvelerini verir. İyi tasarlanmış bir deneyin bulgularını çözümlmek kolaylaşır, çoğu zaman sonuçlar hemen göze çarpar. Buna karşılık deney kötü tasarlanmışsa elde edilen veriler başlangıçta sorulan sorulara yanıt veremez.

Bir deney, genellikle iki (bazen daha çok) etmen arasında bir nedensellik ilişkisinin araştırılması için kurulur. Bu etmenlerden birindeki bir değişme ötekinde de değişmeye yol açıyor mu? Açılırsa bu değişim ne yönde gerçekleşiyor? Değişimin büyüklüğü ne kadar? Her zaman benzer büyüklükte bir etki doğuyor mu? Bu tür soruları yanıtlamaya uğraşan araştırmacı, bir ya da bir kaç girdinin, öteki adıyla **bağımsız değişkenin** çeşitli değerlerini seçerek, her bir durumda çıktının ya da **bağımlı değişkenin** değerlerine ne olduğunu gözleyebileceği bir deney tasarlar. Amaç, girdi ile çıktı arasındaki ilişkiyi anlamak, bunun gerisinde yatan süreci açıklamaya çalışmaktır. Öyleyse bir deney tasarımının özü şu üç soruya yanıt bulmaktır: Hangi etmenleri ele almalıyız? Bunların düzeyleri nasıl değişmelidir? Bu düzeyler bir arada nasıl kullanılmalıdır?

İyi bir deney tasarımının istatistik bakımından belkemiği **rassallık** olmalıdır. Sözelimi iki gübre türünün verimliliği karşılaştırılacaksa, bir tarlayı ikiye bölüp yarısına bir gübreden, öbür yarısına öteki gübreden koymak iyi bir deney olamaz. Çünkü su geçirgenliği, toprak yapısı, mineral içeriği, uzun ömürlü yabancı otların varlığı ya da öngörülemeyen başka özellikler tarlanın her parçasında aynı değildir. O zaman tarlayı küçük küçük bir sürü parçaya ayırıp bunlardan hangi parçaya hangi gübrenin konacağına rassal olarak karar vermek en doğrusudur. Böylece denetim altına alamadığımız çeşitli etmenlerin, iki gübrenin kullanıldığı tarla parçalarının verimlilikleri üzerindeki etkilerini birbirlerine eşitlemiş ya da eşite bir hayli yaklaştırmış, böylelikle bu etmenlerin, verimliliği yansıtan veriler içindeki rassal paylarını azaltarak gerçekliğe yaklaşmış oluruz.

Aynı düşünceler tarım dışındaki deneyler için de geçerlidir. Bir süper market yöneticisi belli bir ürüne ayrılmış raf uzunluğunun o ürünün satışını

etkilediğini düşünüyorsa, rassal seçilmiş günlerde o ürünün dizildiği raf uzunluklarını değiştirerek satışların bundan nasıl etkilendiğini görebileceği bir deney düzenleyebilir. Bu yolla raf uzunluğu dışındaki etmenlerin satışları etkileyerek gerçekliği perdelemesini büyük ölçüde engelleyebilir.

Bir tarladan alınan ürün miktarında çeşitli gübrelerin etkileri farklı mı? Yeni geliştirilen üretim yöntemi, verimliliği gerçekten artırır mı? Fareler üzerinde denemeleri yapılacak olan yeni ilaç sağaltımında bir ilerleme sağlayacak mı? Yeni öğretim yöntemi öğrencilerin daha çabuk, daha özümseyerek öğrenmelerinde rol oynayabilir mi? Sigara kanserin nedeni midir? Belli bir iş kolundaki işçilerin kent ya da kır kökenli olmaları verimliliklerinde bir fark yaratır mı?

Bunlar gibi bir sürü sorunun yanıtı yapılacak deneylerle bulunmaya çalışılabilir. Ama deneyler öyle tasarlanmalıdır ki yalnızca etkili olduğu düşünülen değişkenin etkisini yakalayabilmeli, öbür etmenlerin işe karışmasını engelleyebilmelidir. Bunu sağlayacak olan da iyi bir deney tasarımıdır. Deney iyi tasarlanmazsa istatistik çözümlmeleri yapacak kişinin önüne içinden çıkılmaz, yararsız veri kümeleri gelebilir. Daha ileri gidilerek eğer verilerin bu özellikleri bilerek ya da bilmeyerek görmezden gelinirse varılan sonucun yanlışlığı fark edilmeden doğruymuş gibi savunulmasına yol açılabilir.

1920’de bir bahar günü İngiltere’deki Cambridge Üniversitesi’nin bahçesinde öğretim üyeleriyle eşleri sütlü çaylarını yudumlarken kadınlardan birinin, fincana önce çay mı, yoksa süt mü konduğunu çayın tadına bakarak anlayabileceğini öne sürmesi bıyık altından gülümsemelere yol açar. Ama Profesör Fisher bu savı ciddiye alıp enikonu bir deney tasarımı hazırlar. Kadına kaç fincan çay tattırılacak, bunlardan kaçına önce çay, kaçına önce süt konacak, fincanlar hangi sırayla sunulacaktır? Bu sorulara bilimsel yanıtlar vererek hazırlanan deney sonucunda kadının haklı çıkıp çıkmadığını bu öyküyü anlatan yazara bırakalım (Salzburg, 2001:1-3). Vurgulamak istediğimiz, deney tasarımı konusuna ilk büyük katkıları sağlayan R. A. Fisher’in işini ne kadar ciddiye aldığıdır.

Giriş düzeyinde bir istatistik kitabında deney tasarımı yeterince işlemek olanaksızdır. Bu yüzden bir örnek vermekle yetinelim. Bir iplik fabrikasındaki üretim sürecinde, püskürtülen sıcaklığın derecesi ile üretim bandının hareket hızı denetlenen iki bağımsız değişken olarak alınmış, bunların ürünün niteliği (bağımlı değişken) üzerindeki etkileri, öbür bütün etmenler elden geldiğince sabit tutularak incelenmiştir (Sennaroğlu, 2001). Üretim şeridi değişik hızlarda farklı püskürtme sıcaklıklarıyla çalıştırılmış, her iki (bağımsız) etmenin çeşitli bileşimlerinde üretilen ürünlerin sağlamlığına (bağımlı değişken) ilişkin bulgular aşağıda gösterilmiştir.

ÇİZELGE 6.1 İplik fabrikasında sağlamlık deneyi sonuçları

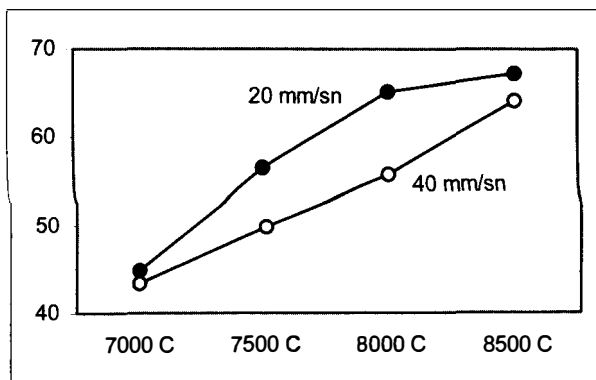
	700° C	750° C	800° C	850° C
20 mm/sn	44,9	56,5	65,1	67,2
40 mm/sn	43,4	49,9	55,8	64,1

Sennaroğlu, 2001: 23'ten uyarlanmıştır.

Bu bulguları, dik eksene sağlamlık ölçüsünü koyarak küçük bir çizimle aşağıdaki gibi gösterebiliriz.

Bu çizimde her iki hızda da sıcaklık arttıkça sağlamlığın da arttığı, ama 20 mm/sn hızda her sıcaklıkta sağlamlığın, 40 mm/sn hıza göre daha yüksek olduğu görülmektedir. Bu çizelgeden, hız-sıcaklık etmenlerinin karşılıklı etkileşim içinde olup olmadıkları da görülebilir ama henüz bu konuları öğrenmediğimiz için ayrıntıya girmeden şu kadarını söyleyelim. Çizimde iki değişik hızla ilişkin kırık çizgiler kesişmiyor. Öyleyse iki etmen etkileşim içinde değil. Eğer iki kırık çizgi kesişseydi o zaman sağlamlığın düzeyini, hız ile sıcaklığın kendi aralarındaki etkileşimle belirledikleri sonucuna varılabilirdi.

Sıcaklıklara bakılmaksızın iki hız arasında gerçekten anlamlı bir sağlamlık farkı yakalanabilir mi sorusu *t* **sınaması** kullanılarak, hız gözardı edildiğinde çeşitli sıcaklık dereceleri farklı sağlamlık düzeyleri yaratabilir mi sorusu da **varyans çözümlemesi** yöntemiyle yanıtlanabilir. Hem hız, hem sıcaklık değişirse sağlamlık düzeyleri bundan nasıl etkilenir sorusunun yanıtı da **çoklu varyans çözümlemesi** yöntemiyle verilebilir. Ayrıca bu sorulara **basit** ya da **çoklu bağlanım** (regresyon) yöntemleriyle de yaklaşılabılır.

**ÇİZİM 6.2** İplik fabrikasında deney sonuçları

Okuma Önerileri

- Barnett, V. (1991) *Sample Survey Principles and Methods*, Edward Arnold.
- Box, G. E., Hunter, W. G., Hunter, J. S. (1978) *Statistics for Experimenters*, Wiley.
- Çömlekçi, N. (1988) *Deney Tasarımı ve Çözümlemesi*, Anadolu Üniversitesi.
- Groebner, D. F., Shannon, P. W., Fry, P. C., Smith, K. D. (2001) *Business Statistics: A Decision Making Approach*, 5. b., Prentice Hall (9-25).
- Hays, W. L. (1988) *Statistics*, 4th ed. Holt, Rinehart and Winston (384-536).
- İdil, O. (1980) *Örneklem Teorisi ve İşletme Yönetiminde Uygulanması*, İÜ İşletme Fakültesi.
- Montgomery, D. (1997) *Design and Analysis of Experiments*, 4th ed. Wiley.
- Stuart, A. (1984) *The Ideas of Sampling*, Charles Griffin.
- Yamane, T. (1967) *Elementary Sampling Methods*, Prentice-Hall (Çev. Esin, A., Bakır, M. A., Aydın, C., Gürbüzsel, E.(2001) *Temel Örneklem Yöntemleri* Literatür).
- Yoğurtçugil, K. (1975) *Örneklem:Yöntemler ve Uygulama*, İÜ İktisat Fakültesi.

ALİŞTIRMALAR

- 6.1 19 Mart 2003 Milli Piyango çekilişinde 1 milyar TL kazanan 20 numara aşağıdadır. Bu numaralardan en büyüğü 688867'dir. Öbür ödülleri kazanan bütün numaralar arasında da 699264'ten büyük numara yoktur. Öyleyse satılan bilet sayısının 700.000'i aşmadığı varsayılabilir.

004469	106774	119223	148366	150374	175120	190071
201948	211641	217447	296950	325939	330615	412990
429038	434896	438313	551377	627296	688867	

- Bu numaralarda 120 rakam vardır. Yüz binler basamağında yalnızca 0-6 arasında rakamlar olduğundan bunları dışlarsak sayı 100'e iner. Numaralar rassal çekilmişse bu rakamların yaklaşık yarısının tek, yarısının çift olması gerekmez mi? O halde sayalım bakalım bu rakamların kaç tek, kaç çifttir?
- Kazanan numaraların rassal olması için bu rakamların yaklaşık yarısının 0-4, öbür yarısının da 5-9 arasında olması gerekmez mi? 0-4 arasında kaç, 5-9 arasında kaç rakam vardır?
- Bu numaraların birler basamakları için de benzer bir sayım yapınız. Burada yaklaşık 2 tane 0, 2 tane 1, ... 2 tane 9 olmalıdır. Sayım sonucunuz bunu doğruluyor mu? Bu numaralar sizce rassal çekilmiş midir?
- Benzer sayımları, onlar, yüzler, binler, on binler basamakları için de yapınız. Sizce burada rassallık geçerli mi?

- e) 0-9 arasındaki her rakamın çekilme şansı eşit olduğuna göre, yüzbinler basamağını dışarıda bıraktığımızda bu numaralarda tam olmasa bile yaklaşık 10 tane 0, 10 tane 1, ..., 10 tane 9 olması gerekmez mi? Ama sözelimi 11 tane 0 var. Ötekileri de siz sayarak her birinden kaç tane olduğunu bulunuz. Bulgunuzu nasıl yorumluyorsunuz?
- f) Numaraları ikişer basamaklı sayılara ayırınız. Her numaranın ilk iki basamağını yukarıda açıklanan nedenle dışlayınız. Geriye 40 tane iki basamaklı sayı kalır. Öyleyse 00-09 arasında yaklaşık 4, 10-19 arasında yaklaşık 4, ... 90-99 arasında yine yaklaşık 4 tane sayı olmalıdır. Benzer bir sayımı iki basamaklı sayı kümeleri için de yapınız. Bulgularınızı yorumlayınız.
- g) Bütün bu işlemleri yaptıktan sonra Milli Piyango çekilişinde 1 milyar TL kazanan numaraların rassal çekildiklerini düşünüyor musunuz? Neden?

6.2 Öğrenci sayısı 1127 olan bir bölümün öğrencileri arasından 30 birimlik bir basit rassal örneklem çekmeniz istenmiştir. Öğrencilerin 1'den 1127'ye kadar olan dökümü size verilmiştir. Bu iş için gerekli rassal sayıları çizelge, el hesap makinesi ya da bilgisayar kullanarak çekiniz. Bunların hangi öğrencilere karşılık geldiğini belirleyiniz.

6.3 Yukarıda basit rassal örnekleme açıklanırken kullanılan öğrenci seçme örneğine dönelim. Orada aşağıdaki rassal sayılar

100	850	353	396	231	695	243	516	375	168
904	055	450	054	489	084	584	665	148	321

kullanılarak şu sıra numaralarındaki öğrenciler seçilmişti.

100	186	21	64	231	31	243	184	43	168
240	55	118	54	157	84	252	1	148	321

Bu öğrencilerin bu rassal sayılara karşılık geldiğini gösteriniz.

6.4 Alistırma 6.3'teki verilerle atlamalı rassal bir örneklem çekmeniz istenmiş olsun. Hesaplamalarınızı göstererek çektiğiniz örnekleme giren öğrencilerin sıra numaralarını yazın.

6.5 Yukarıda 3.2 altbölümünde **rassal örnekleme – rassal olmayan örnekleme** başlığı altında verilen örneğe dönelim.

- a) Orada altı birimli bir anakütleden dördü örneklem çekilip her birinin örneklem ortalaması hesaplanmış, bunların en küçüğünün 6,5, en büyüğünün 10 olduğu bulunmuştu. Ayrıca "*örneğimizde dördü değil de beşli örneklem çekseydik elde edeceğimiz örneklem ortalamalarının en kü-*

çüğü 7,2, en büyüğü 9,2 olurdu” denmişti. Bu sözün doğruluğunu gösteriniz.

- b) Üçer gözlemler örnekler çekilince ortaya çıkacak örneklem ortalamalarını bulunuz. Etkinliğin azaldığını gösteriniz.

6.6 Yukarıda katmanlı rassal örnekleme başlığı altında iki katmandan eşit sayıda birimin örnekleme alındığı örneğimize dönelim. Orada katmanlar

Katman I	A, B, D	2, 6, 10
Katman II	C, E, F	8, 10, 12

biçiminde ayrılmıştı. Acaba katmanlar öyle değil de aşağıdaki gibi değişik biçimlerde ikiye ayrılırdı etkinlik yine artar mıydı? Hesaplamalarınızı yaparak her bir durumda etkinliğe ne olduğunu bulup düşüncelerinizi yazın.

- a) Katman I: 2, 10, 12 Katman II: 6, 8, 10
b) Katman I: 2, 6, 8 Katman II: 10, 10, 12
c) Katman I: 2, 8, 12 Katman II: 6, 10, 12

6.7 Yaşadığınız semtte oda sayısı, kalorifer, asansör, bahçe, vb. bakımından ortak özellikler taşıyan evler arasından bir örnekleme yaparak kira bedellerini öğrenmek istediğinizi düşünelim. Örneklemeyi nasıl yapacağınızı ayrıntılı olarak açıklayın.

6.8 Çalıştığınız ilaç deposunda yıl sonu sayımı yapılacaktır. Çeşitli fiyatlarda bir sürü değişik ilaç depo raflarında istiflenmiştir. Bunların hepsini tek tek saymanın, fiyatlarını okuyup toplam değerini hesaplamanın çok uzun, pahalı, yorucu bir iş olduğunu düşünen yöneticiniz rassal bir örnekleme toplam ilaç tutarını kestirmenizi istemektedir. Nasıl bir yol izleyeceğinizi açıkça yazın.

6.9 Çalıştığınız işletmenin 1673 müşterisinin adlarıyla borçları 56 sayfalık uzun bir dökümlerle gösterilmiştir. Siz bunlar arasından rassal bir örneklem yaparak 100 müşteriyi seçip borçlarını zamanında ödemeyenlerin oranıyla bunların gecikme sürelerini bulacaksınız. Döküm müşteri adlarının ilk harflerine göre dir. Çok acil bir rapor hazırlamanız gerektiğine göre örnekleminizi nasıl seçersiniz? Rassal sayı seçerek hangi firmaların örnekleme girdiğini belirtiniz.

6.10 Yukarıdaki iplik fabrikası örneğinde sağlamlığın yanı sıra ipliğin gevşeklik-sıkılık düzeyi de ölçülmüştür. Buna ilişkin bulgular aşağıdadır. Bu bulguları küçük bir çizimle gösteriniz. Dört ayrı püskürtme düzeyinde, iki ayrı şerit hızında gevşeklik düzeyini gösteren kırık çizgiler kesişiyor mu? Kesişiyorsa bunun anlamı nedir? Bütün sıcaklıklarda belli bir hızda üretilen ürünlerin gevşeklik özelliği, öbür hızdakinden hep daha yüksek oluyor mu? Yoksa bir hız belli sıcaklıklarda, öteki hız öbür sıcaklıklarda daha düşük gevşeklikte ürünler mi üretiyor? Başka bir deyişle şerit hızı ile püskürtme sıcaklığı arasında bulunan bir etkileşim ipliğin gevşeklik-sıkılık niteliğini etkiliyor mu?

Gevşeklik	700° C	750° C	800° C	850° C
20 mm/sn	1,60	1,21	1,13	1,00
40 mm/sn	2,58	1,55	0,84	0,37

Sennaroğlu, 2001: 23'ten uyarlanmıştır.

6.11 Bir para 20 kez atılmış, 10 kez yazı (Y), on kez de tura (T) gelmiştir.

- a) Bu paranın hileli olmasından kuşkulandır mısınız? Neden?
b) Yazı-tura geliş sırası aşağıdaki gibi olsaydı bu paranın hileli olmasından kuşkulandır mıydınız? Neden?

Y Y Y Y Y Y Y Y Y Y T T T T T T T T T T T T

- c) Yazı-tura geliş sırası aşağıdaki gibi olsaydı yine kuşkulandır mıydınız? Neden?

Y T Y T Y T Y T Y T Y T Y T Y T Y T Y T

- d) Sıralanma aşağıdaki gibi olsaydı yine kuşkulandır mıydınız? Neden?

Y Y T Y T Y Y T T T Y T Y T T Y T T Y Y

6.12 Bir kalite denetim birimi fabrikada üretilen ipliklerden 6 cm uzunluğunda parçaları gererek koparmakta, kopma noktasının belli bir uçtan uzaklığını ölçmektedir. İplik uzunluğu boyunca her noktada kopma şansının aynı olması beklenmektedir. Rassal seçilmiş 20 iplik parçasıyla yapılan bir denemede kopma noktalarının sol uçtan uzaklığı mm cinsinden aşağıdaki gibi bulunmuştur. Sizce yukarıdaki beklenti gerçekleşmiş midir?

18	6	8	34	39	12	25	31	44	19
11	34	14	18	22	24	59	52	49	29

Bu Çizelge, Bu Çizim Ne Anlatıyor?

Düşünmedeki açıklık, yetkinlik, verilerin sunumundaki açıklıkla yetkinliğe çok benzer.

E. Tufte

7.1 VERİLERİN SUNULMASI

İstatistikler ilk derlendiklerinde ham veri niteliğindedir. Özellikle gözlem sayısının ya da değişken sayısının büyük olduğu durumlarda ham verilere bakan bir kimse ağaçlara bakmaktan ormanı göremez. Bir başka deyişle ham veriler, inceleyen kimseye pek bir şey söylemez. Nüfus sayımında derlenen milyonlarca verinin herhangi bir düzenlemeye gidilmeden incelenmeye, bunlardan anlam çıkarılmaya çalışıldığını düşünün. Böyle bir işin altından kalkmak olanaksızdır. Öyleyse verilerin ne söylediklerini anlayabilmek için düzenlenmeleri, öbeklendirilmeleri, özetlenmeleri, çizelgelere dökülmeleri, çizimlere aktarılmaları gerekir.

Yalnız ham veriler değil, basılı kaynaklardan alınmış, düzenlenmiş, çizelgelere dönüştürülmüş veriler bile, bazen yeniden özetlenmeyi, yeniden çizilmeyi gerektirirler. Veri kümelerinin yapılarını doğru anlayabilmek, başkalarına doğru aktarabilmek için, bu işleri yaparken de bazı kurallara uymak yerinde olur. Ham verilerle yapılabilecek ilk iş bunlara bir öbekleme uygulamak olabilir. Diyelim 2000 genel nüfus sayımında derlenen cinsiyetle, yaşla, medeni durumla, eğitim düzeyiyle ilgili milyonlarca veri uygun öbeklere ayrılabilir. Cinsiyet için kadın-erkek; yaş için 0,1,2, ...; medeni durum için hiç evlenmemiş, evli, eşi öldü, boşandı; eğitim düzeyi için okuryazar olmayanlar, hiç okul bitirmemiş okuryazarlar, ilkokul bitirmişler, ... gibi öbekler kullanılabilir.

Bu öbeikleme sonucunda her bir değişken için ayrı bir çizelge oluşturularak, değişkenin çeşitli değerlerine giren gözlem sayıları gösterilebilir. Böylelikle de “gözlem birimi olan kişilerin kaç erkek, kaç kadın öbeğine girer; kaç 0, kaç 1, kaç 2, ... yaş öbeklerine düşer; kaç hiç evlenmemiş, kaç evli, ... medeni durum öbeklerine ayrılır; kaç okuryazar değil, kaç hiç

okul bitirmemiş okuryazar, ... eğitim düzeyi öbeklerinde yer alır?” sorularının yanıtları bulunabilir. Bölüm 3’te Çizelge 3.1 söz konusu gözlemlerin medeni duruma göre yapılmış bir öbeklendirilmesinin sonuçlarını yansıtmaktadır. Çizelge 3.2 ile 3.4 de benzer öbeklemeleri göstermektedirler.

Bu tür bir çizelge, eğer ad ölçeğiyle ölçülmüş kümeler içeriyorsa, verilerin hangi kümede yoğunlaştıklarını gösterebilmelidir. Bir başka deyişle ensıkdeğerin hangi kümede olduğunu yansıtabilmelidir. Çizelge 3.1’de bu kümenin “evli” satırı olduğu anlaşılmaktadır.

Eğer veriler sıra ölçeğindeyse böyle bir çizelge, en kalabalık kümeye ek olarak, ortalama düzey konusunda da kabaca bir fikir vermelidir. Yani ortancanın hangi kümede olduğunu, ipucu biçiminde de olsa gösterebilmelidir. Çizelge 3.2’de ortancanın “beklediğim düzeyde” kümesi içinde olduğu görülebilir. Çünkü bu çizelgedeki 5.826 öğrenci “beklentilerine göre İTÜ’yü beğenme düzeyleri” bakımından sıraya konulsa ortaya denk gelen, yani 2913-2914. sıralarda yer alan öğrenciler sorulan soruya “beklediğim düzeyde” yanıtını vermişlerdir.

Eğer çizelge aralık ya da oran ölçeğindeki veriler içerirse, yukarıdakilerden başka, yansıttıkları veri öbeğinin ortalama düzeyi, hangi değerlerde yoğunlaştığı, ortalamadan uzaklaşma derecesi, çarpık olup olmadığı, aşırı değer içerip içermediği gibi konularda kaba da olsa bazı bilgiler verebilecek biçimde düzenlenmelidir. Çizelge 3.3 bu konuda çok başarılı sayılmaz. Çünkü belli bir bölgenin illeri ortalama hava sıcaklıklarına göre bir sıraya dizilmemiştir. Bu da ortancanın görülmesini zorlaştırmaktadır. Ama Çizelge 3.4’te durum farklıdır. Burada hem ensıkdeğerin, hem ortancanın 4 kişilik haneler olduğu kolayca saptanabilir. Hane büyüklüğünün 4-5 kişilik düzeylerde yoğunlaştığı, toplam verilerin yaklaşık yarısının bu iki düzeyde toplandığı hemen görülebilir. Hanedeki kişi sayısı değişkeninin gittikçe azalan sıklık sayılarıyla büyük değerlere doğru daha çok uzadığı da gözden kaçmamaktadır. Bu da bize verilerin ortalama etrafındaki dağılımının çarpıklığına ilişkin önemli bir ipucu verir. Bu arada 9’dan çok kişi içeren hanelerin sayısı ayrı ayrı gösterilmediği için aşırı değer olup olmadığını anlamadığımızı da söyleyelim.

Burada çok kaba biçimde ölçtüğümüz bu özellikleri ileride daha ayrıntılı olarak ele alacağız.

Şimdi birden çok değişkenli çizelgelere geçelim. Aynı gözlem birimi birden çok özelliğiyle, birden çok değişkenle ele alınabilir. Nüfus sayımına konu olan bir yurttaş yukardaki pek çok özelliği bakımından incelenmiştir. Öyleyse aynı anda iki ya da daha çok özelliği bakımından öbeklere ayrılabilir. Bir başka deyişle veriler aynı anda birden çok değişkenin değerlerine göre öbeklenebilir. Yaşa-cinsiyete göre; cinsiyete-medeni duruma göre; yaşa-medeni duruma-cinsiyete göre öbeklemeler buna örnek olarak gösteri-

lebilir. Bu örnekleri yansıtan çizelgeleri, DİE tarafından yayımlanan herhangi bir Türkiye İstatistik Yıllığı'nda görebilirsiniz.

Sözü edilen değişkenlerden yaş dışındakilerin alabileceği değişik değerlerin sayısı sınırlıdır. Cinsiyet iki, medeni durum dört, eğitim düzeyi de istenen ayrıntıya göre 5-10 öbekte özetlenebilir. Yaş değişkeninin öbek sayısı ise yüzü aşabilir. Bunların hepsine birden bakıp bir şeyler anlamaya çalışmak hâlâ zor olabilir. O zaman bu yaş öbekleri, diyelim 0-4, 5-9, 10-14, ... gibi, daha geniş başka öbeklerde toplulaştırılabilir. Böylece bazı bilgi ayrıntıları yitirilebilir ama karşılığında veri yapısının kolay kavranması gibi bir yarar sağlanabilir.

Bu tür iki değişkenli çizelgelerde, değişkenlerin birlikte değişip değişmedikleri konusunda ipucu elde edilebilir. Sözelimi cinsiyet-eğitim düzeyi çizelgesinde kadınlar, erkeklere göre genellikle daha alt öğretim düzeylerinde toplanıyorlarsa iki değişken arasında bir ilişki olduğu, bir başka deyişle bunların birbirinden bağımsız olmadıkları düşünülebilir. Bir örnek olmak üzere 1990'da 18 yaşından büyüklerin cinsiyetleri ile eğitim durumlarını gösteren Çizelge 7.1'i ele alalım.

Satır toplamlarından kadınların payının % 44, erkeklerinkinin % 56 olduğu hesaplanabilir. Her eğitim düzeyinde bu oranlar geçerli olsaydı, eğitim bakımından kadınlarla erkekler arasında fark olmadığını söyleyebilirdik. Oysa lise mezunları arasında kadınların oranı % 37'nin, yüksek öğrenim görmüşlerde % 29'un altındadır. Demek ki bu çizelgeden, üst öğretim düzeylerine ulaşmada kadınların erkekler kadar şanslı olmadıkları görülebilir.

Ailelerin gelirleriyle tüketim harcamalarını gösteren bir çizelgeyi başka bir örnek olmak üzere ele alalım. Varsayımsal bir topluluktan derlenen veriler Çizelge 7.2'deki gibi olsun.

ÇİZELGE 7.1 Cinsiyet ile eğitim düzeyi ilişkisi (1990 sayımı) (milyon kişi)

	Bitirilen okul					Toplam
	Yok	İlkokul	Ortaokul	Lise	Yüksek	
Kadın	3,8	10,5	1,3	1,4	0,4	17,5
Erkek	4,1	12,2	2,4	2,4	1,0	22,1
Toplam	7,9	22,7	3,7	3,8	1,4	39,6

DİE, 2002a:64-65. Yuvarlatma nedeniyle kimi toplamlar tutmamaktadır.

ÇİZELGE 7.2 Gelir ile tüketim harcamaları arasındaki ilişki (aile sayısı)

Gelir düzeyi	Tüketim harcamaları düzeyi			Toplam
	Düşük	Orta	Yüksek	
Düşük	25	10	-	35
Orta	5	40	10	55
Yüksek	-	-	10	10
Toplam	30	50	20	100

Bu çizelge incelendiğinde gelirle tüketim arasında bir **ilişki** olduğu açıkça görülebilir. Düşük gelirli aileler az, yüksek gelirli aileler çok tüketim harcaması yapma eğilimindedirler. Bu da ailelerin tüketim harcamalarının gelirlerinden **bağımsız** olmadığı önerisini getirir.

Gerek Çizelge 7.1'deki, gerek 7.2'deki verilerden kaba olarak edindiğimiz bu izlenimlerin doğruluğu, sözgelimi **ki-kare** (χ^2) gibi yöntemlerle sınanabilir, iki değişken arasında bir ilişki olup olmadığına belli hata paylarıyla karar verilebilir.

Bazı çizelgeler de zaman serisi verileri içerebilir. Böyle çizelgelere bakılarak ilgili değişken(ler)in zaman içinde nasıl bir yol izlediği görülebilir. Sözgelimi Türkiye dışalımıyla dışsatımının 1980-2001 arasındaki gelişmesini gösteren Çizelge 7.3'ü ele alalım.

ÇİZELGE 7.3 Türkiye'nin dışalımıyla dışsatımı 1980-2001(milyar \$)

	Dışalım	Dışsatım	Dışalım	Dışsatım
1980	7,9	2,9	1991	21,0
1981	8,9	4,7	1992	22,9
1982	8,9	5,7	1993	29,4
1983	8,2	5,7	1994	23,3
1984	10,8	7,1	1995	35,7
1985	11,3	8,0	1996	43,6
1986	11,1	7,5	1997	48,6
1987	14,2	10,2	1998	45,9
1988	14,3	11,7	1999	40,7
1989	15,8	11,6	2000	54,5
1990	22,3	13,0	2001	41,4

Bu çizelgeden hem dışalımın, hem dışsatımın zaman içinde arttığı kolayca görülür. Bu dönemde dışalım 5 katına, dışsatım da 10 katına çıkmıştır. Önceleri oldukça düzgün olan bu gelişmenin 1990'ların ortalarından bu yana inişli çıkışlı bir yola girdiği anlaşılmaktadır. **Zaman serileri çözümlemesi** yöntemleriyle bu veriler incelenebilir. Yıllık ortalama artışları belli **güven aralıklarında** kestirilebilir, **genel eğilim** denklemleri bulunabilir, bunlara dayanılarak bir sonraki yılda alabilecekleri değerler belli hata paylarıyla hesaplanabilir.

Aşağıda ilk altbölümde iyi bir çizelgenin özellikleri incelenecek, ikinci altbölümdeyse çizimler ele alınacaktır.

7.2 İYİ BİR ÇİZELGE NASIL OLMALI?

Yayımlanmış çoğu çizelge, yazarınca anlaşılmamıştır. Aslında yazar, okuyucuya "İşte sana verileri veriyorum. Ben onları anlamadım, belki sen anlarsın" demektedir.

A.S.C.Ehrenberg

Çizelgeler, ham ya da işlenmiş verilerin başkalarının kullanımına sunulması için düzenlenebileceği gibi, incelenen konunun yapısını ortaya koymak, belli bir iletiyi okuyucuya ulaştırabilmek amacını da taşıyabilir. Veriler bir çizelgeye dökülürken bu amaçlardan hangisinin gözönünde tutulduğu önemlidir. İlk durumda çizelgenin fazla veri barındırması, yatay ya da dikey sıralarının çok sayıda olması gerekebilir. Ama ikinci amaç önemliyse çizelgeler olabildiğince yalınlaştırılmalı, ulaştırılması istenen ileti gereksiz verilerle boğulmamalıdır. Bu amaçlara uygun özenli bir düzenleme yapılırsa çizelgeler çok yararlı birer araç olabilirler.

Her çizelgede uyulması gereken bazı kurallar şöyle özetlenebilir (Gürtan (1987:107-109), Seyidoğlu (1995:172-182)):

- Her çizelgeye bir başlık verilmelidir. Başlık kısa, ama çizelgenin içeriği konusunda yanlışlığı yol açmayacak biçimde olmalıdır. Aynı yazıda birden çok çizelge kullanılacaksa çizelgeye bir de sıra numarası konulmalıdır.
- İncelenen konudaki değişken(ler)in aldığı değişik konumlar satır ya da sütunlarda gösterilirler. Bunlara verilecek başlıklarda da çizelge başlığında olduğu gibi davranılmalıdır. Gerekirse bu satırlarla sütunlar da numaralandırılabilir.
- Hangi değişkenin satırda, hangisinin sütunda gösterileceğine karar verilirken değişkenlerin aldıkları değerlerin sayısı yardımcı olabilir.

Az sayıda değer alan değişkene satırda, çok değer alana sütunda yer vermek genellikle uygun görülebilir.

- Tane, kg, 1995 fiyatlarıyla milyon TL gibi ölçü birimleri satır ya da sütun başlıklarında belirtilmelidir.
- Bir değişkenin çeşitli konumları nasıl sıralanmalıdır?
 - Sıralama konusunda zaman serilerinde sorun yoktur. Burada zaman sırası belirleyicidir.
 - Sıra, aralık ya da oran ölçeklerinden birinde ölçülmüş değişkenlerde, bunlar kesit verileri bile olsa sorun yoktur. Konumlar sayısal sıraya göre dizilebilirler.
 - Eğer kesit verileri ad ölçeğine göre öbeklendirilmişlerse konumların sıralandırılmasında sorun çıkabilir. Böyle bir durumda (hiç evlenmemiş, evli, dul, boşanmış gibi) akla uygun bir sıralama varsa bu uygulanabilir. Yoksa sıralamada sıklık ya da değişken değeri büyüklüğü temel alınabilir.
 - Eğer amaç veri sunmaktan öteye bir bilgi iletmekse, sıklık ya da değişken değeri büyüklüklerine göre sıralama çoğu zaman en uygundur.
- Gerektiğinde toplam, ortalama, yüzdeler çizelgeye konmalıdır.
- Verilerin alındığı kaynak belirtilmelidir.
- Gerektiğinde açıklamalara yer verilmelidir.

Aşağıda birkaç çizelge örneği verilmiştir. Bu örneklerde yukarıdaki genel kuralların varlığı araştırılacak, gerekirse yapılabilecek değişik düzenlemeler gösterilecektir.

2001-2002 yıllarında satılan araba sayılarını gösteren Çizelge 7.4'ü inceleyelim. Başlık, kaynak yerli yerinde. Sütunlar yıllara, satırlar araba markalarına ayrılmış. Buraya kadar her şey güzel. Ancak sayıların çoğunun dört basamaklı verilmesi iletinin algılanmasını güçleştiriyor. Markalar 2002 satış sayılarına göre düzenlenmiş gibi görünüyor ama Seat'ın sıralamadaki yeri yanlış. Acaba satış sayısı mı hatalı sorusu akla geliyor (Haber kaynağına göre bu sayı doğrudur). Son iki markanın 2002'de aynı sayıda satış yapması, veriler aktarılırken bir yanlışlık yapıldığı kuşkusunu uyandırıyor (Gerçekten de haberin kaynağını araştırdığımızda Nissan'ın 2002 satışının 1.097 olduğu görülmüştür). Ayrıca gazetedeki çizelgede, araba anahtarı gibi bir takım resimlerin gereksiz kalabalık yaratarak görüntüyü kirlettiğini de eklemeliyiz.

Çizelge 7.5'teki gibi bir düzenleme, söylenmek isteneni daha iyi iletebilir mi acaba? Bu çizelgede önce satış sayısını "bin" olarak değiştirip basamak sayısını azalttık. Çizelgenin başlığında bu belirtildi. Ayrıca Seat markasını büyüklük sırasında doğru yere yerleştirip Nissan'ın satış sayısını düzelttik.

ÇİZELGE 7.4 Kim, kaç adet sattı?

	2001	2002
Renault	15.196	16.672
Opel	12.231	10.824
Fiat	1.366	10.646
Peugeot	5.917	8.023
Volkswagen	9.671	7.812
Toyota	805	7.606
Ford	8.268	6.917
Honda	629	5.505
Hyundai	707	4.561
BMW	1.276	2.641
Mercedes	1.119	1.914
Skoda	1.712	1.845
Seat	1.507	2.504
Citroen	1.179	1.356
Nissan	1.714	1.356

Radikal gazetesi, 25 Şubat 2003.

ÇİZELGE 7.5 Kim, kaç bin tane sattı?

	2001	2002
Renault	15,2	16,6
Opel	12,2	10,8
Fiat	1,4	10,6
Peugeot	5,9	8,0
Volkswagen	9,7	7,8
Toyota	0,8	7,6
Ford	8,3	6,9
Honda	0,6	5,5
Hyundai	0,7	4,6
BMW	1,3	2,6
Seat	1,5	2,5
Mercedes	1,1	1,9
Skoda	1,7	1,8
Citroen	1,2	1,4
Nissan	1,7	1,1

Çizelge 7.4'ten uyarlanmıştır.

Hangi çizelgeyi daha kolay algılayabiliyorsunuz? Hangi çizelgeyi incerseniz bu satış sayıları aklınızda daha kolay kalır? Fiat, Toyota, Honda, Hyundai'nin bir yılda büyük ataklar yaptıklarını, Opel'in, Volkswagen'ın, Ford'un satışlarının az da olsa gerilediğini hangi çizelgeden daha kolay görebiliyorsunuz?

Şimdi markaların adlarından çok satışlarıyla ilgilendiğimizi düşünelim. On beş markayı bir kez 10 binden çok, 5-10 bin arası, 5 binden az sayıda satılan arabalar olarak üç öbekte, bir kez de biraz daha ayrıntılı olarak Çizelge 7.6 ile 7.7'deki gibi özetleyelim.

Bu çizelgeler ne işe yarar? İki yıl arasındaki gelişmeyi özetlerler. Böylece Çizelge 7.6'dan 2001'de 10 binden çok satan 2 markanın bir yıl sonra 3'e çıktığını görürüz. Benzer biçimde 5-10 bin arasında satan markaların sayısı da aynı dönemde 3'ten 5'e çıkmıştır. Buna karşılık 5 binden az satanların da 10'dan 7'ye inmesi doğaldır. Yalnız bu son öbek 2001'de toplam marka sayısının üçte ikisi gibi büyük bir paydır. Bunu biraz daha ayrıntılı görmek istersek Çizelge 7.7'deki küçük değişikliği yapar, bu öbeği 2-5 bin arası, 1-2 bin arası, binden az satan diye üçe bölebiliriz. Bu durumda binden az satan üç, 1-2 bin arasında satan üç markanın satışlarını artırarak üst öbeklere geçtiğini görürüz. Böylece küçük bir düzenlemeyle verilerden biraz daha bilgi edinebildik.

ÇİZELGE 7.6 Marka sayıları

Satış sayısı	2001	2002
10 binden çok	2	3
5-10 bin arası	3	5
5 binden az	10	7
Toplam	15	15

Çizelge 7.4'ten uyarlanmıştır.

ÇİZELGE 7.7 Marka sayıları

Satış sayısı	2001	2002
10 binden çok	2	3
5-10 bin arası	3	5
2-5 bin arası	0	3
1-2 bin arası	7	4
Binden az	3	0
Toplam	15	15

Çizelge 7.4'ten uyarlanmıştır.

Çizelge gözlerindeki sıklık sayıların zamanla az satış düzeylerinden çok satış düzeylerine doğru kaydıklarına göre, burada da “zaman” ile “araba satışları” arasında bir ilişkinin varlığını gözlemiş olduk.

Bir başka örneğe geçelim. Çizelge 7.8’de, bir iktisadi kesimin yurtiçi nihai mal talebinde bir birim artış olduğunda Türkiye ekonomisinde toplam ne kadar bir dışalım gereksinimi doğacağı beş ayrı yıl için gösterilmiştir. Bu değerlere dışalima bağımlılık katsayıları denir.

Çizelge 7.9’da bu katsayılar bir-iki basamağa indirgenmiş, gereksiz çizgileri kaldırılmıştır. Ayrıca her kesimin her yıl için ortalama düzeyleri birer ortancayla gösterilmiş, üretim kesimleri bu ortancanın aldığı değerlere göre sıraya konmuştur. Ortanca, büyüklük sıralamasında tam ortaya düşen değerdir.

Böylece ilk çizelgede pek göze çarpmayan kimi gelişmeler, kimi yapılar gizli olmaktan çıkmış, veriler kolayca karşılaştırılabilir hale gelmiştir. Dışa bağımlılığı en yüksek ya da en düşük kesimler belli olmuştur. En alt satırda yer alan yıl ortancaları da, bütün ekonominin dışalima bağımlılığının bir göstergesi gibi düşünülebilir.

Buradan 1980 sonrası dışa açık politikaların dışalima bağımlılığı az da olsa artırdığı izlenimi doğmaktadır. Kesimlerin sıralanması görece olarak fazla değişmemiştir. Bir kaç kesim çeşitli yıllarda bir-iki sıra kayma göstermekle birlikte genellikle üst sıralar da, alt sıralar da yerlerini korumuşlardır. Petrol kesimi açık farkla hemen her dönemde dışalima bağımlılığı en yüksek kesim olma özelliğini sürdürmüştür. Onun peşinden imalat sanayinin daha teknoloji yoğun kesimlerinin gelmesi ilginçtir. Ayrıca besin, dokuma, hatta tarım gibi geleneksel dışsattım kesimlerimizin son dönemlerde dışalima bağımlılıklarının arttığı da göze çarpmaktadır.

ÇİZELGE 7.8 Dışalığa bağımlılık katsayıları-1996 sabit fiyatlarıyla (%)

	1973	1979	1985	1990	1996
Tarım	2,53	2,78	5,04	3,76	7,13
Maden	10,14	4,53	8,33	5,34	4,60
Besin	3,39	3,93	6,85	10,30	10,99
Dokuma, giyim	6,67	3,66	12,84	10,67	17,54
Ağaç, mobilya	4,64	3,38	7,76	6,48	14,43
Kâğıt, basım	14,53	12,92	19,47	25,43	16,00
Kimya	19,09	18,57	19,33	26,57	26,28
Petrol	38,01	55,51	84,00	72,33	35,53
Lastik, plastik	21,99	10,98	18,65	25,58	32,87
Cam, çimento	13,94	15,66	23,28	18,71	11,64
Ana metal	17,47	14,71	22,98	32,76	24,83
Metal ürünleri	13,55	7,22	17,15	14,53	20,99
Makine	16,29	12,46	15,41	16,87	17,93
Elektrikli makine	13,60	7,05	7,35	17,72	21,88
Ulaştırma araçları	17,50	13,20	18,95	15,06	20,23
Diğer imalat	37,26	15,74	24,28	126,34	25,13
Elektrik, gaz, su	10,87	14,21	24,53	9,14	13,96
İnşaat	11,68	5,88	13,21	14,13	13,81
Ticaret	1,19	0,94	2,48	3,35	3,56
Ulaştırma, haberleşme	7,71	8,22	13,81	12,17	9,91
Bankacılık, sigortacılık	2,37	1,75	6,17	5,38	3,61
Kişisel hizmetler	2,92	3,00	8,92	16,06	9,00
Kamu hizmetleri	0,00	0,00	0,00	0,00	0,00
Konut sahipliği	0,27	0,60	0,30	0,66	2,52

Şenesen, Günlük-Şenesen, 2003:533-559.

Bütün bunları ilk çizelgeye bakıp çıkarmak nerdeyse olanaksızdır ama bir-iki küçük değişiklikle verilerin dili çözülmüş, kendilerini bize anlatmaya başlamışlardır. Elbette edinilen bu izlenimler birer ipucu niteliğindedir. Daha kesin konuşabilmek için başka istatistik yöntemleri gerekmektedir. İki boyutlu çizelgelerin incelenmesinde kullanılabilecek **ortanca işleme** gibi kimi yöntemler ileride açıklanacaktır.

Bu yöntemleri henüz bilmiyor olabiliriz ama yine de bazı basit çizelgelerle bu verileri biraz daha konuşturabiliriz. Yukarıda araba satışları örneğindeki benzer bir uygulamayı dışalığa bağımlılık verileri için de kullanırsak Çizelge 7.10'u elde ederiz.

ÇİZELGE 7.9 Dışalima bağımlılık katsayıları (1996 sabit fiyatlarıyla-%)

	1973	1979	1985	1990	1996	Ortanca
1 Petrol	38	56	84	72	36	56
2 Diğer imalat	37	16	24	126	25	25
3 Ana metal	17	15	23	33	25	23
4 Lastik, plastik	22	11	19	26	33	22
5 Kimya	19	19	19	27	26	19
6 Ulaştırma araçları	18	13	19	15	20	18
7 Makine	16	12	15	17	18	16
8 Kâğıt, basım	15	13	19	25	16	16
9 Cam, çimento	14	16	23	19	12	16
10 Metal ürünleri	14	7	17	15	21	15
11 Elektrikli makine	14	7	7	18	22	14
12 Elektrik, gaz, su	11	14	25	9	14	14
13 İnşaat	12	6	13	14	14	13
14 Dokuma, giyim	7	4	13	11	18	11
15 Ulaştırma, haberleşme	8	8	14	12	10	10
16 Kişisel hizmetler	3	3	9	16	9	9
17 Besin	3	4	7	10	11	7
18 Ağaç, mobilya	5	3	8	6	14	6
19 Maden	10	5	8	5	5	5
20 Tarım	3	3	5	4	7	4
21 Bankacılık, sigortacılık	2	2	6	5	4	4
22 Ticaret	1	1	2	3	4	2
23 Konut sahipliği	0	1	0	1	3	1
24 Kamu hizmetleri	0	0	0	0	0	0
Ortanca	11,5	7	13,5	14,5	14	13,5

Çizelge 7.8'den uyarlanmıştır.

Bu çizelgede de çeşitli bağımlılık düzeylerindeki kesim sayılarının öbekler arasındaki dağılımlarının zaman içinde nasıl değiştiğini izleyebiliriz. % 20 – 30 arasındaki kesim sayısının genel olarak arttığını, % 10'un altındakilerin sayısının daha düşük bir eğimle de olsa azaldığını söyleyebiliriz.

ÇİZELGE 7.10 Türkiye’de iktisadi kesimlerin dışalım bağımlılık düzeylerine göre dağılımı (1996 sabit fiyatlarıyla)

Bağımlılık katsayısı (%)	1973	1979	1985	1990	1996	Ortanca
30 +	2	1	1	3	2	2
20 – < 30	1	0	4	3	6	3
10 – < 20	11	9	9	10	8	9
0 – < 10	10	14	10	8	8	10
Toplam	24	24	24	24	24	

Çizelge 7.8’den uyarlanmıştır.

Burada da alttaki iki öbek, kesimlerin üçte ikisi ile yarısı arasında bir payı içermektedir. Öyleyse bu iki öbek biraz daha bölünerek Çizelge 7.11 elde edilir.

Çizelge 7.11’de dışalım bağımlılığı % 20’nin altında olan kesimler daha ayrıntılı olarak izlenebilmektedir. Buradan anlaşıldığına göre asıl hareketlilik alt kesimlerde. Gerçekten de dışalım pek bağımlı olmayan kesimler iyice azalmıştır. Bağımlılığı artan pek çok kesim üst öbeklere doğru tırmanma yolundadır. Bu gözlemi Çizelge 7.10’a bakarak da belki yapabiliydik ama bu hayli zor olurdu.

Çizelge 7.10’un önemli bir özelliği, öbek aralıklarının biri dışında hepsinin eşit olmasıdır. Sol yandaki öbek aralıklarını karşılaştırmaya gerek kalmaksızın sıklık sayıları kolayca karşılaştırılabilir. Bu üstünlük Çizelge 7.11’de yoktur. Eğer biraz dikkatsiz davranır da aralıkları farklı iki öbeğin sıklık sayılarını karşılaştırsak yanlış yorumlar yapabiliriz.

ÇİZELGE 7.11 Türkiye’de iktisadi kesimlerin dışalım bağımlılık düzeylerine göre dağılımı (1996 sabit fiyatlarıyla)

%	1973	1979	1985	1990	1996	Ortanca
30 +	2	1	1	3	2	2
20 – < 30	1	0	4	3	6	3
15 – < 20	4	3	6	6	3	4
10 – < 15	7	6	3	3	5	5
5 – < 10	2	4	7	5	3	4
0 – < 5	8	10	3	4	5	5
Toplam	24	24	24	24	24	

Çizelge 7.8’den uyarlanmıştır.

Her iki gösterim de değişik açılardan bakıldığında elverişli bulunabilirler. Çizelge düzenleyici bu gösterimlerden birini kendi amacı için daha uygun bulabilir, ya da bunlar dışında başka bir öbeklendirme yoluna gidebilir. Her iki gösterimde de verilerin değişik düzeylerdeki dağılımı görülebilmekte, hangi düzeylerde yığılma olduğu, dağılımın bakışık olup olmadığı anlaşılabilir. Buna karşılık iktisadi kesimlerin tekil kimlikleri silinmiştir. Dolayısıyla çeşitli öbeklerde kaç kesim bulunduğu görülmesine karşın bunların hangi kesimler olduğu anlaşılamamaktadır. Ayrıca veriler sayısal ayrıntılarını da yitirmektedirler. Her öbek aralığına giren kesimlerin dışalıma bağımlılık düzeylerinin ne olduğu da artık görülememektedir.

Çizelge 7.4, 7.5, 7.8, 7.9'un satırlarında ad ölçeği kullanılmıştı. Bu çizelgelerin asıl gövdesindeki gözler sürekli değişken değerlerinden oluşmaktaydı. **Sıklık dağılımı** da denebilen 7.6, 7.7, 7.10, 7.11 çizelgelerindeyse satırlar sıra ölçeğiyle ölçülmüş, gözlerin içindeki değişken değerleri yerlerini her göze düşen sıklık sayılarına bırakmıştır. Sekiz çizelgenin hepsinde sütunlar zamanı, yani sürekli bir değişkenin, uzunluğu bir yıl olan belli aralıklardaki değerlerini temsil etmektedir. Dolayısıyla bütün bu çizelgeler aslında belli değişkenlerin zaman içinde izlenmesi amacıyla düzenlenmiştir. Bir anlamda **zaman serisi** özelliği sergilemektedirler. Yine de her bir sütun ilgili yıla ilişkin **kesit verilerinden** oluşmaktadır.

Ancak karşımıza çıkacak her çizelgede eksenlerden biri zaman olmayaabilir. Herhangi bir yıla ilişkin dışalıma bağımlılık verilerini, kesimleri başka niteliklerine göre öbeklere ayırarak da gösterebiliriz. Sözelimi 1990 gözlemlerini bir yandan birincil üretim kesimleri (tarım, madencilik), hafif imalat sanayii (besin, dokuma, ağaç, kâğıt), ağır imalat sanayii, elektrik-gaz-su + inşaat, hizmetler diye öbeklere ayırırken, öte yandan Çizelge 7.11'deki bağımlılık düzeylerine göre de öbeklersek Çizelge 7.12'yi düzenleyebiliriz.

Bu çizimden hemen görüldüğü gibi 1990 yılında dışalıma bağımlılığı yüksek kesimler genellikle ağır sanayi kesimleridir. Birincil üretim kesimleriyle hizmet kesimleri bu anlamda bağımlılığı en düşük kesimlerdir. Benzeri çizelgeleri öteki dönemler için hazırlayan okuyucu bu yapının zaman içinde nasıl değiştiğini izleyebilir.

ÇİZELGE 7.12 Türkiye’de iktisadi kesimlerin hem alanlarına hem dışalıma bağımlılık düzeylerine göre dağılımları (1990)

%	Birincil üretim	Hafif sanayi	Ağır sanayi	E-G-S İnşaat	Hizmet kesimleri	Toplam
30 +			3			3
20 – < 30		1	2			3
15 – < 20			5		1	6
10 – < 15		1		1	1	3
5 – < 10	1	2		1	1	5
0 – < 5	1				3	4
Toplam	2	4	10	2	6	24

Çizelge 7.8’den uyarlanmıştır.

Son üç çizelgede dışalıma bağımlılık katsayısının % 30’un üzerinde olduğu öbeklerde üst sınır verilmemiştir. Öbek aralığını belirleyen alt, üst sınır değerlerinden biri belirsiz bırakılan bu tür öbeklere **açık uçlu öbekler** denir. Böyle açık uçlu bir öbek kullanılmasının nedeni dışalıma bağımlılığı % 125’i aşan bir kesimin varlığıdır. Eğer açık uçlu öbek kullanmazsak ya 30 – 130 arasında çok geniş bir öbek tanımlamak zorunda kalacak ya da çoğu boş (hiç gözlem içermeyen) 30 – 40, 40 – 50, ..., 120 – 130 gibi gerksiz öbekler kurgulayacaktık.

Yukarıda iki değışkene (dışalıma bağımlılık katsayısı ile çalışma alanına) göre dağılımın birlikte ele alındığı çizelgelere bir örnek görmüş buluyoruz. Uygulamada üç ya da daha çok değışkenin aynı anda gösterilmeye çalışıldığı çizelgeler de kullanılmaktadır. Ancak bir çizelgeye ne kadar çok değışkene ilişkin veri yüklenmeye çalışılırsa çizelgenin kolay anlaşılır olma özelliğinden de o denli uzaklaştığını unutmamak gerekir. Yalnızca veri sergilemek amacı güden DİE gibi kuruluşlar, bu tür çizelgeler düzenlemekte haklı olabilirler, ancak belli bir konuda açıklama yapmak isteyenlerin çizelgelerini yalın tutmalarında yarar vardır.

Bir istatistikçiye göre “... ayrıntılı çizelgeler yalnızca üzerinde çalışma yapmak amacıyla düzenlenmelidir, varılan sonuçları iletmek için böyle çizelgeler gerekmez. ... Bazı aykırı durumlar dışında ‘hiçbir raporun ayrıntılı veriler sunan çizelge içermemesi’ kuralı şu anlama gelir. Eğer bir çizelge büyük çapta veri sunuyorsa bunu hazırlayanın ya bir özet sunmak ya da daha sonra yapılabilecek çalışmalar için ayrıntılı veri sergilemek gibi özel bir amacı olmalıdır.” (Ehrenberg, 1975:38). Çizelge 7.13, bu söylenenlere uyan güzel bir örnek olarak “Yoksulluk artık yalnızca Güney ülkelerine özgü değil, Kuzey’e de sıçradı” (UNDP, 2000:34) diyen bir rapordan alınmıştır.

ÇİZELGE 7.13 Bazı OECD ülkelerinde yoksulluk

Ülke	Yoksulluk sınırı altındaki halk (%)
ABD (1997)	17
İtalya (1995)	13
Avustralya (1994)	12
Kanada (1994)	11
İngiltere (1995)	11

UNDP, 2000: 34. Harcanabilir hanehalkı geliri eşdeğer ortancasının % 50'si yoksulluk sınırı olarak alınmıştır.

Buraya kadar gördüğümüz çizelgeler hemen her gün çeşitli yayınlarda rastlanan türdendir. Sayıların arkasında yatanı anlamak amacıyla, aynı veriler kullanılarak, çözümlmeye yönelik başka türlü gösterimler de hazırlanabilir. Bunlara daha sonra değinilecektir.

7.3 İYİ BİR ÇİZİM NASIL OLMALI?

Bir çizim on bin sözcüğe bedeldir.

Çin Atasözü

Ölçülmüş ya da sayılmış büyüklükler olan verileri, noktalar, çizgiler, ek-senler, sayılar, simgeler, sözcükler, renkler, renk koyulukları kullanarak göze yönelik bir biçimde sunmak için çizimlere başvurulur. İstatistik verilerinin çizimlerle gösterilmesini bugünkü anlamıyla geniş ölçüde ilk kullanan İngiliz iktisatçı William Playfair (1759-1823) olmuştur (Gürtan, 1987:148). 1786'da çıkan *The Commercial and Political Atlas* (Ticaret ve Siyaset Atlası) adlı kitabındaki 44 çizim ile çağdaş çizim ilkelerinin temellerini atmıştır.

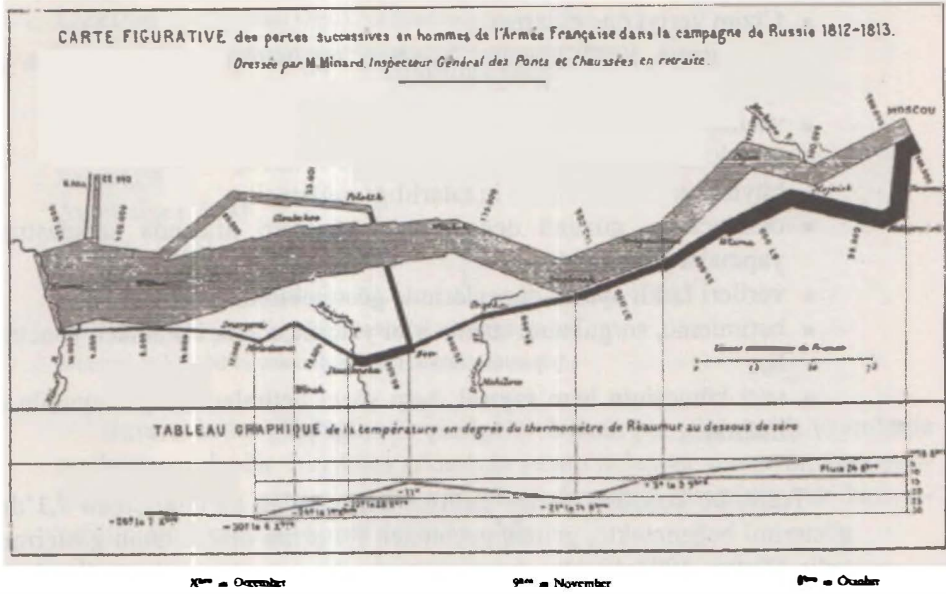
Günümüzde bilgisayar olanaklarının gelişmesi, kolayca çok çeşitli çizimler yapabilme yollarını açmıştır. Ancak daha sonra göreceğimiz gibi, bazı bilgisayar çizimleri kimi temel ilkelere uymamaktadır.

Bir çizim yapılırken en temel ilke verinin öne çıkarılmasıdır (Tuft, 1983:92; Cleveland, 1985:24) İyi bir çizim dikkati kendi üzerine çekmez, bakanı istatistik verilerinin niteliği konusunda bilgilendirir. İyi çizim bilgi demektir (Walgren vd. 1996:3). Öteki çizim öğelerinin verinin önüne geçmesi, onun söyleyeceklerini engellemesi ya da en azından iletinin gücünü zayıflatması çok rastlanılan bir durumdur. Tuft'e (1983:13) göre istatistik çizimlerinde yetkinlik, karmaşık düşüncelerin açık, doğru, etkin bir biçimde okuyucuya iletilmesidir.

- Çizim veriyi öne çıkarmalı,
- okuyucuyu, kendi yöntemi, tasarımı, uygulandığı gibi noktalardan çok işin özüne yöneltmeli,
- verilerin anlamını çarpıtmamalı,
- küçük bir alanda çok sayıda veriye yer vermeli,
- büyük veri kümelerinde iç tutarlılığı sağlamalı,
- okuyucunun gözünü değişik veri kümeleri arasında karşılaştırma yapmaya yöneltmeli,
- verileri farklı ayrıntı düzeylerinde gösterebilmeli,
- betimleme, sorgulama, sunma gibi yeterince açık bir amaca yönelmeli,
- veri kümesinin hem sayısal, hem sözel betimlenmesiyle uyumlu olmalıdır.

Tufte, bu koşulları yerine getirme bakımından en çok Çizim 7.1'deki gösterimi beğenmekte, gelmiş geçmiş en iyi çizim olarak bunu göstermektedir (Tufte, 1983:40-41). Aynı çizim başka yazarlarca da övülmektedir (Wonnacott, Wonnacott, 1990:59-60; Walgren vd. 1996:9). Napolyon'un subaylarından birinin düzenlediği bu çizimde 1812 Rusya seferinin öyküsü çok çarpıcı bir biçimde anlatılmaktadır. Çizim üzerinde ordunun büyüklüğü, haritadaki yeri, gidiş yönü, çeşitli tarihlerdeki hava sıcaklıkları konusunda bilgi verilmektedir. Yola çıkan 400.000 askerin nerelerdeki çarpışmalarda kaç kişiye indiği, ancak dörtte birinin Moskova önlerine varabildiği, geri çekilirken hangi tarihte, hangi ırmağı geçerken, eksi kaç derece soğukta, ne kadarının kırıldığı, geriye neredeyse donmuş 10.000 kişinin ancak dönebildiği, tek bir çizimle, Tolstoy'un *Savaş ve Barış*'ında bin sayfada anlattığına eşdeğer vuruculukta sunulmaktadır.

Bir çizimde nicel, nitel bilgiler simgelere, geometrik biçimlere, renklere, renk koyuluklarına dönüştürülerek deyim yerindeyse "kod"lanır. Bir çizimin iletisi algılanırken bu bilgiler göz yardımıyla kodlarından arındırılır. Çizimin okuyucu tarafından algılanması, çizimin asıl amacı olmalıdır. Bir çizim algılanamıyorsa, veriler ne denli zekice seçilmiş, ne denli ustaca kodlanmış, teknik olarak ne denli etkileyici çizilmiş olursa olsun, o çizim boşa gitmiş bir çabadır. Çizimlerdeki kodlanmış bilgileri kodlarından ayıklayıp yeniden ilk anlamlarına dönüştürürken çeşitli konularda bir takım algılama kararları veririz.



ÇİZİM 7.1 Napolyon'un 1812 Rusya seferi

Kimi deneyler, çeşitli çizimlerin, bizim algılayışımızdaki doğruluk derecesine göre şöyle sıralanabileceğini göstermiştir (Cleveland, 1985:254):

1. Ortak bir eksene göre konum,
2. Aynı ölçekte ama aynı hizada olmayan eksenlere göre konum,
3. Uzunluk, uzaklık,
4. Açı, eğim,
5. Alan
6. Oylum (hacim),
7. Renk, ton, koyuluk.

Öyleyse bir çizim düzenlenirken bunlar arasından olabildiğince ilk sıraldakiler seçilmelidir. Bir başka deyişle, zorunluluklar dışında, sayısal bilgileri, renk, ton ya da koyuluk farklarına dönüştürmek pek doğru olmaz. Aynı biçimde oylum, alan, açı ya da eğim farkları da algılamayı zorlaştıran kodlama biçimleri olarak düşünülmelidir. Olanaklar elverdiği sürece çizimlerimizde bilgileri, bir eksene göre konuma ya da uzaklığa dönüştürmek en iyisidir. Çeşitli çizim türleri arasında seçim yaparken bunların hangi kodlamayı kullandığına bakıp, yukarıda verilen sıralamada elden geldiğince en üstlerdeki yeğlersek, ulaştırmak istediğimiz iletinin okuyucu tarafından daha kolay algılanmasını sağlayabiliriz.

Başarılı bir çizim yaratabilmek için çizimin içeriği, biçimi, bağlamı konusunda iyice düşünülmelidir. İletinizi ulaştırmayı amaçladığınız kimseler için önemli olan neyse, çizimin içeriği onların gereklerine göre düzenlenmelidir. Bu noktaya daha verilerin işlenmesi aşamasında dikkat edilmelidir. Çizime hangi değişkenlerin konacağı, bunun için hangi hesaplamaların yapılması gerektiği en başta planlanmalıdır. Büyük bir çizelgenin içeriği çizimde yoğun ama açıkça anlaşılır bir biçimde verilmelidir. Okuyucu çizim üzerinde karşılaştırma yapabilmeli, önemli gelişmeleri, veride gizli yapısal özellikleri görebilmelidir. Konunun çeşitli yönleri değişik çizimler gerektirebilir.

Çizim konuyu hiç bilmeyenlere yönelikse basit kavramlarla genel bilgi vermeli, konuyu iyi bilenlere yönelikse daha çok ayrıntı içermeli, daha çok karşılaştırmalara yönelik olmalıdır.

Bir çizimin “iyi” olup olmadığı şu sorulara verilecek yanıtlarla anlaşılabilir (Walgren vd. 1996:66).

- Çizimde veri baskın mı? Asıl çizim alanının çizim içindeki payı, başlığın, kenar açıklamalarının, eksen adlarının, eksen ölçeklerinin toplamından büyük olmalı, yatay ya da dikey ölçek çizgileriyle boğulmamalıdır. Gereksiz mürekkep harcanmamalıdır. Tufte’nin (1983:93) bu konuda “verilere harcanan mürekkep / çizimin bütününe harcanan mürekkep” ya da “1 – bilgi yitirmeden silinebilecek mürekkep oranı” diye tanımladığı **veri mürekkebi oranı** 1’e yakın olmalıdır.
- Çizimdeki birim alan başına veri değeri sayısı anlamına gelen **veri yoğunluğu** düşük mü, yüksek mi? Çizim alanınız veri sayısına göre ayarlanmalıdır. Göstereceğiniz veri sayısı azsa çiziminiz küçük, çoksa çiziminiz de büyük olmalıdır. Zaman serileri, haritalar gibi çizimlerde veri yoğunluğu daha yüksek olur.
- Çizim verilerin açık bir resmini sergiliyor mu? Bir çizimde çok sayıda çubuk varsa başka tür bir çizim uygun olabilir, bir zaman serisi çizimi çok sert iniş çıkışlarla doluysa düzletilmesi gerekebilir, aynı çizimde gösterilen eğrilerin ya da noktaların sayısı yüksekse bunlar birden çok çizime bölünebilir.
- Çizim verilerin içerdiği gerçeği yansıtıyor mu? Bir çizim istatistik bilgiyi doğru vermelidir. Özellikle resimli çizimlerde bu noktaya özen gösterilmeli, iki değişik değer için çizilen resimler farklı boyutlarda çizilmemeli, büyüklükler arasındaki oran aynı resmin yinelenmesiyle verilmelidir. Çubuk çizimlerinde eksenin ölçeği “0” değerinden başlamalıdır ki karşılaştırmalar doğru yapılabilir. Pasta çizimlerinden, özellikle de üç boyutlu olanlarından olabildiğince kaçınılmalıdır, çünkü karşılaştırma yapmakta zorluk çıkarırlar. Üç değişkene ilişkin

veriler için bile çizimlerde üç boyutluluğa başvurulmamalı, iki değişken iki eksene yerleştirilip üçüncü değişken değişik simgelerle, değişik eğrilerle gösterilmelidir.

Bir örnek üzerinde bu söylenenleri açıklayalım. Türkiye'deki çeşitli bölgelerin 2000'de yarattıkları katma değerlerin sabit fiyatlarla GSYİÜ değeri içindeki % paylarını ele alalım.

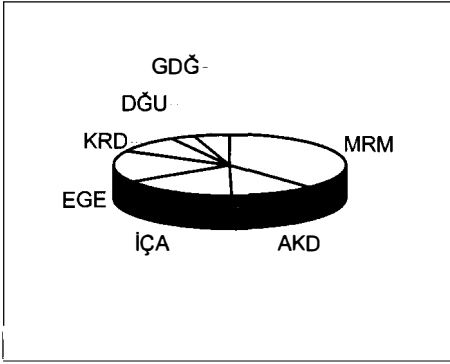
ÇİZELGE 7.14 Bölgelerimizin GSYİÜ içindeki payları-2000 (1987 fiyat. %)

Marmara	38,0	Ege	16,8	Akdeniz	11,7	İç A.	16,1
Karadeniz	9,1	Doğu A.	3,3	Güneydoğu A.	5,1		

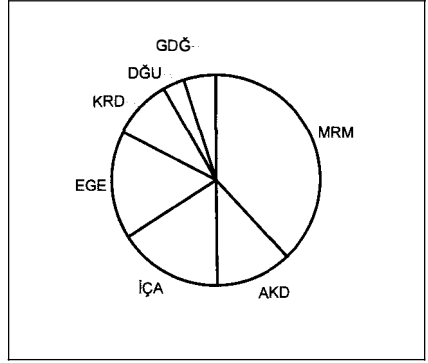
Yuvarlatma nedeniyle toplam 100,0 etmemektedir. DİE, 2002a:663.

Bu bilgileri çeşitli çizim türleriyle gösterebiliriz. Bunlardan hangileri okuyucunun kolayca algılayabildiği, hangileri algılamakta zorluk çektiği çizimlerdir? Çizim 7.2 (a), 7.2 (b) gibi çizimlere **pasta çizimi** ya da bölünmüş daire denmektedir (Gürtan, 1987:168). (a)'daki üç, (b)'deki iki boyutlu pasta çizimleridir. Bu tür çizimlerde açı, alan, oylum kodlamaları kullanıldığından, bölgelerin paylarını doğru algılayabilmek için bir takım açılar, alanları, oylumları karşılaştırmak gerekir. Bunu yaparken göz kolaylıkla yanılabilir.

Gerçekten de (a)'da gözümüze yakın olan Akdeniz ile İç Anadolu bölgelerinin merkez açıları ya da oylumları, Ege'ninkinden daha büyükmüş gibi algılanabilir. Bilgisayar yardımıyla türetilmiş bu çizimde herhangi bir ölçüm ya da çizim yanlışlığı yoktur. Bu tür üç boyutlu pasta çizimlerinin doğası gereği gözümüz yanılmaktadır. Bu çizimde yalnızca 7 bölge vardır. Ya bir de pasta çiziminin daha fazla parçaya ayrıldığı durumları düşünün. Burada açılar arasındaki büyüklük farklarını algılamak neredeyse olanaksız hale gelebilir.



(a)



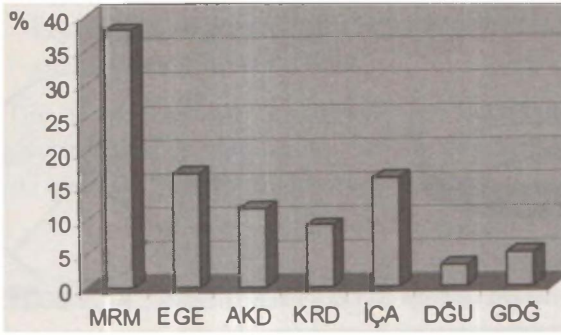
(b)

ÇİZİM 7.2 Bölgelerimizin GSYİÜ içindeki % payları-2000 (1987 fiyatları)
Çizelge 7.14'teki veriler kullanılarak düzenlenmiştir.

Çizim (b)'de durum pek az iyileşmiştir. Her ne kadar artık oylumları karşılaştırmak gereğinden kurtulmuşsak da açıları, alanları karşılaştırma zorunluluğu sürmektedir. Akdeniz'le Karadeniz, Ege'yle İç Anadolu arasındaki farklar çizimde öylesine yitirilmiştir ki bunları algılamamız çok zorlaşmıştır.

Bu iki pasta çiziminde de % paylar gösterilmemiştir ama çoğu çizimde bu sayılara hemen her zaman yer verilmektedir. Veriler ayrıca sayılarla gösterilecekse çizime ne gerek var ki? Bir çizelge bu bilgileri daha etkin sunabilir. Çünkü çizelgede yazılı sayısal değerleri okuyucu doğrudan algılar; çizimdeyse önce açıları ya da alanları algılaması, daha sonra bunları yeniden sayısal bilgiye dönüştüren zihinsel bir işleminden geçirmesi gerekir. Bu da hem algılamayı güçleştirir, hem yanlışlığa yol açar.

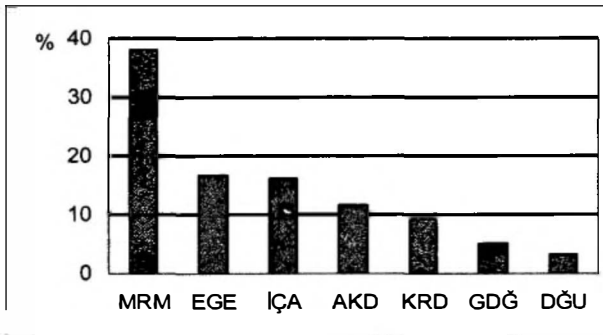
Aynı veriler Çizim 7.2 (c) ile (d)'de, bu kez **çubuk çizimleriyle** gösterilmiştir.



ÇİZİM 7.2 (c) Bölgelerimizin GSYİÜ içindeki % payları-2000 (1987 fiyatları)
Çizelge 7.14'teki verilerle düzenlenmiştir.

Burada algılama, pasta çizimlerine göre daha kolay gibi görünmektedir, ama yine de algılamanın perdelendiği söylenebilir. Yukarıda okuyucunun, çizimin yöntemi, tasarımı, uygulamasından çok, özüne yöneltilmesi gereği vurgulanmıştı. Bu çizimde bunun tam tersi yapılmıştır. Üç boyutlu çubuklar, arka-yan duvarlar, sol tarafta gereğinden sık ölçek sayıları, çizime bakanın dikkatini verilerden çok başka öğelere çekmekte, algılamanın zorlaşmasına neden olmaktadır. Özellikle sol yan duvardaki çizgiler düpedüz yanıltıcıdır. İç Anadolu'nun payını gösteren çubuğun arka hizası sanki % 15'i, Karadeniz'ininki de % 7-8 arasında bir değeri gösteriyor gibidir. Oysa verilerden bildiğimiz gibi bu değerlerden ilki % 16,0, ikincisi % 9,1'dir. Ayrıca gerek çubukların yan yüzlerindeki, üst yüzlerindeki, gerek tabandaki gölgeler de dikkatleri kendi üzerlerine çekmektedir. Oysa bu çizim öğelerinin verilerle hiç bir ilişkisi yoktur.

Çizim 7.2 (d)'de yukarıda eleştirilen bir çok özellik düzeltilmiştir.

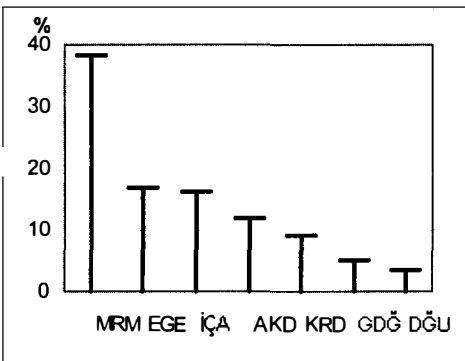


ÇİZİM 7.2 (d) Bölgelerimizin GSYİÜ içindeki % payları-2000 (1987 fiyatlarıyla)
Çizelge 7.14'teki verilerle düzenlenmiştir.

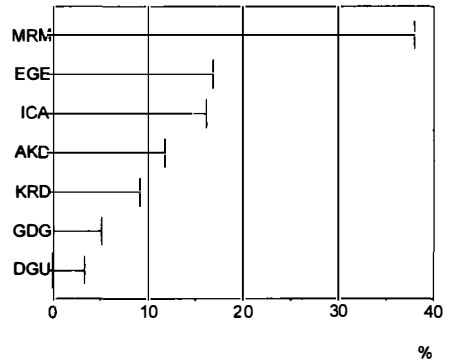
Üç boyutluluk kaldırılmış, böylece yan duvardan kurtulunmuştur. Dik eksendeki ölçek sayıları seyreltilip daha kolay okunmaları sağlanmıştır. En önemlisi, bölgeler, GSYİÜ içindeki paylarına göre sıraya dizilmiş, böylelikle her bölgenin büyüklük sıralamasındaki yeri kolayca algılanabilir duruma gelmiştir. Ama ölçek sayıları hizasındaki yatay ızgara çizgileri gereğinden daha koyu çizilmiştir. Bunların göz alıcı olmaları değil, arka planda kalmaları, yalnızca karşılaştırmalara yardımcı olmaları beklenir.

Bu düzeltmeleri de yaparak Çizim 7.2 (e) elde edilebilir. Bu çizimde öncekilere oranla çok daha az mürekkep harcanmıştır ama verilerin içerdiği bilgi hiç eksiltilmeden çizime yansıtılmıştır. Bu anlamda veri mürekkebi oranı 1'e yakındır. Her bölgenin payı, ortak bir eksenle uzaklığa dönüştürülmüş, o uzunluğun ucuna da konum belirten bir çizgi konmuştur. Bu çizimde hemen hemen hiç bir çizim öğesi gereksiz değildir. Gerçi Tufte bu çizimi görseydi çerçeveden pek hoşlanmazdı ama çizimler konusunda kuramsal çalışmalar yapan başka bazı yazarlar da (Cleveland, 1985; Chambers, vö,1983) "veri alanı" olarak tanımladıkları bölgenin, çizimin öteki öğelerinden ayrılmasını önermektedirler. Bu çizimde bu ikinci öneriye uyulmuştur.

Çizim 7.2 (f) ise aynı çizimin eksenleri yer değiştirmiş biçimidir. Göz, büyüklük karşılaştırmalarını yana doğru değil de yukarıdan aşağıya doğru daha rahat yapabilir. Bu çizimde bu sağlanmaya çalışılmıştır. Aynı zamanda okuyucuya bir karşılaştırma olanağı sağlamak için Tufte'nin hoşlanacağı biçimde davranılıp dış çerçeve çizilmemiştir.



(e)



(f)

ÇİZİM 7.2 Bölgelerimizin GSYİÜ içindeki % payları-2000 (1987 fiyatları)

Çizelge 7.14'teki verilerle düzenlenmiştir.

Bu çizimlerde verinin baskın olduğundan söz edilebilir mi? Bütün çizimlerde başlık çizim alanını çevreleyen dikdörtgenin altındadır. Çevrili alanın içindeyse verilerden başka hiçbir şey yoktur. Öyleyse bunlar verinin baskın olduğu çizimlerdir diyebiliriz.

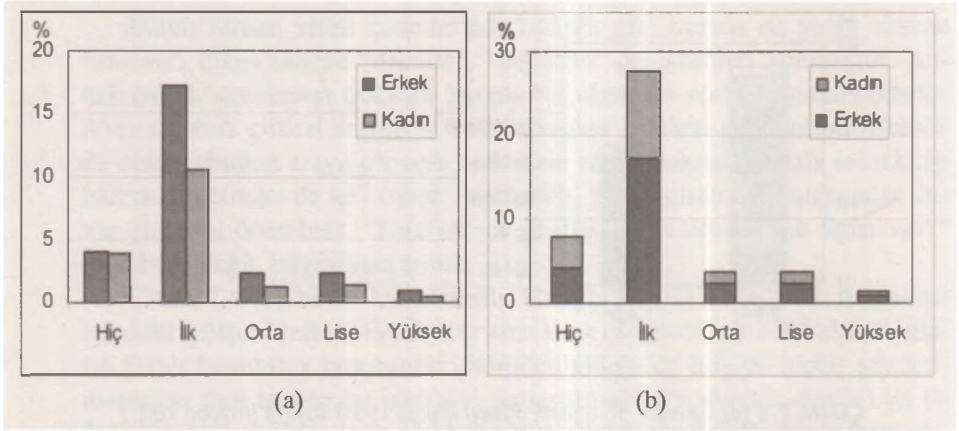
Bütün bu çizimlerde gösterilen veri sayısı, 7 bölgenin adı ile GSYİÜ içindeki payı olmak üzere 14'tür: Yaklaşık olarak Çizim 7.2 (a) 28, (b) 26, (c) 50, (d) 43, (e) 32, (f) ise 20 cm^2 'dir. Demek ki veri yoğunluğu (a)'da $14/28 = 0,5$, (b)'de $14/26 = 0,54$, (c)'de $14/50 = 0,28$, (d)'de $14/43 = 0,23$, (e)'de $14/32 = 0,44$ (f)'de ise $14/20 = 0,7$ gibi düşük sayılardır. Bir araştırmaya göre (Tuft, 1983:167) bu yoğunluk bilimsel dergilerde 2-6, gündelik yayınlarda 0,03-1,5 bulunmuştur.

Yukarıdaki çizimlerde iki değişken kullandık. Bunlardan biri ad ölçekli "bölge" değişkeni, öteki oran ölçekli "GSYİÜ içindeki pay" değişkeniydi. Çubuk çizimleriyle pasta çizimleri bu değişkenleri göstermeye uygundur. Bu tür çizimlerde iki değişken varsa bunlardan biri ad ya da sıra ölçeğinde, yani nitel bir ölçekte, öbürü de aralık ya da oran ölçeğinde, yani nicel bir ölçekte ölçülmüş olmalıdır.

Çubuk çizimleri tek değişken için düzenlenirse o değişken nitel olmalı, çubuk boylarını ölçen öbür eksen de sıklık, göreceli sıklık, birikimli sıklık ya da göreceli birikimli sıklıklar yer almalıdır. Pasta çizimlerinde tek değişken kullanılırsa da durum benzerdir. O değişken nitel ölçekli olmalı, pasta içindeki paylar da sıklıklar ya da paylarla ilgili büyüklükleri göstermelidir. Ayrıca bu sıklıkların toplamı bir bütünü oluşturmaldır. Sözgelimi Çizelge 7.13'teki yükseklik verileri de birer paydır ama toplamaları bir bütünü göstermediği, yani 100 etmediği için pasta çizimiyle gösterilmeleri anlamsız olur.

Çubuk çizimleri ikiden çok değişken için de çizilebilir. O zaman nitel ölçekli değişken bir eksene konur, öbür eksen de öteki değişkenlere ayrılır. Her bir değişken için çubuklar farklı renkte ya da koyulukta boyanır. Bazen de aynı çubuk üzerinde değişik renkte ya da koyulukta parçalar oluşturulur. Pasta çizimleri zorunlu olarak en çok iki değişkenlidir. Çizime daha çok değişken katılmak istenirse ayrı bir pasta çizimi gerekir. Bütün bu durumlara birer örnek verelim.

1990 nüfus sayımına göre Türkiye'de cinsiyet ile eğitim düzeyi ilişkilerini açıklayan Çizelge 7.1'deki veriler Çizim 7.3'te üç değişik biçimde gösterilmiştir. Burada üç değişken vardır. "Cinsiyet" ad, "eğitim düzeyi" sıra, "kişi sayısı" oran ölçeğindedir. Sonuncuya sıklık da diyebilirsiniz.



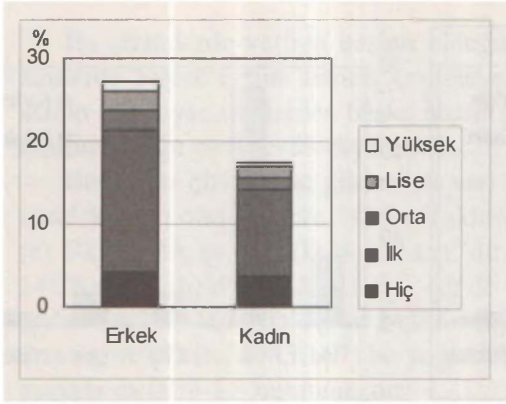
ÇİZİM 7.3 Cinsiyet ile eğitim düzeyi ilişkisi – 1990 sayımı (milyon kişi)
Çizelge 7.1'deki verilerle çizilmiştir

(a)'da her çubuk çifti eğitim düzeyini, farklı koyulukta boyanmış her iki çubuk cinsiyeti, çubuk boyları da sıklıkları gösterir. Bu çizimde belli bir eğitim düzeyindeki kadın-erkek karşılaştırması kolayca yapılabilir. Çeşitli eğitim düzeylerindeki toplam kişi sayıları karşılaştırılabilirse de bu biraz daha zordur.

(b)'de her eğitim düzeyindeki kadın-erkek sayıları aynı çubuk üzerinde farklı renklerle gösterilmiştir. Burada her eğitim düzeyindeki toplam kişi sayılarını birbirleriyle, erkeklerin sayılarını da birbirleriyle kolayca karşılaştırabilmemize karşın kadın sayıları için aynı şeyi yapmak o kadar kolay olmaz. Kolay olmayan bir başka karşılaştırma da aynı eğitim düzeyindeki kadın-erkek sayılarıdır.

(c)'de yatay eksenindeki değişken artık eğitim düzeyi değil, cinsiyettir. Eğitim düzeyleri farklı koyulukta boyanarak çubuklarda üst üste dizilmiştir. Bu çizimde kolay olan karşılaştırmalar kadın-erkek toplam sayılarıyla hiç okul bitirmemiş kadın-erkek sayılarıdır. Bunlardan ilkinin yapmak için böyle karışık bir çizime zaten gerek yoktur. Öteki eğitim düzeylerinde kadın-erkek sayılarını karşılaştırmak hayli zordur.

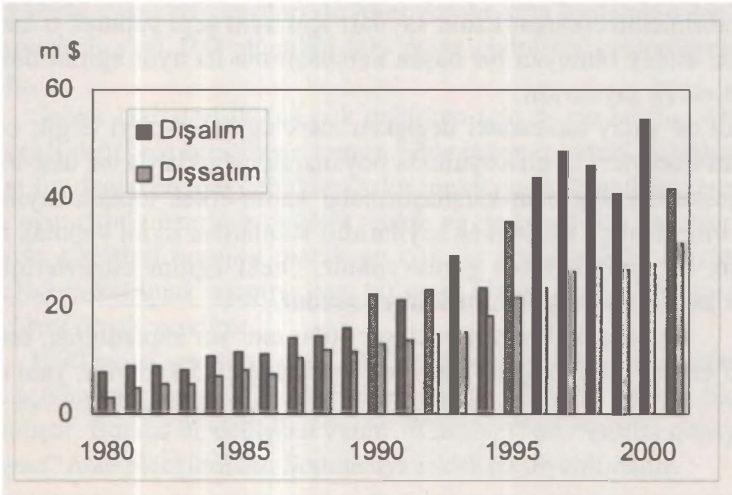
Üç çizimin her biri yaklaşık 30'ar cm² yer kapladığına, her çizimde de 15'er veri bulunduğuna göre veri yoğunluğu 0,5 kadardır, yani düşüktür.



ÇİZİM 7.3 (c) Cinsiyet ile eğitim düzeyi ilişkisi 1990 sayımı (milyon kişi)
Çizelge 7.1'deki verilerle çizilmiştir.

Bu verilerle pasta çizimi yapmaya kalksak kadınlar için ayrı, erkekler için ayrı birer çizime gerek duyardık. Bunları yapıp yorumlamak okuyucuya bırakılmıştır.

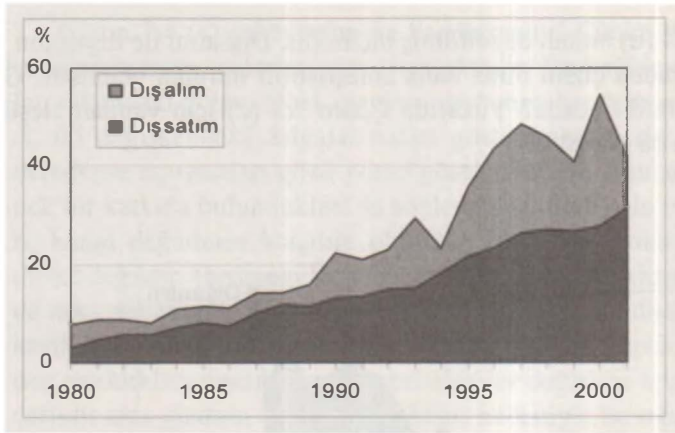
Çizelge 7.3'te dışalım-dışsatım verileri iki zaman serisi olarak yer almaktaydı. Bu verileri tek bir çizimde nasıl gösterebiliriz? Bunlar herhangi bir bütünün parçaları olmadıkları için pasta çizimi söz konusu bile değildir. Acaba çubuk çizimi bu verilere uygun mudur? Çizim 7.4 (a) bize bir fikir verebilir.



ÇİZİM 7.4 (a) Türkiye'nin dışalımıyla dışsatımı 1980-2001 (milyar \$)
Çizelge 7.3'teki verilerle düzenlenmiştir.

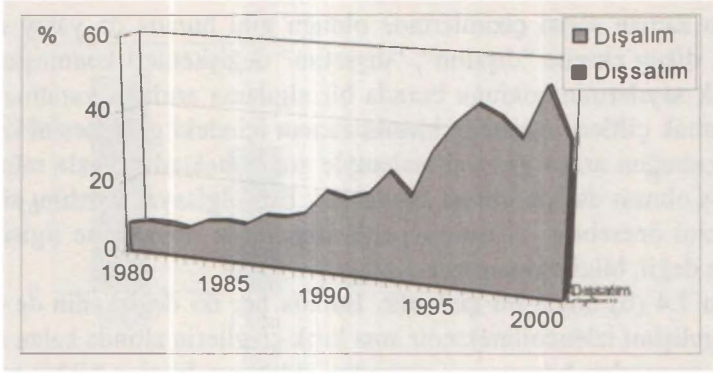
Bütün zaman serisi çizimlerinde olduğu gibi burada da yatay eksene “zaman”, dikey eksene “dışalım”, “dışsatım” değişkenleri konmuştur. Ancak çubuk sayılarının çokluğu burada bir algılama zorluğu yaratmaktadır. Ayrıca çubuk çiftleri arasından birinin zaman içindeki gelişmesini izlemek de öbür çubuğun araya girmesi nedeniyle zor olmaktadır. Fazla mürekkep harcanmış olması da işin cabası sayılabilir. Bir bilgisayar yazılımı size duvar çizimini önerebilir. “Dışalım ya da dışsatımın duvarla ne ilgisi var?” diye bana değil, bilgisayara sorun.

Çizim 7.4 (b) böyle bir çizimdir. Burada her iki değişkenin de zaman içindeki gelişimi izlenebilmektedir ama kırık çizgilerin altında kalan alanların farklı koyulukta boyanması çizimden aldığımız bilgiye hiçbir şey katmamakta, tam tersine bu izlemeyi zorlaştırmaktadır. Ayrıca dışalım ya da dışsatım gibi değişkenlerin alanlarla gösterilmesinin ne gibi bir anlamı olabilir ki? Ya mürekkebin bol bol akıtılmasına ne dersiniz?

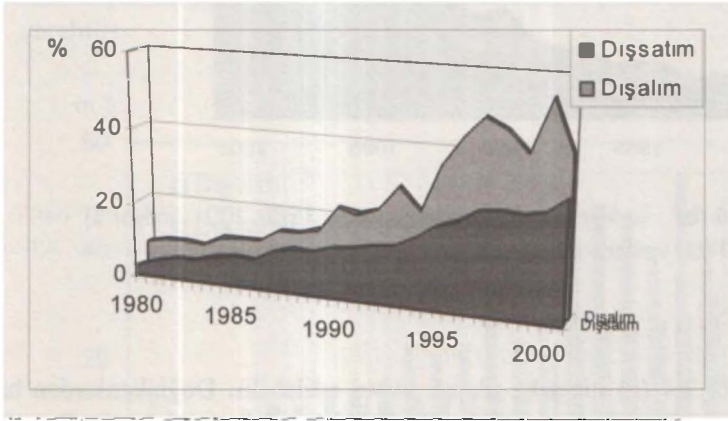


ÇİZİM 7.4 (b) Türkiye'nin dışalımıyla dışsatımı 1980-2001 (milyar \$)
Çizelge 7.3'teki verilerle düzenlenmiştir.

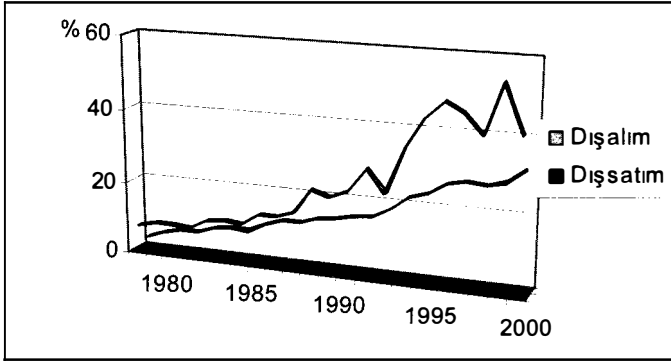
Çizim 7.4 (c) bir şaka olarak kabul edilebilir. Değişkenlerden birinin, ötekini ördeği duvarın arkasında kalması görünmesini engellemektedir. Bu da dışalima ilişkin herhangi bir bilgi elde edilmesini olanaksız kılmaktadır. Böyle bir çizim sunmayı kimse aklından bile geçirmez umarım. Ama biz gene de bir uyaralım.



Çizim 7.4 (d) bunun düzeltilmiş biçimidir. Dış satım ile dış alımın yerleri değiştirildiğinden çizim biraz daha anlaşılabilir duruma gelmiştir. Gerçekten de öyle midir acaba? Yukarıda Çizim 7.2 (c) için yapılan eleştirilerin hepsi burada da geçerlidir.



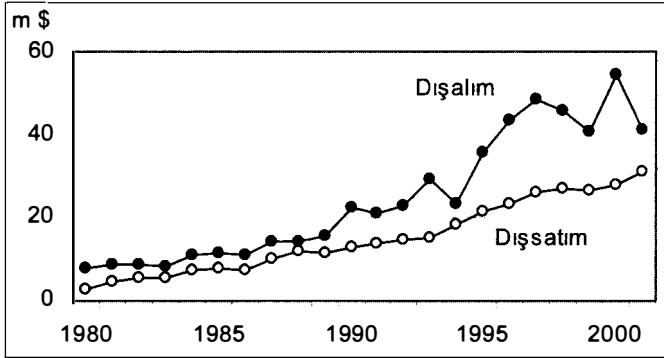
Bilgisayarın önerebileceği bir başka çizim türü de **uçuşan şeritler**dir. Çizim 7.4 (e) bu tür bir çizimdir. Bunun “uçuk” bir gösterim olduğu söylenebilir.



ÇİZİM 7.4 (e) Türkiye'nin dışalımıyla dışsatımı 1980-2001 (milyar \$)
Çizelge 7.3'teki verilerle düzenlenmiştir.

Çizim 7.4 (c) şakaysa bu da karabasandır. Çünkü bu çizim üç boyutlu çizimlerin bütün sakıncalarını taşımaktadır. Bu bağlamda, arka-yan duvarlar, tabandaki gölge, ölçek çizgilerinin hem tabanda hem duvarda uzatılması, iki değişkendeki dalgalanmaları gösteren şeritlerin üç boyutlu uzayda neredeyse uçuşmaları gözü yoran gösterimlerdir. Bunların çizimin iletilisine pek bir katkıda bulundukları da söylenemez. Şeritlerin boşluktaki konumları, hangi değerlerin karşılığı oldukları konusunda okuyucuya hiç de yardımcı değildir. Herhangi bir yılda değişkenlerden herhangi birinin, özellikle de arka duvardaki ölçek çizgilerine daha uzak olan dışalımın büyüklüğünü kestirmek neredeyse olanaksızdır. Bunun nedeni hepsinin ortak bir eksen den uzaklıkları yansıtmamalarıdır. Her iki değişken için de aynı taban geçerlidir ama çizimin üç boyutlu olması nedeniyle bu taban, okuyucuya eğik düzlem olarak görünmekte, böylece değişkenlerin yerden yükseklikleri, aynı hizadan başlamamaktadır. Dolayısıyla bu çizim bakana hemen hiç bir şey söylememektedir.

Öyleyse zaman serileri için yaptığımız bütün bu çizimleri bir yana koyabiliriz. Çizim 7.4 (f)'ye bakarsak her iki değişkenin de zaman içindeki değerlerinin birer noktaya dönüştüğünü, bu noktaların aralarının da zaman içindeki bağlantıyı sağlamak üzere birleştirildiğini görebiliriz. Böylece hem belli yıllardaki değerlerin dikey eksenindeki konumu, hem de bu değerlerin dönem boyunca değişimi çok kolaylıkla izlenebilmektedir. Ayrıca dışalım ile dışsatımın aynı yıla ilişkin değerleri kolayca karşılaştırılabilmekte, arasındaki farkın boyutu algılanabilmekte, zaman içindeki genel eğilimleri konusunda bir fikir edinilebilmektedir.



ÇİZİM 7.4 (f) Türkiye'nin dışsalımıyla dışsatımı 1980-2001 (milyar \$)
Çizelge 7.3'teki verilerle düzenlenmiştir.

Değişken adları ayrı bir kutunun içinde değildir, çizimin içinde çizgilerin hemen yakınına, bir karışıklığa yol açmayacak biçimde yerleştirilmiş, böylece her çizginin hangi değişkeni gösterdiğini anlamak için yandaki açıklamalara başvurma gereği ortadan kaldırılmıştır. Son çizim şimdiye kadar gösterilen zaman serisi çizimleri içinde en iyisi olarak değerlendirilebilir. Elbette okuyucu bundan daha iyisini yapabilir. Bu çizimlerin veri yoğunluklarını bulmak okuyucuya bırakılmıştır.

Son çizim bize ne anlatmaktadır? Dönem boyunca dışsalım değeri hep dışsatımın üzerindedir. Her iki seri de genel olarak artmaktadır. Ancak dışsalım daha büyük dalgalanmalar göstermekte, ekonomik ortamdan daha çok etkilenmektedir. 1994, 1999, 2001 bunalımlarının dışsalımı olumsuz etkilediği açıktır. Üreticilerin böyle dönemlerde can havliyle dışsatıma yöneledikleri izlenimi doğmaktadır.

Dışsalımla dışsatımın farkı anlamındaki dış ticaret açığı da bu çizimden izlenebilir. Ancak eğimlerin büyük olduğu yıllarda bunların farklarını karşılaştırmak yanıltıcı olabilir. Sözelimi bu açığın 1993'te mi, 1995'te mi, 1999'da mı daha fazla olduğu sorulsa bu çizime bakarak nasıl bir yanıt verirdiniz? Önce yanıtınızı verin, sonra Çizelge 7.3'ten bu farkları hesaplayıp doğru yanıtı öğrenin. Küçük bir şaşkınlık geçirebilirsiniz. Açığı yanıltmadan görebilmek için önerilebilecek bir yol, ayrı bir çizim düzenlemektir. Bunu yapması okuyucudan istenmektedir.

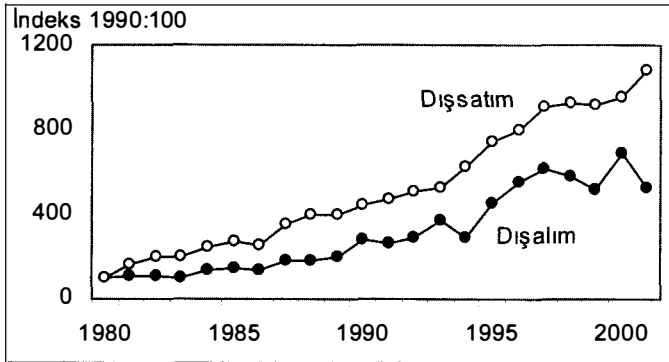
Bu çizimde dışsalım-dışsatım mutlak değerlerinin gelişmesini izledik. Bunların göreceli değişimlerini görebilmek için, aynı verilerden hesaplanan birer indeksin yardımına başvurup her bir değişkenin kendi 1980 düzeyine oranla nasıl geliştiğine bakabiliriz. Bunun için her değişkenin her yıldaki milyar \$ değerini 1980 değerine bölüp 100 ile çarpmak yeterlidir. Çizelge 7.15 ile Çizim 7.5'te bu yolla hesaplanmış indeks değerleri yer almaktadır.

ÇİZELGE 7.15 Dışalım ile dışsatımın indeks değerleri – 1980-2001 (1980:100)

Dışalım Dışsatım			Dışalım Dışsatım		
1980	100	100	1991	266	469
1981	113	162	1992	290	507
1982	113	197	1993	372	528
1983	104	197	1994	295	624
1984	137	245	1995	452	745
1985	143	276	1996	552	800
1986	141	259	1997	615	907
1987	180	352	1998	581	931
1988	181	403	1999	515	917
1989	200	400	2000	690	959
1990	282	448	2001	524	1079

Çizelge 7.3'ten hesaplanmıştır.

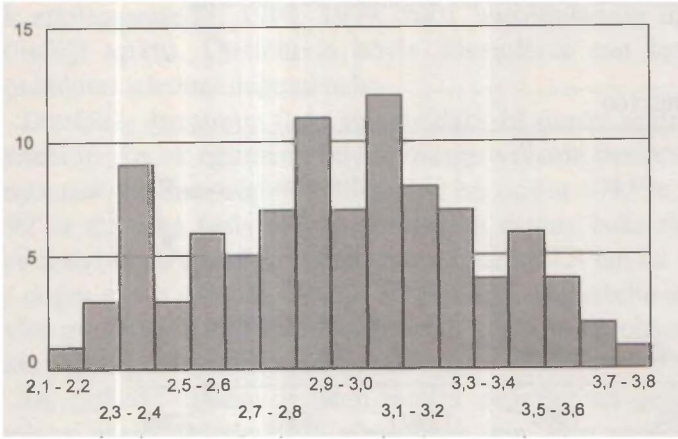
Çizelge 7.15'teki indeks değerleri (sonuncusu dışında) üç basamağa yuvarlatılmıştır. Ama Çizim 7.5'te yuvarlatmadan önceki değerler kullanılmıştır. Gerçi yuvarlatılmış değerler de kullanılsaydı hiçbir göz aradaki farkı algılayamazdı. Bu çizelgeye bakınca hemen şunu görebiliyoruz. 1980'den 2001'e kadar olan dönemde dışalım 5 katını, dışsatım 10 katını aşmıştır. Yani biri %400'den, öbürü % 900'den çok artmıştır. Bunu Çizelge 7.15'in en altında sağdaki iki değerden okuyabiliriz.

**ÇİZİM 7.5** Dışalım-dışsatım indeks değerleri 1980-2001 (1980:100)

Çizelge 7.15'teki verilerle düzenlenmiştir.

Buraya kadar verdiğimiz örneklerde ad ölçeğinde ölçülmüş bir değişkenin kümelerindeki sıklık sayılarını kullanarak çubuk çizimi yapabileceğimizi gördük. Acaba öbür ölçeklerde ölçülmüş sürekli bir değişkenin sıklık sayılarıyla ne yapabiliriz? Burada değişken sürekli bir olguyu yansıttığı için çubukları birbirine bitişik çizebiliriz. Böyle bir çizime **histogram** ya da **alan çizimi** denir. Çünkü birbirine bitişik çubukların oluşturduğu toplam alan gözlem sayısına eşit olacak biçimde çizilir. Her bir çubuğun kapladığı alan da değişkenin o değer kümesindeki sıklık sayısına karşılıktır. Dolayısıyla böyle çizimlere sıklık çizimi de diyebiliriz. Bu tür çizimlere günlük basında pek rastlanmaz ama istatistik kullanan bilimsel yazılarda sıkça görülür.

Bunu Alistırma 7.7'deki verilerle örnekleyelim. Oradaki çizelgede İTÜ İşletme Mühendisliği bölümü 1996-97 yaz yarıyılı mezunlarının ortalama notları gösterilmişti. Çizim 7.6'dan görüldüğü gibi bunun çubuk çiziminden farkı, not değişkeninin değerleri yatay eksende her noktada yer alabileceği için, çubukların birleştirilmiş olmasıdır. Bu amaçla yatay eksen 0,1'lik 17 aralığa bölünmüş, 2,1 ile 3,8 arasındaki her değer düştüğü aralık içinde yer almış, çubukların boyları o aralıktaki mezun sayısına göre ayarlanmıştır. Sözgelimi 2,2 ile 2,3 arasında 3 kişi vardır. Elbette aralıklar birbirine eşitse çubuk boyları sıklık sayılarını gösterir. Aralıklar eşit değilse ne yapılacağına ileride değinilecektir.



BITİRME NOTLARI

ÇİZİM 7.6 Bitirme notları için histogram
Alistırma 7.7'deki verilerle düzenlenmiştir.

Histogram bize bitirme notlarının 2,7 ile 3,3 arasında yoğunlaştığı, en çok öğrencinin 3,0 – 3,1 aralığına düştüğünü göstermektedir. Bazı not aralıklarındaki öğrenci sayısı komşularına göre daha düşük görünmektedir. Bu olgu sizce nasıl açıklanabilir? Çukur oluşturan bu not aralıklarının 2,5, 3, 3,5 gibi notlara geçilirken görülmesi ne anlama gelir? İTÜ’de 2,5 – 3 arasındaki ortalamaların *iyi*, altındakilerin *orta*, üstündekilerin *pekiyi* sayıldığını söylemem size yardımcı olabilir mi?

Notların dağılımı hayli **bakışık** (simetrik) sayılabilir. **Çarpıklık** konusunda bir şey söylemek zor. 2,3 – 2,4 aralığındaki yoğunluk olmasa belki birazcık sola çarpıklıktan söz edilebilir. Yani küçük değerlere doğru gidildikçe öğrenciler az da olsa seyreliyor denebilir. Aşırı düşük ya da aşırı yüksek bir not yoktur. Zaten mezun olmak için en az 2 ortalamanın gerektiği, 4’ten yüksek ortalama da olamayacağı için bu doğaldır. Belki siz bu yorumları geliştirebilirsiniz.

Şimdi aynı notları **dal-yaprak** denilen başka bir çizimle gösterelim. Çizim 7.7’deki gösterim doksan derece döndürülüp bakılırsa az önceki alan çizimine benzeyeceği açıktır. Tek fark oradaki çubukların yerine bir takım sayıların yazılmış olmasıdır. Burada dal başlığı altındaki 17 tane sayıyı ilk çizimin 17 aralığına benzetebiliriz. Çubukların boyları da aynı hizada dizilmiş rakamlara karşılıktır.

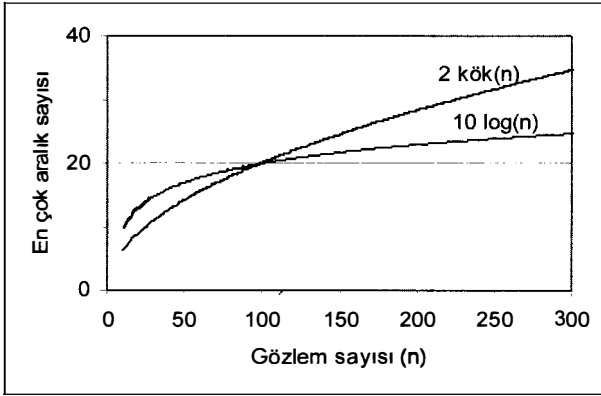
Bu gösterim, bir bilgisayar çıktısı Türkçeleştirilip bazı bölümleri atılarak düzenlenmiştir. Dal başlığı altında notların tam sayılarıyla onda birlik bölümleri yer almakta, yüzde birlik bölümleri yaprak başlığı altında sıralanmaktadır. Sözelimi ilk değer 2,10, ikinci satırdakiler 2,20, 2,22, 2,27 diye okunmalıdır. Bunlar histogramın soldan ikinci çubuğundaki 3 kişinin ortalama bitirme notlarıdır. Yukarıdaki histogramda görebildiğimiz her şey bu dal-yaprak gösterimi ile de görülebilir. Ek olarak her öğrencinin ortalama notu da tek tek okunabilir.

Dal	Yaprak
21	0
22	027
23	01223
24	167
25	00455 8
26	13468
27	23568 89
28	02334
29	00255 77
30	03777
31	00001
32	22356 88
33	1238
34	02346 7
35	0225
36	15
37	1
0,1	0,01

ÇİZİM 7.7 Bitirme notları için 0,1 aralıklı dal-yaprak gösterimi
Alıştırma 7.7'deki verilerle düzenlenmiştir.

Hem histogramda hem dal-yaprak gösteriminde 17'şer aralık kullandık. Bunun nedeni aralıkların 0,1 gibi yuvarlak bir sayıya karşılık gelmesiydi. Daha dar ya da daha geniş aralıklar kullanırsak aralık sayısının da çoğalıp azalacağını düşünebiliriz. Acaba böyle çizimlerde aralık sayısı kaç olmalıdır? Önce alt sınır için genel kabul gören sayıyı söyleyelim: 5. Öyleyse aralık sayısı 5'ten küçük gösterimlere gitmeyelim. Elbette 5 ya da 5'e yakın sayılar gözlem sayısı az veri kümeleri için uygundur. Gözlem sayısı arttıkça aralık sayısını da çoğaltmamız uygun olur, aksi halde veri kümesinin içinde taşıdığı yapıya ilişkin önemli özellikleri gözden kaçırabiliriz. Peki en çok kaç kadar çıkabiliriz? Bu da yine gözlem sayısına bağlıdır. Emerson, Hoaglin (1983a) yaptıkları bir araştırmada bu konuda iki öneriye rastlamıştır. Bunlardan biri $10 \log(n)$, öbürü de $2\sqrt{n}$ 'dir. Burada n gözlem sayısıdır. Bu iki öneriye göre çeşitli n değerleri için *en yüksek* aralık sayıları Çizim 7.8'de gösterilmiştir. Bu çizimde n 100'den küçükse $10 \log(n)$ önerisinin, 100'ün üstündeyse $2\sqrt{n}$ önerisinin daha yüksek sayılar verdiği görülmektedir. Emerson, Hoaglin (1983a) bu durumda hep alttaki eğrinin gösterdiği sayıların kullanılmasından yanadır. Yani 100'den az gözlemler

için $2\sqrt{n}$, 100'den çok gözlemler için de $10 \log(n)$ ile hesap yapmayı öğütlerler. Bu da aralık sayısının üst sınırını aşağıda tutar. n 100 iken her iki yol da 20 sayısını vermektedir. Unutmayalım ki bu hesaplamalarla bulunan sayılar en üst sınırlardır. Uygulamalarda bu sınırın altında kalınabilir. Çizimde n 10'dan başlayarak değerler almaktadır. Bu da bize 10'dan az gözlemimiz varsa bu tür çizimlere gitmenin pek doğru olmadığı yolunda bir uyarıdır. Örneğimizdeki 96 gözlem için kullanılabilecek en yüksek aralık sayısı $10 \log 96 = 19$ 'dur. Öyleyse kullandığımız 17 aralık sayısı uygundur.



ÇİZİM 7.8 Histogramlarla dal-yaprak gösterimlerinde en yüksek aralık sayısı Emerson, Hoaglin (1983a)'daki verilerden düzenlenmiştir.

Acaba bitirme notlarını 0,2 gibi daha geniş aralıklara ayırsaydık ne olurdu? Bunun yanıtı Çizim 7.9 (a)'da verilmiştir. Burada her satırda 0,1'lik dilimlerden ikişer tanesi gösterilmiştir. Sol sütundaki 3, 2 sayıları birler basamağındaki değerlere karşılıktır. Onların yanındaki ikişer harf de 0,1'lik dilimler için konmuştur. “sb” sıfır-biri, “iü” iki-üçü, “db” dört-beşi, “ay” altı-yediyi, “sd” sekiz-dokuzu temsil eder. Böylece ilk satırda 3,6-3,7 ile başlayan bitirme notları yer almıştır. Bunların yüzde bir basamakları yuvarlatılmamış, kesilip atılmıştır. Böyle bir uygulamadan kesme-yuvarlatma konuları anlatılırken söz edilmişti. Başlangıçtaki üç basamaklı asıl sayıya geri dönmek gerekirse bunun kolayca yapılabilmesi için yuvarlatma yerine kesme daha uygundur.

3 ay	667
3 db	44444 45555
3 iü	22222 22333 3
3 sb	00000 00011 11111 11111
2 sd	88888 88888 89999 999
2 ay	66666 77777 77
2 db	44455 5555
2 iü	22233 33333 33
2 sb	1
1	0,1

ÇİZİM 7.9 (a) Bitirme notları için 0,2 aralıklı dal-yaprak gösterimi
Alıştırma 7.7'deki verilerle düzenlenmiştir.

Son dal-yaprak gösterimindeki bir başka yenilik, daldaki notların büyükten küçüğe doğru dizilmesidir. Bunun da nedeni çizimlerde büyük sayılara geleneksel olarak eksenin üst bölümlerinde yer verilmesidir. Böylece öbür çizimlerle uyum sağlanır.

İki dal-yaprak gösteriminden hangisi kullanılmalıdır? Bu soruya kesin bir yanıt vermek zordur. Şu kadarını söyleyelim ki deneyimli uzmanlar bile tek bir dal-yaprak ile yetinmeyip iki tane çizer, bunların ikisinden de verilerin yapısı konusunda farklı bilgiler edinebilirler. Burada ikinci çizim verileri daha özetlenmiş gösterir, 2,3 ile başlayan notlar dışında hafifçe sola çarpık bir dağılım sergiler, ilkindeyse 2,8'le, 3,1'le başlayan notların yüksek sıklıkları daha göze batar.

Dal-yaprak gösterimleriyle bir veri kümesinin ortalama düzeyi, ortalama-uzaklaşma derecesi, çarpıklık, basıklık gibi özellikleri, varsa aşırı değerleri kolayca incelenebilir, aynı değişkenli birden çok veri kümesi birbirleriyle karşılaştırılabilir. 10. bölümde sorgulayıcı veri çözümlemesi yapılırken dal-yaprak gösterimleri ayrıntılı olarak yeniden ele alınacaktır.

Histogramın yatay eksenindeki küme aralıkları eşit değilse ne yapacağımıza da kısaca değinelim. Bu durumda o aralığa düşen alanın verinin gerçek değerini yansıtması temel kuraldır. Sözelimi bir aralık öbür aralıkların iki katıysa o aralığa düşen sıklık ikiye bölünerek bulunan yükseklikle gösterilir. Böylece çizilen alan verinin değerini doğru yansıtır.

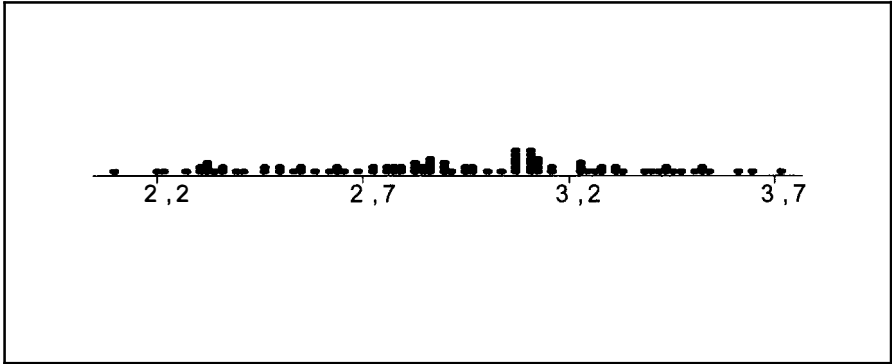
Bitirme notlarına X dersek, $X < 2,50$ orta, $2,50 \leq X < 3,50$ iyi, $3,50 \leq X < 4,00$ pekiyi anlamına geldiğini biliyoruz.. Bunu yansıtan bir histogram nasıl çizilir? Dal-yapraktan 16 kişinin fakülteyi orta, 73 kişinin iyi, 7 kişinin de pekiyi ile bitirdiği kolayca görülebilir. Öyleyse 16 kişiyi $2,5 - 2 = 0,5$ birimlik, 73 kişiyi $3,5 - 2,5 = 1$ birimlik, 7 kişiyi de $4 - 3,5 = 0,5$ birimlik tabanı olan dikdörtgenlerle göstereceğiz demektir. Bu dikdörtgenlerin

yüksekliklerini kişi sayısını taban uzunluğuna bölerek bulabiliriz. Bunları sırasıyla $16 / 0,5 = 32$, $73 / 1 = 73$, $7 / 0,5 = 14$ birim yükseklikte göstermek gerekir. Çizimin yapılması okuyucuya bırakılmıştır.

Histogram, sıklık değerleri yerine görelî sıklıklar kullanılarak da çizilebilir. Bu durumda çizimin içeriği hiç değişmez, yalnızca dik eksen de sıklık sayıları yerini görelî sıklıklara bırakır.

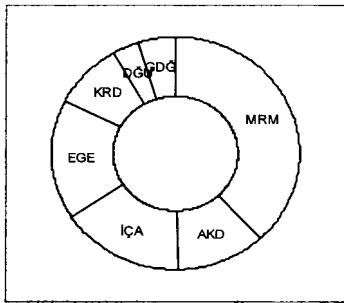
Gözlem sayısı çok olan histogramlarda verilerin öbeklendikleri aralıklar iyice daralır. Çizimdeki dikdörtgenlerin tabanları çok küçülürse genişlikleri birer noktaya, dikdörtgenler de birer çizgiye dönüşür. Bu durumda dikdörtgenleri çizmek yerine birer nokta olan tavanlarının oluşturduğu eğriyi çizebiliriz. Buna **sıklık eğrisi** denir. İstatistikte çok sık kullanılan **sürekli olasılık dağılımları**, bu yolla çizilmiş görelî sıklık çizimlerinden başka bir şey değildir.

Örneğimizdeki bitirme notlarını Çizim 7.9 (b)'deki gibi bir **nokta çizimi**yle de gösterebiliriz. Bu çizimin yapılması çok kolaydır. Bir veri değerinin yatay eksen de konumu hizasına bir nokta koymak yeterlidir. Aynı değerde birden çok veri varsa noktalar üstüste konur. Bu çizimle verilerin tekil konumları, kümenin biçim özellikleri belirgin biçimde görülmektedir.



ÇİZİM 7.9 (b) Bitirme notları için nokta çizimi

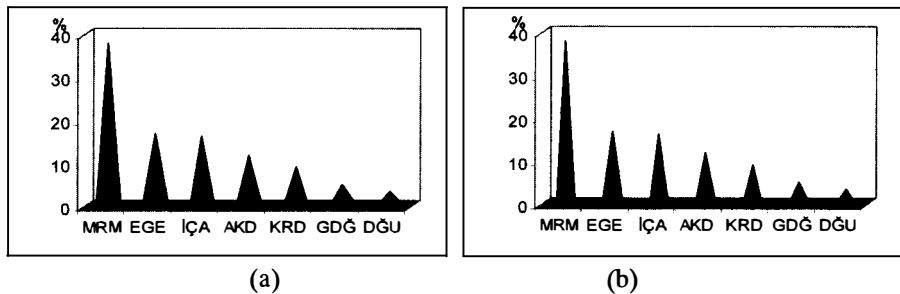
Şimdi başka çizim türlerine dönelim. Aslında bir bilgisayar yazılımının önerebileceği pek çok çizim türü vardır. Ama uygulamada bunların pek azı kullanılır. Sözelimi yüz sayfası bütünüyle çizimlerle çizelgelerden oluşan bir kitapta (Sönmez, 2002) 66 çubuk çizimi, 57 zaman serisi çizimi, 38 pasta çizimi yer almaktadır. Yine çizimlerle dolu bir başka kitap (Sutcliffe, 2002) 58 çubuk, 31 zaman serisi, 6 pasta, 6 serpilme çizimi, 13 harita, 5 başka tür çizimden oluşmaktadır.



ÇİZİM 7.10 Bölgelerimizin GSYİÜ içindeki % payları-2000 (1987 fiyatlarıyla)
Çizelge 7.14'teki verilerle düzenlenmiştir.

Pasta çizimine uygun verilerle **simit çizimler** de yapılabilir. Ayrıca gerek pasta, gerek simit çizimler kimi parçaları bütünden kopmuş olarak da gösterilebilir. Çizim 7.10 bölgelerimizin GSYİÜ içindeki payını simit çizimle gösteren bir örnektir. Çizim 7.2 (b) ile bu çizimi karşılaştırmak okuyucuya bırakılmıştır.

Çubuk çizimlerle gösterilebilen veriler, **piramit, koni** gibi çizimlere de konu olabilmektedir. Çizim 7.2 (c, d) ile karşılaştırılması okuyucudan beklenen Çizim 7.11 (a) ile (b) de bu tür çizimlere birer örnektir.



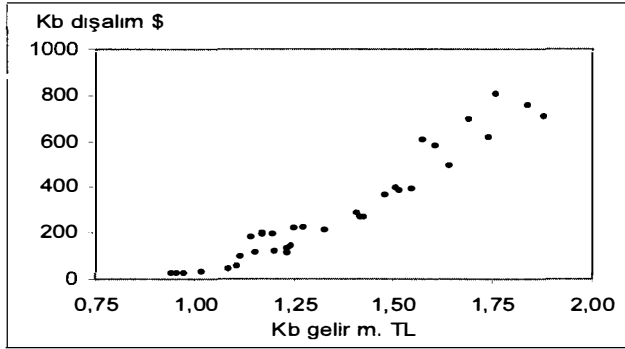
ÇİZİM 7.11 Bölgelerimizin GSYİÜ içindeki % payları-2000 (1987 fiyatlarıyla)
Çizelge 7.14'teki verilerle düzenlenmiştir.

Özellikle borsa ile ilgilenenlerin pek tuttukları bir çizim de günlük en düşük, en yüksek fiyatlarla kapanış fiyatlarının gösterildiği çizim türüdür. **Düşük-yüksek çizimi** diye adlandırılacak bu çizim yalnız borsa verilerine değil, her türlü veriye uygulanabilir. Aşağıda 2002 yılında iktisat programlarına giren devlet üniversiteleri 2. öğretim öğrencileriyle vakıf üniversiteleri bursuz öğrencilerinin taban-tavan puanlarını gösteren Çizim 7.12'yi buna bir örnek olarak alabilirsiniz.

lara bir de daha yüksek puanlı burslu öğrenciler eklendiğinde puan eşitsizliği iyice artmaktadır.

Ayrıca vakıf üniversitelerinin 2002'de taban puanı sıralamasına göre pek de başarılı olamadıkları anlaşılmaktadır. Koç Üniversitesi 6., Bilkent 10. sırada yer bulabilirken, Sabancı Üniversitesi, Kahramanmaraş Sütçü İmam Üniversitesi'nin hemen önünde 35. sıraya girmektedir. Ötekilerin sıralamadaki yerleri bunlardan sonra gelmektedir. Bunda 2001 iktisadi bunalımının rolü önemli olabilir.

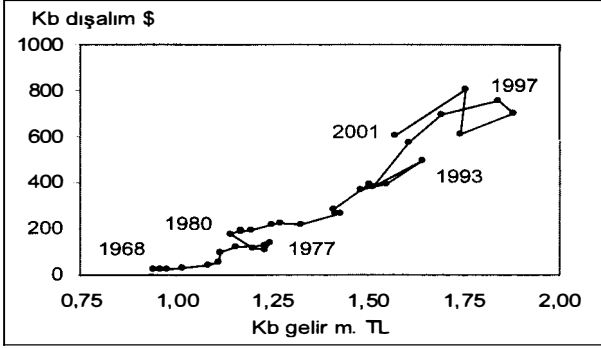
Buraya kadar bazı değişkenlerin zaman içindeki izlerine bakılmakla ya da kesit verilerinin dağılımları incelenmekle yetinilmiştir. Bir sonraki adım, değişkenlerin aralarındaki ilişkilerin incelenmesi olacaktır. Bu amaçla 1968-2001 döneminde kişi başına (kb) gelirle kişi başına dışalım arasındaki ilişkiyi gösteren serpilme çizimi Çizim 7.13 (a)'da gösterilmiştir.



ÇİZİM 7.13 (a) Kb gelir (1987 sabit fiyatlarıyla m. TL) ile kb dışalım (\$) ilişkileri 1968-2001

DİE, 2002a'dan hesaplanan verilerle düzenlenmiştir.

Bu çizimde yatay eksene kb gelir, dikey eksene de kb dışalım konmuştur. Her nokta belli bir yılın gelir-dışalım veri çiftini göstermektedir. Sözgeçlimi en yukarıdaki nokta kb gelir yaklaşık 1,75 m. TL iken kb dışalımın yaklaşık 800 \$ olduğu anlamına gelmektedir. Kb gelirin büyük değerlerine kb dışalımın büyük değerlerinin karşılık geldiği gözlenebilir. Bunun anlamı kb gelirimiz arttıkça daha çok dışalım yapmaya başladığımızdır. Özellikle kb gelirin 1,60 m. TL'den büyük değerlerine karşılık gelen dışalım değerlerinin genişçe bir aralık gösterdiğini söyleyebiliriz. Aynı olgu daha küçük ölçekte 1,15-1,25 m. TL kb gelir aralığında da geçerlidir. Bunların hangi yılların değerleri olduğunu görmek istersek Çizim 7.13 (b)'ye bakabiliriz.



ÇİZİM 7.13 (b) Kb gelir(1987 sabit fiyatlarıyla m. TL) ile kb dışalım (\$) ilişkileri 1968-2001

DİE, 2002a'dan hesaplanan verilerle düzenlenmiştir.

Peşpeşe gelen yılların değerlerinin birer çizgiyle birbirine bağlandığı bu çizimde şunu gözleyebiliriz. Bunalım yıllarında hem gelir hem dışalım düşmekte, bunalımdan çıkınca aynı gelir düzeyinde daha yüksek dışalım yapılmaktadır. 1980 öncesi, 1994, 1999 bunun örnekleridir. Dolayısıyla iki değişken arasındaki ilişki ayrıntılı olarak incelenirken bu olgunun göz önüne alınması gerekir. İki değişken arasındaki ilişkiyi gösteren serpilme çizimleri gündelik basında pek görülmez; daha çok raporlar, mesleki dergiler, bilimsel yayınlarda boy gösterir. Oysa kamuoyu böyle çizimlerden pek çok şey öğrenebilir. Durum dünya basını için de pek farklı değildir. Tufte'nin 1974-80 dönemini kapsayan araştırmasına göre (1983:83) birden çok değişken içeren ama zaman serisi çizimi olmayan gösterimlerin toplam çizim sayısına oranı dünya basınında şöyledir:

<i>Akalata (Kızıl Bayrak)</i> (Japonya, günlük, satış 30.000)	%9,3
<i>Asahi Shimbun</i> (Japonya, günlük, 8.000.000)	7,6
<i>Der Spiegel</i> (Almanya, haftalık, 1.000.000)	5,7
<i>The Economist</i> (İngiliz, haftalık, 170.000)	2,0
<i>Nihon Keizai Shimbun</i> (Japonya, günlük finans, 1.700.000)	1,7
<i>Le Monde</i> (Fransa, günlük, 440.000)	0,7
<i>Business Week</i> (ABD, haftalık, 800.000)	0,6
<i>New York Times</i> (ABD, günlük, 900.000)	0,5

Görüldüğü gibi serpilme çizimlerinde Japon basını ön sıralardadır. İngiliz *The Times*, Amerikan *Wall Street Journal* gazeteleriyle Amerikan

Time dergisinde bu dönemde basılan yüzlerce çizimden biri bile serpilme çizimi değildir.

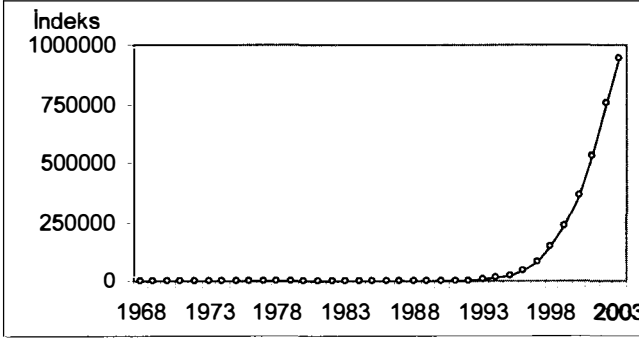
Buraya kadar gördüğümüz çizimlerdeki bütün eksenler verileri hiç bir **dönüştürme**ye uğratmadan oldukları gibi ölçmekteydi. Oysa öyle durumlar olabilir ki çizmeyi amaçladığımız değişkenlerden birinin (ya da bir kaçının) içerdiği değerlerin en küçüğü ile en büyüğü arasındaki oran çok büyük olabilir. Böyle bir değişkenin çiziminde eksen en büyük değeri de içerecek biçimde düzenleneceğinden, küçük değerler arasındaki küçük ama anlamlı sayılabilecek farklar gözden kaçabilir. Bunu önlemek amacıyla bu tür çizimlerde eksenin bir ya da ikisi de gözlem değerlerini oldukları gibi değil, dönüştürerek gösterecek biçimde düzenlenir. Bu işlerde en çok kullanılan, **logaritmali dönüştürme**dir. Gözlem değerlerinin asıl değerlerinin değil logaritmalarının eşit aralıklara ayrıldığı ölçeklerin kullanıldığı böyle eksenlere **logaritmali eksen** adı verilir. İTO İstanbul ücretliler geçinme indeksi-ne ilişkin böyle bir örnek aşağıdaki çizelgeyle çizimlerde verilmiştir.

ÇİZELGE 7.16 İTO İstanbul ücretliler geçinme indeksi – 1968:1,00

Yıl	İndeks	Yıl	İndeks	Yıl	İndeks	Yıl	İndeks
1968	1,00	1977	3,96	1986	135,59	1995	21890,47
1969	1,05	1978	6,66	1987	195,45	1996	40751,88
1970	1,15	1979	11,79	1988	344,95	1997	78728,17
1971	1,34	1980	20,73	1989	593,38	1999	233808,79
1972	1,52	1981	28,05	1990	984,44	2000	362780,08
1973	1,73	1982	37,72	1991	1651,97	2001	528783,68
1974	2,13	1983	48,34	1992	2866,55	2002	753224,30
1975	2,59	1984	69,45	1993	5066,79	2003	941923,51
1976	3,02	1985	99,72	1994	11257,26		

İTO, 2004:14.

Çizelge 7.16'daki verilere bakıldığında 2002'de indeks değerinin, 1968 değerinin 750 bin katını aştığı görülmektedir. Böyle geniş bir aralıkta sıralanan veriler aynen çizilirlerse Çizim 7.14 (a) elde edilir. Bu çizime göre 1968-90 arasında İstanbul'da fiyatlar sanki hiç artmamış gibidir. Oysaki çizelgede 1990'ların başında fiyatların 1968'in bin katına ulaştığı görülmektedir. Bu çizimi daha iyi bilgi sunar duruma getirebilmek için dik eksenini logaritmali alabiliriz.

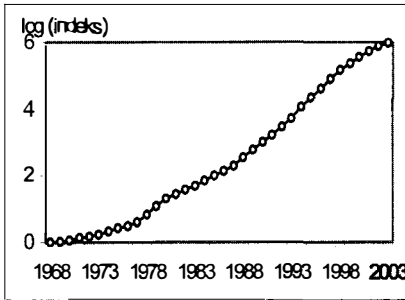


ÇİZİM 7.14 (a) İTO İstanbul ücretliler geçinme indeksi (1968:1,00)

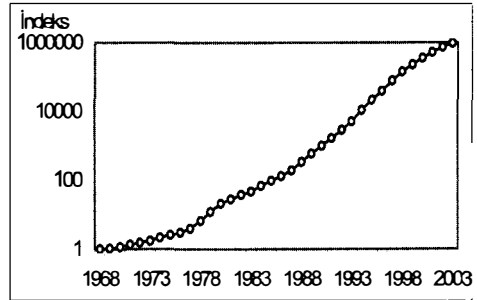
Çizelge 7.16'dan düzenlenmiştir.

Çizim 7.14(b)'de dik ekseninde fiyatların logaritma değerleri yer almaktadır. Buna göre sol eksenin en üstündeki 6 rakamı 10^6 yani 1.000.000 değerine karşılıktır. Bu da bize, bu eksenindeki ölçümlerin, verilerin kendi değerleri bakımından değil, birbirlerine göre katları bakımından eşit uzaklıkta gösterildiğini anlatır.

Çizim 7.14 (c)'de ise aynı veriler bu kez logaritmaya göre düzenlenmiş bir dik ekseninde gösterilmiştir. Böylece 1980'lerin ortalarında fiyatların 1968'e göre 100 katına, 1990'ların ortalarında da 10.000 katına çıktığını söyleyebiliriz. Demek ki iki nokta arasındaki eğim, fiyatların o yıl % kaç arttığının bir göstergesidir. Eğrinin dikleştiği dönemlerde fiyatlar daha hızlı artmıştır. 1970'lerin ortalarına kadar görece olarak yumuşak artış gösteren fiyatlar birden hızla artmaya başlamış, 1980'den sonra bir süre artış yavaşlamışsa da 1980'lerin ortalarında yeniden hızlanmış, 1990'ların ortalarında bu hız daha da artmıştır.



(b)

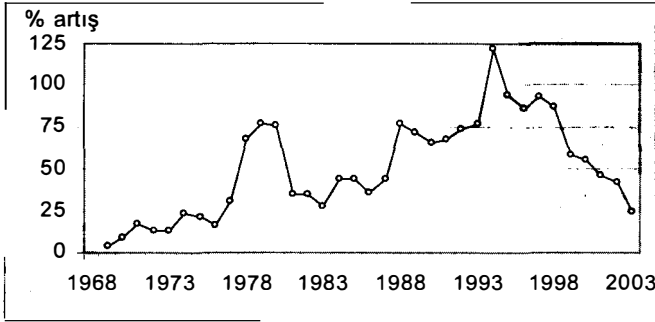


(c)

ÇİZİM 7.14 İTO logaritmalı ücretliler geçinme indeksi (1968:1,00)

Çizelge 7,16'daki verilerin logaritması alınarak düzenlenmiştir.

Bu çizim bize fiyatların genel gidişi konusunda iyi bir fikir vermektedir ama tek tek yıllar arası fiyat artışlarını karşılaştırmak kolay olmamaktadır. Her yılın fiyat artış oranlarını birbiriyle karşılaştırabilmek için dik eksen, artış yüzdelelerini gösterecek biçimde çizmek önerilebilir. Çizim 7.14 (d) bize indeksin her yıl % kaç değiştiğini gösterir. Buradan 1980 öncesi yıllarda fiyatların hızla yükseldiği kolayca anlaşılır. 1980'lerin ilk yarısında görece olarak hız kesen fiyat artışlarının ikinci yarıda hemen 1979 düzeyine tırmandığı, 1994 bunalım yılında % 120'yi aştığı görülür. Son yıllarda enflasyon oranı % 50'nin altına inmiştir.



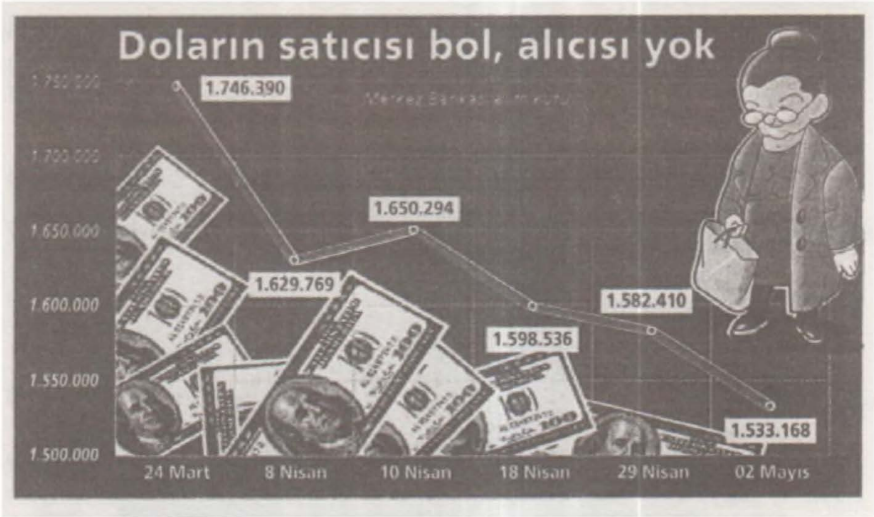
ÇİZİM 7.14 (d) İTO İstanbul Ücretliler geçinme indeksindeki % değişimler
Çizelge 7.16'dan düzenlenmiştir.

7.4 KÖTÜ ÖRNEKLER

İyi bir çizim için uyulması gereken özellikler yukarıda anlatıldı. Ancak rastladığımız kimi çizimlerde bu ilkelere uyulmadığı anlaşılmaktadır. En çok yapılan yanlışları örnekler üzerinde görelim. Çizim 7.15 (a) bu konuda bize hayli yardımcı olacağı benzer. Gündelik basında karşılaştığımız çizimlerin en önemli sorunu çizimin içine, yanına, köşesine bir yere ille de bir resim koyma zorunluluğu var sanılmasıdır. Bu çizimde de oraya buraya serpiştirilen resimlerden çizimin kendisi zor görülmektedir. Sağdaki kadın, dolarları alıp elindeki pazar çantasına dolduracak gibi mi geliyor size? Okuyucu başlıkta geçen dolar sözcüğünü anlamaz diyerek mi çizime dolar resimleri dolduruluyor? Çizimlere gereksiz hiçbir resim konmaması kuralını birinin çizirler anlatması gerekiyor.

Altbaşlık okunmuyor mu dediniz? Çizimin gazetedeki aslında da okunamıyor zaten. Zemine boca edilen mürekkep o kadar yoğun ki incecik

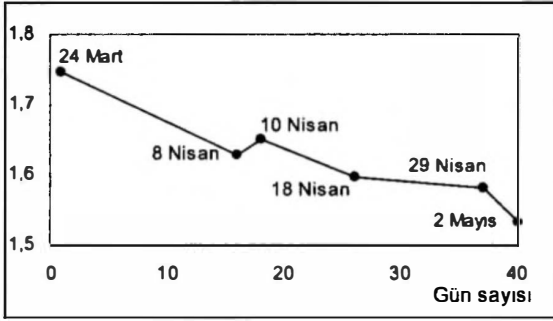
harfli yazının içine taşarak onu neredeyse görünmez kılmış. Oysa “Merkez Bankası alım kuru” altbaşlığı ev kadınından da, dolarlardan da daha önemli bir bilgi içeriyor. Hele bir de yatay eksendeki tarihlere bakınca insanın saçını başını yolası geliyor. Eşit aralıklarla sıralanan bu tarihlerin aralıkları hiçbir yerde eşit değil. Kimi 15, kimi 2 günlük aralar. Çizime göre doların TL fiyatı ilk iki tarih arasında en hızlı düşmüş görünüyor. Ama gerçek öyle değil. Çünkü bu düşüş 15 günde olmuş. Demek ki bu süre içinde günde ortalama $(1.746.390 - 1.629.769) / 15 = 7.775$ TL düşüş olmuş. Buna karşılık 29 Nisan – 2 Mayıs arasındaki son üç günde ortalama düşüş 16.414 TL. İlkinin iki katından çok ama çizim bize tersini söylüyor.



ÇİZİM 7.15 (a) Kötü bir çizim örneği

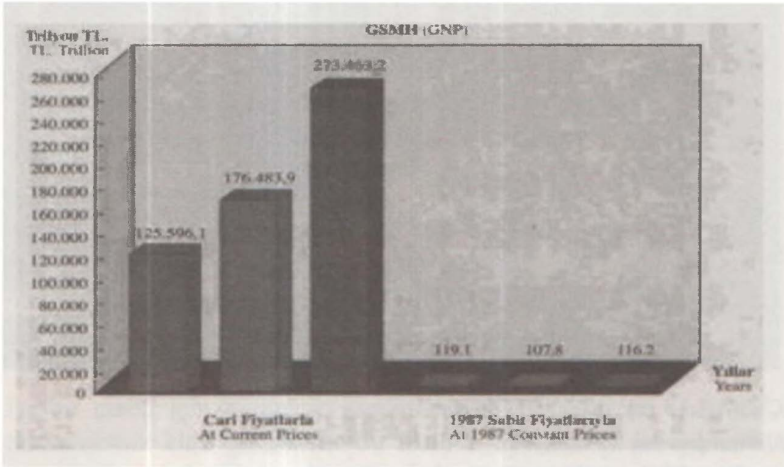
Milliyet gazetesi, 2 Mayıs 2003.

Bu çizimi yeniden düzenleyecek olsak Çizim 7.15 (b)'deki bir görünümle karşılaşırız. Bu düzeltilmiş çizimde bilgi doğru aktarılıyor. Son günlerdeki hızlı düşüş burada çok daha iyi görülebiliyor. Karmakarışık resimlerle bakanın dikkati dağıtılmıyor.



ÇİZİM 7.15 (b) Çizim 17.15 (a)'nın düzeltilmiş biçimi

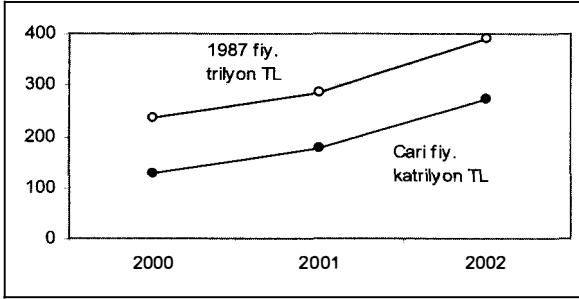
Sık yapılan bir başka yanlışlık, değerlerinin büyüklükleri çok farklı iki değişkeni aynı eksende göstermektir. Çizim 7.16 (a) bu konuda iyi bir örnektir.



ÇİZİM 7.16 (a) Yanlış bir çizim örneği

İTO, 2003b:94.

Bu çizimde gayrisafi ulusal ürün (GSUÜ) hem cari, hem sabit fiyatlarla aynı eksende yer almaktadır. Dolayısıyla sabit fiyatlarla GSUÜ'nün artmış mı, azalmış mı olduğu çizimden anlaşılamamaktadır. Daha önce eleştirdiğimiz arka duvar, koyu yer, üç boyutlu gösterim gibi özelliklerin hepsi bu çizimde de vardır. Ayrıca yatay eksenin hangi yıllar olduğu da belli değildir. Bu iki değişken ayrı eksenlerde gösterilseydi çok daha anlaşılır olabilirdi. Çizim 7.16 (b)'de bu yapılmak istenmiştir.



ÇİZİM 7.16 (b) Çizim 7.16 (a)'nın düzeltilmiş biçimi

Görüldüğü gibi 1987 fiyatlarıyla GSUÜ her iki yılda da yükselmiştir. Değerler yazılmış olmasa bunu ilk çizimden anlamak olanaksızdır. Hiçbir şey söylemeyen bir çizim yapmak için harcanan emeğe, zamana, mürekkebe, kâğıda yazık değil mi?

Okuma Önerileri

- Chambers, J. M., Cleveland, W. S., Kleiner, B., Tukey, P. A. (1983) *Graphical Methods for Data Analysis*, Wadsworth & Brookes.
- du Toit, S. H. C., Steyn, A. G. W., Stumpf, R. H. (1986) *Graphical Exploratory Data Analysis*, Springer-Verlag.
- Cleveland, W. S. (1985) *The Elements of Graphing Data*, Wadsworth.
- Ehrenberg, A.S.C.(1975) *Data Reduction: Analysing and Interpreting Statistical Data*, Wiley.
- Ehrenberg, A.S.C. (1982) *A Primer in Data Reduction*, Wiley.
- Groebnèr, D. F., Shannon, P. W., Fry, P. C., Smith, K. D. (2001) *Business Statistics: A Decision Making Approach*, 5. b., Prentice Hall (37-72).
- Gürtan, K. (1987) *İstatistik ve Araştırma Metodları*, 4. b., İstanbul Üniversitesi (145-207).
- Huff, D. (1973) *How to Lie with Statistics*, Penguin (58-71, 94-109) (Çeviren: Ko-paran, E. (1995) *İstatistik ile Nasıl Yalan Söylenir?*, Sarmal) (61-74, 97-111).
- Jaffe, A. J., Spirer, H. F. (1987) *Misused Statistics*, M. Dekker (64-84).
- McNeill, D. R. (1977) *Interactive Data Analysis*, Wiley.
- Tufte, J. W. (1983) *The Visual Display of Quantittative Information*, Graphics Press.
- Tufte, J. W. (1990) *Envisioning Information*, Graphics Press.
- Tufte, J. W. (1997) *Visual Explanations*, Graphics Press.

ALİŞTIRMALAR

- 7.1 Çeşitli yayınlardan seçtiğiniz en az beş değişik çizelgeyi inceleyerek işlevlerini yerine getirme başarılarını irdelleyiniz.
- 7.2 2002 yılında Türkiye'deki devlet üniversitelerinde lisans düzeyinde iktisat programlarına öğrenci alan bölümlere ilişkin bazı bilgiler şöyledir. Bu bölümlerden 46'sında 1. öğretimde (gündüz), 35'inde 2. öğretimde (gece) Türkçe, 8'inde 1. öğretimde yabancı dilde öğretim yapılmaktadır. 2002'de 14 vakıf üniversitesi iktisat programına hem burslu, hem burssuz öğrenci almıştır. Ayrıca bir vakıf üniversitesinin iktisat öğrencilerinin hepsi bursuzdur. Bu bölümlere ilişkin en düşük, en yüksek giriş puanları, kontenjanlar, kaydolan öğrenci sayıları bulunmaktadır (ÖSYM, 2002).
- a) Bütün bu verileri bir ya da birden çok çizelgede nasıl sunardınız? Yalnız çizelge ad(lar)ını, sütun başlıklarını, satır adlarını yazarak gösteriniz.
- b) Bu düzenlemeleri yaparken amacınızın ne olduğunu belirtiniz.
- 7.3 Aşağıdaki iki çizelgeyi karşılaştırınız.

ÇİZELGE A. Geçen yılın üçer aylık bölgesel satışları (bin YTL)

Dönem	Marmara	Ege	Akdeniz	İç Anadolu	Toplam
I	97 876	75 042	50 406	48 129	271 453
II	91 785	74 969	56 804	42 456	266 014
III	101 302	99 912	80 235	50 003	331 452
IV	90 482	75 423	51 347	38 658	255 910
Toplam	381 445	325 346	179 246	238 792	1 124 829

Ehrenberg (1975:29)'dan esinlenerek düzenlenmiştir.

ÇİZELGE B. Geçen yılın üçer aylık bölgesel satışları (milyon YTL)

Dönem	Marmara	Ege	Akdeniz	İç Anadolu	Ortalama
I	98	75	50	48	68
II	92	75	57	42	67
III	101	(100)	(80)	50	(83)
IV	90	75	51	39	64
Ortalama	95	75*	53*	45	67*

* Bu ortalamalara Temmuz-Eylül satışları katılmamıştır.

- Hangisi daha kolay anlaşılabilir?
- Hangisi akılda kolay kalabilir?
- Hangisi daha ayrıntılı bilgi verir?
- Satışları genel olarak izlemek isteyen üst düzey bir yöneticiye bilgi sunmakla görevli olsanız hangi çizelgeyi kullanırdınız? Neden?
- Çizelge B’de “Toplam” yerine “Ortalama” kullanılması çizelgenin ilet-tiği bilgilerin anlaşılmasına katkı yapıyor mu? Gerekçenizi açıklayınız?
- Çizelge boyutu küçültülerek sayıların birbirine yaklaştırılması karşılaştırmaları kolaylaştırıyor mu?

7.4 Aşağıdaki çizelge önde gelen ülkelerde kadınların siyasette nasıl temsil edildiğini göstermektedir.

Kadınların siyasete katıldığı önde gelen ülkelerde temsil oranları, 1994 (%)

Ülkeler	Parlamento	Belediye	Hükümet
<i>Sanayileşmiş</i>			
Danimarka	33	28	29
Finlandiya	39	30	39
Hollanda	29	22	51
Norveç	39	28	35
İsveç	34	34	30
OECD	13	16	15
<i>Gelişmekte olan</i>			
Küba	23	14	4
Guyana	20	22	11
Jamayka	12	13	5
Nikaragua	16	13	10
Trinidad-Tobago	18	21	19
Bütün gelişmekte olan	10	...	5
Dünya	10	...	0

Bu veriler parlamento için Haziran 1994, belediye için 1990-94 arası, hükümet için Mayıs 1994'teki durumu yansıtmaktadır. UNDP (1995:42)

- Bu çizelgenin hazırlanış nedeni sizce ne olabilir?
- Bu amacına ulaşmakta sizce başarılı mıdır? Neden?
- % değerlerinin basamak sayısının artırılmasını ister miydiniz? Neden?
- Bu çizelgenin altındaki açıklama sizce gerekli mi? Neden?
- Bazı satırların koyu, bazılarının yatık yazılması sizce çizelgenin anlaşılmasına katkı sağlıyor mu? Nasıl?
- Bu çizelgenin iletmek istediği bilgi sizce nedir? Açıklayınız.

7.5 Bir sınavda 23 öğrencinin 100 üzerinden aldığı notlar aşağıdadır. Bu notları bir çizelgede özetlemek gerekse:

- a) Hangi öbek aralıklarını kullanırdınız? Neden?
 b) Başka öbek aralıkları kullanarak çizelgeyi yeniden düzenleyiniz. Hangisi sınıfın not dağılımının en kolay biçimde anlaşılmasını sağlıyor? Nedenlerinizi açıklayınız.

Öğrenci	1	2	3	4	5	6	7	8	9	10	11	12
Not	5	40	90	45	50	55	70	30	30	65	50	35

Öğrenci	13	14	15	16	17	18	19	20	21	22	23
Not	70	75	75	55	80	80	20	75	90	70	65

7.6 Bir gazeteden aynen alınmış aşağıdaki çizelgeyi inceleyiniz. Bu çizelgede 1992-94 yıllarında başlıca ülkelerle Türkiye'nin buğday üretimine ilişkin veriler gösterilmiştir.

Ülke	Ekim alanı (bin ha)			Verimlilik (kg/ha)			Üretim (bin ton)		
	1992	1993	1994	1992	1993	1994	1992	1993	1994
Çin	30497	30327	30501	3331	3519	3318	101591	106395	101205
ABD	25399	25379	24998	2635	2569	2526	66222	65210	63141
Hindistan	23262	24431	24437	2364	2323	2420	55690	56762	59131
Rusya	24284	24666	22146	1901	1764	1449	46167	43500	32094
Fransa	5082	4515	4591	6404	6479	6676	32546	29252	30652
Kanada	11366	13830	12377	2635	2569	2526	29871	27232	23350
Türkiye	9473	9716	9600	2039	2163	1823	19318	21016	17500
Dünya	221682	222497	215921	2548	2535	2445	564837	564065	527982

Usumi, S. "Buğday krizinde vurgun hesabı", *Cumhuriyet*, 15 Mayıs 1996.

- a) Sizce bu çizelgeyle verilmeye çalışılan ileti(ler) nedir?
 b) Bu ileti ya da iletileri daha etkin bir biçimde sunabilmek için bu çizelgede ne gibi değişiklikler önerirdiniz? Neden?
 c) Bu verilerle birden çok çizelge düzenlemeyi düşünür müsünüz? Neden?
 d) Önerdiğiniz değişiklikleri uygulayınız. Öngördüğünüz etkinliği ne ölçüde sağlayabildiniz? Değerlendiriniz.
 e) Bu çizelgenin içinde yer aldığı yazıda dünya buğday üretiminde 1994'te bir düşüş başladığı ileri sürülmektedir. Düzenlemeniz bu savı iyice belirginleştiriyor mu?

7.7 İTÜ İşletme Fakültesi İşletme Mühendisliği programını 1996-97 yaz yarısında bitiren 96 öğrencinin notları aşağıdadır. Bu verileri uygun bir çizelge üzerinde özetleyiniz:

3,71	3,46	3,31	3,16	3,10	3,07	2,92	2,85	2,78	2,63	2,47	2,32
3,65	3,44	3,28	3,15	3,10	3,07	2,90	2,84	2,76	2,61	2,46	2,32
3,61	3,43	3,28	3,13	3,10	3,03	2,90	2,83	2,75	2,58	2,41	2,31
3,55	3,42	3,26	3,13	3,10	3,00	2,89	2,83	2,73	2,55	2,39	2,30
3,52	3,40	3,25	3,12	3,08	2,97	2,86	2,82	2,72	2,55	2,36	2,27
3,52	3,38	3,23	3,12	3,08	2,97	2,86	2,80	2,68	2,54	2,35	2,22
3,50	3,33	3,22	3,11	3,08	2,95	2,86	2,79	2,66	2,50	2,34	2,20
3,47	3,32	3,22	3,11	3,07	2,95	2,86	2,78	2,64	2,50	2,33	2,10

7.8 Aşağıdaki çizelge son beş yılda borçlulara göre dış borç stokumuzu göstermektedir.

	1998	1999	2000	2001	2002*
DIŞ BORÇ STOKU	96408	102980	119692	115186	127535
KISA VADELİ	20774	22921	28301	16241	14129
TCMB	905	686	653	590	470
Hükümet	0	0	1000	0	0
Ticari bankalar	11159	13172	16900	7097	5706
Diğer sektörler	8710	9063	9748	7654	5706
ORTA-UZUN VADELİ	75634	80059	91397	98945	113406
Kamu sektörü	39881	42381	48809	46277	61684
TCMB	12073	10312	13429	23753	20384
Özel sektör	23680	27366	29153	28915	31338

* Geçici (Eylül 2002).TCMB (2003:267)'den özetlenmiştir.

- Bu çizelge ne anlatıyor? Bunu algılamakta güçlük çekiyor muyuz?
- Bu çizelgeyi daha yalın hale getirerek iletinin daha kolay anlaşılmasını sağlayabilir misiniz? Nasıl?

7.9 Aşağıdaki çizelge bir gazetedeki köşe yazısından alınmıştır. Bu yazıda Grup 7 ülkeleri (ABD, Japonya, Almanya, İngiltere, Fransa, Kanada, İtalya) ile yeni yükselen 7 ekonominin (Türkiye, Rusya, Çin, Arjantin, Brezilya, Hindistan, Endonezya) dünya toplamı içindeki payları tartışılmaktadır.

- Çizelgede “Dünya payı” başlıklı iki sütun var. Yazarın bununla ne demek istediği anlaşılıyor mu?
- Bu iki sütunun başlığında “%” işaretinin yer almasını ister miydiniz? Neden?

- c) Dünya ihracatı, dünya ithalatına eşit değil midir? Çizelgedeki farkı nasıl açıklayabiliriz?
- d) Son satırdaki sayılarda herhangi bir yanlışlık göze çarpıyor mu? Bunu nasıl açıklayabilirsiniz?
- e) Vermek istediği iletiyi daha iyi yansıtacak biçimde bu çizelgeyi istediğiniz gibi yeniden düzenleyiniz.

	Dünya	Grup 7	Dünya payı	7 Yeniler	Dünya payı
Toprak (bin m ²)	148.940	20.226	14	43086	29
Nüfus (mn)	6.234	700	11	2988	48
GSYİH (PPP mr \$)	47.000	20.979	45	12183	26
GSYİH yıllık büyüme (%)	2,2	1,5		3,3	
Kişi başına GSYİH (\$)	7.600	29.974		4.078	
İşsizlik (%)	30	7,2		10,4	
İhracat (mr \$)	6.300	2.786	44	586	9
İthalat (mr \$)	2.000	970	48	1.095	55

Eğilmez, M. "Bir ekonomik karşılaştırma", *Radikal* Gazetesi, 23 Mart 2003.

7.10 "2002'de kişi başına sadece 45 gram pastırma tüketildi" başlıklı bir haberde şu çizelge yer almaktadır.

Yıllar	Sucuk	Sosis	Salam	Pastırma
1993	10.955	5.338	6.827	720
1994	9.960	4.952	5.963	632
1995	11.237	5.917	5.403	642
1996	11.391	6.130	7.474	567
1997	11.589	7.882	7.116	601
1998	12.883	8.197	7.717	912
1999	14.251	8.713	8.849	988
2000	14.867	9.517	9.581	813
2001	10.536	8.556	7.162	384
2002	9.890	8.352	7.242	321

Sabah, 1 Mart 2003.

- a) Bu çizelgede ölçü birimi ton mu, TL mi, başka bir şey mi?
- b) Sayılar üretimi mi, tüketimi mi gösteriyor?
- c) Çizelgedeki basamak sayısını azaltmak işe yarar mı?
- d) Basamak sayısı azaltılmış bir çizelge düzenleyiniz.

- e) Satır ya da sütunlara toplam ya da ortalama bilgisi eklemeyi uygun bulur musunuz? Neden?
- f) Her et ürününün zaman içindeki gelişmesini kendi çizdiğiniz çizelgeden izleyerek açıklayınız.
- g) Çeşitli et ürünlerinin tüketimlerini kolayca karşılaştırabiliyor musunuz? Bunun için sütunlara ortanca eklemek işe yarar mı?

7.11 Arabaların ikinci el fiyatlarını gösteren aşağıdaki çizelgeyi yeniden düzenleyip fiyatların karşılaştırılmasını kolaylaştırınız. (PALIO 1.4 EL'nin 2002 fiyatlarında yanlışlık olabilir mi?)

AUDI		DAEWOO		PALIO 1.4 EL	
A4 1.8T		MATIS 0.8 SE		2002	14-46 milyar
2002	32-34 milyar	2002	9-11 milyar	2001	11-13 milyar
2001	28-31 milyar	2001	8-9 milyar	2000	10-11 milyar
2000	25-27 milyar	2000	7-8 milyar		
CITROEN		NUBIRA		PALIO 1.616V HL	
SAXO-2 1.4I SK SK		2002	7-8 milyar	2002	12-14 milyar
2002	13-15 milyar	2001	6-7 milyar	2001	11-12 milyar
2001	11-13 milyar	2000	5-6,5 milyar	2000	10-11 milyar
2000	9-11 milyar				
FIAT		TIPO 1.4 S		SIENA 1.2 S	
XSARA 1.6I CLIM		1999	9,5-10 milyar	2002	14-16 milyar
2002	15-16 milyar	1998	8-9 milyar	2001	11-13 milyar
2001	13-14 milyar	1997	7-8 milyar	2000	10.5-11.5 milyar
1999	11-12 milyar				
XANTIA 2.0 SK		TIPO 1.4 IE SX		SIENA 1.4 EL	
2001	26-28 milyar	1999	12-13 milyar	2002	12-14 milyar
2000	19-22 milyar	1998	10-11 milyar	2001	11-12 milyar
1999	17-19 milyar	1997	8-10 milyar	2000	9-11 milyar

Cumhuriyet, 30 Mart 2003.

7.12 Çizelge 7.12'yi yorumlarken “bu çizelgeden hemen görüldüğü gibi 1996 yılında dışarıya bağımlılığı yüksek kesimler genellikle ağır sanayi kesimleri- dir. Birincil üretim kesimleriyle hizmet kesimleri bu anlamda bağımlılığı en düşük kesimlerdir” demiştik. Benzer çizelgeleri öteki dönemler için de çizerek bu yapının zaman içinde değişip değişmediğini, değiştiyse bu değişikliğin ne olduğunu açıklayınız. (Gerekli veriler Çizelge 7.8'dedir.)

7.13 ABD'nin Irak'a saldırmasından önce Avrupa halklarının savaşa karşı olma oranları aşağıdaki gibiydi.

Ülke	%	Ülke	%	Ülke	%	Ülke	%
Türkiye	94	İsviçre	83	Almanya	71	Slovakya	57
İspanya	91	Macaristan	82	Romanya	70	Danimarka	53
İngiltere	90	Bulgaristan	81	İrlanda	68	İsveç	50
Norveç	90	Hollanda	80	Portekiz	65	Avusturya	49
Yunanistan	88	Fransa	77	Çek C.	65		
Belçika	84	Polonya	72	İtalya	61		

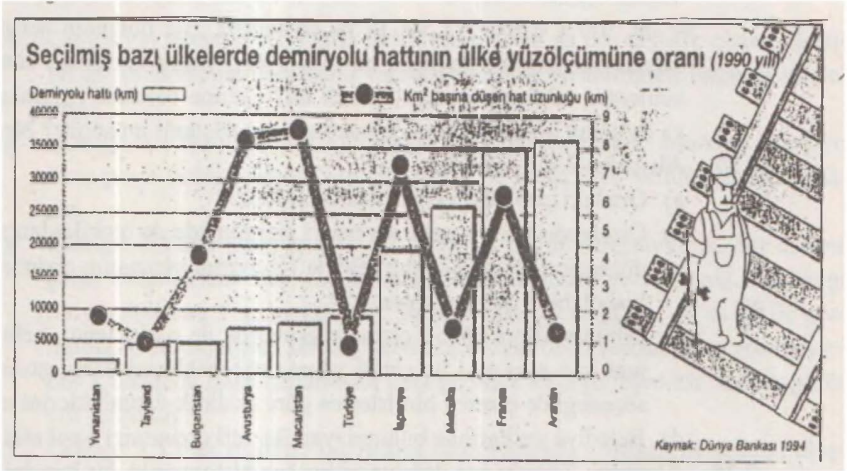
Le Monde, 14 Şubat 2003, aktaran G. Uras, "Fransızın 'hayır'ı bizden çok farklı", *Milliyet*, 18 Şubat 2003.

- a) Bu çizelgeyi yorumlayınız.
 - b) Çizelgede herhangi bir yalınlaştırmaya gidebilir misiniz? Neden gide-
mezsiniz ya da nasıl gidersiniz?
- 7.14** Çeşitli yayınlardan seçtiğiniz farklı türlerde en az beş değişik çizimi eleştirel gözle inceleyip işlevlerini yerine getirmede gösterdikleri başarıyı irdeleyiniz..
- 7.15** Çizelge 7.1'deki verileri kullanarak kadınlar için ayrı, erkekler için ayrı birer pasta çizimi yapınız. Bu çizimlerin ne kadar "iyi" olduklarını irdeleyiniz.
- 7.16** Çizelge 7.2'den yararlanarak bir serpilme çizimi düzenleyiniz. Bir eksene gelir, öbür eksene tüketim harcamaları düzeylerini yerleştiriniz. Herhangi bir gözde sıklık 5'se bunu $\curvearrowright$ ile, 10'sa $\curvearrowleft$ ile, 25'se $\times$ ile, 40'sa $\boxtimes$ ile gösteriniz. Gördüklerinizi iki değişken arasındaki ilişki bakımından yorumlayınız. Çizelgede göze çarpmayan ama çizimde belirginleşen bir özellik görüyor musunuz?
- 7.17** Çizelge 7.3'teki verilerle dış ticaret açığı (dışalım – dışsatım) serisini oluşturup çiziniz. Gördüklerinizi yorumlayınız.
- 7.18** Çizelge 7.5'teki veriler için bir pasta çizimi, bir de çubuk çizimi düzenleyiniz. Bunları karşılaştırıp ne kadar "iyi" çizimler olduğunu tartışınız.
- 7.19** Çizelge 7.5'teki verilerle her iki yıl için ayrı ayrı birer histogram, birer dal-yaprak, birer nokta çizimi düzenleyiniz. Bunları karşılaştırınız.
- 7.20** Alıştırma 7.19'da çizdiğiniz gösterimleri Çizelge 7.7 ile karşılaştırınız. Verilmek istenen ileti çizelgede mi, yoksa çizimlerde mi daha kolay algılanıyor. Tartışınız.
- 7.21** Çizelge 7.8'deki verilerle her yıl için ayrı ayrı birer histogram, birer dal-yaprak, birer nokta çizimi düzenleyiniz. Bunları karşılaştırınız.

- 7.22** Alıştırma 7.14'te çizdiğiniz gösterimleri Çizelge 7.11 ile karşılaştırınız. Verilmek istenen ileti çizelgede mi, yoksa çizimlerde mi daha kolay algılanıyor? Tartışınız.
- 7.23** Yatay eksene yılları, dikey eksene dışılma bağımlılık katsayılarının düzeylerini koyarak Çizelge 7.10'daki verilerle bir serpilme çizimi düzenleyiniz. Bir gözdeki sıklık 1 – 2 ise • ile, 3 – 4 ise ile, 5 – 6 ise ile, 7 – 8 ise ile, 9 – 10 ise× ile, 11 – 12 ise× ile, 13 – 14 ise× ile gösteriniz. Gördüklerinizi iki değişken arasındaki ilişki bakımından yorumlayınız.
- 7.24** Çizelge 7.13'teki verileri uygun bir çizimde gösteriniz. Çizelge mi, yoksa çizim mi daha kolay anlaşılıyor? Tartışınız.
- 7.25** Çizim 7.4 (a, b, c, d, e, f), 7.5'teki veri yoğunluklarını bulunuz. Çizim 7.2, 7.3'tekilerle karşılaştırınız.
- 7.26** Alıştırma 7.3'e dönelim. Bir eksene bölgeleri, öbür eksene üçer ayları koyarak Çizelge B'deki verilerle bir serpilme çizimi düzenleyiniz. Bir gözdeki satışlar < 40 trilyon TL ise • ile, 40 – < 50 trilyon TL ise | ile, 50 – < 60 trilyon TL ise ∩ ile, 70 – < 80 trilyon TL ise + ile, 90 – < 100 trilyon TL ise × ile, 100 + trilyon TL ise * ile gösteriniz. Her bölgenin her mevsimdeki satışları benzer bir değişme mi gösteriyor, yoksa her bölge satışının mevsimlerden etkilenmesi farklı farklı mı? Çizime bakarak yorumlayınız. Satır ya da sütun ortalamalarını çizime katmayı düşünür müsünüz? Neden?
- 7.27** Alıştırma 7.4'e dönelim.
- Oradaki çizelgeyi çizime dönüştürünüz.
 - Çizelgede koyu yazılmış verileri çiziminizde de ötekilerden ayıracak bir düzenek düşündünüz mü? Böyle bir ayırım yapmanın çiziminize bir katkısı olabilir mi? Açıklayınız.
 - Bütün verileri tek bir çizimde gösterme ile parlamento, belediye, hükümet için ayrı ayrı üç çizim yapma seçeneklerini karşılaştırınız. Her iki seçeneği de çizerek birbirlerine göre eksiklik-üstünlüklerini tartışınız.
 - Belediye verilerinde bulunmayan iki veri çiziminizi nasıl etkiledi?
- 7.28** Alıştırma 7.5'teki not dağılımını bir kez histogramla, bir kez dal-yaprak, bir kez de nokta çizimiyle gösteriniz. Üç çizimi karşılaştırıp birbirlerine göre eksik ya da üstün yanlarını tartışınız.
- 7.29** Alıştırma 7.6'daki verilerle uygun çizimler düzenleyiniz. Bunlar için pasta çizimini mi, yoksa başka bir çizimi mi yeğlersiniz? Neden? Çizimlerinizin çizelgeye göre üstün yanları var mı? Açıklayınız.
- 7.30** Alıştırma 7.7'deki bitirme notlarını bir kez histogramla, bir kez dal-yaprak, bir kez de nokta çizimiyle gösteriniz. Üç çizimi karşılaştırıp birbirlerine göre eksik ya da üstün yanlarını tartışınız.
- 7.31** Alıştırma 7.8'de gösterilen bütün verileri tek bir çizimde göstermek gerekirse nasıl bir çizim düşünürsünüz? Birden çok çizim daha uygundur diyorsa-

nız bu çizimleri düzenleyiniz. İletinin alınması açısından en uygun yol hangisidir? Tartışınız.

- 7.32 Alıştırma 7.9'daki verileri tek bir çizimde nasıl düzenlerdiniz? Bununla çizelgeyi karşılaştırınız, hangisi daha etkin? Tartışınız.
- 7.33 Alıştırma 7.10'daki verileri bir zaman çiziminde göstererek gelişmelerini, birbirlerine göre konumlarını, değişme hızlarını karşılaştırınız. Ne gördüğünüzü yazınız.
- 7.34 Alıştırma 7.11'deki araba fiyatları verileri için düşük-yüksek çizimini kullanmayı uygun bulur musunuz? Düzenlediğiniz çizime bakarak yorumunuzu yazınız.
- 7.35 Alıştırma 7.13'teki çeşitli ulusların savaşa karşı olma oranlarını uygun bir çizimle gösteriniz. Gördüklerinizi yorumlayınız.
- 7.36 Çizim 7.10 ile Çizim 7.2 (b)'yi karşılaştırarak yorumlayınız.
- 7.37 Çizim 7.11 (a,b) ile Çizim 7.2 (c,d,e,f)'yi karşılaştırınız.
- 7.38 Bir gazeteden alınmış olan çeşitli ülkeler demiryollarına ilişkin aşağıdaki çizime bakınız.



- a) Bu çizimi çeşitli bakımlardan eleştiriniz.
- b) Daha uygun bir ya da birden çok çizim düzenleyiniz.
- 7.39 Ülkemizde otomobil üretimi 1986-90 arasında aşağıdaki gibidir.
- a) Bunu uygun bir çizimle gösterip yorumlayınız.
- b) 1986 yılını 100 alınız. Öbüryılların verilerini 82'ye bölüp 100'le çarparak bir üretim indeksi oluşturunuz. Bu indeks değerlerini çizimle gösteriniz.
- c) Yıllık artış oranlarını hesaplayıp bir çizimle gösteriniz? Hangi çizim hangi bilgileri daha iyi iletebilmektedir? Tartışınız.

Üretim		Üretim		Üretim	
Yıl	(bin tane)	Yıl	(bin tane)	Yıl	(bin tane)
1986	82	1991	196	1996	196
1987	107	1992	256	1997	236
1988	112	1993	343	1998	224
1989	118	1994	209	1999	221
1990	166	1995	222	2000	302

DİE, 2001:309; DİE 2002a:349.

7.40 Aşağıda rassal olarak seçilmiş 18 ülkenin savunma harcamalarıyla sağlık harcamalarının, merkezi hükümetin toplam harcamaları içindeki payı verilmiştir (1991).

Harcamalar (%)			Harcamalar (%)		
Ülkeler	Savunma	Sağlık	Ülkeler	Savunma	Sağlık
Avustralya	8,5	12,8	Namibya	5,5	11,1
Botswana	11,6	4,8	Nepal	6,0	4,8
Ekvator	12,9	11,0	Norveç	8,0	10,4
Filipinler	11,0	4,1	Papua-Yeni Gine	4,7	9,4
Finlandiya	4,6	10,8	Paraguay	13,3	4,3
Fransa	6,7	15,2	Sri Lanka	7,4	5,4
Gana	3,2	9,0	Türkiye	11,7	3,6
G. Kore	25,8	2,2	Uruguay	9,2	4,5
Kenya	7,8	5,4	Zaire	6,7	0,7

World Bank (1992:238-39)

- Birer dal-yaprak gösterimi düzenleyerek bu iki harcamanın bu ülkelerdeki konumunu inceleyiniz. Bulgularınızı açıklayınız.
- Bir eksene savunma, öbür eksene sağlık harcamalarının toplam harcamalar içindeki payını koyarak aynı verileri bir serpilme çiziminde gösteriniz. Harcamalar arasında bir ilişki var mı?
- Savunma harcamaları = sağlık harcamaları eşitliğinin geometrik yerini gösteren 45° açılı bir doğru çiziniz. Çizimde bir ülkenin konumunu gösteren noktanın bu doğrunun üstünde ya da altında olması ne anlama gelir?
- Çiziminizi yorumlayınız.
- Aynı verilerin daha yenilerini bulup benzeri işlemleri yapınız. Eski verilerle karşılaştırınız.

- 7.41** Çizelge 7.10'daki dışılıma bağımlılık katsayısını yatay eksene koyarak oradaki beş yıldan istediğiniz biri için histogram çiziniz. Aynı yıl için Çizelge 7.11'deki verilerle yeni bir histogram çiziniz. İkisi arasında bir çelişki görülüyor mu? Eğer bir çelişki varsa ikinci histogramınızı yanlış çizmişsiniz demektir, düzeltiniz. (İpucu: Histogramlardaki dikdörtgenlerin alanları sıklık sayılarını gösterir. Yatay eksen e eşit olmayan aralıklar varsa bu sıklıklar alana nasıl çevirirsiniz?)
- 7.42** Eurostat verilerine göre Avrupa Birliği'ne üye ülkelerde 1990 yılında evlenme oranlarıyla boşanma oranları aşağıda verilmiştir.

Ülkeler	Boşanma oranı %	Evlenme oranı %
Almanya	2,2	6,5
Belçika	2,0	6,5
Danimarka	2,7	6,1
Fransa	1,9	5,1
Hollanda	1,9	6,4
İngiltere	2,9	6,8
İrlanda	0,0	5,0
İspanya	0,6	5,5
İtalya	0,4	5,4
Lüksemburg	2,3	6,1
Portekiz	0,9	6,9
Yunanistan	0,6	5,4

- a) Bunları birer dal-yaprak çiziminde gösterip bu çizimi yorumlayınız.
- b) Ötekilere aykırı konumda ülkeler varsa bunun nedenleri üzerinde düşününüz.
- 7.43** Bir ürünün pazar payı ile o ürüne yapılan reklam harcamalarının benzer ürünlere yapılan reklam harcamaları içindeki payı son on yılda şöyle olmuştur.
- a) Bu verilerle bir serpilme çizimi düzenleyiniz.
- b) Bu iki değişken arasında bir ilişki var mı? Sizce bu ilişki doğrusal mı, değil mi?
- c) Serpilme çizimindeki noktaları temsil ettiğini düşündüğünüz bir doğruyu göz kararıyla çiziniz. Noktalarla çizdiğiniz doğru arasındaki dikey uzaklıkların büyüklük derecesi doğrunuzun gözlemleri ne kadar iyi temsil ettiğini gösterir. Sizce doğrunuz iki değişken arasındaki ilişkiyi iyi temsil ediyor mu? Tartışınız.

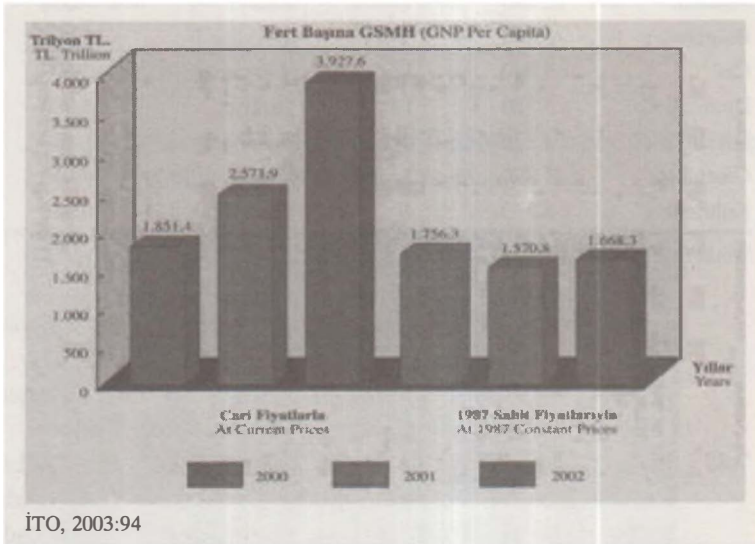
Pazar payı (%)	13,5	12,5	12,0	8,6	7,9	7,6	6,5	5,9	5,1	3,9
Reklam payı (%)	10,3	7,2	7,1	7,7	8,6	4,7	6,0	5,0	7,0	5,3

7.44 Belli bir model binek arabasından rassal olarak seçilen 8 tanesinin yaşı ile ikinci el fiyatı aşağıdadır.

- Bir serpilme çizimiyle arabaların yaşının fiyatı üzerinde nasıl bir etki yaptığını gösteriniz.
- Çizimdeki noktalar arasından göz kararıyla bir doğru geçirip bunun temsil niteliğini tartışınız.

Yaş (yıl)	1	4	10	7	5	6	7	1
Fiyat (mr TL)	18	10	3	13	8	10	9	17

7.45 Aşağıdaki çizimde gördüğünüz yanlışları düzeltiniz.



7.46 DİE'nin İstanbul tüketici fiyatları indeksi şöyledir:

Yıl	İndeks	Yıl	İndeks	Yıl	İndeks	Yıl	İndeks	Yıl	İndeks
1968	1,00	1975	2,38	1982	30,68	1989	464,01	1996	24761,03
1969	1,08	1976	2,79	1983	39,62	1990	739,11	1997	46424,70
1970	1,15	1977	3,54	1984	59,68	1991	1235,27	1998	86161,72
1971	1,33	1978	5,14	1985	88,52	1992	2054,86	1999	138080,04
1972	1,49	1979	8,17	1986	119,38	1993	3394,01	2000	213134,97
1973	1,72	1980	17,17	1987	168,36	1994	7014,48	2001	328431,98
1974	1,99	1981	23,45	1988	292,00	1995	13704,52		

DİE, 2001:484; son üç yıl DİE 2002a:557'den hesaplanmıştır.

- a) Bu verileri Çizim 7.14 (b) ya da (c)'deki gibi bir zaman serisi çizimiyle gösteriniz.
- b) Çiziminizi Çizim 7.4 (b) ya da (c) ile karşılaştırınız. DİE ile İTO verileri enflasyonu ölçmekte uyuyorlar mı? Tartışınız.

7.47 Rassal seçilmiş bazı ülkelerde 2000 yılında (satınalma gücü eşdeğeri olarak) kişi başına ulusal gelir (KBUG) ile yetişkinler arası okuma-yazma bilmeyen oranı aşağıdadır.

Ülkeler	Okuma-yazma KBUG bilmeyen yetişkin-		Ülkeler	Okuma-yazma KBUG bilmeyen yetiş-	
	(\$)	ler %		(\$)	kinler %
Arjantin	12.090	3	Malavi	600	41
Bulgaristan	5.530	2	Meksika	8.810	9
Çin	3.940	17	Nikaragua	2.100	32
Dominik C.	5.720	17	Türkiye	7.050	15
Gambiya	1.540	64	Uganda	1.230	34
Guatemala	3.770	32	Ürdün	4.040	11
Kolombiya	5.890	9	Yunanistan	16.940	3

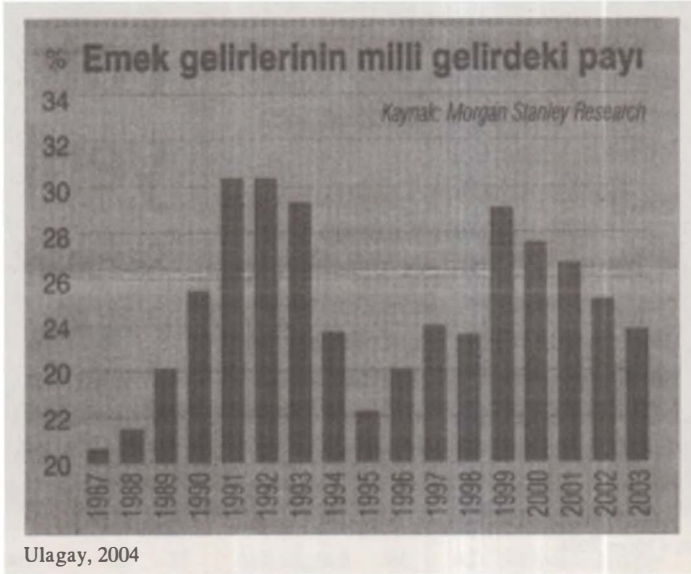
World Bank, 2001:232-233.

- a) Bu verileri bir serpilme çiziminde gösteriniz. İki değişken arasındaki ilişki sizce bir doğruyla iyi temsil edilebilir mi? Neden?
- b) KBUG değerlerinin logaritmalarını alıp serpilme çizimini yeniden düzenleyiniz. Şimdi doğrusal ilişkiden söz edilebilir mi?
- 7.48** Birleşmiş Milletler Kalkınma Programı (UNDP) dünya ülkelerini her yıl insani gelişmişlik düzeylerine göre sıralamaktadır. 2001 yılında kimi ülkelere bu sıra numaralarıyla kişi başına gelire (satınalma gücü eşdeğeri, \$) göre sıra numaraları aşağıdadır.
- a) Yatay eksene kişi başına gelir (KBG), dikey eksene insani gelişmişlik indeksi (İGİ) sıra numaralarını koyarak bir çizim düzenleyiniz. Sizce bu iki gösterge arasında bir ilişki var mı?
- b) Her iki sıra numarasının birbirine eşit olduğu noktaların geometrik yerini gösteren 45° açılı bir doğruyu bu çiziminize ekleyiniz. Bir ülkenin konumunun bu doğrunun altında ya da üstünde kalması ne anlama gelir? Türkiye'nin konumunu bu bakımdan inceleyiniz.

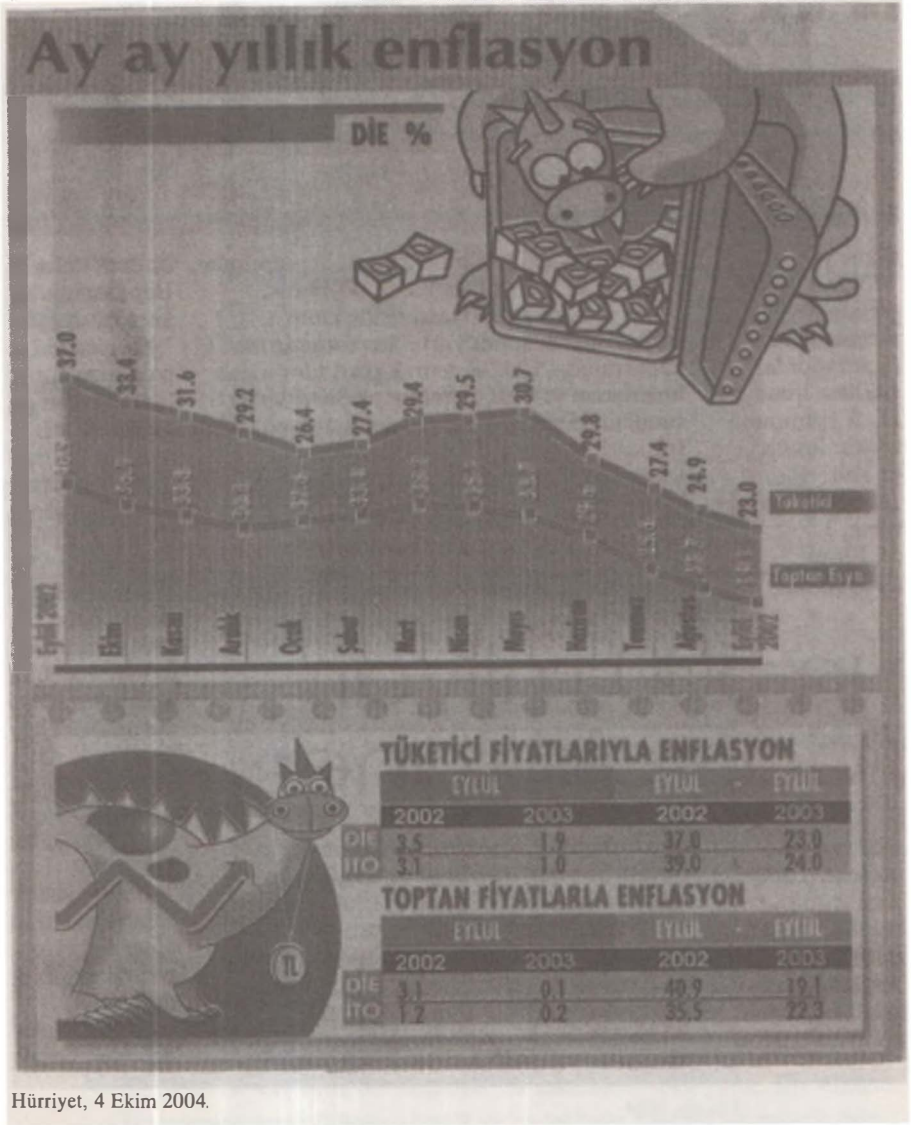
Ülke	İGİ	KBG	Ülke	İGİ	KBG
Norveç	1	5	Kuveyt	46	29
İzlanda	2	4	Hırvatistan	47	54
ABD	7	2	Küba	52	90
Japonya	9	14	Bulgaristan	57	69
İngiltere	13	20	Rusya	63	66
İtalya	21	16	Brezilya	65	64
Yunanistan	24	31	Romanya	72	81
Kıbrıs	25	22	S. Arabistan	73	40
G. Kore	30	37	Kazakistan	76	81
Çek Cum.	32	39	Lübnan	83	101
Macaristan	38	42	Arnavutluk	95	106
Şili	43	53	Türkiye	96	80

UNDP, 2003:237-238.

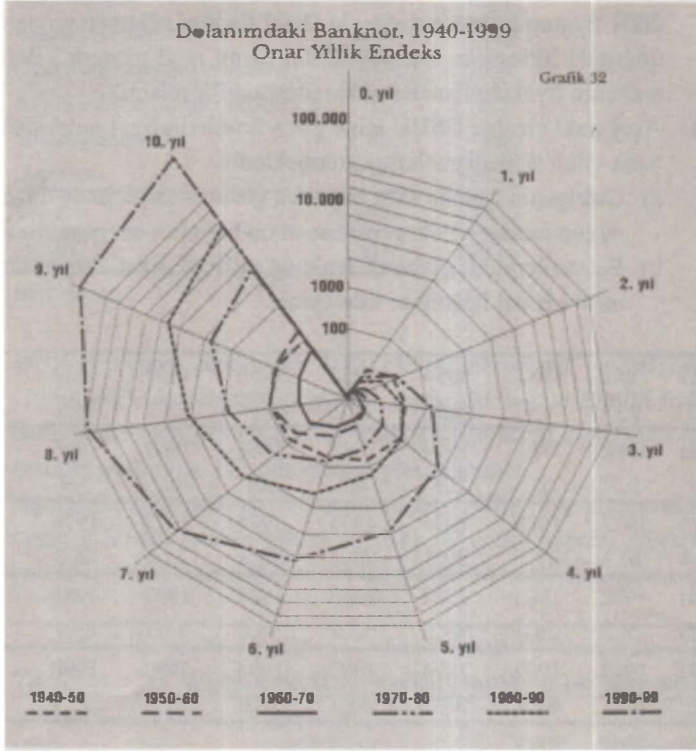
7.49 Aşağıdaki çizim bir köşe yazısından alınmıştır (Ulagay, 2004). Bu çizimdeki yanlışları bulup düzelterek çizimi yeniden yapınız.



7.50 Aşağıdaki çizimi eleştirerek bütün yanlışlarını düzeltiniz.



7.51 Aşağıdaki çizimden (Yenal, 2003: 140) anladığınızı açıklayınız.



7.52 2002-03 döneminde Türkiye süper futbol liginde takımların gol sayıları aşağıdadır. Bir eksene *atılan* (A), öbürüne *yenen* (Y) gol sayılarını koyarak bir çizim yapınız. Atılan gollerle yenilenlerin aynı sayıda olduğu noktaların geometrik yerini gösteren 45° açılı bir doğruyu bu çiziminize ekleyiniz. Bir takımın bu çizimin altında ya da üstünde yer alması ne anlama gelir?

Takım	A	Y	Takım	A	Y	Takım	A	Y
Beşiktaş	63	21	Trabzonspor	44	33	Elazığspor	40	59
Galatasaray	61	27	Ankaragücü	44	42	Diyarbakırspor	34	47
Gençlerbirliği	76	40	İstanbulspor	42	47	Bursaspor	42	61
Gaziantepspor	61	41	Denizlispor	37	42	Altay	48	69
Malatayaspor	56	45	Adanaspor	44	54	Göztepe	32	57
Fenerbahçe	55	42	Samsunspor	42	59	Kocaelispor	32	66

7.53 Çizelge 7.5'teki verilerle bir çizelge düzenleyiniz. Bu çizelgede dik eksenle araba fiyatları, yatay eksenle 2001, 2002 yılları yer alsın. Her bir arabanın 2001 fiyatını gösteren noktayla, 2002 fiyatını gösteren noktanın arasını birer doğruyla birleştirin. Bu doğruların eğimi neyi gösterir? Buna göre her bir arabanın fiyat değişimlerini karşılaştırabilir misiniz?

7.54 Aşağıdaki çizelge GSUÜ içine giren ürünlerin fiyat indeksinde 1950'den bu yana yıllık % değişimleri göstermektedir.

- Çizelgenin zamanı satır boyunca gösterecek biçimde düzenlenmesi sizce uygun mudur? Yoksa zaman sütun boyunca mı gösterilmeli? Neden?
- Bu verilerle bir çizim düzenleyip yıllık % fiyat artışlarının zaman içinde nasıl bir yol izlediğini irdeleyiniz.

1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
6,5	2,8	4,9	5,2	11,3	11,7	23,3	14,3	19,8	3,3
1961	1962	1963	1964	1965	1966	1967	1968	1969	1970
4,1	9,5	5,7	2,6	4,3	6,4	6,5	3,9	7,1	8,5
1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
17,4	10,2	21,1	30,5	21,2	15,3	24,0	46,7	75,6	89,6
1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
44,3	28,3	26,0	48,5	52,9	35,6	33,5	69,7	75,5	57,6
1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
59,2	63,5	67,4	107,3	87,2	78,0	81,2	75,3	55,8	50,9
2001	2002	2003*							
55,3	43,8	24,4*							

* Kestirim. DPT, 2004:3.

7.55 Aşağıdaki çizelgede GSYİÜ değerini oluşturan harcama kalemleri % paylarla verilmiştir.

- Bunları uygun birer çizimle gösteriniz.
- Bu verileri pasta çizimine dönüştürebilir misiniz? Size engel olan nedir? Açıklayınız.
- Bu veriler için pasta çizimi dışında önerdiğiniz çizimin üstün ya da eksik yanlarını tartışınız.
- On yıl içindeki en önemli değişme hangi kalemlerde? Bunun iktisadi anlamını tartışınız.

Harcama	1992	2002
Özel tüketim	67,3	66,4
Kamu tüketimi	12,4	14,0
Özel sabit sermaye oluşumu	15,7	11,0
Kamu sabit sermaye oluşumu	7,2	5,7
Sok değişimleri	0,2	4,6
Dışsatım	13,9	28,7
Dışalım	-16,8	-30,4
GSYİÜ	100,0	100,0

DPT, 2004:12.

- 7.56** DPT ağçevrimi adresinden DPT (2004) *Ekonomik ve Sosyal Göstergeler (1950-2003)* kaynağına girip, 16. sayfadaki sanayi üretim indeksinin hem özel, hem kamu kesimi için üçer aylık değerlerini alınız. Bunları uygun bir çizimle gösteriniz. Gördüklerinizi yorumlayınız.
- 7.57** Aşağıdaki çizelgede dış borçlara ilişkin veriler yer almaktadır. Uygun bir çizim üzerinde bunları gösteriniz. Tek bir çizim hepsini göstermek için uygun mu, yoksa birden çok çizim mi gerekli? Gerekçenizi açıklayınız.

Yıl	Kısa erimli dış borç (m \$)	Uzun erimli dış borç (m \$)	Cari işlem geliri içinde Dış borç ödemesi (%)
1950	373		11,9
1960	992		30,7
1970	1.929		25,9
1980	2.505	13.722	47,4
1985	4.759	20.717	30,7
1990	9.500	39.535	27,6
1995	15.701	57.577	28,0
2000	28.301	90.401	37,1
2002	15.190	116.010	47,0

DPT, 2004:47.

Aşağı Yukarı Ne Kadar?

*Ortalamaları herkes anlar,
ama kullanmaya gelince iş değişir.*
W. J. Reichman

8.1 ORTALAMA DÜZEY

Derlenmiş verilerin ne söylediklerinin iyice anlaşılması için bu verilerin düzenlenmesi, sergilenmesi önemli bir adımdır ama yetmez. Bunların belli başlı özelliklerinin ortaya konulması gerekir. Tek değişkenli her veri kümesinde ortak olan özellikler şöyle sıralanabilir:

- **Ortalama düzey.**
- **Yayılım, yayıklık, oynaklık, değişkenlik** derecesi, bir başka deyişle verilerin ortalama düzeyden uzaklaşma ölçüsü.
- **Verilerin dağılma biçimi.** Histogram, dal-yaprak, nokta çizimi bu biçimin resimleri gibi algılanabilir. Biçim birkaç bakımdan ele alınabilir.
 - a) Verilerin yoğunlaştıkları düzeyler, yani **tepe sayısı** tek mi, birden çok mu?
 - b) Dağılım **bakışık** mı, **çarpık** mı? Ortalamanın iki yanında benzer biçimde mi uzanıyor ?
 - c) Dağılım **basık** mı, **sivri** mi?
- **Veriler arasında olağandışı değerler** var mı?

Ayrıca birden çok değişkenin ya da veri kümesinin aralarındaki **ilişki** de incelemeye konu olabilir. Bu bölümde bu özelliklerden yalnızca ortalama düzeyi inceleyeceğiz. Öteki özellikler daha sonraki bölümlerde ele alınacaktır.

Bir veri kümesinin ortalama olarak hangi düzeylerde bulunduğunu gösteren değerlere genel bir adla **ortalama** diyoruz. İstatistikte çok çeşitli ortalama ölçüleri vardır, bunları daha sonra göreceğiz. Bu ortalamalarla günlük yaşamımızın hemen her alanında hem de sık sık karşılaşırız. Ortalamasız

bir tek günümüz bile geçmez desek yeridir. Radyoda ya da televizyonda hava durumunu izlerken yağışların ya da hava sıcaklığının son günlerde mevsim normallerinin (siz bunu ortalamaların diye anlayabilirsiniz) altına düştüğünü ya da üstüne çıktığını duyarız. Yazılı ya da görsel basından her ay başında fiyatların yüzde kaç arttığını, böylece aylık ya da yıllık ortalama artışın yüzde kaçta ulaştığını öğreniriz. Öğrenciysek dönem not ortalamamızı yükseltip yükseltmeyeceğimizi, okulu bitirmişsek genel not ortalamamızın iş bulmamızda nasıl bir etkisi olacağını kestirmeye çalışırız. İlgi alanımıza göre ya tuttuğumuz takımın maç başına gol ortalamasına, ya da kişi başına ortalama ulusal gelirimize bakıp üzülebiliriz. Türkiye’de her hanede ortalama 5,1 kişinin oturmakta (Esmer, vö., 1986:24) olduğunu görüp her hanedeki altıncı kişinin yalnızca yıllık izinlerini evinde geçirdiğini (!) düşünebiliriz. Sendikacıysak reel ortalama işçi ücretlerinin düşüşü, işverensek ya da bir işletmede yöneticiyse ortalama maliyetlerin yükselişi bizi endişelendirebilir. Satın aldığımız tuvalet kâğıdının, markasına bağlı olarak ya ortalama 120 gram ağırlığında ya da ortalama 33,32 metre uzunluğunda olduğunu öğrenip bilinçli (!) bir seçim yaparız. Tiryakiyse günde ortalama kaç sigara içtiğimizi, hekimsek bu kadar sigara içen birinin ortalama ömrünü biliriz. Sigortalı işçiler arasında ortalama ölüm yaşının 1980’lerin ortalarından sonra bir düşüş gösterdiği de ilgimizi çekebilir (SSK, 1995:99). DİE’nin çıkardığı Türkiye İstatistik Yıllıkları’ndan birine göz atarsak kentsel kesimde ortalama her 6 bebekten, kırsal kesimde her 5 bebekten birinin bir yıl içinde yaşamını yitirdiğini (DİE, 2002a:97), her yıl kendi canına kıyanlardan ortalama her üç kişiden birinin 15 – 24 yaşları arasında olduğunu (DİE, 2002a:111), cezaevine giren ortalama her 8 hükümlüden birinin lise ya da yüksek okul bitirdiğini (DİE, 2002a:185) görürüz. Meraklıları için İstanbul Menkul Kıymetler Borsası (İMKB) indeksinin de bir tür ortalamadan başka bir şey olmadığı belirtilebilir.

Bu kadar çok kullandığımız ortalama, her şeyden önce bir özetdir. Bir sürü sayının yerine geçip onları temsil etmeye çalışan, o sayıların hepsi adına kısa bir bilgi veren tek bir sayıdır. Bu işte ne denli başarılı olduğu, temsilcilik görevini ne denli iyi yerine getirdiğine bakılarak anlaşılabilir. Eğer bir ortalama, bize, temsil ettiğini ileri sürdüğü sayılar yerine kullanabileceğimiz anlamlı bir özet veremiyorsa, ne denli karmaşık yöntemlerle hesaplanmış, ne denli duyarlı bir biçimde dile getirilmiş olursa olsun başarılı bir ortalama sayılmaz. Öte yandan başarılı bir ortalama da tek başına fazla bir şey söyleyebilir ama kendisini destekleyen yayılıma, biçime ilişkin ek bilgilerle çok işe yarar özetlemeler yapabilir, benzer başka veri kümelerinin ortalama düzeyleriyle karşılaştırmalarda işimize yarayabilir. Aslında ortalama en yalın istatistik modelidir. Bir sonraki bölümde ele ala-

çağımız dağılımın biçimine ilişkin göstergelerle bu modeli zenginleştirip gerçekliğe daha çok yaklaşabiliriz.

Sayıları çok fazla olan ortalamalardan en yaygın kullanılanları üç öbeğe ayırabiliriz:

- **Ensikdeğer** (tepedeğer, mod), **ortanca** (medyan) gibi bütün gözlem değerlerinden etkilenmeyen **berk ortalamalar**,
- **aritmetik, geometrik, ters** (harmonik), **kareli** (kuadratik), **küplü** (kübik) ortalama gibi bütün gözlem değerlerinden etkilenen **duyarlı ortalamalar**,
- **kırpık ortalamalar** gibi yukarıdaki iki öbek ortalamasının sakıncalarını gidermeye yönelik olarak hesaplanan ortalamalar.

Bu ortalamaları teker teker incelemeye geçmeden önce bütün ortalamaların ortak bir özelliğini belirtmemiz gerekir. Bir ortalama, en küçük gözlem değerinden daha küçük, en büyük gözlem değerinden daha büyük olmaz. Bunu bir kez daha yazalım:

$$\text{En küçük gözlem değeri} \leq \text{Ortalama} \leq \text{En büyük gözlem değeri}$$

Herkese çok açık gibi görünen bu olgu, özellikle sınav heyecanı içinde gözden kaçabilmekte, masum (!) bir işlem hatası sonucunda ortalama yukarıdaki sınırlar dışında bulunabilmektedir. Hesaplama işlemi bittikten sonra yapılacak küçük bir sağlama bizi böylesi yanılgılardan kolayca koruyabilir.

8.2 BERK ORTALAMALAR

Yanlış soruya kesin yanıt yerine, doğru soruya yaklaşık yanıt bulmak daha iyidir.

John W. Tukey

Bu kesimde göreceğimiz ortalamalar, gözlem değerlerinden bazıları önemli ölçüde değişse bile bundan hemen hiç etkilenmeyebilirler. Bu özellikleri nedeniyle aşırı büyük ya da aşırı küçük değerlerin ortalamayı saptırmasını engellerler. İşte bu nedenle, dış etmenlerden kolay kolay etkilenmeyen, güçlü, sağlam anlamında, İngilizcedeki “robust” karşılığı **berk** ortalama adını alırlar. Deyim yerindeyse *vurdumduymaz* diye bile nitelenebilirler. Berk ortalamalar en çok, verileri ilk incelemeye aldığımız **sorgulayıcı çözümleme** aşamasında kullanılırlar. İleride ayrıntılı olarak açıklanacak bu çözümleme aşamasında veriler, bazı ipuçlarının bulunması, araştırmaya değer bazı özelliklerinin ortaya çıkarılması, verilerin yapısına ilişkin çabuk ama kaba bilgiler elde edilmesi amaçlarına yönelik olarak incelenirler. Da-

ha sonra doğrulayıcı çözümleme yapılırken bu bulgular daha ayrıntılı ele alınıp karar vermede kullanılırlar. Hesaplanmaları hızlı, kavranmaları kolay olduğu için bu tür ortalamaların sorgulayıcı aşamada çok işe yarayacakları söylenebilir.

Bu kesimde berk ortalamalar içinde en çok tanınan **ensıkdeğerden, ortancadan**, bir de daha az kullanılmakla birlikte son yıllarda önem kazanan **dördebölenler ortalamasından** söz edeceğiz. Ayrıca, tam anlamıyla ortalama olmasalar bile, gerek son ortalamanın hesabında, gerek veri dağılımlarının yayıklığını, çarpıklığını, basıklığını ölçmede kendilerinden yararlanılması nedeniyle **bölenler** (kantiller) üzerinde durulacaktır.

Ensıkdeğer (ESD)

Bir tür ortalama düzey göstergesi olan bu kavram, ensıkdeğer adının yanı sıra tepe değer, egemen değer, mod gibi adlar da almaktadır. Verilerin en yoğun olarak hangi gözlem değeri düzeyinde toplandığını gösterir. *ESD* kavramının bazı batı dillerindeki karşılığı olan “mod” (*mode*) sözcüğü, moda ile ortak kökenlidir. En yaygın görülen, en sık rastlanan, moda olan anlamına gelir. Sözelimi Türkiye’de en sık rastlanan ten rengi buğday, göz rengi kahverengi/siyah, saç rengi siyah (Efes Pilsen, 2002:9) dediğimizde bu ortalama söz etmiş oluruz.

ESD’ye bir örnek vermek için Alistırma 3.12’de bebek ölüm oranlarını gösteren çizelgeye dönelim. Bu veriler aşağıda Çizelge 8.1’de yeniden gösterilmiştir. Bunlarla Çizim 8.1’deki gibi bir dal-yaprak gösterimi düzenleyelim. Hem çizelgede, hem çizimde görüldüğü gibi gözlem değerlerinin en yoğun olarak toplandıkları düzey ‰ 6’dır. Bir başka deyişle 16 ülke içinde en sık rastlanan bebek ölüm oranı ‰ 6’dır. Bu oran 6 ülkede görülmektedir. Bir rastlantı sonucu hem *ESD*, hem de ona karşılık olan sıklık sayısı (frekans) aynı (6) bulunmuştur. Bu karışıklığa yol açabilir. Konuyu yeni öğrenmekte olan kimi öğrenciler sıklık sayısını yanlışlıkla *ESD* sanmaktadırlar. Bu yanlışlığa yol açmamak için dikkatli olmak gerekir.

ÇİZELGE 8.1 15 AB ülkesiyle Türkiye’de 2000 yılı bebek ölüm oranları (‰)

Ülke	‰	Ülke	‰	Ülke	‰
Almanya	6	Hollanda	7	Lüksemburg	6
Avusturya	6	İngiltere	7	Portekiz	8
Belçika	7	İrlanda	7	Yunanistan	8
Danimarka	6	İspanya	6	Türkiye	43
Finlandiya	5	İsveç	4		
Fransa	6	İtalya	7		

World Bank, 2003: 234-235.

(43)

8	00
7	00000
6	00000 0
5	0
4	0
Binde	On binde

ÇİZİM 8.1 15 AB ülkesiyle Türkiye’de 2000 yılı bebek ölüm oranları
Çizelge 8.1’deki verilerle düzenlenmiştir.

ESD aşırı gözlem değerlerinden pek etkilenmez. Gerçekten de bu örnekte bebek ölüm oranı öteki Avrupa ülkelerine göre çok yüksek olan Türkiye’nin bu değeri ister 4’e insin, ister 50’ye çıksın *ESD* değişmez, yine 6 kalır. Bu önemli bir özelliktir.

Bu örnekte *ESD* iyi bir temsilci midir? Eğer *ESD*’nin ne olduğu iyi kavranırsa, söylediğini iyi söyleyen bir ortalama olduğu ileri sürülebilir. *ESD*, gözlem değerleri arasında en çok yinelenen değerdir. Örneğimizde 16 ülkeden 6’sı aynı bebek ölüm oranını göstermektedir. Demek ki *ESD*’ye karşılık gelen sıklık sayısının öteki bütün sıklık sayıları toplamından ($16 - 6 = 10$ ’dan) daha büyük olma zorunluluğu yoktur. Bunu ikiden çok adaylı bir seçim sonucuna benzetebiliriz. En yüksek oyu alan kimsenin aldığı oy sayısının, %50’yi aşmasının gerekmediği seçimlerde durum böyledir. Bu yolla seçilen bir aday ne denli iyi bir temsil ediciyse, *ESD* de o denli iyi temsil edicidir. Elbette bir aday toplam oyların büyük çoğunluğuyla seçilmişse temsil ediciliği yüksek olur. *ESD*’nin sıklık sayısının toplam içindeki oranı yükseldikçe temsil gücü de artar.

Tarihte bilinen en eski kullanımlarından birini aktararak bu ortalamanın temsil edicilik işlevini açıklayabiliriz. İsa'dan önce 5. yüzyılın ikinci yarısında yaşamış tarihçi Tukidides, *Pelepenez Savaşlarının Tarihi* adlı yapıtında şöyle bir olay anlatır. Yüksek surlarla çevrili bir kenti kuşatan bir ordu, duvarları aşmak üzere yeterince uzun merdivenler yapacaktır. Merdivenlerin en uygun uzunluğunu belirleyebilmek için surların yüksekliğinin bilinmesi gerekmektedir. Ne var ki askerler ok yağmuru altında surlara yaklaşıp bunların kaç sıra kesme taştan yapıldığını doğru olarak saymakta güçlük çekmektedirler. Bunun üzerine sayma işiyle görevlendirilen her askerin söylediği sayı bir yere yazılır, bunlardan en çok yineleneni kesme taşların doğru sayısı olarak benimsenir. Bir kesme taşın yüksekliği bilindiğinden, merdivenler için en uygun yükseklik belirlenebilir. Burada kullanılan ortalama türünün *ESD* olduğu açıktır.

ESD'yi öteki ortalamalardan ayıran önemli bir özelliği, gözlem değerlerinin sayısal bir değere ilişkin olma zorunluluğunun bulunmamasıdır. Daha önce belirtildiği gibi **ad ölçeği**nde ölçülmüş veriler için bile *ESD* hesaplanabilir. Çizelge 3.1'de en sık rastlanan medeni durum *evli* olmaydı. Dolayısıyla bu veri kümesinin, başka bir deyişle *medeni durum* değişkeninin *ESD*'si *evli* olma durumudur. Çizelge 3.2'de **sıra ölçekli beğenme düzeyi** değişkeni için de *ESD*, *beklediğimden kötü* yanıtıdır. Bu özelliği nedeniyle en çok satılan model, en tutulan renk, en beğenilen koku gibi kavramlara çok iyi uyum sağlar.

Ancak bu özellik, *ESD*'nin nicel değişkenler için de kullanılmasını engellemez. Bir bölgedeki evlerin oda sayısı, hanedeki kişi sayısı, işteki günlük devamsızlık sayısı gibi durumlarda da *ESD* başarıyla kullanılabilir. Bu örneklerdeki gözlem değerlerinin hep tamsayı olması dikkatinizi çekmiştir. Bunun nedeni, *ESD*'nin özellikle tamsayı değişkenleri için uygun bir ortalama niteliği taşımasıdır. Böyle değişkenlerde gözlem değerlerinden biri *ESD*'dir. **Oran ölçeği**ne örnek olarak verilen Çizelge 3.4'teki bir hanede yaşayan *kişi sayısı* değişkeni için de *ESD* 4 kişidir. Çünkü en çok 4 kişilik hanelere rastlanmaktadır.

Tamsayı olmayan değişkenlerde her gözleme yalnızca bir kez rastlandığı, dolayısıyla sıklık sayılarının hepsinin bir olduğu durumlar görülebilir. O zaman *ESD* diye bir ortalama da yok demektir. Ama eğer tamsayı olmayan gözlem değerleri belli öbeklerde toplanırlarsa *ESD* hesaplanabilir, en yüksek sıklık sayısını içeren öbek, *ESD* öbeği olarak adlandırılabilir. Çizelge 3.3'teki **aralık ölçeği**ne göre ölçülmüş hava sıcaklıkları içinde *ESD* hesaplamak istersek hiçbir gözlemin birden çok sayıda olmadığını görürüz. Ama eğer verileri birer C derecelik aralıklarda öbeklersek verilerin 14 – <15° C aralığında yoğunlaştığını, dolayısıyla *ESD*'nin bu aralık olduğunu anlarız.

ÇİZELGE 8.2 Türkiye nüfusunun yaşa göre dağılımı (1990)

Yaş aralıkları	Nüfus içindeki pay (%)
0 – < 10	22,8
10 – < 20	23,2
20 – < 30	17,6
30 – < 40	13,4
40 – < 50	8,8
50 – < 60	7,0
60 +	7,1
Bilinmeyen	0,1
Toplam	100,0

DİE, 2002a:57-59'deki verilerle düzenlenmiştir.

Bunu bir başka örnekle de gösterelim. 1990 sayımına göre nüfusun yaş öbeklerine göre dağılımı Çizelge 8.2'deki gibidir. Dikkat edilirse burada sıklık sayıları, kişi sayısı olarak değil, % olarak verilmiştir. Bu tür sıklık gösterimlerine **görelî sıklık** ya da **görece sıklık** dendiğini biliyoruz. *ESD* hem sıradan, hem görelî sıklık sayılarıyla bulunabilir. Örneğimizde en yüksek görelî sıklık (% 23,2) 10 ile 20 yaşları arasındadır. Öyleyse 10 – < 20 yaş aralığı *ESD* öbeğidir.

ESD'yi daha incelikli bulabilir miyiz? 10 – < 20 arasındaki hangi yaşın *ESD* olduğunu anlayabilmek için Çizelge 8.2'deki verilere bir göz atalım. 10 – < 20 yaş öbeğinin iki yanındaki 0 – < 10 ile 20 – < 30 yaş öbeklerinin sıklıkları birbirlerine eşit olsaydı, *ESD* öbeği içindeki insanların yaşlarının bu aralık içinde bakışık dağıldığı varsayımını yapabilir, *ESD*'yi bu aralığın orta noktasını gösteren 15 olarak bulabilirdik. Oysa durum böyle değildir. En genç öbeğin görelî sıklığı, öbür komşu öbeğinkine göre hayli yüksektir. Öyleyse *ESD*, bu genç öbeğe daha yakın bir yaşta olmalıdır.

Peki ama kaçtır? İki komşu öbeğin sıklıklarının *ESD* öbeği sıklığına göre farkı bizi bir sonuca götürebilir. Bu amaçla her iki yandaki sıklıkların 23,2'ye göre farklarını alalım, her bir farkın toplam fark içindeki payını bulalım. *a*, *b* diye adlandırılan bu paylar bize çarpıklığın hem yönünü, hem boyutunu gösterebilir. Bunu Çizelge 8.3'teki gibi özetleyelim.

ÇİZELGE 8.3 Öbeklenmiş verilerde *ESD*'nin hesaplanması

Yaş öbekleri	Sıklıklar %	Farklar %	Fark payları
0 – < 10	22,8	0,4	$0,4 / 6,0 = 0,067 = a$
10 – < 20	23,2	–	–
20 – < 30	17,6	5,6	$5,6 / 6,0 = 0,993 = b$
Toplam		6,0	$1,000 = a + b$

Çizelge 8.2'den hesaplanmıştır.

Fark payları sütunundaki değerler *ESD*'nin 10 – < 20 aralığı içinde alt sınırdan 0,067, üst sınırdan 0,993 oranında bir uzaklıkta bulunduğunu gösterir. Öyleyse bu bilgileri kullanarak aşağıdaki iki denklemden biri yardımıyla *ESD*'nin değerini daha incelikli olarak bulabiliriz.

$$20 - ESD = \frac{b}{a + b} (20 - 10) = 9.33$$

$$ESD - 10 = \frac{a}{a + b} (20 - 10) = 0.67$$

Bunların herhangi biri bize *ESD*'yi 10,67 olarak verecektir. Öyleyse *ESD* yaklaşık 10 yıl 8 aydır, daha kaba söylersek 11 yaşıdır. *ESD*'yi genel olarak gösterelim:

$$ESD = Y_{alt} + \frac{a}{a + b} \times \text{öbek aralığı} = ESD = Y_{üst} - \frac{b}{a + b} \times \text{öbek aralığı}$$

burada Y_{alt} *ESD* öbeğinin alt sınır değeri, $Y_{üst}$ üst sınır değeridir.

a , b fark paylarının yaş dağılımının histogramı üzerinde işaretlemesi, yukarıdaki hesaplamaların benzer üçgenlere dayandığını göstermesi okuyucudan istenecektir.

Bu örnekte de *ESD*'nin aşırı değerlerden etkilenmediği anlaşılmaktadır. En yaşlı kimsenin yaşı 100 de olsa 120 de olsa *ESD* değişmeyecektir.

Dikkatli okuyucunun anlamış olabileceği gibi *ESD* yalnız tek tepeli dağılımlar için elverişlidir. Eğer bir dağılımda birden çok tepe varsa, gözlem değerleri birden çok yerde yoğunlaşıyorlarsa, ilk akla gelmesi gereken olasılık, gözlem değerlerinin türdeş bir topluluğu yansıtmadığıdır. Bir sınıftaki öğrencilerin ağırlıklarını gösteren dağılımda iki tepe görüyorsak sınıfta hem kız, hem erkek öğrenciler bulunuyor denebilir. Benzeri bir örneği Çizelge 8.4'teki verilerle inceleyelim.

ÇİZELGE 8.4 Üniversitelerimizde öğretim üyesi başına öğrenci sayısı (1986)

Öğretim üyesi başına öğrenci sayısı	Üniversite sayısı	
10 – < 20	1	
20 – < 30	5	1. tepe
30 – < 40	3	
40 – < 50	7	
50 – < 60	9	2. tepe
60 – < 70	2	
70 – < 80	1	
Toplam	29	

ÖSYM, 1987'den hesaplanmıştır.

Bu dağılımda iki tepe görülmektedir. Üniversitelerimizin 1986'da öğretim üyesi başına öğrenci sayısı bakımından biri 20-30, öteki 50-60 olan iki öbekte yoğunlaştıkları anlaşılmaktadır. Her ne kadar bu iki yoğunlaşma eşit değilse de (birinde 5, öbüründe 9 üniversite varsa da) yine de yerel iki tepeden söz edebiliriz. Bu da bize o yıl üniversitelerimizin söz konusu oran açısından türdeş olmadıkları, büyük kentlerdeki üniversitelerle o sıralarda çoğu yeni kurulmuş olan öteki üniversitelerin birbirinden ayrı iki öbek oluşturdukları izlenimini vermektedir. Yeni verilerle bu farklılığın zaman içinde ortadan kalkıp kalkmadığını incelemek okuyucudan beklenmektedir.

Şimdi bir başka ortalama ölçüsüne geçelim.

Ortanca (Ortc)

Gözlem değerlerinin hepsinden etkilenmeyen ortalamalardan biri de ortanca'dır. Medyan adı da verilen bu ortalamanın hesaplanması çok kolaydır. Adından da anlaşılacağı gibi *Ortc*, gözlem değerleri büyüklük sırasındayken tam ortaya düşen değerdir. Nasıl ki üç kardeşten yaşı ortada olana ortanca diyorsak istatistikte de bu terime aynı anlamı yüklüyoruz. Kardeş sayısı tek değilken ortanca kardeşten söz edilemese de istatistikte çift sayıda gözlem değerlerinin ortalaması vardır. Bunlar büyüklük sırasına dizildiğinde, tam ortaya gelen iki gözlem değerinin aritmetik ortalaması *Ortc* olur. *Ortc*'nin genel gösterimi şöyledir:

$$N \text{ tekse } Ortc = Y_{\frac{N+1}{2}} \quad N \text{ çiftse } Ortc = \frac{1}{2} \left(Y_{\frac{N}{2}} + Y_{\frac{N+1}{2}} \right)$$

Burada N gözlem sayısını, Y değişkenin o gözlemdeki değerini, altimlerde o gözlemin büyüklük sıralamasındaki yerini gösterir. Sözgelimi $N = 9$ iken $Ortc$, $(9+1)/2 = 5$. sıradaki gözlemin değeri, yani Y_5 'tir. $N = 10$ olursa, $Ortc$, 5. ile 6. gözlemlerin ortalamasıdır, yani $(Y_5 + Y_6) / 2$ 'ye eşittir.

Şimdi Çizelge 8.1'deki örneğe dönüp bebek ölüm oranlarının $Ortc$ değerini bulalım. Bu örnekte 16 ülke ($N = 16$) olduğuna göre, bunlar söz konusu değişkenin değerlerine göre sıraya dizilirse 8. ile 9. ülkelerin bebek ölüm oranlarının ortalaması bize $Ortc$ değerini verir. Bunun için **birikimli sıklık** adı verilen yeni bir kavramı kullanacağız. Belli bir değere eşit ya da daha küçük gözlemlerin sayısına birikimli sıklık denir. Önce Çizelge 8.5'te bu hesaplamaları görelim.

ÇİZELGE 8.5 15 AB ülkesiyle Türkiye'de 2000 yılı bebek ölüm oranları için ortanca hesabı

	Ortc ↓					
Bebek ölüm oranı ‰	4	5	6	7	8	43
Sıklık (ülke sayısı)	1	1	6	5	2	1
Birikimli sıklık	1	2	8	13	15	16
	←	8 ülke	→	↑	←	8 ülke →

Çizelge 8.1'den hesaplanmıştır.

Birikimli sıklıklardan 8. ülkenin bebek ölüm oranının ‰ 6, bir sonrakinin ‰ 7 olduğu anlaşılmaktadır. Öyleyse $Ortc = ‰ 6,5$ diyebiliriz. ESD gibi $Ortc$ da aşırı değerlerden pek etkilenmez. Bu örnekte aşırı değer olan Türkiye bebek ölüm oranı ister ‰10, ister ‰100 olsun, $Ortc$ aynı kalır. Ama Türkiye'yi aşırı değerinden dolayı dışlarsak kalan 15 ülke içinde $Ortc$, 8. sırayı alan ‰ 6 olur.

Çizelge 8.2'de verilen yaş dağılımına dönelim. Bunun $Ortc$ değerini nasıl bulabiliriz? Bu örnekte gözlem değerlerinin sıklıkları göreceli olarak verilmiştir. Yaşı bilinmeyenlerin genellikle yaşlı oldukları, dolayısıyla $Ortc$ 'nin üstünde yer aldıkları varsayılabilir. Şimdi birikimli sıklıkları yazalım. Çizelge 8.6'da bu iş yapılmıştır. Bu durumda $Ortc$, verileri % 50 – % 50 olarak ikiye ayıran değerdir. Çizelgeden bunun 20 – < 30 yaş öbeği içinde olduğu görülmektedir.

ÇİZELGE 8.6 1990 Türkiye nüfusunun yaşa göre dağılımının ortanca hesabı

Yaş aralıkları	Nüfus içindeki pay (%)	Birikimli görelî sıklık	
0 – < 10	22,8	22,8	
10 – < 20	23,2	46,0	
20 – < 30	17,6	63,6	← <i>Ortc</i> aralığı
30 – < 40	13,4	77,0	
40 – < 50	8,8	85,8	
50 – < 60	7,0	92,8	
60 +	7,1	99,9	
Bilinmeyen	0,1	100,0	
Toplam	100,0		

DİE, 2002a:57-59'daki verilerle düzenlenmiştir.

Bu değer daha incelikli olarak bulunabilir mi? 20 yaş altındakilerin birikimli görelî sıklığı % 46'dır. Demek ki % 50'ye ulaşmak için *Ortc* aralığının alt sınırına bitişik % 4'lük bir bölümün hangi yaşa kadar uzandığını bulursak *Ortc*'yı da buluruz. % 17,6'lık *Ortc* dilimini *Ortc*'nin altındakiler, *Ortc*'nin üstündekiler olarak şöyle ikiye ayıralım. Nüfusun % 46'sı 0 – < 20 yaşları arasında olduğuna göre *Ortc* öbeği içindeki % 17,6'nın (50,0 – 46,0 =) 4'ü *Ortc*'nin altında, (17,6 – 4,0 =) 13,6'sı *Ortc*'nin üstündedir. *Ortc*'nin altındaki gözlemlerin, bu öbeğin içindeki gözlem sayısına oranı (4,0 / 17,6 =) 0,227; üstündekilerin oranı da (13,6 / 17,6 =) 0,773 kadardır. Bu öbek içindeki verilerin tekdüze dağıldığı varsayımıyla *Ortc*'nin, 20 – < 30 yaş aralığını da aynı oranlarda ikiye ayırdığını söyleyebiliriz. Bu durumda *Ortc* 22,27 olarak bulunur. Öyleyse nüfusun yarısı 22 yıl 99 günlük ya da daha genç, öbür yarısı da 22 yıl 99 günlük ya da daha yaşlıdır. Kabaça söylersek *Ortc* yaş 22'dir.

Ortc'nin genel gösterimi şöyledir:

$$Ortc = Y_{alt} + \frac{0.5 - \sum_i^{Y_{alt}} f_i}{f_{Ortc}} \times Ortc \text{ öbeğinin aralığı}$$

Bir başka örnek olarak Alıştırma 7.7'deki verilerle Çizim 7.9'daki dal-yaprak gösterimini aşağıdaki gibi yeniden düzenleyip *Ortc* hesaplayalım.

3	3 ay	667
13	3 db	44444 45555
24	3 iü	22222 22333 3
44	3 sb	00000 00011 11111 11111
(18)	2 sd	88888 88888 89999 999
34	2 ay	66666 77777 77
22	2 db	44455 5555
13	2 iü	22233 33333 33
1	2 sb	1
1		0,1

ÇİZİM 8.2 Bitirme notları için 0,2 aralıklı dal-yaprak gösterimi
Alıştırma 7.7'deki verilerle düzenlenmiştir.

Bu gösterimin daha öncekinden farkı en sola eklenen sütundur. Bu sütundaki sayılar **birikimli sıklıklar**ı gösterir. Yalnız bu birikimli sıklıklar iki uçtan başlayıp ortaya doğru ilerler. İlk sayı o satırda 3 gözlem olduğunu gösterir, yani o aralıktaki sıklık sayısı 3'tür. Bir sonraki 13 sayısı da ilk iki satırın birikimli sıklığıdır. Toplam 96 gözlemimiz bulunduğu göre *Ortc* 48. ile 49. gözlemlerin orta noktasıdır.

Üstten 44'e gelince durmak gerekir. Çünkü bir alt satırın sıklığı da buna eklenirse ($44 + 18 = 64$) *Ortc*'nin sıralamadaki yeri aşılacaktır. Aynı şey alttan başlayarak gelen sıklık sayıları için de geçerlidir. Burada da 34'e ulaşınca ilerleme sürmez. Çünkü $34 + 18 = 52$, *Ortc*'nin sıralamadaki yerini geçmektedir. Öyleyse bu satıra birikimli sıklık yerine yalnızca o satırın sıklığını koyar, bunu belirtmek için de ayrıç içine alır, (18) yazarız. Demek ki *Ortc* değeri bu satırdadır. Burada bir sağlama da yapılabilir. Bu üç sıklık $44 + 34 + 18$ diye toplanırsa gözlem sayısı 96'yı vermelidir.

Acaba *Ortc* kaçtır? Bunu da şöyle buluruz. 44'ten 48'e ulaşmak için 4 sayı daha aşağıya doğru inmemiz gerekir. Üstten 44 sayının en küçüğü 3,0 ile başlayan bir nottur. Bunun bir altındaki not 2,9'la başlar, öyleyse bu satırdaki sağdan dördüncü 9 ile beşinci 9'un arası *Ortc*'dir. Öyleyse *Ortc* 2,9'dur. Üç basamaklı asıl değerlere dönecek olursak 2,9 ile başlayan notlardan üstten dördüncüsü 2,95, beşincisi 2,92'dir. Öyleyse *Ortc* bu ikisinin orta noktası, yani $(2,92 + 2,95) / 2 = 2,935$ 'tir. *Ortc*'yi böyle 4 basamaklı bulmak doğru mudur? Sayıların kendileri 3 anlamlı basamaktan oluşurken *Ortc*'nin 4 anlamlı basamak içermesini nasıl karşılırsınız? Yuvarlatma konusunda söylenenleri anımsayarak bu soruyu siz yanıtlayın.

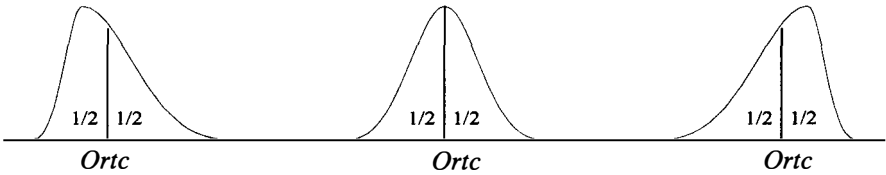
Çizim 8.2'den *Ortc*'yi böyle kolayca hesaplayabildiğimiz gibi **ensıkdeğer** konusunda da bir fikir edinebiliriz. Çizimi biraz incelersek 3sb

satırında 12 tane 1 olduğunu görürüz, bu da en çok yinelenen sayıdır. Demek ki *ESD* 3,1 ile başlayan notların bulunduğu aralıktır.

Bitirme notları 2,8 ile başlayanlarla 3,1 ile başlayanlar arasında yoğunlaşmıştır. Bu olguyu bir önceki dal-yaprakta da saptamıştık.

Dal-yaprak gösteriminden not dağılımının ortalama düzeyi konusunda kabaca da olsa bilgilenmiş olduk. Tek tek notların bu ortalama düzeyden uzaklaşma dereceleri de yine bu çizimden anlaşılabilir. Bu uzaklaşmalar iki yöne doğru bakışık mıdır, yoksa bir çarpıklık mı söz konusudur? Yine bu çizime bir göz atıp hemen anlayabiliriz. Bu konunun uzantılarına ileride ayrıntılı olarak değineceğiz.

Ortc'nin en önemli özelliği, bir dağılımdaki gözlem sayısını iki eşit parçaya ayırmasıdır. Çizim 8.3'te bu özellik sergilenmektedir.



ÇİZİM 8.3 *Ortc* gözlem sayısını iki eşit parçaya böler

Bölenler

Bir veri kümesini eşit sayıda parçalara ayıran değerlerin genel adı **bölenler**dir. Kantil de denen bu değerlerden biri olan *Ortc*'yi daha önce gördük. *Ortc* gözlem sayılarını iki eşit parçaya ayırıyordu. Öteki bölenler de aynı ilkeye göre bulunur. Gözlemleri dört eşit parçaya bölen değerlere **dördebölen** (*D*) (kartil), beş eşit parçaya bölenlere **beşebölen** (kentil), on eşit parçaya bölenlere **onabölen** (desil), yüz eşit parçaya bölenlere de **yüzebölen** (santil) vb. diyoruz.

Yukarıda Çizim 8.2'deki verileri örnek olarak kullanıp dördebölenleri (*D*) hesaplayalım. Bu örnekte mezun sayısı 96'dır. $(96 + 1)/2 = 48,5$ olduğuna göre *Ortc*, 48. ile 49. sıradaki öğrencilerin notlarının ortalamasıdır. Bunun da 2,9 dolayında (daha incelikli söylersek 2,935) olduğunu bulmuştuk. Demek ki 96 öğrenciden 48'inin bitirme notu 2,935'in altında, 48'inin de üstündedir.

D'ler üç tanedir. İlk % 25 dilimin üst sınırındakine **alt dördebölen** denir, D_a ya da D_1 ile gösterilir. *Ortc*, **ikinci dördebölen**dir, dolayısıyla D_2

biçiminde de yazılabilir. Son % 25'lik dilimin alt sınırındaki D de **üst dördebölen** adıyla anılır, D_u ya da D_3 diye kısaltılır.

D 'ler üç değişik yolla hesaplanır. Bunlardan biri geleneksel yol, biri **Tukey menteşeleri**, üçüncüsü de bizim önerdiğimiz basit yoldur. Tukey uç değerlerle üç D 'yi M harfinin beş köşesine gelecek biçimde yerleştirip bir gösterim hazırlar. Bu köşeleri de katlanabilen bir pancurun menteşe yerlerine benzetir. Bu yüzden onun hesap yoluna bu ad verilmiştir. (Tukey, 1977:33) Her üç yolla da bulunan $Ortc$ aynıdır, öbür D 'lerde küçük farklar olabilir.

İlk yola göre gözlem sayısının bir fazlasının dörtte biri D_a 'nın, dörtte üçü D_u 'nün konumunu belirler. Yani

$$D_a = Y_{\frac{N+1}{4}} \quad D_u = Y_{\frac{3(N+1)}{4}}$$

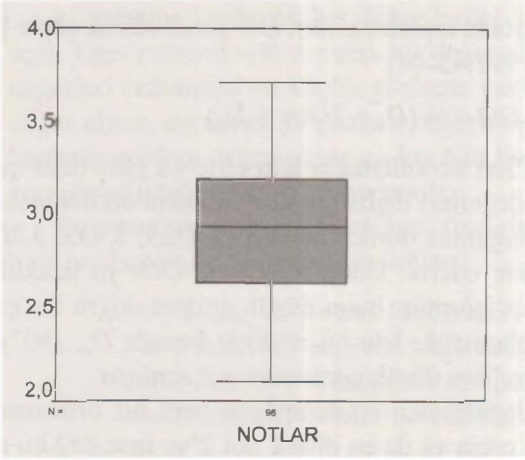
yazabiliriz. Burada da altimler büyüklük sıralamasındaki konumu gösterir.

Örneğimizde $\frac{1}{4} (N + 1) = \frac{1}{4} (96 + 1) = 24,25$ olduğuna göre D_a , 24. gözlemle 25. gözlem arasındadır, 24. gözlem değerine aradaki farkın $\frac{1}{4}$ 'ü kadar uzaktır. 24. gözlem 2,63 değerine, 25. gözlem 2,64'e karşılık gelir. Aradaki fark 0,01'dir. Bunun $\frac{1}{4}$ 'ü 0,0025 olduğuna göre $D_a = 2,6325$ 'tir.

Benzer biçimde $\frac{3}{4} (N + 1) = \frac{3}{4} (96 + 1) = 72,75$ eder. 72. gözlem 3,16 notuna, 73. gözlem 3,22'ye karşılık geldiğinden aradaki fark 0,06 olur. Bunun $\frac{3}{4}$ 'ü 0,045 eder. Demek ki $D_u = 3,16 + 0,045 = 3,205$ 'tir. Öyleyse bütün mezunların $\frac{1}{4}$ 'ü (yani 24 kişi) 2,6325'ten düşük, $\frac{1}{4}$ 'ü 2,6325 – 2,935 arasında, $\frac{1}{4}$ 'ü 2,935 – 3,205 arasında, son $\frac{1}{4}$ 'ü de 3,205'ten yüksek notlar almışlardır.

D 'lerin hesaplanmasında kullanılan öbür iki yol sorgulayıcı veri çözümleme yöntemleri anlatılırken açıklanacaktır.

Dikkat edilirse bulunan D_a 'nın $Ortc$ 'ya uzaklığıyla D_u 'nünki tam aynı değildir. İkincisi azıcık da olsa daha kısadır. Bu da, $Ortc$ 'nın hemen altındaki 24 öğrencinin notlarının 0,325 (= 2,935 – 2,6325) not aralığında yayılmasına karşın, $Ortc$ 'nın hemen üstündeki $\frac{1}{4}$ öğrencinin notlarının 0,27 (= 3,205 – 2,935) gibi biraz daha dar bir not aralığında yer aldığı anlamına gelir. Buradan $Ortc$ 'nın iki yanındaki 24'er öğrencinin (toplamın yarısının) tam bakışık olmadığı, dağılımın $Ortc$ 'nın hemen altındaki notları içeren bölgenin çok az da olsa daha uzun olduğu anlaşılır. Bunu Çizim 8.4'ten de görebiliriz. **Kutu çizimi** denen bu gösterimi daha önce çizimleri incelerken görmedik, çünkü dördebölenler kavramını öğrenmemiştik.



ÇİZİM 8.4 Bitirme notlarının kutu çizimi
Aıştırma 7.7'deki verilerle düzenlenmiştir.

Bu çizim bize bitirme notları veri kümesinin özellikleri konusunda hayli bilgi verir. Ama önce nasıl çizildiğini anlatalım. Dik eksen *bitirme notu* değişkeninin aldığı değerleri gösterir. Ortadaki kutunun alt çizgisi D_a 'nın, üst çizgisi de D_u 'nün hizasındadır. Kutuyu ikiye bölen çizgi de $Ortc$ değerine karşılıktır. Kutunun iki yanında uzayan uçlar D 'lerin dışında kalan gözlem değerlerini temsil eder. Uç değerler kısa birer yatay çizgiyle işaretlenmiştir. Bu örnekte dışadüşen olmadığı için şimdilik bu konuya değinmeyeelim.

Çizime bakarak $Ortc$ 'nın, D 'lerin, Y_{ek} , Y_{eb} değerlerinin nerelerde olduğunu kolayca anlayabiliriz. Kutu içindeki iki bölümün birbirlerine göre büyüklükleri, orta bölgedeki bakışım; kutunun iki yanındaki kolların uzunluklarıysa kutu dışındaki bakışım konusunda bilgi verir. Her iki bölgede de hayli bakışık bir görüntüyle karşı karşıyayız. Ancak dikkatli bakılırsa kutunun alt yarısının üst yarısından daha uzun olduğu görülür. Bu da az önceki bulgumuzu doğrular. Bu konuya yeniden döneceğiz.

Dördebölenler ortalaması (DO)

Bir veri kümesinin ortalama düzeyine ilişkin bilgi edinebilmek için çok kolay hesaplanabilen başka bir ortalama da tartılı bir ortalama olan dördebölenler ortalamasıdır, DO ile gösterilebilir. D_a ile D_u birer, $Ortc$ iki ağırlıkla ortalamaya katılır. $Ortc$ 'nın iki ağırlık almasının nedeni gözlem

değerlerinin genellikle ortada toplanmasıdır. *DO* genel olarak şöyle bulunur (Erickson, Nosanchuk, 1992:43-44):

$$DO = \frac{1}{4} (D_a + 2Ortc + D_{\bar{u}})$$

Bu ortalama, öbür *D*'leri de kullandığı için *Ortc*'ya göre daha çok bilgi içerir. Öte yandan aşırı değerleri dışladığı için onlardan etkilenmez. Çizim 8.4'teki veriler için bulduğumuz dördebölenleri (2,6325, 2,935, 3,205) kullanırsak $DO \approx 2,927$ elde ederiz. Dördebölenlerin *Ortc*'ya uzaklıklarının eşit olmadığını daha önce görmüş, bunu düşük notlara doğru bir çarpıklık göstergesi olarak yorumlamıştık. İşte bu nedenle burada D_a , *DO*'yu daha çok etkileyerek onun *Ortc*'dan düşük çıkmasına yol açmıştır.

Bu ortalama da uç değerlerden etkilenmeyen berk bir ortalamadır. En yüksek bitirme notu 4'e çıksa ya da en düşük not 2'ye inse *DO* bu değişikliklerden hiç etkilenmez.

Ortc'yı hiç hesaba katmadan yalnızca iki *D*'nin ortalaması alınırsa buna **yarı dördebölenler ortalaması** (*YDO*) denir. Kimi yazarlarca **dördebölenler ortası** diye de adlandırılan bu ortalama şöyle bulunur:

$$YDO = (D_a + D_{\bar{u}}) / 2$$

Bitirme notları örneğimizde $YDO = (2,6325 + 3,205) / 2 = 2,9375$ 'tir.

8.3 DUYARLI ORTALAMALAR

Bu tür ortalamalar, adlarından da anlaşılacağı gibi, ortalaması hesaplanacak gözlem değerlerinin hepsinin hesaba katılması sonucunda bulunurlar, dolaşısıyla da gözlem değerlerinden bir tanesi bile bir değişikliğe uğrasa, bundan az ya da çok etkilenirler.

$$M_r = \left(\sum_{i=1}^N Y_i^r / N \right)^{1/r}$$

Bu öbeğe giren ortalamaların hepsi tek bir tanıma bağlanabilir. Bu genel tanım şöyledir: Gözlem değerlerinin *r*'nci kuvvetleri toplamının gözlem sayısına oranının $1/r$ 'nci kuvveti.

Burada

M_r : bütün gözlem değerlerinden etkilenen *r*'ye bağlı ortalama,

Y_i : gözlem değerlerinden *i*'ncisi,

r : $-\infty$ ile $+\infty$ arasında değerler alabilen bir sayı,

N : gözlem sayısıdır (Ana kütle için *N*, örneklem için *n* kullanılır).

r terimine verilecek her değer başka bir ortalamanın bulunmasına yol açar. Eğer r terimi $-\infty$ 'a yakın bir değer alırsa, ortalamanın değeri gözlem değerleri arasındaki en küçük gözleme yaklaşır. Tersine r , $+\infty$ 'a yakın bir değer alırsa, ortalama da gözlem değerlerinin en büyüğüne yaklaşır. Ortalamanın gözlem değerlerinin en küçüğü ile en büyüğü arasında yer alması zorunluluğu burada da görülmektedir.

Bu formülde r 'ye verilecek bazı özel değerler, yaygın olarak kullanılan bazı ortalamaların bulunmasını sağlar:

$r = -1$ iken M_{-1} = ters (harmonik) ortalama,

$r = 0$ iken M_0 = geometrik ortalama,

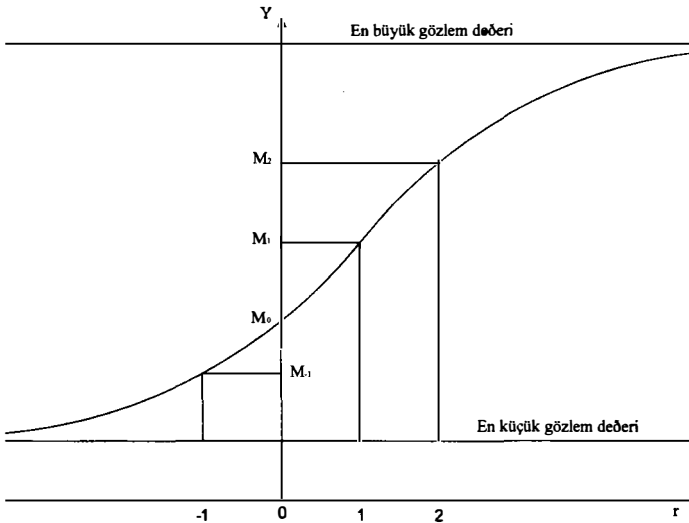
$r = 1$ iken M_1 = aritmetik ortalama,

$r = 2$ iken M_2 = kareli ortalama, vb. olur.

Bu ortalamaların büyüklük sıralaması Çizim 8.4'ten görüleceği üzere r 'nin büyüklük sıralaması gibidir. Çok kullanılan dört ortalama için bu sırayı şöyle yazabiliriz.

Ters ortalama $\leq$ Geometrik ortalama $\leq$ Aritmetik ortalama $\leq$ Kareli ortalama
 $TO \leq GO \leq AO \leq KO$

Her ne kadar Çizim 8.5 bakışık veri kümeleri için çizilmişse de bakışık olmayan dağılımlar için de biraz değiştirilerek kullanılabilir.



ÇİZİM 8.5 Bütün gözlemleri artı değerli, bakışık veriler için ortalamalar

Şimdi bu ortalamaları teker teker görelim. Bunu yaparken en yaygın olarak kullanılan, hemen herkes tarafından bilinen aritmetik ortalamadan başlayacağız. Daha sonra öteki ortalamaları ele alacağız.

Aritmetik ortalama (AO)

Bütün gözlem değerlerinden etkilenen ortalamalar içinde, anlaşılması kolay olduğu için en yaygın kullanılan aritmetik ortalamadır. Bu ortalamayı belki de farkında olmadan en çok kullananlar mesleği aşçılık olanlardır. Bir aşçı kaç kişiye, sözgelimi pilav pişireceğini bilmek ister. Gerisi onun için kolaydır. Çünkü kişi başına ne kadar pirinç kullanması gerektiğini bilir. Bunu kişi sayısıyla çarpınca ne kadar pirinç kullanacağını bulur. Öbür malzemeler için de benzerini yapar.

AO, gözlem değerlerinin toplamı gözlem sayısına bölünerek hemen bulunabilir. Gözlem değerleri Y_i (anakütle için $i = 1, 2, \dots, N$, örneklem için $i = 1, 2, \dots, n$) ile gösterilirse, AO

$$\text{ana kütle için } \mu = \frac{1}{N} \sum_{i=1}^N Y_i \quad \text{örneklem için } \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

ifadeleriyle hesaplanabilir. Burada μ (mü) ana kütledeki, $\bar{Y}$ (Y çizgi) ise örneklemdeki AO değerini simgelemektedir. Dikkatli bir okuyucu bu ifadenin ortalamalar genel formülünde r yerine 1 konulmasıyla elde edilebileceğini görmüştür.

$$M_1 = \bar{Y} = \left(\sum Y_i^1 / n \right)^{1/1} \quad (n \text{ yerine } N \text{ konulursa } M_1 = \mu \text{ olur})$$

AO'nun hesaplanmasını bir örnek üzerinde gösterelim. Çizelge 3.3'te Marmara Bölgesi'ndeki bazı merkezlerde ölçülmüş hava sıcaklıkları şöyleydi: Balıkesir 14.5°, Bursa 14.6°, Çanakkale 14.8°, Edirne 13.4°, Göztepe 14.0°. Bu beş merkezin ölçümlerine göre Marmara Bölgesi'nin ortalama hava sıcaklığı

$$\bar{Y} = (14.5 + 14.6 + 14.8 + 13.4 + 14.0) / 5 = 71.3 / 5 = 14.26^\circ \approx 14.3^\circ$$

bulunur.

Çizelge 8.1'deki bebek ölüm oranlarının AO'sunu hesaplayalım:

$$\mu = (4 + 5 + 6 \times 6 + 5 \times 7 + 2 \times 8 + 43) / 16 = 139 / 16 = 8,6875 \approx 8,7$$

Bu ortalama sizce bu veri kümesini çok iyi özetleyip temsil edebiliyor mu? Türkiye dışındaki bütün ülkelerin bebek ölüm oranları bulduğumuz AO'nun altında kalıyor. Daha önce aynı veriler için $ESD = 6$, $Ortc = 6,5$

bulmuştuk. Bu üç ortalama arasında sonuncunun ötekilerden hayli farklı çıkması dağılımın çarpık olduğunun ya da veriler arasında aşırı uç değer(ler)in bulunduğu bir göstergesi olarak yorumlanmalıdır. Öyleyse burada tek başına Türkiye değeri ortalama üzerinde baskın bir etki yaratıp onu kendi düzeyine doğru çekmektedir. Türkiye'yi dışarıda bırakıp kalan 15 AB ülkesi için *AO* hesaplırsak

$$\mu = (4 + 5 + 6 \times 6 + 5 \times 7 + 2 \times 8) / 15 = 96 / 15 = 6,4$$

elde ederiz. Bu ortalama 4 – 8 arasında yayılan değerler için iyi bir özet gösterge olarak kullanılabilir diyebilir miyiz? Daha önce Türkiye dışındaki ülkeler için *ESD* = 6, *Ortc* = 6 bulduğumuzu düşünürsek bu soruya evet deme eğilimi ağır basar. Ama gözden kaçırmamamız gereken önemli bir noktayı atlamayalım. Bu değerler bu ülkelerde canlı doğan bebekler içinde ilk beş yılda ölenlerin oranıdır. 15 AB ülkesinde doğan bebek sayıları eşit olsaydı bulduğumuz *AO* da güvenilir olurdu. Oysa durum böyle değil. Daha iyi bir *AO* bulmak istiyorsak bu oranların paydasını oluşturan canlı doğmuş bebek sayılarını elde etmemiz gerekir. Bu yeni verilerle *AO* şöyle hesaplanır. Her ülkenin bebek ölüm oranını o ülkede canlı doğan bebek sayısı ile çarpıp bunları toplarız. Bu toplamı bu ülkelerde doğan toplam bebek sayısına böleriz. Aslında bu bölmenin payı AB ülkelerinde ilk beş yıl içinde ölen bebek sayısıdır. Öyleyse oran AB’de ölen bebek sayısının doğan bebek sayısına oranından başka bir şey değildir. İşte gerçekten aradığımız *AO* da budur. Bu ülkeler arasında nüfusları (dolayısıyla canlı doğan bebek sayıları) görece yüksek olan Almanya, Fransa, İngiltere, İspanya, İtalya’nın bebek ölüm oranları 6 – 7 olduğuna göre aslında bulunacak *AO*’nun bizim bulduğumuz 6,4’ten çok farklı olmayacağı düşünülebilir. Yine de oranların *AO*’sunu hesaplarken bu noktaya özellikle dikkat etmemiz gerekir.

Bu örnekte ülkelerin bebek ölüm oranlarını canlı doğan bebek sayıları ile çarpıp bu oranları farklı ağırlıklarla ortalamaya katmış olduk. Böyle hesaplanan ortalamalara **ağırlıklı** ya da **tartılı ortalama** denir. Örneğimizde ağırlıklar ya da tartılar her ülkede canlı doğan bebek sayıları oldu.

Tartılı *AO*’nun genel gösterimi şöyledir:

$$\bar{Y} = \frac{\sum_{i=1}^k f_i Y_i}{\sum_{i=1}^k f_i} \quad n = \sum_{i=1}^k f_i \quad (n \text{ yerine } N \text{ konulursa sonuç } \mu \text{ olur})$$

f = her öbekteki gözlem sayısı (sıklık)

k = öbek sayısı

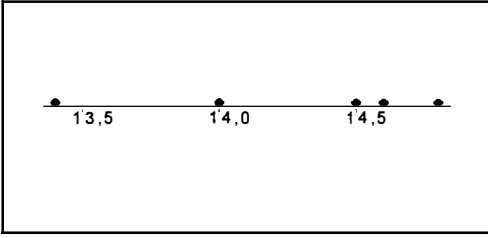
Öyleyse AO 'sunu hesaplayacağımız sayıların ne anlama geldiğini iyice bilmemiz gerekmektedir. Zaten istatistik yöntemlerini doğru kullanmanın belki de en önemli önkoşulu her sayıyı, her yöntemi, her sonucu, her zaman aklımızın süzgecinden geçirmek, sağduyumuzun denek taşına vurmaktır.

Şimdi de öbekli veri kümelerinde AO 'yu nasıl hesaplayacağımızı görelim. Bu amaçla Çizelge 8.2'deki yaş dağılımını örnek olarak alalım. O çizelgede yaşlar onar yıllık aralıklara ayrılmış, her aralığa düşen nüfus yüzdesi verilmişti. Her bir aralık içindeki dağılımın bakışık olduğu varsayılırsa, aralıkların orta noktaları o aralıktaki bütün yaşların temsil edicisi olur. Bundan yararlanarak, sözgelimi "0 - < 10 yaşları arasında bütün nüfusun % 22,8'i yer alır" demek yerine "bütün nüfusun % 22,8'i 5 yaşındadır" denebilir. Yukarıdaki varsayım geçerliyken AO hesabı bakımından bu ikisi arasında fark yoktur. Öyleyse her bir aralığın orta değerini o aralığın sıklığıyla çarpıp bulduğumuz toplamı sıklık toplamına bölerek AO 'yu elde edebiliriz. Bu örnekte olduğu gibi sıklıklar göreliyse bölme işlemine de gerek kalmaz. Dikkatli bir okuyucu yukarıda verilen tartılı ortalamayı hesapladığımızı ayırt etmiştir. Örneğimizdeki verilerle bu işlemleri şöyle yaparız:

$$AO = 5 \times 0,228 + 15 \times 0,232 + 25 \times 0,176 + 35 \times 0,134 + 45 \times 0,088 \\ + 55 \times 0,070 + 65 \times 0,072 = 26,2$$

Burada iki küçük ayrıntıyı açıklamak gerekli. İlki, çizelgedeki bilinmeyen kaleminin görelî sıklığını, yaşını bilmeyenler genellikle yaşlıdır düşüncesiyle, 60 + öbeğine ekledik. İkincisi 60 + öbeğini temsil eden yaş olarak 65'i aldık. Bulduğumuz sonuç bu iki varsayıma dayanıyor. 60 + yaş öbeğinin çok yaşlıları da içerdiği ileri sürülürse 65 yerine daha yüksek bir yaş seçilebilir. Sözgelimi 70 kullanılsa $AO = 26,56$; 75 kullanılsa $AO = 26,92$ olur. Bu örnekteki verilerin histogramı çizilirse çarpıklığın yönü görülebilir. Burada dağılımın büyük değerlere doğru olan kuyruğu daha uzundur. Şimdilik bu olguyu vurgulamakla yetinelim. Geometrik ortalamayı anlatırken bu noktaya döneceğiz.

Nasıl hesaplanacağını artık öğrenmiş bulduğumuz AO 'yu hangi tür gözlemler için kullanabiliriz? Her şeyden önce şunu söylemek gerekir ki AO 'nun anlamlı olabilmesi için gözlemlerin aralık ya da oran ölçeklerden biriyle ölçülmüş olması zorunludur. Ad ya da sıra ölçekleriyle ölçülmüş verilerin AO 'su anlamlı olmaz. Ayrıca AO , bakışık ya da bakışığa yakın bir dağılım gösteren veriler için daha anlamlı bir ortalama değildir. Yukarıda Marmara Bölgesi'nin ortalama hava sıcaklığını bulmuştuk. Bu verilerin dağılımının bakışık olup olmadığını **nokta çizimi** denilen şöyle bir çizimle kabaca görebiliriz (Bakışımın ölçülmesi daha sonra ele alınacaktır).



ÇİZİM 8.5 Marmara Bölgesi'ndeki beş merkezin hava sıcaklıklarının nokta çizimi

Bu çizime göre beş verinin dağılımı tam bakışık değildir ama aşırı bir çarpıklık da göze çarpmamaktadır. $Ortc = 14,5$ olduğunu düşünürsek bu örnekte bulunan ortalama hava sıcaklığını ($\sim 14,3$) bölge verilerini temsil etmede bir ölçüde başarılı sayabiliriz..

AO özellikle bakışık dağılımlarda çok başarılı bir ortalamadır, ama dağılımın çarpıklığı arttıkça AO da ortalama düzeyi gösterme işlevini yerine getiremez olur. Ayrıca birden çok tepeli dağılımlarda AO hiç de iyi bir temsil edici değildir. Bunu küçük bir örnekle gösterelim. Çizelge 8.4'teki Türkiye üniversitelerinde 1986'da öğretim üyesi başına düşen öğrenci sayısıyla ilgili iki tepeli dağılımı alıp ortalamamızı hesaplayalım.

$$AO = (1 \times 15 + 5 \times 25 + 3 \times 35 + 7 \times 45 + 9 \times 55 + 2 \times 65 + 1 \times 75) / 29 \\ = 43,45$$

buluruz. Oysa çizelgeden açıkça görüldüğü gibi üniversitelerimiz, öğretim üyesi başına biri 40'tan az öğrencili büyük kentlerdeki köklü kurumlar, öbürü 40'tan çok öğrencili yeni kurumlar diye iki farklı kümeye ayrılabilir. Bu iki kümeden her birinin ayrı ayrı ortalamalarını hesaplamak daha anlamlı olur. Bu yapılsa biri 20 – 30, öbürü 50 – 60 arasında iki ortalama bulunur. Yaklaşık 43 kişilik genel bir ortalama her iki küme için de yanıltıcı bir fikir verir.

AO 'nun ne tür veriler için uygun olduğunu gördükten sonra şimdi de bazı önemli özelliklerine değinebiliriz.

Yukarıda AO 'nun bakışık dağılımlar için çok uygun olduğundan söz edilmişti. Bunun bir uzantısı olarak, veriler arasında aşırı büyük ya da aşırı küçük olağan dışı değerlerin AO 'yu ciddi ölçüde etkilediklerini söyleyebiliriz. Bir basketbol takımına çok uzun yeni bir oyuncunun katılması takımın boy ortalamasını yukarı doğru çeker. İşte AO 'nun bu özelliğinden dolayı, yarışmacıların başarılarının, hakemlerin teker teker verdikleri puanların ortalaması alınarak saptandığı buz pateni, trampelen atlama, jimnastik gibi

spor dallarında, hem en düşük, hem en yüksek puanları veren hakemlerin yargıları ortalama hesabına katılmaz. Böylece herhangi bir hakemin verebileceği aşırı yüksek ya da aşırı düşük puanların ortalamayı etkilemesi engellenmiş olur.

Çizim 8.5'e yeniden bir göz atacak olursak uç değerlerdeki değişmelerin ortalamayı nasıl etkilediğini görebiliriz. Bu çizim bakışık veri dağılımları için çizildiğinden AO , en küçük ve en büyük gözlem değerlerinin tam ortasına düşmektedir. Öteki gözlem değerleri aynı kalırken bir tek gözlem değerinin aşırı büyüdüğünü, bunun en yüksek değerden çok daha yukarılara çıktığını düşünelim. Bu değişiklik çizimi şöyle etkiler. En büyük gözlem değerini gösteren yatay doğru yukarı kayar, ortalamaları gösteren M , eğrisi de bakışıklığını bozarak yeni duruma uyum sağlar, üst yarısı daha çok yukarı doğru uzar. Böylece bütün ortalamalar bu değişiklikten az ya da çok etkilenirler. Çizimde adı geçen dört ortalama arasında en çok etkilenen kareli ortalama, aritmetik ortalama onu izler. Buradan da anlaşılacağı gibi aşırı değer varlığı bakışımı bozar. Dolayısıyla bu iki olgu bir arada düşünülmelidir.

Bir işyerinde çalışanların gelirlerinin dağılımı, yüksek ücret alan birkaç üst düzey yöneticinin varlığı yüzünden bakışımından uzak, büyük değerlere doğru sağa çarpık olabilir. Bu durumda AO , o işyerinde çalışanların çoğunun gelirinin üstünde çıkabilir. Bu da gerçekleri yansıtmaktan uzak bir görünüm çizebilir. Toplu sözleşme görüşmelerinde işyerindeki ortalama ücretten söz edilirken çok dikkatli olmak gerekir. Buradaki ortalamanın AO mu yoksa başka bir ortalama mı olduğunu iyice anlamak bizi yanılgılardan korur. Aynı gerekçeyle hanehalkı başına ortalama gelir ya da kişi başına gelir gibi kavramlara da kuşkuyla yaklaşılmalıdır. Çünkü bu değerler de yalın birer AO 'dan başka bir şey değildir, üstelik hesaplanmalarında kullanılan verilerin bakışık olduğu da pek söylenemez. Bu tür gelir ortalamaları aslında gerçek gelir düzeyini genellikle abartır. Gelir dağılımları birbirine benzer iki ülke için kişi başına gelirlerine bakılarak karşılaştırma yapmak pek sakıncalı olmasa bile aynı karşılaştırma gelir dağılımındaki çarpıklıkları farklı iki ülke için gönül rahatlığıyla yapılamaz. Çünkü bunların kişi başına gelirleri aynı bile olsa, gelir dağılımı daha bakışığa yakın ülkenin gelir düzeyinin daha iyi olduğunu söylemek için fazla zorlanmak gerekmez.

AO 'nun bir başka özelliği, tanımı gereği, gözlem sayısı ile AO 'nun çarpımının gözlem değerlerinin toplamını vermesidir. Bu özellik uygulamada çok işe yarar. Daha önce de söz ettiğimiz gibi her aşçı, AO kavramıyla hiç karşılaşmamış bile olsa bu özelliği kullanır. Bir kişilik yemeğe konacak malzemenin miktarını deneyimleriyle bildiği için o gün o yemekten kaç kişilik pişirilecekse bu sayıyı ortalama malzeme miktarıyla çarpması, top-

lam gerekli malzeme miktarını bulmasına yeter. Bu özelliği şöyle gösterebiliriz:

$$\mu = \Sigma Y_i / N \quad \Sigma Y_i = N\mu$$

Gözlem değerlerinin AO 'dan farklarının cebirsel toplamının sıfıra eşit olması, AO 'nun bir başka özelliğidir. Dönem içi sınav notu düşük olan her öğrenci bu özelliğin ayrımındadır. Ortalama başarı notunu, diyelim 50'yi tutturabilmesi için dönem sonu sınavından 50'nin üzerinde not alması gerektiğini bilir. Dönem içiyle dönem sonu sınavının tartılarının aynı olduğunu varsayarsak, başarı notunun tam 50 olabilmesi için, dönem içi notu elliden ne kadar düşükse, dönem sonu notu da elliden o kadar büyük olmalıdır. Tartıların farklı olması durumunda, bu farklar da tartılarla çarpılmalıdır. AO 'nun tanımından yola çıkarak bu özellik kolayca gösterilebilir.

Aynı özelliği göstermenin bir başka yolu da gözlem değerlerini AO 'dan küçükler, AO 'dan büyükler diye iki öbeğe ayırıp, gözlem değerleri ile AO arasındaki farkları da u ile göstermektir. Gözlemlerden biri ya da bazıları tam AO 'ya eşitse bu gözlem(ler)i iki öbekten birine katmakta sakınca yoktur, bunun için $u = 0$ 'dır.

$$\mu = \Sigma Y_i / N \quad \Sigma Y_i = N\mu \quad \Sigma Y_i - N\mu = 0 \quad \Sigma Y_i - \Sigma \mu = 0 \quad \Sigma (Y_i - \mu) = 0$$

$$Y_i < \mu \text{ iken} \quad u_i = \mu - Y_i \quad i = 1, \dots, k$$

$$Y_i > \mu \text{ iken} \quad u_i = Y_i - \mu \quad i = k + 1, \dots, b$$

Bu durumda AO şöyle gösterilebilir:

$$\sum_{i \leq k} u_i = \sum_{k < i \leq b} u_i$$

$$\mu = \frac{\sum_{i \leq k} Y_i + \sum_{k < i \leq b} Y_i}{k + b}$$

Bu yeni tanımdan

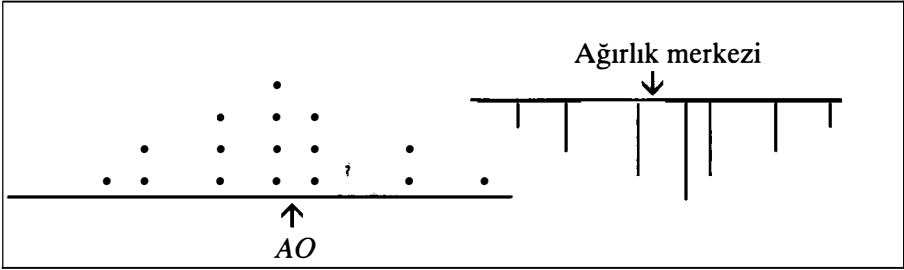
$$(k + b)\mu = \sum_{i \leq k} Y_i + \sum_{k < i \leq b} Y_i = \sum_{i \leq k} (\mu - u_i) + \sum_{k < i \leq b} (\mu + u_i)$$

$$= k\mu - \sum_{i \leq k} u_i + b\mu + \sum_{k < i \leq b} u_i = (k + b)\mu - \sum_{i \leq k} u_i + \sum_{k < i \leq b} u_i$$

işlemleri yardımıyla şu sonuç elde edilir:

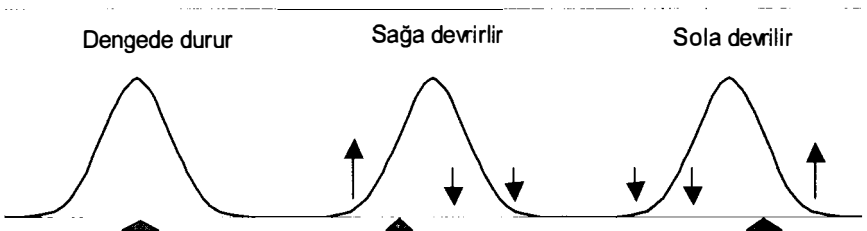
$$\sum_{i \leq k} u_i = \sum_{k < i \leq b} u_i$$

Bu özellik çocuk parklarındaki tahterevallilere benzetilebilir. Bu oyunda ortadaki destek noktası AO 'ya, iki çocuğun oturduğu yerlerin destek noktasından uzaklıkları gözlem değerlerine, çocukların ağırlıkları da tartılara karşılıktır. Bu benzetme daha genelleştirilerek denebilir ki, AO , gözlem değerleri dağılımının ağırlık merkezidir. Gözlem değerlerinin AO 'dan farkları, fizik derslerinden bildiğimiz kuvvetlerin uygulama noktalarının ağırlık merkezinden uzaklıklarına; gözlem değerlerinin tartıları ya da sıklıkları, bu kuvvetlere karşılıktır. Çizim 8.7 bu benzerliği varsayımsal bir örnek üzerinde göstermektedir.



ÇİZİM 8.7 Aritmetik ortalama ile ağırlık merkezi benzerliği

Eğer bir gözlem dağılımını, özgül ağırlığı her yerinde aynı olan bir metal tabakaya çizerek kesersek, bu kesilmiş parçanın tabanı üzerindeki noktalardan yalnız AO 'ya karşılık olanı bu metal parçasını dengede tutabilecek noktadır. Taban üzerindeki öteki bütün noktalarda metal parçası AO 'nun (ağırlık merkezinin) bulunduğu yana devrilir. Çizim 8.8'de bu durum gösterilmektedir.



ÇİZİM 8.8 Aritmetik ortalama denge noktasına karşılıktır

AO 'nun en çok kullanılan özelliklerinden biri de gözlem değerlerinin AO 'ya göre farklarının karelerinin toplamı, AO dışındaki herhangi bir değere göre farklarının karelerinin toplamından daha küçük olmasıdır. Bu özellik **standart sapma**, **varyans** gibi yaygınlık ölçülerinin hesaplanmasında, **bağlanım (regresyon)** denkleminin bulunmasında büyük önem kazanır.

Bu özelliği şöyle gösterebiliriz. Gözlem değerlerinin herhangi bir değere (M) göre farklarının kareleri toplamı T olsun. Bu durumda M değiştikçe T de buna bağlı olarak değişir, demek ki T , M 'nin fonksiyonudur. Eğer $T = f(M)$ fonksiyonunun en düşük (minimum) noktasındaki M değerinin AO 'ya eşit olduğunu gösterebilirsek bu özelliği kanıtlamış oluruz.

$$T = \sum (Y_i - M)^2 = \sum (Y_i^2 - 2MY_i + M^2) = \sum Y_i^2 - 2M \sum Y_i + NM^2$$

Bu fonksiyonun en düşük noktasında M 'ye göre türevi sıfıra eşittir, ayrıca ikinci türevi de artı değer taşır. Bu özellikleri taşıyan noktada

$$dT / dM = -2 \sum Y_i + 2NM = 0$$

$$2NM = 2 \sum Y_i$$

$$M = \left(\sum Y_i \right) / N = \mu$$

M değeri AO 'ya eşittir. İkinci türeve bakarsak

$$d^2T / dM^2 = 2N > 0$$

bunun da artı değerli olduğunu görürüz. Demek ki gözlem değerlerinin herhangi bir M değerinden farklarının kareleri toplamı, bu M değeri AO 'ya eşit olursa en düşük bulunmaktadır. Başka bir deyişle M yerine AO 'dan farklı bir değer koyarsak bu eşitlik bozulur.

Aynı sayıda gözlem değeri içeren iki dizinin karşılıklı gözlemleri ikiye ikiye toplanarak (ya da farkları alınarak) yeni bir dizi oluşturulursa, yeni dizinin AO 'su, önceki iki dizinin AO 'larının toplamına (ya da farkına) eşittir. Bu özellikten yararlanılarak bazı işlemler kısaltılabilir. Sözgelimi birim kârı, birim satış fiyatıyla birim maliyetin farkı biçiminde tanımlarsak, ortalama fiyat ile ortalama maliyeti bildiğimizde, ortalama birim kârı bu ikisinin farkını alarak kolayca bulabiliriz. Bu özelliği de cebirsel olarak göstereyim.

$$Z_i = X_i + Y_i \text{ ya da}$$

$$Z_i = X_i - Y_i \quad i=1,...,N$$

$$\sum Z_i = \sum X_i + \sum Y_i$$

$$\sum Z_i = \sum X_i - \sum Y_i$$

$$\Sigma Z_i / N = \Sigma X_i / N + \Sigma Y_i / N \quad \Sigma Z_i / N = \Sigma X_i / N - \Sigma Y_i / N$$

$$\mu_Z = \mu_X + \mu_Y$$

$$\mu_Z = \mu_X - \mu_Y$$

AO'nun bir başka özelliği, gözlem değerlerinin her biri sabit bir sayı ile toplanır ya da çıkarılırsa, veri kümesinin *AO*'su da bu sabit sayı kadar artar ya da eksilir. Bir öğretim üyesi sınava giren herkesin notunu 5 puan yükseltirse sınıfın not ortalaması da 5 puan yükselir. Bir malın bütün çeşitlerine aynı tutarda zam yapılırsa ortalama fiyat da yapılan zam kadar artar. Gözlem değerlerinin sabit bir sayıyla çarpılması ya da bölünmesi durumunda *AO* da aynı sabit sayıyla çarpılmış ya da bölünmüş olur. Sınav notlarının ya da fiyatların hepsinin %10 arttırıldığını varsayarsak, sınıfın not ortalaması ya da malların ortalama fiyatı % 10 yükselir. Bunların cebirsel gösterimi şöyle olabilir.

$$\text{Toplama-çıkarmalarda: } Z_i = Y_i \pm a \quad \Sigma Z_i = \Sigma Y_i \pm N a \quad \mu_Z = \mu_Y \pm a$$

$$\text{Çarpma-bölmelerde: } Z_i = b Y_i \quad \Sigma Z_i = b \Sigma Y_i \quad \mu_Z = b \mu_Y$$

Uyarı: Yukarıda gözlem değerleri %10 arttırıldığında $b = 1.1$ olduğunu unutmayalım. %10 indirim yapmak için $b = 0.9$ alınır.

Bir bütünün parçalarını oluşturan iki öbek gözlem değerinin ayrı ayrı *AO*'larının tartılı ortalaması alınırsa bütünün *AO*'su bulunur. Burada tartılar öbeklerdeki gözlem sayılarıdır. Aynı malı birden çok fabrikada üreten bir işletmenin genel maliyet ortalaması, her fabrikanın ortalama maliyetlerinin tartılı ortalamasıdır. Burada tartılar her bir fabrikanın üretim miktarlarıdır. Sözgelimi 2002 bütçesinde vergi gelirleri % 3,0, vergi dışı gelirler % 36,7 oranında tahminlerin üzerinde gerçekleşti. Bunların basit *AO*'su % 19,85 eder. Öyleyse (vergi + vergi dışı) toplam gelirler, tahminlerin % 20'ye yakın bir oranda üstünde gerçekleşmiştir diyebilir miyiz? Kesinlikle hayır. Çünkü iki tür gelirin toplam içindeki payları aynı değildir. Dolayısıyla bu paylara göre tartılı ortalama hesaplamak uygun olur. Bu yapıldığında bulunan gerçek sapma % 8,4 kadardır. Bunun az önce hesaplanan % 20'den hayli uzak bir oran olduğu açıktır.

Uzun süre aynı malı üreten bir işletmenin her ay rassallık kurallarına uyarak üretiminden belli büyüklükte bir örneklem seçtiğini, bu örneklem giren birimlerin belli bir özelliğini ölçüp bunun *AO*'sunu bulduğunu düşünelim. İlk ayın örneklem ortalamasıyla ikincinininkini toplayıp ikiye bölerek ilgililer iki aylık daha genel bir ortalama üretebilirler. Bu genel ortalama 2 katsayısını, üçüncü ayınkine de 1 katsayısını tartı olarak verip bu iki değerlerin tartılı *AO*'sunu hesaplarılarsa üç aylık üretim için bir ortalama bulunabilir. Bu sürdürülürse, i 'nci örneklem çekildiğinde bunun ortalamasının

tartısını 1 alıp, daha önceki $i - 1$ sayıdaki örneklem için hesaplanmış genel ortalamanın tartısını da $i - 1$ alarak tartılı bir AO hesaplanabilir, bu da i aylık üretimin genel ortalamasını verir. Bu yolla hesaplanan ortalamalar kalite denetimi uygulayan işletmelerde anakütle ortalaması olarak kullanılabilirler. Burada örneklemin büyüklüğünü değişmez varsaydık. Kuşkusuz örneklem büyüklüğü aydan aya değişebilir, ancak o zaman tartıların hesabında bunu da göz önüne almak gerekir.

Geometrik ortalama (GO)

Orta değerlerde yoğunlaşmış, bakışık dağılımlı veriler için AO 'nun çok iyi bir temsil edici olduğunu vurgulamıştık. Biraz çarpık gözlem değerleriyle karşılaştığımızda geometrik ortalamayı kullanmak uygun olabilir. Zaman içindeki artış-azalışların ortalamalarını bulmakta da GO 'dan yararlanabiliriz. Aşağıda bunları açıklayacağız ama önce genel ortalama gösteriminde GO 'nun özel konumuna bir göz atalım.

Ortalamalar genel formülünde r terimine 0 verilirse GO 'nun elde edileceğinden söz etmiştik. Oysa iş o denli kolay değildir. Çünkü bu işlem yapılırsa M_0 belirsizdir. Bu durumdan kurtulabilmek için r 'ye doğrudan sıfır değeri vermek yerine sıfıra yaklaşırken aldığı limit değerini kullanabiliriz. Önce genel formülün her iki yanının logaritmasını alalım, sonra da $r \rightarrow 0$ için limit hesaplayalım.

$$\log M_r = [\log(\sum Y_i^r) - \log N] / r = (\log N - \log N) / 0 \quad \log M_r = \frac{0}{0}$$

Bu durumda bir belirsizlikle karşılaştığımıza göre l'Hospital (opital) kuralını uygulayıp pay ile paydanın r 'ye göre ayrı ayrı türevlerini alır,

$$\lim_{r \rightarrow 0} \log M_r = \lim_{r \rightarrow 0} \sum Y_i^r \log Y_i / \sum Y_i^r$$

buluruz. Buradan da

$$\log M_0 = (\sum \log Y_i) / N$$

elde edilir. Şimdi de logaritmadan kurtulup başa dönelim.

$$M_0 = \left(\prod_{i=1}^N Y_i \right)^{1/N} = \sqrt[N]{\prod_{i=1}^N Y_i}$$

Bu GO formülüdür. Öyleyse GO , N sayıda gözlem değerinin çarpımının N 'nci dereceden kökü olarak ya da gözlem değerlerinin logaritmalama-

rının AO 'sunun ters logaritması olarak tanımlanabilir. Demek ki gözlemlere önce bir logaritma dönüştürmesi uygulanır. Dönüştürülmüş logaritmali değerlerin AO 'su bulunur. Daha sonra da bu ortalamanın ters logaritması alınır. Böylece GO 'ya ulaşılır. Bu dönüştürme işlemine öbür ortalamaları incelerken de başvuracağız. Aslında dönüştürmeyi çarpık bir dağılımı bakışık duruma getirmek için yaptığımızı düşünürsek durumu daha kolay algılayabiliriz. Bakışıklaştırılmış veri kümesinin, böyle kümeler için uygun olan AO 'su alınır da, gözlemler ters dönüştürmeyle ilk konumlarına çevrilirlerse, çarpık bir veri kümesi için uygun bir ortalama bulunmuş olur. Öyleyse hemen bu bulguyu yorumlayalım. Eğer çarpık bir veri kümesi logaritması alınarak bakışık yapılabiliyorsa burada GO kullanmak çok uygun olabilir.

GO hesaplaması için bir örnek verelim. Bir işletme 64 milyar TL ödeyerek aldığı bir makine için azalan bakiyeler yöntemine göre amortisman hesaplıyor olsun. Bu yöntemle göre ilk yıl alış bedelinin belli bir oranı kadar amortisman ayrılmakta, sonraki yıllarda kalan defter değeri üzerinden aynı oranda amortisman ayrılması sürdürülmektedir. Bu örnekte kullanılan oran $\frac{1}{2}$ olsun. Dolayısıyla bu makinenin defter değeri yıllar boyunca şöyle olacaktır:

Yıl	1	2	3	4	5	6	7
Defter değeri (mr TL)	64	32	16	8	4	2	1

Acaba bu yedi yıllık süre boyunca makinenin ortalama defter değeri kaç liradır? Eğer AO hesaplayacak olursak

$$(64+32+16+8+4+2+1) / 7 = 18,14 \text{ milyar TL}$$

bulunur. Bu değer iyi bir temsilci sayılabilir mi? Bunun için defter değerlerinin bakışım durumuna bir göz atmamız iyi olur. Çizim 8.9'a baktığımız zaman dağılımın bakışımdan çok uzak olduğunu hemen görebiliriz. Aritmetik ortalamanın sağında iki, solunda ise beş değer bulunmaktadır. Öyleyse aritmetik ortalama defter değerlerinin çoğundan büyük, yalnızca iki tanesinden küçüktür.

GO (milyar TL)	AO (milyar TL)
0	0
5	5
10	10
15	15
30	30
70	70

ÇİZİM 8.8 Bir makinenin defter değeri

Makinenin defter değerlerinin şöyle bulunduğunu biliyoruz.

$$64 \quad \frac{64}{2} \quad \frac{64}{2^2} \quad \frac{64}{2^3} \quad \frac{64}{2^4} \quad \frac{64}{2^5} \quad \frac{64}{2^6}$$

bu sayılar geometrik bir dizi oluşturmaktadırlar. Bu özelliği taşıyan bir veri kümesi, adı üstünde, geometrik bir ortalamayla daha iyi temsil edilemez mi? Deneyelim! Az önce, çarpık bir gözlem değerleri kümesi, logaritmaları alındığında bakışığa dönüşüyorsa *GO* bu durumda çok uygundur demiştik. Gözlem kümemizi oluşturan sayıların logaritmaları sırasıyla şunlardır. 1,806, 1,505, 1,204, 0,903, 0,602, 0,301 0. Logaritmali değerlerin nokta çizimi Çizim 8.10'da gösterilmiştir.

•	•	•	•	•	•	•	•	•
0,00	0,25	0,50	0,75	1,00	1,25	1,50	1,75	2,00

ÇİZİM 8.10 Makinenin defter değerinin logaritmaları

Bunların tam bakışık oldukları görülmektedir. Öyleyse bu bakışık dağılım için *AO* uygun olacaktır.

$$(1,806 + 1,505 + 1,204 + 0,903 + 0,602 + 0,301 + 0) / 7 = 0,903$$

Bu da beklendiği gibi tam ortadaki değerdir. 0,903'ün 8'in logaritması olduğunu bildiğimize göre, *GO* 8 milyar TL olarak bulunmuş olur. Aslında *GO* hesaplamaya karar verdiğimizde, logaritmaları hiç işe karıştırmadan defter değerleri üzerinden şu işlemi yapabiliirdik.

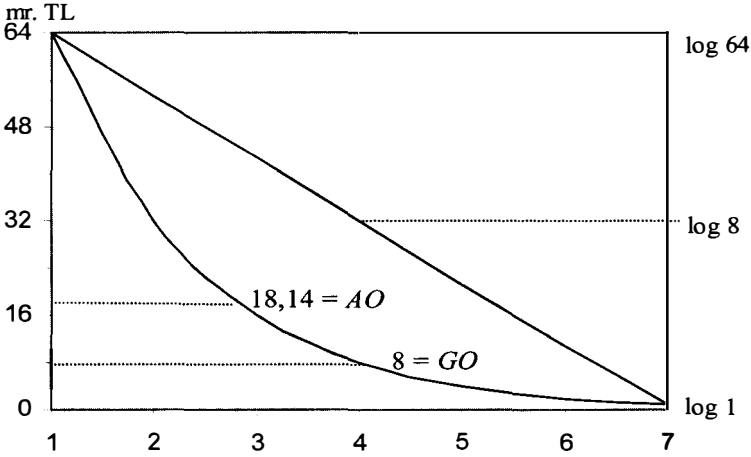
$$64 \times 32 \times 16 \times 8 \times 4 \times 2 \times 1 = 2\,097\,152 \quad \sqrt[8]{2\,097\,152} = 8 \text{ milyar TL}$$

Elbette bunu yaparken veri sayısının ya da gözlem değerlerinin büyük sayılara ulaşması hesaplama olanaklarımızı zorlayabilir, giderek olanaksız bile kılabilir. Son 30 yılın ulusal gelir rakamlarını hiç yuvarlatmadan birbiriyle çarptığınızı düşünün. Bunu hesap makinenize sığdıramazsınız. Ya da binlerce gözlemde oluşan bir kümenin değerler çarpımını alabilseniz bile, bu denli yüksek bir dereceden kökünü almanız elinizdeki makineyle olanaksız olabilir. Bu nedenle logaritmaya başvurmak sorunu çözecektir. Zaten *GO*'yu kavrayabilmek için logaritma kullandığımıza göre hesaplanmasında da logaritma kullanmak hiç de yadırgatıcı olmaz.

Uyarı: Logaritma alırken 64.000.000.000 yerine 64'ün logaritmasını aldık. Bunu işi kolaylaştırmak için yaptık. Eğer milyarları işin içine katsaydı logaritma değerlerimizin her birine 9 ekleyecektik. Toplam 7×9 kez daha büyük olacak, bunu 7'ye bölünce sonuç, 0.903 yerine 9.903 bulunacaktı.

Bunun da 8.000.000.000 sayısının logaritması olduğunu biliyoruz. Dolayısıyla milyar sıfırlarını atarak iş görmek bizi kısa yoldan doğru sonuca ulaştırdı.

Bu örneği zamanı da içeren bir başka çizimle, Çizim 8.11 ile bir kez daha inceleyelim. Makinenin defter değeri geometrik bir diziyle azalarak 64'ten 1'e inmektedir. Bu, *eğri* üzerinde AO ile GO işaretlenmiştir. Görüldüğü gibi GO yedi yıllık sürenin tam ortasına düşmekte, AO ise üçüncü yılın öncesindeki bir defter değerini göstermektedir. Bu çizimin asıl ilginç olan yönü, dizi terimlerinin logaritmalarının da gösterilmiş olmasıdır. 64'ten 1'e inen *doğru* üzerinde sıralanan bu değerler sağdaki eksende ölçülmektedirler. Bu logaritmali dizinin AO 'sunun da sürenin tam ortasına düştüğüne dikkat edelim. Demek ki logaritmali değerler tam bir bakışım içindedirler, dolayısıyla GO tam orta noktada yer almaktadır.

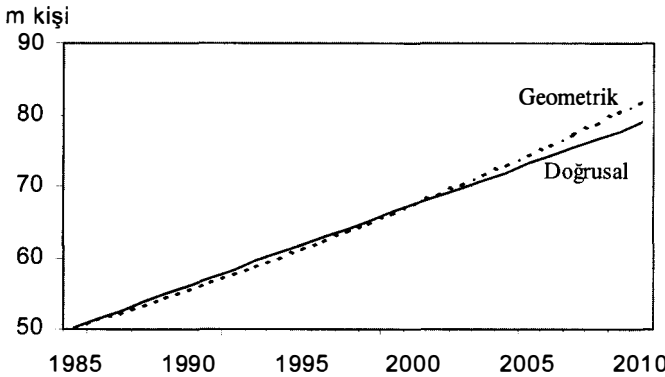


ÇİZİM 8.11 Makine defter değerlerinin AO , GO , logaritmali AO değerleri

Yukarıda incelediğimiz veri kümesinin belki de en önemli özelliği gözlem değerlerinin her birinin bir önceki değerini belli bir orana karşılık gelmesiydi. Gerçek yaşamda böyle dizilere ne kadar sık rastlanabilir ki geometrik ortalamaya gerek doğsun denebilir. Hiç de öyle değil! Dünyada bizim sandığımızdan çok daha fazla olay, sözü edilen özelliğe uymaktadır. Hemen akla gelebilecek bir kaç örnek vermek gerekirse, bileşik faiz uygulamaları, göçler dışlanır da doğum-ölüm oranlarının zaman içinde değişmediği varsayılırsa nüfus artışları (bunu yıllar önce Malthus dememiş miy-

di?), biyolojik her türlü çoğalma, ivme en yetkin örnekler olarak sıralanabilir. Kaldı ki biraz sonra göreceğimiz gibi bir gözlem değerleri dizisinin en uygun ortalamasının *GO* olması için ille de her gözlem değerinin, örneğimizdeki $\frac{1}{2}$ gibi değişmez bir sayıyla çarpılmış olması gerekmez. Eğer biz gözlem değerleri arasındaki değişme oranlarıyla ilgileniyor, bunların ortalama düzeyini arıyorsak *GO* yine en uygun araçtır diyebiliriz.

Bunu bir örnek üzerinde görelim. Türkiye'nin nüfusu sayım yılları olan 1985'te 50,306 milyon, 2000'de 67,469 milyon bulunmuştur. Bunlar yıl ortası nüfuslarıdır. Acaba ara yılların nüfuslarını bu iki rakamdan yola çıkarak kestirebilir miyiz? Dahası 2000 sonrası yıllar için de kestirimlerde bulunabilir miyiz? Bunun için Çizim 8.12'ye bir göz atmalıyız.



ÇİZİM 8.12 Türkiye nüfusunun gelişimi
DİE, 2002a:44'teki verilerle düzenlenmiştir.

Bu çizimde nüfus artışları biri doğrusal, biri geometrik iki ayrı seçenekle gösterilmiştir. 1985 nüfusu 50,306 milyon, 2000 nüfusu da 67,469 milyon kişidir. 1985-2000 arası yıllık ortalama nüfus artışı 1,1442 milyon kişi olarak hesaplanmış, bütün yıllarda nüfusun her yıl bu kadar artacağı varsayımıyla kesiksiz doğru çizilmiştir. Kesikli eğri ise 1985-2000 nüfus sayımlarından hesaplanan yıllık ortalama artış *oranına* göre nüfus tahminlerini gösterir. Gerçek bunlardan ikincisine daha yakındır. Öyleyse nüfusun her yıl aşağı yukarı aynı oranda arttığı varsayımını benimseyerek 1985-2000 sayımları arasındaki yıllar için bu değişmez oranı bir *GO* gibi hesaplayabiliriz. Daha sonra 1985 yılı nüfusunu her yıl bu oranda artırarak (ya da 2000 nüfusunu her yıl bu oranda azaltarak) 1985-2000 arasındaki yılların nüfuslarını kestirebiliriz. Aynı yolla 2010 yılına kadar nüfusumuzun

önkestirimini yapabiliriz. Çizimde hem doğrusal, hem oransal yolla yapılan kestirimler yer almaktadır. Doğrusal artış yolu 1985-2000 arasında daha yüksek, 2000'den sonra daha düşük değerler vermektedir. İki kestirim dizi si doğal olarak 2000 yılında kesişmektedir.

Hesaplamak istediğimiz ortalama yıllık nüfus artış oranına a diyelim. 1986 yılı nüfusu $50,306(1+a)$ ile, 1987 nüfusu $50,306(1+a)^2$ ile gösterilir. Aynı akıl yürütmeyle 2000 yılı nüfusu da $50,306(1+a)^{15}$ olur. Oysa bu sayının 67,469 milyon kişi olduğunu biliyoruz. Demek ki $50,306(1+a)^{15} = 67,469$ 'dur. Buradan a oranını hesaplayabiliriz.

$$(1+a)^{15} = 67,469 / 50,306 = 1,341172027$$

Bu sayı bize Türkiye nüfusunun bu on beş yılda toplam $\sim \% 34,12$ arttığını gösterir. Bunun bir yıllık karşılığı ise 15'e bölünerek değil, 15. kökü alınarak bulunur.

$$(1+a) = \sqrt[15]{1,341172027} = 1,019762332$$

$$a = 0,019762332$$

Demek ki bu on beş yıllık süre içinde nüfusumuz yılda ortalama $\sim \% 1,976$ artmıştır. Elbette her yıl nüfusun tam bu oranda değiştiği savını ileri sürüyor değiliz. Bazı yıllar bunun altında, bazı yıllar bunun üstünde artış gerçekleşmiştir, ama bu oranların ortalaması $\sim \% 1,976$ 'dır. Buna göre çeşitli yılların nüfusları şöyle kestirilebilir:

1986 :	$50,306 (1,019762332)^1 = 51,300$ milyon kişi
:	:
1990 :	$50,306 (1,019762332)^5 = 55,477$ milyon kişi
:	:
1995 :	$50,306 (1,019762332)^{10} = 61,180$ milyon kişi
:	:
2000 :	$50,306 (1,019762332)^{15} = 67,469$ milyon kişi
:	:
2010 ..	$50,306 (1,019762332)^{25} = 82,053$ milyon kişi

Dikkatli bir okuyucu burada geometrik ortalama kullandığımızı anlamış olabilir. Gerçekten de $50,306 (1+a)^{15} = 67,469$ yerine yıllık artış oranlarını a_i ($i = 1, \dots, 15$) ile gösterip

$$50,306 (1+a_1) (1+a_2) \dots (1+a_{15}) = 67,469$$

yazabiliriz. Öyleyse $(1+a)^{15} = (1+a_1)(1+a_2) \dots (1+a_{15})$ eşitliğinden $1+a$ değerinin $1+a_i$ değerlerinin geometrik ortalaması olduğu görülebilir. Çünkü yapılacak iş

$$1+a = \sqrt[15]{\prod_{i=1}^{15} (1+a_i)}$$

eşitliğini kullanarak $1+a$ değerini bulmaktır. Bu eşitliğin de GO tanımıyla aynı olduğu ortadadır. Burada dikkatli okuyucunun kolayca anlamış olabileceği gibi GO 'su hesaplanan değişken a değil, $(1+a)$ 'dır.

Meraklı okuyucular için şunu da ekleyelim ki nüfus istatistikçileri büyüme oranını yukarıdaki gibi değil, sürekli büyüme kavramını kullanarak yaparlar. Matematik derslerinizden anımsayacaksınız, sürekli büyüme, $(1+a)$ gibi bir sayıyla değil, e^b gibi bir sayıyla gösterilir. Burada e doğal logaritma tabanı, b ise yeni hesaplayacağımız yıllık büyüme oranıdır. Öyleyse

$$e^{15b} = 67,469 / 50,306 = 1,341172027$$

$$15b = \ln 1,341172027 = 0,293543878$$

$$b = 0,019569591$$

Demek ki bu yeni hesaplamayla yıllık nüfus artış oranı $\sim \% 1,957$ bulunmaktadır. a ile b eşit değildir ama dikkat edilirse $(1+a)$ ile e^b 'nin birbirine eşit oldukları görülecektir. Gerçekten ikisi de tam 1,019762332 eder. Dolayısıyla gerek a , gerek b kullanılarak yapılacak kestirimler tamı tamına aynı kişi sayılarını verecektir. Bir örnek olması için 2010 yılının nüfusunu sürekli büyüme oranıyla kestirirsek

$$50,306 e^{25(0,019569591)} = 50,306 e^{0,489239775} = 50,306 (1.631075764) = 82,053$$

milyon kişi daha önce bulduğumuz nüfusla tam bir uyuşum içindedir. Daha genel olarak şunu yazabiliriz.

$$Y_0(1+a)^n = Y_n$$

$$Y_0 e^{nb} = Y_n$$

Bu eşitliklerinden ilki kesikli, ikincisi sürekli büyüme için kullanılır. İlkinde 10 tabanına, ikincisinde ise e tabanına göre logaritma alınır.

Buraya kadar öğrendiklerimizi bir kaç örnek üzerinde sınavalım. İlk örneğimiz aşınma payı düşülmemiş ulusal ürün değerlerimizin büyüme oranlarıyla ilgili. Söz konusu büyüme oranları (% olarak)

1998: 3,9

1999: - 6,1

2000: 6,3

2001: - 9,4

kadardır (DİE,2002a:643). Acaba dört yıllık ortalama büyüme oranı % kaçtır? Görüldüğü gibi bu kez elimizde başlangıç ve bitiş dönemlerinin büyüklükleri yoktur. Yalnızca ekonominin yıllık büyüme hızları vardır. Bunlara göre 1998 yılının değeri 1997 yılının 1,039 katına ulaşmıştır. 1999 yılının büyüme oranı eksi işaretli olduğuna göre bu dönemde bir büyümeden değil bir küçülmeden söz etmek daha doğru olur. Öyleyse 1999 değeri, 1998 değerinin $(1 - 0.061) = 0,939$ katına eşittir. Bu yolla 1998 değerinin kaç katının 2001 değerine eşit olduğunu bulabiliriz.

$$Y_{1998} (1,039)(0,939)(1,063)(0,906) = Y_{2001}$$

Bu dört oranın yerine onların geometrik ortalaması olan $1+a$ oranını geçirirsek

$$Y_{1998} (1+a)^4 = Y_{2001}$$

eşitliğini yazabiliriz. Buradan

$$(1+a)^4 = (1,039)(0,939)(1,063)(0,906) = 0,9396$$

bulunur. Dolayısıyla

$$1 + a = \sqrt[4]{0,9396} = 0,9845 \quad a = \% - 1,55$$

ortalama büyüme, daha doğrusu küçülme hızı elde edilir. Yani Türkiye’de 1998-2001 döneminde üretim yılda ortalama % 1,55 kadar düşmüştür.

Fiyat artışlarına ilişkin başka bir örnek alalım. İstanbul ücretliler geçinme indeksi 2002 yılında ayda ortalama % 3,18; 2003’ün ilk üç ayında da sırasıyla % 0,1; 1,4; 2,2 artış göstermiştir (İTO, 2003a:8, 28). Bu 15 ay boyunca ortalama aylık fiyat artışı % kaçtır?

2002 yılında 12 ay boyunca fiyatların hep % 3,18 arttığını varsaymanın bu soruyu yanıtlamak bakımından hiç bir sakıncası yoktur. Öyleyse 15 aylık fiyat artış oranını $(1,0318)^{12} (1,001) (1,014) (1,022)$ çarpımı olarak gösterebiliriz. Bunun aylık ortalaması

$$\sqrt[15]{(1,0318)^{12}(1,001)(1,014)(1,022)} = \sqrt[15]{1,5157} = 1,028$$

olarak bulunur. Demek ki bu 15 ay boyunca aylık ortalama tüketici fiyat artış oranı % 2,8 olmuştur. Burada **tartılı** *GO*’nun bir örneğini incelemiş bulunuyoruz. Tartının, *AO*’daki gibi gözlem değerinin bir çarpanı değil, kuvveti olarak kullanıldığını da böylece görmüş olduk. Gözlem değerlerinin sıklık sayıları farklı olan veri kümeleri için de benzer bir yol izlenerek sıklık sayıları ilgili gözlemlerin kuvveti olarak

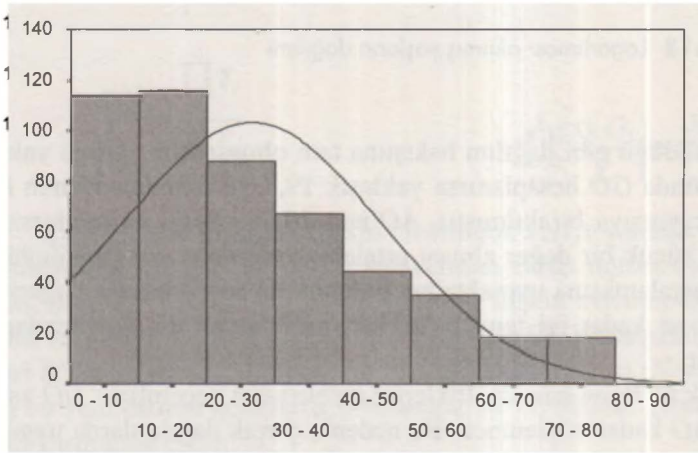
$$M_0 = \sqrt[n]{\prod_{i=1}^k Y_i^n}$$

$$N = \sum_{i=1}^k f_i$$

k : gözlem değerlerinin öbek sayısı

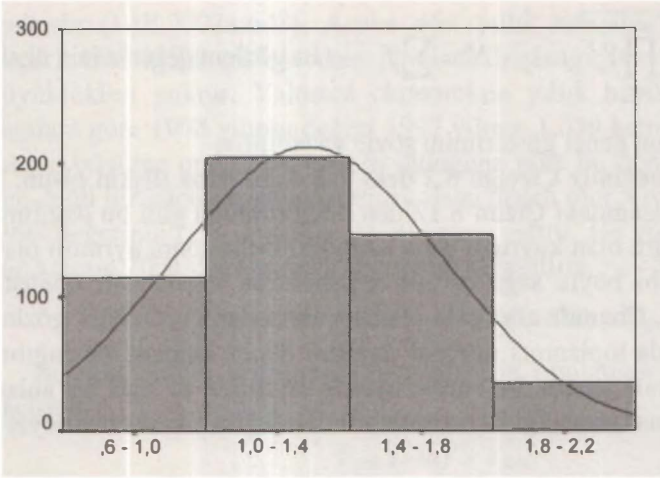
alınır. Bunun genel gösterimini şöyle yazabiliriz.

Son örneğimiz Çizelge 8.2'deki yaş dağılımına ilişkin olsun. Bu çizelgeden de, aşağıdaki Çizim 8.12'den de görüldüğü gibi bu dağılımın büyük yaşlara doğru olan kuyruğu daha uzundur. Daha sonra ayrıntılı olarak göreceğimiz gibi böyle sağa çarpık dağılımlarda logaritmali dönüştürme işe yarayabilir. Çizimde alan çubuklarının arasından geçen eğri, gözlem değerleri ortalarında toplanmış **normal dağılım** denen bakışık bir dağılımı gösterir. Histogram az çok bu eğriye uyarsa dağılımın da özel bir anlamda normal dağılıma benzediği ileri sürülebilir. Örneğimizde durum böyle değildir.



ÇİZİM 8.12 Türkiye'nin 1990 nüfusu yaş dağılımı

Dağılımı biraz olsun bakışığa yaklaştırmak amacıyla gözlem değerlerinin (aslında Çizelge 8.2'deki öbek aralıklarının orta noktalarının) logaritmaları alınarak Çizim 8.13'teki histogram elde edilmiştir.



ÇİZİM 8.13 Logaritması alınmış yaşların dağılımı

Görüldüğü gibi dağılım bakışına tam olmasa bile epeyce yaklaşmıştır. Bu durumda GO hesaplanırsa yaklaşık 19,3 yıl bulunur. Bunun doğrulanması okuyucuya bırakılmıştır. AO 'nun 26,56 çıktığı anımsanırsa GO 'nun bundan küçük bir değer alması ortalamaların, daha önce gördüğümüz, büyüklük sıralamasına uymaktadır. Bulunan bu son ortalama değer in yaş dağılımını ne kadar iyi temsil ettiğini yorumlamak da okuyucudan beklenmektedir.

Artık GO 'nun bazı özelliklerini incelemeye geçebiliriz. GO aşırı değerlerden AO kadar etkilenmez. Bu nedenle çarpık dağılımlarda uygulanabilir. Ancak çarpıklığın logaritma alınarak düzeltilebilmesi gerekir.

GO 'nun gözlem sayısı N kadar kuvveti alınır sa gözlem değerlerinin çarpımına ulaşılır. Bu özellik AO 'nun N ile çarpılınca gözlem değerlerinin toplamını vermesi özelliğine benzetilebilir. Eğer gözlem değerleri yerine logaritmaları kullanılırsa aynı özellikten söz ettiğimiz daha iyi anlaşılır.

$$GO^N = \prod Y_i \quad N \times \log GO = \sum \log Y_i$$

Bu özelliği yukarıda tartılı geometrik ortalamayı anlatırken kullanmıştık. 2002 yılı aylık ortalama fiyat artışının 12. kuvvetini almış, $(1,0318)^{12}$ değerini 2002 yılı aylık artışlarının çarpımı olarak kullanmıştık.

Geometrik ortalamanın bir başka özelliği, gözlem değerlerinin teker teker geometrik ortalamaya oranlarının çarpımının bire eşit olmasıdır.

$$\prod \left(\frac{Y_i}{GO} \right) = \frac{\prod Y_i}{GO^N} = \frac{GO^N}{GO^N} = 1$$

Bu özelliğin uzantısı olarak şu özellikten de söz edilebilir. Geometrik ortalama kendinden küçük gözlem değerlerine teker teker oranlanır da bu oranların çarpımı alınırsa bulunan değer, geometrik ortalamadan büyük değerlerin geometrik ortalamaya oranlarının çarpımına eşit çıkar. Bunu da şöyle gösterebiliriz. GO 'dan küçük gözlem değerlerinin sayısı k , büyüklerin sayısı da b olsun.

$$GO = \sqrt[k+b]{\left(\prod_{i=1}^k Y_i \right) \left(\prod_{j=1}^b Y_j \right)} \quad (GO)^{k+b} = \left(\prod_{i=1}^k Y_i \right) \left(\prod_{j=1}^b Y_j \right)$$

$$\frac{GO^k}{\prod_{i=1}^k Y_i} = \frac{\prod_{j=1}^b Y_j}{GO^b} \quad \prod_{i=1}^k \left(\frac{G.O.}{Y_i} \right) = \prod_{j=1}^b \left(\frac{Y_j}{G.O.} \right)$$

Aynı gözlem birimine ilişkin birden çok veriye **eşlenik veri** adı verilir. Bunlar aynı öğrencinin çeşitli sınavlardan aldığı notlar, aynı ülkenin üretim, dışalım, dışsatım değerleri, aynı yıla ilişkin evlenme-boşanma sayıları gibi verilerdir. Eşlenik veri içeren iki dizinin (X, Y) karşılıklı gözlem değerleri ikişer ikişer çarpılıp yeni bir dizi (Z) oluşturulursa $(X_i \times Y_i = Z_i \quad i = 1, \dots, n)$ bu yeni dizinin geometrik ortalaması, ilk iki dizinin geometrik ortalamaları çarpımına eşittir.

$$\prod (X_i Y_i) = \prod Z_i \quad \prod X_i \prod Y_i = \prod Z_i \quad \sqrt[N]{X_i} \sqrt[N]{Y_i} = \sqrt[N]{Z_i}$$

$$GO_X GO_Y = GO_Z$$

Bir veri kümesindeki gözlem değerlerinin terslerinin geometrik ortalaması, asıl gözlem değerlerinin geometrik ortalamasının tersine eşittir.

$$\sqrt[N]{\prod \frac{1}{Y_i}} = \sqrt[N]{\frac{1}{\prod Y_i}} = \sqrt[N]{\frac{1}{GO^N}} = \frac{1}{GO}$$

Bütün bu özellikleri bilmek, yeri geldiğinde kullanabilmek çeşitli çözümler sırasında önemli kolaylıklar sağlar.

Ters (harmonik) Ortalama (TO)

Gözlem değerlerine duyarlı ortalamalardan biri de ters ortalamadır. Buna harmonik ortalama da denmektedir. Bu adı almasının nedeni, eski çağlardan bu yana incelenen müzik-matematik ilişkisine bağlanabilir. Pisagorcular 9 birim uzunluğunda bir telin çıkardığı sesin, 12 birimlik bir telinkine göre dörtte bir oranında daha tiz olduğunu bulmuşlardı. Bu tizlik 8 birimlik bir telde beşte bir oranında, 6 birimlik bir teldeyse bir oktav oluyordu. Burada bizi ilgilendiren 12 ile 6'nın AO 'sunun 9, TO 'sunun 8 olmasıdır. Bu uyum (ahenk, harmoni ya da armoni) nedeniyle ters ortalama harmonik ortalama denmiştir. Yine aynı nedenle geometrik anlamdaki küp, 6 yüzü, 8 köşesi, 12 kenarıyla harmonik bir cisim sayılmıştır. Ünlü gökbilimci Kepler gezegenlerin hızlarıyla müzikteki armoni arasında ilişki kurmuş, her bir gezegen için birer beste yapmıştır. (Pappas, 1991:132, 1993:86, 1999: 191-195, 2002:192).

Ortalamalar genel gösteriminde r yerine -1 koyarsak TO 'yu elde ederiz.

$$M_{-1} = \left(\frac{\sum Y_i^{-1}}{N} \right)^{-1} = \frac{N}{\sum Y_i^{-1}} = \frac{N}{\sum (1/Y_i)} = TO$$

Görüldüğü gibi TO , gözlem değerlerinin terslerinin AO 'sunun tersidir. GO 'da olduğu gibi burada da gözlem değerleri dönüştürülerek daha bakışık biçime getirilmekte, bu bakışık dağılımın AO 'su alınmakta, sonra bu ortalamanın yeniden tersi alınarak gözlem değerlerinin ilk biçimine geri dönülmektedir. Yalnız burada uygulanan dönüştürme logaritma değil, sayıların tersini almadır. Bu kadar karışık bir ortalama ne gerek var diye düşünebilirsiniz. Aynı kanıdaki şu satın alma görevlisinin yaptığı ortalama hesabına bir göz atın, bakalım ona ne diyeceksiniz?

Bir işletmeye çok ivedi olarak gereken bir mal satın alınacaktır. Bu iş için 2 milyar TL para ayrılmıştır. Satın alma görevlisi hemen kolları sıvar, bu malı bulunduran iki işletmeyi arayarak şu bilgileri edinir. İlk işletme bu malı 25'erlik kutularda satmakta, kutu başına 10 milyon TL fiyat istemektedir. Ancak ellerinde 100 kutu mal vardır. İkinci işletmede stok boldur ama 20'erlik kutular 10 milyon TL fiyatla satılmaktadır. Görevli her iki işletmeden de 100'er kutu mal alarak ivedi gereksinimi karşılar. İşini başarmanın mutluluğuyla hemen bir hesap yapar. İlk işletmeden tanesi (10 milyon TL / 25) 400 bin TL'den, ikinciden de tanesi (10 milyon TL / 20 tane) 500 bin TL'den aldığı malların ortalama fiyatı $(400 + 500) / 2 = 450$ bin TL eder. Kendisinden haber bekleyen üstüne durumu anlattığında hesapta bir yanlışlık olduğu, ortalama fiyatın 450 bin TL'den düşük olması gerektiği yanıtını alır. Kim yanılmaktadır?

Kuşkusuz yanılan satın alma görevlisidir. Çünkü ilk işletmeden 2500, ikinciden 2000, toplam 4500 tane mal almış, 2 milyar TL ödemiştir. Öyleyse malların tanesi ortalama $(2 \text{ milyar TL} / 4500 \text{ tane}) 444.444, \bar{4}$ TL'dir. Satın alma görevlisi nerede yanlış yapmıştır? Tanesi 400 ya da 500 bin TL olan malların her ikisinden eşit sayıda alınsaydı görevlinin hesabı doğru olacaktı. Ama bu sayılar eşit olmadığı için tartı kullanmak gerekirdi. Gerçekten de

$$(2500 \text{ tane} \times 400 \text{ bin TL} + 2000 \text{ tane} \times 500 \text{ bin TL}) / 4500 \text{ tane} = 444.444, \bar{4} \text{ TL / tane}$$

bulunur. Demek ki AO doğru kullanılıncaya doğru sonuç bulunmaktadır.

Öyleyse TO 'ya ne gerek var? Bu soruyu yanıtlamadan önce TO kullanarak bu örneği bir kez daha inceleyelim. Örneğimizde ortalama fiyat aradığımıza göre Y_i gözlem değerleri fiyatlardır. Yukarıdaki formülü kullanarak

$$TO = \frac{N}{\sum \frac{1}{Y_i}} = \frac{2}{\frac{1}{40} + \frac{1}{50}} = \frac{2}{0.045} = 444, \bar{4} \text{ bin TL ya da } 444.444, \bar{4} \text{ TL}$$

bulabiliriz.

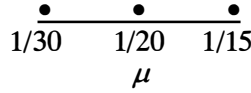
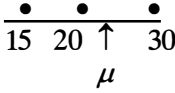
Bu durumda denilebilir ki, madem bu yanıt tartılı ortalama ile elde edilebilir, öyleyse ne diye işe TO gibi ters türs şeyler karıştırılmakta, bu arada bizim de kafamız karmakarışık edilmektedir? Belki de haklısınız ama bazı durumlarda tartılar bilinmeyebilir, o zaman da tartılı AO hesaplayamayız. Böyle durumlarda TO kullanılabilir. Bir örnekle bunu gösterelim.

Bir resmi kuruluşu son üç yılda ısıtma giderleri için hep aynı miktar ödenek verilmiştir. Bu kuruluş yakıtın tonunu ilk yılda 15, ikinci yılda 20, üçüncüde 30 milyon TL fiyatla almıştır. Burada her yıl alınan yakıt miktarı bilinemediğinden tartılı AO kullanamayız. Ama TO için bu örnek uygun bir uygulama alanı olabilir. Her yıl ayrılan ödenek A TL olsun. Üç yılda toplam $3A$ TL ödenmiştir. Bu tutarı toplam yakıt miktarına bölersek ortalama fiyatı buluruz:

$$\frac{3A}{\frac{A}{15} + \frac{A}{20} + \frac{A}{30}} = \frac{3}{\frac{1}{15} + \frac{1}{20} + \frac{1}{30}} = 20 \text{ milyon TL}$$

Bu işlem de TO hesabından başka bir şey değildir. Bu yaptıklarımızı bir kez de küçük çizimlerle gösterelim. Aşağıda solda fiyatlar, sağda ise bunların tersleri yer almıştır. Asıl fiyatların çarpık, terslerinin ise bakışık dağılı-

dıkları göze çarpmaktadır. Çarpık dağılımlar için uygun olmayan aritmetik ortalama, çarpıklık düzeltildikten sonra işe yaramaktadır.



Yukarıdaki iki örnekte de ortalaması aranan değişken, fiyattır; fiyat da tutar / miktar oranıdır. Ayrıca her alım-satımda ödenen toplam tutarlar aynıdır. Bu özellikleri aklımızda tutarak başka bir örneğe geçelim.

Bir sürücü bir kentten ötekine saatte ortalama 75 km hızla gitmiş olsun. Dönüşte aynı yolu kötü hava koşulları yüzünden saatte ortalama 60 km ile aldığını varsayalım. Acaba bu sürücünün gidiş-dönüş bütün yol boyunca ortalama hızı saatte kaç km'dir? Bu örnekte gidiş yoluyla dönüş yolunun aynı uzunlukta olduğuna dikkat edelim. İki satıcıya eşit paylaştırılan tutarları, her yıl aynı olan ödenekleri anımsayalım. Ayrıca hızın da fiyat gibi bir oran (yol / zaman) olduğunu düşünürsek burada da *TO* kullanmak uygun olabilir mi? Yolun uzunluğuna Y diyelim. Gidiş-dönüş yol uzunluğu $2Y$ km olur. Yolda geçen süreler gidişte $Y / 75$, dönüşte $Y / 60$ saattir. Öyleyse ortalama hız

$$\frac{2Y}{\frac{Y}{75} + \frac{Y}{60}} = \frac{2}{\frac{1}{75} + \frac{1}{60}} = 66.\bar{6} \text{ km/saattir.}$$

Peki, ne zaman *TO* kullanacağımızı nasıl bileceğiz? Ortalaması bulunacak değişken bize bir ipucu verebilir. Örneklerimizde değişkenler (fiyat ya da hız) hep birer orandı. Buradan her oran değişkene *TO* uygulanır sonucuna varabilir miyiz? Elbette hayır, çünkü daha önce hem *AO*'yu, hem *GO*'yu oran değişkenlerde kullandık. Peki, buradaki oran değişkeninin bir özelliği mi var? Evet. Fiyat, hız oranlarının payında yer alan büyüklükler soruda değişmeden kalıyorlarsa *TO* için uygun bir ortam var demektir. Her üç örneğimizde de bu özellik geçerlidir. Fiyat = Tutar / Miktar oranında *tutar*, Hız = Yol / Zaman oranında *yol* değişmeden kaldığı için *TO* iyi bir temsil edici olabilmıştır.

Hız örneğini biraz değiştirelim. Bir sürücü iki saat süren bir yolculuğun ilk saatinde 75 km, ikincisinde 60 km yol alsa ortalama hız ters ortalamayla değil, aritmetik ortalamayla daha doğru bulunur. Çünkü burada Yol / Zaman oranının payındaki yol büyüklüğünün değiştiğini, paydasındaki zaman değişkeninin aynı kaldığını görüyoruz.

Durumu genellemek için Y değişkenini X / Z oranıyla gösterelim. X aynı kalırken Z değişsin. Y_i değerlerinin ters ortalamasını bulalım. Böylece TO ile AO arasındaki ilişkiyi de görmüş olacağız.

$$Y_i = \frac{X}{Z_i} \quad \text{ya da} \quad \frac{1}{Y_i} = \frac{Z_i}{X} \quad i = 1, \dots, n$$

$$TO(Y) = \frac{N}{\sum \frac{1}{Y_i}} = \frac{N}{\sum \frac{Z_i}{X}} = \frac{N}{\frac{1}{X} \sum Z_i} = \frac{X}{\sum \frac{Z_i}{N}} = \frac{X}{AO(Z)}$$

Duruma yeni bir öge katan şu örneği alalım. Bir işyerinde çalışan 6 kişiden biri bir birim üretimi 15 dakikada, iki kişiden her biri aynı üretimi 20'şer dakikada, kalan üç kişiden her biri de 30'ar dakikada tamamlamaktadır. Bu işyerinde herkesin aynı süre çalıştığını düşünelim. Bir birim üretim ortalama kaç dakika sürer? Burada bir tür verim (süre / üretim) oranından söz edebiliriz. Süre, çalışanların her biri için aynıdır. Öyleyse oranın payı değişmez. Demek ki TO uygulanabilir.

$$TO = \frac{6}{\frac{1}{15} + \frac{1}{20} + \frac{1}{20} + \frac{1}{30} + \frac{1}{30} + \frac{1}{30}} = \frac{6}{\frac{1}{15} + \frac{2}{20} + \frac{3}{30}} = 22,5 \text{ dakika / ürün}$$

Bu örneğin öncekilerden tek farkı, bazı gözlem değerlerinin birden çok kez karşımıza çıkması, sıklık sayılarının birden büyük olmasıdır. Sıklık sayılarına f_i dersek tartılı TO 'nun genel gösterimini şöyle yazabiliriz.

$$TO = \frac{N}{\sum \frac{f_i}{Y_i}}$$

TO 'nun genel gösteriminden, gözlem değerlerinin sıfır ya da eksi işaretli olmaması gerektiği anlaşılabilir. Y_i 'lerden biri sıfır olursa, TO her zaman sıfır çıkar. (Neden?) Yine Y_i 'lerden biri sıfırdan küçükse TO anlamsız olur. Sözelimi $-1, 1, 2$ sayılarının TO 'su 6'dır; $-5, 10, 20$ sayılarının TO 'su da -60 çıkar. İlk ortalama en büyük gözlem değerinden büyük, ikincisi en küçüktür, dolayısıyla ikisi de anlamsızdır.

Bütün bu örneklerde ilk değişkenimizin tersini aldık. Bu gerçek yaşamda ne anlama gelir? *Fiyat* değişkeni *Tutar / Miktar* oranı olduğuna göre $1 / \text{Fiyat} = \text{Miktar} / \text{Tutar}$, yani sabit bir tutarla alınabilen miktar demektir. *Hız* değişkeni *Yol / Zaman* olduğuna göre *Hız*'ın tersi *Zaman / Yol*, başka bir deyişle belli uzunlukta bir yolu alma süresi anlamına gelir. Son örnekteki

verim = *süre* / *üretim* değişkeninin tersini aldığımızda belli bir sürede ne kadar üretim yapıldığını gösteren *üretim* / *süre* değişkenine dönmüş oluruz. Anlaşılacağı gibi *TO* hesaplamaya elverişli değişkenlerin tersleri alındığında bunlara gerçek yaşamda birer karşılık bulabiliyoruz. Bu da *TO*'yu anlamamızda, başkalarına anlatmamızda kolaylık sağlar. İlk değişken sağa, büyük değerlere doğru hayli çarpıklık gösterir de tersi alındığında bu çarpıklık ortadan kalkarsa *TO* hesaplamak için çok uygun bir ortam var demektir. Ama değişkenimizin tersi gerçek yaşamda bir anlam ifade etmiyorsa bunu açıklamakta güçlük çekebiliriz. Sözelimi *yaş* değişkeninin tersini aldığımızda tam bakışık bir dağılım elde etsek bile bunu yorumlamakta güçlük çekeceğimiz açıktır.

Kareli ortalama (KO)

Ortalamalar genel gösteriminde r 'yi 2 alırsak kareli ortalamayı elde ederiz.

$$M_2 = \left(\frac{\sum Y_i^2}{N} \right)^{\frac{1}{2}} = \sqrt{\frac{\sum Y_i^2}{N}} = KO$$

Burada Y değişkeninin gözlem değerleri, önce kareleri alınarak bir dönüştürmeye uğratılmakta, dönüştürülmüş değerlerin AO 'su elde edilmekte, bulunan değer ilk dönüştürmenin tersine karekök alınarak ilk ölçü birimine çevrilmektedir. Çarpık bir dağılımdan kareli bir dönüştürmeyle bakışık bir dağılımın elde edilebildiği durumlarda bu ortalamanın iyi bir temsil edici olduğu söylenebilir. Ayrıca ilk dağılım bakışık bile olsa, eksi değerli gözlemlerle artı değerli gözlemlerin bir arada bulunduğu dağılımlarda, özellikle bu gözlemlerin işaretli değerlerinden çok işaretsiz salt değerleri önemliyse (sözelimi bu gözlemler belli bir değere göre farkları gösteriyorsa), KO uygun bir ortalama olarak kullanılabilir.

AO 'yu incelerken ele aldığımız Marmara Bölgesi'nde bazı merkezlerde ölçülmüş hava sıcaklıkları örneğine dönelim. Bu değerler şöyleydi: Balıkesir 14,5, Bursa 14,6, Çanakkale 14,8, Edirne 13,4, Göztepe 14,0. AO da 14,26 bulunmuştu. Her merkezdeki sıcaklığın bu ortalamadan farkını alır da bu farkların ortalamasını, yani her bir merkezin ortalama sıcaklıktan ortalama ne kadar farklı olduğunu bulmak istersek ne yapabiliriz? AO denememiz boşunadır, çünkü gözlem değerlerinin AO 'dan sapmalarının toplamının her zaman sıfır çıktığını öğrenmiştik. Bu da sapmaların AO 'sunun hep sıfır çıkacağı anlamına gelir. Bu bilgi de işimize pek yaramaz doğrusu! Sapmaların işaretleri göz ardı edilerek sapmalar bulunursa, Balıkesir için $|14,5 - 14,26| = 0,24$, Bursa için 0,34, Çanakkale için 0,54, Edirne için

0,86, Göztepe için 0,26 elde edilir. Bunların AO 'su da 0,448'dir. **Ortalama mutlak sapma (OMS)** denen bu ortalama işimize yarayabilir, ama ortalamalar genel gösterimine uymadığı için kullanılışı sınırlıdır.

AO 'nun özelliklerini incelerken, gözlem değerlerinin AO 'dan farklarının kareleri toplamının, başka herhangi bir ortalama farkların kareleri toplamına göre her zaman daha küçük olduğunu görmüştük. Bunun gibi özelliklerinden ötürü burada KO kullanmak daha çok işimize yarayabilir. Bu sapmaların mutlak değerleri yerine karelerini alarak KO 'yu bulalım.

$$KO = \sqrt{\frac{0,24^2 + 0,34^2 + 0,54^2 + 0,86^2 + 0,26^2}{5}} = \sqrt{0,2544} \approx 0,5$$

Demek ki sapmaların KO 'su yaklaşık yarım derece kadardır. Bu tür ortalamaı daha sonra yayıklık ölçüleri arasında standart sapmayı incelerken kullanacağız.

KO çoğunlukla ortalama sapmalar için kullanılır ama bu başka zaman kullanılmayacağı anlamına gelmez. Az önce değindiğimiz gibi, çarpık bir dağılım, değişken değerlerinin kareleri alınarak bakışık kılınabiliyorsa KO uygulamamak için bir neden yoktur? Elbette burada da bir değişkenin karesi alınırsa gerçek yaşamda buna bir karşılık bulunabilir mi sorusunu da unutmamalıyız. Bir ipucu olarak şunu da söylemek yerinde olur ki sağa, büyük değerlere doğru uzayan çarpık dağılımlarda kareli bir dönüştürme yapmak çarpıklığı azaltmak yerine artırır. Kareli dönüştürme sola, küçük değerlere doğru uzayan çarpık dağılımlar için daha elverişlidir. Okuyucu yukarıdaki örneklerde deneyerek bunu kendi görmelidir.

Yarı duyarlı bir ortalama: Kırpık ortalama ($KıO$)

Bu tür ortalama AO ile $Ortc$ arasında bir tür uzlaşmayı yansıtır. AO , bütün gözlem değerlerini kullandığından çok bilgi içerir ama öte yandan aşırı değerlerin etkisi altındadır. $Ortc$ ise tam tersi konumdadır, aşırı değerlere karşı dirençlidir, berktir ama bilgi yükü daha azdır. Bu iki ortalamanın iyi yönlerini bir araya getirmek için $KıO$ geliştirilmiştir. Bu ortalama, iki uçtan eşit sayıda gözlem değerini kırpıp arada kalanların AO 'su alınarak bulunur. Böylece uçlarda yer alan aşırı büyük ya da aşırı küçük değerler dışlandığından ortalamaımız berkleşir, direnç kazanır, öte yandan kalan bütün gözlem değerlerini hesaba kattığından içerdiği bilgi yükü de az değildir.

Burada sorun uçlardan ne kadar kırpma yapılacağına düğümленir. Hiç kırpma yapmazsak AO elde edilir. Yapabileceğimiz en çok kırpmayı yaparsak elimizde yalnız $Ortc$ kalır. Kırpılacak gözlem değerlerinin sayısı bu iki uç arasında, yeterince direnç kazandıracak ama duyarlılıktan da pek ödün

vermeyecek biçimde seçilmelidir. Çok kullanılan bazı kırpma oranları % 5 ile % 10'dur. Bu oranlar her iki uçta da uygulandığında ilk durumda gözlem değerlerinin % 10'u, ikincide % 20'si kırılmış olur. Bazı bilgisayar istatistik paket programları kendi hallerine bırakılırlarsa % 5 oranını uygularken bazı yazarlar (Siegel, 1988:66-68) % 10 oranını önermektedir. Burada gözlem sayısı da önemlidir. Az sayıda gözlem kırpma oranını sınırlar.

Orta yarı ortalaması ya da **dördebölenler arası ortalama** diyebileceğimiz özel bir tür *KırO* ise gözlem değerlerinin yarısını kırpar. D_a ile D_u arasındaki bütün gözlem değerlerinin *AO*'sunu alan bu ortalama da oldukça yaygın kullanılmaktadır (Erickson, Nosanchuk, 1992:45). Bütün bu ortalamalarda kırılan değerlere 0, kırılmayanlara 1 ağırlık verilir.

Bu ağırlıklar değiştirilerek hesaplanan ortalamalara da ***M*-kestirimcileri** (**en yüksek olabilirlik kestirimcileri**) denir (Hoaglin, vd., 1983:292-294). Sözelimi hiç bir değer kırılmazsa *M*-kestirimcisi *AO* olur. Kırılmayan gözlemlere verilen ağırlıklar her *M*-kestirimcisinde uçlara doğru gittikçe azalır. SPSS yazılımı Huber, Tukey, Hampel, Andrew tarafından önerilen *M*-kestirimcilerini hesaplar (SPSS, 1993:192-195). Alıştırma 7.7'deki bitirme notları için $AO = 2,9211$, $Ortc = 2,935$ bulmuştuk. SPSS sonuçlarına göre aynı verilerin *M*-kestirimcileri şöyledir.

Hubert: 2,9333; Tukey: 2,9328; Hampel: 2,9296; Andrew: 2,9328

Görüldüğü gibi *M*-kestirimcileri *AO* ile *Ortc* arasında değer almıştır. Demek ki yukarıda kendilerinden beklenen işlevi yerine getirdikleri söylenebilir. Bulunan bütün sonuçların birbirine hayli yakın çıkması da dağılımda oldukça yüksek bir bakışım bulunduğu göstergesi olarak yorumlanabilir.

Bu aileden bir başka ortalama da **genişletilmiş ortanca** (*G-Ortc*) adını alır (Rosenberger, Gasko, 1983). *G-Ortc*, gözlemlerin yuvarlatılmalarına karşı berktir. Bunu sağlamak için *Ortc* ile iki yanındaki bazı değerlerin ortalaması alınır. Gözlem sayısı 5-12 arasında tek sayıysa ortadaki 3 değerlerin ortalaması, çift sayıysa ortadaki 4 değerlerin sırasıyla 1/6, 1/3, 1/3, 1/6 tartılarıyla ortalaması alınarak bulunur. Gözlem sayısı 12'den büyük tek sayıysa ortadaki 5 değerlerin ortalaması, çift sayıysa ortadaki 6 değerlerin sırasıyla 0,1, 0,2, 0,2, 0,2, 0,2, 0,1 tartılarıyla ortalaması alınarak bulunur. Buradan anlaşılacağı gibi *G-Ortc* da bir tür kırık ortalamadır. Gözlem sayısı

5 ise kırpma oranı $2/5 = 0,4$;

12 ise kırpma oranı $8/12 = 0,75$;

13 ise kırpma oranı $8/13 = 0,62$;

20 ise kırpma oranı $14/20 = 0,70$;

100 ise kırpma oranı $94/100 = 0,94$ kadardır.

Alıştırma 7.7'deki bitirme notları aslında virgülden sonra iki basamağa yuvarlatılmıştır. 96 gözlemlili bu veri kümesinde *G-Ortc* şöyle hesaplanır.

$$2,90 \times 0,1 + 2,90 \times 0,2 + 2,92 \times 0,2 + 2,95 \times 0,2 + 2,95 \times 0,2 + 2,97 \times 0,1 = 2,931$$

Bu değer de *M*-kestirimcileri gibi *Ortc* ile *AO* arasında olup onlara çok yakındır. Yukarıda yapılan yorum burada da geçerli sayılabilir. Çarpık dağılımlarda sonuçlar birbirine bu kadar yakın çıkmaz.

Böylece bütün ortalamaların başlıcalarına değindik. Bir dağılım için uygun ortalamayı seçmek, bir spor karşılaşmasında bizi (okulumuzu, ülkemizi) temsil edecek en iyi takımı seçmek gibidir. Basketbol karşılaşması için masa tenisi ya da güreş takımımızı sahaya sürersek sonuç pek parlak olmayabilir.

Okuma Önerileri

- Erickson, B. H., Nosanchuk, T. A. (1992) *Understanding Data*, 2nd. ed. Open University Press (37-46).
- Gürtan, K. (1987) *İstatistik ve Araştırma Metodları*, 4. b., İstanbul Üniversitesi (209-296).
- Groebner, D. F., Shannon, P. W., Fry, P. C., Smith, K. D. (2001) *Business Statistics: A Decision Making Approach*, 5. b., Prentice Hall (82-95).
- Huff, D. (1973) *How to Lie with Statistics*, Penguin (29-36) (Çeviren: Koparan, E. (1995) *İstatistik ile Nasıl Yalan Söylenir?*, Sarmal (31-38)).
- Jaffe, A. J., Spierer, H. F. (1987) *Misused Statistics*, Marcel Dekker (87-114).
- Marsh, C. (1988) *Exploring Data*, Polity Press (24-40).
- Newbold, P., Carlson, W. L., Thorne, B. (2003) *Statistics for Business and Economics*, 5th ed, Prentice Hall (31-38) (4. basımdan çeviren: Şenesen, Ü. (2000) *İşletme ve İktisat İçin İstatistik*, Literatür (9-15)).
- Ott, R. L., Longnecker, M. (2001) *An Introduction to Statistical Methods and Data Analysis*, 5th ed. Duxbury (69-81).
- Rosenberger, J. L., Gasko, M. (1983) "Comparing location estimators: Trimmed means, medians, and trimean", Hoaglin, Mosteller, Tukey (1983) içinde 297-338.
- Siegel, A. F. (1988) *Statistics and Data Analysis*, Wiley (57-76).
- Siegel, A. F. (1997) *Practical Business Statistics*, Irwin (64-104).

Alıştırmalar

- 8.1** Bir laboratuarda üretilen bir bakteri türünün sayısı üç gün içinde 1000'den 16000'e çıkmıştır. Günlük ortalama artış oranını bulunuz.

- 8.2 Aşağıdaki verilere göre Türkiye'nin 1982, 1989, 1993, 1997, 2003, 2005 yıl ortası nüfusunu kestiriniz.

Nüfus sayımı yılı	1980	2000
Yıl ortası nüfus (bin kişi)	44.439	67.469

- 8.3 KO 'su 25, TO 'su 22.5, AO 'su 23.5 olan bir veri kümesinin GO 'su kaç olabilir?
- 8.4 Aliye parasının hepsini bir işe yatırıyor, üç yıl boyunca hiç çekmiyor. Buradan sabit fiyatlarla ilk yıl % 10, ikinci yıl % 8, üçüncü yıl % 12 getiri sağlıyor. Üçüncü yıl sonunda parasının ulaştığı tutarın aynısını elde edebilmek için ilk parasını üç yıl boyunca değişmeyen hangi faiz oranıyla değerlendirmesi gerekirdi?
- 8.5 Veliye parasını 25 – 25 – 50 oranlarında üç parçaya ayırıp sabit fiyatlarla sırasıyla yılda %10, %8, %12 getiri sağlayan üç ayrı işe yatırıyor, bir yıl tutuyor. Yıl sonunda parasının ulaştığı tutarın aynısını elde edebilmek için ilk parasının hepsini tek bir yere hangi faiz oranıyla yatırması gerekirdi?
- 8.6 Bir malın fiyatı 4 yılda 8.35 katına çıkmıştır. Yıllık ortalama fiyat artış oranı kaçtır?
- 8.7 Aşağıdaki verilere göre Türkiye'de toplam mevduat yılda ortalama % kaç artmıştır?

Yıl	1997	1998	1999	2000
Mevduat (trilyon TL):	11964,6	22.916,8	45.291,8	64.943,0

- 8.8 Hazine'nin İktisadi Devlet Teşekkülleri'nden aldığı gelir 1988'de 70 milyar TL iken 1994'te 2,682 trilyon TL olmuştur. GSÜÜ (Gayrisafi Ulusal Ürün) aynı yıllarda 129,125 trilyon TL'den 3.887,903 trilyon TL'ye çıkmıştır. Hangisinin yıllık artış oranı daha yüksektir?
- 8.9 Aşağıdaki durumların her birinde hangi ortalamanın kullanılması doğru olur? Gerekçenizi açıklayınız?
- Bir sınıftaki kız öğrencilerin not ortalaması,
 - Bir yıl boyunca her ay kâğıt alımları için aynı parayı harcayan bir kurumun her ay farklı fiyatlardan aldığı kâğıdın aylık ortalama fiyatı,
 - Kastamonu'daki özel taşıtların yıllık ortalama artış hızı,
 - Dört kişilik bir bisiklet takımının aynı yarıştaki ortalama hızları,
 - Bursa'nın İnegöl ilçesinde tipik bir ailenin evindeki oda sayısı,
 - Bir öğrencinin bir yıl boyunca her dersten aldığı notlarla yıllık ortalama notu arasındaki farkların ortalaması.
- 8.10 Bir kentten başka bir kente ortalama 200 km/saat hızla uçan bir uçak dönüşte kötü hava koşullarıyla karşılaşmıştır. Uçağın gidiş-dönüş ortalama hızı 177.7 km/saat olduğuna göre dönüş yolundaki ortalama hızını bulunuz.

8.11 1987 sabit fiyatlarıyla düzenlenen aşağıdaki çizelgeyi kullanarak:

- a) GSUÜ'nün 2000 yılı büyüme oranını bulunuz.
b) Her bir kesimin 2000 yılı GSUÜ içindeki payı ne kadardır?

Kesim	1999 GSUÜ içindeki pay (%)	2000 büyüme oranı (%)
Tarım	13,4	3,9
Sanayi	27,9	6,0
Hizmetler	58,7	7,0

8.12 İstanbul nüfusunun 1985-90 arasında % 4,580 olan yıllık ortalama artışı, 1990-2000 arasında % 3,309'dur. 1985'te 5.842.985 kişiyi barındıran İstanbul, 1987, 1990, 1997, 2003 yıllarında kaç kişi olmuştur? Herhangi bir var sayım yaptıysanız bunu açıklayınız.

8.13 Son üç yılda bir bölgede satılan ev fiyatları ortalama olarak sırasıyla 64, 124, 176 milyar TL olmuştur. Aynı dönemdeki ortalama yıllık enflasyon oranı % 39,7'dir. Buna göre ev fiyatlarındaki artış oranı enflasyondan düşük mü, yoksa yüksek mi olmuştur?

8.14 Bir sınavdaki 50 öğrencinin notlarının dağılımı aşağıdaki gibidir. Bu dağılım için:

- a) Ortç,
b) ESD,
c) AO bulunuz.
d) Bu üç ortalama bize ne gibi bilgiler vermektedir?
e) Bu dağılımı uygun bir çizimle gösteriniz.

Notlar	30-< 40	40-< 50	50-< 60	60-< 70	70-< 90	90-< 100
Kişi sayısı	4	6	8	12	16	4

8.15 Bir tiyatronun sekiz günlük bilet satışları aşağıdaki gibiyken ortalama günlük satış gelirlerini bulunuz.

Gün	1	2	3	4	5	6	7	8
Tam (8 milyon TL)	140	136	114	130	123	120	135	134
İndirimli (5 milyon TL)	60	51	82	68	77	77	59	62

8.16 1981 yılında ortalama aylık işçi ücretiyle 40 kg et alınırken, bu sayı 1982'de 32,4 kg olmuştur. İki yılın ortalamasını bulunuz.

8.17 Türkiye nüfusunun yıllık ortalama artış hızları aşağıdaki gibi olduğuna göre 1970-2000 arası **yıllık** ortalama artış oranı kaçtır?

Dönemler	1970-75	1975-80	1980-85	1985-90	1990-2000
Yıllık artış hızı (%)	2,501	2,065	2,488	2,171	1,835

- 8.18** 1990 yılında ülkemizde kadınların % 42,30'u, erkeklerin % 39,71'i köylerde yaşadığına göre 1990'da toplam nüfusun % kaç şehirde yaşamaktaydı? Eksik bilgiyle bulduğunuz sonucu DİE'nin *Türkiye İstatistik Yıllığı*'ndaki oranla karşılaştırınız.
- 8.19** Bir kentte 1 kg ekmek fiyatıyla bir kg et fiyatı arasındaki oran son üç yılda 1/10, 1/12, 1/16 olmuştur.
- a) Üç yılın ortalama "ekmek fiyatı / et fiyatı" oranı nedir?
- b) Üç yılın ortalama "et fiyatı / ekmek fiyatı" oranı nedir?
- c) İki ortalama oran birbirleriyle uyumlu mu, değil mi? Çelişki varsa nerede yanlış yaptığınızı düşünürsünüz?
- 8.20** Aşağıdaki oranlar Türkiye GSÜÜ'nün yıllık büyüme hızlarıdır.

Yıl	%	Yıl	%	Yıl	%	Yıl	%
1969	4,3	1978	1,2	1987	9,8	1996	7,1
1970	4,4	1979	-0,5	1988	1,5	1997	8,3
1971	7,0	1980	-2,3	1989	1,6	1998	3,9
1972	9,2	1981	4,8	1990	9,4	1999	-6,1
1973	4,9	1982	3,1	1991	0,3	2000	6,3
1974	3,3	1983	4,2	1992	6,4	2001	-9,4
1975	6,1	1984	7,1	1993	8,1		
1976	9,0	1985	4,3	1994	-6,1		
1977	3,0	1986	6,8	1995	8,0		

- a) 1968-2001 döneminin ortalama yıllık büyüme hızı kaçtır?
- b) 1980 öncesi dönemin ortalama yıllık büyüme hızı kaçtır?
- c) 1980-2001 döneminin ortalama yıllık büyüme hızı kaçtır? 1980 öncesiyle karşılaştırınız.
- d) 1981'deki GSÜÜ düzeyi 1978'in % kaç üstündedir?
- e) 1999'da % 6,1 azalan, 2000'de % 6,3 artan GSÜÜ'nün 2000 yılı düzeyi 1998 düzeyinin üstünde mi, yoksa altında mıdır? Nasıl hesapladığınızı gösteriniz.
- f) Bu verileri bir zaman serisi olarak çizip gördüklerinizi yorumlayınız.
- 8.21** Kişi başına GSÜÜ 1965-83 arasında Türkiye'de yılda ortalama % 3,0, İrlanda'da % 2,3 artmıştır. Aynı gelişme hızları gelecekte de sürdürülürse iki ülkenin kişi başına GSÜÜ değerlerinin 2187 yılında eşitleneceği hesaplanmıştır.

- a) 1983'te Türkiye'de kişi başına GSUÜ 1240 \$ olduğuna göre İrlanda'da kaç \$'dır? 2187'de kaç \$ olacaktır?
- b) Alıştırma 8.20'deki verileri kullanarak Türkiye kişi başına GSUÜ büyüme hızının 1983'ten sonra ortalama %3'e ulaşip ulaşmadığını bulunuz.
- 8.22** 14 gözlemden oluşan bir veri kümesinde gözlem değerlerinin ondalık logaritmalarının toplamı 34,26'dır. Bu kümenin AO'su 275, 280 ya da 290'dan biri olabilir mi? Gerekçenizi açıklayınız.
- 8.23** Memur maaşlarına 1997 yılbaşında % 30, altı ay sonra bir %30 daha zam yapılmıştır.
- c) 1997 Aralık ayı maaşının satın alma gücünün bir yıl önceye göre azalması için 1997'de enflasyonun ortalama aylık hızı en çok % kaç olmalıdır?
- d) 1996'da 339,1 olan DİE kentsel yerler tüketici fiyatları indeks sayısı 1997'de 629,8 olmuştur (1994:100). Bu durumda memurların satın alma gücü hangi yönde, ne kadar değişmiştir?
- 8.24** Bir işyerindeki toplu iş görüşmeleri sırasında işveren, ücretlerin AO'sunu vererek bunun yüksek bir ücret olduğunu, işçi sendikasıysa *Ortc*'sını ileri sürerek ücretlerin düşük olduğunu savunmaktadır.
- a) Bu çelişkili gibi görünen durumu nasıl açıklayabilirsiniz?
- b) Sendikanın bir işyeri temsilcisi pazarlığın, ortanca yerine ensıkdeğer temeline oturtulmasını istemenin sendika yararına olacağı kanısındadır. Siz ne dersiniz?
- 8.25** Dokuz kişi belli bir işi ayrı ayrı yapmakla görevlendirilmişlerdir. Her birinin bu işi tamamlama süresi aşağıdaki gibidir.
- | | | | | | | | | | |
|---------|---|-----|---|-----|----|------|----|---|-----|
| Dakika: | 1 | 2,5 | 3 | 1,5 | 2, | 1,25 | 30 | 1 | 0,9 |
|---------|---|-----|---|-----|----|------|----|---|-----|
- a) Bu işin ortalama süresini çeşitli düzey ölçüleriyle bulunuz, sonuçları karşılaştırınız.
- b) Hangi ortalamayı kullanmak sizce daha uygundur? Neden?
- 8.26** Bir hastanede yatan 10 hastanın ateşi ölçülmüş, aşağıdaki ölçümler elde edilmiştir.
- | | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|------|
| 40,0 | 40,4 | 38,7 | 42,2 | 39,9 | 37,9 | 39,1 | 38,6 | 38,2 | 40,1 |
|------|------|------|------|------|------|------|------|------|------|
- a) AO'yu bulunuz. %10'luk *Kıro* bu değerden çok farklı çıkar mı? Hesap yapmadan yanıtlayınız.
- b) *Kıro*'yu hesaplayınız, yukarıdaki yanıtınızla karşılaştırınız.
- 8.27** 2000 Olimpiyatlardan bir yıl önce 28 Temmuz – 12 Ağustos günleri arasında oyunların yapılacağı bir alanda ozon düzeyi ölçülmüş, ppm cinsinden şu sonuçlar bulunmuştur.

10 14 13 18 12 22 14 19 22 13 14 16 3 6 7 19

- Aritmetik ortalamayı bulunuz.
- Ortancayı bulunuz.
- En küçük ile en büyük gözlemleri atarak kırpık ortalamayı bulunuz. Bu ortalama, AO ile $Ortc$ 'dan daha iyi bir temsil edici midir? Neden?
- İkinci bir alandaki ölçümler de aşağıdaki gibidir. Bu alan için de AO ile $Ortc$ 'yı bulunuz. Bunları ilk alaninkilerle karşılaştırınız. Hangisinde ozon düzeyi daha yüksektir? Neden?

8 13 10 9 16 12 13 14 17 13 9 16 5 9 8 12

- Her iki dağılımın histogram, dal-yaprak, nokta çizimlerini düzenleyerek karşılaştırınız.

8.28 Bir süpermarket müşterilerinden rassal seçilmiş 30'unun para ödeme kuyruğunda bekledikleri süreler dakika olarak aşağıdadır. Çeşitli ortalamalar kullanarak ortalama bekleme süresini bulunuz. Bu ortalamalardan hangisi verileri en iyi temsil etmektedir? Açıklayınız.

8 6 8 11 6 12 7 11 8 7 1 2 13 7 7
6 6 12 7 5 13 6 6 9 17 5 5 2 2 3

8.29 “Bir kentte oturanların % 65'inden çoğunun yıllık geliri o kentin yıllık ortalama gelirinin altındadır”. Bu söz doğru olabilir mi? Gerekçenizi açıklayınız.

8.30 Bir vergi dairesinde yapılan bir araştırmaya göre para ödeme kuyruğunda bekleme süreleri şöyle dağılmaktadır.

Süre (dak)	0	<05	0.5-<1	1-<2	2-<3	3-<5	5-<10	10-<15
Müşteri (%)	5	20	32	19	11	7	4	2

- Dağılımın histogramını çizin.
- Bu dağılımın ortalama düzeyini AO , ESD , $Ortc$, $KırO$ kullanarak bulunuz. Hangi ortalamanın en uygun olduğunu tartışınız

- 8.31 TC Merkez Bankası'nın 1986 yılı araştırmasına göre Almanya'daki yurttaşlarımızın aylık tasarrufları aşağıdaki gibidir.

Aylık tasarruflar (DM)	Kişi sayısı
0 – < 250	165
250 – < 500	674
500 – < 1000	1021
1000 – < 2000	441
2000 – < 3000	75
3000 – < 5000	34
5000 +	17
Toplam	2427

- Bu dağılımın histogramını düzenleyiniz. Eğer bir varsayım yapmak gerekiyorsa bunu açıkça belirtiniz.
 - Dağılımın biçimini yorumlayınız.
 - Dağılımın *AO*'sunu hesaplayınız. Bu ortalama bu dağılıma uygun mudur? Ne düşünüyorsunuz?
 - Daha uygun bir ortalama düşünürseniz onu da hesaplayınız.
 - ESD*'yi bulunuz.
 - Ortc*'yi bulunuz.
 - 2427 kişinin bir aylık toplam tasarrufu kaç DM tutmaktadır?
 - Bu yurttaşlardan % kaçının aylık tasarrufu 2000 DM ya da daha çoktur? Bu yurttaşların toplam tasarruf içindeki payı ne kadardır?
 - Bu yurttaşlardan en **düşük** tasarruf yapan $\frac{1}{4}$ 'ünün toplam tasarrufun % kaçını yaptığını bulunuz.
 - Bu yurttaşlardan en **yüksek** tasarrufu yapan $\frac{1}{4}$ 'ünün toplam tasarrufun % kaçını yaptığını bulunuz.
 - Dördebölenler aralığına giren $\frac{1}{2}$ oranındaki yurttaşın aylık tasarrufu en az kaç, en çok kaç DM'dir? Bu yurttaşların tasarruf tutarı, toplam tasarrufun % kaçındır?
- 8.32 “Aylık ortalama enflasyon oranının % 3 olduğu bir dönemde yıllık % 36 gelir artışı satın alma gücünü aynen korur” sözünü tartışınız.

8.33 Dolaşımdaki para + vadesiz mevduat + vadeli mevduat = M2 biçiminde tanımlanan geniş para arzı 1970-2000 arasında aşağıdaki gibi gelişmiştir (milyar TL).

Yıl	M2	Yıl	M2	Yıl	M2	Yıl	M2
1970	44	1978	328	1986	12.173	1994	630.348
1971	57	1979	528	1987	17.648	1995	1.256.632
1972	71	1980	882	1988	27.194	1996	2.924.893
1973	90	1981	1.637	1989	47.139	1997	5.658.800
1974	113	1982	2.554	1990	71.570	1998	11.423.198
1975	147	1983	3.288	1991	117.118	1999	21.992.654
1976	181	1984	5.179	1992	190.736	2000	31.109.334
1977	244	1985	8.146	1993	282.442		

- Beşer yıllık dönemlerin hangisinde para arzının yıllık ortalama artış hızı en yüksektir?
- Beşer yıllık dönemlerin her birinde yıllık ortalama artış hızı kaçtır?
- 1970’li yıllarda geçerli olan yıllık ortalama artış hızı daha sonra da ay-nen sürseydi, 1995 yılında M2 kaç TL olurdu? 2000 yılında kaç TL olurdu? Yorumlayınız.

8.34 Türkiye’de cari fiyatlarla kişi başına ulusal ürün (KBUÜ) ile fiyat şişkinliklerini gösteren “deflatör” 1970-2000 arasında beşer yıllık aralıklarla aşağıdaki gibi değişmiştir.

- Her beş yıllık dönemin sabit fiyatlarla KBUÜ değerlerini bulunuz.
- Bunların her birinin bir önceki döneme göre ne yönde, % kaç değiştiğini gösteriniz.
- Sabit fiyatlarla KBUÜ’nün her beş yıllık dönem içindeki yıllık ortalama artış hızlarını bularak karşılaştırınız.

Yıl	KBUÜ (TL)	“Deflatör” (1987:100)
1970	5.884	0,60
1975	17.239	1,49
1980	119.335	10,42
1985	702.706	56,24
1990	7.066.839	469,52
1995	127.423.385	7.931,99
2000	1.851.412.612	10.541,54

8.35 Asgari (en düşük yasal) ücretle çalışan birinin çeşitli yıllarda 1 kg ekmek alabilmek için kaç dakika çalışması gerektiği hesaplanmış, aşağıdaki sonuçlar elde edilmiştir (Petrol-İş, 2000:494).

- Asgari ücretli birinin 1995-99 döneminde 1 kg ekmek almak için ortalama kaç dakika çalışması gerektiğini bulunuz. Bu kimse günde 8 saat çalışırsa kaç ekmek alabilir?
- Bu kimsenin günde 8 saat çalışarak kaç ekmek alabileceğini 1995-99 arasındaki her yıl için bulunuz. Bulduğunuz beş değerin uygun bir ortalamasını alarak (a)'da bulduğunuz sonuçla karşılaştırınız. Aynı mı çıktı, yoksa farklı mı? Bunun nedenini açıklayınız.
- Bir asgari ücretlinin 1 kg ekmek alabilmesi için çalışması gereken süre 1963-99 arasında yılda ortalama % kaç değişmiştir?

Yıl	1963	1970	1980	1995	1996	1997	1998	1999
Süre (dak)	49	32	96	97	110	112	99	101

8.36 1990 sayımında nüfusu 750.882 kişi olan Denizli'de 337.416 kişi kentte oturmaktaydı. 2000 sayımında toplam nüfus 1990'a göre yılda ortalama % 1,240, kent nüfusu % 3,043 artmıştır.

- Son sayımda nüfusun % kaç kentte sayılmıştır?
- Köy nüfusu yılda ortalama % kaç artmıştır?
- Toplam nüfus on yılda % kaç artmıştır?

8.37 1990, 2000 nüfus sayımının bazı illerdeki sonuçlarına göre düzenlenen aşağıdaki çizelgede “?” ile gösterilen gözleri doldurunuz.

		Kastamonu	Sinop	Tunceli	Ardahan
1990	Kent	148.861	?	50.799	?
sayımı sonucu	Köy	?	178.712	?	?
	Toplam	423.206	?	?	163.731
2000	Kent	?	101.285	?	39.725
sayımı sonucu	Köy	201.456	?	?	?
	Toplam	?	225.574	93.584	?
Yıllık ortalama % artış	Kent	15,61	15,84	6,99	15,45
	Köy	?	?	-74,97	?
	Toplam	?	?	?	-20,22

- 8.38** Alıştırma 3.5'e dönerek bölgelerin nüfus sayımı verileri için *ESD*'yi bulun. Bunun ne anlama geldiğini açıklayın.
- 8.39** Alıştırma 3.6'ya dönerek meteoroloji istasyonlarının yağışlı gün sayısı verileri için *ESD*'yi bulun. Bunun ne anlama geldiğini açıklayın.
- 8.40** Alıştırma 3.11'e dönelim.
- Bu alıştırmadaki *kesilen ceza* değişkeninin uygun bir ortalamasını bulmak isterseniz hangi ortalamayı kullanırdınız? Neden?
 - Bu değişkenin cari fiyatlarla ölçülmüş, yani enflasyonun etkisinden arındırılmamış olması ortalamamızın yorumlanmasını etkiler mi? Aynı soruyu ödenen *ceza değişkeni* için de yanıtlayınız.
- 8.41** Alıştırma 4.6'ya dönerek hem *satış miktarı*, hem *satış geliri* değişkenleri için uygun ortalamaların neler olabileceğini tartışınız.
- 8.42** Alıştırma 4.7'ye dönerek hem *nüfus*, hem *yüzölçümü* değişkenleri için uygun ortalamaları bulunuz, anlamlarını tartışınız.
- 8.43** Alıştırma 4.8'deki öğrenci payları için uygun ortalama ne olabilir?
- 8.44** Çizelge 7.3'e dönelim.
- Dışalım ile dışsatımın yıllık ortalama düzeylerini nasıl bulursunuz? Kullandığınız ortalamamızın gerekçesini açıklayınız.
 - Genel olarak artan dizilerde böyle bir ortalama hesaplamamızın anlamlı olup olmadığını tartışınız.
 - Bu verilerle yıllık ortalama artış yüzdesini bulunuz. Bunun anlamlı olup olmadığını tartışınız.
- 8.45** Çizelge 7.4'e dönelim.
- Oradaki arabaların yıllık ortalama satışlarını bulunuz. Neden o ortalamayı kullandığınızı açıklayınız.
 - Çizelge 7.5'teki verilerle aynı ortalamayı hesaplayıp (a)'daki bulgunuzla karşılaştırınız. Yuvarlatma işlemi sonucunuzu nasıl etkiliyor? Tartışınız.
- 8.46** Çizelge 7.8'e dönelim.
- Herhangi bir yılı seçerek o yıla ilişkin uygun bir ortalama hesaplayınız. Bu ortalama değer ne anlama gelir?
 - (a)'da sorulan soruda neden *GO* hesaplanamayacağını tartışınız.
 - Başka bir yıl için aynı ortalamayı hesaplayınız. İki dönem arasındaki değişmeyi bu ortalamalardan anlayabilir misiniz? Açıklayınız.
 - Sol sütundan *kamu hizmetleri* dışında herhangi bir kesimi seçip 1973-96 dönemi ortalamasını hesaplayınız. Niye o ortalamayı seçtiğinizi açıklayınız.
 - (d)'deki soruyu Çizelge 7.9'daki yuvarlatılmış verilerle yanıtlayınız. Aradaki farkın önemli olup olmadığını tartışınız.
 - (a)'daki soruyu bir de Çizelge 7.11'deki öbeklenmiş verilerle yanıtlayınız.

8.47 Alıştırma 7.3'e dönelim.

- Çizelge A'daki verilerle bölgelerin yıllık ortalama satışlarını hesaplayıp yuvarlatınız. Aynı işlemi Çizelge B'deki yuvarlatılmış verilerle yapıp iki sonucu karşılaştırınız. Fark var mı? Varsa önemli mi?
- İşletmenin bütün bölgelerinin üçer aylık toplam satışlarının ortalamasını hem A, hem B çizelgelerinden hesaplayınız. Bulgularınızı yukarıdaki gibi karşılaştırınız.

8.48 Alıştırma 7.10'a dönelim.

- Her bir et ürünü için yıllık ortalama tüketimi hesaplayınız. Neden o ortalama seçtiğinizi açıklayınız.
- Herhangi bir yıla ilişkin çeşitli et ürünleri verileriyle ortalama hesaplamak anlamlı olur mu? Neden?

8.49 Alıştırma 7.12'deki savaşa karşı olma oranlarıyla uygun bir ortalama hesaplanabilir mi? Hangi ortalama neden seçtiğinizi belirtiniz.

8.50 Çizelge 7.14'teki verilerle bölgelerimizin GSYİÜ içindeki ortalama payını hesaplayınız. Neden o ortalama kullandığınızı açıklayınız.

8.51 Alıştırma 7.39'daki verilerle ortalama savunma harcamaları payını, ortalama eğitim harcamaları payıyla karşılaştırınız. Neden o ortalama kullandığınızı açıklayınız?

8.52 Alıştırma 7.38'deki verilere göre araba üretimi için ortalama hesaplamak anlamlı olabilir mi? İrdeleyiniz.

8.53 Alıştırma 7.42'deki verilerle ortalama boşanma oranını, ortalama evlenme oranıyla karşılaştırınız. Kullandığınız ortalamanın gerekçesini açıklayınız.

8.54 Çizelge 8.2'deki verilerle bir histogram çizin. Çizelge 8.3'te sözü edilen a , b fark paylarını bu histogram üzerinde işaretleyiniz. Oradaki hesaplamanın benzer üç genlere dayandığını gösteriniz.

8.55 Çizelge 8.4'teki veriler incelenirken "Bu da bize o yıl üniversitelerimizin söz konusu oran açısından türdeş olmadıkları, büyük kentlerdeki üniversitelerle o sıralarda çoğu yeni kurulmuş olan öteki üniversitelerin birbirinden ayrı iki öbek oluşturdukları izlenimini vermektedir" demiştik. Bu farklılığın zaman içinde ortadan kalkıp kalkmadığını yeni verilerle inceleyiniz.

Ne Kadar Oynak? Bakışık mı, Çarpık mı? Sivri mi, Basık mı?

*Bir makine sıradan elli kişinin yaptığı işi yapabilir.
Hiçbir makine sıradışı bir insanın yaptığı işi yapamaz.*

E. G. Hubbard

9.1 OYNAKLIK – YAYIKLIK – YAYILIM – DEĞİŞKENLİK

Bundan önceki bölümde verilerin bazı özelliklerinin incelenmesi gerektiğinden söz etmiştik. Bu özelliklerden biri olan ortalama düzeyin nasıl özetlenebileceğini gördük. Bu ortalamalardan uygun birini ya da birkaçını kullanarak elimizdeki veri kümesinin hangi düzeylerde olduğunu kısa, öz bir biçimde belirtebilmenin yanı sıra birden çok veri kümesinin ortalama düzeylerinin birbirlerine göre konumlarını da karşılaştırabilme olanağını elde ettik. Ancak bir veri kümesinin yalnızca düzeyi bakımından özetlenmesi çoğu zaman yeterli olmaz. Bu veri kümesinin ortalamadan daha küçük ya da daha büyük değerlere doğru ne denli uzaklaşabildiği, bu uzaklaşmaların ortalamanın iki yanında birbirine benzer olup olmadığı, gözlem değerleri dağılımının kaç tepeli olduğu, tepe sayısı tekse bunun sivri mi yoksa basık mı olduğu gibi özellikleri de bilmek gerekir. Bu özelliklerin ölçülmesi, birer özet değerle belirtilmesi karşılaştırmaları kolaylaştırmak bakımından uygun olur. Bu bölümde verilerin yayıklık, çarpıklık, basıklık gibi özelliklerinin nasıl ölçüleceği, ölçüm sonuçlarının nasıl özetlenebileceği konuları ele alınacaktır.

Örnek olarak bir işletmenin günlük satış miktarlarının iki aylık dağılımını birbirleriyle karşılaştırmak istediğimizi düşünelim. Bulduğumuz ortalama düzey değerleri birbirine çok yakın çıkarsa bu iki ayın satışlarının birbirine çok yakın olduğunu söyleyebiliriz, ama bu bulgu, iki aydaki günlük satış dağılımlarının her bakımdan birbirine benzediği anlamına gelme-

yebilir. Bu aylardan ilkinde günlük satış miktarları ortalamanın pek az altında ya da pek az üstünde değerler alabilirken ikincisinde ortalamadan uzaklaşmalar büyük boyutlara varıyorsa ortada yakıcı bir sorun var demektir. Böyle bir durumda işletmenin satışlarının büyük oynamalar gösterdiği, bir başka deyişle kararlılıktan uzaklaştığı söylenebilir.

Başka bir örnek alalım. Diyelim ki bir tüketiciyi koruma derneği, iki ayrı marka televizyonun tüplerinin dayanıklılığı ile ilgilenmektedir. Her iki markadan rassal olarak seçilen dörder televizyonun tüplerinin dayanma süreleri ölçülmüş, *A* markası için 80, 83, 85, 88 ay; *B* için 68, 79, 90, 101 ay bulunmuştur. Her iki marka için de ortalama dayanıklılık süresi 84 aydır. (Burada hangi ortalamanın kullanıldığını bulmak okuyucuya bırakılmıştır.) Bu durumda iki marka arasında dayanıklılık süresi bakımından hiç fark yoktur diyebilir miyiz? *B* markasının dayanıklılık süresinin çok daha büyük bir değişkenlik gösterdiği ortadadır.

Kişi başına ulusal gelir düzeyleri aynı olan iki ülkede gelir dağılımlarının da aynı olduğu söylenemez. Yurttaşlar arasındaki gelir farklılıkları her iki ülkede aynı olmayabilir. Birinde gelir eşitliğe daha yakın dağılmışken ötekinde kişilerin gelir düzeyleri arasında uçurumlar bulunuyorsa bu iki ülkede gelir dağılımının benzeşmezliği farklı sorunlara yol açabilir, dolayısıyla farklı iktisat politikalarına gereksinim duyulabilir.

Bu örneklerden de görüleceği gibi veri kümelerinin yalnızca ortalama düzeylerini bilmek yetmez. Çünkü ortalama çok basit bir modeldir. Bu verilerin dağılımlarının ortalama düzeyden ne ölçüde uzaklaştığı, dağılımın bakışık olup olmadığı, tepe sayısı, ne kadar sivri ya da basık olduğu, uçlarda aşırı değer bulunup bulunmadığı da bilinmelidir. İşte bu bölümde açıklanan ölçüler bu özellikleri yansıtmayı amaçlayan göstergeler olarak algılanmalıdır Böylece modelimiz zenginleşmiş, gerçeklik daha iyi açıklanmış olur.

Bir veri kümesinin oynaklık, yayıklık ya da değişkenlik özelliği denince anlaşılması gereken, gözlem değerlerinin sayı ekseninde ne kadar toplu ya da yayılmış biçimde bulunduklarıdır. Başka bir deyişle tekil gözlem değerlerinin ortalama düzeyinden ne ölçüde uzaklaşabildikleridir.

9.2 NASIL ÖLÇÜLÜR?

Bu özelliğin ölçülmesi çeşitli yollarla yapılabilir. Aşağıda bu ölçülerden önemlileri açıklanacaktır. Bu ölçülerden kimileri hem çok çabuk hesaplanabilir hem de kolayca anlaşılır. Ötekilerin hesaplanması için hesap makinesi ya da bilgisayar gerekir, anlaşılmaları da görece olarak biraz daha zordur. Kolaylardan başlayarak açıklamalarımızı sürdürelim.

Aralık (Ar)

Bir veri kümesinin yayıklığının en kolay hesaplanabilen ölçüsü, en büyük değerle en küçük değerin farkıdır. Buna **aralık** denir. Öyleyse aralığı şöyle gösterebiliriz:

Aralık = En büyük gözlem değeri – en küçük gözlem değeri

$$Ar = Y_{eb} - Y_{ek}$$

Çizim 7.12’de çeşitli üniversitelerin iktisat programlarına giren öğrencilerin 2002’deki taban-tavan puan aralıklarına yer vermiştik. O çizimde her bir program için *Ar* değerlerinin birer yatay doğru parçasıyla gösterildiğini söyleyebiliriz. Bu çizimden, özellikle vakıf üniversitelerindeki burssuz öğrencilerin geniş bir puan aralığında yer aldıkları gözlemlenebilir.

Yukarıdaki televizyon tüplerinin dayanıklılığı örneğinde *A* markası için $Ar = 8$, *B* markası için ise 33 aydır. Bu iki sayıya bakarak iki markanın dayanıklılık süresinin değişkenliğine ilişkin bilgi edinebiliriz. *A* markasını alan bir kimse 84 ay dolayında bir süre dayanıklılık beklerken, *B* markasını seçenler ya çok şanssız olup 80 aydan daha kısa sürede tüp değiştirirler ya da çok şanslı olup 90 aydan daha uzun bir süre bu işi yapmak zorunda kalmazlar. Bu durumda *B* markasını almanın daha **riskli** olduğu söylenebilir.

Çizelge 7.5’teki araba satışlarının 2002 yılı *Ar* değerini bulmak istersek, en çok satan Renault’nun 16,6 binlik satışından en az satan Nissan’ın 1,1 binlik satışını çıkarmamız gerekir. Bunun sonucunda $Ar = 15,5$ bin arabadır. Aynı çizelgedeki verilerle 2001 yılı için *Ar*’ın kaç olduğunu bulup iki yıl arasında satış oynaklığının nasıl değiştiğini incelemek okuyucuya bırakılmıştır.

Görüldüğü gibi bu ölçü yalnızca iki uçtaki gözlem değerlerine bağlıdır. Bunlardan birinin ya da ikisinin birden olağandışı değer almaları durumunda *Ar* da olağanüstü büyür. Ara değerler nasıl dağılırsa dağılsın iki uçtaki değer her şeyi belirlemektedir. Demek ki *Ar*, uç gözlem değerlerinden çok etkilenen bir ölçüdür. Bir başka deyişle **berk** değildir. Bunun çok önemli bazı sakıncalar yaratabileceği ortadadır. Yukarıdaki günlük satışlar örneğine dönelim. Diyelim ki Mart - Nisan aylarında günlük satış miktarları (ton olarak) incelenip şöyle bir sonuç bulunmuştur:

	Mart	Nisan
Ay içinde en çok günlük satış	39	65
Ay içinde en az günlük satış	30	29

Bu durumda *Ar* Mart ayında 9 ton, Nisan ayında 36 tondur. Yalnızca bu iki *Ar* değerini karşılaştırmak bizi yanıltabilir. Sözelimi Nisan ayındaki 65 tonluk günlük satış o ay içinde yalnız bir tek gün gerçekleşmiş, ondan sonra gelen ikinci en yüksek satışlı günde bu miktar 38 tonda kalmış olsun. Bu olağandışı günü saymazsak her iki ayın da satış aralığı birbirine eşittir. Öyleyse bu olağandışı yüksek satışlı gün özel olarak incelenmeden yalnızca *Ar* değerlerine bakıp sonuca varmak yanıltıcıdır.

Çizelge 8.1'de yer alan çeşitli ülkelerin bebek ölüm oranlarında da bu özelliğe rastlarız. 16 ülke için $Ar = \% 39$ iken Türkiye çıkarıldığında 15 AB ülkesi için bu değer $\% 4$ 'e düştüğünü görürüz. Bu özelliğinden dolayı *Ar* ölçüsü, çok çabuk hesaplanabilme, çok kolay anlaşılabilme niteliklerine karşın dikkatli kullanılmalıdır. **Olağandışı gözlemlerin** varlığı göz önüne alınmalıdır.

Öte yandan uç değerlerin doğal sınırlar içinde kaldığı veri kümelerinde *Ar*'ın bu sakıncası büyük ölçüde ortadan kalkar. Sözelimi sınav notları alttan da üstten de sınırlıdır. Belli sayıda takımın katıldığı bir şampiyonada takımların alabilecekleri toplam puanlar belli sınırlar arasında kalır. Bu tür veri kümelerinde olağandışı gözlem değerlerine pek rastlanmaz. Dolayısıyla böyle durumlarda *Ar*, karşılaştırmalar için çok elverişli olabilir.

Dördebölenler Aralığı (*DAr*)

Ar ölçüsünün uç değerlerden etkilenmesi sakıncasından kurtulmak, aynı zamanda çabuk hesaplanma, kolay anlaşılma özelliklerini yitirmemek için bir ara çözüme gidilmektedir. En büyük değerle en küçüğün farkını almak yerine, üst dördebölenle alt dördebölenin farkı alınabilir. Dördebölenler aralığı denilen bu ölçüyü şöyle gösterebiliriz:

Dördebölenler aralığı = Üst dördebölen – Alt dördebölen

$$DAr = D_u - D_a \quad \text{ya da} \quad DAr = D_3 - D_1$$

Böylece gözlem değerleri dağılımının iki ucundan gözlemlerin $\frac{1}{4}$ 'ü atılmakta, gözlemlerin tam ortadaki yarısının *aralığı* bulunmaktadır. Bu uygulama hem uç değerlerden kurtulmayı sağlamakta, hem de kolaylıkla çabuktan ödün vermemektedir.

Çizelge 8.1'deki 16 ülkenin bebek ölüm oranları için *DAr* hesaplamak istersek önce *D*'leri bulmalıyız. Bunun için 8. bölümde üç yöntemden söz etmiştik. Geleneksel yöntemle $Dar = 1$ bulunur. Bu değer hesaplanmasında Türkiye'nin dışta bırakılması sonucu değiştirmez. Bu hesaplamaların yapılıp sonucun doğrulanması okuyucudan beklenmektedir.

Çizelge 7.5'teki 2002 yılı araba satışları için geeleneksel yöntemle $DAr = 6,1$ çıkar. Bunun doğrulanması da okuyucuya bırakılmıştır.

DAr hesaplanırken gözlem değerlerinin yarısının dışlanması bir sakınca olarak görülebilir. Bu durumda daha az sayıda gözlemin atılmasına yol açan **onabölenler aralığı** kullanılabilir. 9. onabölen ile ilk onabölenin farkı olarak ölçülen bu değer gözlem değerlerinin her iki uçtan % 10 (toplam % 20) kadarının kırpılması sonucunu doğurur. Bu oranlara karşılık gelen gözlem konumları tamsayı değilse iki yol izlenebilir. Bu oranın arasına düştüğü iki değer basit ya da tartılı ortalaması alınır.

Bebek ölüm oranları örneğinde her iki uçtan % 10 ülke dışlanmaya kalkılırsa toplam ülke sayısı 16 olduğunda her iki uçtan atılacak ülke sayısı $(16 + 1) 0,1 = 1,7$ olur. Az önce sözü edilen iki yol kullanılırsa onabölenler, iki uçtaki ilk iki değer $\frac{1}{2}$ oranlarında basit ya da 0,7-0,3 oranlarında tartılı ortalaması alınarak hesaplanabilir. Bu durumda uçlardan iki ülkeden birinciyle ikincinin değerleri arasındaki farkın ya yarısı ya da % 70'i dışarıda bırakılarak onabölenler bulunabilir. Öyleyse ilk yolla onabölenler aralığı % 21, tartılı ortalamayla % 13,8 bulunur. Okuyucu bu sonuçların doğruluğunu sinamalıdır. Bu örnekte görüldüğü gibi onabölenler aralığı olağandışı Türkiye değerinin izlerini hâlâ taşımaktadır. Çünkü üst onabölenin hesabında % 43 bebek ölüm oranımız hesaba girmiştir.

Türkiye dışında 15 ülke düşünülürse onabölenler aralığı basit yolla % 1,5, tartılı ortalamayla % 1,6 çıkar.

Bu örnekte görüldüğü gibi 20'den az sayıda gözlem değerinin bulunduğu örneklerde, eğer bir ya da her iki uçta bir olağandışı değer varsa, onabölenler aralığı hâlâ sakıncalı olabilmektedir.

Araba satışları örneğindeyse onabölenlerin basit ya da tartılı ortalamayla bulunmasına bağlı olarak onabölenler aralığı da 12,45 ya da 11,44 bulunur. Bu aralığın 2001 yılı verileriyle hesaplanıp iki yılın sonuçlarının karşılaştırılması okuyucuya düşmektedir.

Dördebölenler Sapması (DS)

Kimi zaman dördebölenler aralığının yarısı alınarak dördebölenler sapması (DS) hesaplanabilir. Böylelikle alt dördebölenle üst dördebölenin ortanca- dan ortalama olarak ne kadar saptıkları gösterilebilir. Bu ölçü, dördebölenlerin ortancaya göre bakışık olmaları durumunda çok iyi bir göstergedir, ama bu bakışım özelliğinden uzaklaşıldığı ölçüde yanıltıcı olabilir. Dördebölenler sapmasını şöyle gösterebiliriz:

$$\text{Dördebölenler sapması} = \text{Dördebölenler aralığı} / 2$$

$$DS = DAr / 2 = (D_3 - D_1) / 2 = (D_u - D_a) / 2$$

Ar açıklanırken ele aldığımız günlük satışlar örneğinde Mart ayı *D*'lerinin sırasıyla 32, 34, 36; Nisan ayındakilerin de 31, 34, 35 olduğunu varsayalım. Bu durumda *DAr* her iki ay için de 4 ton, *DS* ise 2 ton olarak bulunur. Mart ayında *D*'ler *Ortc*'ya göre bakışık olduğundan iki *D*'nin ortancadan uzaklıkları da tam bu kadardır. Oysa Nisan ayında durum böyle değildir. *D_a*'nın *Ortc*'dan sapması 3, *D_a*'ninki ise 1 tondur. Dolayısıyla *DS*, Mart için çok daha uygun bir gösterge niteliğindedir.

Çizelge 8.1'deki bebek ölüm oranı değişkenini bir başka örnek olarak alırsak *DS* = 0,5 bulunabilir.

Ortalama Mutlak Sapma

Aralık, dördebölenler aralığı, dördebölenler sapması gibi ölçüler çabuk hesaplanma, kolay anlaşılma özelliklerini taşırlar, ama veri kümesinden az sayıda gözlem değerini kullanırlar. Son ikisi olağandışı gözlem değerlerinden kurtulmayı sağladıkları için yararlı birer gösterge olabilirler, ama eğer kümede olağandışı gözlem bulunmuyorsa, çok eksik bilgiyle iş gördükleri de açıktır. Böyle olağandışı gözlem içermeyen veri kümeleri için bütün gözlem değerlerini, dolayısıyla veri kümesinin taşıdığı bütün bilgiyi kullanan ölçülere gerek duyulabilir. Bu amaçla gözlem değerlerinin her birinin belli bir ortalama dan farklarının ortalaması alınarak sapmalar ortalaması ya da ortalama sapma hesaplanabilir. Burada kendisine göre sapmaların hesaplandığı ortalama genellikle *AO* olur. Pek sık rastlanmamakla birlikte *Ortc* da bu amaçla kullanılabilir. Buradaki açıklamalarımızda biz de ortalama düzey ölçüsü olarak genellikle *AO* kullanacağız, ancak bunun dönüştürülmüş gözlem değerlerinin aritmetik ortalaması olabileceği de gözden uzak tutulmamalıdır.

Herhangi bir veri kümesinde *AO*'dan sapmaların toplamının sifıra eşit olduğunu görmüştük. Dolayısıyla bu sapmaların ortalaması da her zaman sıfır çıkar. Eğer anlamlı bir ortalama sapma bulmak istiyorsak bu sapmaların işaretlerinden kurtulmamız gerekir. Bu da iki yolla yapılabilir. Ya sapmaların işaretleri göz ardı edilerek mutlak değerler kullanılır, ya da kareleri alınarak gözlemlerin hepsi artı işaretli değerlere dönüştürülür. Bunlardan ilki bize ortalama mutlak sapma, ikincisi varyans denilen ölçüyü verir. Varyansın karekökü de standart sapma adıyla anılır.

Ortalama mutlak sapma = Sapmaların mutlak değerlerinin ortalaması

$$OMS = \sum |Y_i - \mu| / N = \sum y_i / N$$

Anakütle için yapılan bu hesaplamada *N* yerine *n* geçirilirse *OMS*, örneklem için elde edilmiş olur. *OMS*, bütün gözlem değerlerini hesaba katar,

ama daha ileri düzeylerde çözümlemelere elverişli değildir. Bu nedenle kullanım alanı sınırlıdır.

Çizelge 8.1'deki 16 ülkenin bebek ölüm oranları için $AO = 8,6875$ alınırsa $OMS \approx 4,3$ bulunur. Ama bu yanıltıcıdır, çünkü Türkiye'den başka bütün ülkelerin bebek ölüm oranları AO 'dan küçük, dolayısıyla AO 'dan sapmaları eksi işaretli çıkar. Bunların mutlak değerleri arasında da bir tek İsveç'inki OMS 'den büyüktür. Aynı hesaplamaları Türkiye dışındaki 15 ülke için yaparsak, $AO = 6,4$, $OMS \approx 0,8$ bulunur. Bu da tek tek ülkelerin bebek ölüm oranlarının % 6,4 olan AO 'dan ortalama olarak % 0,8 saptığı anlamına gelir. Bu örnekten anlaşılacağı üzere olağandışı değerler içeren çarpık veri kümeleri için AO gibi OMS de pek uygun değildir. Bütün bunları doğrulamak okuyucuya bırakılmıştır.

Standart Sapma – Varyans

Gözlem değerlerinin ortalamadan sapmalarının karelerinin alınması ise **kareli ortalama** kavramına başvurulması demektir. KO hesaplarırken kullanılan gözlem değerleri yerine bu kez ortalamadan sapmalar geçirilecek olursa standart sapma denilen yayıklık ölçüsüne ulaşılır. Bu ölçü, anakütle için hesaplanırsa σ (*sigma*) ile, örneklem için hesaplanırsa s ile gösterilir.

Standart sapma = Aritmetik ortalamadan sapmaların kareli ortalaması

$$\sigma = \sqrt{\frac{\sum (Y_i - \mu)^2}{N}}$$

Kimi durumlarda sapma karelerinin aritmetik ortalaması alınmakla yetinilir, sonucun yeniden karekökü alınmaz. Bir başka deyişle standart sapmanın karesi bulunur. Bu ölçüye de **varyans** adı verilir:

Varyans = Aritmetik ortalamadan sapmaların kareli ortalamasının karesi

$$\sigma^2 = \frac{\sum (Y_i - \mu)^2}{N}$$

Varyansın ölçü birimiyle, gözlemlerin ölçü birimi aynı değildir. Gözlem değerlerinin kareleri alınırken ölçü birimlerinin de kareli biçimine geçilmiş olur. Dolayısıyla gözlem birimleri metre ile ölçülüyorsa varyans m^2 ile, gözlemler TL ile ölçülüyorsa varyans TL^2 (o da ne demekse?) ile belirtilmiş olur. Standart sapmada böyle bir sorun yoktur.

Anakütle varyansı, gözlemlerin kareli ortalamasının karesinden, aritmetik ortalamasının karesi çıkarılarak da bulunur:

$$\sigma^2 = KO^2 - AO^2$$

Bu özelliğin geçerliliğini şöyle gösterebiliriz:

$$\begin{aligned}\sigma^2 &= \frac{\sum (Y_i - \mu)^2}{N} = \frac{\sum (Y_i^2 - 2\mu Y_i + \mu^2)}{N} = \frac{\sum Y_i^2}{N} - 2\mu \left(\frac{\sum Y_i}{N} \right) + \frac{N\mu^2}{N} \\ &= \frac{\sum Y_i^2}{N} - 2\mu\mu + \mu^2 = \frac{\sum Y_i^2}{N} - 2\mu^2 + \mu^2 = \frac{\sum Y_i^2}{N} - \mu^2 = KO^2 - AO^2\end{aligned}$$

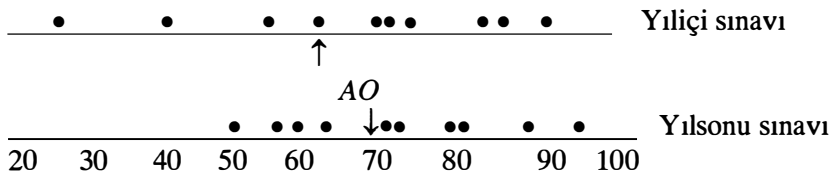
Gözlem sayısı biraz büyüyünce varyans ya da standart sapma hesaplamasının elle ya da yalnızca dört işlem yapan hesap makineleriyle yürütülmesi güçleşir. Bunun için standart sapma hesaplamak üzere özel olarak programlanmış hesap makinesi ya da bilgisayar kullanmak gerekir. Ancak okuyucunun varyans ya da standart sapma kavramlarını özümseyebilmesi amacıyla bu işlemleri az sayıda veri kullanarak, özel hesap makinesi ya da bilgisayara gerek duymadan yapması uygun olur. Aşağıda bu işe yarayan örnekler verilmiştir.

Bir sınıftaki öğrencilerin yılıçi, yılsonu sınavlarında aldıkları notlar aşağıdaki gibiiken bu sınıfın başarı düzeyinin yılıçi-yılsonu karşılaştırmasını yapalım.

ÇİZELGE 9.1 Varsayılan bir sınıfın sınav notları (Tam not 100)

	Anı	Banu	Can	Çağ	Dağ	Efe	Fuat	Giz	Hür	Işıl
Yılıçi	74	76	63	57	92	40	86	73	84	25
Yılsonu	72	80	65	55	95	62	81	71	89	60

Kaba bir incelemeyle kimi öğrencilerin notlarının yükseldiğini, kimi öğrencilerin notlarının da düştüğünü görürüz. Sınıfın not ortalamasını hesaplasak yılıçi sınavında 67, yılsonu sınavında da 72 buluruz. Bu iki dağılımı görsel olarak izleyebilmek için Çizim 9.1'i düzenleyelim.



ÇİZİM 9.1 Varsayılan bir sınıfın sınav notları (Tam not 100)

Bu çizime baktığımızda ne görüyoruz? Sınıfın notları ortalama olarak yükselmiştir. Ancak bir gelişme daha göze çarpmaktadır. Yılsonunda notlar yılıçine oranla birbirine daha çok yaklaşmıştır, başka bir deyişle not dağılımının yayıklığı azalmıştır. Gerçekten de *Ar* yılıçinde $92 - 25 = 67$ iken yılsonunda $95 - 55 = 40$ 'a düşmüştür. *DA*, yılsonunda yılıçinin yaklaşık yarısına inmiştir. (Bunun doğrulanması okuyucuya bırakılmıştır.)

Çizelge 9.2 yardımıyla bu notların, *OMS*, σ^2 , σ değerleri hesaplanmıştır.

ÇİZELGE 9.2 Varsayılan bir sınıfın sınav notlarının *OMS*, σ^2 , σ değerlerinin hesaplanmasına yardımcı çizelge

Yılıçi sınavı				Yılsonu sınavı			
X_i	$x_i = X_i - \mu_x$	$ x_i $	x_i^2	Y_i	$y_i = Y_i - \mu_y$	$ y_i $	y_i^2
74	7	7	49	72	0	0	0
76	9	9	81	80	8	8	64
63	-4	4	16	65	-7	7	49
57	-10	10	100	55	-17	17	289
92	25	25	625	95	23	23	529
40	-27	27	729	62	-10	10	100
86	19	19	361	81	9	9	81
73	6	6	36	71	-1	1	1
84	17	17	289	79	7	7	49
25	-42	42	1764	60	-12	12	144
670	0	166	4050	720	0	94	1306

Son satırdaki toplamlar kullanılarak şu değerler elde edilir:

	<i>OMS</i>	σ^2	σ
Yılıçi sınavı	16,6	405,0	20,1
Yılsonu sınavı	9,4	130,6	11,4

Buradan anlaşılacağı gibi yılsonu sınav notlarının dağılımı her üç ölçüye göre de daha az yayıklık göstermektedir. Öğrencilerin notları biraz daha birbirine yaklaşmıştır. Öğrenciler arasında başarı durumu ortalama düzeyden daha az sapma göstermektedir.

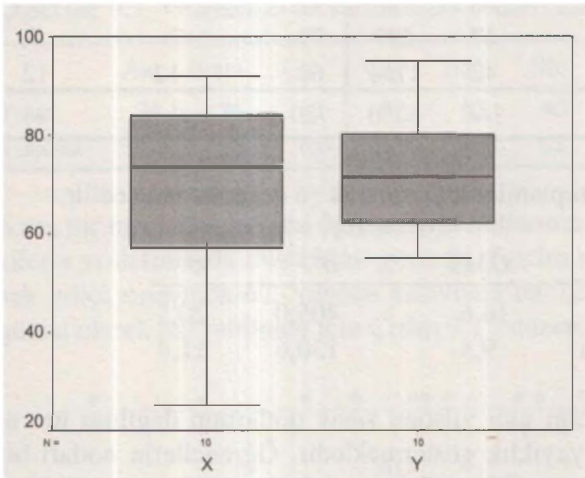
Benzer gözlemleri Çizim 9.2'deki dal-yaprak, Çizim 9.3'teki kutu çizimleriyle de yapabiliriz. Her iki gösterimde de göze çarpan bir özellik

yılıçi notlarındaki *Ortc*'nin yılsonu sınavında düşmesidir. Oysa az önceki hesaplarımıza göre *AO* artmıştı. Bu da bize *AO*'nun uç değerlerden etkilene özelliğini bir kez daha anıttmaktadır.

Gerek dal-yaprak, gerek kutu çiziminden gördüğümüz gibi yıl içi sınavının yayıklık derecesi yılsonu sınavında azalmış, öğrencilerin notları birbirine daha yakın konumlarda toplanmıştır. Kutu çiziminde iki uç değeri arasındaki uzaklık *Ar*, ortadaki kutunun yüksekliği *DA*'dır. Her ikisinin de yılsonu sınav notları içindeki değeri yılıçine göre düşmüştür. Bunu ilgili uzunluklara bakarak anlayabiliyoruz.

Yılıçi notları <i>X</i>			Yılsonu notları <i>Y</i>		
(25)					
1	2	9	5	1	
3	46	8	01	3	
(3)	346	7	129	(3)	
4	3	6	025	4	
3	7	5	5	1	
2	0	4			
Bir			On		
On			Bir		

ÇİZİM 9.2 İki sınavın notlarının dal-yaprak gösterimi



ÇİZİM 9.3 İki sınavın notlarının kutu çizimi

Buraya kadar σ^2 , σ hesaplamaları, anakütledeki bütün gözlem değerleri dikkate alınarak yapılmıştır. Eğer hesaplama örneklemden elde edilen verilerle yapılarak, bulunan standart sapma ya da varyans, anakütleninkinin yerine onun bir **kestirimcisi** olarak kullanılacaksa, hesaplamalarda küçük bir değişikliğin yer alması gerekir. Sapma kareleri örneklemin gözlem sayısına değil, onun bir eksiğine bölünür. Bunun nedeni, olasılık kuramı öğrenildikten sonra tam olarak anlaşılabilir. Burada şu kadarını söyleyelim. Örneklemden bulunan AO , anakütle AO 'suymuş gibi kullanılırsa örneklemdeki değerlerin bir tanesi bu anakütle ortalamasını verecek biçimde bir değer almak zorundadır. Yani serbestçe seçilmiş sayılamaz. Dolayısıyla $n - 1$ tane gözlem değeri serbestçe seçilmiştir. Zaten bu nedenle $n - 1$ sayısına **serbestlik derecesi** denir. Demek ki ortalamadan sapmaların da $n - 1$ tanesi serbestçe belirlenmiş olmaktadır. Böylece sapmalar toplamı n 'ye değil, serbestlik derecesi $n - 1$ 'e bölünmelidir. Ancak o zaman örneklemin standart sapması olan s , anakütle standart sapması olan σ 'nın **yansız bir kestirimcisi** olabilir.

İstatistik işlemleri yapan hesap makinelerinde standart sapma hesabı için iki ayrı düğme bulunmasının nedeni budur. Düğmelerden biri anakütle standart sapması içindir. Öbür düğme anakütle standart sapmasının yansız bir kestirimcisi olan örneklemin standart sapmasını bulmakta kullanılır. Anakütle standart sapmasının örneklemden elde edilen kestirimcisini şöyle yazabiliriz (Buradaki $\hat{\sigma}$, bulunacak değerin σ 'nın kestirimcisi olduğunu gösterir).

$$\hat{\sigma} = s = \sqrt{\frac{\sum (Y_i - \bar{Y})^2}{n - 1}}$$

Gerek standart sapma, gerek varyans, istatistikte en çok kullanılan yayılım ölçüleridir. Olasılık kuramında, örneklem almada, normal dağılımla ilgili konularda, iki değişken arasındaki bağlanımın ya da ilişkinin ölçülmesinde bu kavramlara sık sık başvurulur.

Standart sapma ortada toplanmış bakışık dağılımlar için çok uygun bir yayıklık göstergesidir. Ancak gözlem değerlerinin AO 'ya göre bakışımından uzaklaşması durumunda kullanılabilirliği azalır. AO konusunu incelerken gördüğümüz gibi AO da ortada toplanmış çarpık olmayan dağılımlar için çok uygun bir düzey ölçüsüyü. Standart sapmanın hesaplanmasında AO 'dan yararlanıldığına göre, çarpık dağılımlarda benzeri sakıncaların, bu değişkenlik ölçüleri için de geçerli olması doğaldır. Daha önce ele aldığımız dönüştürme yollarıyla bakışım sağlanabilirse, varyans ya da standart sapma kullanmanın sakıncaları da ortadan kalkmış olur. Bu konuya ileride döneceğiz.

Yukarıdaki sözlerden bakışık olmayan dağılımlarda standart sapmanın hiç bir işe yaramayacağı anlamı çıkarılmalıdır. Dağılımın biçimi ne olursa olsun standart sapmadan yararlanılabilir. Bu konuda bize **Çebişev eşitsizliği** yardımcı olur. Bir Rus matematikçisi olan Çebişev (Чебышев), kendi adıyla anılan şu eşitsizliğin her türlü dağılım için geçerli olduğunu kanıtlamıştır:

Çebişev eşitsizliği: k en az 1 iken, dağılım biçimi ne olursa olsun her veri kümesinde, AO 'nun her iki yanında k standart sapma kadar uzanan aralıktaki gözlem sayısının, toplam gözlem sayısına oranı **en az** $1 - 1/k^2$ kadardır. Başka bir gösterimle

$$k \geq 1 \text{ iken } (\mu \pm k\sigma \text{ aralığındaki gözlem sayısı}) / N \geq \left(1 - \frac{1}{k^2}\right).$$

Bu eşitsizliğe göre gözlem değerlerinin **en az**

$$1 - (1/2)^2 = 1 - 1/4 = 3/4 = 0,75 \quad \text{kadarı} \quad \mu \pm 2\sigma \text{ arasında,}$$

$$1 - (1/3)^2 = 1 - 1/9 = 8/9 = 0,89 \quad \text{kadarı} \quad \mu \pm 3\sigma \text{ arasında yer alır.}$$

Bu söylediklerimizi bir örnek üzerinde açıklayalım. Bu amaçla daha önce çarpık olduğunu gördüğümüz Çizelge 8.2'deki nüfusun yaşa göre dağılımını gösteren verileri kullanalım. Ancak burada tek tek herkesin yaşlarını göremiyoruz. Belli yaş aralıklarında nüfusun % kaçının bulunduğunu anlayabiliyoruz. Böyle kümelenmiş veriler için standart sapmayı şöyle bulabiliriz.

$$\sigma = \sqrt{\sum f_i (Y_i - \mu)^2 / \sum f_i}$$

Burada f_i , i 'ninci dilimdeki sıklığı, Y_i , i 'ninci dilim aralığının orta noktasındaki yaşı gösterir. Örneğimizde sıklık toplamı % 100, yani 1 ettiğine göre paydadaki terimden de kurtuluruz. Bu durumda hesaplamayı yukarıda gördüğümüz kestirme yolla şöyle de yapabiliriz.

$$\sigma = \sqrt{KO^2 - AO^2} = \sqrt{\sum f_i Y_i^2 - \mu^2}$$

Böyle öbeklenmiş verilerle AO hesaplarken sözünü ettiğimiz bir varsayımı burada da kullandığımız dikkatli bir okuyucunun gözünden kaçmamıştır. Bu varsayım belli bir yaş dilimindeki herkesin aynı yaşta olduğuydu. Bu yaş için de o aralığın orta noktasını kullanıyoruz. Örnek olarak $0 - < 10$ yaş diliminin orta noktasını 5 yaş, $10 - < 20$ yaş diliminin orta noktasını 15 yaş, vb. alıyoruz. “Bilinmeyenler” satırını en yaşlılara katarak “60+” yaş öbeğinin üst sınırını 80 olarak belirleyip öbek aralıklarının orta noktalarını

kullanalım. Daha önce bu varsayımlar altında $\mu = 26,56$ bulmuştuk. Bu durumda

$$\sigma^2 = 0,228 \times 5^2 + 0,232 \times 15^2 + 0,176 \times 25^2 + 0,134 \times 35^2 + 0,088 \times 45^2 + 0,07 \times 55^2 + 0,072 \times 70^2 - 26,56^2 = 369,3664$$

$\sigma = 19,22$ bulunur.

Eğer dağılımın tam bakışık olduğuna güven duyuyorsak Çebişev eşitsizliğiyle bulunan oranları ortalamanın iki yanına eşit olarak bölebiliriz. Burada yaş dağılımının bakışık olmadığı biliniyor. Yine de Çebişev eşitsizliği bize doğru bilgiler verir. Ancak dağılımın çarpık olması nedeniyle bu oranları iki yana eşit olarak bölemeyiz.

Bu örnekte $26,56 - 19,22 = 7,34$; $26,56 + 19,22 = 45,78$ olduğuna göre $7,34 - 26,56$ yaş arasındaki nüfusla $26,56 - 45,78$ yaş arasındaki nüfusun birbirine eşit olduğunu söyleyemeyiz ama “bütün nüfusun en az % 75’i $26,56 \pm 2 \times 19,22 = 0 - 65$ yaş aralığındadır, başka bir deyişle nüfusun en az % 75’i 65 yaşından küçüktür” diyebiliriz.

Eğer elimizdeki dağılım, daha sonra ayrıntılarıyla inceleyeceğimiz normal dağılım ölçülerine uygunsa o zaman Çebişev eşitsizliğinden daha güçlü söylemler kullanabiliriz. Böyle bir durumda $\mu \pm k\sigma$ aralığındaki gözlemlerin oranını tam olarak bilebiliriz.

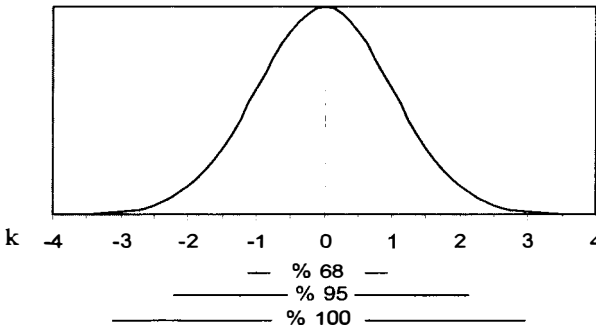
Verilerimiz **normal dağılıma** uygunsa bunların

~ % 68’i $\mu \pm 1\sigma$ arasında,

~ % 95’i $\mu \pm 2\sigma$ arasında,

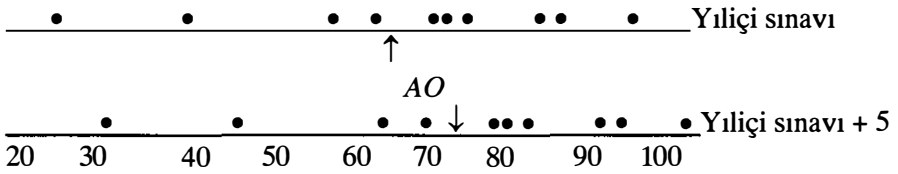
~%100’ü $\mu \pm 3\sigma$ arasında

yer alır. Bu konuyu ileride daha ayrıntılı inceleyeceğiz. Şimdilik yukarıda gösterilen oranların yaklaşık olduğunu belirtmekle yetinelim. Söz konusu oranlar ayrıca Çizim 9.4’te de gösterilmiştir.



ÇİZİM 9.4 Normal dağılımda $\mu \pm k\sigma$ arasındaki alanlar

Gözlem değerlerinin her birine sabit bir sayı ekleme (ya da çıkarmanın) AO üzerindeki etkilerini görmüştük. Benzeri incelemeyi standart sapma için de yapalım. Yukarıdaki iki sınavlı örnekte bütün öğrencilerin notlarına 5 eklendiğini düşünelim. Sınıfın not ortalaması da 5 birim artacaktır ama standart sapma değişmeyecektir. Çünkü sınıfın ortalama not düzeyi yükselmiştir ama yayıklığında bir değişiklik olmamıştır. Çizim 9.5 ortalamanın kaydığını ama yayıklığın aynı kaldığını gözler önüne sermektedir. (Okuyucu bütün notları 5 birim yükseltip standart sapmayı yeniden hesaplamalı, değerinin değişmediğini görmelidir.)



ÇİZİM 9.5 Varsayılan bir sınıfın yılıçi sınav notlarına 5 eklenmesi

Bunu daha genel olarak da gösterebiliriz. Bütün gözlemlere a sabit sayısını eklersek AO 'nun a kadar artacağını biliyoruz. Öyleyse gözlem değerlerimiz $Y_i + a$ olacak, $AO = (\mu + a)$ bulunacaktır. Bunları eşitlikte yerine koyarsak σ 'nın değişmediğini görürüz.

$$\sigma = \sqrt{\frac{\sum [Y_i + a - (\mu + a)]^2}{N}} = \sqrt{\frac{\sum (Y_i - \mu)^2}{N}}$$

Peki bütün notlar %5 yükseltirse (hepsi 1,05 ile çarpılsa) σ değişir mi? Bu durumda 40 alan birine 2 not eklenirken 80 alana 4 not eklenecektir. Demek ki düşük notlara daha az, yüksek notlara daha çok ekleme yapılmaktadır. Bu da dağılımın yayılımını arttıracak, σ 'nın değerini büyütecektir. Gerçekten de gözlem değerlerimizi a sabit sayısı ile çarparsak aY_i buluruz, aritmetik ortalamamız da $a\mu$ olur. Bunları formülde yerine koyarsak σ 'nın a ile çarpıldığını görürüz. Yeni standart sapmayı σ' ile gösterelim.

$$\sigma' = \sqrt{\frac{\sum (aY_i - a\mu)^2}{N}} = \sqrt{\frac{a^2 \sum (Y_i - \mu)^2}{N}} = a \sqrt{\frac{\sum (Y_i - \mu)^2}{N}} = a\sigma$$

Demek ki notlar % 5 yükseltirse ($a = 1,05$ olsa) σ da % 5 büyür.

Değişkenlik Katsayısı (DK)

Buraya kadar gördüğümüz yayıklık ölçüleri bir dağılımın yayıklığını, değişkenliğini aynı ölçü birimi cinsinden gösterir. Varyans bunun dışında sayılabilir ama o da aynı ölçü biriminin karesiyle ölçülür. Böyle bir özellik, iki farklı dağılımın yayılımlarını karşılaştırırken sorunlar çıkarabilir. İki dağılım farklı ölçü birimleriyle (diyelim biri TL, öbürü € ile) ölçülmüşse, standart sapmalar ya da varyanslar karşılaştırılmaz. Daha ileri giderek diyebiliriz ki iki dağılım aynı ölçü birimiyle ölçülmüş gözlemlerden oluşsa bile eğer bunların AO 'ları birbirine eşit ya da yakın değilse yine sorunlar doğar. İki pay senedinin fiyatlarının oynaklık derecesini karşılaştırmak istediğimizi düşünelim. Birinin borsadaki ortalama fiyatı 1 m. TL, ötekini 10 m. TL olan iki pay senedinin belli bir sürede alım-satımına konu olan fiyatlarının standart sapmaları karşılaştırılabilir mi? İlk pay senedi için $\sigma_1 = 100$ bin TL, ikincisi için de $\sigma_2 = 500$ bin TL olsun. Hangi pay senedi daha risklidir? Yalnızca σ değerlerini karşılaştırsak yargımız yanıltıcı olur, çünkü Çebişev eşitsizliğine göre fiyatların en az %75'i ilk pay senedinde 0,8 – 1,2 m. TL arasında oynarken, ikincide 9 – 11 m. TL arasında değişecektir. Buradan anlaşılacağı gibi ilk pay senedi *görelî* olarak daha risklidir. Çünkü beklenen fiyatların en az % 75'inin oynaklık aralığı, 0,8 m. TL olan alt sınırın yarısı kadardır ($(1,2 - 0,8) / 0,8 = 0,5$). Oysa ikinci pay senedinde bu oran 2/9, yani yaklaşık 0,22 kadardır. Öyleyse bize *görelî* riski ölçecek bir yayıklık ölçüsü gereklidir. İşte bu ölçü değişkenlik katsayısıdır (DK). Standart sapmanın aritmetik ortalamaya oranlanmasıyla bulunur.

Değişkenlik katsayısı = Standart sapma / Aritmetik ortalama

$$DK = \sigma / \mu$$

Örneğimizde iki hisse senedinin değişkenlik katsayıları $DK_1 = 0,1$, $DK_2 = 0,05$ olarak bulunur. Dolayısıyla ilk pay senedinin *görelî* riski ikincinin iki katıdır diyebiliriz.

DK 'nin karesi bize *görelî varyans* denilen kavramı verir. Bu kavram, *görelî* sapmalar üzerinden hesaplanan varyanstan başka bir şey değildir. Sapmaları $Y_i - \mu$ ile gösterdiğimizizi anımsarsak, *görelî* sapmaları da $(Y_i - \mu) / \mu$ ile gösterebileceğimiz ortaya çıkar. Bunların kareli ortalaması *görelî varyansı* verir (Yamane, 1973:78-79).

$$\text{Görelî varyans} = \frac{\sum \left(\frac{Y_i - \mu}{\mu} \right)^2}{N} = \frac{\sum (Y_i - \mu)^2}{N \mu^2} = \frac{\sigma^2}{\mu^2} = DK^2$$

Gerçek yaşamda karşılaştığımız çoğu dağılımda μ ile σ birlikte değişirler. Dolayısıyla *DK* yaklaşık olarak aynı kalır. Böyle durumlarda aynı olguyla ilgili olarak değişik zamanlarda ya da değişik yerlerde toplanan verilerin *DK*'leri hesaplanarak birbirlerine yakın olup olmadıklarına bakılır. Bu da veri derlemenin bir tür denetimidir.

DK, örneklemin varyansından (s^2) anakütle varyansının (σ^2) kestirimi yapılırken, kestirimin güvenilirliğini denetlemede işe yarar. Eğer *DK* büyük çıkıyorsa, dağılım çok çarpık olabilir, bu da güvenilir σ kestirimi yapabilmek için örneklem büyüklüğünün yüksek olmasını gerektirir.

Görelî Fark (GF)

İki ya da daha çok veri kümesinin ortalamaları arasındaki fark sayısal olarak dile getirildiğinde pek anlamlı olmayabilir. Sözgelimi bir futbol liginde takımların attığı ortalama gol sayısı ile başka bir ligdeki ortalama arasındaki fark tek başına bize çok açıklayıcı gelmeyebilir. Oysa iki ligdeki bütün takımların attığı gol sayılarının tek bir standart sapmasını hesaplayıp, iki ligin ortalamaları arasındaki farkı bu yayılım ölçüsüne bölsek, söz konusu farkın görelî bir büyüklüğünü elde ederiz. Bu da bize farkın kaç standart sapmaya karşılık geldiğini gösterir. Öyleyse şunu yazabiliriz.

$$GF = |\mu_1 - \mu_2| / \sigma$$

Genellikle $GF = 0,2$ ise düşük, $0,5$ ise orta, $0,8$ ise yüksek olarak nitelendirilir (Spatz, 2001:73).

2002-03 döneminde her birinde 18 takım olan Türkiye, Almanya, İtalya liglerinde atılan ortalama gol sayıları (*Radikal Futbol*, 13 Mayıs 2003, 4 Haziran 2003'teki verilere göre) sırasıyla 47,39; 43,22; 40,67 olmuş, üç ligin ortak standart sapması 10,89 bulunmuştur. Ortalama gol farkları Türkiye-Almanya arasında 4,17, Türkiye-İtalya arasında 6,72, Almanya-İtalya arasında 2,55 kadardır. Bu farklar az mıdır, çok mudur kestiremeyiz. Ama *GF* değerleri hesaplanırsa sırasıyla 0,38; 0,62; 0,23 bulunur. Demek ki Türkiye liginde takım başına atılan gol sayısının İtalya ortalamasından ortalayla yüksek arasında, Almanya ortalamasından düşükle orta arasında fazla olduğunu söyleyebiliriz. Almanya ortalamasıysa İtalya'ninkinden düşük ölçüde fazladır.

9.3 ÇARPIKLIK ÖLÇÜLERİ

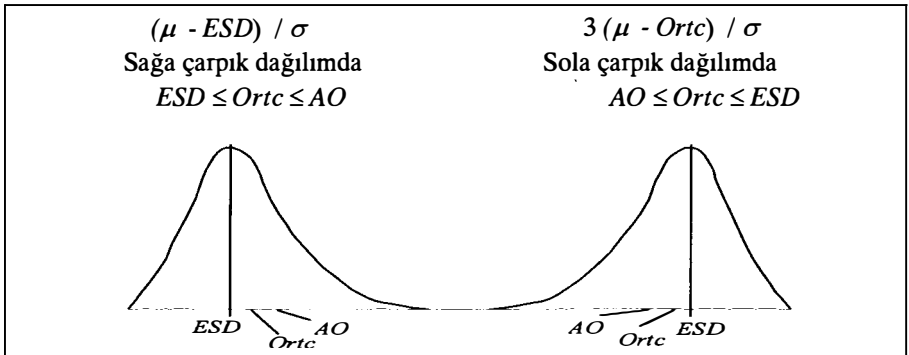
Bir gözlemler kümesinin incelenmesi gereken özelliklerinden ortalama düzey ile yayılım özellikleri görmüştük. Şimdi de dağılımın biçim özelliklerine değinecek, bakışım özelliğini bu altbölümde, basıklık özelliğini bir sonrakinde inceleyeceğiz.

Gözlem değerlerinin ortalamanın iki yanında eşit uzaklıklarda benzer biçimde bulunduğu dağılımlara **bakışık**, daha çok küçük değerler dolayında yoğunlaştığı, büyük değerlere, yani sağa doğru gittikçe seyredildiği dağılımlara **sağa çarpık**, tersine büyük değerler dolayında yoğunlaştığı, küçük değerlere doğru, yani sola gidildikçe daha büyük aralıklara rastlandığı dağılımlara **sola çarpık** dağılımlar denir. Burada inceleyeceğimiz ölçütler bir dağılımın çarpıklığının hem yönünü göstermeli, hem büyüklüğünü ölçebilmelidir.

Ortalamalarla Çarpıklık

Bütünyle bakışık bir dağılımda AO , ESD , $Ortc$ gibi ortalamaların hepsi birbirine eşittir. Çarpıklık arttıkça bu üç ortalamanın değerleri de birbirinden uzaklaşır. Sağa çarpık dağılımlarda $Ortc$, ESD 'den büyüktür, AO ise ikisinden de büyüktür. Sola çarpık dağılımlarda ise durum tam tersinedir. AO bu üçünün en küçüğü, ESD ise en büyüğüdür. $Ortc$ yine ikisi arasında yer alır. Öyleyse bu üç ortalamanın birbirlerine göre konumlarına bakılarak çarpıklığın yönüne ilişkin bilgi edinilebilir. Çizim 9.6'da bütün bu özellikler gösterilmiştir.

Çarpıklığın yalnızca yönünü bilmek yeterli olmayabilir, büyüklüğünün de bilinmesi gerekebilir. Bu ortalamalar arasındaki ilişkileri kullanarak görelî bir çarpıklık ölçüsünü *Karl Pearson* şöyle göstermiştir (Gürtan, 1987:329; Kendall, Stuart, Ord, 1987:106):



ÇİZİM 9.6 Çarpık dağılımlarda üç ortalamanın birbirlerine göre konumları

Ortalamalara dayalı çarpıklık ölçüleri

Aritmetik ortalama - Ensıkdeğer

Standart sapma

$$(\mu - ESD) / \sigma$$

3 (Aritmetik ortalama - Ortanca)

Standart sapma

$$3(\mu - Ortc) / \sigma$$

Bakışık dağılımlarda bu ölçüler sıfıra eşittir. Sağa çarpık dağılımlarda artı, sola çarpık olanlarda ise eksi değer alırlar. Çarpıklık arttıkça bu ölçülerin aldığı değerler de sıfırdan uzaklaşır. Ölçü biriminden bağımsız olan bu ölçüler yalnızca ortalamalara dayanır. Bakışma ilişkin çabuk bir bilgi edinmek için kullanılabilirler. Tek tepeli dağılımlarda yukarıdaki ilk ölçü için

$$[(\mu - ESD) / \sigma] \leq \sqrt{3}$$

olduğu gösterilmiştir (Kendall, Stuart, Ord, 1987:106). Kümelenmiş verilerde *ESD*'nin bulunmasında doğabilecek güçlükler, bu ölçülerden ilkinin kullanımında caydırıcı olabilir.

Çizelge 8.2'deki nüfusun yaş dağılımında $\mu = 26,56$, $\sigma = 19,22$ idi. $Ortc = 22,27$, $ESD = 10,67$ bulgularını da bunlara eklersek ortalamalara dayalı çarpıklık ölçüleri, bu dağılım için şöyle bulunur.

$$\begin{aligned} (\mu - ESD) / \sigma &= (26,56 - 10,67) / 19,22 = 0,83 \\ 3(\mu - Ortc) / \sigma &= 3(26,56 - 22,27) / 19,22 = 0,67 \end{aligned}$$

Bu iki ölçü birbirine eşit ya da yakın çıkarsa çarpıklığın oldukça az olduğu söylenebilir. Bu iki ölçüyle yapılan incelemeler genellikle yeterli bulunmaz. Daha iyi çarpıklık ölçülerine aşağıda değinilecektir.

Dördebölenlerle Çarpıklık

Bakışık dağılımlarda dördebölenlerden her birinin *Ortc*'ya uzaklığı birbirine eşittir. Veri kümesi sola çarpık ise alt dördebölenin *Ortc*'ya uzaklığı, üst dördebölenin *Ortc*'ya uzaklığından çok olur. Sağa çarpık dağılımlarda ise durum tersinedir.

Bakışık dağılımlarda	$Ortc - D_1 = D_3 - Ortc$
Sola çarpık dağılımlarda	$Ortc - D_1 > D_3 - Ortc$
Sağa çarpık dağılımlarda	$Ortc - D_1 < D_3 - Ortc$

Bu özellikten yola çıkan İngiliz istatistikçisi A. L. Bowley kendi adıyla anılan bir çarpıklık ölçüsü geliştirmiştir. Bu ölçü, dördebölenlerin *Ortc*'dan uzaklıklarının farkı, dördebölenler aralığına (*DAr*) bölünerek bulunur. Bir başka deyişle bu farkın *DAr* içindeki payı hesaplanır.

$$\text{Bowley çarpıklık ölçüsü} = \frac{(D_3 - Ort_c) - (Ort_c - D_1)}{DAr}$$

Ölçü biriminden bağımsız olan bu çarpıklık ölçüsü -1 ile $+1$ arasında değişir. Eksi değer sola, artı değer sağa çarpıklığı gösterir. Ölçü sıfıra ne kadar yakınsa dağılım da bakışığa o kadar yakındır, -1 ya da $+1$ sayılarına doğru sıfırdan uzaklaştıkça çarpıklığın büyüklüğü de artar.

Okuyucu, Çizelge 8.2'deki nüfusun yaş dağılımı verilerini kullanarak yapacağı bir hesaplamayla dördebölenleri kolayca bulabilir ($D_1 = 10,22$, $D_3 = 49,80$). Ort_c da daha önce $22,27$ olarak bulunmuştu. Bu durumda çarpıklığı şöyle ölçebiliriz:

$$\text{Bowley çarpıklık ölçüsü} = \frac{49,80 - 22,27 - (22,27 - 10,22)}{49,80 - 10,22} = 0,39$$

Beklemlerle (momentlerle) Çarpıklık

Bu çarpıklık ölçüleri, ileri düzeydeki duyarlı hesaplamalarda en yaygın kullanılan ölçülerdir. Veri kümesindeki bütün gözlem değerleri bu ölçülerin hesaplanmasına katılır. Dolayısıyla bu ölçüler, çarpıklığı arttıran aşırı uç değerlerden hayli etkilenirler. Bu ölçülerin nasıl hesaplandığını açıklayabilmek için önce **beklem (moment)** kavramını görmemiz gerekecektir.

Beklem, gözlem değerlerinin belli bir noktadan uzaklıklarının belli bir kuvvetinin alınarak toplanması, toplam sonucunun da gözlem sayısına bölünmesiyle bulunan değer demektir. Değişmez olarak alınan belli nokta sıfır değeri olabileceği gibi AO da olabilir. Bu belli noktanın neresi olduğuna bağlı olarak beklemler de **sıfır dolayındaki beklemler** ya da **AO dolayındaki beklemler** gibi adlar alırlar. Bu tür beklemler şöyle gösterilebilir:

Sıfır dolayındaki r 'ninci beklem

AO dolayındaki r 'ninci beklem

$$m_r = \frac{\sum Y_i^r}{N}$$

$$m_r = \frac{\sum (Y_i - \mu)^r}{N}$$

Yukarıdaki gösterimler, rassal örneklemede karşılaştığımız bir kavram olan **beklenen değer**le de yazılabilirler. *Beklem* sözcüğünün kökeni *beklenen* değer kavramıdır.

$$m_r = E(Y_i)$$

$$m_r = E(Y_i - \mu)^r$$

Bu gösterimlerden de kolayca anlaşılacağı gibi

$$m_1 = \frac{\sum Y_i}{N} = \mu,$$

$$m_2 = \frac{\sum Y_i^2}{N} = KO^2,$$

$$m_1 = \frac{\sum (Y_i - \mu)}{N} = \frac{\sum Y_i}{N} - \mu = 0$$

$$m_2 = \frac{\sum (Y_i - \mu)^2}{N} = \sigma^2 \text{ olur.}$$

Bizi asıl ilgilendiren *AO* dolayındaki 2., 3., 4. beklemlerdir. Bu beklemler kullanılarak çarpıklık, basıklık ölçülerini yazabiliriz. Bilgisayar olmadan sıfır dolayındaki beklemleri hesaplamak, *AO* dolayındakileri hesaplamaktan daha kolaydır. Bu nedenle iki tür beklemler arasında geçiş sağlayabilmek için şu eşitlikleri bilmekte yarar olabilir (Gürtan, 1987:336):

$$m_2 = m_2 - m_1^2$$

$$m_3 = m_3 - 3 m_1 m_2 + 2 m_1^3$$

$$m_4 = m_4 - 4 m_1 m_3 + 6 m_1^2 m_2 - 3 m_1^4$$

İlgili bir okuyucunun ayırımına kolayca varabileceği gibi *AO* dolayındaki m_3 'ün işareti çarpıklığın yönünü gösterebilir. Gerçekten de sola çarpık dağılımlarda küçük değerlerle *AO* arasındaki uzaklıklar, göreceli olarak ortalamanın öbür yanındaki uzaklıklardan daha büyüktür. Bir de küpleri alındığında bu değerlerin daha da büyüyecekleri açıktır. Ancak bu uzaklıkların (elbette küplerinin de) eksi işaretli oldukları unutulmamalıdır. Dolayısıyla sola çarpık dağılımlarda $m_3 < 0$ olması doğaldır. Sağa çarpık dağılımlarda da aynı beklemler değerinin artı işaretli çıkması beklenir. Öyleyse m_3 bir çarpıklık ölçüsü olarak kullanılabilir. Ancak beklemler değerleri gözlem değerlerinin ölçü birimlerine bağlı olduklarından karşılaştırmalarda kullanılmaları pek doğru olmaz. Ayrıca bir dağılım bakışık olmadığı halde bazen $m_3 = 0$ bulunabilir (Saraçoğlu, Çevik, 1989:327).

İşte bu nedenlerle beklemler değerlerini ölçü biriminden bağımsız kılmak gerekir. Bu amaçla *Karl Pearson*, *Pearson* dağılımları adıyla anılan geniş bir dağılımlar kümesi için geçerli olmak üzere, *AO* dolayındaki 2., 3., 4. beklemleri kullanarak iki ayrı katsayı türetmiştir (Kendall, Stuart, Ord, 1987:106).

$$\beta_1 = \frac{m_3^2}{m_2^3}$$

$$\beta_2 = \frac{m_4}{m_2^2}$$

Bunlardan ikincisi aşağıda **basıklık katsayısının** bulunmasında işe yarayacaktır. İlki de *Ronald A. Fisher* tarafından karekökü alınarak bir **çarpıklık katsayısına** dönüştürülmüştür. Uygulamada önce β_1 katsayısı alınıp sonra karekökü bulunmak yerine doğrudan m_3 değeri, standart sapmanın küpüne oranlanır. Böylece ölçü biriminden bağımsız (ya da ölçü birimi

taşımayan), α_3 (alfa-3) (Gürtan, 1987:340) ya da γ_1 (gamma-1) (Kendall, Stuart, Ord, 1987:107) adıyla anılan bir çarpıklık ölçüsü elde edilmiş olur.

Beklemlere dayalı çarpıklık katsayısı (ÇK)

$$\text{ÇK} = \alpha_3 = \sqrt{\beta_1} = \frac{m_3}{m_2^{3/2}} = \frac{m_3}{\sigma^3}$$

Bu çarpıklık ölçüsü bakışık dağılımlarda ($m_3 = 0$ olduğundan) sıfır değerini alır, sağa çarpık dağılımlarda +, sola çarpık dağılımlarda da – değer taşır. Her ne kadar bir üst sınırı yoksa da yukarıda gördüğümüz öteki çarpıklık ölçülerine göre daha güvenilir sonuçlar verir. $|\alpha_3| > 0,5$ olması, güçlü bir çarpıklığın göstergesi sayılabilir.

Okuyucu nüfusun yaş dağılımını gösteren verileri kullanarak AO dola-yındaki 3. beklemini hesaplırsak

$$m_3 = 5500,57$$

bulacaktır. Öte yandan $\sigma = 19,22$ olduğu bilindiğine göre

$$\alpha_3 = m_3 / \sigma^3 = 0,77$$

elde edilir. Bunun göz ardı edilemez bir çarpıklık göstergesi olduğunu söyleyebiliriz. Demek ki 1990 nüfusuna göre yaş dağılımı sağa doğru çarpıktır. Bunun da anlamı yaş ilerledikçe yaşlıların sayısının gittikçe azaldığıdır. Türkiye için başka bir dönemde ya da başka ülkeler için herhangi bir dönemde benzeri veriler derlenir de bu verilerle α_3 bulunursa iki veri kümesinin karşılaştırılması güvenilir bir biçimde yapılabilir. Böylece ülkemizde yaş dağılımının zaman içindeki gelişmesi izlenebilir ya da ülkelerin yaş dağılımları arasında karşılaştırmaya gidilebilir.

Daha önce ortalamalara dayalı çarpıklık katsayısı olarak verilen, AO ile ESD farkının σ 'ya bölünmesiyle bulunan ölçü, Pearson'un yukarıda sözü edilen β_1 , β_2 katsayılarıyla şöyle yazılabilir (Kendall, Stuart, Ord, 1987:106):

$$\frac{\mu - ESD}{\sigma} = \frac{\sqrt{\beta_1}(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)}$$

Bu ölçünün tek tepeli dağılımlarda $\sqrt{3}$ değerinden küçük olduğu daha önce belirtilmişti.

9.4 BASIKLIK ÖLÇÜLERİ

Buraya kadar bir dağılımın ortalama düzeyi, yayıklığı, çarpıklığı gibi özelliklerini ele alıp nasıl ölçtüklerini gördük. Şimdi de çarpıklık gibi yine bir biçim özelliği olan basıklıktan söz edeceğiz. Basıklık normal dağılımla karşılaştırılarak ölçülür.

Daha önce sözü edildiği gibi **normal dağılım**, bütün dağılımlar içinde çok özel bir dağılımdır. Tam bakışıktır. Uç değerlere doğru inceler. Görünüşü bir çanı andırır. Bu nedenle **çan dağılımı** ya da **çan eğrisi** diye de anılır. AO 'suyla σ 'sı bilinirse her türlü aralıkta yer alan gözlem sayılarının toplam içindeki payı bulunabilir. Bu özelliğinden dolayı da istatistikçilerin, deyim yerindeyse gözdesidir. İstatistikçiler başka dağılımların basıklıklarını bu gözdeleleriyle karşılaştırarak ölçerler. Çan dağılımı için Pearson'un yukarıda sözü edilen β_2 katsayısı 3'e eşit çıkar. Bunun için Roland A. Fisher tarafından geliştirilen ölçü, β_2 katsayısından 3 çıkarılarak bulunur. Bu ölçü α_4 (Kendall, Stuart, Ord, 1987:107) diye adlandırılır.

Beklemlere dayalı basıklık katsayısı (BK)

$$BK = \beta_2 - 3 = \frac{m_4}{m_2^2} - 3 = \frac{m_4}{\sigma^4} - 3$$

Normal dağılıma uyan veri kümelerinde $\alpha_4 = 0$ değerini alır. Bu ölçü eksi değerler aldığı anda normal dağılıma göre daha basık, artı değerler aldığı andaysa normal dağılıma göre daha sivri bir dağılımı gösterir.

Çizelge 8.2'deki yaş dağılımı verileri için AO dolayındaki 4. beklem $m_4 = 366.442,19$ bulunur. Standart sapmanın da $\sigma = 19,22$ olduğunu bildiğimize göre $\alpha_4 = 2,69$ elde edilir. Bunun da basıkça bir dağılımın göstergesi olduğu açıktır.

Çarpıklık-Basıklık Ölçüsü

Yukarıda beklemlere dayanan çarpıklık, basıklık ölçülerini gördük. Bunların her ikisi birden kullanılarak bir dağılımın ne ölçüde çan dağılımına (normal dağılıma) benzediği gösterilebilir. Bu amaçla C. M. Jarque ile A. K. Bera'nın birlikte geliştirdikleri Jarque-Bera ölçüsü (JB) kullanılabilir (Jarque, Bera, 1987). Bu ölçü şöyledir:

Jarque-Bera çarpıklık - basıklık ölçüsü

$$JB = n \left(\frac{\alpha_3^2}{6} + \frac{\alpha_4^2}{24} \right)$$

Normal dağılımda $\alpha_3 = 0$, $\alpha_4 = 0$ olduğuna göre, JB ölçüsü de sıfıra eşit çıkacaktır. Gerek çarpıklık, gerek basıklık bakımlarından normal dağılıma

benzeme azaldıkça *JB* ölçüsü de büyür. Aslında bu ölçü, örneklem büyüdükçe (*n* arttıkça) **ki-kare dağılımına** yaklaşıp. Bu nedenle de eldeki örneklemi oluşturan veri kümesinin dağılımının, normal dağılıma uyan bir anakütleden çekilip çekilmediğinin sınanması yapılabilir. Bu sınamaya Bowman-Shelton sınaması da denir (Newbold, vö., 2003:565). **Önsav sınamalarının** nasıl yapıldığını öğrenmeden bu konuyu daha ayrıntılı ele alamayız.

Özellikle örneklemelerin normal dağılmış bir anakütleden çekilmiş olup olmadıklarını anlamak için çizimlerden ya da çeşitli önsav sınamalarından yararlanılır. Bu tür çizimlere daha sonra değineceğiz. **Ki-kare, Kolmogorov-Smirnov, Lilliefors, Shapiro-Wilks** gibi sınamalar da bu amaçla kullanılır. Bu sınamalarda uygulanan yöntem gözlenen değerlerin beklenen değerlerle şu ya da bu biçimde karşılaştırılmasına dayanır.

9.5 GELİR DAĞILIMINDA ÇARPIKLIK ÖLÇÜLERİ

Gelir dağılımı, Türkiye Cumhuriyeti vatandaşları olanların meslek hastalığıdır.

Ümit Kıvanç

Gelir dağılımı kavramı çeşitli anlamlarda kullanılmaktadır. Bunlardan biri, gelirin emek ile emek-dışı üretim etmenleri arasındaki bölüşümünü, başka biri gelirin kişiler (ya da hanehalkları) arasında dağılımını, daha başkaları da iller, bölgeler, meslekler, vb. arasında dağılımını gösterir. Bu kullanımlardan ilkinde **işlevsel gelir dağılımı** ya da **gelirin bölüşümü** denir. Gelirin kişiler arasındaki dağılımına da **kişisel gelir dağılımı** ya da kısaca **gelir dağılımı** denilebilir (DPT, 1976:1). Burada da gelir dağılımı terimi bu ikinci kullanıma uymaktadır.

Bir toplumu oluşturan kişilerin gelir düzeyleri birbirine ne denli yakınsa o toplumun gelir dağılımında o denli eşitlik bulunduğundan söz edilebilir, tersine kişilerin gelirleri arasındaki farklar büyüdükçe de gelir dağılımında eşitlikten uzaklaşıldığı söylenebilir. İşte gelir dağılımındaki çarpıklık ölçüleri bu durumu saptamayı amaçlar. Aynı ölçüler gelir dağılımları dışında, servet, toprak mülkiyeti gibi dağılımlar ya da sermayenin, üretimin, satışların işletmeler arasındaki dağılımları için de geçerlidir.

Çeşitli ülkelerin gelir dağılımları incelendiğinde genellikle üç ortalama ölçüsünün $ESD < Ortc < AO$ gibi bir büyüklük sırasına dizildiği görülür (Lambert, 1993:30-31). Bu da gelir dağılımlarının hemen her ülkede sağa çarpık olduklarının bir göstergesidir.

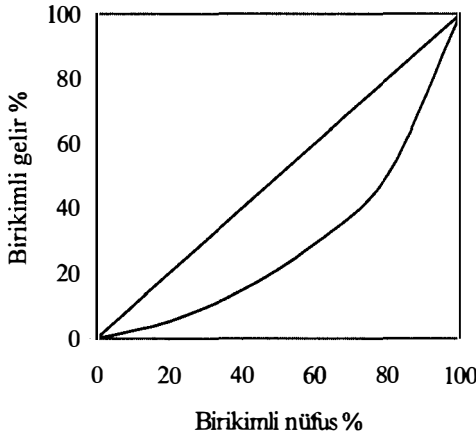
Gelir dağılımındaki çarpıklığın boyutlarını, gelir düzeyleri farklı kişi öbeklerine toplam gelirden ne kadar pay düştüğüne bakarak inceleyebiliriz. Gerek çeşitli ülkelerin gelir dağılımlarının, gerek aynı ülkenin değişik zamanlardaki gelir dağılımlarının karşılaştırılmasında kolaylık sağlamak amacıyla genellikle kişilerin (ya da hane-halklarının) beşebölenleri, onabölenleri, ..., yüzebölenleri kullanılır.

1984'te yapılan uluslararası bir karşılaştırmaya göre (Lambert, 1993: 31) 39 ülkedeki kişilerin en yüksek gelir elde eden % 10 öbeğinin gelirden aldıkları pay % 28 (Kuzey Kore) ile % 56,3 (Kenya) arasında değişmektedir. Dünya Bankası'nın 1995 sonrası verilerine göre söz konusu oranların en düşüğü Macaristan'da (% 20,5), en yükseği Brezilya'dadır (% 48,0) (World Bank, 2003:237-238). Türkiye'de 1994 gelir dağılımı araştırmasının sonuçlarına göre en yüksek geliri elde eden % 10'luk nüfusun gelir payı % 32,3 iken en düşük geliri alan % 10'luk nüfusun gelir payı yalnızca % 2,3 düzeyinde kalmaktadır (World Bank, 2003:238).

Yukarıda gördüğümüz çarpıklık ölçüleri gelir dağılımındaki eşitsizliği ölçmekte kullanılabilirler. Bunlar dışında, özellikle gelir dağılımı araştırmacılarının geliştirdikleri ölçüler de vardır. Burada en yaygın kullanılan **Lorenz eğrisi** ile **Gini yoğunlaşma katsayısı**'dan söz edilecektir. Bu ölçülerin hesaplanmalarında az önce sözü edilen çeşitli gelir düzeylerindeki toplum kesimlerine düşen gelir payları temel alınır. Ayrıca hanehalkının ortak tüketiminden kaynaklanan ölçek ekonomilerini dikkate alan *bireysel eşdeğer gelir* kavramı kullanılarak hesaplanan eşitsizlik ölçüleri de vardır (Gürsel, vö. 2000).

Lorenz Eğrisi

Amerikalı istatistikçi M. Lorenz'in adıyla Lorenz eğrisi olarak bilinen bir çizim biçimi bu gelir paylarını göstermeye yöneliktir. Bu çizimin yatay ekseninde nüfusun, dik ekseninde ise gelirin birikimli yüzde payları yer alır. Böyle bir Lorenz eğrisi Çizim 9.7'de gösterilmiştir.



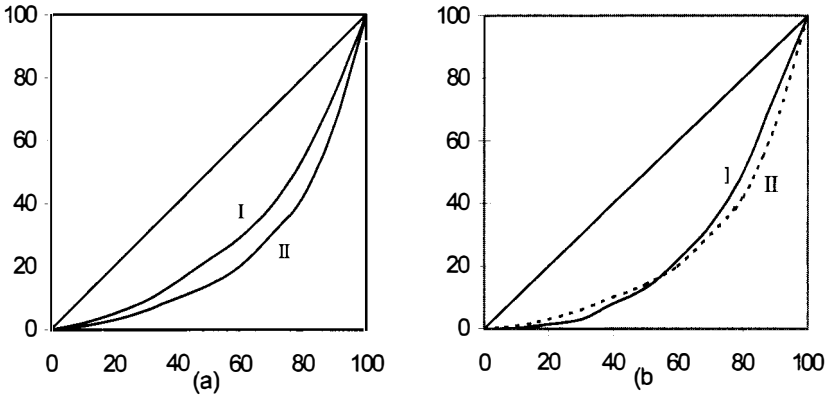
ÇİZİM 9.7 Bir Lorenz eğrisi örneği

Bu çizimde köşegen doğrusu gelir dağılımında tam eşitlik durumunu yansıtır. Sol alt köşeden başlayıp önce yatay, sonra dikey eksen boyunca ilerleyen ters L biçimindeki iki doğru parçasının konumu da tam bir eşitsizlik durumunu belirtir. Bu ikisi arasında yer alıp karşı iki köşeyi birleştiren eğri de **Lorenz eğrisi**dir. Lorenz eğrisi köşegene ne kadar yakınsa gelir dağılımı da eşitliğe o kadar yakın demektir. Gelir dağılımı eşitlikten uzaklaştıkça, eğri köşegenden uzak konumlara doğru kayar. Gelir dağılımında eşitsizlik varolduğu sürece, nüfusun en yüksek gelirli öbeğinin, diyelim % 10'unun gelirden aldığı pay % 10'dan büyük olacaktır. Benzer biçimde nüfusun en düşük gelirli % 10'unun gelirden aldığı pay da % 10'dan az olacaktır. Demek ki Lorenz eğrisi her zaman köşegenin altında çıkar.

Gini Yoğunlaşma Katsayısı

Aynı çizim üzerinde iki ayrı ülkenin gelir dağılımına ya da aynı ülkeye ilişkin iki ayrı tarihteki gelir dağılımına yer verilebilir. Böylece karşılaştırmanın gözle kolayca yapılması sağlanabilir. Eğer bir dağılımın Lorenz eğrisi, her noktasında ötekinin Lorenz eğrisinin üstündeyse ilk dağılım, ikinciye göre eşitliğe daha yakındır diyebiliriz.

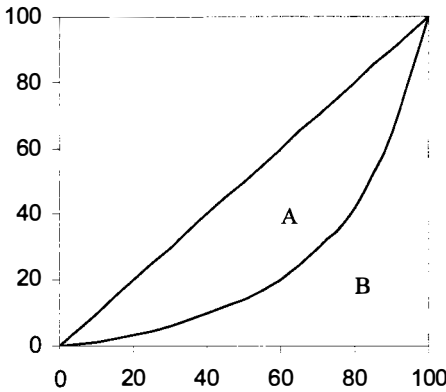
Çizim 9.8 (a) böyle bir durumu göstermektedir. Burada I dağılımı, II dağılımına göre eşitliğe daha yakındır. Ancak her zaman durum bu kadar açık olmayabilir. Bir Lorenz eğrisi, bazı bölümlerinde başka bir Lorenz eğrisinin üstünde yer alırken başka bazı bölümlerinde altına düşebilir.



ÇİZİM 9.8 İki Lorenz eğrisinin karşılaştırılması

Çizim 9.8 (b) böyle bir durumu yansıtmaktadır. Burada hangi dağılımının eşitliğe daha yakın olduğunu kolayca göremeyiz. İşte Gini yoğunlaşma katsayısı bu durumda yardımımıza koşar.

İtalyan istatistikçi C. Gini tarafından geliştirilen Gini katsayısı, Lorenz eğrisiyle köşegen arasında kalan alanın (Çizim 9.9'da karpuz dilimine benzeyen A alanının), köşegen altındaki dik üçgenin alanına (çizimde $A+B$ alanına) oranlanmasıyla bulunur. Bu oran 0 ile 1 arasındadır, eşitliğe yaklaştıkça sıfıra yakın, eşitlikten uzaklaştıkça bire yakın değerler alır.

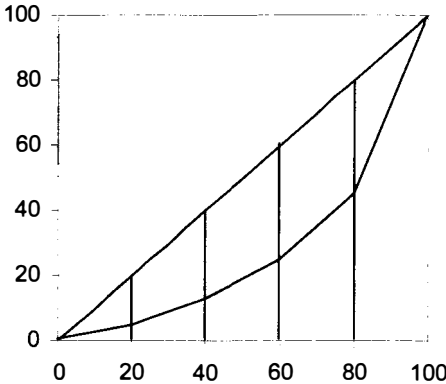


ÇİZİM 9.9 Gini katsayısı = $A / (A + B)$

Bütün dörtgenin alanına 1, dolayısıyla $A+B$ alanına $\frac{1}{2}$ birim diyecek olursak Gini katsayısını (GK) şu yollarla bulabiliriz:

$$GK = A / (A+B) = A / \frac{1}{2} = 2A = 2(\frac{1}{2} - B) = 1 - 2B$$

Eğer dağılım sürekli ise bu alan tümlev (integral) ile hesaplanabilir, kesikli ise üçgen, yamuk alanları toplanarak sonuca ulaşılabilir. Çizim 9.10'da süresiz bir dağılım için Gini katsayısının hesaplanmasında kullanılacak alanlar gösterilmiştir. Bu alanlardan yararlanarak GK değişik biçimlerde bulunabilir. Bir yol köşegen ile Lorenz eğrisi arasındaki üçgenlerle paralel kenarların alanlarını toplayıp ikiyle çarpmaktır. Bu yol yukarıdaki $2A$ hesaplanmasına karşılıktır. Bir başka yol Lorenz eğrisinin altında kalan alanların toplamının iki katını bulup 1'den çıkarmaktır. Bu da sürekli eğri çizimindeki $1 - 2B$ hesabına karşılık gelir. Daha uzun bir yol da köşegen ile Lorenz eğrisi arasındaki alanı, köşegenin altındaki alana bölmektir. Bu yol yukarıdaki $A / (A + B)$ hesabına karşılıktır. Bu yollardan en kolay uygulananı ikincisidir. Çünkü gelir dağılımı verileri her birikimli nüfus yüzdesinin eline geçen birikimli gelir yüzdesini gösterecek biçimde düzenlenmiştir. Dolayısıyla bu verilerle, Lorenz eğrisi altındaki alanın hesaplanması, köşegen ile eğri arasındaki alanın hesaplanmasına göre daha kısa sürer. Bu nedenle burada da bu kısa yol kullanılacaktır.



ÇİZİM 9.10 Kesikli dağılımda Gini katsayısı hesabında kullanılan üçgenlerle yamuklar

Kesikli dağılımlarda Gini katsayısı,

$$GK = \frac{\sum_i \sum_j |Y_i - Y_j| f_i f_j}{2N^2 \mu}$$

eşitliğiyle de bulunabilir (Kendall, Stuart, Ord, 1987:58-60). Burada

$|Y_i - Y_j|$ i 'ninciyle j 'ninci gelir düzeyleri arasındaki mutlak farkları,
 $f_i f_j$ bu gelir düzeylerindeki nüfus yüzdeleri,
 N gelir düzeylerinin toplam sayısını,
 μ gelirlerin aritmetik ortalamasını gösterir.

Bu eşitlikte μ dışındaki terimler her gelir düzeyi ile bütün öteki gelir düzeylerinin farklarının, işaretlerine bakılmaksızın alınmış bir aritmetik ortalaması olduğunu gösterir. *Ortalama fark* diyebileceğimiz bu ortalama, aritmetik ortalamaya bölünürse Gini katsayısını verir. Bir başka deyişle *GK*, ortalama fark ile *AO* arasındaki orandır.

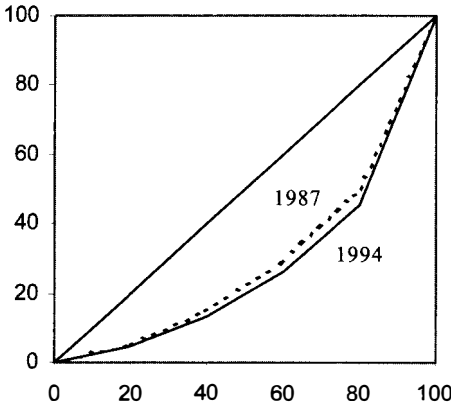
Bu anlatılanları bir örnek üzerinde görelim. Türkiye'de son iki gelir dağılımı araştırması 1987 ile 1994 yıllarında yapılmıştır. Çizelge 9.3'te gösterilen bu iki araştırmanın verileri kullanılarak çizilen Lorenz eğrileri de Çizim 9.11'de yer almışlardır. Buradan görüldüğü gibi 1994 yılında Türkiye'de gelir dağılımı, 1987 yılına göre eşitlikten biraz daha uzaklaşmıştır. Ancak bu iki dağılımın eşitlikten uzaklık derecelerini karşılaştırabilmek için Gini katsayılarını hesaplamamız gerekmektedir.

ÇİZELGE 9.3 Türkiye'de hanehalklarının %20'lik dilimlere göre yıllık gelirleri

Hanehalkı	Birikimli hanehalkı	1987		1994	
		Gelir %	Birikimli gelir %	Gelir %	Birikimli gelir %
İlk % 20	% 20	5,24	5,24	4,9	4,9
2. % 20	% 40	9,61	14,85	8,6	13,5
3. % 20	% 60	14,06	28,91	12,6	26,1
4. % 20	% 80	21,15	50,06	19,0	45,1
5. % 20	% 100	49,94	100,00	54,9	100,0
Toplam		100,00		100,0	

Kaynak: DİE (1990:32); DİE (1996:1)

Çizelge 9.3'teki birikimli hanehalkı yüzdeleri, Çizim 9.11'de yatay ekseninde yer almıştır. Birikimli gelir yüzdeleri de, bu yatay eksenin ilgili noktasından çıkılacak dikmelerin Lorenz eğrisine kadar olan bölümlerine karşılıktır. Dolayısıyla ardarda olan iki birikimli gelir yüzdesi değeri çizimdeki bir yamuğun alt – üst kenarlarını gösterir. Her % 20'lik nüfus diliminin karşılığı da yamuğun yüksekliğidir.



ÇİZİM 9.11 1987, 1994 yıllarında Türkiye gelir dağılımının Lorenz eğrileri

Bu durumda alanları kullanarak 1987 yılı için Gini katsayısını hesaplayalım. Lorenz eğrisi ile yatay ekseninde

$$\begin{aligned}
 &0 - 0,20 \text{ arasındaki üçgenin alanı: } 0,2 \times 0,0524 / 2 = 0,00524 \\
 &0,20 - 0,40 \text{ arasındaki yamuğun alanı: } 0,2 \times (0,0524 + 0,1485) / 2 = 0,02009 \\
 &0,40 - 0,60 \text{ arasındaki yamuğun alanı: } 0,2 \times (0,1485 + 0,2891) / 2 = 0,04376 \\
 &0,60 - 0,80 \text{ arasındaki yamuğun alanı: } 0,2 \times (0,2891 + 0,5006) / 2 = 0,07897 \\
 &0,80 - 1,00 \text{ arasındaki yamuğun alanı: } 0,2 \times (0,5006 + 1,0000) / 2 = \underline{0,15006} \\
 &\text{Toplam } B \text{ alanı} = 0,29812
 \end{aligned}$$

olur. Öyleyse Türkiye'nin 1987 yılı gelir dağılımı için Gini katsayısı

$$GK = 1 - 2B = 1 - 2 (0,29812) = 0,40376$$

bulunur.

Okuyucu benzer işlemleri 1994 yılı gelir dağılımı için yinelerse Gini katsayısını 0,4416 olarak bulacaktır. Buradan da gelir dağılımında eşitlikten uzaklaşmanın nicel boyutu ortaya çıkar.

ÇİZELGE 9.4 Türkiye'de hanelerin 1987 yılında %20'lik dilimlere göre yıllık gelirleri

Hanehalkı	Gelir %	Y	f
İlk %20	5,24	0,2620	0,2
2. %20	9,61	0,4805	0,2
3. %20	14,06	0,7030	0,2
4. %20	21,15	1,0575	0,2
5. %20	49,94	2,4970	0,2
Toplam	100.0		

Kaynak: DİE (1990:32)

Aynı sonuçları yukarıda verilen GK eşitliğiyle de bulabiliriz. Bunun için önce Çizelge 9.3'teki verileri bu işlemlere uygun duruma getirmek gerekir. Çizelge 9.4, 1987 yılı verileri için bu değişimleri göstermektedir. Y ile gösterilen sayılar gelir yüzdelерinin hanehalkı yüzdelерine bölünmesiyle bulunmuştur. f başlığı altındaki sayılar ise % 20'lik nüfus dilimlerinden başka birşey değildir. Y değerleri her % 20'lik dilimdeki hanelerin ortalama görelі gelir düzeylerini, f değerleri bu gelir düzeyindekiilerin görelі sıklıklarını gösterir. Dolayısıyla aynı satırdaki Y ile f çarpımı o dilimdeki hanelerin toplam gelirden aldıkları payı verir. $\sum Yf$ toplamını da gelirlerin toplamı, yani % 100 = 1 olarak algılayabiliriz. İşte bu Y değerleriyle f değerleri aşağıdaki GK eşitliğinde kullanılan simgelere karşılık gelirler. Burada $\mu = 1$, $N = 1$, $f_i f_j = 0,2 \times 0,2 = 0,04$ olduğunu görmek kolaydır.

Böylece şunu yazabiliriz:

$$GK = \frac{\sum_i \sum_j |Y_i - Y_j| f_i f_j}{2N^2 \mu} = \frac{0,04 \sum_i \sum_j |Y_i - Y_j|}{2}$$

Şimdi $|Y_i - Y_j|$ terimlerine bir göz atalım. Örnek olarak $i = 1$, $j = 2$ iken $|Y_1 - Y_2|$ biçiminde yazılan bu terim, $i = 2$, $j = 1$ iken $|Y_2 - Y_1|$ olarak gösterilir. Mutlak değerler söz konusu olduğuna göre bu ikisi birbirine eşittir. Öyleyse aynı değer toplama iki kez girer. Hesaplamalarımızı kısaltmak için bunları toplama bir kez alırsak paydadaki 2 sayısından da kurtulmuş oluruz. Bu durumda Gini katsayısı şöyle bulunur:

$$GK = 0,04 \sum_i \sum_{j \geq i} |Y_i - Y_j|$$

$i \geq j$ iken aşağıdaki değerler toplanır

$$\begin{array}{ll}
|0,4805 - 0,2620| = 0,2185 & |0,7030 - 0,4805| = 0,2225 \\
|1,0575 - 0,7030| = 0,3545 & |0,7030 - 0,2620| = 0,4410 \\
|1,0575 - 0,4805| = 0,5770 & |2,4970 - 0,7030| = 1,7940 \\
|1,0575 - 0,2620| = 0,7955 & |2,4970 - 0,4805| = 2,0165 \\
|2,4970 - 0,2620| = 2,2350 & |2,4970 - 1,0575| = 1,4395
\end{array}$$

$$\sum_i \sum_j |Y_i - Y_j| = 10,0940 \text{ bulunur.}$$

Demek ki $GK = 0,04 \times 10,0940 = 0,40376$ 'ya eşittir. Bu da daha önce alanlar toplamı yoluyla bulduğumuz değerin tıpkısıdır.

Toplanma oranı

Gini katsayısına benzeyen ama onun daha kolay hesaplanabilen bir biçimi toplanma oranı denilen bir ölçüdür (Gürtan, 1987: 134-135). Bu ölçü de Lorenz çiziminde gösterilebilir. Çizim 9.10'daki dikey çizgilerin, Lorenz eğrisi ile köşegen arasındaki bölümlerinin uzunluklarının toplamı alınıp, aynı çizgilerin yatay eksenle köşegen arasındaki uzunlukları toplamına bölünmesiyle toplanma oranı elde edilir. Bu bölme işleminin paydasındaki sayı, birikimli nüfus yüzdelерinin % 100 dışında kalanlarının toplamıdır. Pay ise birikimli nüfus % değerleri ile birikimli gelir % değerleri arasındaki farkların toplamıdır. 1987 gelir dağılımı verilerini kullanarak toplanma oranını Çizelge 9.5'ten hesaplayabiliriz.

Bu hesaplama yöntemi Gini Katsayısı hesaplanmasına çok benzer. Orada Lorenz eğrisi ile köşegen arasındaki alan, köşegenin altında kalan üçgenin alanına bölünüyordu. Burada da Lorenz eğrisi ile köşegen arasındaki doğru parçalarının toplam uzunluğu, köşegenin altında kalan doğru parçalarının uzunluğuna bölünüyor.

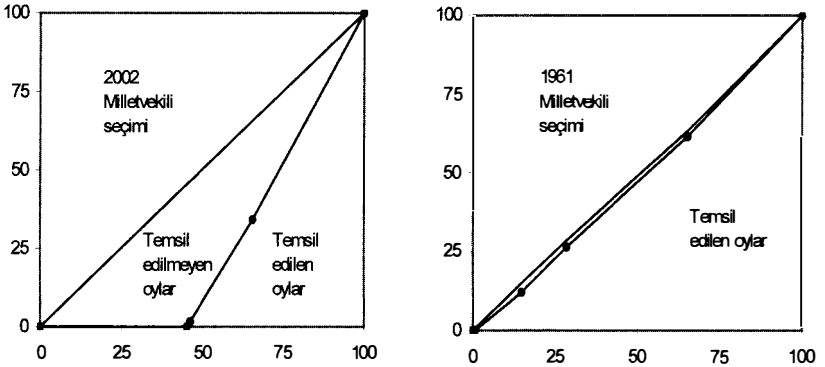
ÇİZELGE 9.5 1987 Gelir dağılımı için Toplanma Oranı hesabı

Birikimli Hanehalkı %	Birikimli gelir %	Farklar
20	5,24	14,76
40	14,85	25,15
60	28,91	31,09
80	50,06	29,94
200	99,06	100,04

$$\text{Toplanma Oranı} = 100,94 / 200 = 0,5047$$

Burada da kestirme işlem yapılabilir. Tek tek her doğru parçası üzerindeki birikimli hane halkı % ile birikimli gelir % farklarını almak yerine, bu ikisinin toplamı arasındaki farkı bir kerede almak yoluna da gidilebilir. Örneğimizde, Çizelge 9.5'te *Farklar* başlığı altındaki sayıları hiç hesaplamadan bu farkların toplamını $200 - 99,06 = 100,94$ biçiminde de bulabilirdik. Bu daha kestirme bir yol olurdu. Ya da bu farkı doğrudan hesaplama- dan, Toplanma Oranı = $1 - 99,06 / 200 = 0,5047$ olarak da elde edilebilirdi. Hangi kestirme yol kullanılırsa kullanılsın sonuç elbette aynı olurdu. Okuyucu 1994 yılı için bu ölçüyü kendi hesaplayıp 1987 ile karşılaştırmalıdır.

Gelir dağılımındaki çarpıklığı ölçmek üzere geliştirilmiş Lorenz eğrisi ya da Gini katsayısı araçları daha sonra eğitimdeki eşitsizliği ölçmek amacıyla da kullanılmıştır (Thomas, vö., 2000). Biz de buradan esinlenerek aynı araçları seçimlerin temsil niteliğini ölçmekte kullanabiliriz. Bunun için yatay eksene partilerin seçimde aldıkları oy oranlarını, dikey eksene de seçimde çıkardıkları milletvekili sayılarını birikimli olarak yerleştirelim. Her partinin konumunu bu çizimde işaretleyerek Lorenz eğrisini çizebilir, yukarıda gördüğümüz yöntemlerle Gini katsayısını hesaplayabiliriz. Çizim 9.12 temsil nitelikleri bakımından iki seçimin sonuçlarını karşılaştırmaktadır. Bu çizimde kullanılan veriler Alıştırma 9.47'de verilmiştir.



ÇİZİM 9.12 Seçimlerin temsil niteliklerini yansıtan Lorenz eğrileri

Okuma Önerileri

- Emerson, J. D., Strenio, J. (1983) “Boxplots and batch comparison”, Hoaglin, Mosteller, Tukey (1983) içinde 58-93.
- Erickson, B. H., Nosanchuk, T. A. (1992) *Understanding Data*, 2nd. ed. Open University Press (46-56).
- Groebner, D. F., Shannon, P. W., Fry, P. C., Smith, K. D. (2001) *Business Statistics: A Decision Making Approach*, 5. b., Prentice Hall (97-111).
- Gürsel, S., Levent, H., Selim, R., Sarıca, Ö. (2000) *Türkiye’de Bireysel Gelir Dağılımı ve Yoksulluk*, TÜSİAD.
- Gürtan, K. (1987) *İstatistik ve Araştırma Metodları*, 4. b., İstanbul Üniversitesi (297-345).
- Hoaglin, D. C. (1983) “Letter values: a set of selected order statistics”, Hoaglin, Mosteller, Tukey (1983) içinde 33-55.
- Huff, D. (1973) *How to Lie with Statistics*, Penguin (29-36) (Çeviren: Koparan, E. (1995) *İstatistik ile Nasıl Yalan Söylenir?*, Sarmal(31-38)).
- Marsh, C. (1988) *Exploring Data*, Polity Press (24-40).
- Newbold, P., Carlson, W. L., Thorne, B. (2003) *Statistics for Business and Economics*, 5th ed, Prentice Hall (38-46) (4. basımdan çeviren: Şenesen, Ü. (2000) *İşletme ve İktisat İçin İstatistik*, Literatür (16-30)).
- Ott, R. L., Longnecker, M. (2001) *An Introduction to Statistical Methods and Data Analysis*, 5th ed. Duxbury (81-93).
- Siegel, A. F. (1988) *Statistics and Data Analysis*, Wiley (77-124).
- Siegel, A. F. (1997) *Practical Business Statistics*, Irwin (105-140).

Alıştırmalar

- 9.1** Pay senetleri A ile B’nin son bir ay içindeki fiyatlarının çeşitli ortalamalarıyla standart sapmaları aşağıda verilmiştir.

(m. TL)	Pay senedi A	Pay senedi B
AO	10	100
ESD	9,1	92
Ortc	9,7	97,5
σ	1	5

- Hangi pay senedi daha risklidir? DK değerlerini düşündünüz mü?
- Çarpıklık ölçülerini hesaplayınız. Bulgularınızı yorumlayınız.
- Satın almak için hangi pay senedi daha uygundur? Neden?

9.2 Bir sınıftaki kız öğrencilerin boylarıyla ağırlıkları şöyle olsun:

Boy (m)	Ağırlık (kg)
$AO = 1,60 \quad \sigma = 0,08$	$AO = 48 \quad \sigma = 1,72$

Bu öğrencilerin boyları mı, yoksa kiloları mı daha değişkendir? Neden? Hesaplamanızı gösterip gerekçenizi açıklayınız.

9.3 İki ayın ilk haftalarındaki günlük en yüksek hava sıcaklıkları ($^{\circ}\text{C}$) şöyledir. Sizce hangi haftada sıcaklık daha değişkendir? Neden? Hesaplamanızı gösterip gerekçenizi açıklayınız.

Şubat	-1	0	2	2	2	4	2
Mart	1	0	3	2	1	7	7

9.4 Gözlem sayısı 110 olan bir veri kümesinden bulunmuş şu toplamaları kullanarak bu dağılımın ortalamasıyla biçimi (yayıklığı, çarpıklığı, basıklığı) konusunda neler söyleyebilirsiniz?

$$\Sigma Y = 330 \quad \Sigma Y^2 = 1.650 \quad \Sigma Y^3 = 9.900 \quad \Sigma Y^4 = 60.830$$

9.5 Bir televizyon tüpü üreticisi rassal olarak seçilmiş 400 tüpün saat cinsinden ömrünü ölçmüş, şu dağılımı bulmuştur. Dağılımın ortalama düzey, biçim (yayıklık, çarpıklık, basıklık, tepe sayısı) özelliklerini inceleyerek yorumlayınız.

Tüp ömrü (saat)	Tüp sayısı	Tüp ömrü (saat)	Tüp sayısı
300 - < 400	14	800 - < 900	62
400 - < 500	47	900 - < 1000	49
500 - < 600	58	1000 - < 1100	22
600 - < 700	75	1100 - < 1200	2
700 - < 800	68		

9.6 Gözlem sayısı N olan bir veri kümesinde $\sum (Y_i - \bar{Y})^3 / N = 0,08$, $\sum (Y_i - \bar{Y})^4 / N = 47,6$, $\sigma^2 = 4$ bulunmuştur. Bu dağılımın normal dağılıma uygunluğu konusunda neler söyleyebilirsiniz?

9.7 Bir halının rassal olarak seçilmiş 10 ayrı yerinde yapılan sayımlarda 1 cm^2 'ye düşen düğüm sayıları aşağıdadır. Bu dağılımın biçimini inceleyiniz.

33 32 31 36 34 35 30 34 33 32

9.8 Aynı sınava giren iki sınıfın notlarına ilişkin özet bilgiler aşağıdadır.

- a) Bu sınıflardan birinde 90'dan yüksek not alan kimse yokken öbüründe üç kişi vardır. Bu üç kişi hangi sınıftadır? Neden?
- b) A sınıfında bir öğrenciye yanlışlıkla düşük verilmiş olan not 30'dan 50'ye çıkarılınca bu sınıfta AO , ESD , $Ortc$, σ değişir mi? Neden? Değişirse kaç olur?

Sınıf	AO	ESD	$Ortc$	σ
A	45	40	43	10
B	45	50	47	13

9.9 Bir işyerindeki 400 işçinin aylık ücretleri için $AO = 630$, $\sigma = 15$ milyon TL bulunmuştur. İşçilere yılbaşında aşağıdaki çizelgede gösterildiği gibi zam yapılacaktır.

- a) Ücretlerin normal dağılıma uyduğu varsayılırsa bu zammın işverene getireceği ek yük ne kadardır?
- b) Normal dağılım varsayımı yapılamazsa ek yük konusunda neler söyleyebilirsiniz?

Ücret düzeyi	Zam tutarı	Ücret düzeyi	Zam tutarı
< 600 m. TL	40 m. TL	645 – < 660 m. TL	55 m. TL
600 – < 615	45	660 +	60
615 – < 645	50		

9.10 Bir işyerinde haftalık ücretlere zam yapılacaktır. Toplu sözleşme görüşmelerinde A ile B gibi iki ayrı zam önerisi ortaya çıkmıştır. İki öneriye göre hesaplanan özet göstergeler aşağıdadır.

- a) İşveren, b) yüksek ücretli çalışan, c) düşük ücretli çalışan
d) sendika temsilcisi hangi öneriyi destekler? Neden?

(m. TL)	AO	ESD	$Ortc$	σ
Zam öncesi	170	165	167	14
A önerisi	177	172	174	14
B önerisi	177	171	173	15,4

9.11 Bir alışveriş yerinde yapılan bir araştırmaya göre müşterilerin para ödeme kuyruğunda bekleme süreleri şöyle dağılmaktadır. Bu dağılımın ortalama düzeyini, biçim özelliklerini (yayıklığını, çarpıklığını, basıklığını) inceleyiniz.

Süre (dak.)	Müşterilerin yüzdesi	Süre (dak.)	Müşterilerin yüzdesi
0	5	2 – < 3	11
0 – < 0,5	20	3 – < 5	7
0,5 – < 1	32	5 – < 10	5
1 – < 2	19	10 – < 15	1

9.12 İki bölgede yapılan kamuoyu yoklamalarında deneklere bir önceki hafta kaç saat TV seyrettikleri sorulmuş, alınan yanıtlar aşağıda özetlenmiştir. İki bölgenin TV seyretme alışkanlıklarını karşılaştırınız.

A bölgesi		B bölgesi	
Saat	Kişi	Saat	Kişi
0 – < 4	5	0 – < 5	8
4 – < 4	25	5 – < 10	12
8 – < 4	30	10 – < 15	25
12 – < 4	11	15 – < 20	9
16 – < 4	4	20 – < 25	4
20 – < 4	2	25 +	2
24 +	3		
	80		60

9.13 Bir su dolum işletmesinde şişelere konan su miktarı $AO = 400$ gr, $\sigma = 5$ gr ile normal dağılıma uymaktadır.

- Doldurulan bütün şişeler içinde
 - 380 gr'dan az,
 - 390 gr'dan az,
 - 390 – 400 gr
 - 410 gr'dan az,
 - 420 gr'dan çok
 su içeren şişelerin oranı yaklaşık ne kadardır?
- Normal dağılım varsayımı yapılamazsa yukarıdaki sorular yanıtlanabilir mi? Bu durumda söz konusu oranların hiç olmazsa alt sınırlarını bulabilir misiniz?

9.14 Her anakaranın en yüksek tepesi aşağıda gösterilmiştir. Bu veri kümesinin ortalama düzeyi, yayıklığı, çarpıklığı, basıklığı konusunda ne söyleyebilirsiniz? Bulduğunuz sayıların ne anlama geldiğini açıkça belirtiniz.

Anakara	Tepe	Yükseklik (m)
Afrika	Kibo	5895
Antarktika	Vinson Massif	5139
Asya	Everest	8848
Avrupa	El'brus	5642
Avustralya	Kosciusko	2228
Güney Amerika	Aconcagua	6960
Kuzey Amerika	McKinley	6194

- 9.15** Beşer gözlemleri öyle iki veri kümesi bulunuz ki
a) AO'ları aynı, standart sapmaları farklı,
b) AO'ları farklı, standart sapmaları aynı olsun.
- 9.16** Bir normal dağılımda bütün gözlemlerin yaklaşık yarısı $\mu \pm 2/3\sigma$ aralığında bulunur. Bu bilginin dördebölenler aralığı (DAr) ile ilgisi nedir?
- 9.17** Aşağıdaki veriler 1 Mayıs 2003 gazetelerinden derlenmiş olup değişken yatırım fonlarından rassal olarak seçilmiş A tipinden 20, B tipinden 20 tane fonun yılbaşından o güne kadar sağladıkları % getirileri göstermektedir.

A tipi değişken fon getirisi (%)					B tipi değişken fon getirisi (%)				
22,5	15,2	8,8	9,1	3,7	15,2	15,6	13,7	12,0	11,2
2,5	9,9	9,0	9,8	12,9	14,0	14,2	6,5	10,0	13,0
10,5	6,4	9,3	-8,4	6,8	13,7	14,5	13,0	11,6	14,5
12,3	19,4	5,0	14,8	11,2	11,3	13,4	7,3	14,8	14,0

Aşağıdaki ölçüleri kullanarak iki tip fonun getirilerini ortalama düzey, yayılım bakımlarından karşılaştırınız, hangi tip fonun ortalama getirisinin daha yüksek olduğunu, hangi tip fonun daha riskli sayılabileceğini açıklayınız.

- a) AO b) s c) DK d) $Ortc$ e) DA f) $DAr/Ortc$

- 9.18** Çeşitli tarihlerde yapılan araştırmalara göre Türkiye'de gelir dağılımını gösteren özet bilgiler aşağıdadır.
- a) Her araştırma için ayrı bir Lorenz eğrisi çiziniz.
- b) Her araştırmanın Gini katsayılarını hesaplayınız.
- c) Araştırmacıların ayrıntılı verilerle buldukları Gini katsayıları çizelgede gösterilmiştir. Bunlarla sizin sonuçlarınız arasında ne kadar fark var? Bu farkın nedeni ne olabilir?
- d) Gelir dağılımının zaman içine nasıl değiştiğini irdeleyiniz. Bunu yaparken 1987 bulgularına fazla güvenilmemesi gerektiğini, DPT ile DİE dışındaki kurumların yöntemlerinin farklı olduğunu unutmayınız.

- e) Her araştırmadaki gelir paylarının anlamlı basamak sayıları sizce neden farklıdır?

Hanehalkı dilimlerinin gelirden aldıkları paylar %								
Hanehalkı dilimleri	I '63	II '68	III '73	IV '78	V '83	VI '86	VII '87	VIII '94
İlk % 20	4,5	3	3,5	2,9	2,7	3,9	5,24	4,9
2. % 20	8,5	7	8,0	7,4	7,0	8,4	9,61	8,6
3. % 20	11,5	10	12,5	13,0	12,6	12,6	14,06	12,6
4. % 20	18,5	20	19,5	22,1	21,9	19,2	21,15	19,0
5. % 20	57,0	60	56,5	54,7	55,8	55,9	49,94	54,9
Gini katsayısı	0,55	0,56	0,51	0,51	0,52	0,46	0,43	0,49

I: DPT, 1966, II: Bulutay, vö.1971, III: DPT, 1976, IV-V: Celasun, 1986, 1989,

VI: Esmer, vö. 1986, VII-VII: DİE, 1990,1997a.

- 9.19** Birer litrelik kutulara meyva suyu koyacak biçimde ayarlanmış bir maki-neyle doldurulan kutulardan 100 tanesi rassal olarak seçilip içlerindeki meyva suyu miktarları (Y_i) ölçülmüş, şu sonuçlara ulaşılmıştır.

$$\Sigma Y_i = 102,00$$

$$\Sigma Y_i^2 = 105,03$$

- Kutulardaki meyva suyu miktarlarının AO, s, DK değerlerini bulunuz.
 - Bu miktarların normal dağılıma uyduğu varsayımıyla doldurulan bütün kutulardan içinde 1 litreden az meyva suyu bulunanların oranını bulunuz.
 - Normal dağılım varsayımı yapılamazsa bu oran için ne söyleyebilirsiniz?
- 9.20** Seçilmiş bazı ülkelerdeki gelir dağılımı verileri aşağıdadır.
- Her ülke için ayrı birer çizim düzenleyerek iki yıla ilişkin Lorenz eğrile-rini gösteriniz. Gelir dağılımının zaman içinde ne yöne doğru değiştiğini gözleyiniz. Her ülke için değişimin ne yönde olduğunu söyleyebiliyor musunuz?
 - Her gelir dağılımı için ayrı birer Gini katsayısı hesaplayınız. Her ülkenin kendi gelir dağılımının zaman içinde nasıl değiştiğini bu katsayılarla göre yorumlayınız. Bulgularınız Lorenz eğrileri ile yaptığınız yorumlarla uyumlu mu?
 - On yıla yakın sürede ortadirek diye adlandırılan kesimin konumu bu dört ülkede nasıl değişmiştir?

Hanehalkı Dilimleri	Hanehalkı dilimlerinin gelirden aldıkları paylar %							
	Brezilya		Hindistan		Spain		ABD	
	1983	1995	1985	1994	1981	1990	1985	1994
İlk % 20	2,4	2,5	8,1	9,2	6,9	7,5	4,7	4,8
2. % 20	5,7	5,7	12,3	13,0	12,5	12,6	11,0	10,5
3. % 20	10,7	9,9	16,3	16,8	17,3	17,0	17,4	16,0
4. % 20	18,0	17,7	22,0	21,7	23,2	22,6	25,0	25,5
5. % 20	62,6	64,2	41,4	39,3	40,0	40,3	41,9	45,2

World Bank, 1990:236-237; 1999:198-199.

9.21 DİE'nin 1987, 1994 araştırmalarına göre Türkiye'de kır-kent kesimlerindeki gelir dağılımı verileri aşağıdadır.

Hanehalkı dilimleri	1987		1994	
	Kır	Kent	Kır	Kent
İlk % 20	5,21	5,43	5,57	4,83
2. % 20	10,03	9,33	10,14	8,19
3. % 20	14,98	13,60	14,80	11,87
4. % 20	21,97	20,71	21,79	17,90
5. % 20	47,82	50,93	41,70	57,22

DİE, 1990:525, 597; 1997:6, 8.

- 1987 yılı Lorenz eğrilerini aynı çizimde göstererek yorumlayınız.
- 1994 yılı Lorenz eğrilerini aynı çizimde göstererek yorumlayınız.
- Kır kesiminin 1987, 1994 Lorenz eğrilerini aynı çizimde göstererek yorumlayınız.
- Kent kesiminin 1987, 1994 Lorenz eğrilerini aynı çizimde göstererek yorumlayınız.
- Dört ayrı Gini katsayısı hesaplayarak iki kesimin iki ayrı yıldaki gelir dağılımlarını karşılaştırınız.
- Dört ayrı toplanma oranı hesaplayarak iki kesimin iki ayrı yıldaki gelir dağılımlarını karşılaştırınız.

9.22 DİE'nin 1987, 1994 araştırmalarına göre Türkiye'de bölgelere göre gelir dağılımları aşağıdaki gibidir.

Hanehalkı dilipleri '87	Marmara, İç Anadolu			Doğu, Güneydoğu	
	Ege	Akdeniz	Karadeniz		
İlk % 20	5,47	5,45	5,28	5,83	5,33
2. % 20	9,78	9,88	9,66	11,58	9,25
3. % 20	14,21	14,52	14,07	16,14	19,17
4. % 20	20,96	21,06	21,29	23,15	20,11
5. % 20	49,58	49,09	49,70	43,79	52,14

DİE, 1990:345.

Hanehalkı dilipleri '94	İç			Doğu		G.doğu
	Marmara	Ege	A.	Akdeniz	Karadeniz	
İlk % 20	4,33	5,39	5,07	5,25	5,13	6,30
2. % 20	7,47	9,69	9,11	9,05	9,47	11,29
3. % 20	10,69	14,09	13,77	13,23	13,77	16,20
4. % 20	16,35	20,89	21,89	19,59	20,16	23,12
5. % 20	61,16	49,93	50,16	52,86	51,49	43,08

DİE, 1997a:12, 14, 16, 18, 20, 22, 24.

- Her bölgenin 1987, 1994 Lorenz eğrisini çizerek hem bölgeler arası, hem dönemler arası gelir dağılımlarını karşılaştırınız.
- Her bölgenin 1987, 1994 Gini katsayısını hesaplayıp hem bölgeler arası, hem dönemler arası gelir dağılımlarını karşılaştırınız.
- Her bölgenin 1987, 1994 toplanma oranını hesaplayıp hem bölgeler arası, hem dönemler arası gelir dağılımlarını karşılaştırınız.

9.23 Bir bölgedeki tarım işletmelerine ilişkin veriler aşağıdaki gibi derlenmiş olsun. Lorenz eğrisi, toplanma oranı, Gini katsayısı gibi araçlarla toprak dağılımı eşitsizliğini yorumlayınız.

İşletme büyüklüğü (ha)	İşletme sayısı (tane)	Ekili alan (ha)	İşletme sayısı (%)	Ekili alan (%)
< 1	20	10	10,0	0,1
1 – < 5	35	100	17,5	1,0
5 – < 10	42	310	21,0	3,1
10 – < 25	31	540	15,5	5,4
25 – < 50	25	940	12,5	9,4
50 – < 100	20	1500	10,0	15,0
100 – < 250	15	1600	7,5	26,0
250 +	12	4000	6,0	40,0
	200	10000	100,0	100,0

9.24 1989 sonunda bankalardaki tasarruf mevduatı verilerinin hesap büyüklüğüne göre dağılımı aşağıdadır.

Hesap büyüklüğü (milyon TL)	Hesap sayısı (bin tane)	Mevduat tutarı (milyar TL)
< 3	27869	7178
3 – < 6	827	3716
6 – < 25	522	5971
25 – < 50	117	2558
50 – < 100	24	1566
100 +	11	1641
Toplam	29370	22830

- Bu dağılımın Lorenz eğrisini çiziniz.
- Gini katsayısını hesaplayınız.
- Toplanma oranını hesaplayınız.
- Dağılımın eşitsizlik derecesini yorumlayınız.
- Hesap büyüklüklerinin histogramını çiziniz. Çeşitli çarpıklık ölçüleri hesaplayıp çizimde görülen çarpıklığı yorumlayınız.
- Ortalama hesap büyüklüğünü en iyi hangi ortalama(lar)la gösterebilirsiniz? Bu ortalamayı ya da ortalamaları bulunuz.
- Bu konudaki daha yeni tarihli verileri bularak aynı soruları yanıtlayınız.
- İki veri kümesinin derlendiği tarihler arasında mevduat dağılımındaki eşitsizliğin ne yönde geliştiğini inceleyiniz.

- 9.25** Çizelge 3.4'e dönelim. Kristal-İş Sendikası'na üye işçilerin yaşadıkları haneyi paylaştıkları kişi sayısı dağılımının ortalama düzeyiyle biçim özelliklerini inceleyiniz.
- 9.26** Rassal örnekleme-rassal olmayan örnekleme başlıklı 6.2 altbölümünde altı gözlemlili bir anakütleden çekilen dörder gözlemlili 15 örnekleme bunların AO 'ları aşağıdaki gibiydi.

2	2	2	2	2	2	2	2	2	2	6	6	6	6	8
6	6	6	6	6	6	8	8	8	10	8	8	8	10	10
8	8	8	10	10	10	10	10	10	10	10	10	10	10	10
10	10	12	10	12	12	10	12	12	12	10	12	12	12	12
6,5	6,5	7	7	7,5	7,5	7,5	8	8	8,5	8,5	9	9	9,5	10

Bu AO 'ların AO 'su da anakütle ortalamasının yansız bir kestirimcisiydi. Ayrıca 6,5 – 10 arasındaki 3,5 birimlik Ar değerinin başka ortalamalarla bulunabilecek Ar 'lardan daha küçük olacağı ileri sürülmüş, dolayısıyla AO 'nun burada yansız kestirimciler arasında en etkin kestirimci olduğu da belirtilmişti.

- Bu 15 örneklemin GO değerlerini hesaplayıp bunların Ar değerinin 3,5 birimden büyük olduğunu gösteriniz.
 - Aynı yöntemle 15 örneklemin TO , KO değerlerini de hesaplayıp bunların her birinin Ar değerinin 3,5 birimden büyük olduğunu gösteriniz.
 - 15 AO için σ değerini bulunuz. Bunu 15 GO , 15 TO , 15 KO için hesaplayacağınız σ değerleriyle karşılaştırınız. Etkinlik Ar yerine σ ile ölçülürse en etkin ortalama yine AO mudur?
- 9.27** Yine 6.2 altbölümünde katmanlı rassal örnekleme başlığı altında yapılan açıklamalara göre, anakütle iki katmana ayrılıp her katmandan alınan ikişerli gözlemlili örneklemlemlerle bunların AO 'ları aşağıdaki gibi bulunmuştu.

2	2	2	2	2	2	6	6	6
6	6	6	10	10	10	10	10	10
8	8	10	8	8	10	8	8	10
10	12	12	10	12	12	10	12	12
6,5	7	7,5	7,5	8	8,5	8,5	9	9,5

Bunların $9,5 - 6,5 = 3$ olan Ar değerinin daha önce bulunan 3,5'ten küçük çıkmasına bakılarak etkinliğin basit rassal örnekleme göre arttığı ileri sürülmüştü. Aynı AO 'ların yayılımını Ar yerine σ kullanarak ölçünüz. Hesapladığınız değerin, basit rassal örnekleme yöntemine dayanan Alıştırma

9.26 c'de 15 AO için bulduğunuz σ değerinden küçük olduğunu göstererek aynı savı doğrulayınız.

- 9.28** Alıştırma 6.5'e dönelim. Gerek 3'lü, gerek 5'li örneklemelerde AO'ların yayılımı *Ar* değerleriyle bulunmuştu. Şimdi de aynı AO'ların σ değerlerini hesaplayınız. Etkinlikleri yorumlayınız.
- 9.29** Alıştırma 6.6'ya dönelim. Orada katman alma yolumuzu değiştirip *Ar* hesaplayarak etkinliği artırıp artıramayacağımız sorulmuştu. *Ar* yerine σ geçirerek aynı soruyu yeniden yanıtlayınız.
- 9.30** Çizelge 7.5'e dönerek araba satışlarının biçim (yayıklık, çarpıklık, basıklık) özelliklerini çeşitli ölçütleri hesaplayarak yorumlayınız.
- 9.31** Çizelge 7.8'deki dışalıma bağımlılık katsayılarının biçim (yayıklık, çarpıklık, basıklık) özelliklerini her dönem için ayrı ayrı hesaplayarak yorumlayınız.
- 9.32** Alıştırma 9.30'u bir kez de Çizelge 7.9'daki verileri kullanarak yanıtlayınız. İki çizelgeden elde edilen bulgular arasında herhangi bir fark var mı? Açıklayınız.
- 9.33** Alıştırma 7.3'teki Çizelge B'nin verilerini kullanarak şu soruları yanıtlayınız.
- Dört bölgedeki satışların mevsimlik dağılımlarının biçim özelliklerini bulup karşılaştırınız. Satışları mevsimlere göre en oynak bölge hangisidir?
 - Dört mevsimlik satışların bölgelere göre dağılımlarının biçim özelliklerini bulup karşılaştırınız. Satışları bölgelere göre en değişken mevsim hangisidir?
- 9.34** Alıştırma 7.5'e dönelim.
- 23 öğrencinin not dağılımının ortalama düzeyiyle biçim (yayıklık, çarpıklık, basıklık) özelliklerini inceleyiniz. Yorumunuzu yazınız.
 - Her öğrencinin notuna 10 eklendiğini düşünelim. Yukarıda bulduğunuz değerler bundan nasıl etkilenir? Hesaplamaları yeniden yapmadan açıklayınız.
 - Bu kez her öğrencinin notunun % 10 artırıldığını düşünelim. (a)'da bulduğunuz değerler bu işlemten nasıl etkilenir? Hesaplamaları yeniden yapmadan açıklayınız.
- 9.35** Alıştırma 7.6'daki buğday üretimi çizelgesine bakınız. Buradaki verimlilik değerlerinin ortalama düzeyiyle biçim (yayıklık, çarpıklık, basıklık) özelliklerini her yıl için ayrı ayrı hesaplayarak karşılaştırınız. Bu ülkelerdeki buğday üretim verimliliği zaman içinde ne yönde değişmektedir?
- 9.36** Alıştırma 7.12'ye dönelim. 22 Avrupa ülkesinde ABD'nin Irak'a saldırmasından önce savaşa karşı olma orannının ortalama düzeyini Alıştırma 8.49'da hesaplamıştınız.

- a) Ülkelerin bu ortalamadan uzaklaşmalarını ölçmek için hangi ölçütü kullanırsınız? Neden? Bu ölçütü hesaplayınız.
- b) Bu dağılımın çarpıklığı, basıklığı konusunda neler söyleyebilirsiniz?
- 9.37** Alıştırma 7.40'a dönelim.
- a) 18 ülkenin savunma harcamalarıyla sağlık harcamalarının oynaklıklarını karşılaştırınız.
- b) Bu iki harcama türünün çarpıklıkları aynı yöne mi, yoksa farklı yönlere doğru mudur? Hangi harcama daha çarpıktır?
- c) Hangi harcama türü daha basıktır? Açıklayınız.
- 9.38** Alıştırma 7.41'e dönelim. 12 ülkenin evlenme-boşanma oranlarını biçim özellikleri bakımından karşılaştırınız.
- 9.39** Alıştırma 8.25'teki iş sürelerinin yayıklık, çarpıklık, basıklığını ölçünüz. Bulduğunuz çarpıklık, orada kullandığınız ortalama türü ile ilgili kararınıza uyuyor mu? Açıklayınız.
- 9.40** Alıştırma 8.26'daki hastaların ateş ölçümlerinin biçim özelliklerini inceleyiniz.
- 9.41** Alıştırma 8.27'deki iki ayrı ozon ölçümünün biçim özelliklerini karşılaştırınız.
- 9.42** Alıştırma 8.28'deki bekleme sürelerinin yayıklık, çarpıklık, basıklığını ölçünüz. Bulduğunuz çarpıklık, orada kullandığınız ortalama türü ile ilgili kararınıza uygun mu? Açıklayınız.
- 9.43** Alıştırma 8.31'deki tasarruf verilerinin biçim özelliklerini bulup irdeleyiniz.
- 9.44** Çizelge 9.3'teki 1994 verilerini kullanarak Gini katsayısını hesaplayınız. Bulgunuzu 1987 *GK* ile karşılaştırınız.
- 9.45** Çizelge 9.3'teki 1994 verilerini kullanarak toplanma oranını hesaplayınız. Bulgunuzu 1987 *TO* ile karşılaştırınız.
- 9.46** Aşağıdaki çizelge iki ayrı genel seçimde partilerin oy oranlarıyla çıkardıkları milletvekili sayılarını göstermektedir.

1961 seçimi			2002 seçimi		
Parti	Oy %	MV	Parti	Oy %	MV
Bağımsız	0,8	0	Öbürleri	35,33	0
YTP	13,7	65	Bağımsız	1,00	9
CKMP	14,0	54	CHP	19,39	178
AP	34,8	173	AKP	34,28	363
CHP	36,7	158			
Toplam	100,0	450	Toplam	100,0	450

Bu bölümün son satırlarında bu verilerle çizilmiş iki Lorenz eğrisi verilmiştir.

- Aynı verileri kullanarak iki Gini katsayısını hesaplayınız.
- Bu bağlamda GK nasıl yorumlanabilir? Sıfıra ya da bire yakın olması ne anlama gelir?
- İki GK'yi karşılaştırınız.

9.47 Aşağıdaki çizelgede 2002-03 döneminde üç futbol liginde oynayan takımların attıkları gol sayılarıyla aldıkları puanlar görülmektedir.

Takım	Türkiye		Almanya		İtalya	
	Gol	Puan	Gol	Puan	Gol	Puan
1	63	85	70	75	64	72
2	61	77	53	59	64	65
3	76	66	51	58	55	61
4	61	57	46	56	57	60
5	56	52	52	54	55	56
6	55	51	51	52	38	56
7	44	51	46	49	51	55
8	44	49	39	46	55	49
9	42	43	55	45	36	42
10	37	40	44	45	40	42
11	44	40	47	43	39	41
12	42	39	43	42	36	38
13	40	37	35	41	30	38
14	34	36	40	40	35	38
15	42	36	47	40	38	38
16	48	35	35	36	44	30
17	32	26	33	30	29	24
18	32	22	34	30	23	21

Radikal Futbol, 25 Mayıs 2003, 4 Haziran 2003.

- Üç ligin puanlarını dal-yaprak, kutu çizimle yayılım, çarpıklık, basıklık bakımlarından karşılaştırınız. Bunun için öğrendiğiniz bütün göstergeleri kullanınız. Bulgularınızı yorumlayınız.
- Aynı çözümlemeyi atılan goller bakımından yapınız.

Sorgulayalım, Bakalım Ne Çıkacak?

Bir işin ne kadar iyi yapıldığını ölçmeyi öğrenmeden önce, ne yapılabileceğini anlamak çok daha önemlidir.

John W. Tukey

10.1 SORGULAYICI VERİ ÇÖZÜMLEMESİ

Bundan önceki iki bölümde her veri kümesinde ortak olan bazı özelliklerin ortaya çıkarılması üzerinde durduk. Bu özellikleri ölçebilmek için gerekli bazı araçları inceledik. Veri kümesinin ortalama düzeyini bulmak için kullanabileceğimiz çeşitli ortalamaları, yayıklık, çarpıklık, basıklık ölçülerini öğrendik. Bu ölçüler bize ölçtükleri özelliklere ilişkin özet, ama yalnızca özet bilgiler verebilirler. Bu bölümde göreceğimiz yalın, kolay, hızlı bazı gösterimlerden yararlanarak bu özet bilgilerin bize sağlayacağı yararları çok artırabiliriz. Bir gösterimle gözlemlerin hangi düzey(ler)de yığıldığını, veri kümesinin bakışık olup olmadığını, basıklık ya da sivrililiğinin derecesini bir bakışta görebilir, bir başka gösterimde gözlemlerin dördebölenlerle ortanca arasında nasıl yayıldığını, çarpıklığın yönü ile gücünü, aşırı bir gözlem olup olmadığını hemen anlayabiliriz. Görüldüğü gibi bu araçlar bize aynı anda birden çok özelliği, üstelik de veri kümesinin bütününe yapısını gözden yitirmeden aktarabilirler. Bu nedenle de çok yararlıdırlar.

Sorgulayıcı veri çözümleme teknikleri, istatistik yazınında oldukça yenidir. Tukey'in 1970'lerin başlarında Princeton Üniversitesi'nde geliştirdiği yolların bir kitapta (Tukey, 1977) toplanmasıyla ilk kez geniş okur kitlesine sunulan bu tekniklerin gerisinde yatan ana düşünce, bir veri kümesine değişik inceleme araçlarıyla yaklaşmaktır. Vurgulama, yalın aritmetik işlemlerle göze yönelik gösterimlerdedir. Bir araştırmacı, toplumsal bilim kuramlarını sınamak için toplanmış verileri, kullanmadan önce iyice tanımak zo-

rundadır (Hartwig, Dearing, 1979:5). İşte sorgulayıcı veri çözümlemesi araçları burada önem kazanır, araştırmacıya daha sonra çok işine yarayabilecek değerli ipuçları sağlayabilirler. Her araştırmada önce bu araçları kullanarak işe başlamak doğru olur. Bu nedenle bu tekniklere **ön çözümleme** adı da verilmektedir (Chatfield, 1986).

“Sorgulayıcı veri çözümlemesi, sayısal dedektifliktir. ... Bir suçu araştıran birinin hem araçlara, hem durumu anlamaya gereksinimi vardır. Eğer elinde kullanabileceği parmak izi tozu yoksa, pek çok yüzeyde aradığı izleri bulamayacaktır. Eğer suçlunun nerelerde parmak izi bırakabileceğine ilişkin bir fikri yoksa, bu kez de doğru yerlere bakmayacaktır. Veri çözümleyicisi de aynı durumdadır, hem araçları olması, hem olayı anlaması gerekir.” (Tukey, 1977:1).

Bu bölümde yalnızca tek bir değişkene ilişkin verilerin incelenmesi ele alınacak, iki ya da daha çok değişken arasındaki ilişkilerin araştırılmasına bir sonraki bölümde değinilecektir. Dolayısıyla sorgulayıcı veri çözümlemesi araçlarını yalnızca bu bölümde ele alınanlarla sınırlı sanmak yanlış olur. Tam tersine istatistiğin hemen her alanında kullanılabilecek bu teknikler çeşitli yazarlarca geliştirilmektedir. Bu konuda geniş bir yazın oluşmuş, çeşitli sorgulayıcı veri çözümlemesi araçları ders kitaplarına girmiştir. Bu yayınlar araştırmacıların verilere yaklaşımına yeni bir boyut katmıştır. Özellikle, **SPSS**, **MİNİTAB** gibi bazı bilgisayar yazılımlarına son yıllarda bu araçların da eklenmesi, sorgulayıcı veri çözümlemesinin kullanım alanını daha da yaygınlaştırmıştır (Monk, 1991; Norusis, 1993:181-200).

Sorgulayıcı veri çözümlemesi araçları, dört temel öğeye dayanır (Velleman, Hoaglin, 1981:xv):

- *Gösterimler*, verilerin davranış biçimini, yapısını gözler önüne serer.
- *Kalıntılar*, bir çözümleme sırasında veriler üzerinde yapılan işlemlerden sonra verilerden geriye kalan büyüklüklere dikkat çeker.
- *Dönüştürmeler*, logaritma, kare, karekök gibi kolay matematik işlemlerini kullanarak verilerin yapısını yalın bir biçimde ortaya koyar.
- *Berklik*, olağandışı gözlem(ler) nedeniyle çözümleme sonuçlarının sapmasını önler.

Bu temel öğeler aşağıda yeri geldikçe açıklanacaktır.

9.2 TEK DEĞİŞKENLİ VERİLERİN SORGULANMASI

Sorgulayıcı veri çözümleme yöntemlerini önce tek değişken üzerinde inceleyecek, birden çok değişken içeren veri kümeleriyle ilgili çözümlemeleri bir sonraki bölümde ele alacağız.

Dal-Yaprak

Tek değişkenli bir veri kümesini incelemenin ilk adımı, bu verileri uygun bir çizimle göstermektir. Bu gösterimden beklenen, verilerin yapısını, örüntüsünü ya da genel eğilimini yansıtmasıdır. Gösterimler bazen, daha önce ayırımına varılmamış, beklenmedik, şaşırtıcı öğeleri gözler önüne serebilirler. İşte çok kolay düzenlenebilen dal-yaprak adlı bir gösterim, bu özellikleriyle taşır. Daha önce çizelgeler, çizimler açıklanırken değindiğimiz dal-yaprak gösterimine bakarak şunları görebiliriz:

- Gözlem değerleri nerelerde yoğunlaşmıştır?
- Verilerin yayılma aralığı ne kadardır?
- Küme bakışık mıdır, çarpık mıdır? Ne kadar çarpıktır?
- Veriler basık mı dağılmaktadır, yoksa sivri mi?
- Veri kümesi tek tepeli midir, yoksa çok tepeli mi?
- Gözlem değerleri arasında belirgin boşluklar var mıdır?
- Verilerden herhangi biri ya da bir kaç ötekilerden belirgin biçimde uzak mıdır?

Bir dal-yaprak gösteriminde gözlem değerleri büyüklük sırasına konurlar. Bu sıra, gözün yukarıda sorulan soruların yanıtlarını kolay algılayabileceği bir biçime sokulur. Az sonra açıklanacak kurallarla her gözlem değeri dal basamaklarıyla yaprak basamaklarına ayrılır. Sözgelimi 15210 sayısını alalım. Bu gözlem değeri aşağıdaki gibi değişik biçimlerde dal-yaprak basamaklarına ayrılabilir.

1 | 5 15 | 2 152 | 1 1521 | 0 1 | 52 15 | 21 152 | 10

Dik çizginin solundaki sayılar dal, sağındakiler yapraklardır. Dalın kaç basamak olacağı veri kümesi içindeki bütün gözlem değerlerinin basamak sayılarına bağlı olarak belirlenir. Az sonra buna ayrıntılı olarak değineceğiz. Önce Çizelge 10.1'deki verilerin dal-yaprak gösterimine nasıl dönüştürüleceğine bir göz atalım.

ÇİZELGE 10.1 OECD ülkelerinde vergi gelirlerinin GSYİÜ'ye oranı % (1997)

Ülke	%	Ülke	%	Ülke	%	Ülke	%
Almanya	26,7	Danimarka	33,7	İrlanda	32,4	Portekiz	31,1
ABD	18,5	Finlandiya	28,4	İspanya	28,3	Türkiye	15,2
Avustralya	23,2	Fransa	39,2	İsveç	36,9	Y. Zelanda	31,2
Avusturya	34,4	Hollanda	42,7	İtalya	42,2	Yunanistan	20,6
Belçika	43,0	İngiltere	33,4	Kanada*	18,5		

World Bank, 2000:256-257. * Kanada verisi 1996 yılına ilişkindir.

Bütün gözlem değerlerindeki onlar basamağının dal basamağı, birler basamağının yaprak basamağı olmasına, onda birleri gösteren son basamağın da atılmasına karar verdiğimizizi varsayalım. Buna nasıl karar verdiğimizi sonra göreceğiz. Bu durumda en küçük dal basamağı 1 (ABD, Kanada, Türkiye), en yükseği de 4 (Belçika, Hollanda, İtalya) olacak demektir. Öteki ülkelerin dal basamakları da 2, 3 olur. Bunları büyüklük sırasına göre iki değişik biçimde yazabiliriz.

4		1
3	ya da	2
2		3
1		4

Bilgisayar paket programları küçük sayıyı yukarıya koyarlar. Her iki gösterimin de kendine göre artıları, eksileri var. Büyükten küçüğe doğru sıralama, daha sonra yapılabilecek kutu ya da serpilme gibi bazı çizimlerle uyumludur. Öte yandan küçük sayıdan başlayarak aşağıya doğru gidilerek elde edilen dal-yaprak çizimine, saatin tersi yönünde doksan derece çevriler (bir başka deyişle dikey çizgi yatay duruma getirilerek) bakılırsa, dal-yaprağın histograma benzediği görülecektir. Bu döndürmeyle yaprak değerleri, sıklıkları gösteren birbiri üstüne konmuş kutucuklar gibi algılanabilir. Bu verilerin bir histogramını çizip, sözü edilen döndürmeyi yaparak iki gösterimin birbirine benzediğini görmek okuyucuya bırakılmıştır.

Daha sonra tek tek her gözlem değerinin çizimde gösterilecek yaprak basamakları, ilgili dalın satırına, dik çizginin sağına yazılırlar. Almanya'nın 26,7 değeri, 2 | dalına, 2 | 6 biçiminde yerleştirilir. Bu sırada 26,7'nin 0,07 kadarı kesilip atılmış olur. Böylece bu ayrıntıyı yitiririz. ABD, Avustralya, Avusturya, Belçika gözlem değerleri de benzer biçimde ilgili satırlarda yerlerini alırlar. İlk beş ülkeden sonra dal-yaprak gösterimi şu duruma gelir:

4	3		1	9
3	4		2	63
2	63	ya da	3	4
1	9		4	3

Öteki ülkeleri de buna eklersek dal-yaprağımızı ham biçimiyle elde ederiz:

4	322		1	985
3	43932611		2	63880
2	63880	ya da	3	43932611
1	985		4	322

Bunun üzerinde iki düzeltme, bir de ekleme yapmak gerekir. Düzeltmelerden birincisi aynı satırdaki sayıları büyüklük sırasına dizmek, ikincisi de saymayı kolaylaştırmak üzere her satırdaki yaprak sayılarını beşer beşer öbekleyerek aralarına küçük boşluklar yerleştirmektir. Ekleme ise dal ile yaprak basamaklarının ölçüm birimlerini en altta göstermektir. Böylece Çizim 10.1'e ulaşırız.

4	223		1	589
3	11233 469		2	03688
2	03688	ya da	3	11233 469
1	589		4	223
%10	% 1		%10	% 1

ÇİZİM 10.1 Çizelge 10.1'deki verilerle iki dal-yaprak örneği
Çizelge 10.1'deki verilerden düzenlenmiştir.

Bu gösterimle onda birler basamağındaki ayrıntıları yitirdik ama verileri bir bütün olarak çok daha kolay algılama şansını kazandık. Çizelge 10.1'de fark edemediğimiz şu özellikler, Çizim 10.1'deki dal-yaprak gösterimlerinin ikisinden biri alınarak kolayca görülebilir:

- OECD ülkelerinde vergi gelirlerinin GSYİÜ içindeki payları % 15 ile % 43 arasındadır.
- Bu oranlar %20'li, özellikle de %30'lu düzeylerde yoğunlaşmıştır. 19 ülkeden 13 tanesi bu satırlardadır.
- Dağılım tek tepeli gibi görünmektedir.
- Gözlemler biraz çarpıklık göstermektedir. Küçük oranlara doğru uzama daha çoktur.
- Ülke oranları arasında belirgin boşluklar görülmemektedir.
- OECD ülkelerinden herhangi biri ötekilerden çok uzak değildir. Yani hiç bir ülkenin görece vergi yükünün olağanüstü ölçüde aşırı ya da az sonra göreceğimiz teknik deyimle bir **dışadüşen** olmadığı söylenebilir.
- Bizi özellikle ilgilendiren bir nokta Türkiye'nin konumudur. Dal-yaprak gösteriminden kolayca anladığımız gibi Türkiye'de görece vergi yükü bütün OECD ülkelerinden daha düşüktür.

Aynı verilerle farklı dal-yaprak gösterimleri hazırlanabilir. Bunun için satır sayısı değiştirilir. Yukarıdaki örnekte dal basamakları birer farkla sıralanıyordu. Eğer uygun görürsek bu farkları dal basamağının yarısına indirebiliriz. %10, 20,...,40 değerlerini yansıtan dal basamaklarını ikiye, daha

da ileri giderek beşe bölebiliriz. Elbette gösterimdeki satır sayıları da bunlara uygun olarak artar. Bu örnekler Çizim 10.2’de gösterilmiştir.

		4iü	223
		4 sb	
		3 sd	9
		3 ay	6
		3 db	4
		3 iü	233
		3 sb	11
		2 sd	88
		2 ay	6
		2 db	
		2 iü	3
		2 sb	0
		1 sd	89
		1 ay	
		1 db	5
4	223		
3	69		
3	11233 4		
2	688		
2	03		
1	589		
% 10	% 1	% 10	% 1

ÇİZİM 10.2 Çizelge 10.1’deki verilerle iki dal-yaprak örneği
Çizelge 10.1’deki verilerden düzenlenmiştir.

Önce sol yandaki dal-yaprağı ele alalım. Dal basamakları (1 ile 4 dışında) ikişer kez yazılmışlardır. Bunlardan % 1 basamakları 0 – 4 arasında olanlar bir satıra, 5 – 9 arasında olanlar öteki satıra yerleştirilmiştir. %10 ile başlayan üç ülke verisinin üçü de %15,0 – 19,9 aralığındadır (ABD % 19,8, Kanada % 18,5, Türkiye % 15,2). Dolayısıyla % 10’u gösteren dal satırı iki tane değil bir tanedir. Eğer % 10,0 – 14,9 arasında herhangi bir ülke olsaydı, o zaman bir tane daha 1 satırı eklenecekti. Benzer nedenle % 40 satırı da tektir.

Sağdaki dal-yaprak gösteriminde, % 10 değerleri beş satıra ayrılarak yazılmıştır. İlk satırda % 1 basamakları 0 – 1, ikincide 2 – 3, ..., beşincide 8 – 9 olan ülkeler yer almıştır. Bu çizimde sözgelimi % 20 ile başlayan değerler için kullanılan beş satırın her birinin dal basamağında 2 sayısı bulunmakta, ayrıca % 1’ler basamağı için sıfırla bir sb, ikiyle üç iü, dörtle beş db, altıyla yedi ay, sekizle dokuz sd harf çiftleriyle temsil edilmektedir. İngilizce’de bunlar yerine sırasıyla •, T (two-three), F (four-five), S (six-seven), * kullanılmaktadır. Elbette bu aralıklardan bazılarında hiçbir ülke yoksa o satır boş bırakılmıştır. % 16,0 – 17,9, % 24,0 – 25,9, % 40,0 –

41,9 aralıklarında hiçbir gözlem yer almadığından 1ay, 2db, 4sb satırları boş kalmıştır.

Çizim 10.1 ile 10.2'deki dal-yaprak gösterimlerinden hangisi kullanılmalıdır? Bu sorunun yanıtının veri kümesinin büyüklüğüne bağlı olarak değiştiğini 7. bölümde çizelgeleri incelerken görmüş, küçük kümelerde az satır, büyük kümelerde çok satırın daha uygun olduğunu öğrenmiştik. Daha da ileri giderek satır sayısının üst sınırını, gözlem sayısı (n) 100 ya da daha azsa $2\sqrt{n}$ ile, 100'den çoksa $10 \cdot \log(n)$ ile bulabiliyorduk. Örneğimizde 19 gözlem olduğundan satır sayısının en çok $2\sqrt{19} = 8,6$, yani 8 olması uygundur. 6 satırlık olan dal-yaprağı kullanabiliriz ama 15 satırlık olan gereğinden çok ayrıntılıdır. Hangi dal-yaprak gösteriminin daha iyi olduğunu verilerin özellikleriyle araştırmacının gereksinimleri belirler. Deneyimli veri çözümleyicileri bile tek bir dal-yaprak gösterimiyle yetinmez, bir kaç gösterime bakarak veri özelliklerini daha iyi yakalamaya çalışırlar.

Az veriyle çok satır kullanılırsa ünlü deyimdeki gibi “ağaçlara bakmaktan ormanı göremeyiz”. Çizim 10.2'nin sağ yanındaki dal-yaprakta bu durum açıkça bellidir. Aralarında bir sürü boşluk bulunan sağa-sola dağılmış değerler, verilerin nerede yoğunlaştığını, bakışımın derecesini gözlerden saklamakta, özellikle alt uçta % 15, üst uçta % 42 – 43 gözlem değerleri olan ülkeleri OECD içinde olağandışı ülkeler gibi göstermektedir. Oysa, az sonra göreceğimiz gibi, bu ülkelerin hiçbiri **dışadüşen** değildir. Bu gösterimdeki boş satırlar, komşu iki gözlem arasında en az % 2 fark olduğu anlamına gelir. Bu da yararlı bir bilgidir. Özellikle büyük kümelerle çalışırken bu bilgi önemli olabilir. Aynı çizimin sol yanındaki gösterimde yer alan her boşluk en az % 5'lik fark anlamındadır. Dolayısıyla ancak bu büyüklükte farklara dikkat çekilmiş olmaktadır.

Bu açıklamalardan anlaşılabilceği gibi dal-yaprak gösterimleri ancak üç değişik biçimde dal basamakları içerebilirler. Bu basamaklar ya 1, ya 2, ya da 5 kez yinelenir. Dal basamaklarımız ister milyar TL, ister binde çocuk sayısı olsun, her zaman sağda ya da soldaki sıfırlarından arındırılarak yazılır. Yaprak basamakları da genellikle tek basamak olarak alınır. Eğer daha ayrıntılı bir gösterim istenirse yaprak basamakları ikiye çıkarılabilir, ama o zaman gözlem değerlerini birbirinden ayırmak için aynı satırdakilerin aralarına birer virgül konur. Çizim 10.3'te böyle bir örnek görülmektedir. Bu çizimde Çizelge 10.1'deki veriler aynen korunmuş, hiçbir bilgi yitirilmemiştir. Az sayıda veri için kullanılabilecek bu yolun, gözlem sayısı arttıkça elverişliliği azalır. Çünkü satırların uzunlukları iki katına çıkar.

4	22, 27, 30
3	69, 92
3	11, 12, 24, 34, 37, 44
2	67, 83, 84
2	06, 32
1	52, 85, 98
% 10	% 1, % 1

ÇİZİM 10.3 OECD ülkeleri vergi yükü
Çizelge 10.1'deki verilerden düzenlenmiştir.

Dal-yaprak çiziminin histogramla benzeştiğinden söz etmiştik. Aslında dal-yaprakın histograma göre bazı üstünlükleri olduğu söylenebilir. Bir kere dal-yaprak çok daha kolay düzenlenir. Bu küçümsenebilir bir üstünlük değildir. İnsanlar kolay olanı uygulama eğiliminde olduklarından iki çizim arasında dal-yaprak seçebilirler. İkincisi, dal-yaprak, histograma göre daha ayrıntılı bilgi verir. Çünkü komşu gözlem değerleri arasındaki farklar histogramda yiter gider, oysa dal-yaprakta yapraklarda sayılar yer aldığından komşu iki gözlem arasındaki farklar (atılan basamaklar dışında) gözle görülebilir. Örneğimizdeki ilk dal-yaprak gösteriminde bile % 15 ile % 18 arasında, %20 ile % 23 arasında, % 28 ile % 31 arasında, ya da % 39 ile % 42 arasında boşluklar olduğu açıkça görülürken sıklık çizimi olan histogramlarda bunu yakalamak olanaksızdır. Dolayısıyla dal-yaprak gösterimi aynı verilerle düzenlenmiş sıklık çizimine göre daha çok bilgi içerir (Hartwig, Dearing, 1979:18).

Dal-yaprak çizimlerinin bir başka üstünlükleri de iki veri kümesinin karşılaştırılmalarında sağladıkları kolaylıktır. Bunu da örneğimizi sürdürerek görelim. Aynı ülkelerin Çizelge 10.2'de verilen 1980 yılı verilerine bakalım. Arada geçen sürede OECD ülkelerinde merkezi hükümetlerin topladıkları vergilerin kendi GSYİÜ rakamlarına oranlarının nasıl bir değişim geçirdiklerini inceleyelim.

ÇİZELGE 10.2 OECD ülkelerinde vergi gelirlerinin GSYİÜ'ye oranı % (1980)

Ülke	% Ülke	% Ülke	% Ülke	% Ülke	%
Almanya	*	Danimarka	30,7	İrlanda	30,9
ABD	18,5	Finlandiya	25,1	İspanya	22,1
Avustralya	19,5	Fransa	36,7	İsveç	30,1
Avusturya	31,2	Hollanda	44,1	İtalya	29,2
Belçika	41,2	İngiltere	30,6	Kanada	16,0
				Portekiz	24,1
				Türkiye	14,5
				Y. Zelanda	30,7
				Yunanistan	22,6

World Bank, 2000:256-257. * 1980'de iki Almanya olduğundan veri yazılamamıştır.

Bu verileri 1997 verileriyle birlikte sırt sırta birleştirilmiş iki dal-yaprak gösterimine yerleştirebiliriz. Çizim 10.4'te bu yapılmıştır.

1980		1997
41	4	223
6	3	69
1 00000	3	11233 4
95	2	688
422	2	03
986	1	589
4	1	
% 1	%	% 1
	10	

ÇİZİM 10.4 OECD ülkelerinde vergi yükünün 1980 – 1997 karşılaştırması
Çizelge 10.1, 10.2'deki verilerden düzenlenmiştir.

İki dal-yaprak çizimini böyle sırt sırta koymak karşılaştırmayı gerçekten de çok kolaylaştırır. 1997 verilerini daha önce incelemiştik. 1980 verilerinde en yüksek-en düşük oranlar arasındaki fark (*Ar*) daha büyüktür. 1980'de yaklaşık % 30 olan *Ar*, 1997'de % 27'ye inmiştir. Öyleyse OECD ülkeleri 1980'den 1997'ye geçerken görece vergi yükü açısından az da olsa birbirlerine yaklaşmıştır denebilir mi? Bilmiyoruz çünkü bu sonuç yalnızca iki uçtaki ülkenin vergi yükü oranlarının birbirine yaklaşmasıyla sağlanmıştır. Bu ikisi dışındaki hemen her ülkede bu oran artmıştır. % 20'den düşük oranlı ülke sayısı 4'ten 3'e inmiş, % 40'tan yukarı olanlar da 3'ten 4'e yükselmiştir. Hollanda ile Yunanistan dışında bütün hükümetler 1997'de topladıkları vergi oranlarını 1980'e göre yukarı çekmişlerdir. Bu ilginç bir bulgudur. Küreselleşme yandaşlarının savundukları konulardan biri devleti küçültmektir. Ama 1980 – 1997 döneminde OECD ülkeleri içinde bunu yapan pek yoktur. İşte bu önemli bulgu dal-yaprak gösteriminin gücünü kanıtlamaktadır. Hiçbir aritmetik işlem yapmadan, yalnızca sayıları büyüklük sırasına koyup uygun bir biçimde düzenlemek bizi bu çarpıcı sonuca götürebilmiştir.

Burada bir başka ilginç saptama daha yapabiliriz. Her iki yılda da vergi yükü en düşük ülke Türkiye'dir. Üstelik kayıt dışı ekonominin büyük boyutlara vardığı düşünülürse yukarıdaki oranların aslında daha da düşük olduğu anlaşılır. Sürekli olarak vergi yükünün ağırlığından dem vurulan ülkemizin uluslararası karşılaştırmalardaki konumu bu yakınmaları hiç de

haklı göstermemektedir. Üstelik bu vergilerin önemli bir bölümü iç borç faizi adı altında yine özel kesime aktarıldığına göre ülkemizde vergi adaleti bozuk demektir.

Veri sayısının çok yüksek olduğu durumlarda dal-yaprağın elle çizilmesi zor olabilir. O zaman zaten bilgisayara başvurulacaktır. Dolayısıyla dal-yaprak gösterimi de bilgisayarla çizilebilir.

Dal-yaprak gösterimiyle ikiden çok veri kümesinin birbirleriyle karşılaştırılması da olanaklıdır. Bu konuda kullanacağımız örnek Türkiye’de 1975 – 2000 yılları arasında illerimizin kişi başına gelir düzeyleri olacaktır (Şenesen, 2002). Bu dal-yapraklar Çizim 10.5’te gösterilmiştir. Bu çizime bakarak neler söylenebilir?

- Ortalama düzey yıllar boyunca az da olsa yükselmektedir. Yoğunlaşma ilk dönemlerde 6 – 7 yüz bin TL düzeyindeyken son dönemlerde 8 – 9 yüz bin TL’ye doğru kaymıştır.
- Gelir dağılımı hep büyük değerlere doğru çarpıktır.
- İllerimizin kişi başına gelir dağılımının yayıklığı zamanla artmaktadır. 1975’te illerimizin büyük çoğunluğu kişi başına 400 bin ile 1,2 milyon TL arasında gelir elde ederken, daha sonraki yıllarda bu aralık açılmaya başlamış, 2000 yılında 4 milyon TL’yi geçmiştir. Bunun da anlamı illerimiz ortalama gelir düzeylerinin gitgide farklılaştığıdır.
- Daha da önemlisi ilk dönemlerde tek tepeli olan dağılım zamanla iki, giderek üç tepeliye dönüşmektedir. Bu da illerin kişi başına gelir bakımından kutuplaşmalar oluşturmaya başladığı anlamına gelmektedir.

	1975	1980 (3,5)	1985 (4,1)
3sb			
2sd			
2ay	78		
2db	4		5
2iü			2
2sb		11	
1sd		9	
1ay	6		66667 77
1db	445	4445	44455 55
1iü	22222	22223 33	22233 33
1sb	00000 00011 111	00001 11111	001
0sd	88888 88999 999	88888 89999 999	88899 99999
0ay	66666 66777 77777 77	66666 66777 777	66666 67777 7777
0db	44444 44555 55	44455 55555 5	44448 55555 5
0iü		33333 3	22333 3
m	yüz bin TL	yüz bin TL	yüz bin TL
	1990 (3,9)	1995 (4,0)	2000 (4,4 3,7 3,4)
3sb			1
2sd		9	89
2ay		6	6666
2db	5	5	45
2iü	33	3	3
2sb	001	00111 11	011
1sd	88999	8888	88899
1ay	7	66777	66666 777
1db	44455 555	4455	4455
1iü	22223 33	22333 33	22333 33
1sb	00000 00111 1111	00001 11111	00001 11111
0sd	89999 9	88889 9999	88888 89999 9
0ay	66667 77777	66667 77777 77	66667 7777
0db	44555 555	44455 5	44555 5
0iü	22333	22333 3	33333
m	yüz bin TL	yüz bin TL	yüz bin TL

ÇİZİM 10.5 İllere göre kişi başına GSYİÜ değerlerinin beşer yıllık aralarla dal-yaprak gösterimi (1987 fiyatlarıyla)

Özütün, 1980, 1988, DİE, 1997b, 2003b.

Bu yorumlamalara döneceğiz ama bu kadarıyla bile dal-yaprak gösteriminin ne kadar güçlü bir araç olduğu görülmektedir.

Beşli Özet

Bir veri kümesi için dal-yaprak gösterimini hazırlamak bize yararlı bilgiler sağlayabilir ama birkaç küçük işlem, birkaç çizgi verilere ilişkin bilgilerimizi daha da zenginleştirebilir. Bu yolda kolayca kullanılabilecek bir araç **beşli özet**tir.

Beşli özet, veri kümesindeki iki uç değerin, iki **dördebölenin**, bir de **ortancanın** bulunup alt alta yazılmasından başka bir şey değildir. Öyleyse burada yepyeni bir kavramla karşı karşıya değiliz. Yalnızca eskiden bildiğimiz kavramları yeni bir kullanıma sokmuş oluyoruz.

Beşli özeti oluşturan değerleri bulabilmek için gözlem değerlerini büyüklük sırasına koymamız gerekir. Ama **dal-yaprak** gösterimini zaten düzenlemişsek, bu sıraya koyma işlemi de yapılmış demektir.

Ancak önce dal-yaprak gösterimine küçük bir ekleme yapmak işlerimizi kolaylaştırabilir. Bu ekleme, dal-yaprak gösteriminde her satırın yanına o satırdaki gözlem sayısını yazmaktır. Dal-yaprak çizimini yapan bilgisayar yazılımları bu sıklık sayılarını da kendiliklerinden çizime eklemektedirler. Eğer dal-yaprak çizimi elle yapılmışsa sıklık sayılarını her iki uçtan başlayarak birikimli sıklık olarak yazmak işi biraz daha kolaylaştırabilir (Tukey, 1977:30-35). Bu da iki uçtan başlayarak gözlemleri saymakla eşdeğerdir. Yalnız bunu yaparken görelî sıklık sayısı, *Ortc*'nin sıra sayısını içerecek bir büyüklüğe ulaşmışsa o satıra ayrıca içinde yalnız o satırın sıklık sayısı yazılmalıdır. Eğer *Ortc* iki satırın arasına düşüyorsa, o zaman birikimli sıklıklar hem alttan, hem üstten bu iki satıra kadar yazılır, sonra *Ortc*'nin bu ikisinin arasında olduğunu belirtmek için araya bir çizgi konur.

OECD ülkelerinin merkezi hükümet vergileri örneğine dönersek Çizim 10.6'daki iki ayrı yıla ilişkin dal-yaprak gösterimlerinden beşli özetleri bulabiliriz. Yukarıda açıklanan birikimli sıklık sayılarının yazımı bu çizimde uygulanmış, her dal satırının hizasına bu sayılar yazılmıştır. Böylece iki uçtan başlayarak ortadaki değerlere doğru giden bir gözlem sayma işlemi yapılmış olmaktadır.

1980		1997	
2	41	4	223
3	6	3	69
9	1 00000	3	11233 4 (6)
9	95	2	688
7	422	2	03
4	986	1	589
1	4	1	
% 1		% 10	% 1

ÇİZİM 10.6 OECD ülkelerinde vergi yükünün 1980 – 1997 karşılaştırması
Çizelge 10.1, 10.2'deki verilerden düzenlenmiştir.

1980 kümesi için *Ortc* iki satırın arasına düşmüş, dolayısıyla ikisi de 9 olan birikimli sıklık sayılarının arası bir çizgiyle ayrılmıştır. Bunun anlamı bu çizginin üstünde de, altında da 9'ar gözlem olduğudur. Buradan toplam gözlem sayısının 18 olduğu kolayca anlaşılır. Demek ki *Ortc*, bu yatay çizginin üstündeki satırın en küçük değeriyle altındaki satırın en büyük değeri arasındadır. Bunu dal-yapraktan okursak kabaca $(30 + 29)/2 = 29,5$, Çizelge 10.2'den bakarsak $(30,1 + 29,2)/2 = 29,65$ buluruz. Sorgulayıcı veri çözümleme yöntemleri için dal-yaprağa bakmak yeterlidir. Bilgisayar yazılımları daha hassas olan değeri hesaplar.

Ortc bulunduktan sonra sıra *D* 'lere gelir. Onlar da hesaplanınca

$$\begin{aligned}
 &Y_{eb} \text{ (en büyük değer)} \\
 &D_{\bar{a}} \\
 &Ortc \\
 &D_a \\
 &Y_{ek} \text{ (en küçük değer)}
 \end{aligned}$$

alt alta sıralanıp beşli özet yazılmış olur.

Daha önce *D*'lerin üç ayrı yolla bulunabileceğini görmüş, bunlardan geleneksel yolu incelemiştik. Şimdi de öteki yolları açıklayalım.

$\frac{1}{2} (N + 1)$, *Ortc*'nin sıralamadaki yerini gösterir. *N* tekse bu konum tamsayıdır, çiftse değildir. Bu sıra numarasının tamsayı kısmına bir ekleyerek ikiye bölersek *D*'lerin iki uca göre konumunu elde ederiz. Bu konumdaki *D*'lere **Tukey menteseleri** denir. Daha da ileri gidilerek gözlemleri sekize, on altıya, otuz ikiye bölen değerler benzer yolla bulunurlar (Hoaglin, 1983:34-38). Dikkatli bir okuyucu bu yola başvurulduğunda

Ortc'nin hem üst hem alt dördebölen bulunurken iki kez hesaba katıldığını anlamıştır.

Bizim önerdiğimiz daha basit hesap yoluna göre, *Ortc*'nin altında kalan gözlemlerin ortancasına D_a , üstünde kalanların ortancasına da $D_{\bar{u}}$ denir.

Başka bir deyişle *Ortc* bir kez hesaplandıktan sonra dışlanmakta, ne alt ne de üst dördebölenin hesabına girmektedir.

Dördebölenleri hesaplamada bu en basit yolu kullandığımızda *Ortc* ile dördebölenlerin kaç tanesinin gözlem değerlerine tam eşit olacağı, kaç tanesinin iki gözlem değerinin arasına düşeceği kolayca görülebilir. Toplam gözlem sayısı N , 4'ün tam katı ($4k$), tam katının bir fazlası ($4k + 1$), iki fazlası ($4k + 2$), üç fazlası ($4k + 3$) olabilir (k tamsayı). $4k$ 'ye eklenen 0, 1, 2, 3 sayıları, *Ortc* ile dördebölenlerden kaçının veri kümesindeki gözlem değerlerine tam karşılık geldiğini gösterir.

Eğer $N = 4k$ ise, ne *Ortc*, ne dördebölenler, gözlem değerlerinden herhangi birine eşit olur, hepsi iki komşu gözlem değerinin arasına düşer, bu iki değer ortalaması olarak bulunurlar. Bu söylediğimizi gözlem değerlerini büyüklük sırasına dizerek şöyle gösterebiliriz. Y_1 en küçük, Y_N en büyük değer olsun.

$$\begin{array}{ccccccc} Y_1 & Y_2 & \dots & Y_k & \uparrow & Y_{k+1} & \dots & Y_{2k} & \uparrow & Y_{2k+1} & Y_{2k+2} & \dots & Y_{N-k-1} & \uparrow & Y_{N-k} & Y_{N-k+1} & \dots & Y_{N-1} & Y_N \\ \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow \\ D_a & D_{\bar{u}} \end{array} \rightarrow = 4k$$

$N = 4k + 1$ ise *Ortc* gözlem değerlerinden tam ortadakine eşittir. *Ortc*'yi toplam gözlem sayısından çıkarırsak $N - 1 = 4k$ olur. O zaman da ortancanın her iki yanındaki gözlem sayısı ($2k$) çift olur. Dolayısıyla dördebölenlerden hiçbirisi herhangi bir gözlem değerine eşit olmaz, komşu iki gözlemin tartısız ortalaması olarak bulunur. *Ortc*, $2k+1$ 'inci gözlem değeri olur; D_a , k ile $k+1$ 'inci; $D_{\bar{u}}$ de $N - k$ ile $N - k+1$ 'inci gözlem değerlerinin ortalamalarıdır.

$$\begin{array}{ccccccc} Y_1 & Y_2 & \dots & Y_k & \uparrow & Y_{k+1} & \dots & Y_{2k} & Y_{2k+1} & Y_{2k+2} & \dots & Y_{N-k-1} & \uparrow & Y_{N-k} & Y_{N-k+1} & \dots & Y_{N-1} & Y_N \\ \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow \\ D_a & D_{\bar{u}} \end{array} \rightarrow = 4k + 1$$

$N = 4k + 2$ ise çift bir sayıdır, o halde *Ortc* bir ara değerdir. Gözlemler ikiye bölündüğünde *Ortc* arada kalır. *Ortc*'nin hem altında hem üstünde kalan gözlem sayıları ($\frac{1}{2}N = 2k + 1$) tek olduğu için dördebölenler bunların tam ortalarındaki gözlem değerlerine eşittir. Yani D_a , $k + 1$ 'inci, $D_{\bar{u}}$ ise $N - k = 3k + 2$ 'inci gözlem değerleridir.

$$\begin{array}{ccccccc} Y_1 & Y_2 & \dots & Y_k & Y_{k+1} & \dots & Y_{2k} & Y_{2k+1} & \uparrow & Y_{2k+2} & \dots & Y_{N-k-1} & Y_{N-k} & Y_{N-k+1} & \dots & Y_{N-1} & Y_N \\ \leftarrow k \text{ tane} \rightarrow & D_a & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow & \uparrow & \leftarrow k \text{ tane} \rightarrow \\ & D_{\bar{u}} \end{array} \rightarrow = 4k + 2$$

Ortc

$N = 4k + 3$ ise hem *Ortc*, hem D' 'ler, birer gözlem değerine eşittir. Tam ortaya denk gelen *Ortc* gözlem toplamdan çıkarılır da kalanın yarısı alınırsa $(\frac{1}{2}(N - 1) = 2k + 1)$ tek bir sayıya ulaşılır. Bu da bize dördebölenlerin de birer gözlem değerine eşit olduğunu gösterir.

$$Y_1 Y_2 \dots Y_k \quad Y_{k+1} \dots Y_{2k} \quad Y_{2k+1} \quad Y_{2k+2} \dots Y_{N-k-1} \quad Y_{N-k} \quad Y_{N-k+1} \dots Y_{N-1} \quad Y_N \\ \leftarrow k \text{ tane} \rightarrow D_a \leftarrow k \text{ tane} \rightarrow \text{Ortc} \leftarrow k \text{ tane} \rightarrow D_u \leftarrow k \text{ tane} \rightarrow = 4k + 3$$

Bu anlatılanları küçük bir çizelgede örnekleyip üç hesaplama yoluyla bulunacak konumları karşılaştırabiliriz (Çizelge 10.3).

ÇİZELGE 10.3 Çeşitli yollara göre D_a , *Ortc*, D_u 'nün sıra konumları

Hesaplama yolu	$N = 4k$ $N = 16$	$N = 4k + 1$ $N = 17$	$N = 4k + 2$ $N = 18$	$N = 4k + 3$ $N = 19$
Geleneksel	4,25–8,5–12,75	4,5–9–13,5	4,75–9,5–14,25	5–10–15
Tukey menteşeleri	4,5–8,5–12,5	5–9–13	5–9,5–14	5,5–10–14,5
Basit	4,5–8,5–12,5	4,5–9–13,5	5–9,5–14	5–10–15

Görüldüğü gibi her yol her zaman aynı *Ortc* değerine ulaşmaktadır. Geleneksel yolun D_a 'ları Tukey menteşelerine göre hep daha küçük, D_u 'leri hep daha büyük çıkma eğilimindedir. Bizim önerdiğimiz yol, tek sayılı veri kümelerinde geleneksel yolla, çift sayılı veri kümelerinde Tukey menteşeleriyle aynı sonucu vermektedir.

Şimdi örneğimizdeki D' 'leri bulalım. Önce geleneksel yolu kullanalım. Gözlem sayısı $N = 18$ olduğuna göre $\frac{1}{4}(18 + 1) = 4,75$ konumundaki değer D_a 'yı, $\frac{3}{4}(18 + 1) = 14,25$ konumundaki de D_u 'yü verir. Öyleyse 4. ile 5. gözlem değerleri arasındaki farkın % 75'i 4. gözleme eklenmelidir. Dal-yaprakta alttan 4. gözlem 19'dur, 5. gözlem 22'dir. $22 - 19 = 3$ 'ün % 75'i olan 2,25'i 19'a ekleyip $D_a = 21,25$ buluruz. Benzer biçimde alttan 14. gözlem 30'dur, 15. gözlem 31'dir. Demek ki $D_u = 30,25$ olur.

Tukey menteşelerini ise şöyle hesaplarız. *Ortc*'nin konumu 9,5 olduğundan bunun tamsayı kısmına 1 ekleyip ikiye bölelim. $\frac{1}{2}(9 + 1) = 5$. Alttan 5. gözlem değeri D_a 'yı, üstten 5. gözlem değeri de D_u 'yü verir. Bunları da dal-yapraktan $D_a = 22$, $D_u = 30$ olarak okuruz.

Bizim önerdiğimiz basit yola göre *Ortc*'nin altındakiyle üstündeki dokuzar değerın tam ortalarına denk gelen beşinci gözlemler dördebölenlerdir. Öyleyse $D_a = 22$, $D_u = 30$ 'dur.

Çizim 10.6'daki 1997 verilerinde ise *Ortc* üçüncü satırdaki gözlemlerden biri olmuş, dolayısıyla bu satırın hizasına birikimli sıklık değil, yalnız-

ca o satırın sıklık değeri (6) yazılmıştır. Bu satırın üstünde 5, altında 8 gözlem bulunmaktadır. Bu üç sayının toplamı bize 1997’de 19 gözlem olduğunu gösterir. Öyleyse *Ortc* (her iki uçtan 10. gözlem), o satırdaki 5. en büyük ya da 2. en küçük değerdir. Bunun da dal-yaprğa göre 31 olduğu açıktır.

Geleneksel yola göre alttan $\frac{1}{4} (19 + 1) = 5$. gözlem değeri dal-yapraktan $D_a = 23$, benzer biçimde alttan $\frac{3}{4} (19 + 1) = 15$. gözlem değeri yine aynı çizelgeden $D_{\bar{a}} = 36$ bulunur.

Tukey menteşelerini hesaplamak istersek önce *Ortc*’nın konumunu bulmamız gerekir. Bu da $\frac{1}{2} (19 + 1) = 10$ ’dur. Öyleyse *D*’lerin iki uca göre konumları da $\frac{1}{2} (10 + 1) = 5,5$ olarak bulunur. Alttan 5. gözlem 23 ile 6. gözlem 26’nın orta noktası $D_a = 24,5$ çıkar. Üstten 5. gözlem 36 ile 6. gözlem 34’ün orta noktası da $D_{\bar{a}} = 35$ bulunur.

Basit yola başvurursak ortancanın hem altında, hem üstünde dokuzar gözlem olduğunu görürüz. Dolayısıyla bunların ortalarındaki 5. gözlemler *D*’lerdir. Bunların dal-yapraktan okunan değerleri de $D_{\bar{a}} = 36$, $D_a = 23$ ’tür. Bütün bulgular Çizim 10.4’teki gibi özetlenebilir.

ÇİZELGE 10.4 Vergi / GSYİÜ verileri için beşli özetler
(Dal-yaprak kesilmiş değerleri kullanılarak)

1980			1997		
Gele- neksel	Tukey menteşeleri	Basit yol	Gele- neksel	Tukey menteşeleri	Basit yol
44	44	44	43	43	43
30,25	30	30	36	35	36
29,5	29,5	29,5	31	31	31
21,25	22	22	23	24,5	23
14	14	14	15	15	15

Çizim 10.6’dan düzenlenmiştir.

Yukarıda “... her yol her zaman aynı *Ortc* değerine ulaşmaktadır. Geleneksel yolun D_a ’ları Tukey menteşelerine göre hep daha küçük, $D_{\bar{a}}$ ’leri hep daha büyük çıkma eğilimindedir. Bizim önerdiğimiz yol, tek sayılı veri kümelerinde geleneksel yolla, çift sayılı veri kümelerinde Tukey menteşeleriyle aynı sonucu vermektedir” demiştik. 1980’de 18, 1997’de 19 gözlem olduğuna göre örneğimizde bu söylenenlerin doğruluğunu okuyucu denetlemelidir.

Bildiğimiz gibi dal-yaprak gösterimi düzenlenirken, gözlem değerlerinde fazla basamak varsa atılır, kalan basamaklar kullanılır. Dolayısıyla Çizelge 10.4'teki bulgular hep kesilmiş verilerden elde edilmiştir. Oysa bu değerler bazı basamaklar kesilmeden önceki asıl gözlemler üzerinden bulunsaydı, geleneksel yol ile Tukey menteşeleri yukarıdaki değerlerinden az da olsa farklı çıkabilirdi. Bizim önerdiğimiz basit yol her zaman kesilmiş (dal-yaprak) verileri üzerinden hesaplanmalıdır.

Çizelge 10.5'te aynı verilerin kesilmemiş asıl değerlerinden bilgisayar yazılımı kullanılarak bulunmuş beşli özetler yer almaktadır.

ÇİZELGE 10.5 Vergi / GSYİÜ verileri için beşli özetler
(Kesilmemiş asıl değerler kullanılarak)

1980				1997			
Gele- neksel	Tukey menteşeleri	Basit yol		Gele- neksel	Tukey menteşeleri	Basit yol	
44,1	44,1	44,1	Y_{eb}	43,0	43,0	43,0	
30,975	30,9	30,9	$D_{\bar{u}}$	36,9	35,65	36,9	
29,65	29,65	29,65	$Ortc$	31,2	31,2	31,2	
21,45	22,1	22,1	D_a	23,2	24,95	23,2	
14,5	14,5	14,5	Y_{ek}	15,2	15,2	15,2	

Çizim 10.6'dan düzenlenmiştir.

İki çizelge karşılaştırılırsa bulunan sonuçların pek az farklı oldukları görülür. Burada sorgulayıcı veri çözümleme yöntemlerini açıkladığımızdan basit yolla bulunan beşli özetleri kullanacağız.

Dışadüşenler

Beşli özet oluşturulduktan sonra **dördebölenler aralığı** (DAr) hesaplanır. DAr kullanılarak gözlemler arasında **dışadüşen** olup olmadığı anlaşılabilir. DAr 'ın bir buçuk katı bir **Adım** olarak adlandırılır. Bu **Adım** değeri alt dördeböleden (D_a) çıkarılır, üst dördebölene ($D_{\bar{u}}$) eklenir, böylelikle **alt eşik** (E_a) ile **üst eşik** ($E_{\bar{u}}$) değerleri bulunur. Bu eşik değerlerinin dışında kalan gözlemler dışadüşen olarak nitelenir. Bu eşikler birer **Adım** daha genişletilirse **uzak eşikler** bulunur. Bunların dışına düşen değerlere de **uzak dışadüşen** denir. Dışadüşen, özel ilgi bekleyen bir gözlem değeri olarak algılanmalıdır. Buradaki 1,5 katsayısı, birçok veri kümesiyle yapılan deneylerden elde edilmiştir (Emerson, Strenio, 1983:62). Normal dağılıma tam

uyan bir veri kümesinde eşik dışında kalan gözlem sayısı, toplam gözlemlerin % 7'si kadardır.

1980 için hemen şu hesaplamaları yapabiliriz:

$$D_{Ar} = 30 - 22 = 8$$

$$Adım = 1,5 \times D_{Ar} = 1,5 \times 8 = 12$$

$$E_{\bar{u}} = 30 + 12 = 42$$

$$E_a = 22 - 12 = 10$$

$E_{\bar{u}} = 42$ 'den büyük olan 44 gözlem değeri bir dışadüşendir. Demek ki Hollanda'nın 1980 yılında topladığı vergilerin GSYİÜ'ye oranı, OECD ülkeleri içinde ötekilerden epeyce yüksektir. Bir başka deyişle bu ülke öbür OECD ülkeleri kümesinden dışa düşmekte, deyim yerindeyse sürüden ayrı durmaktadır. Alt uçta ise dışadüşen yoktur, çünkü en küçük değer bile alt eşik içinde, yani $E_a < Y_{ek}$.

1997 için de şu hesaplamalar yapılabilir:

$$D_{Ar} = 36 - 23 = 13$$

$$Adım = 1,5 \times D_{Ar} = 1,5 \times 13 = 19,5$$

$$E_{\bar{u}} = 36 + 19,5 = 55,5$$

$$E_a = 23 - 19,5 = 3,5$$

Burada dışadüşen yoktur. 1997'de bütün OECD ülkeleri, vergi gelirlerinin GSYİÜ'ye oranı bakımından az çok birbirlerine benzerler. Az önceki deyimle sürüden ayrılan yoktur.

Yukarıdaki beşli özete bu bulgular eklenerek daha da genişletilmiş Çizelge 10.6 elde edilir.

ÇİZELGE 10.6 Vergi / GSYİÜ verileri için genişletilmiş beşli özetler

1980		1997
44	Y_{eb}	43
30	$D_{\bar{u}}$	36
29,5	$Ortc$	31
22	D_a	23
14	Y_{ek}	15
18	N	19
8	D_{Ar}	13
12	$Adım$	19,5
42	$E_{\bar{u}}$	55,5
10	E_a	3,5
44	Üst dışadüşen	yok
yok	Alt dışadüşen	yok

Çizim 10.6'dan düzenlenmiştir.

Bu çizimde *Adım*'ın yanı sıra eşikler de gösterilmiş, dışadüşen varsa değeri yazılmış, yoksa olmadığı belirtilmiştir. Uzak eşikler gerekmediği için gösterilmemiştir.

Bulduğumuz iki beşli özete göre neler söylenebilir? 17 yıl içinde OECD ülkelerinde merkezi hükümetlerin topladıkları vergilerin GSYİÜ'ye oranlarının D_a düzeyinde ciddi, $Ortc$, D_w , Y_{ek} düzeylerinde de küçük artışlar olmuş, Y_{eb} düzeyiyse biraz düşmüştür. Demek ki en üst düzeydeki ülke dışında vergi / GSYİÜ oranı OECD ülkelerinde genellikle yükselmiştir. Bu yükseliş üst sıralardaki ülkelerde daha büyük boyutlardadır.

Dar da önemli ölçüde büyümüştür. Bu da OECD ülkelerinin incelenen değişken bakımından aralarındaki farkların arttığı anlamına gelir. 1980 verilerinde Hollanda öteki ülkelere çok farklı, aykırı bir ülke olarak göze çarparken 1997'de böyle bir ülkeye rastlanmamıştır.

Dördebölenler arasındaki ülkelerin gözlem değerlerinin *Ortc*'ya göre çarpıklığını **Bowley çarpıklık ölçüsü**yle bulabileceğimizi görmüştük. Bu ölçüyü örneğimizdeki verilerle hesaplırsak

$$1980: ((30 - 29,5) - (29,5 - 22)) / (30 - 22) = -7/8$$

$$1997: ((36 - 31) - (31 - 23)) / (36 - 23) = -3/8$$

buluruz. Öyleyse çarpıklık 1980'de ortanca altı değerlere doğru ciddi ölçülerdeyken, 1997'de bu çarpıklık yine ortanca altı değerlere doğru olmuş, ama bu kez hafiflemiştir. Başka bir deyişle vergi / GSYİÜ oranı düşük olan ülkeler 1980'de ortalamadan iyice uzak konumdayken 1997'de ortalama düzeye doğru yaklaşmışlardır. Bu yorumlara ileride yeniden göz atılacaktır.

Şimdi başka bir örneğe, illerimizin kişi başına GSYİÜ verilerine dönelim. 1975, 1980, 1985 yıllarının her birinde 67, 1990'da 71, 1995'te 76, 2000 yılında 81 ilimiz vardır. $67 = 4 \times 16 + 3$, $71 = 4 \times 17 + 3$, $76 = 4 \times 19$, $81 = 4 \times 20 + 1$ olduğu düşünülürse, ilk dört döneme ilişkin beşli özetlerde hem *Ortc*, hem de *D*'lerin birer ilimizin verilerine karşılık geldiği anlaşılır. 1995 yılında ise bunlardan hiçbirinin, hiçbir ille eşleşmediği, gerek *Ortc*, gerek *D*'lerin iki il değerinin arasında bulunacağı anlaşılmaktadır. 2000 yılında yalnızca *Ortc* bir ilin değerine eşit olmakta, *D*'ler iki il arasına düşmektedir. Şimdi beşli özetlerimizi bulalım, sonra da yorumlayalım (Çizelge 10.7). E_a her dönem en küçük gözlemin altında olduğundan çizelgede gösterilmemiştir.

ÇİZELGE 10.7 İllerimizde kişi başına GSYİÜ (1987 fiyatlarıyla) beşli özetleri

(100 bin TL)	1975	1980	1985	1990	1995	2000
N	67	67	67	71	76	81
Y_{eb}	27	35	41	39	40	44
$D_{\bar{u}}$	10	11	13	14,5	16,5	18
Ortc	8	8	8	10	10,5	11
D_a	6	5,5	5,5	7	7	7,5
Y_{ek}	4	3	2	2	2	3
D_{Ar}	4	5,5	7,5	7,5	9,5	10,5
Adım	6	8	11	11	14	16
$E_{\bar{u}}$	16	19	24	25,5	30,5	34
Uzak $E_{\bar{u}}$	22	27	35	36,5	44,5	50
Dışadüşen	yok	21, 21	25	yok	40	34, 37, 44
Uzak dışadüşen	24, 27	35	41	39	yok	yok

Çizim 10.5'teki verilerden düzenlenmiştir.

- *Ortc* değeri artmakta, dolayısıyla illerimizde kişi başına gelirin ortalama düzeyi zaman içinde yükselmektedir.
- *D_{Ar}* genişlemekte, iller arasında gelir farklılıkları büyümektedir.
- Hem D_a hem $D_{\bar{u}}$ 1975'te *Ortc*'ya eşit uzaklıktayken daha sonra bu eşitlik bozulmuş, $D_{\bar{u}}$ *Ortc*'dan daha fazla uzaklaşmıştır. Bunun anlamı *Ortc* üstündeki illerin birbirlerine göre gelir farklılıklarının alttakilere oranla daha hızlı büyümekte olduğudur.
- Her dönemde ya dışadüşen ya da uzak dışadüşen bulunduğu göze çarpmaktadır. Son yıllara doğru uzak dışadüşenlerin görülmemesi dikkat çekicidir. Bu da, öteki illerden kopuk, aşırı yüksek kişi başına geliri olan bir kaç ilin, geriden gelen hızlı bazı illerce yakalanmak üzere oldukları anlamına gelir.

67-81 il için 26 yılın gelir düzeylerini sayısal olarak gösteren bir çizelgeye bakarak bu tür ipuçları elde etmek neredeyse olanaksızdır. Oysa çok yalın, çok çabuk bazı işlemlerle oluşturulan dal-yapraklar, beşli özetler verilerde gizli yapıları ele vermektedirler.

Kutu-Çizimi

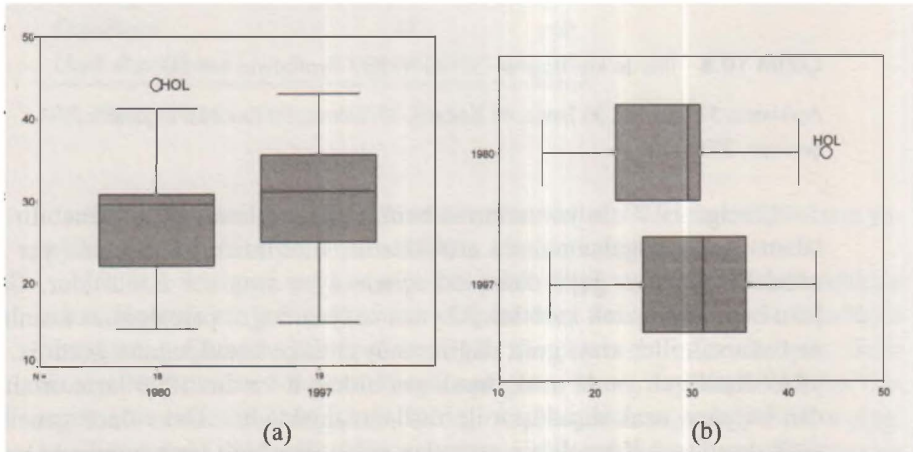
Beşli özetle bulunanların bir de çizimle gösterilmesi anlamayı iyice zenginleştirebilir. Kutu gösterimi bu iş için biçilmiş kaftandır. Bu çizim, beşli özetle elde edilen bilgilerle yetinir, ama onları çok daha göze yönelik biçimde sergiler. Kutu çizimleri, adından da belli olduğu gibi, dikdörtgen bir

kutuyla bundan iki yana doğru uzayan iki çizgiden oluşur. Gözlemlerin ölçüldüğü ölçek istenirse dik, istenirse yatay eksene yerleştirilebilir. Bilgisayar yazılımlarında genellikle her iki yol da olanaklıdır. Biz burada duruma göre her iki eksen kullanacağız. Ölçme eksenini hem Y_{eb} hem Y_{ek} değerlerini içerecek biçimde çizilir, bu iki uç nokta uygun yerlere işaretlenir. Sonra D' 'ler birer çizgiyle gösterilirler. Bu iki çizginin karşılıklı uçları birleştirilerek bir dikdörtgen kutu elde edilir. $Ortc$ 'nin yeri de bu kutu içinde bir çizgiyle işaretlenirse kutu $Ortc$ düzeyinden ikiye bölünür. Daha sonra dışadüşen olmayan en uzak değerler birer doğruyla bu kutuya birleştirilip, varsa dışadüşenler de belirtilirse çizim bitmiş olur. Bu tür bir çizimi daha önce 9. bölümde kısaca görmüştük.

Ancak veriler arasında dışadüşen varsa yer(ler)inin belirtilmesi gerekir. Bu durumda kutudan dışadüşenin bulunduğu yöne doğru giden doğru parçası uç değere kadar uzatılmaz, dışadüşen(ler)e en yakın ama eşğin içinde kalan gözlem değerinde kesilir.

Çizim 10.7(a) ile (b), 1980, 1997 yıllarında OECD ülkelerindeki vergi / GSYİÜ oranları için böyle bir çizimi bir kez dik, bir kez de yatay eksen kullanarak göstermektedir.

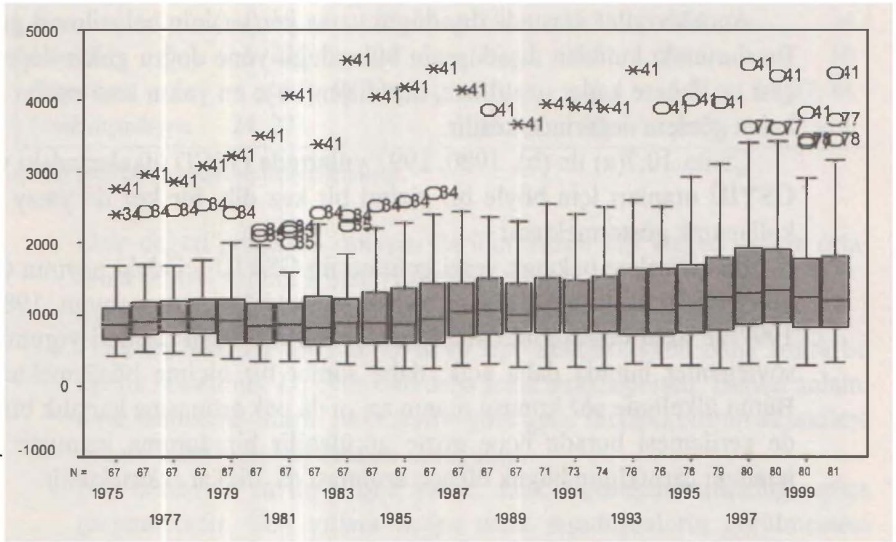
Bu çizimlere bakarak vergi gelirlerinin GSYİÜ içindeki payının OECD ülkelerinde nasıl bir dağılım yapısı gösterdiğini, bu yapının 1980'den 1997'ye nasıl değiştiğini daha iyi anlayabiliriz. Beşli özetleri yorumlarken söylenenler burada daha açık, daha somut bir biçime bürünmektedirler. Bütün ülkelerde söz konusu oranın az ya da çok artmasına karşılık bir ülkede gerilemesi burada iyice gözle görülebilir bir duruma gelmiştir. Kutu içindeki çarpıklığın büyük ölçüde azalması da dikkat çekmektedir.



ÇİZİM 10.7 OECD ülkelerinde vergi / GSYİÜ 1980 – 1997 karşılaştırması

1980’de dışadüşen ülke olan Hollanda’nın ayrı gösteriliş biçimine dikkat ediniz. Kutudan çıkan çizgi Hollanda’ya kadar değil, ondan bir alt değerdeki ülkeye kadar uzatılmıştır. Eğer bir gözlem uzak dışadüşen ise, o gözlemi gösteren noktanın daha da belirginleştirilmesi önerilir (Tukey, 1977:47-48). Ayrıca dışadüşen gözlemlerin hangileri olduğu ya adları ya da sıra numaraları yanlarına yazılarak belirtilir.

Öteki örneğimize döner de iller düzeyinde gelir dağılımı verileri için kutu çizimlerini yaparsak uzak dışadüşenlerin de bu çizimde nasıl gösterileceklerini öğrenmiş oluruz. Çizim 10.8, 1975-2000 yıllarında illere göre kişi başına GSYİÜ düzeylerini göstermektedir.



ÇİZİM 10.8 İllerde kişi başına GSYİÜ (1987 fiyatlarıyla bin TL)

Açıklama: 34 İstanbul, 35 İzmir, 41 Kocaeli, 77 Yalova, 78 Karabük’ü gösterir. Şenesen, 2002:119.

Çizelge 10.7’nin yorumlarına benzer şeyler burada da söylenebilir. Ortalama gelir dalgalanmalarla artmaktadır. İllerimizin kutu içinde yer alan ortadaki yarısının gelir düzeyleri içinde aynı artış söz konusudur. Gerek kutu boyunun, gerek kutudan iki yana uzayan doğru parçalarının uzunluğuna bakarsak iller arası gelir dağılımının gittikçe bozulduğunu görürüz. Her yılda dışadüşen ya da uzak dışadüşen birkaç il vardır. 1990’ların ortalarından bu yana uzak dışadüşen ile rastlanmamaktadır. Daha önce genellikle uzak dışadüşen Kocaeli’nin ardından gelen iller hızlı bir büyümeyle bu ilin

belirtilen özelliğini ortadan kaldırmıştır. Alt sınırdaki illerin gelir düzeyleri 1980'den sonra kötüleşmiş, daha sonra da bir türlü düzelememiştir.

Bu arada bütün değerlerin 1987 fiyatlarıyla verildiğini bir kez daha belirtelim. Okuyucuya bu gelir düzeyleri çok düşük gelebilir. Bir fikir vermek amacıyla 1987'de uzak dışadüşen Kocaeli'nde 4.150.724 TL olarak hesaplanan kişi başına gelirin 4.757 \$'a karşılık geldiği söylenebilir.

Birinde dışadüşen, öbüründe uzak dışadüşen birer il olan 1988, 1989 yıllarını örnek olarak alıp *Adım*, *Eşik* büyüklüklerini, dışadüşenleri hesaplayalım. Bu yılların beşli özetleri Çizelge 10.8'de verilmiştir. İki yılda da Y_{ek} 'nin altında kalan E_a hesaplanmamıştır.

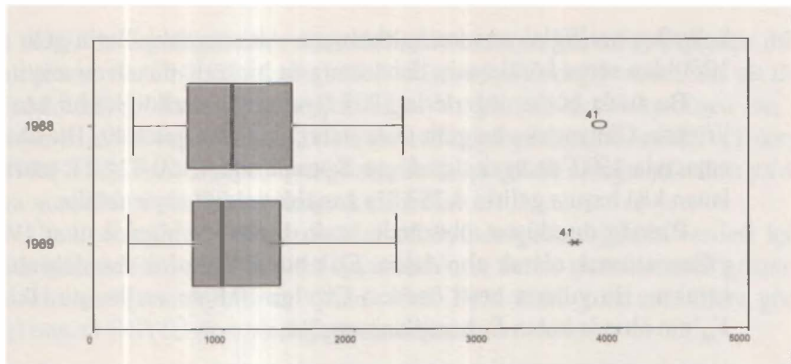
ÇİZELGE 10.8 İller kb GSYİÜ beşli özetleri 1988, 1989
(1987 fiyatlarıyla 100 bin TL)

	1988	1989
N	67	67
Y_{eb}	38	36
$D_{\bar{u}}$	15	14
Ortc	10	9
D_a	7	7
Y_{ek}	3	2
<i>DAr</i>	8	7
<i>Adım</i>	12	10,5
$E_{\bar{u}}$	27	24,5
<i>Uzak $E_{\bar{u}}$</i>	39	35
<i>Dışadüşen</i>	38	yok
<i>Uzak dışadüşen</i>	yok	36

DİE, 1997:712-715'ten hesaplanmıştır.

Bütün bunları Çizim 10.9 üzerinde gösterelim. Bu çizimde kutuların yatay uzunlukları *DAr*'ye, bunun 1,5 katı *Adım*'a karşılıktır.

1988'de kutunun sağ kenarına bir *Adım* değil ama iki *Adım* eklendikten sonra Kocaeli'nin gelir düzeyine ulaşılmakta, dolayısıyla bu il dışadüşen olmaktadır. Çizimde bu özellik minik bir çemberle gösterilmiştir. Gelir düzeyi bakımından Kocaeli'nden sonra gelen İzmir, İstanbul ilk *Adım*'dan sonra elde edilen $E_{\bar{u}}$ 'nin içinde kaldığından kutudan sağa doğru çıkan çizgi bu illerin gelir düzeyine kadar uzatılmış, bu noktadan Kocaeli düzeyine kadar olan bölüm çizimde boş bırakılmıştır.



ÇİZİM 10.9 İller kb GSYİÜ – 1988, 1989 (1987 fiyatlarıyla bin TL)

DİE, 1997:712-715'ten düzenlenmiştir.

1989'da da benzer bir yol izlenmiş, bu kez uzak E_u dışında kalan Kocaeli uzak dışadüşen olmuş, bu da çizimde bir yıldızla gösterilmiştir. Yine E_u 'nün içinde kalan İstanbul, İzmir illerinin gelir düzeyine kadar uzatılan sağ çizgi burada kesilmiş, Kocaeli yıldızı bir boşluktan sonra çizilmiştir.

Çizimden görüldüğü gibi Kocaeli'nin gelir düzeyi 1988'de 1989'dan yüksek olmasına karşın bu il 1988'de dışadüşen, 1989'da uzak dışadüşen niteliğini taşımaktadır çünkü önemli olan bu niteliğin belirlenmesinde yalnızca il gelirinin mutlak düzeyi değil, öbür illerin gelirlerine göre saptanan konumudur.

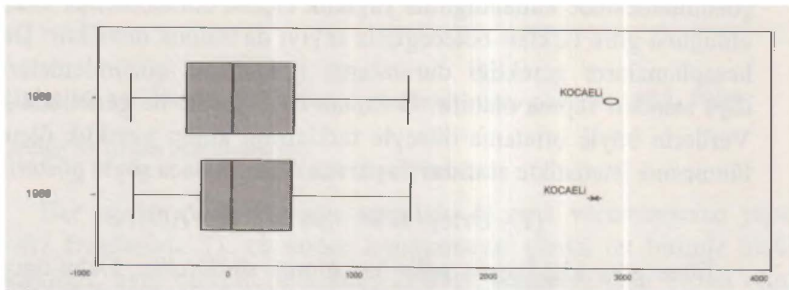
Buraya kadar kutu gösterimlerinin nasıl çizileceğini, nasıl yorumlanacağını gördük. Her zaman olduğu gibi veri yapısının incelenmesine ilk önce ortalama düzeyden başladık. Bu özellik belirlenip bir kenara yazıldıktan sonra dağılımın biçim özelliklerine daha açık bir görüşle eğilebilmek için, özellikle birden çok dağılımın karşılaştırılmasında, düzey farklılıklarının ortadan kaldırılması gereği karşımıza çıkar. Bunu da istatistikte çok yaygın kullanılan bir yolla, gözlem değerlerinin ortalama düzeyden uzaklıklarının ele alınmasıyla sağlayabiliriz. Sorgulayıcı veri çözümlemelerinde, hesaplanma kolaylığından dolayı *Ortc* kullandığımız için gözlem değerlerinin *Ortc*'dan farklarının alınması, bu işlemin her kümede yinelenmesi, kümeler arasındaki düzey farklarını giderir. Çabuk hesap yapmak istediğimize göre bu fark alma işlemini yalnızca beşli özetle uygulamamız yeterlidir. Böylece gözlem değerlerinin ortancadan farklarından oluşan yeni beşli özetler, dolayısıyla da yeni kutu çizimlerine gidebiliriz. Bu da düzey farkları ortadan kalkmış iki ya da daha çok veri kümesini, dağılımlarının biçimi bakımından daha kolay karşılaştırabilmemizi sağlar.

Böyle bir incelemeye örnek olmak üzere illere göre kişi başına gelir dağılımlarını alalım. Az önceki 1988 ile 1989 yıllarına ilişkin beşli özetler üzerinde gereken işlemleri yaparsak Çizelge 10.9'daki gösterimle karşılaşırız. Bu beşli özetlerin kutu çizimleri yapılırsa Çizim 10.10'a ulaşılır. Okuyucu bunları kendi de yaparak sınamalı, veri dağılımlarının biçim özelliklerinin gerçekten daha iyi ortaya çıkıp çıkmadığını incelemelidir.

ÇİZELGE 10.9 İllerin kb GSYİÜ – *Ortc* beşli özetleri 1988, 1989

	100 bin TL			
	<i>Y</i>		<i>Y – Ortc</i>	
	1988	1989	1988	1989
Y_{eb}	38	36	$38 - 10 = 28$	$36 - 9 = 27$
D_u	14	14	$14 - 10 = 4$	$14 - 9 = 5$
<i>Ortc</i>	10	9	$10 - 10 = 0$	$9 - 9 = 0$
D_a	6	7	$6 - 10 = -4$	$7 - 9 = -2$
Y_{ek}	3	2	$3 - 10 = -7$	$2 - 9 = -7$

Çizelge 10.8'den düzenlenmiştir.



ÇİZİM 10.10 İllerin kb GSYİÜ – *Ortc* kutu çizimleri 1988, 1989

Çizelge 10.9'dan düzenlenmiştir.

Çizelge 10.9'dan DAr , $Adım$, eşik değerleriyle illerin dışadüşen konumlarının değişmediğini kolayca görebiliriz. Çizim 10.9 ile Çizim 10.8 karşılaştırıldığında ikisi arasındaki farkların ölçekteki değişme ile iki yılın ortalama düzeylerinin aynı hizaya gelmesi olduğu hemen anlaşılabilir.

Birkaç sözcükle yorum yapmak gerekirse şunlar söylenebilir. Kocaeli dışındaki 66 ilin Ar değeri her iki yılda da neredeyse aynıdır. Yani iki uçtaki illerin gelir düzeylerinin *Ortc* ilden uzaklıkları 1987 fiyatlarıyla TL cin-

sinden ele alınırsa 1988'den 1989'a değişmediği görülür. Benzer bir yorum kutu içinde kalan iller için yapılamaz. Kutunun boyu 1989'da bir önceki yıla göre kısalmıştır. En kaba ölçümle bile 1988'de $14 - 6 = 8$ olan *DAr*, ertesi yıl $14 - 7 = 7$ 'ye inmiştir. Bu da illerin ortadaki yarısının kendi aralarında gelir yakınlaşmaları olduğu anlamına gelir. Bu yakınlaşmaların önemli bölümünün kutunun sol yarısında gerçekleştiği de açıktır. *Ortc*'ya göre baktığımızda 1988'de $10 - 6 = 4$ olan bu fark, bir yıl sonra $9 - 7 = 2$ 'ye düşmüştür. Bu kendi içinde olumlu bir gelişmedir. İllerimizin *Ortc*'nın hemen altındaki dörtte birinin gelir düzeyi ortalamaya doğru kaymıştır. Ama kutunun sağ yarısında durum böyle bir eğilim yoktur. Burada *Ortc*'ya göre farkların pek değişmediğini görüyoruz.

Tam “bu çizime bakılarak çarpıklık da yorumlanabilir mi?” diyecektim ki “acaba yayıklık farkını da giderip bakışım özelliğini daha iyi inceleyebilir miyiz?” diye bir soru duyar gibi oldum. Bu soruyu sorduğunuza göre siz artık bu işten anlıyorsunuz demektir.

Standartlaştırma

Yayıklık farklarını ortadan kaldırmanın en kolay yolu, az önce bulduğumuz *Ortc*'ya göre farkları, bir yayıklık ölçüsüne bölmektir. Sorgulayıcı veri çözümlemesinde kullandığımız yayıklık ölçüsü dördebölenler aralığı (*DAr*) olduğuna göre farkları böleceğimiz sayıyı da bulduk demektir. Daha kesin hesaplamaların gerektiği durumlarda (çıkarsama çözümlemelerinde) bu ölçü standart sapma olabilir. O zaman *Ortc* yerine de genellikle *AO* gelir. Verilerin böyle ortalama düzeyle farklarının alınıp yayıklık ölçüsüne bölünmesine istatistikte **standartlaştırma** denir. Kısaca şöyle gösterilebilir:

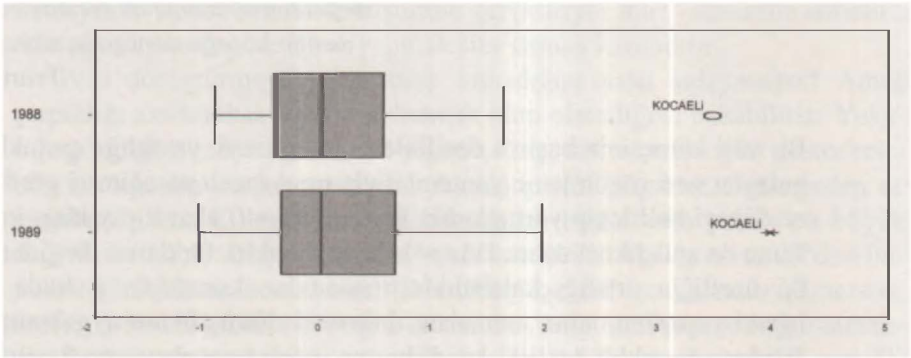
$$(Y - Ortc) / DAr \quad \text{ya da} \quad (Y - AO) / \sigma$$

İllere göre kişi başına gelir örneğimizi sürdürelim. Daha önce gözlem değerlerinin ortancayla farklarını almıştık, şimdi de bu farkları *DAr*'ye bölelim. Sonuç Çizelge 10.10 ile Çizim 10.11'deki gibidir. Okuyucu benzer işlemleri başka yılların verilerine uygulayarak bulgularını yorumlayabilir.

ÇİZELGE 10.10 İller kb GSYİÜ standartlaştırılmış beşli özetleri 1988, 1989

	Y		Y - Ortc		(Y - Ortc) / DAr	
	1988	1989	1988	1989	1988	1989
Y_{eb}	38	36	28	27	$28 / 8 = 3,5$	$27 / 7 = 3,9$
$D_{\bar{u}}$	14	14	4	5	$4 / 8 = 0,5$	$5 / 7 = 0,7$
Ortc	10	9	0	0	0	0
D_a	6	7	-4	-2	$-4 / 8 = -0,5$	$-2 / 7 = -0,3$
Y_{ek}	3	2	-7	-7	$-7 / 8 = -0,9$	$-7 / 7 = -1$

Çizelge 10.9'dan düzenlenmiştir.

**ÇİZİM 10.11** İller kb GSYİÜ standartlaştırılmış kutu çizimleri 1988, 1989

Çizelge 10.10'dan düzenlenmiştir.

Her şeyden önce şu vurgulanmalıdır ki artık yorumlarımızı yaparken 1987 fiyatlarıyla TL cinsinden konuşamayız çünkü bu birimle ölçülmüş değerleri, aynı birimle belirtilmiş DAr 'ye bölmekle ölçü birimi olmayan büyüklüklere geçmiş olduk. Dikkat ederseniz her iki yıl için de $Ortc = 0$, $DAr = 1$ çıkmaktadır. Öyleyse bu aşamada yorumlarımızı kutu uzunluğunun katları cinsinden yapabiliriz. Bu da bize hem ortalama düzeyleri, hem yayıklıkları aynı ölçülere getirilmiş iki veri kümesinin çarpıklığını karşılaştırma olanağı verir. Buna da **görelî çarpıklık** karşılaştırması diyebiliriz.

Bu resimler bize çarpıklık konusunda neler söyleyebilir? İlk göze çarpan özellik iki uç ilin de $Ortc$ 'ya görelî uzaklığının 1989'da artmış olmasıdır. Tek başına bu saptama bize zengin illerin görece daha zenginleştiği, yoksul illerin görece daha yoksullaştığını gösteriyor. Bu da 1989'da gelir dağılımında önemli bir bozulmanın göstergesi olarak yorumlanabilir. Daha önce dışadüşen ya da uzak dışadüşen olmanın mutlak değerden çok görelî

konuma bağlı olduğunu söylemiştik. Bu çizimde Kocaeli'nin iki yıldaki konumu ne demek istediğimize iyi bir örnek. Bu ilin 1989'daki konumu ilk iki çizimde 1988'e göre daha gerideydi ama görelî bir karşılaştırma yapılmınca bu çizimde öne çıktı. Böylece 1988'de dışadüşen olmasına karşın 1989'da niye uzak dışadüşen olduğu daha iyi anlaşılmaktadır.

1989'da kutu içi çarpıklığın da arttığını hemen fark ediyoruz. Demek ki gelir dağılımı yalnızca iki uçtaki iller arasında değil orta düzeye yakın iller arasında bile görelî olarak bozulmuş.

Dönüştürme

*Gregor Samsa, bir sabah tedirgin
düşlerden uyandığında kendini yatağında
dev bir böceğe dönüşmüş olarak buldu.*

Franz Kafka

Bir veri kümesinin başlıca özellikleri olan düzeyi, yayıklığı, çarpıklığı sorgulayıcı veri çözümleme yöntemleriyle nasıl inceleyeceğimizi gördük. Önce düzeyi belirleyip yorumladık. Sonra $Ortc = 0$ alarak yayıklığı irdeledik. Bunu da anladıktan sonra $DAR = 1$ deyip yayıklık farklarını da gidererek bu iki özelliğin ortalığı kalabalıklaştırmasından kurtulduk. Aslında her iki işlemi yaparken, adını anmadan, **doğrusal dönüştürme** uygulamış olduk. Böylece çarpıklık özelliği kendi başına incelenme olanağına kavuştu. Çarpıklık özelliğinden doğan karmaşayı da ortadan kaldırmak istersek veri kümesinin dağılımını bakışık kılmamız gerekir. Bunu sağlamanın yolu **üstel dönüştürmedir**. Bir veri kümesi çarpık ise üstel dönüştürme yoluyla bakışma yaklaştırılabilir. Bu, her veri kümesinin tam bakışık kılınabileceği anlamına gelmez. Ama bakışım yönünde önemli ilerlemeler sağlanabilir. Öyleyse dönüştürmenin bir amacı, çarpık bir veri kümesini bakışığa yaklaştırmak, böylece daha etkin çözümleme yapabilmektir (Siegel, 1988:141; Siegel, 1997:38-40).

Bu amaç, tek bir veri kümesi için tek bir değişken incelendiğinde anlamlıdır. Aynı değişkenin çeşitli veri kümeleri için karşılaştırılması söz konusu olduğunda yayıklıkların benzeştirilmesi başka bir amaç olarak karşımıza çıkar. Üstel dönüştürme burada da işe yarar.

Aynı veri kümesi için iki değişkenin arasındaki ilişki incelenecekse bunun doğrusallaştırılması uygun olabilir. Üstel dönüştürmeyle bu doğrusallık da sağlanabilir.

Nitel ölçekle ölçülmüş iki değişkenin çeşitli öbeklerine düşen sıklıkları iki boyutlu çizelgelerle göstermiştik. Bu çizelgelerin içerdiği örtük veri

yapısını ortaya çıkarmaya yönelik basit bir model kurulmak istendiğinde üstel dönüştürme yine yardımımıza koşar.

Öyleyse üstel dönüştürme bu dört amacı gerçekleştirmede işimize yarar (Emerson, Stoto, 1983:126; Emerson, 1983:247).

Ortalamaların ele alındığı 8. bölümde bazı dönüştürmelerden söz edilmişti. Orada gözlem değerlerinin logaritmasını, tersini, karesini almak gibi uygulamalara gidilmişti. Bu işlemler aslında birer üstel dönüştürmeden başka bir şey değildir. Ortalamalar genel gösteriminde yer alan $1/r$ kuvveti bunun bir göstergesidir. Burada aynı konu daha ayrıntılı biçimde açıklanacaktır. Yukarıda belirtilen dört amaç için neler yapılacağı sırasıyla gösterilecektir.

Önce bir uyarı yapalım. Hiç bir zaman gözü kapalı dönüştürmeye girişmeyiniz. Önce “veriler dönüştürme gerektiriyor mu?” sorusunu sorunuz. Alacağınız yanıtı göre işlem yapmak size zaman kazandırır.

Peki dönüştürmeye gerek olup olmadığını nasıl anlayacağız? Amaç çarpıklığı azaltmaksa dağılımın bakışık olup olmadığına bakabiliriz. Yukarıda gördüğümüz histogram, dal-yaprak, kutu, nokta çizimi gibi gösterimler bize bu konuda yardımcı olabilir. Ayrıca çeşitli çarpıklık ölçülerinden de yararlanabiliriz. Kolayca kullanılabilecek bir başka kaba yol da en büyük değerin en küçük değere oranına bakmak olabilir. Y_{eb}/Y_{ek} oranı 20’den büyükse dönüştürme zorunludur; 2’den küçükse bir işe yaramaz (Emerson, Stoto, 1983:125). Bizce bu oran 10’dan büyükse dönüştürme uygulanmalıdır. Bu kaba yol yalnız oran ölçeğindeki veriler için kullanılabilir, sıra ölçeğiyle ölçülmüş verilerde anlamsızdır.

İkinci bir uyarı kimi veri kümelerinde dönüştürme işleminin hangi verilere uygulanacağına ilişkindir. Sözgelimi kâr gelirden gider çıkarılarak bulunur. Dış ticaret açığı, dışsatımla dışalımın farkıdır. Böyle durumlarda fark değişkeni değil onu doğuran değişkenler dönüştürülmelidir. $Y_i - Ortc$ farkı için de aynı kural geçerlidir.

Dönüştürme işlemi, her gözlem değerine belli bir kurala göre hesaplanmış yeni bir değer vermek demektir.

Doğrusal dönüştürmelerde gözlem değerleri bir sabit sayıyla çarpılıp başka bir sabit sayıyla toplanır. Dönüşmüş değer $aY + b$ olur. Buradaki b , ölçekteki sıfır noktasına yer değiştirir. Daha önce düzey farklarından kurtulup yayıklığı incelemek amacıyla $Y - Ortc$ işlemini uygularken sıfır noktasını $Ortc$ ’ya taşıdığımızı anımsayalım. Y ’nin önündeki a katsayısı ise ölçüm aralıklarını değiştirir. Böylece ölçü birimimiz a ile çarpılmış olur. Yayıklık farklarından kurtulmak için kullandığımız $(Y - Ortc) / DAr$ dönüştürmesinde $a = 1 / DAr$ olduğunu görebiliriz. Böylece ölçü birimlerimiz DAr ’ye bölünmekte, dönüşmüş Y , artık Y ’nin kaç kg, kaç kişi, kaç m, kaç TL olduğunu değil DAr ’nin kaç katı olduğunu göstermektedir.

Üstel dönüştürmelerin genel gösterimi $aY^r + b$ biçimindedir (Emerson, Stoto, 1983:99). Ancak bu dönüştürme $r \neq 0$ iken geçerli olur, $r = 0$ iken $c \log Y + d$ kalıbı kullanılır. Bu dönüştürmelerde a, b, c, d katsayıları herhangi bir kısıtlama altına girmeden kolaylık sağlayacak biçimde seçilir. Tek kısıt $r < 0$ iken $a < 0$, $r > 0$ iken $a > 0$ olması gereğidir. Böylece gözlem değerlerinin dönüştürme öncesindeki sırası bozulmadan korunmuş olur. Üstel dönüştürmede asıl belirleyici olan r 'dir. Bu nedenle uygulamada a, b, c, d katsayılarının hepsi bire eşitlenerek dönüştürme uygulanabilir. Bu durumda genel gösterim şöyle olur:

$$r \neq 0 \text{ ise } Y^r \quad r = 0 \text{ ise } \log Y \quad r < 0 \text{ ise } -Y^r$$

Üstel dönüştürmeler

- veri kümesindeki gözlem değerlerinin sıralarını değiştirmez,
- böylece dönüştürülmemiş verilerin *Ortc*'sı ya da bölenleri dönüştürülünce, yeni dizinin *Ortc*'sı ile bölenleri elde edilir.

Üstel dönüştürme yapabilmek için önce r 'nin değerini saptamak gerekmektedir. Bu iş için önerilen bir yol "bakışım amaçlı dönüştürme çizimi" yapmaktır (Emerson, Stoto, 1983:107-110). Bu yol r 'nin değerini çok yakın bulabilmektedir, ancak uygulanması zaman alıcı bazı hesaplamalar gerektirdiğinden burada daha kolay bir yol olan **dönüştürme merdiveni**'nden yararlanacağız (Tukey, 1977;89). Bu merdiven bize hangi durumlarda hangi r değerlerini kullanarak dönüştürme yapacağımızı gösterir. Dağılımın çarpıklığı azsa Tukey merdiveninde 1'e yakın bir r seçilir, çarpıklık arttıkça seçilecek r 'nin değeri de büyür.

ÇİZELGE 10.11 Tukey dönüştürme merdiveni

Çarpıklığın yönü	r	Y^r	Dönüştürme
Sola	3	Y^3	Küp
"	2	Y^2	Kare
"	3/2	$\sqrt{Y^3}$	3/2. kuvvet
Bakışık	1	Y	Yok
Sağa	1/3	$\sqrt[3]{Y}$	Küpkök
"	1/2	$\sqrt{Y}$	Karekök
"	0	$\log Y$	Logaritma
"	$-1/2$	$-1/\sqrt{Y}$	Karekökten sonra ters
"	-1	$-1/Y$	Ters
"	-2	$-1/Y^2$	Kareden sonra ters

0'dan küçük r değerlerinde, dönüştürülen değerlerin işaretinin de değiştirilmesinin nedeni özgün verilerdeki büyüklük sırasını korumaya yöneliktir.

Herhangi bir r değeri yeterli bakışımı sağlayamıyorsa daha güçlü bir r değeri denenebilir. Ancak burada dönüştürme yapıldıktan sonra elde edilen yeni dizinin gerçek yaşamda bir karşılığı bulunmasına dikkat edilmelidir. Aşağıda örnekler verilirken bu konuya döneceğiz.

Uygulamaya uygun düşerse, r değeri dönüştürme merdiveninde daha küçük aralıklarla gösterilebilir.

Tek veri Kümesini Dönüştürme

Artık dönüştürmeyi uygulamaya geçebiliriz. Önce tek bir veri kümesini ele alalım. Bunun için kullanılacak örnek, illerimizin 2000 yılı nüfuslarıdır. Acaba illerimizin nüfus dağılımı bakışık mıdır? Bunu anlayabilmek için Y_{eb}/Y_{ek} oranına, dal-yaprak, kutu, nokta çizimi gibi gösterimlere bakabiliriz. Burada bütün verilere yer vermeden yalnızca bu göstergeleri inceleyelim.

$Y_{eb}/Y_{ek} = 1003 / 9 = 111$ hayli büyüktür, demek ki ya çok çarpık bir dağılımla karşı karşıyayız ya da en kalabalık il uzak dışadüşendir. Ayrıca Çizim 10.12'deki dal-yaprak gösterimi bize dağılımın büyük değerlere doğru hayli çarpık olduğunu göstermektedir.

(100, 40, 34, 22, 21)	
1	8
1	67
1	4
1	22222 3
1	0000
0	88888 8999
0	66666 67777 7
0	44444 45555
0	22222 22223 33333 33333 33
0	01111 11111
milyon	Yüz bin kişi

ÇİZİM 10.12 İllerimizin 2000 yılı nüfusları

DİE, 2002a:50'deki verilerden düzenlenmiştir.

Şimdi de Çizelge 10.12'deki bölenler özetini görelim. Bu özette 2'nin kuvveti olan bütün bölenler Tukey menteşeleri yoluyla hesaplanmıştır. Çizelge 10.12'nin en sol sütunu bu yöntemle bulunan konumları (hem üstten, hem alttan sıra numaralarını) gösterir. İkinci sütun çeşitli bölen adlarını içermektedir. *Ortc* tektir. *D*'ler 3, sekizebölenler 7, onaltıyabölenler 15, ... tanedir. Sağdaki üç sütunda bu bölenlerin en alttakinin değeri "alt", en üsttekinin değeri de "üst" başlığı altında yer almaktadır. Bu iki değerlerin orta noktası, bir başka deyişle ikisinin toplamının yarısı "orta" başlığı altında gösterilmiştir. Buna **bölenler ortası** diyebiliriz. 8. bölümde **dördebölenler ortası** (ya da **yarı dördebölenler ortalaması**) kavramını görmüştük. Burada "orta" başlığı altındaki değerler de benzer biçimde yorumlanmalıdır.

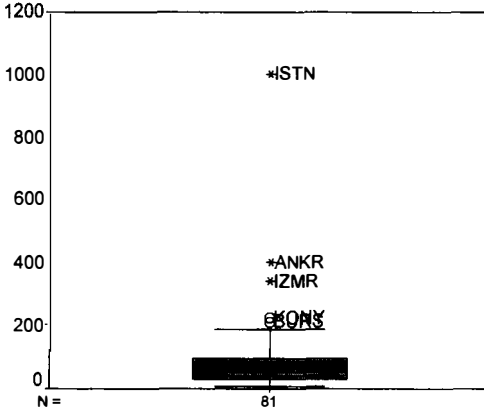
ÇİZELGE 10.12 İllerimizin 2000 yılı nüfusları için bölenler özeti (10 bin kişi)

Sıra konumu	Bölen	Üst	Orta	Alt
$(81+1)/2 = 41$	<i>Ortc</i>	—	53	—
$(41+1)/2 = 21$	<i>D</i>	96	64	31
$(21+1)/2 = 11$	<i>8'ebölen</i>	129	76	22
$(11+1)/2 = 6$	<i>16'ıyabölen</i>	167	92	17
$(6+1)/2 = 3,5$	<i>32'yebölen</i>	280	196	12
$(3+1)/2 = 2$	<i>64'ebölen</i>	401	206	10
$(2+1)/2 = 1,5$	<i>Uç değer</i>	702	356	10

DİE, 2002a:50'deki verilerden düzenlenmiştir.

Böyle bir gösterime gidilmesinin nedeni şudur: Eğer bütün bu bölenlerin orta noktaları yaklaşık aynı değeri taşıyorsa veriler bakışığa yakındır. Yok eğer bu orta değerler *Ortc*'dan gittikçe uzaklaşıyorsa dağılım çarpıktır. Uzaklaşma büyük değerlere doğruysa sağa, küçük değerlere doğruysa sola çarpıklık var demektir. Orta değerler belli bir bölene kadar *Ortc*'ya yakın kalıp daha sonra uzaklaşıyorsa çarpılma yalnızca o uç bölgede var demektir.

Çizelge 10.12'de orta noktaların büyük değerlere doğru *Ortc*'dan hemen, hem de hızla uzaklaştığı görülmektedir. Buradan da dağılımın her bölgesinde sağa çarpıklığın güçlü olduğu anlaşılmaktadır. Bunu Çizim 10.12'deki kutu gösterimi de doğrulamaktadır.



ÇİZİM 10.12 2000 sayımı il nüfusları kutu çizimi

DİE, 2002a:50'deki verilerden düzenlenmiştir.

Ele aldığımız bütün gösterimler dağılımın çarpıklığını ortaya koymuştur. Bu durumda bakışım sağlamak için bir dönüştürme yapılmasının gereği ortadadır. Dağılım büyük değerlere doğru, yani sağa çarpık olduğuna göre dönüştürme merdiveninden 1'den büyük bir r seçilmelidir.

Bunu yapmanın bir yolu verilerin beşli özetini çıkarıp bunlar üzerinde denemelere giderek uygun dönüştürmeyi bulmaktır. Örneğimizdeki dağılıma bunu uygularsak Çizelge 10.13'teki gösterimi elde ederiz. En sağdaki sütunda verilerin hepsinin birden küçük olmasını önlemek için dönüştürülmüş değerler 10'la çarpılmıştır.

Bu gösterimin en alt satırında D 'lerle $Ortc$ arasındaki farkların birbirine oranı yer almaktadır. Bu oranın bire yakın olması dağılımın *hiç olmazsa* orta bölgesinde bakışım sağlandığını gösterir. Örneğimizde bu oranın bire en yakın olduğu r değeri 0'dır, demek ki dönüştürme logaritma alınarak yapılabilir.

ÇİZELGE 10.13 Beşli özetlerle dönüştürme denemeleri

	$r = 1$	$r = \frac{1}{2}$	$r = 0$	$r = -\frac{1}{2}$
	Y	$\sqrt{Y}$	$\log Y$	$-10 / \sqrt{Y}$
Y_{eb}	1003	31,7	3,00	-0,32
$D_{\bar{u}}$	96	9,8	1,98	-1,02
$Ortc$	53	7,3	1,72	-1,37
D_a	31	5,6	1,49	-1,80
Y_{ek}	9	3,0	0,95	-3,33
$D_{\bar{u}} - Ortc$	43	2,5	0,26	0,35
$Ortc - D_a$	22	1,7	0,23	0,43
DAr bakışım oranı	2,0	1,5	1,1	0,8

DİE, 2002a:50'deki verilerden düzenlenmiştir.

Dönüştürme merdiveninden r seçmenin bir başka yolu yukarıda kullandığımız bölenler özetini, Çizelge 10.14'teki gibi, değişik r değerleriyle düzenlemek olabilir.

ÇİZELGE 10.14 Bölen özetleriyle dönüştürme denemeleri

	$r = \frac{1}{2}$			$r = 0$			$r = -\frac{1}{2}$		
	$\sqrt{Y}$	Y	$\log Y$	$\sqrt{Y}$	Y	$\log Y$	$\sqrt{Y}$	Y	$\log Y$
Bölen	Üst	Orta	Alt	Üst	Orta	Alt	Üst	Orta	Alt
$Ortc$	-	7,3	-	-	1,72	-	-	1,37	-
D	9,8	7,7	5,6	1,98	1,73	1,49	-1,02	-1,41	-1,80
$8'ebölen$	11,4	8,0	4,7	2,11	1,72	1,34	-0,88	-1,50	-2,13
$16'yabölen$	12,9	8,5	4,1	2,22	1,72	1,23	-0,77	-1,60	-2,43
$32'yebölen$	16,7	10,1	3,5	2,45	1,76	1,08	-0,60	-1,74	-2,89
$64'ebölen$	20,0	11,6	3,2	2,60	1,80	1,00	-0,50	-1,83	-3,16
$Uç değer$	26,6	18,9	3,2	2,85	1,92	1,00	-0,38	-1,77	-3,16

DİE, 2002a:50'deki verilerden düzenlenmiştir.

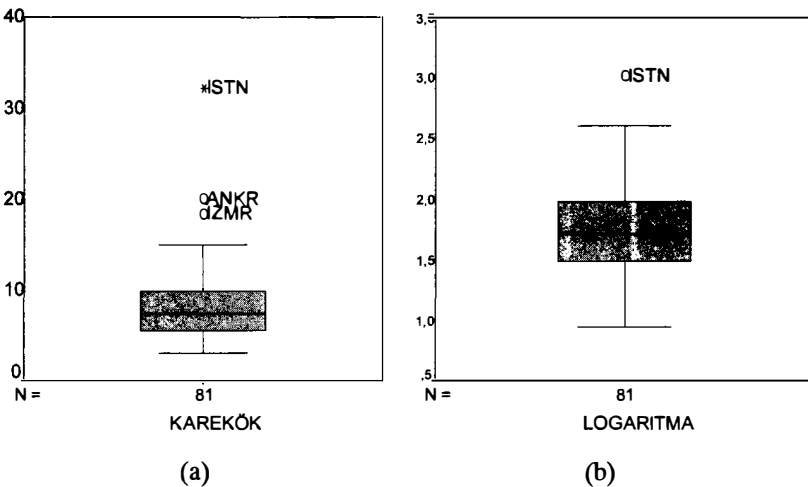
Bu çizelge hazırlanırken üst-alt sütunlarındaki değerlerin her biri üç kez dönüştürülmüş, birincide karekökleri, ikincide logaritmaları alınmış, üçüncü de ise önce karekökleri alınmış, sonra ters çevrilip onla çarpılmış, önlerine bir de eksi eklenmiştir. “Orta” başlıklı sütunlar, bu dönüştürülmüş değerlerin orta noktaları hesaplanarak doldurulmuştur.

Görüldüğü gibi özgün verilerde orta noktaların $Ortc$ 'dan uzaklaşma hızı karekök dönüşümüyle azalmış, logaritma dönüşümüyle neredeyse durdurulmuştur. Logaritma dönüştürmesi orta değerlerini $16'ya$ bölenler satırına

kadar hemen hiç değiştirmemiş ondan sonra biraz artırmıştır. Çizelge 10.12'de $16^{\text{yabölenler}}$ aralığının dışında her iki uçtan 5'er il kaldığına göre 81 ilden ortadaki 71'i için bakışım sağlanabilmektedir. Gerçekten de dönüştürülmüş gözlem değerleriyle düzenlenen Çizim 10.13'teki dal-yaprak, Çizim 10.14'teki kutu gösterimleri bu önemli gelişmeyi doğrulamaktadır.

$r = \frac{1}{2}$		$r = 0$	
	(31, 20, 18)		(3,0)
1	44	2	6
1	2233	2	5
1	0000011111	2	22233
0	88888 88999 99999 9	2	00000 00111 1
0	66666 66666 67777 7777	1	88888 88999 99999 9
0	44444 44445 55555 55555 5555	1	66666 67777 7777
0	3333	1	44444 45555 55555 55
On	Bir	1	2222223333 3
		1	001
		0	9
		Bir	Onda bir

ÇİZİM 10.13 2000 sayımı dönüştürülmüş il nüfuslarının dal-yaprakları
DİE, 2002a:50'deki verilerden düzenlenmiştir.



ÇİZİM 10.14 2000 sayımı dönüştürülmüş il nüfuslarının kutu çizimleri
DİE, 2002a:50'deki verilerden düzenlenmiştir.

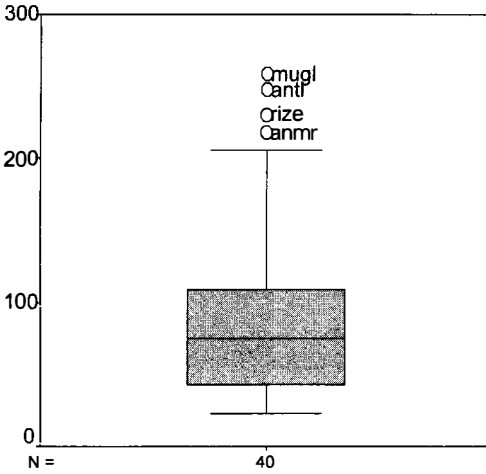
Her iki gösterimi inceleyip bakışında ortaya çıkan değişimleri yorumlamak okuyucuya bırakılmıştır. Okuyucu ayrıca $-10 / \sqrt{Y}$ ile dönüştürmüş değerler için dal-yaprak, kutu çizimi hazırlayabilir, bu dönüştürmenin çarpıklığı sola doğru çevirdiğini görüp, gereğinden güçlü bir dönüştürme olduğunu anlayabilir.

Böylece tek bir veri kümesinde bakışım sağlamak için nasıl bir dönüştürme yapılacağını görmüş olduk Şimdi yine aynı konuda bu kez daha az kullanılan bir r değeriyle, bir başka örnek inceleyelim.

Bu örneğimiz Türkiye’deki çeşitli meteoroloji istasyonlarında Ocak ayı için ölçülmüş yağış miktarları olsun. Burada da bütün verileri göstermek yerine özet göstergelerle yetinelim. $Y_{eb} / Y_{ek} = 258,8 / 23,2 = 11,1$ çıkmaktadır. Çarpıklık aşırı değildir. Çizim 10.15’te bunların dal-yaprak gösterimi, Çizim 10.16’da kutu çizimi, Çizelge 10.15’teyse bölenler özeti verilmiştir.

	258, 247, 230, 218
2	0
1	
1	6
1	4
1	23
1	001
0	88899 99
0	777
0	44444 44555
0	22223 333
1	10 mm/m ²
00	

ÇİZİM 10.15 Yağış miktarları (mm)
DİE, 2002a:24’teki verilerden düzenlenmiştir.



ÇİZİM 10.16 Yağış miktarları (mm)

DİE, 2002a:24'teki verilerden düzenlenmiştir.

Bunlardan anlaşıldığı gibi çarpıklık sağa doğrudur, dışadüşenler vardır, ama bunlar bir önceki örnektekinin tersine uzak dışadüşen değildir. Bu verilere nasıl bir dönüştürme uygulanabileceğini bulmak okuyucuya bırakılmıştır.

Okuyucu uygun dönüştürmeyi araya dursun, biz uygulanabilecek dönüştürmelerden biri üzerinde biraz konuşalım. Yanlış anlaşılmasın, aşağıda sözünü edeceğimiz dönüştürme en elverişlisi midir, değil midir bilmiyoruz. Yalnızca böyle bir dönüştürme uygulanırsa nasıl yorumlanabileceği konusuna değinmek istiyoruz.

ÇİZELGE 10.15 Yağışlar için bölenler özeti (mm)

Sıra konumu	Bölen	Üst	Orta	Alt
$(40 + 1) / 2 = 20,5$	<i>Ortc</i>		7	
		-	5	-
$(20 + 1) / 2 = 10,5$	<i>D</i>		7	
		108	6	43
$(10 + 1) / 2 = 5,5$	<i>8'ebölen</i>		1	
		185	10	36
$(5 + 1) / 2 = 3$	<i>16'yabölen</i>		1	
		230	30	29
$(3 + 1) / 2 = 2$	<i>32'yebölen</i>		1	
		247	36	24
$(2 + 1) / 2 = 1,5$	<i>Uç değer</i>		1	
		258	40	23

DİE, 2002a:24'teki verilerden düzenlenmiştir.

Diyelim ki bu verilere küpkök dönüştürmesi uygulamayı gerekli gördük. Peki ama belli bir alana düşen yağışı mm cinsinden gösteren özgün verilerin küpkökü ne anlama gelebilir? Bunların birer hacim ölçüsü olduğu açıktır. Her biri, tabanı 1 mm^2 , yüksekliği gözlem değerine eşit olan yağmur suyuyla dolu bir küpün hacmi gibi düşünülebilir. Öyleyse sözgelimi Afyon'un 43 mm olan yağış miktarına 43 mm^3 dersek, bunun küpkökü 3,5 mm olur. Demek ki bir mm^2 'ye düşen 43 mm yüksekliğindeki yağmur suyu, bir kenarı 3,5 mm olan bir küpü tam olarak doldurur. Yani küpkök almakla, bir dikdörtgenler prizmasını aynı hacimde küpe dönüştürmüş oluyoruz. Özgün veri, bu prizmanın yüksekliğini gösterirken, bunun küpkökü olan sayı aynı miktar suyla dolu küpün bir kenarının uzunluğunu verir. Böylece küpkök dönüştürmesinin ölçü birimini nasıl değiştireceğini anlamış bulunuyoruz. Logaritma kullanılarak farkların göreceli farklara dönüştürdüğünü de daha önce görmüş, böylece büyüme miktarları yerine büyüme hızlarını kullanabildiğimizi öğrenmiştik.

Birden Çok Veri Kümesini Hep Birlikte Dönüştürme

Aynı değişkenin ölçüldüğü birden çok veri kümesinin hep birlikte dönüştürülmesi gerekebilir. Çünkü çeşitli kümelerin ortalama düzeyleri arasında istatistik bakımından bir fark olup olmadığını anlamak için kullanılan *t* sınaması, **değişke çözümlemesi (varyans analizi)** gibi sınamalar, her kümenin hem normal dağılıma uyduğu, hem de değişkenliklerinin (varyanslarının) eşit olduğu varsayımına dayanır. Eğer kümelerin yayıklık

dereceleri arasında çok fark varsa bu sınamalar geçersiz olur. Demek ki bu tür yöntemlerden yararlanabilmek için veri kümelerinin bir yandan bakışık kılınması, öte yandan yayıklıklarının birbirlerine yaklaştırılması gerekir. İşte dönüştürme bu iki işlevi birden görür.

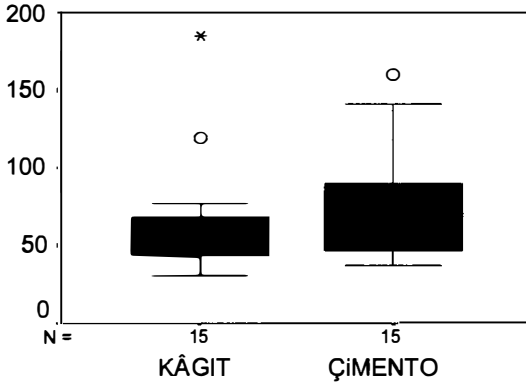
Yalnız bir uyarının yapılması gerekir. Bütün veri kümelerine Çizelge 10.11'deki dönüştürme merdiveninden seçilen tek bir r değeri uygulanmalıdır. Bir veri kümesi için logaritma kullanıp, öbürü için karekök almak kesinlikle kaçınılması gereken bir yoldur. Aynı ölçü birimleriyle ölçülmüş veri kümelerinin sayısı 2-3 taneyse bunlar için ortak bir r seçmek pek zor olmaz, bir-iki denemeyle en uygununu bulabiliriz. Ama küme sayısı arttıkça bu seçimi bir kurala bağlamak uygun olur.

Pek çok çalışma alanında kullanılan veri kümelerinin ortalama düzeyleriyle yayıklıkları arasında aynı yönlü bir ilişkiden söz edilebilir. Başka bir deyişle ortalama düzey yükselirken değişkenlik de artar. Sözelimi çeşitli gelir düzeylerindeki ailelerin tüketim harcamalarını ele alalım. Bunlardan düşük gelirli olanların tüketim harcaması tutarları birbirine yakındır, pek oynaklık göstermez. Ama gelir düzeyi yükseldikçe benzer gelirli ailelerin harcama tutarlarındaki farklar da artar.

Önce iki küme için bir örnek verelim, sonra çok kümeli örneklere geçebiliriz. 2002 yılında iki ayrı işkolunda en büyük 500 sanayi kuruluşu arasına girenlerin üretimden satış tutarları Çizelge 10.16'da gösterilmiştir. Bu verilerin kutu çizimleri Çizim 10.16'da yer almaktadır. Çizimden anlaşıldığı gibi her iki işkolundaki dağılımlar da yüksek değerlere doğru hayli çarpıktır. Ayrıca her iki işkolunda da dışadüşenler göze çarpmaktadır. Şimdi her iki işkolu için ortak bir dönüştürme bulmaya çalışalım. D_a ile $Ortc$ arasındaki farkın $Ortc$ ile D_a arasındaki farka oranı bize bakışım konusunda bir ipucu verebilir. Bu oranın 1'e yakın olması bakışımın en azından orta bölgede sağlandığını gösterir.

ÇİZELGE 10.16 2000 yılında iki işkolunda en büyük 500 sanayi kuruluşu arasına girenlerin üretimden net satış tutarları (milyar TL)

Kâğıt			Çimento			
185	64	44	160	85	49	34
119	61	42	141	72	44	33
77	53	37	112	68	43	32
69	50	32	94	63	38	32
67	45	30	85	52	37	



ÇİZİM 10.16 İki işkolunda en büyük kuruluşların üretimden net satış tutarları (milyar TL)

Bu amaçla verilerin beşli özetlerini çıkarıp bunlar üzerinde Çizelge 10.17'deki gibi çeşitli dönüştürmeleri deneyelim.

ÇİZELGE 10.17 İki işkolunun satışlarını dönüştürme denemeleri

	$r = 1$ Y		$r = 0$ $\log Y$		$r = -1$ $100/Y$	
	Kâğıt	Çimento	Kâğıt	Çimento	Kâğıt	Çimento
Y_{eb}	185	160	2,27	2,20	0,54	0,62
$D_{\bar{u}}$	69	85	1,84	1,93	1,45	1,18
Ortc	53	52	1,72	1,72	1,89	1,92
D_a	42	37	1,62	1,57	2,38	2,70
Y_{ek}	30	32	1,48	1,51	3,33	3,12
$D_{ortası}$	56	61	1,73	1,75	1,90	1,96
$8'ebölen\ ortası$	66	72	1,76	1,78	1,99	1,96
$16'yabölen\ ortası$	92	86	1,63	1,83	1,96	1,92
$Uç\ değerler\ ortası$	108	96	1,87	1,86	1,93	

Denenen her iki dönüştürme sonucunda bulunan değerlerle Çizim 10.17 (a) ile (b) düzenlenmiştir. Hem çizelge, hem çizim bize ters alma işleminin daha uygun bir dönüştürme olduğunu anlatmaktadır. Tek sorun tersi alınmış satış tutarlarının nasıl yorumlanacağıdır. Bunu şöyle açıklayalım. Tersi alınmış bir değer, 100 milyar / özgün satış değeridir. Sözgelimi kâğıt işkolunun en büyük kuruluşunun satış tutarının tersi alınınca $100/185 = 0,54$ bulunmuştur. Bunun anlamı 100 milyarlık bir satış bu kuruluşun satışının

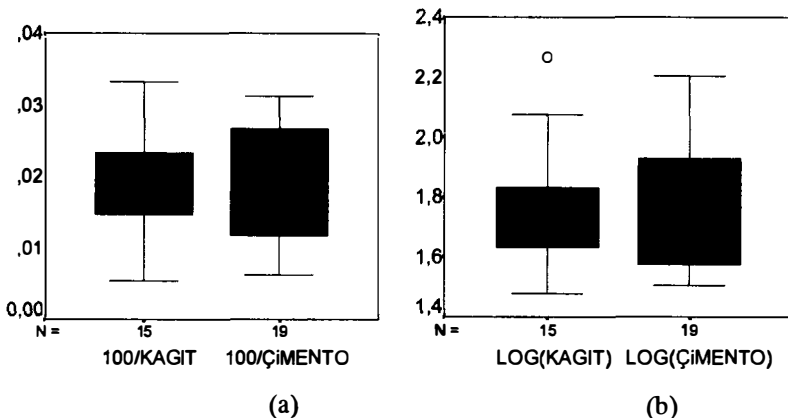
%54'ü kadardır. Benzer biçimde çimento işkolunda *Ortc* kuruluşun satışının tersi olan 1,92, bu kuruluşun satışının 1,92 katının 100 milyar TL olduğu anlamına gelir.

Böylece iki kümeli verilere dönüştürme uygulama örneğimizi tamamladık. Şimdi çoklu kümelere geçebiliriz. Yukarıda “Pek çok çalışma alanında kullanılan veri kümelerinin ortalama düzeyleriyle yayıklıkları arasında aynı yönlü bir ilişkiden söz edilebilir. Başka bir deyişle ortalama düzey yükselirken değişkenlik de artar” demiştik.

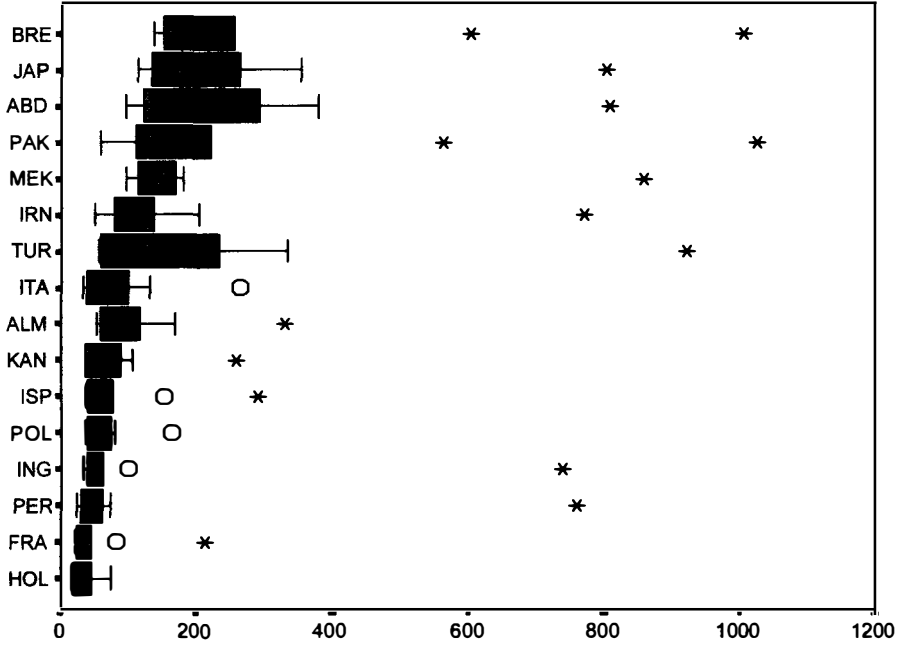
Verilerimizde bu özelliğin bulunup bulunmadığını anlamak için ne yapabiliriz? Önce ortalama düzeyi *Ortc*, yayılımı *DAr* ile gösterelim. Bu ikisi kullanılabilecek en yalın göstergelerdir. Yatay eksene *Ortc*, dikey eksene *DAr* konarak, her kümenin bu iki değeri bir serpilme çiziminde birer noktayla işaretlenirse, söz konusu ilişki gözlenebilir. Eğer noktalar gelişigüzel serpilmişlerse ilişki yok demektir. Diyelim ki düzeyle birlikte yayılım da artmakta; üstelik serpilme noktaları yaklaşık bir doğruyla temsil edilebilmektedir. İşte bu doğrunun eğiminden 1 çıkarılınca bulunan sayı bizim aradığımız *r* değeridir (Emerson, Strenio, 1983:79).

Bunu bir örnek üzerinde açıklayalım. 16 ülkenin 2002’de en büyük 10 kentinin nüfuslarını ele alalım. Elimizde 16 veri kümesi olur. Bu veri kümelerini Çizim 10.18’deki kutu çizimiyle gösterelim. Çizimden görüldüğü gibi 17 ülke içinde dev kentleri olan Brezilya, Pakistan gibi ülkelerin yanı sıra Hollanda gibi bunun tam tersi ülkeler de vardır.

Çizimde ülkeler sıralanırken kent nüfuslarının *Ortc* düzeyi temel alınmıştır. Dik eksende yukarıya doğru gidildikçe ortalama düzeyle birlikte yayıklık da genel olarak artmaktadır. Hemen her ülkede dağılımın çarpık olduğu, bir çoğunda uzak dışadüşenlerin bulunduğu göze çarpmaktadır.



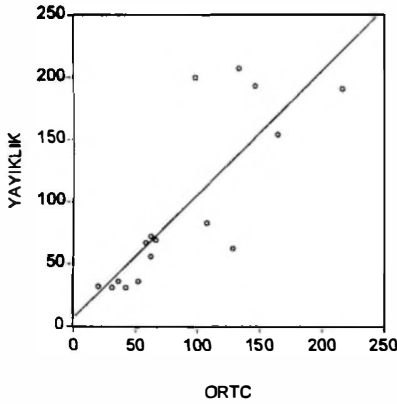
ÇİZİM 10.17 İki işkolunun satışlarını dönüştürme denemeleri



ÇİZİM 10.18 16 ülkedeki en kalabalık 10 kent nüfusunun kutu çizimi

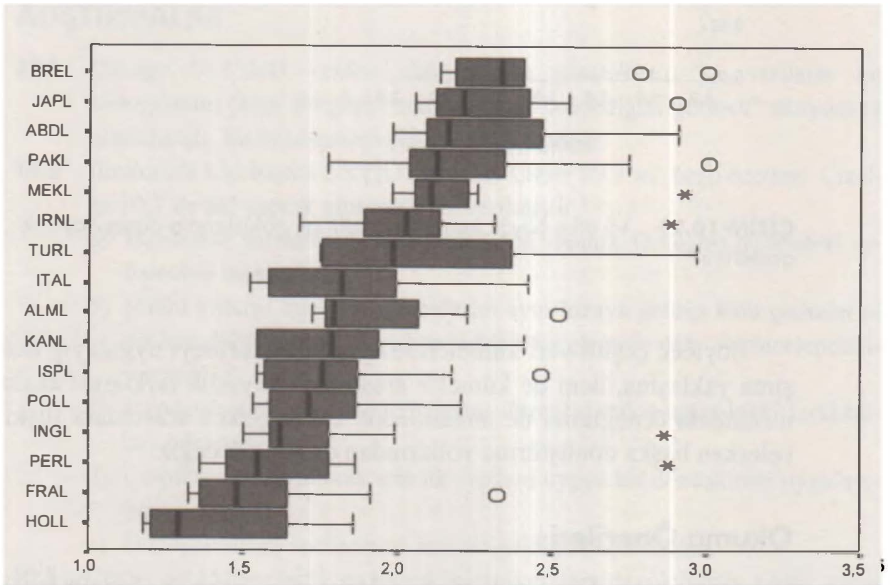
www.mongabay.com/igapo adresindeki verilerden düzenlenmiştir.

Şimdi yapmamız gereken şudur. Öyle bir dönüştürme uygulayalım ki hem her ülkedeki çarpıklık, hem de ülkeler arasındaki yayılım farklılıkları azalsın. Bu amaçla her ülke için *Ortc* ile *D_{Ar}* değerlerini bulup, bunları Çizim 10.19'daki gibi bir serpilme çizimine yerleştirelim. Bu çizimde düzey ile yayıllık arasındaki ilişki bir doğruyla temsil edilebilecek gibi görünmektedir. Bu doğruyu göz kararıyla bile çizsek eğiminin 1'den uzak olmadığını söyleyebiliriz.



ÇİZİM 10.19 16 ülkenin en kalabalık 10 kentinde nüfusların düzey-yayılım gösterimi

Çizim 10.19'u veren EViews yazılımı bu doğrunun eğimini 0,995 bulmuştur. Bu eğimden 1 çıkardığımızda $r = -0,005$ elde ederiz. Bunun sıfıra çok yakın olması logaritmali dönüştürmenin uygun olduğunu gösterir. Nüfusların logaritmaları alınarak Çizim 10.19'daki kutu gösterimine ulaşılır.

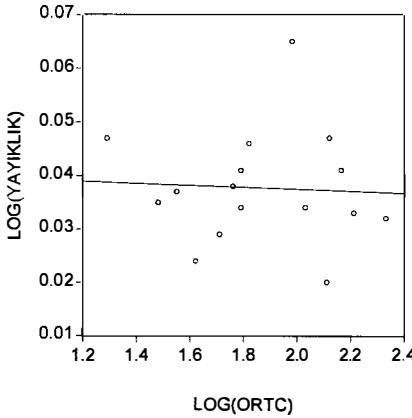


ÇİZİM 10.20 16 Ülkedeki büyük kentlerin logaritmali nüfusları

www.mongabay.com/igapo adresindeki verilerden düzenlenmiştir.

Bu kutu çizimi bize çarpıklıkların, tam olmasa bile, büyük ölçüde azaldığını, ayrıca ülkeler arasındaki yayılım farklarının da hayli küçüldüğünü göstermektedir. Burada ilginç bir gözlem de yayılımı en geniş ülkelerden birinin Türkiye olmasıdır. Çizim 10.18’de Türkiye, ortalama düzeyi kendisine yakın ülkeler arasında hayli geniş bir yayılma aralığı sergilemekteydi. Bu aykırı konum dönüştürme sonucunda iyice su yüzüne çıkmış oldu.

Logaritması alınmış verileri Çizim 10.21’de, Çizim 10.19’dakine benzer biçimde gösterdiğimizde artık ortalama düzeyle yayılım arasında herhangi bir örüntü kalmadığını, bütün noktaların iki eksen arasında rassal dağıldığını gözler önüne serebiliriz. Bu da yaptığımız dönüştürmenin başarılı olduğunun bir göstergesidir.



ÇİZİM 10.21 16 ülke büyük kentler logaritmali nüfuslarının düzey-yayıklık gösterimi

Böylece çeşitli veri kümelerine aynı dönüştürmeyi uygulayıp hem bakışma yaklaşma, hem de kümeler arasındaki yayıklık farklarını azaltma konusundaki örneğimizi de tamamladık. İki değişken arasındaki ilişkileri inceleyen başka dönüştürme yollarından da söz edeceğiz.

Okuma Önerileri

Chambers, J. M., Cleveland, W. S., Kleiner, B., Tukey, P. A. (1983) *Graphical Methods for Data Analysis*, Wadsworth & Brookes (9-73).

Cleveland, W. S. (1985) *The Elements of Graphing Data*, Wadsworth (103-143).

- Chatfield, C. (1988) *Problem Solving: A Statistician's Guide*, Chapman and Hill (22-47, 93-116).
- Devore, J, Peck, R. (2001) *The Exploration and Analysis of Data*, 4th ed., Duxbury (41-125).
- du Toit, S. H. C., Steyn, A. G. W., Stumpf, R. H. (1986) *Graphical Exploratory Data Analysis*, Springer-Verlag (13-35).
- Erickson, B. H., Nosanchuk, T. A. (1992) *Understanding Data*, 2nd. ed. Open University Press (1-113).
- Hartwig, F., Dearing, B. E. (1979) *Exploratory Data Analysis*, Sage (9-30).
- Hoaglin, D. C., Mosteller, F., Tukey, J. W. (1983) *Understanding Robust and Exploratory Data Analysis*, Wiley (1-127).
- Marsh, C. (1988) *Exploring Data*, Polity Press (1-60, 97-122).
- McNeill, D. R. (1977) *Interactive Data Analysis*, Wiley (1-46).
- Siegel, A. F. (1988) *Statistics and Data Analysis*, Wiley (17-55, 127-170).
- Siegel, A. F. (1997) *Practical Business Statistics*, Irwin (32-40).
- Tukey, J. W. (1977) *Exploratory Data Analysis*, Addison-Wesley (1-123).
- Velleman, P. F., Hoaglin, D. C. (1981) *Applications, Basics, and Computing of Exploratory Data Analysis*, Duxbury Press (1-81).

ALİŞTIRMALAR

- 10.1** Çizelge 10.1'deki verileri dal-yaprakla gösterirken “bu verilerin bir histogramını çizip, iki gösterimin birbirine benzediğini görmek” okuyucuya bırakılmıştı. Bu benzeşmeyi çizimlerle gösteriniz.
- 10.2** İllerimizin kişi başına GSYİÜ değerleri Çizim 10.5'te, beşli özetleri Çizelge 10.7'de dal-yaprak gösteriminde verilmiştir.
- Bunlardan yararlanarak kutu çizimlerini yapınız. Ortalama düzeydeki gelişmeleri inceleyiniz.
 - Bütün yılların ortalama düzeylerini aynı hizaya getirip kutu çizimlerini bir kez daha yapınız. Yayılım farkları bu çizimde daha iyi incelenebiliyor mu?
 - Standartlaştırılmış kutu çizimlerini düzenleyerek çarpıklıktaki farklılıkları irdeleyiniz.
 - Çarpıklıktan kurtulmak için ilk verilere uygun bir dönüştürme uygulayınız.
 - Dönüştürülmüş verilerin ne anlama geldiğini açıklayınız.
- 10.3** Çizim 10.15'te çeşitli meteoroloji istasyonlarında ölçülmüş Ocak yağış miktarları (mm) verilmiştir. Bu verileri bakışık kılmak amacıyla uygun bir dönüştürme uygulayınız.

- 10.4** Çizelge 7.5'teki araba satışı verileri için biri 2001, öbürü 2002'ye ilişkin iki beşli özet düzenleyiniz. Bunu yaparken Alıştırma 7.10'da bulduğunuz dal-yaprak gösterimlerinden yararlanabilirsiniz.
- İki yıl için birer kutu çizimi hazırlayınız. Bunlara bakarak çarpıklıklardan nasıl kurtulabileceğinizi belirtiniz.
 - Gerekli dönüştürmenin ne olabileceğini bulunuz. Bunu uygulayarak dağılımları çarpıklıktan kurtarınız.
 - Dönüştürülmüş verilerin ne anlama geldiğini açıklayınız
- 10.5** Çizelge 7.8 ya da 7.0'daki verileri kullanarak her yıl için birer beşli özet bulunuz. Bunun için Alıştırma 7.21'de çizdiğiniz dal-yaprak gösterimlerini kullanabilirsiniz.
- Beş yıl için birer kutu çizimi hazırlayınız. Bu gösterimdeki çarpıklıkları tartışınız.
 - Uygun bir dönüştürmeyle çarpıklıkların ne kadarı giderilebilir? Gösteriniz.
 - Dönüştürülmüş verilerin ne anlama geldiğini açıklayınız.
- 10.6** Alıştırma 7.12'deki savaş karşıtları verilerinin dal-yaprak, beşli özet, kutu çizimi gösterimlerini düzenleyiniz. Çarpıklığı düzeltmek için bir dönüştürme uygulayabilir misiniz?
- 10.7** Alıştırma 7.40'taki sağlık harcamaları için yaptığınız dal-yaprak gösterimini tepe sayısı bakımından inceleyiniz. Gördüklerinizi açıklayınız. Bu ülkeler iki ayrı öbek oluşturabilir mi? Her bir öbeğe giren ülkelerin ortak bir özelliğini söyleyebilir misiniz? Türkiye bu öbeklerden hangisindedir?
- 10.8** Alıştırma 7.40'taki savunma harcamaları için düzenlediğiniz dal-yaprak gösterimini kullanarak beşli özeti bulunuz, kutu çizimini düzenleyiniz. Dönüştürme gerekiyorsa uygulayınız.
- 10.9** Alıştırma 3.12'de ya da Çizelge 8.1'deki bebek ölüm oranlarıyla çeşitli göstergeler hesaplamıştık.
- Bu oranların dağılımındaki çarpıklığı gidermek için uygun bir dönüştürme uygulayınız. Dönüştürülmüş değerleri nasıl yorumlayabiliriz?
 - Türkiye dışındaki 15 ülkenin bebek ölüm oranlarındaki dağılımı bakışık kılabilmek için uygun bir dönüştürme uygulayınız. 16 ülke için bulduğunuz dönüştürme yöntemi ile bunu karşılaştırıp yorumlayınız.
- 10.10** Türkiye nüfusunun Çizelge 8.2'de verilen yaş dağılımını bakışık kılmak için uygun bir dönüştürme bulup uygulayınız. Sonucu yorumlayınız.
- 10.11** Alıştırma 9.5'te verilen televizyon tüpü ömürlerinin dağılımını çarpıklıktan kurtarınız. Bunu yaparken tüp ömrü aralıklarının orta noktalarını kullanınız. Dönüştürdüğünüz değerleri nasıl yorumluyorsunuz?
- 10.12** Alıştırma 9.11'deki alışveriş kuyruğunda bekleme sürelerinin dağılımını bakışık kılmak için dönüştürmeyi bulup uygulayınız. Sonucu yorumlayınız.

- 10.13** Alıştırma 9.12’de iki bölgede deneklerin TV seyretme sürelerine ilişkin veriler gösterilmişti. İki dağılıma aynı uygun dönüştürmeyi uygulayarak bakışimlarını sağlayabilir misiniz?
- 10.14** Alıştırma 9.17’de iki tür yatırım fonunun % getirileri verilmişti. Bunlar için ortak bir dönüştürme uygulayarak çarpıklıklarını gideriniz. Dönüştürülmüş değerler ne anlama gelir? Açıklayınız.
- 10.15** 2002-2003 Futbol süper liginde dönem sonu durumu şöyledir.

	Takım	O	G	B	M	A	Y	P
1	Beşiktaş	34	26	7	1	63	21	85
2	Galatasaray	34	24	5	5	61	27	77
3	G.Birliği	34	19	9	6	76	40	66
4	Gaziantep	34	16	9	9	61	41	57
5	Malatyaspor	34	14	10	10	56	45	52
6	Fenerbahçe	34	13	12	9	55	42	51
7	Trabzonspor	34	13	12	9	44	33	51
8	Ankaragücü	34	15	4	15	44	42	49
9	İstanbulspor	34	12	7	15	42	47	43
10	Denizlispor	34	10	10	14	37	42	40
11	Adanaspor	34	10	10	14	44	54	40
12	Samsunspor	34	10	9	15	41	59	39
13	Elazığspor	34	10	7	17	40	59	37
14	Diyarbakırspor	34	9	9	16	34	47	36
15	Bursaspor	34	9	9	16	42	61	36
16	Altay	34	9	8	17	48	69	35
17	Göztepe	34	5	11	18	32	57	26
18	Kocaelispor	34	6	4	24	32	66	22

O: Oyun, G: Galibiyet, B: Beraberlik, M: Mağlubiyet, A: Attığı gol, Y: Yediği gol, P: Puan.
Radikal, 4 Haziran 2003 (Bir baskı yanlış düzeltilmiştir).

- Galibiyet (G) sayılarının bakışık olup olmadığını, dışadüşen içerip içermediğini inceleyiniz. Bu verilerin bakışık ya da çarpık olması ne anlama gelir?
- Gerek görüyorsanız uygun bir dönüştürmeyle daha bakışık bir dağılım bulunuz. Dönüştürmenizi bir futbol tutkununa açıklayınız. Bulgularınızı yorumlayınız. Dışadüşen varsa dönüştürmenin bunları nasıl etkilediğini inceleyiniz.
- Yukarıdaki soruları B, M, A, Y, P için de yanıtlayınız.
- Beraberlikle bitmeyen karşılaşmalarda bir takım yendiği, öbür takım yenildiğine göre G ile M’nin toplamı, dolayısıyla AO’ları da aynı olma-

lıdır. Peki bu değişkenlerin *ESD*'leri, *Ortc*'ları birbirine eşit midir? Hesaplayıp bakınız. Bunu nasıl açıklarsınız? Bu iki dağılım başka bakımlardan da farklı mıdır?

- e) Öte yandan $AO_G + AO_B + AO_M = AO_O$ olmalıdır. Neden? Aynı değişkenlerin *ESD*'leri, *Ortc*'ları arasında da böyle ilişkiler var mıdır? Bulduğunuz sonucun nedenini açıklayınız.
- f) $AO_A = AO_Y$, eşitliği için de aynı hesaplamaları yapınız. Bulduklarınızı açıklayınız.
- g) Bir karşılaşmada yenen takım 3, berabere kalan 1 puan aldığına göre $3AO_G + AO_B = AO_P$ olması gerekir. Acaba bu değişkenlerin *ESD*'leri, *Ortc*'ları arasında da böyle eşitlikler geçerli midir? Neden? Hesaplamalarınızı yapıp gerekçenizi açıklayınız.

10.16 Alıştırma 9.47'de Avrupa futbol liglerinden üçünün 2002-2003 dönemi sonunda puan durumları verilmişti. Aşağıda bular dal-yaprakla gösterilmiştir. En yüksek-en düşük puanlı takımlar Almanya'da Bayern Münih ile E. Cottbus, İtalya'da Juventus ile Torino'dur.

Almanya		İtalya		Türkiye	
8		8		8	5
7	5	7	2	7	7
6		6	015	6	6
5	24689	5	566	5	1127
4	00123 5569	4	1229	4	0039
3	006	3	08888	3	56679
2		2	14	2	26
o	Bir	on	bir	on	bir
n					

Radikal, 25 Mayıs 2003, 4 Haziran 2003.

- a) Sizce hangi lig genel olarak en çekişmeli geçmiştir? Bunu anlamak için hangi ölçütlere baktınız?
- b) Başa gürşen takımlardan en rahat şampiyon olanı hangi ligdedir? En zor birinci geleni hangi ligdedir? Bunları hangi ölçütlerle anlıyorsunuz?
- c) Hangi ligin dibinde takımlar en çok çekişme yaşamıştır?
- d) Bu liglerde dışa düşen takımlar var mı? Varsa hangileri?
- e) Bu puan dağılımlarını daha bakışık kılmak için nasıl bir dönüştürme uygulanabilir? Dönüştürme sonrasındaki değerler nasıl yorumlanabilir?

10.17 Yine Alıştırma 9.47'de verilen gol sayıları için benzeri bir çözümleme yapınız. Varsa çarpıklıktan kurtulup sonucu yorumlayınız.

10.18 Avrupa futbol liglerinden 20 takımlı olan ikisinde 2002-2003 döneminde atılan-yenilen goller aşağıdaki gibidir. Fransa'da Monaco 66, Lyon 63,

Troyes 23, İngiltere’de Manchester United 74 gol atmış, Fransa’da Sedan 59, Auxerre 29, İngiltere’de Weat Bromwich 65, Sunderland 65, Manchester United 34 gol yemiştir.

- Sorgulayıcı veri çözümleme araçlarıyla iki ülkeyi düzey, yayılım, biçim, dışadüşenler bakımından karşılaştırınız.
- Sizce bu dağılımlardan herhangi biri dönüştürme gerektirir mi? Gerekirse uygulayıp sonucu yorumlayınız. Gerekmezse nedenini açıklayınız.

Fransa			İngiltere		
Atılan gol		Yenilen gol	Atılan gol		Yenilen gol
	7		4	7	
63	6		8531	6	255
97	5	49	821	5	01467 9
76 31100	4	14456 7889	88753 22111	4	12346 7899
98775	3	11136 669		3	48
9973	2	9	91	2	
Bir	on	Bir	bir	on	bir

Radikal, 13 Mayıs 2003, 27 Mayıs 2003.

10.19 İstanbul Sanayi Odası’nın her yıl yayımladığı Türkiye’nin en büyük sanayi kuruluşlarından 2001 yılında ilk 50’ye girenlerin üretimden satışları (trilyon TL) aşağıda verilmiştir.

- Sorgulayıcı veri çözümleme araçlarıyla bu dağılımı ortalama düzey, yayılım, biçim, dışadüşenler bakımından inceleyiniz.
- Bakışım sağlamak için uygun bir dönüştürme uygulayınız. Dönüştürmeyle dışadüşenlere ne olduğunu inceleyiniz.
- Dönüştürmeyle elde ettiğiniz değerlerin büyüklüklerini birbirleriyle nasıl karşılaştırırsınız? Açıklayınız.

5294,74	917,04	514,22	377,47	314,82	268,80	235,24	215,39	194,75
3436,15	865,06	492,02	349,16	301,62	261,12	--	212,87	194,32
1072,83	845,75	460,17	347,37	301,12	260,86	226,67	212,19	
1015,81	818,78	408,43	345,10	300,21	256,00	223,15	206,56	
946,20	776,30	404,03	336,45	283,47	255,02	221,42	197,30	
921,49	754,54	391,73	324,27	273,32	237,17	216,52	196,52	

İSO, 2002:98-9 (38. sıradaki kuruluş verilerinin yayımlanmasına izin vermemiştir).

10.20 Rassal örneklemelerle bir bölgede kırsal kesimden düşük gelirli 50, kentsel kesimden yüksek gelirli 40 kişinin kolesterol düzeyleri (mg/l) aşağıdaki gibi ölçülmüştür.

- İki öbeği düzey, yayılım, biçim açılarından karşılaştırınız.

b) Verilerde herhangi bir çarpıklık göze çarpıyor mu? Ne gördüğünüzü açıklayınız.

Kırsal kesim düşük gelirli					Kentsel kesim yüksek gelirli				
135	158	172	95	204	284	201	155	279	201
231	226	146	144	114	184	179	134	284	199
129	124	189	220	175	200	222	234	244	279
108	108	140	115	139	234	200	205	222	227
162	129	180	192	131	189	234	242	252	241
223	174	171	155	152	170	189	236	206	190
148	144	136	142	143	214	170	196	101	227
152	197	165	136	145	205	205	204	197	188
143	157	158	173	181					
142	131	166	165	194					

Devore, Peck, 1993:19-20.

10.21 Türkiye’de iki ayrı yılda 41 merkezde ölçülen kükürtdioksit (SO_2) oranları ($\mu\text{g}/\text{m}^3$) aşağıdaki gibi bulunmuştur.

- a) İki yılın verilerini sorgulayıcı veri çözümleme araçlarıyla düzey, yayılım, biçim bakımlarından karşılaştırınız. Kirlenme ortalama olarak artmış mı? Yayılımdaki değişme ne anlama gelir?
- b) Çarpıklığı gidermek için dönüştürme yapmak sizce gerekli mi? Neden?

1996					2000						
27	47	110	111	76	151	33	89	52	185	36	87
83	66	26	48	18		71	69	52	52	21	
46	56	60	54	91		47	14	43	56	73	
40	55	102	61	92		49	62	92	34	51	
34	102	52	55	97		17	95	45	91	105	
104	70	97	42	90		69	96	77	68	100	
40	67	83	33	70		37	105	76	49	71	
106	76	97	61	53		74	85	43	37	53	

DiE, 2002a:34.

10.22 Aşağıda 20 ülkenin 1975, 1995 nüfuslarıyla 2015 için yapılan nüfus tahminleri gösterilmiştir.

- a) Üç yıla ilişkin verileri sorgulayıcı veri çözümleme araçlarıyla düzey, yayılım, biçim bakımlarından karşılaştırınız. Ortalama düzey ne kadar yükselmiştir? Yayılımda bir değişme olmuş mudur?

b) Üç yılın verileri tek bir dönüştürmeyle bakışık kılınabilir mi?

	1975	1995	2015		1975	1995	2015
ABD	220,2	263,1	329,7	İsviçre	6,3	7,0	7,0
Arnavutluk	2,4	3,3	3,4	İtalya	55,4	57,2	55,5
Avusturya	7,6	8,1	8,1	İrlanda	3,2	3,6	4,4
Belçika	9,8	10,1	10,5	Kanada	23,1	29,6	34,1
Bulgaristan	8,7	8,4	7,2	Lüksemburg	0,4	0,4	0,5
Finlandiya	4,7	5,1	5,3	Macaristan	10,5	10,2	9,3
Fransa	52,7	58,1	62,8	Norveç	4,0	4,4	4,7
Hollanda	13,7	15,5	16,8	Portekiz	9,1	9,9	10,0
İspanya	35,6	39,2	41,2	Romanya	21,2	22,7	21,6
İsveç	8,2	8,8	9,0	Türkiye	41,0	61,1	82,1

World Bank 1997:214-215; UNDP, 2003:250-251.

10.23 100'ün üzerinde öğretim üyesi olan 52 üniversitenin öğretim üyesi başına düşen Bilimsel Gönderme İndeksi (SCI)'nde yer alan bilimsel yayın sayıları aşağıdaki dal-yaprak gösteriminde verilmiştir.

a) Beşli özeti çıkarıp kutu çizimini yapınız. Gördüklerinizi düzey, yayılım, biçim bakımlarından karşılaştırınız. Dışadüşen varmı? Varsa bu ne anlama gelmektedir?

b) Çarpıklığı gidermek için uygun bir dönüştürme uygulayınız.

3	579
5	02
5	
7	13
9	88
10	3
14	5688
19	01344
(8)	56677
	779
25	12222
	334
17	55777
	7889
8	02234
3	667
	0,01
,1	

10.24 72 ülkenin 2001-2002 yıllarındaki bilimsel yayın sayıları, çarpıklığı çok büyük dağılımlar için kullanılan özel dal-yapraklarla aşağıda gösterilmiştir.

	2001 (247.182)		2002 (244.258)	
$>10^5$	(	7 2	7	1
		6 69	6	47
$10^4 -$		5	5	
10^5 {	4	7	4	5
arası		3 22	3	22
		2 025	2	135
	)	1 00113 5578	1	00123 4688
	(	9 3	9	3
		8	8	03
		7 457	7	23
		6 23	6	1
$<10^4$ {	5	004	5	35
		4 022245	4	01467 9
		3 56	3	799
		2 046	2	0357
		1 00002 33355 99	1	03345 689
	)	0 03333 44455 55555	0	13333 44455 55566
		66666 79		67779 999

Balcı (2003)

- Ülkelerin iki yıldaki yayın sayılarını düzey, yayılım, biçim bakımından karşılaştırınız. Dışadüşen varsa yorumlayınız.
- İki yılın verilerine uyan ortak bir dönüştürme uygulayınız.

10.25 Birleşmiş Milletler Kalkınma Örgütü (UNDP) ülkelerin İnsani Gelişme İndeksi'ni hesaplamaktadır. 175 ülkenin 0 ile 1 arasında sıralanan 2001 yılı indeks değerlerinden aşağıdaki dal-yaprak gösterimi elde edilmiştir.

9	01112 22333 33333 33334 44
8	56778 88899 9
8	00000 11112 22333 33334 44
7	55555 55556 66667 77777 77777
	88899 9
7	00112 22223 33333 34444 44
6	55666 77888 8
6	01233 44
5	5669
5	00001 12244 4
4	56666 67889 999
4	01223 4
3	55667 77889
3	333
2	79
0	0,01
,1	

UNDP, 2003:237-240.

- Bu dağılımı düzey, yayılım, biçim, tepe sayısı bakımlarından inceleyiniz.
- Birden çok tepe sayısı burada ne anlama gelir? Açıklayınız.
- UNDP, ülkeleri İGİ değerlerine göre üçe ayırmakta, 0,800'ün üstündeki ülkeleri insani bakımdan iyice gelişmiş, 0,500-0,799 arasındakileri orta düzeyde, 0,500'ün altındakileri ise düşük düzeyde gelişmiş saymaktadır. Yukarıdaki verilere göre, bu ayırım noktalarının uygun seçildiği söylenebilir mi? Neden?
- Bu dağılımı bir dönüştürmeyle bakışma yaklaştırmak uygun olur mu? Gerekçenizi açıklayınız.

10.26 İGİ değerleri 1992’de 174 ülke için aşağıdaki gibiydi. Bunları Alıştırma 10.25’te verilen 2001 indeksleriyle düzey, yayılım, biçim bakımlarından karşılaştırınız. Bu süre içinde ortaya çıkan değişimleri irdeleyiniz.

	5
	00000 11112 22222 23333
	3333
	55555 66666 67777 77888
	88889 9
	01222 23344 44
	55666 66666 67788 999
	00000 00111 22333 33
	5779
	01111 234
	55777 899
	00123 3334
	55788 8
	00122 233
	56666 6789
	02233 33344
	89999
	02222 2344
	0,01
,1	

UNDP, 1995:155-157.

Aralarında İlişki Var mı?

Bir elin nesi var, iki elin sesi var.
Atasözü

11.1 İKİ BOYUTLU ÇİZELGELER

Buraya kadar genellikle tek değişkenli veri kümelerini inceledik. Yalnız 7. bölümde bazı çizelgeleri ya da çizimleri yaparken iki değişkenli verileri kullandık, son iki bölümde de kimi veri kümelerini karşılaştırırken iki değişkenli örneklerle başvurduk. Sözelimi Çize lge 7.1’de cinsiyet ile eğitim düzeyi, Çizelge 7.2’de gelir ile tüketim harcamaları arasındaki ilişkiler ele alınmıştı. İlkinde cinsiyet değişkeni ad ölçeğinde, eğitim düzeyi ise sıra ölçeğinde ölçülmüştü. İkinci çizelgedeyse hem gelir, hem tüketim harcamaları sıra ölçeğindeydi. Bu çizelgelerde adı geçen değişkenlerden biri satırlara, öbürü sütunlara konmuş, çizelgelerin gözlerine de kişi ya da aile sayıları, bir başka deyişle sıklık değerleri yerleştirilmişti.

Çizelge 7.8’de satırlarda üretim kesimleri ad ölçeğinde, sütunlarda yıl birimiyle ölçülen zaman değişkeni eşaralık ölçeğiyle yer almıştı. Bu çizelgenin gözlerindeki sayılar sıklıklar değil, dışarıya bağımlılık katsayılarıydı. Eğitim düzeyi, gelir, tüketim harcamaları, zaman gibi değişkenler özünde oran ölçeğinde ölçülen sürekli değişkenlerdir ama bu örneklerde öbeklenip toplulaştırılarak kesikli kılınmışlar, böylece çizelge düzenlemeye elverişli konuma getirilmişlerdir. Bu türden gösterimlere **iki boyutlu çizelgeler** adı verilebilir.

7. bölümde çizelgelere bakarak iki değişken arasında bir ilişki var mı sorusunu yanıtlamaya çalışıyorduk ama belli belirsiz ipuçlarından öteye gidemiyorduk. Sözelimi kadınların lise ya da yüksek okul düzeyinde eğitim görme şansının erkeklere göre daha düşük olduğunu; düşük gelirli ailelerin az, yüksek gelirli ailelerin çok tüketme eğiliminde bulunduğunu söyleyebiliyorduk. Aşağıda anlatılacak kimi yöntemlerle bu sorular biraz daha

iyi yanıtlanmaya çalışılacaktır. Önce iki boyutlu çizelgeler üzerinde küçük değişiklikler yaparak **simgeli çizelgeler** düzenleyecek, iki değişken arasında bir ilişki ortaya çıkarsa bunun gücünü daha iyi görmeye çalışacağız. Daha sonra **ortanca işleme** yöntemini uygulayarak, her iki değişkenin gözlem değerleri üzerindeki etkisini ayırıştırma yolunda bir adım daha atacağız. Bu arada kutu çizimlerinden de yararlanarak bazı yorumlar yapacağız.

Simgeli Çizelgeler

İki boyutlu bir çizelgeyi simgeli çizelgeye dönüştürmek için önce hangi satır ya da sütunda yer aldığına bakmaksızın, bütün gözlerdeki sayılarla bir dal-yaprak gösterimi düzenlenir. Bulunacak ortanca, dördebölen, eşik değerlerine göre her gözdeki değerin hangi aralığa düştüğü belirlenir. Bu aralıklarda yer alan her değere yeni düzenlenen çizelgede aşağıdaki gibi özel birer simge verilir (Velleman, Hoaglin, 1981:203-205'ten uyarlanmıştır):

Uzak üst dışadüşenler	⊕
Üst dışa düşenler	#
Üst dördebölen (D_u) ile üst eşik (E_u) arasındakiler	+
Dördebölenler aralığı (DAR) içindekiler	•
Alt dördebölen (D_a) ile alt eşik (E_a) arasındakiler	—
Alt dışadüşenler	=
Uzak alt dışadüşenler	≡

Simgeli çizelgelerle kutu çizimleri, verileri benzer biçimde ele alır. Her ikisi de dördebölenlerle eşikleri kullanarak verilerdeki genel örüntüyü dışa vurur, olağandışı değerleri açığa çıkarır. Dolayısıyla bu iki gösterimin birbirini tamamlamaları şaşırtıcı değildir.

Simgeli çizelge, verilerin bütün içindeki konumlarını korur. Böylece iki boyutlu yapıları göz önüne serer. Her satırın ya da her sütunun DAR 'larını, E 'lerini ayrı ayrı ele alıp simgeli çizelge düzenlemek görünümü karmaşık hale sokar. Satırlar ya da sütunlar arası bir karşılaştırmanın kutu çizimleriyle yapılması daha doğru olur. Kutu çizimlerinde her satır ya da sütun verilerinden dördebölenlerle eşik değerlerinin ayrı ayrı hesaplanması gerekir. Dolayısıyla her veri, yalnız kendi satırı ya da sütunu içindeki konumuna göre değerlendirilir. Bir satır ya da sütunda dışadüşen olabilecek bir gözlem değeri, bütün çizelge dikkate alındığında sıradan bir değere dönebilir.

Anlatılanları bir örnek üzerinde uygulayalım. Bunun için üretim kesimlerinin dışalığa bağımlılıklarını gösteren Çizelge 7.9'u ele alalım. Anımsanacağı gibi bu çizelge üzerinde değişiklikler yaparak verilerin ardında yatan

yapıyı görmeye çalışmıştık. Şimdi bunu biraz daha geliştirelim. Önce bu verileri bir dal-yaprak ile Çizim 11.1'deki gibi gösterelim.

(126)			
2	8	4	
3	7	2	
3	6		
4	5	6	
4	4		
9	3	33678	
23	2	01223 34555 5667	
(48)	1	00011 11222 23333 44444 44445 55556 66667 77888 89999 999	
49	0	00000 00111 12223 33333 33444 44555 55566 67777 77888 8999	

ÇİZİM 11.1 Dışalıma bağımlılık katsayılarının dal-yaprak gösterimi

Çizelge 7.9'dan elde edilmiştir.

Bu gösterimden yararlanıp D , E , uzak E değerleriyle bunların çeşitli aralıklarına vereceğimiz simgeleri şöyle gösterebiliriz:

Y_{eb}	Uzak $E_{\bar{u}}$	$E_{\bar{u}}$	$D_{\bar{u}}$	$Ortc$	D_a	Y_{ek}
126	$\oplus$ 42,25	# 22	+ 18,5	•	12,5	• 5 - 0

Şimdi de bu simgeleri Çizelge 11.1'deki gibi yerlerine koyalım. Ancak bunu yaparken üretim kesimlerini 7. bölümde de yaptığımız gibi beş öbeğe ayıralım. Böylece yorumlamalarımızda bazı kolaylıklar sağlayabiliriz.

Çizelge 11.1'e bakarak her bir kesimin zaman içindeki gelişmesini izleyebildiğimiz gibi, öbür kesimlere göre büyüklük konumunu da görebiliriz. Sözelimi petrol arıtma kesimi, beş yılda da bütün kesimlere göre hep dışadüşen olma özelliğini korumuştur. Bu da Türkiye ekonomisinin üretim yapısının son otuz yıldır petrol alımına ne kadar bağlı olduğunun bir göstergesi olarak yorumlanabilir. Ağır sanayi denebilecek çoğu üretim kesiminin, özellikle seksenli yıllardan bu yana, dışadüşen olması bu kesimlerin dışalıma çok bağımlı duruma geldiklerini yansıtmaktadır. Ayrıca pek çok kesimin, özellikle de hafif sanayi kesimlerinden bazılarının, dışalıma bağımlılığının zaman içinde yükselmesi de göze çarpmaktadır. İlgilenenler tek tek üretim kesimleri için daha ayrıntılı yorumlar üretebilir. Şunu vurgulamakla yetinelim: Bu simgeli çizelge, üretim kesimlerinin dışalıma bağımlılık konusundaki genel yapısını önceki çizelgelerden daha açık bir biçimde gözler önüne serbilmiştir. Dolayısıyla yararı tartışılmazdır.

ÇİZELGE 11.1 Dışalıma bağımlılık katsayıları için simgeli çizelge örneği

		1973	1979	1985	1990	1996
Ağır sanayi	Petrol	+	#	⊕	⊕	+
	Diğer imalat	+	•	+	⊕	+
	Ana metal	•	•	+	+	+
	Lastik, plastik	+	•	+	+	+
	Kimya	+	+	+	+	+
	Ulaştırma araçları	•	•	•	•	+
	Makine	•	•	•	•	•
	Cam, çimento	•	•	+	+	•
	Metal ürünleri	•	•	•	•	+
	Elektrikli makine	•	•	•	•	+
E+G+S İnşaat	Elektrik, gaz, su	•	•	+	•	•
	İnşaat	•	•	•	•	•
Hafif sanayi	Kâğıt, basım	•	•	+	+	•
	Dokuma, giyim	•	—	•	•	•
	Besin	—	—	•	•	•
	Ağaç, mobilya	—	—	•	•	•
Birincil üretim	Madencilik	•	—	•	•	•
	Tarım	—	—	•	—	•
Hizmetler	Ulaştırma, haberleşme	•	•	•	•	•
	Kişisel hizmetler	—	—	•	•	•
	Bankacılık, sigortacılık	—	—	•	•	—
	Ticaret	—	—	—	—	—
	Konut sahipliği	—	—	—	—	—
	Kamu hizmetleri	—	—	—	—	—

Çizelge 7.9'daki verilerden elde edilmiştir.

Bu çizelgenin yorumlanmasını tamamlayabilmek için Çizelge 7.9'daki verilerden yararlanıp her bir yıl için beş kutu çizimini yan yana göstermek okuyucuya bırakılmıştır.

ÇİZELGE 11.2 İTO Ücretliler Geçinme İndeksi (aylık % değişmeler)

Yıl	O	Ş	Mr	N	My	H	T	Ağ	Ey	Ek	K	Ar
1991	2,50	7,20	3,50	9,80	3,90	3,80	-1,30	2,30	6,20	10,80	3,30	4,60
1992	6,20	7,40	5,50	5,30	2,90	0,40	1,10	1,40	7,90	12,60	2,70	1,70
1993	4,30	4,30	5,10	6,30	7,10	2,70	1,90	1,70	4,70	11,40	4,40	1,40
1994	1,30	9,00	5,40	30,40	9,50	0,90	-0,80	-1,30	10,20	13,80	3,50	5,70
1995	3,00	6,20	4,80	11,20	5,40	3,10	1,80	2,00	5,80	15,00	3,90	4,60
1996	5,10	5,90	5,30	10,80	8,00	3,20	1,30	1,10	5,70	14,50	3,10	3,00
1997	6,10	2,20	6,30	9,10	8,50	3,20	2,60	5,50	5,50	13,40	5,10	3,20
1998	5,40	2,00	3,80	9,90	6,10	2,70	2,70	0,50	6,00	12,90	2,40	2,10
1999	0,30	2,30	5,00	7,50	3,80	2,40	2,50	0,10	6,60	9,90	2,40	2,90
2000	4,10	0,80	5,00	7,00	3,20	0,50	-0,10	1,50	3,70	7,80	0,90	2,50
2001	-0,30	1,10	5,60	13,20	5,10	2,80	-0,10	0,10	5,50	14,50	1,40	2,00
2002	2,20	1,10	2,80	6,00	2,50	0,01	-0,80	-2,10	4,00	9,10	1,00	0,20
2003	-0,70	1,30	2,80	9,30	4,20	0,10	-4,50	-0,60	1,60	4,40	-0,20	1,10

İTO, *Fiyat İndeksleri*, Çeşitli aylar.

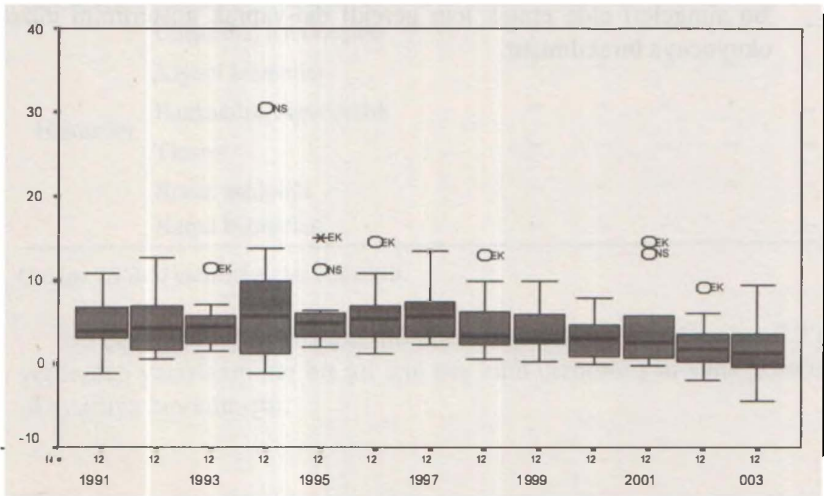
Şimdi de başka bir örnek üzerinde duralım. Çizelge 11.2 İstanbul Ticaret Odası'nın İstanbul'da tüketici fiyatlarını yansıtan Ücretliler Geçinme İndeksi'nin aylık % değişmelerini göstermektedir. Dal-yaprak gösteriminden yararlanılarak düzenlenmiş Çizelge 11.3'teki simgeleri yorumlayalım. Bu simgeleri elde etmek için gerekli dal-yaprak gösterimini düzenlemek okuyucuya bırakılmıştır.

ÇİZELGE 11.3 İTO Ücretliler Geçinme İndeksi için simgeli gösterim

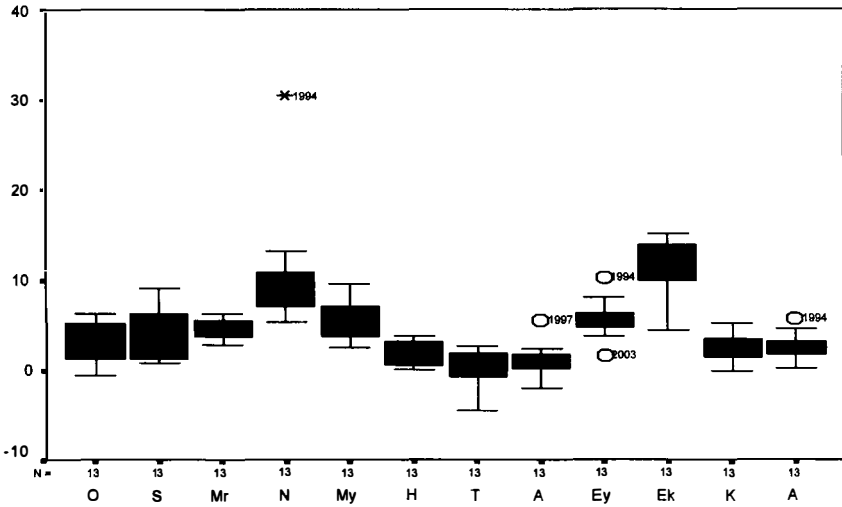
Yıl	O	Ş	Mr	N	My	H	T	Ağ	Ey	Ek	K	Ar
1991	•	+	•	+	•	•	—	•	+	+	•	•
1992	+	+	•	•	•	—	—	—	+	#	•	—
1993	•	•	•	+	+	•	•	—	•	+	•	—
1994	—	+	•	⊕	+	—	—	—	+	#	•	•
1995	•	+	•	+	•	•	—	•	•	#	•	•
1996	•	•	•	+	+	•	—	—	•	#	•	•
1997	+	•	+	+	+	•	•	•	•	#	•	•
1998	•	•	•	+	+	•	•	—	•	#	•	•
1999	—	•	•	+	•	•	•	—	+	+	•	•
2000	•	—	•	+	•	—	—	—	•	+	—	•
2001	—	—	•	#	•	•	—	—	•	#	—	•
2002	•	—	•	•	•	—	—	—	•	+	—	—
2003	—	—	•	+	•	—	—	—	—	•	—	—

Çizelge 11.2'deki verilerden elde edilmiştir.

Simgeli çizelgede iki önemli özellik göze çarpmaktadır. İlki, 1994 ile 2001'de iki kez patlayan enflasyon son yıllarda yavaşlamıştır. İkincisi, fiyatlar mevsimlik dalgalanma göstermektedir. Özellikle nisanla ekim, fiyat artışlarının en ağır yaşandığı aylardır. Yazın hazirandan ağustos sonuna kadar enflasyon hızı ortalamanın epeyce altına düşmektedir. Bu gözlemleri % artışların yer aldığı çizelgeden çıkarmak çok zordur. Sözü edilen iki özellik kutu çizimleriyle gösterilirse Çizim 11.2 ile 11.3'ü elde ederiz.



ÇİZİM 11.2 İTO Ücretliler Geçinme İndeksi aylık % değişimleri-yıllar



ÇİZİM 11.3 İTO Ücretliler Geçinme İndeksi aylık % değişimleri-mevsimler

Simgeli çizelgede göze çarpan ilk özellik Çizim 11.2’de iyice belirgindir. Fiyat artışlarındaki yavaşlamanın 1998’de başladığı, 2001 bunalımından sonra yine yoluna devam ettiği anlaşılmaktadır.

Çizim 11.3’te aylık enflasyon oranlarının yıl içindeki dalgalanması açıkça görülmektedir. Fiyatlar yaz aylarında ya az artmakta ya da düşmektedir. Ekimde doruğa vurduktan sonra birden azalmakta, martta biraz, nisan da daha çok yükselmekte, mayısta yine düşme eğilimine girmektedir. Bu çizimlerle simgeli çizelge birlikte kullanıldığında yorum zenginleşmektedir.

Ortanca işleme

Simgeli çizelgeler gözle algılanmaya uygundur. İki boyutlu veri çizelgelerindeki örüntüyü açığa çıkarırlar. Bu anlamda önemlidirler ama çoğu zaman bu verilerin daha ayrıntılı çözümlemesi gerekir. Genel yapı, yalın bir biçimde özetlendikten sonra kalıntıları incelemek uygun olur. Ortanca işleme yöntemi, bir çizelgedeki gizli yapıyı ya da örüntüyü belirleyebilmede basit bir yol olarak karşımıza çıkar. İki boyutlu çizelgelerdeki örtük yapılar çoğu zaman bütün satırlar ya da sütunlar arasındaki veri farklarıyla tanımlanır (Velleman, Hoaglin, 1981:219). Sözelimi, ortalama değerlerden daha yüksek bir satır dikkat çekebilir.

Bu tür çizelgelerin herhangi bir gözündeki değer genellikle o satırın, o sütunun sıra numarasıyla belirtilir. Buna göre 2. satır 3. sütundaki bir veri y_{23} ile, daha genel olarak i 'ninci satır j 'ninci sütundaki bir değer de y_{ij} ile gösterilir.

İki boyutlu bir çizelgedeki y_{ij} değerlerinin her birini birkaç bileşenin toplamıymış gibi düşünebiliriz. Bu bileşenler **ortak değer**, **satır etkisi**, **sütun etkisi**, **başka etkiler (kalıntı)** diye adlandırılabilir. Sözgelimi az önceki geçinme indeksi örneğinde satır etkisi mevsim dalgalanmaları, sütun etkisi yıllara göre enflasyonun düşmesi ya da çıkmasıdır. Öyleyse bütün bu etkiler sonucunda ortaya çıkan **tepki değeri** y_{ij} aşağıdaki gibi bir **toplamalı modelle** gösterilebilir (Emerson, Hoaglin, 1983c:167).

$$y_{ij} = \text{ortak değer} + i. \text{ satırın etkisi} + j. \text{ sütunun etkisi} + \text{kalıntı}_{ij}$$

$$y_{ij} = c + a_j + b_j + e_{ij}$$

Kalıntılar dışındaki ilk üç terimin toplamı bu modelin uygun bulduğu uyum değeridir. Başka bir deyişle şu eşitlikler yazılabilir:

$$y_{ij} = \text{uyum değeri}_{ij} + \text{kalıntı}_{ij}$$

$$y_{ij} = \hat{y}_{ij} + e_{ij}$$

Buradaki ortak değer, çizelgenin her gözünde değişmeden kalan bir sayıdır. Bütün çizelgedeki veri değerlerinin kökü ya da ortak düzeyi sayılabilir. Satır etkisi, o satırdaki değerlerin ortak değerden ne yönde, ne kadar farklı olduklarının ölçüsüdür. Sütun etkisi de benzer biçimde tanımlanabilir. Uyum değeri, modelin biri satırlarda, öbürü sütunlarda yer alan iki değişkenle yaptığı kestirimidir. Kalıntılar ise bu etkiler dışında ortaya çıkan, dolayısıyla modelin açıklayamadığı bileşendir. Her bir kalıntı değerinin incelenmesi bize bazı ipuçları verebilir. Böylece toplamalı model, singeli çizelgede aradığımız örüntüyü nicel değerlerle gösterir.

Bu anlattıklarımızı bir örnek üzerinde gösterelim. Bunun için 7. bölümde Alıştırma 7.3'teki bölgesel satışları ele alalım. Oradaki verileri Çizelge 11.4'e aktaralım. Bu örnekte her bir gözdeki satış tutarı, ortak değer, yanı sıra üç etkinin toplamıdır: Mevsim (satır) etkisi, bölge (sütun) etkisi, burada adı geçmeyen öbür bütün etkiler.

ÇİZELGE 11.4 Üçer aylık bölgesel satışlar (milyon YTL)

Dönem	Marmara	Ege	Akdeniz	İç Anadolu
I	98	75	50	48
II	92	75	57	42
III	101	100	80	50
IV	90	75	51	39

Alıştırma 7.3'ten alınmıştır.

Üçer aylık dönemleri gösteren “I” satırını $i = 1, \dots$ “II” satırını $i = 2, \dots$, “Marmara” sütununu $j = 1$, “Ege” sütununu $j = 2, \dots$ ile gösterip çizelgeyi yeniden düzenleyelim. En sağa ekleyeceğimiz bir sütuna da satır ortancalarını yazalım. Bunun nedeni aşağıda açıklanacaktır.

ÇİZELGE 11.4 Bölgesel satışlar

	$j = 1$	2	3	4	<i>Ortc</i>
$i = 1$	98	75	50	48	62,5
2	92	75	57	42	66
3	101	100	80	50	90
4	90	75	51	39	63

Şimdilik nasıl hesaplandıklarına aldırmadan ortak değerin 66,75; satır etkilerinin sırasıyla 1, -1, 15, -3,5; sütun etkilerinin de sırasıyla 26,5, 10,5, -10,5, -24 olduğunu düşünelim. Bu durumda modelin bulacağı uyum değerleri Çizelge 11.5’teki gibi hesaplanabilir. Sözelimi Marmara Bölgesi’nde ilk çeyrek satışları aslında 98 milyon YTL iken, bu modelin kestirimi 94,25 milyon YTL çıkmaktadır. Aradaki fark 3,75 milyon YTL burada adı geçmeyen öteki etkilerin sonucudur, başka bir deyişle bu modelin açıklayamadığı kalıntıdır. Öbür gözler için de bu kalıntılar hesaplanıp incelenebilir.

ÇİZELGE 11.5 Toplamalı modelin uyum değerleri

		$j = 1$	2	3	4
$i = 1$	Ortak değer	66,75	66,75	66,75	66,75
	+ satır etkisi	+ 1	+ 1	+ 1	+ 1
	+ sütun etkisi	+26,5	+ 10,5	- 10,5	- 24
	= uyum değeri	= 94,25	= 78,25	= 57,25	= 43,75
2	Ortak değer	66,75	66,75	66,75	66,75
	+ satır etkisi	- 1	- 1	- 1	- 1
	+ sütun etkisi	+26,5	+ 10,5	- 10,5	- 24
	= uyum değeri	= 92,25	= 76,25	= 55,25	= 41,75
3	Ortak değer	66,75	66,75	66,75	66,75
	+ satır etkisi	+ 15	+ 15	+ 15	+ 15
	+ sütun etkisi	+26,5	+ 10,5	- 10,5	- 24
	= uyum değeri	= 108,25	= 92,25	= 71,25	= 57,75
4	Ortak değer	66,75	66,75	66,75	66,75
	+ satır etkisi	- 3,5	- 3,5	- 3,5	- 3,5
	+ sütun etkisi	+26,5	+ 10,5	- 10,5	- 24
	= uyum değeri	= 89,75	= 73,75	= 52,75	= 39,25

Şimdi bütün bu etkilerin nasıl bulunabileceğini görelim. Aslında toplamalı modelle kestirim yapmanın çeşitli yolları vardır. Hangi yol kullanılırsa kullanılsın ilk verilerden yola çıkarak, bir ortak değer, her satır, her sütun için birer etki değeri, her göz için de birer kalıntı elde ederiz. Ortanca işleme süreci tamamlandığında bu bileşenler Çizelge 11.6'da gösterilen yerlerde bulunurlar.

ÇİZELGE 11.6 Ortanca işleme sürecinin sonucu

Kalıntılar	Satır etkileri
Sütun etkileri	Ortak değer

Sorgulayıcı veri çözümlemelerini incelerken *Ortc* kullanmanın aşırı değerlerden etkilenmeyen **berk** sonuçlar verdiğini görmüştük. Burada da *Ortc*'dan yararlanmayı düşünebiliriz. Sözelimi çizelgenin bir satırının *Ortc* değerini, o satırdaki her bir değerden çıkarırsak kalan sayıları, satır etkisinden temizlenmiş **geçici kalıntılar** sayabiliriz. Satır *Ortc*'sı da geçici satır etkisidir. Burada **geçici** sözcüğünü kullanma nedenimiz, daha sonraki

adımlarda bu değerlerin değişebilecek olmasıdır. Bu işlemle satır etkisinin gözlem değerlerine olan katkısını kenara taşımış oluruz.

Söylediklerimizi yukarıdaki örneğe uyguladığımızda Çizelge 11.7 elde edilir. Bu çizelgenin herhangi bir gözündeki değerle o satır *Ortc*'sının toplamı Çizelge 11.4'ün ilgili gözündeki ilk değere eşittir.

ÇİZELGE 11.7 Bölgesel satışlar – İlk adım (a)

	$j = 1$	2	3	4	<i>Ortc</i>
$i = 1$	35,5	12,5	-12,5	-14,5	62,5
2	26	9	-9	-24	66
3	11	10	-10	-40	90
4	27	12	-12	-24	63

Bulduğumuz geçici kalıntıların bu kez de sütunlar boyunca *Ortc*'lerini hesaplar, en alt satıra yazarsak geçici sütun etkilerini elde ederiz. Çizelge 11.8'de bu işlemin sonucu gösterilmiştir. Satır *Ortc*'lerinin da *Ortc*'sı alınıp sağ alt köşedeki göze yazılmıştır. Bunu **geçici ortak değer** sayabiliriz. Gözlerdeki değerler de satır etkilerinden arındırılmış geçici kalıntılardır.

ÇİZELGE 11.8 Bölgesel satışlar – İlk adım (b)

	$j = 1$	2	3	4	<i>Ortc</i>
$i = 1$	35,5	12,5	-12,5	-14,5	62,5
2	26	9	-9	-24	66
3	11	10	-10	-40	90
4	27	12	-12	-24	63
<i>Ortc</i>	26,5	11	-11	-24	64,5

Şimdi de her gözdeki geçici kalıntıdan sütun *Ortc*'sını çıkarıp aradaki farkı aynı göze yazalım. Böylece kalıntıları biraz daha son değerlerine yaklaştırmış oluruz. Bu arada satır ortancalarının ortak değerden farklarını da benzer biçimde bulalım. Bu işlem de bizi satır etkilerine yaklaştırmış olur. Böylece Çizelge 11.9 elde edilir.

ÇİZELGE 11.9 Bölgesel satışlar – İlk adım (c)

	$j = 1$	2	3	4	<i>Geçici satır etkisi</i>
$i = 1$	9	1,5	-1,5	9,5	-2
2	-0,5	-2	2	0	1,5
3	-15,5	-1	1	-16	25,5
4	0,5	1	-1	0	-1,5
<i>Geçici sütun etkisi</i>	26,5	11	-11	-24	64,5

Böylece ilk adımdaki geçici etkilerle kalıntıları bulmuş olduk. Ancak bu yeterli değil, çünkü geçici kalıntıların içinde hâlâ bir takım etkiler kalmış olabilir. Bu nedenle ikinci adıma geçeceğiz. İkinci adımda ilk adımın işlemleri bir kez daha uygulanır. Her türlü etki değeriyle kalıntılar iyice işlenmiş olur. Çoğu zaman iki adımlık işlemler sonuca ulaşmada yeterlidir.

Şimdi de yeni kalıntıların satır ortancalarını bulup sağda açtığımız yeni *Ortc* sütununa koyarsak Çizelge 11.10'a ulaşırız. Yeni *Ortc* değerleri aslında satır etkilerine katılması gereken sayılardır. Bir sonraki aşamada bu yapılacaktır. Bu sütunun en altındaki "0" değeri de o satırın ortancasıdır. Bu değer sıfırdan farklı çıksaydı, ortak değere eklememiz gerekirdi. Böylelikle ortak değeri biraz daha sonuca yaklaştırmış olurduk. Ama örneğimizde buna gerek kalmadı.

ÇİZELGE 11.10 Bölgesel satışlar – İkinci adım (a)

	$j = 1$	2	3	4	<i>Yeni Ortç</i>	<i>Geçici satır etkisi</i>
$i = 1$	9	1,5	-1,5	9,5	5,25	-2
2	-0,5	-2	2	0	-0,25	1,5
3	-15,5	-1	1	-16	-8,25	25,5
4	0,5	1	-1	0	0,25	-1,5
<i>Geçici sütun etkisi</i>	26,5	11	-11	-24	0	64,5

İşlemlerimizi sürdürülecek olursak her gözdeki geçici kalıntıdan aynı satırdaki yeni *Ortc* değerini çıkarırız. Farkı yeni geçici kalan olarak aynı göze yazarız. Bir yandan da geçici satır etkilerine yeni *Ortc* değerlerini ekleyerek bu etkiyi de bir adım daha ileri taşır, Çizelge 11.11'i buluruz.

Eğer ikinci adım sonunda satır ya da sütunların yeni *Ortc*'lerinin hesaplanacak *Ortc*'sı sıfırdan farklı çıkarsa üçüncü adıma geçilir.

ÇİZELGE 11.11 Bölgesel satışlar – İkinci adım (b)

	<i>j</i> = 1	2	3	4		<i>Geçici satır etkisi</i>
<i>i</i> = 1	3,75	-3,75	-6,75	4,25		3,25
2	-0,25	-1,75	2,25	0,25		1,25
3	-7,25	7,25	9,25	-7,75		17,25
4	0,25	0,75	-1,25	-0,25		-1,25
<i>Yeni Ortc</i>	0	-0,5	0,5	0		2,25
<i>Geçici sütun etkisi</i>	26,5	11	-11	-24		64,5

Artık ne yapacağımızı biliyoruz. Yeni *Ortc*'ları ilgili geçici kalıntılardan çıkarıp altlarındaki geçici sütun etkilerine ekleyeceğiz. Bu arada yeni sütun *Ortc*'lerinin *Ortc*'sını üstündeki satırlarda yer alan geçici satır etkilerinden çıkarmayı da unutmayacağız. Bunlar yapılıncaya Çizim 11.12 bulunur.

ÇİZELGE 11.11 Bölgesel satışlar – İkinci adım (c)

	<i>j</i> = 1	2	3	4	<i>Satır etkisi</i>
<i>i</i> = 1	3,75	-3,25	-7,275	4,25	1,00
2	-0,25	-1, 25	1,750	1,25	-1,00
3	-7,25	7,750	8,750	-7,75	15,00
4	0,25	1,250	-1,750	-0,25	-3,50
<i>Sütun etkisi</i>	26,50	10,500	-10,500	-24,00	66,75

Böylece ikinci adımı da bitirdik. Yine de bazı satırlarla sütunlarda yeni bulunacak *Ortc*'lar sıfırdan farklı ama hayli küçük oldukları, bizim de zaten kaba işlemler yaptığımız düşünülecek olursa bir adım daha ileri gitmeye pek gerek yoktur. Alıştırmalar sırasında üçüncü adımın işlemlerini de yapması okuyucudan beklenecektir.

Herhangi bir gözdeki kalıntıya o satırın etkisiyle o sütunun etkisini ekleyip bunlara bir de ortak etkiyi katarsak ilk y_{ij} değerine ulaşırız. Sözgelimi üçüncü satır ikinci sütunu alalım. Satır etkisi 15 ile sütun etkisi 10,5 toplanıp buna ortak değer 66,75 eklenirse elde edilen 92,25 sayısı, o gözün Çi-

zelge 11.5'te de görülen uyum değeridir. Buna ilgili gözdeki 7,75'lik kalıntı eklendiğinde Çizelge 11.4'teki 100 bulunur. Görüldüğü gibi Çizelge 11.11, Çizelge 11.4'ün içerdiği her gözlem değerini, ortak değer, satır etkisi, sütun etkisi, kalıntı diye dört parçaya ayırmıştır. Kalıntılar ne kadar küçükse modelin olguyu açıklama gücü de o kadar yüksektir.

Çizelgenin nasıl yorumlanacağına gelince şunlar söylenebilir. Mevsimin de, bölgenin de, başka herhangi bir değişkenin de hiç bir etkisi olmasaydı çizelgenin her gözünde, ortak değer 66,75 milyon YTL yer alırdı. Bunun anlamı şudur: Her mevsimde, her bir bölgede 66,75 milyon YTL kadar satış gerçekleşebilir. Bu bütün etkilerden arınmış temel satış tutarı gibi görülebilir. Kendi başına çok önemli değildir ama etmenlerin etkileri bu değer temel alınarak incelenir. Başka bir deyişle dönemsel, bölgesel ya da başka tür etkilerin satışlar üzerindeki etkileri bu değerden sapmalara göre ölçülür. Örneğimizde Marmara bölgesi bunun 26,5 milyon YTL, Ege 10,5 milyon YTL üstüne çıkarken, Akdeniz 10,5, İç Anadolu 24 milyon YTL altında kalacaktır. Yılın üçüncü üç ayında her bölgede satışların bu temel değerden 15 milyon YTL kadar daha çok olması beklenir. Öbür mevsimlerdeki oynamalar görece daha küçüktür.

Mevsim ile bölge dışında kalan başka değişkenlerin, ortak değeri aşağıya ya da yukarıya doğru ne kadar çektikleri sorusunun yanıtı kalıntılardadır. İşarete bakılmaksızın kalıntıların en büyük değerleri bütün bölgelere yayılmış olarak yılın Temmuz-Eylül aylarındadır. Bu dönemin kendi etkisinin de büyük olduğu düşünülürse satışlarda en büyük oynamanın bu üç ayda olduğu anlaşılır. Bu arada satışların yükseldiği dönemlerin, Marmara ile İç Anadolu'da Ocak-Mart, Ege ile Akdenizde Temmuz-Eylül olmasının gerisinde ne yatıyor olabilir? İşte size araştırılması gereken bir ipucu. Bölgelerin en düşük satışları da tam tersi dönemlerde ortaya çıkmaktadır. Alın size bir soru daha? Bütün bu ipuçları araştırmayı beklemektedir. Bunlar üzerinde düşünmeye değer mi? Eğer okuyucu bu kalıntılar için simgeli çizelge hazırlarsa söylenenleri kolayca doğrulayabilir. Aslında bu son bulgular, dönemlerle bölgeler arasında karşılıklı bir etkileşim olduğu izlenimini vermektedir. Bütün bu soruların yanıtları **çoklu varyans çözümlemesi** yöntemlerine başvurularak öğrenilebilir.

Bu örneğimizde ortanca alma işine önce satırlardan başladık. Peki sütunlardan başlasaydık sonuç aynı olur muydu? Bunu denemek okuyucuya bırakılmıştır. Ancak şu kadarını söyleyelim ki bu seçim, her zaman tam aynı sonucu vermeyebilir. Ama aradaki fark, sorgulayıcı veri çözümleme amaçları için pek de önemli değildir (Velleman, Hoaglin, 1981:233).

Örneğimizde satır-sütun sayıları az olduğu için işlemleri kolayca yapabildik. Ama büyük çizelgelerle çalışırken, hele iki adımda sonuç elde edilemiyorsa, bu iş iyice sıkıcılaşır. Neyse ki MINITAB yazılımı "median polish" seçeneğiyle bu güçlüğün üstesinden gelir. Böylece iki adımlık süreç

gerekirse istendiği kadar yinelenebilir. Genellikle iki adımın uygulanması yeterli olmaktadır. Bu yazılımla ister satırlardan, ister sütunlardan işe başlama olanağı da bulunmaktadır.

Ortanca işleme, aşırı değerlere karşı dirençli **berk** bir yöntemdir. Veri kümesinde yer alan bir kaç tane çok büyük ya da çok küçük gözlemin, gerek ortak değeri, gerek satır ya da sütun etkilerini değiştirmesini engeller. Böyle aşırılıklar kalıntılar içinde hapsolür. Bu nedenle ortanca işleme yöntemi sorgulayıcı veri çözümlemeleri için çok uygundur. Özellikle dışadüşen(ler) varsa bunlara dikkat çekip üzerinde düşünülmesini sağlar.

Benzer bir inceleme *Ortc* yerine *AO* kullanarak da yapılabilir. Sonuçlar ortanca işleme yönteminin bulgularıyla karşılaştırılırsa, genellikle, çoğu kalıntının ne sifıra yakın yerlerde ne de uçlarda yer aldığı görülür (Emerson, Hoaglin, 1983c:181). Ayrıca *AO* dışında kırpık ortalama (*KırO*), dördebölenler arası ortalama, M-kestirimcileri gibi ortalamalar da benzer biçimde kullanılabilir (Emerson, Hoaglin, 1983c:190-192; Rosenberger, Gasko, 1983; Goodall, 1983).

Bazı iki boyutlu çizelgeler yukarıdaki toplamalı model için uygun olmayabilir. Uygunsuzluğun nasıl anlaşılabilceği, çözüme nasıl ulaşılabilceği bu kitabın sınırlarını aşmaktadır. Bu konuda çok kısa olarak şu söylenebilir. Satır-sütun etkileriyle ortak etki kullanılarak **karşılaştırma değerleri** denen yeni değerler bulunur, bunlar ortanca işlemeyle bulunan kalıntılarla karşılaştırılır. Aralarında **doğrusal ilişki** olup olmadığına bakılır. Eğer ilişki doğrusal yakınsa toplamalı model uygun sayılır, yok eğer ilişki doğrusal dışı bir özellik gösteriyorsa özgün verilere dönüştürme uygulanarak toplamalı model uygun biçime çevrilir. (Emerson, Hoaglin, 1983c:200-204; Marsh, 1988:291-293, Velleman, Hoaglin 1981: 233-240). İlk veri çizelgesinde kimi gözlerin boş olması, yani içinde hiç bir değer görülmemesi durumu için de bazı çözüm yolları üretilmiştir (Emerson, Hoaglin, 1983c:198-200).

Ortanca işleme yöntemini, simgeli çizelgelerde kullandığımız dışalıma bağımlılık örneğine uygularsak Çizelge 11.12'yi elde ederiz. Bu çizelgede üretim kesimleri satır etkilerinin büyüklüğüne göre sıraya dizilmiştir. Ortak değer 12,75 olması, Türkiye ekonomisinde üretim kesimlerinden birinin 1973-96 döneminde sözgelimi 100 YTL değerinde üretim yapabilmesi için ekonominin yurtdışından ortalama 12,75 YTL kadar girdi almak zorunluluğunda kaldığını gösterir.

ÇİZELGE 11.12 Dışalığa bağımlılık katsayıları: Ortanca işleme sonuçları

	1973	1979	1985	1990	1996	Satır etkileri
Petrol	-20,0	0,0	24,5	12,0	-25,0	46,75
Diğer imalat	14,5	-4,5	0,0	101,5	-0,5	11,25
Ana metal	-4,5	-4,5	0,0	9,5	0,5	10,25
Lastik, plastik	0,0	-9,0	-4,5	2,0	8,0	10,75
Kimya	-2,0	0,0	-3,5	4,0	2,0	9,75
Ulaştırma araçları	1,0	-2,0	0,5	-4,0	0,0	5,75
Cam, çimento	-3,0	1,0	4,5	0,0	-8,0	5,75
Makine	1,0	-1,0	-1,5	0,0	0,0	3,75
Kâğıt, basım	0,0	0,0	2,5	8,0	-2,0	3,75
Metal ürünleri	0,0	-5,0	1,5	-1,0	4,0	2,75
Elektrikli makine	0,0	-5,0	-8,5	2,0	5,0	2,75
İnşaat	0,5	-3,5	0,0	0,5	-0,5	0,25
Elektrik, gaz, su	0,0	5,0	12,5	-4,0	0,0	-0,25
Ulaştırma, haberleşme	-2,0	0,0	2,5	0,0	-3,0	-1,25
Dokuma, giyim	-2,0	-3,0	2,5	0,0	6,0	-2,25
Maden	3,5	0,5	0,0	-3,5	-4,5	-4,75
Kişisel hizmetler	-3,0	-1,0	1,5	8,0	0,0	-5,25
Besin	-3,0	0,0	-0,5	2,0	2,0	-5,25
Ağaç, mobilya	0,0	0,0	1,5	-1,0	6,0	-6,25
Tarım	-0,5	1,5	0,0	-1,5	0,5	-7,75
Bankacılık, sigortacılık	-1,0	1,0	1,5	0,0	-2,0	-8,25
Ticaret	0,0	2,0	-0,5	0,0	0,0	-10,25
Konut sahipliği	0,0	3,0	-1,5	-1,0	0,0	-11,25
Kamu hizmetleri	1,5	3,5	0,0	-0,5	-1,5	-12,75
Sütun etkileri	-1,5	-3,5	0,0	0,5	1,5	12,75

Çizelge 7.9'dan ortanca işlemedeki iki adım ikişer kez uygulanarak elde edilmiştir.

Şimdi zaman içindeki gelişmeyi yansıtan sütun etkilerine bakalım. 1973'te dışalığa bağımlılık katsayısı, 12,75'in 1,5 puan altında, yani 11,25 iken, döviz darboğazının etkisiyle 1979'da 2 puan daha düşmüş, 1990'da 0,5, 1996'da ek 1 puan artarak 14,25'e çıkmıştır. Bu bulgu, 1980'lerden bu yana dışalığa bağımlılığımızın nasıl arttığını nicel olarak yansıtmaktadır.

Çeşitli üretim kesimlerinin farklılıklarını gösteren satır etkilerine gelince, dışalığa bağımlılık katsayılarının genellikle ağır sanayi kesimlerinde yüksek, hizmet kesimlerinde düşük olduğu görülüyor.

Özellikle petrol artırımının, ortak değer 46,75 puan üstünde olması çok çarpıcıdır. Bunun anlamı, 1973-96 döneminde petrol artımı kesiminin 100 YTL üretim için, bütün üretim kesimlerinin ortancasından 46,75 YTL daha çok dışalım gereği doğurduğudur. Ana metal, lastik-plastik, kimya kesimleri de ortak değerden yaklaşık 10 YTL yukarıda yer almaktadır. Buna karşılık tarım, bankacılık, ticaret, konut sahipliği, kamu hizmetlerinin dışalima bağımlılık katsayıları genel ortalamadan göze batacak kadar küçüktür. Hatta birininki sıfırdır. (Hangisi?)

Okuyucunun, kalıntıları inceleyebilmek için düzenlemesini beklediğimiz dal-yaprak gösterimine göre 120 gözdeki değerden 113'ü $\pm$ %1 aralığındadır. Bu kalıntılar için bir de simgeli çizelge hazırlayıp hangilerinin önemli ölçüde büyük ya da küçük olduklarını görmek, sonra da bunların olası nedenleri üzerinde düşünmek okuyucuya bırakılmıştır.

Son bir örnek olarak Çizelge 11.2'de yer alan İTO Ücretliler Geçinme İndeksi'ndeki aylık % değişmelere ortanca işleme yöntemi uygulandığında ortak değer 3,915 çıkmıştır. Yıllar boyunca gelişmeyi yansıtan satır etkileriyle mevsim dalgalanmalarının göstergesi olan sütun etkileri de Çizelge 11.13'deki gibi bulunmuştur. Okuyucu bu bulguları doğrulamalı, kalıntıları hesaplamalı, daha sonra hepsini yorumlamalıdır.

ÇİZELGE 11.13 İTO Ücretliler Geçinme İndeksi aylık % değişmeler: Ortanca işleme sonuçları

Satır etkileri				Sütun etkileri			
1991	0	1998	-0,475	Oc	0,040	Tm	-2,815
1992	0,025	1999	-1	Şb	-0,040	Ağ	-2,590
1993	0,345	2000	-2,055	Mr	0,885	Ey	2,285
1994	1,25	2001	-1,225	Ns	5,885	Ek	7,985
1995	0,72	2002	-2,575	My	2,410	Ks	-0,960
1996	0,67	2003	-2,575	Hz	-1,360	Ar	-0,690
1997	1,5						

Çizelge 11.2'deki verilere ortanca işleme yöntemi uygulanarak elde edilmiştir.

11.2 İKİ NİCEL DEĞİŞKEN ARASINDAKİ İLİŞKİ

İki boyutlu çizelgeleri incelerken satır ya da sütunda yer alan değişkenlerin ölçüm düzeylerine dikkat ettiniz mi? Bölgelerle mevsimlere göre satışları ele aldığımız örnekte her iki değişkenin de ölçeği niteldi. Dışalima bağımlı-

lık katsayılarında üretim kesimlerini gösteren değişken de ad ölçeğinde yer alıyordu. İTO ücretliler geçinme indeksi örneğinde boyutlardan biri “yıl”dı, nicel bir ölçekte ölçülüyordu ama öteki aydı, ad ölçeğine uyuyordu. Kısacası çizelge boyutlarından hiç olmazsa biri ad ya da sıra ölçeği gibi nitel bir ölçekle ölçülmüştü.

Her iki boyutun da aralık ya da oran ölçeği gibi nicel bir ölçekte ölçüldüğü durumlarda, gözlem değerlerini belli aralıklarda toplulaştırmadan, iki boyutlu bir çizelgede gösteremeyiz. Böyle veri kümelerinde her gözlem biriminin iki değeri olduğunu, bunların her birinin bir değişkene karşılık geldiği söylenebilir.

Diyelim bir iş kolunda çalışan kimselerin ücretleriyle o iş kolunda çalışma sürelerini inceliyor, bunlar arasında bir ilişki olup olmadığını araştırıyorsak, aynı insan için iki verimiz olur. Biri o kimsenin aldığı ücret, öbürü de iş kolunda çalıştığı süre. Bu iki sayı aynı kişinin iki ayrı özelliğinin ölçümüdür. Uygulamada genellikle bunlardan birine X , ötekine Y değişkeni adı verilir. Araştırma konusu, ikisinden birinin öbürü üzerinde bir etkisi olup olmadığı biçimindeyse, etki yaratana **bağımsız (açıklayıcı) değişken** denir, X ile gösterilir. Etki altında kaldığı düşünülen özellik ise **bağımlı (açıklanan) değişken** diye adlandırılır, Y ile simgelenir. İki değişken arasında bir nedensellik ilişkisi olmamasına karşın birlikte hareket ediyorsa hangisine X , hangisine Y dendiği önemli değildir. Demek ki her gözlem biriminin (burada o iş kolunda çalışan herkesin) bir X , biri Y ile gösterilen iki ayrı değeri vardır. Aynı birime ilişkin bu iki değer X_i , Y_i biçiminde yazılabilir. Gözlem birimini gösteren i , 1’den başlayarak, topluluktaki gözlem sayısı n ’ye kadar olan bütün tamsayı değerlerini alır.

Bu verileri bir bakışta kolayca gözlemleyebilmede kullanılacak en kolay yol bir **serpilme çizimi** düzenlemektir. Bu çizimi 7. bölümde örneklerle ele almıştık. Burada tüketim fonksiyonu diyebileceğimiz başka bir örnek ile incelemeyi sürdürelim.

Türkiye’de 1987-2003 arasında **özel harcanabilir gelirle özel tüketim harcamaları** arasındaki ilişkiyi ele alalım. İki değişken arasında nasıl bir ilişki var görmeye çalışalım. Toplumun geliri artınca tüketimi de artırıyor mu? Ne kadar? Bir liralık gelir artışı, tüketimde ne kadar bir artışa yol açıyor? Bu artış tutarı incelenen dönem boyunca aynı mı kalıyor, yoksa zamanla değişiyor mu? Acaba tüketim değişkeninin tepkisi yüzde olarak mı ele alınmalı? Yani 1987-2003 arasında gelir % 1 artınca, tüketimin buna yanıtı belli bir % oran kadar artış biçiminde mi oluyor? İktisatçıların deyişiyle **tüketimin gelir esnekliği** bu dönem boyunca sabit mi? İşte incelemek istediğimiz sorular bunlar. Ele alacağımız verilerin dökümü Çizelge 11.14’te, serpilme çizimi, kutu çizimleriyle birlikte Çizim 11.2’dedir.

ÇİZELGE 11.14 Türkiye’de özel harcanabilir gelirle (ÖHG) özel tüketim harcamaları (ÖTH) (1987-2003) (1998 fiyatlarıyla katrilyon TL)

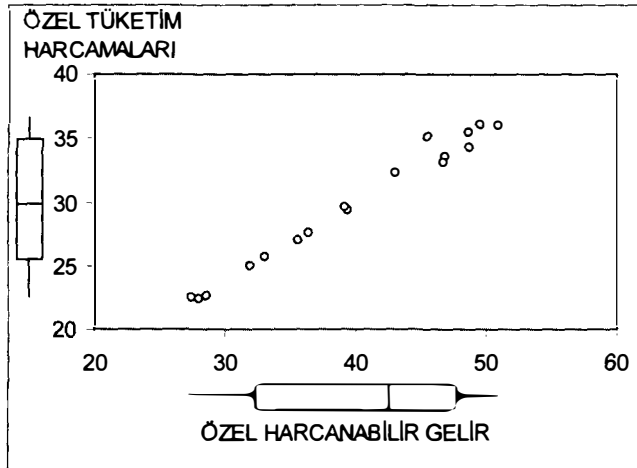
YIL	ÖHG	ÖTH	YIL	ÖHG	ÖTH
1987	27,5	22,5	1996	43,1	32,4
1988	28,0	22,4	1997	45,5	35,1
1989	28,7	22,6	1998	48,6	35,5
1990	32,0	25,0	1999	46,9	33,7
1991	33,1	25,6	2000	49,6	36,1
1992	35,7	27,0	2001	46,8	33,2
1993	39,5	29,3	2002	48,7	34,3
1994	36,5	27,6	2003	50,9	36,1
1995	39,2	29,6			

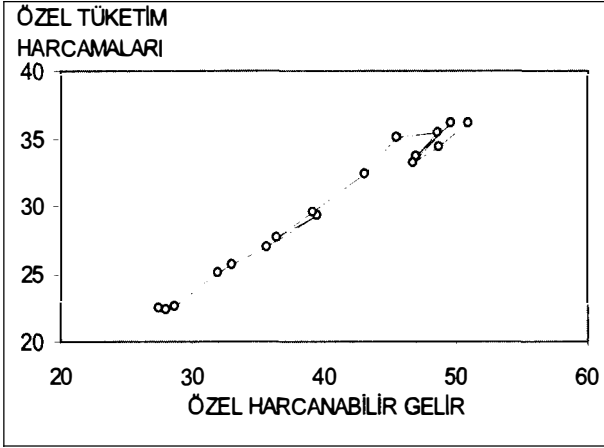
Hazine Müsteşarlığı, 2004:4.

İktisat kuramına göre bu örnekte tüketimin geliri değil, gelirin tüketimi etkilemesini beklediğimizden gelire X , tüketime Y diyerek ilkinin yatay eksene, ikinciye dikey eksene yerleştirdik.

Şimdi bu çizelgeyle çizime bakarsak neler görebileceğimizi düşünelim. Önce kutu çizimlerini teker teker ele alalım. Özel harcanabilir gelir sola çarpık izlenimi vermektedir. Özel tüketim harcamaları görece olarak daha bakışıktır. Bu bulguyu bir kenara yazıp unutmayalım.

Gelirin düşük olduğu yıllarda tüketim de düşük, gelirin yüksek olduğu yıllarda tüketim de yüksek değerler almaktadır. Demek ki Türkiye’de 1987-2003 arasında özel tüketim harcamaları gelire bağlı olarak değişmektedir, yani iktisat kuramının öngörüsü bu örnekte de geçerlidir.

**ÇİZİM 11.2** Harcanabilir gelirle tüketimin serpilme çizimi (1987-2003) (1998 fiyatlarıyla katrilyon TL)



ÇİZİM 11.3 Harcanabilir gelirle tüketimin serpilme çizimi

45 katrilyon TL'nin altındaki gelir düzeylerinde iki değişken doğrusal bir biçimde birlikte hareket eder gibi görünürken daha üst gelir düzeylerinde sanki işler biraz karışmakta, aynı ya da birbirine yakın gelir düzeylerine hayli farklı tüketim harcamaları karşı gelebilmektedir. Sözgelimi 1998'le 2002'nin gelir düzeyleri hemen hemen aynıdır ama tüketim düzeyleri arasındaki fark bir katrilyon TL'den büyüktür. Bu karışık durumları daha iyi görebilmek için Çizim 11.3'e bir göz atalım.

Yeni çizimde gözlem noktaları birbirlerine zaman sırasına göre birer çizgiyle bağlanmıştır. Böylece zaman içindeki hareketleri daha açık izleyebiliriz. Sözgelimi gelir 40 katrilyon TL'ye yaklaşmışken birden düşmüş (1994 bunalımı), tüketim de hemen aynı tepkiyi göstermiştir. Benzer bir durum gelir 50 katrilyon TL'ye yaklaşıncaya (biri 1999, öbürü 2001'de) iki kez yaşanmıştır. Ancak bu son iki bunalımda tüketimdeki düşüşler 1994'e göre daha büyük olmuştur. Son beş-altı yılın tüketim düzeyleri, aynı gelir düzeylerindeki eski yılların tüketimlerinin altında kalmaktadır. İşte size daha ayrıntıya girilip araştırılması gereken bir kaç soru. Bunların nedeni ne olabilir?

Bu soruyu iktisatçılara bırakalım. Gelirle tüketim arasındaki ilişkiyi genel olarak nasıl özetleyebiliriz sorusuna yönelelim. Gelirde bir liralık artış olduğunda tüketimdeki ortalama artış bütün dönem boyunca hep aynı mı kalmaktadır? Yanıt evetse bu ilişkiyi bir doğru üzerinde kayan noktalarımız gibi düşünebiliriz. Bu doğrunun eğimi, bir liralık gelir artışının ne kadar tüketim artışına yol açtığı sorusunun yanıtıdır.

Öyle ya, $\hat{Y} = a + bX$ biçimindeki bir doğrunun b ile belirtilen eğimi, X 'teki bir birim değişme karşısında $\hat{Y}$ 'nın kaç birim değiştiğini gösterdiğine göre, bu değer sorumuza bir yanıt olabilir (Bu denklemlerle kestirilen değerler şapkalı Y ile gösterilir). b 'nin hangi sayıya eşit olduğu biliniyorsa, gelirdeki her bir liralık artış sonucunda $\hat{Y}$ 'nın ne kadar değişeceği hesaplanabilir. Öyleyse bu doğru denklemindeki katsayıları elde etmemiz gerekir. Bu katsayıları elde etmekle, doğruyu bulmak aynı şeydir. Ama çizilecek doğru bu noktaları iyi temsil edebilmeli ki onun eğim değerine güvenebilelim.

Çizimimizdeki noktaların arasından göz kararıyla bir doğru geçirmeye kalkışsak her birimiz az da olsa farklı doğrular önerebiliriz. Bunlardan hangisi en iyisidir? Burada yanıt, “en iyi” ile ne demek istediğimize bağlıdır.

Bir ölçüt, “çizeceğimiz doğrunun bu noktaların her birine olabildiğince yakın bir yerden geçmesi” olabilir. Bunu sağlamak için önerilen bir yol **en küçük kareler yöntemi**dir (EKK). Bu yöntemin ayrıntıları her istatistik kitabından bulunabilir ama şimdilik şu kadarını söyleyelim. EKK'ye göre, çizimimizdeki her noktanın doğruya olan dikey uzaklıklarının kareleri toplamı, öbür bütün doğrular arasında en düşük olmalıdır.

EKK bu ölçüte göre bir doğru bulurken tek tek her noktayı hesaba katar. O zaman da örneğimizde ötekilere göre aykırı konumdaki noktalar bu doğrunun nereden geçeceğini belirlemede gereğinden çok söz sahibi olabilirler. Tıpkı AO hesaplanırken aşırı uç değerlerin bu ortalamayı kendi değerine doğru çekip olumsuz etkilemesi gibi!

Ortalamalar konusunda bu sakıncanın üstesinden *Ortc* kullanarak gelmiştik. *Ortc*, aşırı değerlerden etkilenmeyen **berk** bir ortalama olduğundan burada da işimize yarayabilir. Bunu ilk öneren Tukey (1977:171) olmuştur. Aşağıda açıklayacağımız biçimde üç *Ortc*'dan yararlanılarak, aşırı uç değerlerden etkilenmeyen **dirençli bir doğru** elde edilebilir.

Dirençli doğru bulma

Bu yöntem uyarınca veriler önce X değişkeninin değerlerine göre büyüklük sırasına konulur. Bu yapılırken aynı gözlem birimine ilişkin X ile Y değerlerinin birbirinden ayrılmasına özen gösterilir. Belli bir yıla ilişkin tüketim değerini alıp başka bir yıla ilişkin gelir değerinin yanına yazmanın saçmalığı ortadadır. Daha sonra gözlemler olabildiğince eşit sayıda değer içeren üç öbeğe ayrılır. Düşük değerlere **alt öbek**, yüksek değerlere **üst öbek**, aradakilere **orta öbek** denir. Gözlem sayısı n , 3'ün tam katıysa bu kolaydır. Öbür durumlarda üç öbeğin gözlem sayılarının bakışık olması uygun olur (Velleman, Hoaglin, 1981:123). t bir tamsayı iken her öbekteki gözlem sayısını şöyle gösterebiliriz.

	<u>Alt öbek</u>	<u>Orta öbek</u>	<u>Üst öbek</u>
$n = 3t$	t	t	t
$n = 3t + 1$	t	$t + 1$	t
$n = 3t + 2$	$t + 1$	t	$t + 1$

Görüldüğü gibi gözlem sayısı 3'ün katından bir fazlaysa bu gözlem orta bölgeye katılmakta, iki fazlaysa uç bölgelere paylaştırılmaktadır. Böylece öbeklerin gözlem sayıları arasındaki bakışım sağlanır. Burada bakışımın uzak bir dağılım oluşursa çok gözlemlili öbeğin dirençli doğruyu belirlemede aşırı bir ağırlık taşımasından lorkulur.

Örneğimizde 17 yılın her biri için bir gözlem çifti bulunduğundan bunları 6, 5, 6 biçiminde ayırmak uygun olur. Gözlem değerlerimiz 18 yıllık olsaydı 6, 6, 6 diye, 19 yıllık olsaydı 6, 7, 6 diye öbeklerdik.

Buna aykırı bir öneri orta öbeğin her zaman en fazla gözlem içermesidir (Erickson, Nosanchuk, 1992:190). Bu durumda $n = 3t + 2$ iken üç öbeğin gözlem sayıları sırasıyla, t , $t+2$, t olması gerekir. Yani 17 gözlem öbeklere 5, 7, 5 biçiminde ayrılmalıdır. Burada orta öbeğe daha fazla ağırlık verme amacı güdülmektedir. Çünkü orta bölge doğru denklemindeki a katsayısının hesabında rol oynar, bu nedenle de yeterince gözlem sayısı içermesi uygun olur. Gözlem sayısının büyük olduğu durumlarda iki öneri arasında pek fark yoktur. Gözlemler azsa herhangi bir öbeğin ağırlığını, gereksiz yere artırmamak daha doğru olur. Böyle durumlarda ilk yol tutulabilir.

Burada bir kez daha dikkatinizi çekelim ki bu ayırmaya, kesinlikle gözlem noktaları X değerlerine göre büyüklük sırasına göre dizildikten sonra gidilmelidir. Bilgisayar yazılımlarıyla yapılan serpilme çizimleri gözlemlerimizi kendiliğinden sıraya sokar ama az sonra anlatacağımız nedenlerle bizim bu işi çizelge üzerinde de yapmamız zorunludur.

Öbekleme sırasında dikkat edilmesi gereken bir nokta vardır. Aynı X değerini taşıyan iki gözlem aynı öbeğe konmalıdır. Bu işlem, öbekler arasındaki gözlem sayısı bakışımını bozabilir. Uygulamada nasıl davranılacağı konusunda önerilebilecek tek şey “sağduyunuzu kullanın” demektir.

Şimdi bu söylediklerimizi örneğimize uygulayalım. Önce gözlemlerimizi yatay eksenindeki değişkenin değerlerine göre Çizelge 11.15'teki gibi büyüklük sırasına dizelim.

ÇİZELGE 11.15 Özel harcanabilir gelirin büyüklüğüne göre sıralanmış veriler

YIL	ÖHG	ÖTH	YIL	ÖHG	ÖTH
1987	27,5	22,5	1996	43,1	32,4
1988	28,0	22,4	1997	45,5	35,1
1989	28,7	22,6	2001	46,8	33,2
1990	32,0	25,0	1999	46,9	33,7
1991	33,0	25,6	1998	48,6	35,5
1992	35,7	27,0	2002	48,7	34,3
1994	36,5	27,6	2000	49,6	36,1
1995	39,2	29,6	2003	50,9	36,1
1993	39,5	29,3			

Çizelge 11.14'teki verilerden düzenlenmiştir.

Dikkat edilirse yılların sırasının değiştiği ama her yıla ilişkin gelirle tüketim değerlerinin birbirinden ayrılmadığı görülür. Şimdi de gözlem noktalarını üç öbeğe ayıralım. Burada “niye başka bir sayı değil de üç” diye sorulabilir. İki öbek kullanarak ilişkinin doğrusal olup olmadığını anlayamayız. Bu nokta aşağıda açıklığa kavuşacaktır. Üçten büyük öbek sayısı da işleri karmaşıktırır.

Her öbeğin X değerlerinden en büyüğü ile en küçüğünün farklarını alarak öbek aralıklarını bulalım. Öbeklerden birinin aşırı geniş olması dirençli davranışın katsayılarının sağlıklı saptanmasını engeller. Burada kullanılabilecek bir ölçüt bir öbek aralığının toplam aralığın yarısından büyük olmamasıdır. Böyle durumlarda, ileride anlatılacak dönüştürmeye gitmek uygun olabilir. Çizelge 11.16'da gözlem değerleri üç öbeğe ayrılmış, öbeklerin aralıkları verilmiştir. Her bir aralık alttan üste doğru sırayla 8,2, 9,0, 4,1, toplam aralık 23,4 katrilyon liradır. Öbek aralıklarından hiçbirisi toplam aralığın yarısını geçmediğine göre örneğimiz bu bakımdan sorunlu değildir.

ÇİZELGE 11.16 Gelir-tüketim verilerinin üç öbeğe ayrılması

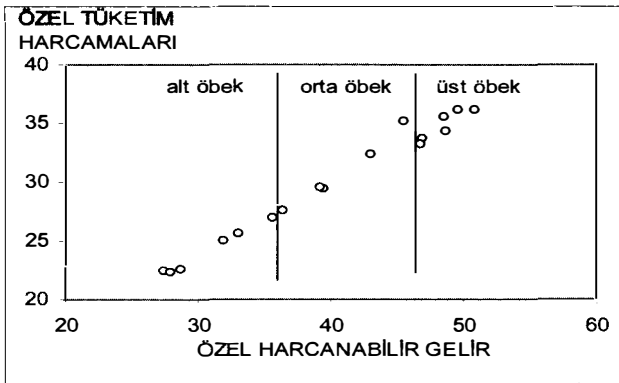
Alt bölge		Orta bölge		Üst bölge	
X	Y	X	Y	X	Y
27,5	22,5	36,5	27,6	46,8	33,2
28,0	22,4	39,2	29,6	46,9	33,7
28,7	22,6	39,5	29,3	48,6	35,5
32,0	25,0	43,1	32,4	48,7	34,3
33,0	25,6	45,5	35,1	49,6	36,1
35,7	27,0			50,9	36,1
8,2	X öbek aralığı	9,0	X öbek aralığı	4,1	
30,35	23,8 Ortç	39,5	29,6 Ortç	48,65	34,8

Bir sonraki adım her öbeğin bir kez X , bir kez Y için $Ortc$ 'lerini ayrı ayrı bulmaktır. $Ortc$ değerleri, eldeki bir gözleme eşit çıkabilir de çıkmayabilir de. Bunun hiç önemi yoktur. Alt öbeğin $Ortc$ değerlerini X_a, Y_a ile, ortadakinin X_o, Y_o ile üstekinin X_u, Y_u ile gösterelim. Bu $Ortc$ değerleri alt öbek için $X_a = 30,35, Y_a = 23,8$, orta öbek için $X_o = 39,5, Y_o = 29,6$, üst öbek için $X_u = 48,65, Y_u = 34,8$ çıkar. Her öbekteki $Ortc$ çiftinin o öbeğe giren noktaları özetleyen yeni bir nokta oluşturduğunu söyleyebiliriz.

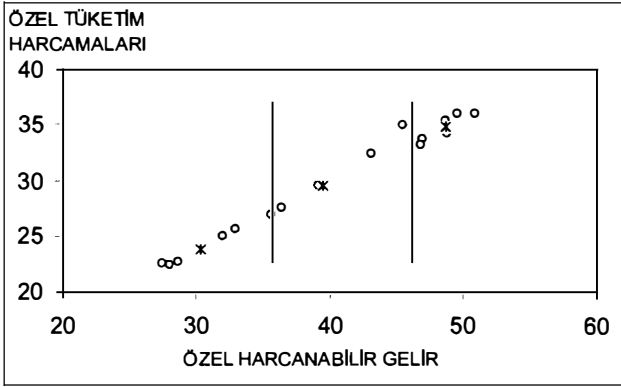
Çizim 11.4 verilerin dağıtıldığı üç öbeği, Çizim 11.5 öbeklerdeki özet noktaların konumunu göstermektedir. Özet noktalar birer * ile işaretlenmiştir. Aslında hiçbir özet nokta gözlem değerlerinden biriyle çakışmamaktadır. Ama orta öbek özet noktası 1993 gözlem noktasına çok yakın olduğundan çizimde aynı yerdeymiş gibi görünmektedir.

Bu çizimlere bakarak yukarıda değindiğimiz bakışım sorunları bir kez daha irdelenebilir. Çizimler bize gerek gözlem sayıları, gerek öbek aralıkları bakımından bakışımın az çok sağlandığını göstermektedir.

Dirençli doğrunun katsayılarını hesaplamadan önce yapmamız gereken bir iş daha var. Bu iki değişken arasındaki ilişkinin doğrusal olduğuna ne kadar güvenilebilir? Çizime bakarak bir ön fikir edinilebilir ama buna daha iyi bir yanıt verebilmek için şöyle bir yol izlenebilir. Alt özet nokta orta özet noktayla, orta özet nokta da üst özet noktayla birer doğru aracılığıyla birleştirilir. Bu iki doğru parçasının eğimlerinin karşılaştırılması bize ilişkinin doğrusal olup olmadığı konusunda bir ipucu verir.



ÇİZİM 11.4 Gelir-tüketim verilerinin üç öbeğe ayrılması



ÇİZİM 11.5 Üç öbekteki özet noktaların bulunması

Önce şunu belirtelim ki iki eğim farklı işaretler taşıyorsa, yani kırık çizginin parçalarından biri yukarı, öbürü aşağı doğru eğimliyse herhangi bir doğrusal ilişkiden söz etmek güçtür. Burada bir tepe ya da bir dip noktası vardır. Eğimler bu noktanın iki yanında ters işaretli olur. Böyle bir durumla karşılaşıldığında, gözlem noktaları arasından bir doğru değil, bir eğri geçirmek daha yerinde olabilir.

İki doğru parçasının eğimleri aynı işaretliyse, yani ikisi de artı ya da ikisi de eksi işaret taşıyorsa işlemlerimizi sürdürürüz. Aynı işareti taşıyan iki eğim değeri birbirine ne kadar yakınsa bizim dirençli doğrumuz da gözlemleri o kadar iyi temsil edecek demektir. Bu yakınlığı iki eğimin birbirine oranını alarak ölçebiliriz. Bu oran 1'e eşit ya da yakın çıkmışsa doğrusallık varsayımı haklılık kazanır. Yok eğer oran 1'den uzaksa o zaman iki değişken arasındaki ilişki doğrusaldır diyemeyiz. Kaba bir kural şöyle olabilir:

İki eğimin arasındaki oran

2'den büyükse
1,25 – 2 arasında
0,8 – 1,25 arasında
0,5 – 0,8 arasında
0 – 0,5 arasında

İlişki

doğrusal değildir,
doğrusal olabilir ya da olmayabilir,
doğrusaldır,
doğrusal olabilir ya da olmayabilir,
doğrusal değildir.

Eğer iki eğimin işaretleri farklıysa eğimler arası oran zaten eksi çıkar. 1,25 – 2 ya da 0,5 – 0,8 arasında ilişkinin doğrusallığı konusunda belirsizlik vardır. Belirsizlikle karşılaşırsa üç ayrı özellik dikkate alınmalı, bunların her biri ayrı ayrı incelenmelidir.

- Gözlem sayısı kaçtır? Bu sayı azsa (sözgelimi 20'den düşükse) modelin yalın olması gerekir. Öyleyse ilişkiyi doğrusal sayma eğilimi ağır basmalıdır. Gözlem sayısı arttıkça doğrusal olmayan bir ilişki göze alınabilir.
- Gözlemler kırık çizginin ne kadar yakınında ya da uzağında yer almaktadır? Eğer gözlemler kırık çizgi çizilmeden de bir dirsek varmış izlenimi veriyorsa ilişkinin doğrusal olması kuşkuludur. Ama gözlem noktaları hayli dağınık serpilmişlerse, yani kırılma belirgin değilse doğrusallık varsayımından vazgeçemeyiz.
- Bulunan oran bu aralıkta hangi sınıra daha yakınsa o seçenek ele alınmalıdır. Sözgelimi eğimlerin oranı 0,54 çıkmışsa doğrusallığa eğilim gösterilebilir, yok eğer 0,78 ise doğrusallık varsayımına pek yüz verilmemelidir.

Örneğimize dönerek bulduğumuz üç özet noktayı anımsayalım. Çizim 11.6'da bu üç özet nokta gösterilmiş, aralarındaki eğimlerin hesabında kullanılan uzaklıklar belirtilmiştir. Bu çizimin bize gösterdiği gibi üst özet noktayla orta özet nokta arasındaki eğim:

$$b_{\bar{u}o} = (Y_{\bar{u}} - Y_o) / (X_{\bar{u}} - X_o) = (34,8 - 29,6) / (48,65 - 39,5) = 5,2 / 9,15;$$

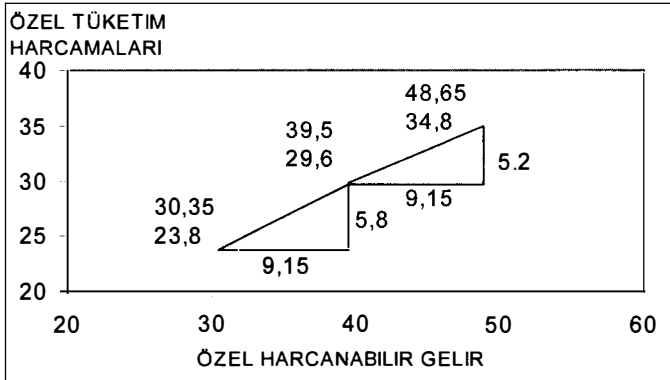
orta özet noktayla alt özet nokta arasındaki eğim:

$$b_{oa} = (Y_o - Y_a) / (X_o - X_a) = (29,6 - 23,8) / (39,5 - 30,35) = 5,8 / 9,15;$$

iki eğimin birbirine oranı:

$$m = b_{\bar{u}o} / b_{oa} = (5,2 / 9,15) / (5,8 / 9,15) = 5,2 / 5,8 = 0,90$$

çıkaktadır. Bu değer yukarıda 0,8 – 1,25 aralığına düştüğüne göre iki değişken arasındaki ilişkinin doğrusallığına güvenebiliriz.



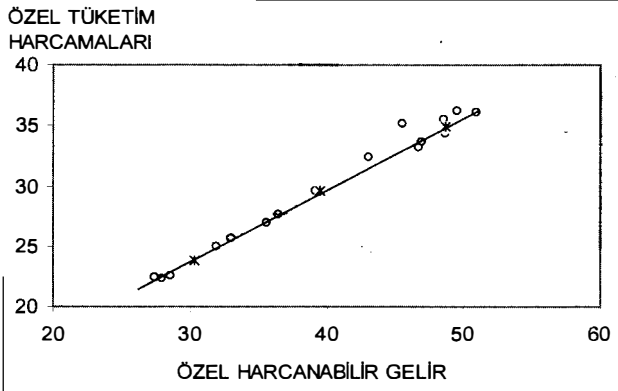
ÇİZİM 11.6 Kırık çizgi eğimleri arasındaki oranın bulunması

Artık bu özet noktalardan yararlanıp dirençli doğruyu elde etmenin zamanı geldi. Önce iki uç öbeğin $Ortc$ değerlerinin gösterdiği konum noktaları arasındaki doğrunun eğimini hesaplayalım. Böylece hem iki uçtan birkaç noktayı hesaplama dışı bırakıp aşırı değerlerin özet noktaların hesaplanmasında yaratabilecekleri olumsuz etkilerden kurtulmuş oluyor, hem de eğimi bulurken birbirinden yeterince uzak iki noktayı kullanmış oluyoruz. Önce dirençli doğru denkleminin b katsayısını, yani eğimini bulalım.

Alt bölgedeki özet noktayla üst bölgedekini, bir doğruyla birleştirilelim. İşte bu doğrunun eğimi bizim aradığımız b değeridir. Bu eğimi lise bilgilerimizle şöyle hesaplayabiliriz:

$$b = \frac{Y_u - Y_a}{X_u - X_a}$$

Örneğimizde $b = (34,8 - 23,8) / (48,65 - 30,35) = 11 / 18,3 = 0,6011 \approx 0,6$ bulunur. Demek ki 100 liralık bir gelir artışı yaklaşık 60 liralık tüketim artışını doğurmaktadır. Çizim 11.7 dirençli doğrunun ilk biçimini göstermektedir. Çizime iyice bakarsak doğrumuzun noktaların pek de ortasından geçmediği izlenimine kapılırız. Gözlemlerimizin pek azı doğrunun altında, pek çoğu doğrunun üstünde kalmıştır. Doğruyu, eğimini hiç değiştirmeden biraz yukarı, orta bölgedeki özet noktanın yakınlarına çekmek iyi olur gibi gelmiyor mu size? Haklısınız, çünkü şimdilik bu doğrunun yalnızca eğimini, yani b katsayısının değerini bulduk. Bize bir de a katsayısının değeri gerekiyor. Bunu nasıl bulabiliriz?



ÇİZİM 11.7 Gelir-tüketim dirençli doğrusunun ilk çizimi

Önce a 'nın anlamı üzerinde düşünelim. Doğru denkleminde X yerine sıfır koyarsak $\hat{Y} = a$ olur. Başka bir deyişle a , Türkiye'nin özel harcanabilir gelirı sıfıra indiğinde yapılabilecek tüketim harcamaları toplamını verir. Gelirdeki her bir liralık artış bu taban tüketimi b kadar artırır. Öyleyse $a = \hat{Y} - bX$ yazabiliriz. Burada X 'e özet noktaların X_{Ortc} değerlerini verip üç tane a değeri elde eder, bunların ortalamasını bularak a değerini kestirebiliriz.

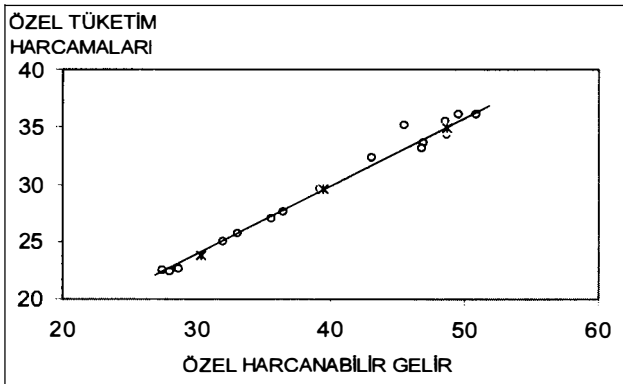
$$\begin{aligned} a &= \frac{1}{3} (Y_{\bar{u}} - bX_{\bar{u}} + Y_o - bX_o + Y_a - bX_a) \\ &= \frac{1}{3} [Y_{\bar{u}} + Y_o + Y_a - b(X_{\bar{u}} + X_o + X_a)] \\ a &= \frac{1}{3} [34,8 + 29,6 + 23,8 - 0,6011(48,65 + 39,5 + 30,35)] \\ &= \frac{1}{3} (88,2 - 0,6011 \times 118,5) = \frac{1}{3} (88,2 - 71,2) \\ &= \frac{1}{3} \times 17 = 5,67 \approx 5,7 \end{aligned}$$

Böylece doğru denklemimizi

$$\hat{Y} = 5,7 + 0,60 X$$

biçiminde buluruz. Çizim 11.8'de bu doğruyu gösterdik. Burada dirençli doğrunun hafifçe yukarı kaymış olduğu görülüyor. Böylece dirençli doğrunun gözlem noktalarının daha ortalarından bir yerden geçtiğini, yani temsil yeteneğini arttırdığını söyleyebiliriz.

Bu doğru denklemi ne anlama gelir? Türkiye'de 1987-2003 arasında ya da bu dönemin dışında yakın yıllarda özel harcanabilir gelir sıfır lira olsaydı özel tüketim harcamaları toplamı 5,7 katrilyon TL olurdu, daha doğrusu olması beklenir. Gelirdeki her 100 liralık artışın tüketim toplamını yaklaşık 60 lira artırması umulur.



ÇİZİM 11.8 Gelir-tüketim dirençli doğrusunun çizimi

Bu denklemdeki X yerine herhangi bir yıldaki özel harcanabilir gelir değerini koyup o yılın özel tüketim harcamalarını kestirebiliriz. Her yıl için bunu yaptığımızda gerçekten harcanan tutarla bu denkleme göre beklenen tutar arasında bir fark bulunabilir. Bu farklara **kalıntı** denir. Kalıntılar, çizimdeki gözlem noktalarıyla dirençli doğru arasındaki dikey uzaklıklardır. Çizim 11.8’de kalıntıların, bir kaç yıl dışında hayli küçük oldukları söylenebilir. Bu yanıltıcı bir görünüm olabilir. Çünkü gerek toplam gelir, gerek toplam tüketim nüfusa bağlı olarak değişiyor olabilir. Başka bir deyişle Türkiye nüfusunun zamanla artması bu iki değişkenin de büyümesine yol açıyor olabilir. Bu nedenle incelenen ilişkinin bir de kişi başına gelir, kişi başına tüketim değerleri kullanılarak yapılması uygun olur. Okuyucudan bunu yapması istenecektir.

Sorgulayıcı veri çözümlemesinin temel adımlarından biri

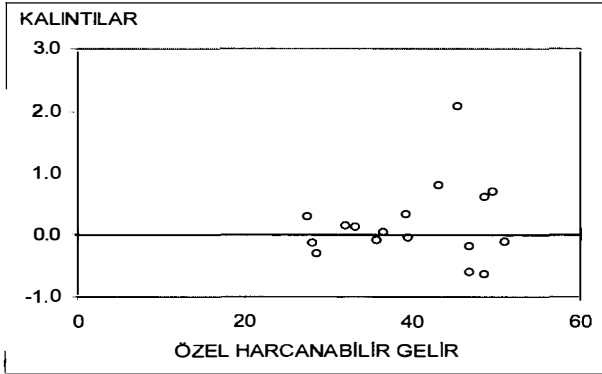
$$\text{kalıntı} = \text{veri} - \text{kestirim} \quad e_i = Y_i - \hat{Y}_i$$

biçiminde hesaplanan kalıntıların incelenmesidir. Kötümser biri için kalıntılar kestirimlerin başarısızlığının kanıtıdır. Bazıları kalıntı yerine **hata terimi** bile der. Ama untmayalım ki hiç kalıntısız kestirimler yapan bir model kuşkuyla karşılanmalıdır. İyimsere göre bu kalıntılar önemli bilgiler içerir. Modelin sunduğu örüntü kalıbına aykırı davranışları ortaya koyar. Aslında her iki görüşün de haklı yanları vardır. İyi modeller iyi kestirimler yapar, kalıntıları küçük olur. Büyük kalıntılar modelin iyi seçilemediği anlamına gelebilir.

Örneğimizin Çizelge 11.17’de bulunan kalıntıları, Çizim 11.9’da gösterilmektedir.

ÇİZELGE 11.17 Özel tüketim harcaması kalıntılarının hesaplanması

YIL	ÖTH	Kestirim	Kalıntı	YIL	ÖTH	Kestirim	Kalıntı
1987	22,5	22,2	0,3	1996	32,4	31,6	0,8
1988	22,4	22,5	-0,1	1997	35,1	33,0	2,1
1989	22,6	22,9	-0,3	1998	35,5	34,9	0,6
1990	25,0	24,9	0,1	1999	33,7	33,8	-0,1
1991	25,6	25,6	0,0	2000	36,1	35,5	0,6
1992	27,0	27,1	-0,1	2001	33,2	33,8	-0,6
1993	29,3	29,4	-0,1	2002	34,3	34,9	-0,6
1994	27,6	27,6	0,0	2003	36,1	36,2	-0,1
1995	29,6	29,2	0,4				



ÇİZİM 11.9 Özel tüketim harcamalarının kalıntıları

Kalıntıları artı işaretli yıllarda beklenenden daha yüksek, eksi olan yıllarda beklenenden daha düşük tüketim harcaması yapılmış demektir. Bu anlamda 1997 yılı beklenenden 2,1 katrilyon daha çok, 2001, 2002 yılları beklenenden 0,6 katrilyon lira daha az özel tüketim harcamasının yapıldığı yıllar olarak göze çarpmaktadır. 1996 yılının özel tüketim harcamaları tutarı da beklenenin 0,8 katrilyon lira üstündedir. Buradan 1990'lı yılların ortalarından sonra tüketim harcamalarının beklenene göre daha büyük sapmalar gösterdiği sonucuna varabiliriz. Acaba bu ekonominin daha az istikrarlı olduğu anlamına gelir mi? 1994'te IMF'yle yapılan anlaşmanın bu istikrarsızlıkla bir ilgisi olabilir mi? Sorular, sorular! Yanıtları nerede?

Yukarıdaki açıklamalarımız arasında a katsayısını, “hiçbir özel harcanabilir gelirin olmadığı bir durumda özel tüketim harcamalarının düzeyi” diye tanımlamıştık. Bu gerçekte pek karşılaşılamayacak bir durumdur. Bunun yerine orta öbeğin $Ortc$ 'sı kadar gelir gösteren bir yıl olsaydı, bu yılın dirençli doğrumuza göre beklenen tüketim düzeyi

$$a^* = a + bX_o$$

eşitliğiyle kestirilebilirdi. Herhangi bir yılın geliriyle bu yılın gelir düzeyi arasındaki farkı b katsayısıyla çarparsak $b(X - X_o)$ elde ederiz. Bunun anlamı bu iki yılın tüketimleri arasında olması beklenen farktır. Öyleyse bu gelir farkını, orta öbeğin $Ortc$ 'sı kadar geliri olan bir yılın beklenen tüketimine (a^*) eklersek, dirençli doğru denklemimiz

$$\hat{Y} = a^* + b(X - X_o)$$

biçiminde yazılabilir. Bu denklem de Çizim 11,8'deki doğruyu verir. Öncekinden tek farklı yanı şudur. Yeni denklemde, bir yılın geliri değil, bu

gelirin $Ortc'$ dan farkı dikkate alınır. Bir başka deyişle, X yerine $X - X_o$ farklarını geçirmek, çizimdeki dik eksen X_o kadar sağa kaydırmak anlamına gelir.

a katsayısının hesaplanmasına hiç gerek kalmadan a^* katsayısı aşağıdaki yolla da bulunabilir.

$$a^* = \frac{1}{3} [Y_{\bar{u}} - b(X_{\bar{u}} - X_o) + Y_o + Y_a - b(X_a - X_o)]$$

Uygulamada her iki yol aynı sonucu verir ama a yerine a^* ile çalışmak, aşağıda anlatılacak işlemler sırasında kolaylık sağlar. Çünkü birden çok kez yapılan yinelenmelerde a^* 'ın bir kez hesaplanması yeterlidir. Oysa a 'nın her aşamada yeniden hesaplanması gerekir.

Örneğimizdeki a^* değeri her iki yolla da aynı bulunmuştur.

$$a^* = 5,7 + 0,6 (39,5) = 29,4$$

$$\begin{aligned} a^* &= \frac{1}{3} [34,8 - 0,6 (48,65 - 39,5) + 29,6 + 23,8 - 0,6 (39,35 - 39,5)] \\ &= \frac{1}{3} (29,31 + 29,6 + 29,29) = \frac{1}{3} (88,2) = 29,4 \end{aligned}$$

Öyleyse dirençli denklemimiz yeni biçimiyle şöyle de yazılabilir:

$$\hat{Y} = 29,4 + 0,6(X - 39,5)$$

Bunun da $\hat{Y} = 5,7 + 0,6 X$ biçimindeki ilk denkleme çevrilebileceği kolayca görülebilir. Demek ki hangi denklem kullanılırsa kullanılsın sonuç değişmez.

Bazı durumlarda dirençli doğru denklemi katsayılarının biraz daha işlenmesi gerekebilir. İlk adımda bulduğumuz b kestirimi en uygun değer olmayabilir. Bunu anlamamanın yolu, kalıntıları (yani doğru denklemine göre beklenen tüketimle, yapılan tüketim düzeyi arasındaki farkları) bulmak, bu e değerleriyle X değerleri (yani gelir düzeyleri) arasındaki ilişkiyi yansıtacak yeni bir doğru çizmektir. Başka bir deyişle, daha önce gördüğümüz dirençli doğruyu Çizim 11.9'daki noktalar için çizip bu doğrunun eğimini hesaplamaktır. Bu yeni doğrunun eğimi sıfırsa ya da sıfıra çok yakınsa ilk doğru yeterince iyidir. Yeni eğim sıfırdan farklıysa, bu değer eski eğime eklenerek düzeltme yapılmalıdır. Çizim 11.9'da görüldüğü kadarıyla bu eğim sizce sıfıra yakın çıkar mı? Hesaplayıp bulunuz.

Kalıntılarla hesaplanan yeni doğrunun eğimi sıfıra iyice yakın çıkıncaya kadar gerekli düzeltmeler yinelenir. Uygulamada ilk eğimin %1'ine ya da %0,1'ine ulaşıncaya kadar hesaplama sürdürülür (Emerson, Hoaglin, 1983b:134). Kimi veri kümelerinde bu işlemlerde güçlüklerle karşılaşılabilir. Sözelimi, yeni eğimlerin sıfıra yaklaşması çok yavaş olabilir ya da bir kaç adım sonra yeni eğimler sıfıra yaklaşacak yerde ters işaretli iki değer arasında gidip gelebilir. Böyle durumlarda ne yapılabileceği konusunda

öneriler geliştirilmiştir (Emerson, Hoaglin, 1983b:137-147) ama MINITAB yazılımı bütün bu güçlüklerin üstesinden gelmektedir.

Şimdi başka örneklerde çıkabilecek sorunları inceleyelim. Birinci örnek futbolla ilgili. 2002-03 döneminde Türkiye süper ligindeki takımların attıkları gol sayısı ile aldıkları puan arasındaki ilişkiyi ele alalım. Bu konuyu seçmemizin nedeni az önceki örnekte doğan ağır havayı biraz olsun yumuşatabilmek. Gol-puan verileri Çizelge 11.18’de gösterilmiştir.

Atılan gol sayısı alınan puanı etkiler, öyleyse gol sayısına X , puana Y diyelim. Verilerimizi X ’lere göre büyüklük sırasına dizip Çizelge 11.19’deki gibi üç öbeğe ayıralım. Toplam aralık $76 - 32 = 44$ olduğuna göre her bir öbeğin aralığı bunun yarısından azdır. Ama üst öbek aralığının öbür öbek aralıklarına oranla hayli büyük olması dikkat çekicidir. Veriler Çizim 11.10’da gösterilmiştir.

ÇİZELGE 11.18 Türkiye süper ligi atılan gol, alınan puan sayıları (2002-03)

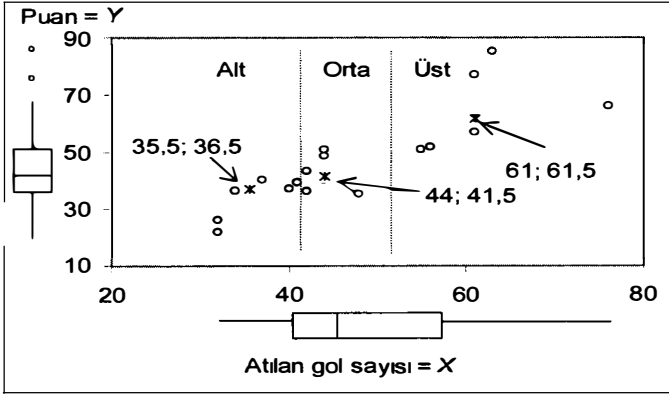
Takım	A	P	Takım	A	P	Takım	A	P
Beşiktaş	63	85	Trabzonspor	44	51	Elazığspor	40	37
Galatasaray	61	77	Ankaragücü	44	49	Denizlispor	34	36
G.Birliği	76	66	İstanbulspor	42	43	Bursaspor	42	36
Gaziantep	61	57	Diyarbakırspor	37	40	Altay	48	35
Malatyaspor	56	52	Adanaspor	44	40	Göztepe	32	26
Fenerbahçe	55	51	Samsunspor	41	39	Kocaelispor	32	22

Alıştırma 10.15’teki verilerden alınmıştır. A: Atılan gol sayısı, P: Puan.

ÇİZELGE 11.19 Atılan gol sayısına göre öbeklere ayırma

Alt öbek	X	Y	Orta öbek	X	Y	Üst öbek	X	Y
Kocaelispor	32	22	Bursaspor	42	36	Fenerbahçe	55	51
Göztepe	32	26	İstanbulspor	42	43	Malatyaspor	56	52
Diyarbakır	34	36	Adanaspor	44	40	Gaziantep	61	57
Denizlispor	37	40	Ankaragücü	44	49	Galatasaray	61	77
Elazığspor	40	37	Trabzonspor	44	51	Beşiktaş	63	85
Samsunspor	41	39	Altay	48	35	G.Birliği	76	66
Aralık	9	—	Aralık	6	—	Aralık	21	—
Ortc	35,5	36,5	Ortc	44	41,5	Ortc	61	61,5

Çizelge 11.18’den düzenlenmiştir.



ÇİZİM 11.10 Gol-puan verilerinin öbekli serpilme çizimi

Dirençli doğrunun hesaplanmasına geçmeden ilişkinin doğrusallığı konusu incelenmelidir. Özet noktaları birleştiren kırık çizgilerin eğimlerini bulup birbirine oranlayalım. Üst özet noktayla orta özet nokta arasındaki eğim:

$$b_{üo} = (Y_{\bar{u}} - Y_o) / (X_{\bar{u}} - X_o) = (61,5 - 41,5) / (61 - 44) = 20 / 17;$$

orta özet noktayla alt özet nokta arasındaki eğim:

$$b_{oa} = (Y_o - Y_a) / (X_o - X_a) = (41,5 - 36,5) / (44 - 35,5) = 5 / 8,5;$$

iki eğimin birbirine oranı:

$$m = b_{üo} / b_{oa} = (20 / 17) / (5 / 8,5) = 1,176 / 0,588 = 2$$

çıkılmaktadır. Bu değer de bize ilişkinin doğrusal sayılamayacağını göstermektedir. Zaten daha önce üst öbek aralığının öbür öbek aralıklarına oranla hayli büyük olmasına dikkat çekmiştik. Aslında bu bize ilişkinin doğrusal olamayabileceği konusunda ilk uyarıydı diyebiliriz. Bu durumda ne yapılabilir? Bu sorunun yanıtını aşağıdaki altbölümde ele alacağız.

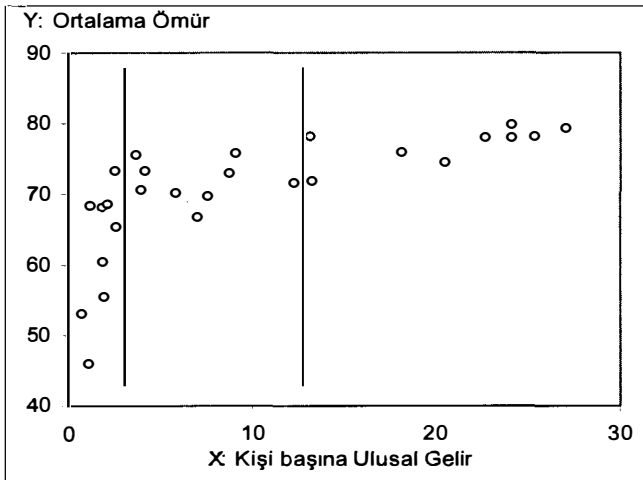
Başka bir örnek alıp çeşitli ülkelerin (1000 \$ cinsinden ölçülmüş) kişi başına ulusal gelirleriyle o ülkelerde yaşayan insanların (yıl olarak) ortalam ömrü arasındaki ilişkiye bakalım. Burada gelir verileri satınalma gücü eşdeğerine göre hesaplanmıştır. Çizelge 11.20 ile Çizim 11.11 bu verileri göstermektedir.

ÇİZELGE 11.20 Kişi başına ulusal gelir (1000 \$)-ortalama ömür (yıl) (2001)

Ülke	X Gelir	Y Ömür	Ülke	X Gelir	Y Ömür
Madagaskar	0.8	53.0	Belarus	7.6	69.6
Burkina Faso	1.1	45.8	Malezya	8.8	72.8
Tacikistan	1.2	68.3	Şili	9.2	75.8
Pakistan	1.9	60.4	Macaristan	12.3	71.5
Solomon Ada.	1.9	68.1	Malta	13.2	78.1
Sudan	2.0	55.4	S. Arabistan	13.3	71.9
Moldovya	2.2	68.5	Portekiz	18.2	75.9
Gürcistan	2.6	73.4	BAE	20.5	74.4
Grenada	2.7	65.3	Singapur	22.7	77.8
Jamayka	3.7	75.5	İngiltere	24.2	77.9
Çin	4.0	70.6	İsveç	24.2	79.9
Lübnan	4.2	73.3	Almanya	25.4	78.0
Türkiye	5.9	70.1	Kanada	27.1	79.2
Rusya	7.1	66.6			

UNDP, 2003.

Görüldüğü gibi yatay eksene göre öbek aralıkları çok dengesizdir. Üst öbek aralığı 13,9; 26,3 olan toplam aralığın yarısından büyüktür. İlişkinin doğrusal olmadığı açıktır. Öbeklerin özet noktalarının hesaplanmasına girilmemelidir. Bu durumda ne yapılması gerektiği bir sonraki altbölümde ele alınacaktır.

**ÇİZİM 11.11** Kişi başına ulusal gelir-ortalama ömür (2001)

11.3 İKİ DEĞİŞKENLİ İNCELEMEDE DÖNÜŞTÜRME

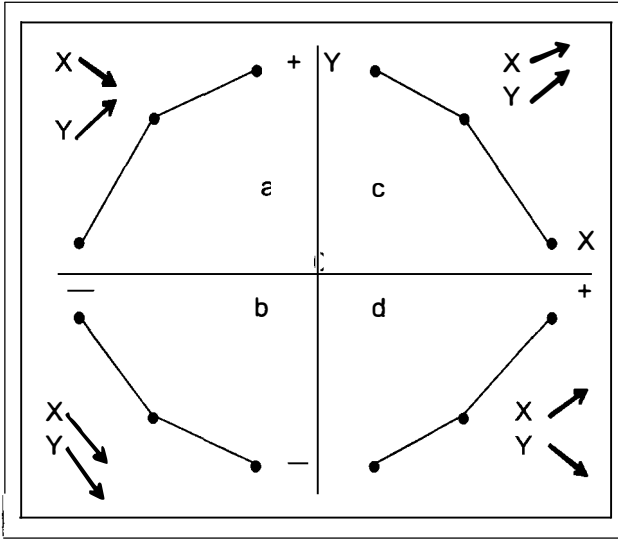
İki değişken arasındaki ilişkinin doğrusal olup olmadığını nasıl anlayabiliriz? Doğrusal olmayan bir ilişkiyle karşı karşıyaysak ne yapabiliriz? Bu sorulara teker teker yanıt arayalım.

Dirençli doğruyu bulmak için attığımız adımlardan ilki verilerden üç tane özet nokta çıkarmak olmuştur. Bu noktalardan iki uçtakini kullanıp dirençli doğrunun eğimini kestirmiştik. Eğimi hesaplamaya kalkışmadan önce bu üç noktadan alt bölgedekini ortadakiyle onu da üst bölgedekiyile birleştirirsek kırık bir çizgi elde edeceğimizi, bu kırık çizgini iki parçasının eğimlerinin karşılaştırılmasıyla ilişkinin doğrusal olup olmadığı konusunda bir ipucu elde edeceğimizi görmüştük.

Önce şunu belirtelim ki iki eğim farklı işaretler taşıyorsa, yani kırık çizginin parçalarından biri yukarı, öbürü aşağı doğru eğimliyse herhangi bir doğrusal ilişkiden söz etmek güçtür. Böyle bir durumla karşılaşıldığında yapılacak iş $\hat{Y} = a + bX$ denklemi yerine doğrusal olmayan bir denklem geçirmek olabilir. Sözelimi $\hat{Y} = a + bX + cX^2$ ya da $\hat{Y} = a + bX + cX^2 + dX^3$ gibi kimi denklemler bazı durumlarda işe yarayabilir. Öte yandan iki eğimin ikisi de aşağı ya da yukarı doğruysa, daha önce gördüğümüz dönüştürme yollarına başvurulabilir. Amacımız dönüştürülmüş veriler için kırık çizgiyi yeniden oluşturup yeni eğimlerin oranını bulmaktır.

Bütün gözlem noktalarını dönüştürmek, özellikle bilgisayar kullanılmadığında, zaman alıcı olabilir. O nedenle yalnız üç özet noktayı dönüştürmekle yetinebiliriz. Dönüştürülmüş yeni özet noktalarla iki eğimi, daha sonra bunlar arasındaki oranı hesaplayıp az önce verdiğimiz aralıklardan hangisine düştüğüne bakabiliriz. Ama önce nasıl bir dönüştürme uygulayacağımıza karar vermeliyiz.

Daha önce de söylediğimiz gibi, dönüştürmeyi ancak kırık çizgilerin eğimleri aynı işareti taşıyorsa yapabiliriz. Böyle durumlarda kırık çizgiler Çizim 11.12'deki dört değişik biçimi alabilirler.



ÇİZİM 11.12 Kırık çizgilerin dört biçimi

Bu aşamada **dönüştürme merdiveni** yeniden devreye girebilir. Bu merdiveni 10. bölümde tek değişkenli veri kümelerini dönüştürmede kullanmıştık. Aynı merdiven iki değişkenli veri kümelerinde de işe yarar. Çizimin dört bölgesinde X , Y 'lerin yanındaki okların yönü, aynı zamanda dönüştürme merdiveninde gidilecek yönü gösterir. Bu merdiven iki değişken için Çizelge 11.21'de yeniden gösterilmiştir.

Daha önce dönüştürme merdivenini anlatırken bu değişkenlerin 1. kuvvetinin hiçbir dönüştürme yapılmayacağı anlamına geldiğini görmüştük. Çizim 11.12'deki okların yönü yanındaki değişkenin üssünün 1'den büyük olup olmadığını yansıtmaktadır. Yukarı doğru bir ok 1. kuvvetten büyük, aşağı doğru bir ok 1. kuvvetten küçük üs kullanılması gereğini yansıtır. Yalnız uygulamada dönüştürme yaparken hemen iki değişken birden dönüştürülmez. Genellikle tek bir değişkenin dönüştürülmesiyle yetinilir. X ya da Y 'den hangisinin dönüştürüleceği konusunda konunun içeriğine bakılarak karar verilir. Eğer tek değişkenin dönüştürülmesi yeterli olmazsa ancak o zaman ikinci değişken için de uygun bir dönüştürme düşünülebilir.

ÇİZELGE 11.21 Tukey dönüştürme merdiveni

r	X'	Y'	Dönüştürme
3	X^3	Y^3	Küp
2	X^2	Y^2	Kare
3/2	$\sqrt{X^3}$	$\sqrt{Y^3}$	3 / 2'nci kuvvet
1	X	Y	Dönüştürme yok
1/3	$\sqrt[3]{X}$	$\sqrt[3]{Y}$	Küpkök
½	$\sqrt{X}$	$\sqrt{Y}$	Karekök
0	$\log X, \ln X$	$\log Y, \ln Y$	Logaritma
- ½	$-1 / \sqrt{X}$	$-1 / \sqrt{Y}$	Karekökten sonra ters
- 1	$-1 / X$	$-1 / Y$	Ters
- 2	$-1 / X^2$	$-1 / Y^2$	Kareden sonra ters

Dönüştürme işlemini bir örnek üzerinde uygulayalım. DİE'nin düzenlediği 1987 için 1,00 kabul edilen kentsel yerler tüketici fiyatları indeksini alalım. İndeksin yıllara göre dökümü Çizelge 11.22'dedir. Burada X değişkeni zaman, Y değişkeni fiyat indeksidir. En alt satır *Ortc*'ları göstermektedir.

ÇİZELGE 11.22 Kentsel yerler tüketici fiyatları indeksi 1987:1,00

Yıl: X	İndeks: Y	Yıl: X	İndeks: Y	Yıl: X	İndeks: Y
1987	1	1993	21,312	1998	518,684
1988	1,737	1994	43,959	1999	847,989
1989	2,836	1995	85,117	2000	1308,034
1990	4,546	1996	152,706	2001	2013,483
1991	7,545	1997	282,494	2002	2915,796
1992	12,831			2003	4227,904
5	Aralık	4	Aralık	5	Aralık
1989,5	3,691	1995	85,117	2000,5	1660,759

DİE, Türkiye İstatistik Yıllığı, çeşitli yıllar.

Öbek aralıklarının görelî büyüklüklerinde sorun yoktur. Öyleyse özet noktaları birleştiren kırık çizgilerin eğimlerini incelemeye geçebiliriz. Üst özet noktayla orta özet nokta arasındaki eğim:

$$b_{üo} = (Y_{ü} - Y_o) / (X_{ü} - X_o) = (1660,759 - 85,117) / (5,5) = 286,48,$$

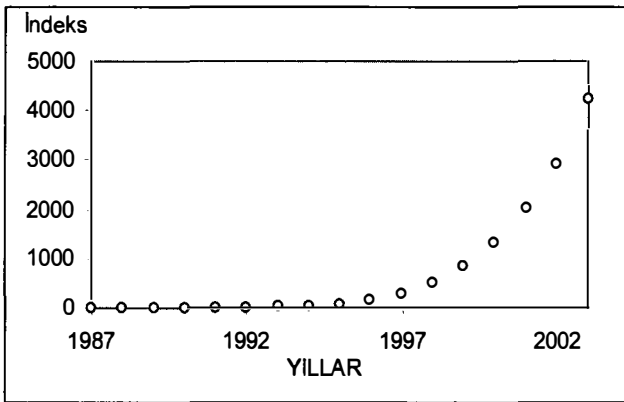
orta özet noktayla alt özet nokta arasındaki eğim:

$$b_{oa} = (Y_o - Y_a) / (X_o - X_a) = (85,117 - 3,691) / (5,5) = 14,80$$

İki eğimin birbirine oranı $286,48 / 14,80 = 19,4$ çıkmaktadır. Bu da çok yüksektir, ilişki doğrusal olamaz. Zaten Çizim 11.13'e bakınca ilişkinin eğrisel olduğu açıkça görülmektedir.

Öyleyse dönüştürme gerekir. Kırık çizginin kolları Çizim 11.12'nin (d) bölümündeki biçime uymaktadır. Yıl eksenindeki değerler tam bakışık, üstelik eşit aralıklarla sıralandığına göre dönüştürmenin indeks sayılarına uygulanması kaçınılmaz görünmektedir. Bu sayıların kökleri, logaritmaları, tersleri alınarak çeşitli dönüştürmeler denenebilir. Bu işlemlerde kolaylık sağlamak amacıyla Çizelge 11.23 düzenlenmiştir. Okuyucu bu çizelgenin gözlerindeki değerleri doğrulamalıdır.

Çizelgenin en alt satırına bir göz atacak olursak 1'e en yakın eğim oranının logaritmali dönüştürme olduğu anlaşılır. Öyleyse logaritma kuulanıldığında en başarılı dönüştürme sağlanacaktır. e tabanlı logaritma yerine 10 tabanlı logaritma kullanılırsa eğimler oranının aynı çıkacağını göstermek okuyucuya bırakılmıştır.



ÇİZİM 11.13 Kentsel yerler tüketici fiyatları indeksi 1987:1,00

ÇİZELGE 11.23 Tüketici indeksiyle önüştürme denemeleri

Öbek	X	Y	$\sqrt{Y}$	$\ln Y$	$-1000/Y$
Üst	2000,5	1660,759	40,75	7,42	-0,60
Orta	1995	85,117	9,23	4,44	-1,17
Alt	1989,5	3,691	1,36	1,31	-27,92
<i>Üst eğim</i>		286,48	5,73	0,54	0,09
<i>Alt eğim</i>		14,80	1,43	0,57	0,49
Eğim oranı		19,4	4,0	0,95	0,2

Peki böyle bir dönüştürme ne anlama gelir? İster inanın ister inanmayın ama logaritma almakla işleri karıştırmadık. Tam tersine fiyatların zaman içindeki değişmesinin kuralını anlama yolunda önemli bir adım attık. Dönüştürme yapmadan önce ilişkiyi yansıtacak dirençli doğru denklemi $Y = a + bX$ biçimindeyken dönüştürmeyle $\ln Y = a + bX$ oldu. Bu denklemde, X bir birim değiştiğinde (bir yıllık bir süre geçtiğinde) $\ln Y$ 'deki değişimin b kadar olduğu hemen görülür. Demek ki $\ln Y$ her yıl sabit bir artış göstermektedir. Matematikle arası iyi olan bir okuyucu, bunun Y 'de sabit bir yüzde artışa denk geldiğini bilir, ama biz yine de bir açıklama yapalım. X değişkenindeki küçük değişimin $\ln Y$ üzerinde yaratacağı etki $\ln Y$ 'nin X 'e göre türeviyle gösterilir, bu türev de b 'ye eşittir. Bu sözlerden pek bir şey anlamadıysa zararı yok. Şunu bilin yeter: $\ln Y = a + bX$ kalıbındaki bir denklemde X bir birim arttığında Y değişkeni $\%100b$ kadar artar. Bunu örneğimiz için şöyle dile getirebiliriz: “Kentsel yerler tüketici fiyatlarının 1987-2003 arasındaki yıllık ortalama artışı sabit bir oran olan $\%100b$ kadardır.” Şimdi dirençli doğrunun eğim katsayısı olan bu sabit oranı hesaplayalım.

$$b = (7,42 - 1,31) / (2000,5 - 1989,5) = 6,11 / 11 = 0,56$$

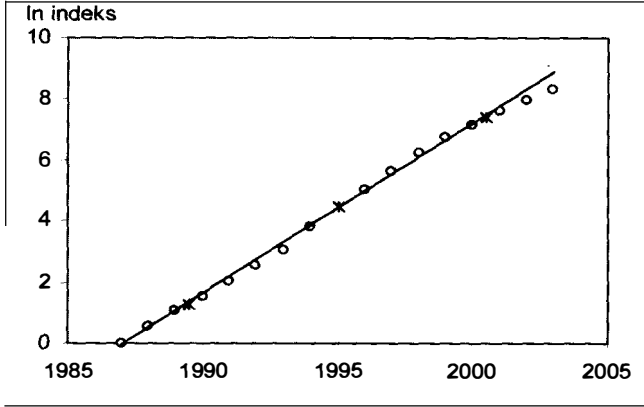
Öyleyse 1987-2003 arasında Türkiye’de kentlerde yıllık enflasyon oranı yaklaşık $\%56$ kadar olmuştur. Bu da Çizim 11.14’teki dirençli doğrunun eğimidir. Aynı doğru, özgün (logaritmasız) indeks sayıları kullanılarak da çizilebilir (Çizim 11.15). Dik eksenindeki farklı iki ölçek dışında iki çizim aynıdır.

Dirençli doğrunun eğim katsayısı dışında bir de a katsayısının hesaplanması gerekir. Burada a , sıfır yılındaki indeksi gösterir. İsa’nın doğum yılındaki indeks bizi ilgilendirmedigine göre a ’yı bulmanın anlamsızlığı ortadadır. Öyleyse sıfır yılı yerine sözgelimi *Ortc* yıl olan 1995’i kullanırsak, daha önce açıkladığımız a^* katsayısını bulabiliriz. Bu katsayının 4.37

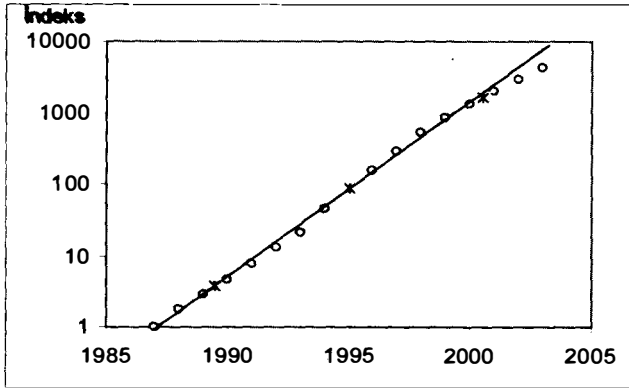
olduğunu okuyucu hesaplayabilmelidir. Öyleyse dirençli denklemimiz 1995 yılındaki kentsel yerler tüketici indeksini 4,37 olarak kestirmektedir. Demek ki dirençli doğru denklemi

$$\ln Y = 4,37 + 0,56 (X - 1995) \quad \text{ya da} \quad \hat{Y} = e^{4,37+0,56(X-1995)}$$

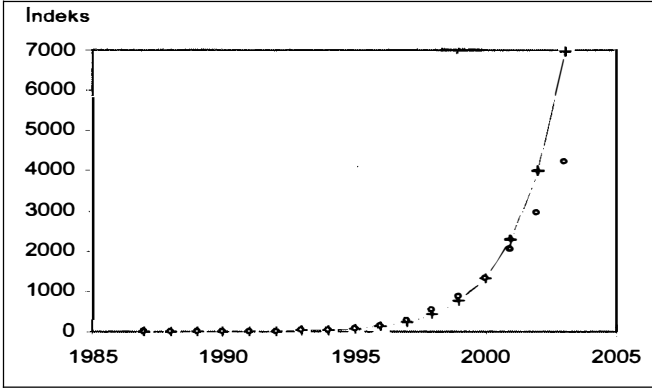
biçiminde yazılabilir.



ÇİZİM 11.14 Tüketici fiyatları indeksi için dirençli doğru



ÇİZİM 11.15 Tüketici fiyatları indeksi için dirençli doğru



ÇİZİM 11.16 Tüketici indeksinin dirençli denklemle kestirimi

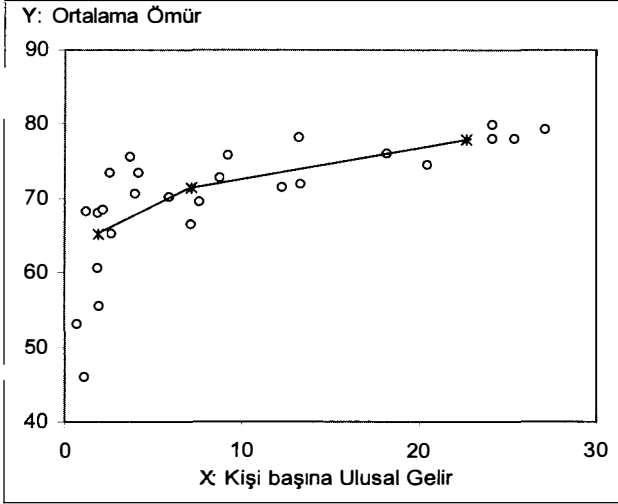
Çizim 11.16 ise logaritma alınmamış özgün verileri logaritmasız bir dik eksenle göstermektedir. Bu çizim aslında ilk serpilme çiziminin aynısıdır, tek farkı bulduğumuz dirençli eğriyle yapılan kestirimleri de içermesidir. Yeni dirençli denklemin, eğrisel bir ilişki gösteren verileri çok daha iyi temsil ettiği açıktır. Son birkaç yıldır gözlenen fiyat indekslerinin dirençli denkleme göre beklenen değerlerin altında kaldığı göze çarpmaktadır. Bu da enflasyonun son yıllarda hızını yitirdiği anlamına gelir. Okuyucu bu çizimdeki kestirimlerden hiç olmazsa birkaçını yukarıdaki denklemlerden birini kullanarak hesaplayabilmelidir.

Şimdi bir önceki altbölümde yarım bıraktığımız bir örneğe, çeşitli ülkelerde kişi başına gelir düzeyiyle beklenen ortalama ömür ilişkisine dönelim. Daha önce X değişkenine göre öbek aralıklarının çok dengesiz olduğunu görerek üç özet noktayı hesaplamaya girişmemiş, ne yapacağımıza karar vermeyi dönüştürme işlemlerini öğrendikten sonraya bırakmıştık. Eğer okuyucu öbeklerin özet noktalarını hesaplayıp bunları birleştiren doğru parçalarının eğimleri arasındaki oranı bulsaydı 0,34 elde edecekti. Bunun doğrusal bir ilişkinin göstergesi olmadığını da görecekti.

Acaba bu örnekte de dönüştürme işlemlerine başvurulsa hem öbek aralıklarının dengesizliği giderilebilir, hem ilişki doğrusal bir dirençli denklem kalıbına oturtulabilir mi? Buna benzer bir soruyla, tek değişkenli veri kümelerinin dönüştürülmesini incelerken karşılaşmış, uygun bir dönüştürmeyle hem her kümedeki çarpıklığın giderildiğini, hem de bütün kümelerin yayılgılığının birbirine yaklaştığını görmüştük. Aynı sonucu burada da elde edebilir miyiz? Deneyelim.

Çizim 11.11'e dönüp özet noktaları işaretleyelim. Kırık çizgiyi oluşturalım. Bu kırık çizginin Çizim 11.12'deki dört biçimden (a) bölgesinde

uyduğunu görürüz. Yani her iki parçanın da eğimi artı işaretlidir, bunlardan soldakinin eğimi, sağdakinden daha büyüktür. Az önce bunlar arasındaki oranın 0,34 bulunacağını da söylemiştik. Bütün bunlar Çizim 11.17’de yer almaktadır.



ÇİZİM 11.17 Gelir-ömür verileri için kırık çizgi

Artık dönüştürme denemelerimize geçebiliriz. Kırık çizginin türü bize, X' e 1'den küçük, Y' ye 1'den büyük kuvvetler vermemin uygun olacağını söyler. Ömrün karesini almak fazla anlamlı olmayacağına göre gelirin logaritmasını ya da tersini almayı deneyebiliriz. Çünkü $\ln X$ ile Y arasında doğrusal bir ilişki kurulursa gelirdeki her %1 artışa karşılık ömrün sabit bir süre kadar uzadığı yorumuna ulaşırız. Öte yandan gelirin tersini almak, yani $1/X$ kullanmak, kişi başına gelir deyimini belli bir gelir tutarını bölüşen kişi sayısına dönüştürmesi anlamına gelir. Bu iki denemenin kırık çizginin eğim oranı üzerindeki etkisini Çizelge 11.24'teki gibi gösterebiliriz.

ÇİZELGE 11.24 Gelir-ömür verileriyle dönüştürme denemeleri

Öbek	X	Y	$\sqrt{X}$	$\ln X$	$-100/X$
Üst	22,7	77,9	4,76	3,12	-4,41
Orta	7,1	71,5	2,66	1,96	-14,1
Alt	1,9	65,3	1,38	0,64	-52,6
<i>Üst eğim</i>	<i>0,41</i>		<i>3,05</i>	<i>5,52</i>	<i>1,52</i>
<i>Alt eğim</i>	<i>1,19</i>		<i>4,84</i>	<i>4,70</i>	<i>0,16</i>
Eğim oranı	0,34		0,63	1,17	9,5

Anlaşılabileceği gibi karekök dönüştürmesi yetersiz kalmakta, ters alma işlemi de dönüştürmeyi abartmaktadır. Gelirin logaritması kullanılacak olursa ilişki kalıbı doğrusal olabilecektir. Söz konusu dönüştürmeyi yapıp $\hat{Y} = a + b \ln X$ dirençli denkleminin katsayılarını hesaplamak okuyucuya bırakılmıştır. Bir ipucu verelim. Eğim katsayısı b bulunduğunda, üst parçanın eğimi 5,52 ile alt parçanın eğimi 4,70 arasında bir değer elde edilmelidir. Daha sonra kalıntılar kullanılarak bu değer daha da işlenebilir.

İşler her zaman bu örneklerdeki gibi yolunda gitmeyebilir. Dönüştürme sonucunda uygun bir kalıp hemen elde edilemeyebilir. Böyle durumlarda gerek X , gerek Y değişkeni için değişik kalıplarla denemelere girmek gerekebilir. Futbol liginden aldığımız örneği yarım bıraktığımız yerden sürdürülim. Daha önce iki eğim arasındaki oranı 2 olarak bulmuş, atılan gol sayısı ile alınan puan arasında doğrusal ilişki bulunmadığı kanısına varmıştık. Okuyucu aşağıda dökümü verilen dönüştürme sonuçlarını ilgili *Ortc*'lar üzerinden hesaplayıp doğrulama yoluna gitmelidir. Burada verilen değerler MINITAB yazılımıyla elde edilmiştir. Okuyucu hesaplamaları daha kaba yapacağı için sonuçları yaklaşık bulması yeterlidir.

Y	X	<u>Eğimler arası oran</u>
ln Puan	Atılan gol	1,53
Ters Puan	Atılan gol	1,19
Puan	Atılan golün karesi	1,51
ln Puan	Atılan golün karesi	1,16

Uygulanacak dönüştürmeyi seçme kararı okuyucuya bırakılmış olup kendisinden sonuçları nasıl yorumlayacağını açıklaması da beklenmektedir.

Okuma Önerileri

- Erickson, B. H., Nosanchuk, T. A. (1992) *Understanding Data*, 2nd. ed. Open University Press (185-216).
- Hartwig, F., Dearing, B. E. (1979) *Exploratory Data Analysis*, Sage (31-51).
- Hoaglin, D. C., Mosteller, F., Tukey, J. W. (1983) *Understanding Robust and Exploratory Data Analysis*, Wiley (129-243).
- Hoaglin, D. C., Mosteller, F., Tukey, J. W. (1985) *Exploring Data Tables, Trends, and Shapes*, Wiley (67-119).
- Marsh, C. (1988) *Exploring Data*, Polity Press (180-217).
- Siegel, A. F. (1988) *Statistics and Data Analysis*, Wiley (373-413).
- Tukey, J. W. (1977) *Exploratory Data Analysis*, Addison-Wesley (331-399).
- Velleman, P. F., Hoaglin, D. C. (1981) *Applications, Basics, and Computing of Exploratory Data Analysis*, Duxbury Press (93-147, 201-242).

ALİŞTIRMALAR

- 11.1 Çizelge 7.1'deki verilerden yararlanıp simgeli çizelgeyi düzenleyiniz. Yorumlarınızı yazınız.
- 11.2 Çizelge 7.9'deki verilerden yararlanıp her bir yıl için beş kutu çizimini yan yana gösteriniz. Buradaki yorumlamalarınızla Çizelge 11.1 için yazılanları karşılaştırınız.
- 11.3 Çizelge 11.3'teki simgeleri elde etmek için gerekli dal-yaprak gösterimini düzenleyip *Ortc*, *D*, *E* değerlerini bulunuz. Çeşitli simgelerin hangi aralıklara konacağını gösteriniz.
- 11.4 Çizelge 11.12'de dışalma bağımlılık katsayılarını için ortanca işlemleriyle kalıntılar bulunmuştu. Bu kalıntıların dal-yaprak gösterimini düzenleyiniz. Gördüklerinizi açıklayınız.
- 11.5 Çizelge 11.12'deki kalıntılarla bir simgeli çizelge düzenleyerek hangilerinin göze batır biçimde büyük ya da küçük olduklarını gösterip yorumlayınız.
- 11.6 Çizelge 11.11'de bölgesel satışlara ilişkin varsayımsal verilerin ortanca işlemlerini ikinci adımını bitirmiştik. Üçüncü adımın işlemlerini yaparak yeni *Ortc*'ları bulunuz. Bunların yeterince küçük olup olmadıklarını inceleyiniz.
- 11.7 Bölge satışlarla ilgili örneğimizde ortanca alma işine önce satırlardan başladık. Peki sütunlardan başlasaydık sonuç aynı olur muydu? Deneyerek sonuçları karşılaştırınız.
- 11.8 Çizelge 11.2'de yer alan İTO Ücretliler Geçinme İndeksi'ndeki aylık % değişimlere ortanca işleme yöntemi uygulandığında ortak değer, yıllar içindeki gelişmeyi yansıtan satır etkileri, mevsim dalgalanmalarının göstergesi

olan sütun etkileri bulunmuştu. Bu bulguların doğru olup olmadığını gösteriniz. Kalıntıları hesaplayınız. Bütün bunları yorumlayınız.

- 11.9 Bir futbol liginde takımların oynadıkları maç, yengi, yenilgi, beraberlik, atılan gol, yenilen gol, alınan puan sayılarını gösteren puan cetveline iki boyutlu çizelge diyebilir miyiz? Neden?
- 11.10 Türkiye nüfusunun zamanla artması hem harcanabilir toplam gelirin, hem toplam tüketim harcamalarının büyümesine yol açıyor olabilir. Bu nedenle iki değişken arasındaki ilişkiyi kişi başına gelir, kişi başına tüketim değerlerini kullanarak bir kez daha inceleyiniz.
- 11.11 Çizim 11.10'da Beşiktaş ile Galatasaray'ın yalnızca puanlara göre dışadüşen oldukları görülmektedir. Bunu doğrulayınız.
- 11.12 Çizelge 11.2'de kentsel yerler tüketici indeksi verileri için çeşitli dönüştürme denemeleri gösterilmiştir. Bu çizelgenin her bir gözündeki değeri doğrulayınız. Logaritmali dönüştürmede e yerine 10 tabanlı logaritma kullanılırsa eğimler arasındaki oranının aynı kalıp kalmadığını inceleyiniz.
- 11.13 Özel tüketim harcamalarıyla harcanabilir gelir arasındaki ilişkiyi yansıtan dirençli doğrunun yaptığı kestirimlerle gerçekleşmiş gözlemler arasındaki kalıntılar Çizim 11.9'da gösterilmiştir. Bu çizime bakarak kalıntı noktaları arasından geçirilecek bir dirençli doğrunun eğiminin sıfıra yakın olup olmadığını belirleyebilir misiniz? Gerekli hesaplamaları yapıp görüşünüzün doğruluğunu deneyiniz.
- 11.14 Çizim 11.16'da tüketici indeksinin dirençli denklemle kestirimi gösterilmiştir. Bu kestirimlerden hiç olmazsa bir kaçını bulunan dirençli denklemi kullanarak hesaplayınız.
- 11.15 Çizim 11.11'de çeşitli ülkelerde kişi başına gelire ortalama ömür değişkenlerinin serpilme çizimi verilmişti. Bu çizimdeki üç özet noktanın konumunu hesaplayıp bunları birleştiren doğru parçalarının eğimleri arasındaki oranı bulup yorumlayınız.
- 11.16 Çizelge 11.24'teki denemeler sonunda gelire ömür arasındaki ilişkiye uygun bir kalıbın $\hat{Y} = a + b \ln X$ biçiminde olabileceği saptanmıştı. Bu dönüştürmeyi yapıp dirençli denklemin katsayılarını hesaplayınız.
- 11.17 Türkiye süper liginde futbol takımlarının attıkları gol sayısı ile aldıkları puan arasındaki ilişkinin doğrusal olmadığı sonucuna varmıştık. Aşağıda dökümü verilen dönüştürme sonuçlarını ilgili $Ortc$ 'lar üzerinden hesaplayıp doğrulayınız. Burada verilen değerler MINITAB yazılımıyla elde edilmiştir. Okuyucu hesaplamaları daha kaba yapacağı için sonuçları yaklaşık bulması yeterlidir

<i>Y</i>	<i>X</i>	Eğimler arası oran
In Puan	Atılan gol	1,53
Ters Puan	Atılan gol	1,19
Puan	Atılan golün karesi	1,51
In Puan	Atılan golün karesi	1,16

Uygulanacak dönüştürmeyi seçiniz. Bulgularınızı yorumlayınız

11.18 Türkiye süper liginde takımların attığı gol sayısı yerine atılan – yenilen gol farkını geçirip ilişkiyi bir de böyle inceleyelim.

Atılan – yenilen gol sayısına göre öbeklere ayırma

Üst öbek	<i>X</i>	<i>Y</i>	Orta öbek	<i>X</i>	<i>Y</i>	Alt öbek	<i>X</i>	<i>Y</i>
Beşiktaş	42	85	Ankaragücü	2	49	Samsunspor	-18	39
G.Birliği	36	66	İstanbulspor	-5	43	Elazığspor	-19	37
Galatasaray	34	77	Denizlispor	-5	40	Bursaspor	-19	36
Gaziantep	20	57	Adanaspor	-10	40	Altay	-21	35
Fenerbahçe	13	51	Diyarbakır	-13	36	Göztepe	-25	26
Malatyaspor	11	52				Kocaelispor	-34	22
Trabzonspor	11	51						
Öbek aralığı	31	–	Öbek aralığı	15	–	Öbek aralığı	16	–
Ortc	20	57	Ortc	-5	40	Ortc	-20	35,5

Çizelgedeki bilgilerden yararlanarak ayrıntılı bir inceleme yapınız. İlişkinin doğrusal olup olmadığını saptayınız. Gerekli çizimlere yer vermeyi unutmayınız. Dönüştürme gerekiyorsa kalıbını bulunuz. Dönüştürülmüş dirençli denklemi elde ediniz. Kestirimleri hesaplayıp kalıntıları bir çizimle gösteriniz. Her aşamada gerekli yorumları yazınız.

11.19 Türkiye savunma bütçesini oluşturan çeşitli harcamaların yüzde payları aşağıdaki çizelgede gösterilmiştir. Bu verilerle simgeli çizelge düzenleyerek yorumlayınız. Aynı verileri bir kez ortanca işleme yöntemiyle irdeleyiniz.

Yıl	Personel	Techizat	Diğer	Yatırım	Transfer
1990	44,1	19,1	33,1	0,6	3,0
1991	49,6	14,4	32,4	1,1	2,5
1992	52,1	14,1	30,7	0,5	2,6
1993	54,0	14,5	28,5	0,5	2,4
1994	50,0	12,8	35,3	0,3	1,6
1995	43,7	12,5	36,0	0,5	7,4
1996	40,6	19,8	35,0	0,8	3,8
1997	41,1	24,6	33,0	0,7	0,7
1998	41,2	20,1	37,7	0,5	0,5
1999	42,2	20,2	36,0	0,7	0,4
2000	40,1	18,3	40,5	0,7	0,4
2001	42,3	22,7	33,8	0,7	0,5

Günlük-Şenesen, 2002:59.

11.20 Çeşitli ülkelerde kamu istihdamının toplam nüfusa (%) oranı aşağıdaki çizelgede gösterilmiştir. Hem simgeli çizelge, hem ortanca işleme yöntemiyle bu verileri yorumlayınız.

	Türkiye	İspanya	Çek C.	İtalya	İngiltere
Merkez	18	18	48	34	31
Yerel	3	20	23	25	34
Eğitim	4	9	7	21	14
Sağlık	3	8	3	12	16
Polis	3	3	-	6	3
Silahlı K.	8	2	11	5	4
KİT	8	-	29	-	-

TÜSİAD, 2003:40.

11.21 Türkiye’de coğrafi bölgelerde çeşitli gelir gruplarının gelirden aldıkları % paylar aşağıdadır. Simgeli çizelge düzenleyerek bu verileri yorumlayınız. Ortanca işleme yönteminin bulgularının anlamını açıklayınız.

Bölge	En zengin Üst % 1	Yüksek % 2 – 6	Üst-orta % 7 - 25	Alt-orta % 26 - 58	Düşük % 59 – 100
Marmara	24	17	25	22	12
Ege	11	18	24	36	11
Akdeniz	14	18	21	35	12
İç A.	8	21	27	32	12
Karadeniz	13	18	19	39	11
Doğu A.	6	14	24	46	10
Güneydoğu A.	9	7	15	45	24

Sönmez, 2001:39-70.

11.22 Aşağıdaki çizelge 1972’den bu yana olimpiyat oyunları erkekler hız koşularında birinci olanların (saniye cinsinden) derecesini gösterir. Hem simgeli çizelge hem ortanca işleme yöntemini kullanarak bu verileri inceleyiniz.

YIL	Yer	100 m	200 m	400 m	800 m	1500 m
1972	Münih	10,14	20,00	44,66	105,90	216,30
1976	Montreal	10,06	20,23	44,26	103,50	219,17
1980	Moskova	10,25	20,19	44,60	105,40	218,74
1984	Los Angeles	9,99	19,80	44,27	103,00	212,53
1988	Seul	9,92	19,75	43,87	103,45	215,96
1992	Barselona	9,96	20,01	43,50	103,66	220,12
1996	Atlanta	9,84	19,32	43,49	102,58	215,78
2000	Sidney	9,87	20,09	43,84	105,08	212,07

AnaBritannica, 2000: 17, 86.

11.23 Türkiye’de 1983-2002 yıllarında kişi başına GSYİÜ (satınalma gücü eşdeğeri-ne göre 1000 \$) değerleriyle kişi başına dışalım (\$) değerleri aşağıdadır. Kişi başına dışalımın kişi başına üretime bağlı olduğu varsayımıyla bir serpilme çizimi düzenleyip üç öbeğin özet noktalarını işaretleyiniz. Bu noktaları birleştiren kırk çizginin iki kolunun eğimleri arasındaki orana bakarak bu iki değişken arasındaki ilişkinin doğrusal olup olmadığını inceleyiniz. Uygun bir kalıp bularak dirençli doğru denklemini belirleyiniz. Kestirimleri yapıp kalıntıları elde ediniz. Bütün bulgularınızı yorumlayınız.

Yıl	GSYİÜ	Dışalım	Yıl	GSYİÜ	Dışalım	Yıl	GSYİÜ	Dışalım
83	2,9	193	90	4,7	397	97	6,5	761
84	3,2	219	91	4,8	368	98	6,3	707
85	3,4	225	92	5,1	392	99	5,9	615
86	3,6	218	93	5,6	496	00	6,2	808
87	4,0	269	94	5,3	385	01	5,7	590
88	4,1	267	95	5,7	580	02	6,2	720
89	4,2	288	96	6,1	696	03	5,7	916

DPT, 2004, Tablo 1.4, 3.4.

Çeşitli üniversitelerin değişik bölümlerinden öğrencilere “Astrolojiye inanıyor musunuz?” diye sorulmuş, aşağıdaki gibi yanıtlar alınmıştır:

- Simgeli çizelgeyle bu yanıtları irdeleyiniz.
- Ortanca işleme yöntemiyle aynı yanıtları yorumlayınız.

Yanıt	A bölümü	B bölümü	C bölümü	D bölümü
İnanıyorum	2	9	15	2
Olabilir	28	24	10	16
İnanmıyorum	26	21	20	22

Helvacıoğlu, E., Kızıler, R., 2001:8-27.

11.24 Aşağıda rassal seçilmiş 15 ülkenin Birleşmiş Milletler Kalkınma Programı Örgütü UNDP tarafından 100 üzerinden hesaplanmış gayrisafi ulusal ürün (GSUÜ), ortalama ömür, eğitim indeksleri gösterilmiştir.

Ülke	GSUÜ	Ömür	Eğitim	Ülke	GSUÜ	Ömür	Eğitim
Belçika	92	89	99	Çin	50	75	80
Barbados	83	86	90	G. Afrika	60	64	87
Bahreyn	82	80	85	Cezayir	66	74	69
Estonya	74	76	94	Zambiya	34	27	68
Dominik	67	70	79	Senegal	34	47	36
Tayland	69	75	84	Ruanda	36	25	57
Jamaika	60	84	78	Nijer	34	33	15
Türkiye	69	74	77				

UNDP, 2001:141-144,

- Bu değişkenler arasındaki ikili ilişkileri görmek için birer serpilme çizimi düzenleyiniz. İlişkiler doğrusal mı?

- b) Herhangi iki değişken arasındaki ilişki doğrusal değilse dönüştürmeyle doğrusallaştırınız.
- c) Uygun dirençli doğru denklemlerini bularak yorumlayınız.
- d) Kalıntıları inceleyerek gerekçelerinizi açıklayınız.

11.25 ABD’de üniversiteler çeşitli ölçütlere göre değerlendirilip sıralanır. Aşağıda rassal seçilmiş dört üniversitenin dört daldaki lisansüstü öğretimlerinin indeks değerleri verilmiştir.

- a) Simgeli çizelge,
- b) Ortanca işleme yöntemlerini kullanarak yorumlayınız

Üniversite	İşletme	Mühendislik	Hukuk	Tıp
Harvard	100	56	91	100
Stanford	100	90	92	57
Pennsylvania	99	49	78	70
Columbia	94	50	86	64

11.26 Büyük bir şirkette çalışanların son bir yılda fazla mesaiye kaldıkları günler aşağıdaki gibidir.

- a) Simgeli çizelge,
- b) Ortanca işleme yöntemleriyle bu verileri yorumlayınız.

Gün sayısı	Erkek	Kadın
0	18	26
1-5	45	47
6-20	27	29
20 +	29	12

11.27 Bölgelerimizde bir günde trafiğe kaydolan araç sayıları aşağıda verilmiştir.

- a) Simgeli çizelge,
- b) Ortanca işleme yöntemleriyle bu verileri yorumlayınız.

	Mrm	Ege	Akdz	Krdz	İÇA	DĞA	GdA
Araba	263	112	78	51	143	18	22
Kamyon	24	16	15	16	25	5	6

11.28 Türkiye’de üretilen otobüs, minibüs, midibüslerin üretim indeksi 1970:100 olacak biçimde aşağıda gösterilmiştir.

- a) Serpilme çizimiyle verileri gösteriniz.

- b) İndeksin zaman içindeki seyrini açıklayınız.
- c) Bu verilerle dirençli doğru denklemi hesaplamak sizce uygun mu? Neden?
- d) Dirençli doğru denklemini bulabiliyorsanız katsayıların anlamını açıklayınız. Kalıntıları hesaplayarak yorumlayınız.

Yıl	Üretim	Yıl	Üretim	Yıl	Üretim	Yıl	Üretim
1983	527	1987	555	1991	896	1995	653
1984	627	1988	468	1992	1009	1996	970
1985	588	1989	469	1993	1124	1997	1362
1986	627	1990	727	1994	462	1998	1414

DPT , 1997:140.

11.29 14 ülkenin kişi başına gelir düzeyleriyle gelir dağılımındaki eşitsizliği yansıtan Gini katsayıları aşağıdadır. Gelir düzeyi arttıkça gelir dağılımı düzelmekte midir? Gerekli incelemeleri yapıp ayrıntılı yanıtınızı yazınız.

Ülke	Gelir \$	Gini	Ülke	Gelir \$	Gini
Avusturya	26310	23,1	Moritanya	1650	37,3
Grundi	580	33,3	Nikaragua	2100	60,3
Kolombiya	5890	57,1	Polonya	9030	31,6
Mısır	3690	28,9	Slovenya	17390	28,4
Yunanistan	16940	32,7	Türkiye	7030	41,5
İsrail	19320	35,5	Yemen	780	33,4
Lao	1530	37,0	Zimbabve	2590	56,8

World Bank, 2001232-235.

*Su yüzünde dolaşır yanlışlar çerçöp gibi,
İnci bulmak isteyen taramalıdır dibi.*

*Errors, like straws, upon the surface flow,
He who should search for pearls must dive below.*

J. Dryden, *All for Love* (1678)

- Acel, A., Sounderpandian, J. (2002) *Complete Business Statistics*, McGraw-Hill.
- Akalın, S. (1977) *İşletme İstatistiği*, 3. b., Ege Üniversitesi.
- Aköz, Y., Büyükeröglü, C., Ercan, H., Erkan, İ., Gökçe, H. E., Kayserilioğlu, A., Kale, R., Özdemir, M. Y., Öztürk, M., Saçaklı, H., Sağlam, H., Sel, S., Sarıalp, R., Şenesen, Ü., Sönmez, M., Sönmez, N., Uzunbaş, Y. (1994) *Beden Eğitimi ve Spor Faaliyetlerine Katılan İTÜ Öğrencilerinin Yapısal ve Fonksiyonel Kapasitelerinin İncelenmesi*, İTÜ.
- Aköz, Y. (1995) (der.) *Türkiye ve Olimpiyat Sempozyumu*, İTÜ.
- Allen, R. G. D. (1964) *Statistics for Economists*, Hutchinson.
- AnaBritannica (2000) Ana Yayıncılık..
- Anderson, D. R., Sweeney, D. J., Williams, T. A. (1996) *Statistics for Business and Economics*, West Publishing Co.
- Arıcı, H. (1988) *İstatistik: Yöntemleri ve Uygulamaları*, Meteksan.
- ATM (1988) *Oto Lastik Bayileri Araştırması*, Basılmamış rapor.
- Babbie, E. (1992) *The Practice of Social Research*, 6th ed. Wadsworth.
- Balcı, M. (2003) "Türkiye 2002 yılında bilimsel makale sayısında 22'nciliğe yükseldi" *Cumhuriyet Bilim- Teknik*, 8 Şubat, 829:8-10.
- Balkan, N., Savran, S. (der.) (2002) *The Ravages of Neo-Liberalism: Economics, Society and Gender in Turkey*, Nova Science Publishers, Inc.
- Barnett, V. (1991) *Sample Survey Principles and Methods*, Edward Arnold.
- Barzun, J., Graff, H. F. (1993) *Modern Araştırmacı*, 7. b., (Çeviren: Dilber, F.) TÜBİTAK.
- Baş, T (2001) *Anket*, Seçkin.
- Başar, T. (1993) *Sözlerin ve Zaman Örgütlenmesinin Sürekli Değişimi – Tekerlemeler Üzerine Bir Alan Çalışması*, basılmamış yüksek lisans tezi, Mimar Sinan Üniversitesi.
- Batuhan, H. (1993) *Bilim ve Şarlatanlık*, Yapı Kredi.

- Becker, W. E. (1997) *Statistics for Business and Economics Using Microsoft Excel 97*, S. R. B: Publishing.
- Berger, P. D. , Maurer, R. E. (2002) *Experimental Design with Applications in Management, Engineering, and the Sciences*, Duxbi-Thomson Learning.
- Bernal, J. D. (1965) *Science in History*, 3rd ed., Penguin, vol. 1 – *The Emergence of Science*, vol. 2 – *The Scientific and Industrial Revolutions*, vol. 3 – *The Natural Sciences in Our Time*, vol. 4 – *The Social Sciences: Conclusion*.
- Bernard, H. R. (2000) *Social Research Methods*, Sage.
- Berndt, E. R. (1991) *The Practice of Econometrics: Classic and Contemporary*, Addison-Wesley.
- Black, T. R. (1999) *Doing Quantitative Research in the Social Sciences*, SAGE
- Blalock, Jr., H.M. (1972) *Social Statistics*, 2.b., McGraw-Hill.
- Bora, T. (2004) “Futbolda istatistik, istatistikte futbol”, *Radikal Futbol*, 20 Ocak.
- Box, G. E., Hunter, W. G., Hunter, J. S. (1978) *Statistics for Experimenters*, Wiley.
- Brookes, B. C., Dick, W. F. L. (1967) *Introduction to Statistical Methods*, Heinemann.
- Bulutay, T., Timur, S., Ersel, H. (1971) Türkiye Gelir Dağılımı 1968, AÜ Siyasal Bilgiler Fakültesi.
- Celasun, M. (1986) “Income distribution and domestic derms of drade in Turkey, 1978-1983: Estimated measures of inequality and poverty”, *ODTÜ Gelişme Dergisi*, 13,1-2:193-216.
- Celasun, M. (1989) “Income distribution and employment aspects of Turkey’s post-1980 adjustment”, *ODTÜ Gelişme Dergisi*, 16,3-4:1-31.
- Chambers, J. M., Cleveland, W. S., Kleiner, B., Tukey, P. A. (1983) *Graphical Methods for Data Analysis*, Wadsworth & Brookes.
- Chatfield, C. (1986) “Exploratory data analysis” *European Journal of Operational Research*, 23:5-13.
- Chatfield, C. (1988) *Problem Solving: A Statistician’s Guide*, Chapman and Hill.
- Chatterjee, S., Handcock, M. S., Simonoff, J. S. (1995) *A Casebook for A First Course in Statistics and Data Analysis*, Wiley.
- Cillov, H. (1976) *İstatistik Tekniği ve Uygulaması*, İstanbul Üniversitesi.
- Cleveland, W. S. (1985) *The Elements of Graphing Data*, Wadsworth.
- Cleveland, W. S., McGill, M. E. (1988) *Dynamic Graphics for Statistics*, Wadsworth & Brookes.
- Cooper, D. R., Schindler, P. S. (2003) *Business Research Methods*, McGraw-Hill.
- Çıngı, H. (1990) Örnekleme Kuramı, Hacettepe Üniversitesi.
- Çoker, D., Karaçay, T. (1983) *Matematik Terimleri Sözlüğü*, Türk Dil Kurumu.
- Çömlekçi, N. (1985) *İstatistiğin Tarihçesi*, Anadolu Üniversitesi.
- Çömlekçi, N. (1988) *Deney Tasarımı ve Çözümlemesi*, Anadolu Üniversitesi.

- Dalal, S. R., Fowlkes, E. B., Hoadley, B. A. (1989) "Risk analysis of space shuttle: Pre-Challenger prediction of failure" *Journal of the American Statistical Association*, aktaran Chatterje vd. (1995:33-36).
- David, F. N. (1955) "Dicing and gaming (A note on the history of probability)", *Biometrika*, 42:1-15, Pearson, Kendall (1970) içinde 1-17.
- David, F. N. (1962) *Games, Gods and Gambling*, Charles Griffin.
- Demirkan, Ö. (2003) "İstatistik doçenti 10 trilyona en yakın rakamları açıklıyor" *Vatan*, 16 Aralık.
- Devore, J, Peck, R. (1993) *Statistics: The Exploration and Analysis of Data*, 2nd ed., International Thompson Publishing.
- Devore, J, Peck, R. (2001) *Statistics: The Exploration and Analysis of Data*, 4th ed., Duxbury.
- Diamantopoulos, A., Schlegelmilch, B. B. (1997) *Taking the Fear out of Data Analysis*, Dryden Press.
- DİE (Devlet İstatistik Enstitüsü) (1973) *Toplumsal ve Ekonomik Gelişmenin 50 Yılı*, DİE Matbaası.
- DİE (1990) *1987 Hanehalkı Gelir ve Tüketim Harcamaları Sonuçları – Gelir Dağılımı*, DİE Matbaası.
- DİE (1993) *Türkiye İstatistik Yıllığı 1991*, DİE Matbaası.
- DİE (1994) *Gayri Safi Milli Hasıla: Kavram, Yöntem ve Kaynaklar*, DİE Matbaası
- DİE (1995) *İl ve Bölge İstatistikleri 1993*, DİE Matbaası.
- DİE (1997a) *1994 Hanehalkı Gelir Dağılımı Anketi Sonuçları*, DİE Matbaası.
- DİE (1997b) *İller İtibariyle Gayrisafi Yurtiçi Hasıla: 1987–1994*, DİE Matbaası.
- DİE (1998) *İstatistik Terimleri Sözlüğü*, DİE Matbaası.
- DİE (1999) *Osmanlı Devleti'nde Bilgi ve İstatistik*, DİE Matbaası.
- DİE (2001a) *İstatistik Göstergeler 1923 –1998*, DİE Matbaası.
- DİE (2001b) *Yayın Kataloğu, 1927-2000*, DİE Matbaası.
- DİE (2002a) *Türkiye İstatistik Yıllığı 2001*, DİE Matbaası.
- DİE (2002b) *Yayınlar ve Hizmetler Kataloğu 2002*, DİE Matbaası.
- DİE (2003a) *İstanbul, 2000 Genel Nüfus Sayımı*, DİE Matbaası.
- DİE (2003b) "İller itibariyle GSYİH sonuçları", Haber bülteni.
- Doane, D: P., Mathieson, K., Tracy, R. L. (2001) *Visiaul Statistics 2.0*, McGraw-Hill, Irwin.
- DPT (Devlet Planlama Teşkilatı) (1966) *Gelir Dağılımı Araştırması*.
- DPT (1976) *Gelir Dağılımı 1973*.
- DPT (1997) *Ekonomik ve Sosyal Göstergeler 1950-1978*.
- DPT (2004) *Ekonomik ve Sosyal Göstergeler (1950-2003)*.
- du Toit, S. H. C., Steyn, A. G. W., Stumpf, R. H. (1986) *Graphical Exploratory Data Analysis*, Springer-Verlag.
- Duverger, M (1980) *Sosyal Bilimlere Giriş*, 2. b., (Çeviren: Oskay, Ü.) Bilgi.
- Efes Pilsen (2002) *Türkiye Profili 2*.

- Efron, B., Thisted, R. (1976) "Estimating the number of unseen species: How many words did Shakespeare know?" *Biometrika*, 63:435-447, aktaran Siegel (1988:2).
- Ehrenberg, A.S.C.(1975) *Data Reduction: Analysing and Interpreting Statistical Data*, Wiley.
- Ehrenberg, A.S.C. (1982) *A Primer in Data Reduction*, Wiley.
- Emerson, J. D. (1983) "Mathematical Aspects of Transformation", Hoaglin, Mosteller, Tukey (1983) içinde 247 – 282.
- Emerson, J. D., Hoaglin, D. C. (1983a) "Stem-and-leaf displays" Hoaglin, Mosteller, Tukey (1983) içinde 7 – 30.
- Emerson, J. D., Hoaglin, D. C. (1983b) "Resistant lines for y versus x " Hoaglin, Mosteller, Tukey (1983) içinde 129 – 165.
- Emerson, J. D., Hoaglin, D. C. (1983c) "Analysis of Two-way Tables by Medians" Hoaglin, Mosteller, Tukey (1983) içinde 166 – 210.
- Emerson, J. D., Stoto, M. A. (1983) "Transforming data" Hoaglin, Mosteller, Tukey (1983) içinde 97 – 128.
- Emerson, J. D., Strenio, J. (1983) "Boxplots and batch comparison" Hoaglin, Mosteller, Tukey (1983) içinde 58 – 96.
- Eraybar, D. (2004) "Düşün ama istatistiksiz de kalma" *Cumhuriyet Spor*, 4 Şubat
- Erickson, B. H., Nosanchuk, T. A. (1992) *Understanding Data*, 2nd. ed. Open University Press.
- Ertek, T. (1996) *Ekonometriye Giriş*, 2. b.; Beta.
- Esmer, Y., Fişek, H., Kalaycıoğlu, E. (1986) *Türkiye’de Sosyo-ekonomik Özellikler, Hane Gelirleri, Harcamaları ve Sosyo-ekonomik İhtiyaçlar Üzerine Araştırma Dizisi, Cilt I: Sosyolojik ve Demografik Yapı ile İlgili Bulgular, Metodoloji ve Uygulama*, TÜSİAD.
- Fisher, R. A. (1971) *Collected Papers of R. A. Fisher*, J. H. Bennett (ed.) The University of Adelaide, aktaran Salzburg (2001:320).
- Frankfort-Nachimas, C., Nachimas, D. (1992) *Research Methods in the Social Sciences*, 4th ed. St. Martin’s.
- Goodall, C. (1983) "M-Estimators of Location" Hoaglin, Mosteller, Tukey (1983) içinde 339 – 400.
- Genceli, M. (2001) *Ekonometri ve İstatistik İlkeleri*, 2. b. Filiz.
- Gonick, L., Smith, W. (1993) *The Cartoon Guide to Statistics*, Harper Perennial.
- Göçmençelebi, K. (1962) *İstatistik Metodları*, Başbakanlık İstatistik Genel Müdürlüğü.
- Göçmençelebi, K. (1976) *İstatistik Metodları*, 4. b., Ongün Kardeşler.
- Graziano, A. M., Raulin, M. L. (1993) *Research Methods*, 2nd ed. HarperCollins.
- Groebner, D. F., Shannon, P. W., Fry, P. C., Smith, K. D. (2001) *Business Statistics: A Decision Making Approach*, 5. b., Prentice Hall.

- Gülçür, F. K. (1946) *Sigorta Muameleleri Matematiği ve Prensipleri*, Milliyet Kitabevi.
- Günlük-Şenesen, G. (1988) "Örnekleme yöntemleri", *Pazarlama Planlaması İçinde Araştırma Teknikleri Semineri*, Türkiye Sanayi Sevk ve İdare Enstitüsü (TÜSSİDE) içinde 127-136.
- Günlük-Şenesen, G. (2002) *1980-2001 Türkiye'de Savunma Harcamaları ve Ekonomik Etkileri*, TESEV.
- Gürsel, S., Levent, H., Selim, R., Sarıca, Ö. (2000) *Türkiye'de Bireysel Gelir Dağılımı ve Yoksulluk*, TÜSİAD.
- Gürtan, K. (1987) *İstatistik ve Araştırma Metodları*, 4. b., İstanbul Üniversitesi.
- Gürüz, K., Şuhubi, E., Şengör, A. M. C., Türker, K., Yurtsever, E. (1994) *Türkiye'de ve Dünyada Yüksek Öğretim, Bilim ve Teknoloji*, TÜSİAD.
- Hamilton, L. C. (2002) *Statistics with STATA 7.0*, Duxbury.
- Hartwig, F., Dearing, B. E. (1979) *Exploratory Data Analysis*, Sage.
- Hays, W. L. (1988) *Statistics*, 4th ed. Holt, Rinehart and Winston.
- Hazine Müsteşarlığı (2004) *Hazine İstatistikleri 1980-2003*.
- Helvacıoğlu, E., Kızılar, R. (2001) "Safata anketini sunarken" *Bilim ve Ütopya*, Ekim, 8-27.
- Hoaglin, D. C. (1973) "Letter values: A set of selected order statistics" Hoaglin, Mosteller, Tukey (1983) içinde 33-55.
- Hoaglin, D. C., Mosteller, F., Tukey, J. W. (1983) *Understanding Robust and Exploratory Data Analysis*, Wiley.
- Hoaglin, D. C., Mosteller, F., Tukey, J. W. (1985) *Exploring Data Tables, Trends, and Shapes*, Wiley.
- Huff, D. (1973) *How to Lie with Statistics*, Penguin (Çeviren: Koparan, E. (1995) *İstatistik ile Nasıl Yalan Söylenir?*, Sarmal).
- İdil, O. (1980) *Örnekleme Teorisi ve İşletme Yönetiminde Uygulanması*, İÜ İşletme Fakültesi.
- İnal, C., Esensoy, Ö., Sözer, M. T., Erar, M. A., Çetinel, B., Kutluk, H. (1983) *İstatistik Terimleri Sözlüğü*, Türk Dil Kurumu.
- İpek, M. (1988) *İstatistiğe Giriş I: Betimsel İstatistik*, Beta.
- İSO (İstanbul Sanayi Odası) (2002) *Türkiye'nin 500 Büyük Sanayi Kuruluşu 2001*, Ağustos.
- İTO (İstanbul Ticaret Odası) (2003) *Fiyat İndeksleri, Mart 2003*.
- İTO, *Fiyat İndeksleri*, Çeşitli aylar.
- İTO (2004) *Aylık Ekonomik Veriler, Aralık 2003*.
- İTÜ Vakfı (1991) *İTÜ Öğrencileri Üzerine Bir Araştırma: Sosyo-ekonomik Özellikler, Tutum ve Davranışlar*, Basılmamış araştırma.
- Jaffe, A. J., Spirer, H. F. (1987) *Misused Statistics*, Marcel Dekker.
- Jarque, C. M., Bera, A. K. (1987) "A test for normality of observations and regression residuals" *International Statistical Review*, 55:163-172.

- Johnson, N. L., Kotz, S. (1997) (eds.) *Leading Personalities in Statistical Sciences*, Wiley.
- Kendall, M. G. (1956) "The beginnings of probability calculus" *Bimetrica*, 43:1-14, Pearson, Kendall (1970) içinde 19-34.
- Kendall, M. G. (1960) "Where shall the history of statistics begin?" *Bimetrica*, 47:447-449, Pearson, Kendall (1970) içinde 45-46.
- Kendall, M., Stuart, A., Ord, K. (1987) *Kendall's Advanced Theory of Statistics*, v. I - *Distribution Theory*, 5th.ed., Griffin.
- Köse, A. H., Şenses, F., Yeldan, E. (der.) (2003) *İktisat Üzerine Yazılar II: İktisadi Kalkınma, Kriz ve İstikrar*, İletişim.
- Kristal-İş (1990) *Cam İşçisi Temmuz 1990 Anket Sonuçları*.
- Kurtuluş, K. (1985) *Pazarlama Araştırmaları*, 3. b. İstanbul Üniversitesi.
- Kutsal, A., Toktamış, Ö. (1981) *İstatistik (Terimler, Tarihçe, Örgütler, Yayınlar, Öğretim)*, Hacettepe Üniversitesi.
- Lambert, P. J. (1993) *The Distribution and Redistribution of Income – A Mathematical Analysis*, 2nd ed. Manchester University Press.
- Lange, O. (1978) *Introduction to Econometrics*, 4th ed., Pergamon.
- Lee, J. C. (2000) *Business and Financial Statistics Using MINITAB 12 and Microsoft Excel 97*, World Scientific Publishing.
- Lind, D. A., Marchal, W. G., Mason, R. D. (2002) *Statistical Techniques in Business & Economics*, McGraw-Hill.
- Marsh, C. (1988) *Exploring Data*, Polity Press.
- McClave, J. T., Benson, P. G., Sincich, T. (2001) *Statistics for Business and Economics*, 8th ed., Prentice Hall.
- McNeill, D. R. (1977) *Interactive Data Analysis*, Wiley.
- Miller, G. A. (1956) "The magical number seven, plus or minus two: Some limits on our capacity of processing information", *The Psychological Review*, 63, 2:81-97.
- Milliyet-İnput (2004) *Almanak 2004*.
- Monk, A. (1991) *Exploring Statistics with MINITAB*, Wiley.
- Montgomery, D. (1997) *Design and Analysis of Experiments*, 4th ed. Wiley.
- Moore, D. S. (1979) *Statistics: Concepts and Controversies*, Freeman.
- Mosteller, F., Kruskal, W. H., Peters, R. S., Rising, G. R. Link, R. F., Zelinka, M. (1973) *Statistics by Sample*, Addison-Wesley, vol.1 – *Exploring Data*, vol. 2 – *Weighing Chances*, vol. 3 – *Detecting Patterns*, vol. 4 – *Finding Models*.
- Mosteller, F., Wallace, D. L. (1972) "Deciding authorship", Tanur vd. (1989) içinde 115-125.
- Neter, J. (1989) "How accountants save money by sampling", Tanur vd. (1989) içinde 151-160.
- Neufeld, J. L. (1997) *Learning Business Statistics with Microsoft Excel*, Prentice Hall.

- Newbold, P., Carlson, W. L., Thorne, B. (2003) *Statistics for Business and Economics*, 5th ed, Prentice Hall (4. basımdan çeviren: Şenesen, Ü. (2000) *İşletme ve İktisat İçin İstatistik*, Literatür).
- Norusis, M. (1993) *SPSS for Windows, Base Sysrem User's Guide*, SPSS.
- Orhunbilge, N., Ersoy, M. Ş., Çatalca, H., Nişel, R. (1997) *Kuramsal ve Uygulamalı İstatistik Terimleri Sözlüğü*, Literatür.
- Ott, R. L., Longnecker, M. (2001) *An Introduction to Statistical Methods and Data Analysis*, 5th ed. Duxbury.
- ÖSYM (Öğrenci Seçme ve Yerleştirme Merkezi) (1987) *1985-1986 Öğretim Yılı Yükseköğretim İstatistikleri*.
- ÖSYM (2002) *2002 – ÖSYS Yükseköğretim Programlarının Merkezi Yerleştirme-deki En Küçük ve En Büyük Puanları*, ÖSYM.
- Özdamar, K. (1999) *Paket Programlar ile İstatistik Veri Analizi*, Kaan.
- Özdemir, D. (2001) *Applied Statistics for Economics and Business*, İstanbul Bilgi Üniversitesi.
- Özötün, E. (1980) *Türkiye Gayrisafi Yurtiçi Hasılası (İller İtibariyle) - Kaynak ve Yöntemler (1973 – 78)*, DİE Matbaası.
- Özötün, E. (1988) *Türkiye Gayrisafi Yurtiçi Hasılasının İller İtibariyle Dağılımı (1979 – 1986)*, İstanbul Sanayi Odası.
- Pappas, T. (1991) *More Joy of Mathematics*, Tetra.
- Pappas, T. (1993) *Yaşayan Matematik*, Sarmal.
- Pappas, T. (1999) *Mathematical Footprints*, Tetra.
- Pappas, T. (2002) *Math Stuff*, Tetra.
- Pearson, E. S., Kendall, M. G. (1970) (eds.) *Studies in the History of Statistics and Probability*, Griffith.
- Petrol-İş (2000) '97-'99 *Petrol-iş Yıllığı*, Petrol-iş Sendikası.
- Rosenberger, J. L., Gasko, M. (1983) "Comparing location estimators: Trimmed means, medians, and trimean", Hoaglin, Mosteller, Tukey (1983) içinde 297-338.
- Salzburg, D. (2001) *The Lady Tasting Tea – How Statistics Revolutionized Science in the Twentieth Century*, Freeman.
- Sandy, R. (1990) *Statistics for Business and Economics*, McGraw-Hill.
- Saraçoğlu, B., Çevik, F. (1989) *Matematiksel İstatistik*, Teori
- Seal, H. L. (1967) "The historical development of the Gauss linear model", *Biometrika*, 54:1-24, Pearson, Kendall (1970) içinde 207-230.
- Sekaran, U. (2000) *Research Methods for Business*, 3rd ed. Wiley.
- Sel, C. (1994) "Ekonomik bilgilerin elde edilebileceği basılı ve on-line kaynaklar", *Ekonomiyi İzleme Sempozyumu 5-6 Mayıs 1994* içinde (sayfa numaraları yok) Hacettepe Üniversitesi.
- Sencer, M. (1981) *Yöntembilim Terimleri Sözlüğü*, TDK.
- Sencer, M., Irmak, Y. (1984) *Toplumbilimlerinde Yöntem*, Say.

- Sennaroğlu, B. (2001) *An Approach to Process Improvement in Industry*, Basılmamış doktora tezi, Marmara Üniversitesi, Temel ve Uygulamalı Bilimler Enstitüsü.
- Serdaroğlu, R. (2001) "Rakam ciddiyetsizliği", *Radikal*, 7 Aralık.
- Seyidoğlu, H. (1995) *Bilimsel Araştırma ve Yazma El Kitabı*, Güzem.
- Sezgin, F. (1993) *Dil ve Edebiyatta İstatistik ve Bilgisayar Uygulamaları*, Dergâh.
- Sheynin, O. B., (1968) "On the early history of the law of large numbers" *Biometrika*, 55:459-467, Pearson, Kendall (1970) içinde 231-239.
- Sınıksaran, E. (2001) *Teori ve Uygulamalarıyla İstatistik Yöntemler*, Filiz.
- Siegel, A. F. (1988) *Statistics and Data Analysis*, Wiley.
- Siegel, A. F. (1997) *Practical Business Statistics*, Irwin.
- Siegel, S., Castellan, Jr., N. J. (1988) *Nonparametric Statistics for the Behavioral Sciences*, 2. b. McGraw-Hill.
- Sönmez, M. (2001) *Gelir Uçurumu*, Om.
- Sönmez, M. (2002) *100 Göstergede Kriz ve Yoksullaşma*, İletişim.
- Spatz, C. (2001) *Basic Statistics: Tales of Distributions*, 7th ed., Wadsworth.
- SPSS (1993) *SPSS for Windows Base System User's Guide*, Release 6.
- SSK (Sosyal Sigortalar Kurumu) (1995) *Sosyal Sigortalar Kurumu 1994 Yıllığı*.
- Stuart, A. (1984) *The Ideas of Sampling*, Charles Griffin.
- Stutely, R. (2003) *The Definitive Guide to Managing the Numbers*, Prentice Hall.
- Sutcliffe, B. (2002) *100 Ways of Seeing an Unequal World*, Zed Books.
- Şenesen, Ü. (1995) "Kısa ve orta mesafe koşularında olimpiyat derecelerinin incelenmesi" Aköz, Y. (1995) içinde 107-119.
- Şenesen, Ü. (2002) "Regional income distribution in Turkey", Balkan, Savran (2002) içinde 117-127.
- Şenesen, Ü., Günlük-Şenesen, G. (2003) "Üretimde dışılıma bağımlılık: 1970'lerden 2000'lere ne değişti?" Köse, Şenses, Yeldan (2003) içinde 533-559.
- Tanur, J. M., Mosteller, F., Kruskal, W. H., Lehman, E. L., Link, R. F., Pieters, R. S., Rising, G. R. (1973, 3rd ed., 1989) *Statistics: A Guide to the Unknown*, 4th ed., Holden Day.
- TCMB (Türkiye Cumhuriyet Merkez Bankası) (2003) *Yıllık Rapor 2002*.
- Thomas, V., Wang, Y., Fan, X. (2000) "Measuring education inequality: Gini coefficient of education", *World Bank Institute Working Paper*.
- Tufte, J. W. (1983) *The Visual Display of Quantitative Information*, Graphics Press.
- Tufte, J. W. (1990) *Envisioning Information*, Graphics Press.
- Tufte, J. W. (1997) *Visual Explanations*, Graphics Press.
- Tukey, J. W. (1977) *Exploratory Data Analysis*, Addison-Wesley.

- TÜBİTAK (Türkiye Bilimsel ve Teknolojik Araştırma Kurumu) (1995) “Bilim ve Teknik Dergisi okuyucu anketi sonuçlandı”, *Bilim ve Teknik Dergisi*, 337, Aralık.
- TÜSİAD (Türkiye Sanayiciler ve İşadamları Derneği) (2001) *Seçim Sistemi ve Siyasi Partiler Araştırması*, c. I *Temel Bulgu ve Sonuçlar*, c. II *Ana Rapor*.
- TÜSİAD (2003) *Karşılaştırmalı Bir Perspektiften Türkiye’de Kamu Sektörü*.
- Tütengil, C. O. (1975) *Sosyal Bilimlerde Araştırma ve Metod*, 3. b., İÜ İktisat Fakültesi.
- Ulagay, O. (2004) “Ekonomimiz dalganın üstünde, aman dikkat” *Milliyet* , 11 Ocak
- UNDP (Birleşmiş Milletler Kalkınma Programı) (1995) *Human Development Report 1995*, Oxford University Press.
- UNDP (2000) *Human Development Report 2000*, Oxford University Press.
- UNDP (2001) *Human Development Report 2001*, Oxford University Press.
- UNDP (2003) *Human Development Report 2003*, Oxford University Press.
- Üster, C. (2003) “Yanlışlıklar komedyası” *Radikal Kitap*, 11 Temmuz.
- Velleman, P. F., Hoaglin, D. C. (1981) *Applications, Basics, and Computing of Exploratory Data Analysis*, Duxbury Press.
- Walgren, A., Walgren, B, Person, R., Jerner, U., Haaland, J. (1996) *Graphing Statistics & Data*, SAGE.
- Wonnacott, T. H., Wonnacott, R. J. (1990) *Introductory Statistics for Business and Economics*, 4th ed. Wiley.
- World Bank (Dünya Bankası) (1993) *World Development Report 1993*, Oxford University Press.
- World Bank (1990) *World Development Report 1990*, Oxford University Press.
- World Bank (1992) *World Development Report 1992*, Oxford University Press.
- World Bank (1997) *World Development Report 1997*, Oxford University Press.
- World Bank (1999) *World Development Report 1998/99 – Knowledge for Development*, Oxford University Press.
- World Bank (2000) *World Development Report 1999/2000 – Knowledge for Development*, Oxford University Press.
- World Bank (2001) *World Development Report 2002 – Building Institutions for Markets*, Oxford University Press.
- World Bank (2003) *World Development Report 2003 – Selected World Development Indicators*, Oxford University Press.
- Yamane, T. (1967) *Elementary Sampling Methods*, Prentice-Hall (Çev. Esin, A., Bakır, M. A., Aydın, C., Gürbüzel, E.(2001) *Temel Örnekleme Yöntemleri*, Literatür).
- Yamane, T. (1973) *Statistics: An Introductory Analysis*, 3rd ed. Harper & Row.
- Yenal, O. (2003) *Cumhuriyet’in İktisat Tarihi*, Homer.
- Yoğurtçugil, K. (1975) *Örnekleme:Yöntemler ve Uygulama*, İÜ İktisat Fakültesi.

Ağçevrimi adresleri:

<http://www.die.gov.tr>
<http://www.hazine.gov.tr>
<http://www.tcmb.gov.tr>
<http://www.imkb.gov.tr>
<http://www.tobb.org.tr>
<http://www.ito.org.tr>
<http://www.iso.org.tr>
<http://www.tbb.org.tr>
<http://www.tusiad.org.tr>
<http://www.musiad.org.tr>
<http://www.turkis.org.tr>
<http://www.disk.org.tr>
<http://www.hakis.org.tr>
<http://www.tisk.org.tr>
<http://www.mess.org.tr>
<http://www.worldbank.org>
<http://www.imf.org>
<http://www.undp.org>
<http://www.unesco.org>
<http://www.unctad.org>
<http://www.oecd.org>
<http://www.opec.org>
<http://www.ilo.org>
<http://www.nyse.com>
<http://www.economist.com>
<http://www.ft.com>
<http://www.wsj.com>

Devlet İstatistik Enstitüsü
 Hazine Müsteşarlığı
 Merkez Bankası
 İstanbul Borsası
 Odalar ve Borsalar Birliği
 İstanbul Ticaret Odası
 İstanbul Sanayi Odası
 Türkiye Bankalar Birliği
 TÜSİAD
 MÜSİAD
 Türk-İş
 DİSK
 Hak-İş
 TİSK
 MESS
 Dünya Bankası
 Uluslararası Para Fonu
 BM Kalkınma Programı
 BM Eğitim, Bilim, Kültür Örgütü
 BM Ticaret, Kalkınma Konferansı
 İktisadi İşbirliği, Kalkınma Örgütü
 Petrol Satan Ülkeler Örgütü
 Uluslararası Çalışma Örgütü
 New York Borsası
The Economist
Financial Times
Wall Street Journal

A

açık uçlu öbek (open ended class interval) 115
açıklanan değişken, *bkz* bağımlı değişken
açıklayıcı değişken, *bkz* bağımsız değişken
ad ölçeği (nominal scale) 27-30, 104, 108, 114, 124, 172, 323, 340
adım (step) 85, 324
ağırlık merkezi (center of gravity) 190
ağırlıklı ortalama, *bkz* tartılı ortalama
alan-çizim, *bkz* histogram
alt dışadüşen (lower outlier) 286, 324
alt dördebölen (lower quartile) 179, 226, 227, 240, 282, 285, 324
alt eşik (lower threshold) 285, 324
alt öbek (lower group) 343
anakütle (population) 25, 75-92, 226-233, 238
anakütle ortalaması (population mean) 75-92
AO dolayındaki beklemler (moments around mean) 241
aralık (range) 225
aralık ölçeği (interval scale) 33-36, 172
aritmetik ortalama (AO) (arithmetic mean) 30, 33, 35, 148, 184-192
aşamalı rassal örnekleme (multistage random sampling) 93-95
atlamalı rassal örnekleme (systematic random sampling) 87-89

B

bağımlı değişken (dependent variable) 7, 96, 340
bağımsız değişken (independent variable) 7, 96, 106, 340
bağlanım (regression) 22, 37, 99, 191
bağlanım katsayısı (regression coefficient) 24
bakışıklık, *bkz* bakışım
bakışım (symmetry) 28, 131, 133, 167, 239, 269-275, 294
basık (platykurtic) 136, 167, 170, 223, 224, 271
basıklık katsayısı coefficient of kurtosis) 37, 242
basıklık ölçüleri (measures of kurtosis) 244
basit rassal örnekleme (simple random sampling) 84-86
başka etkiler (other effects) 330
beklem (moment) 241
beklemlere dayalı basıklık katsayısı (measure of kurtosis based on moments) 244
beklemlere dayalı çarpıklık katsayısı (coefficient of skewness based on moments) 243
beklenen değer (expected value) 77, 79, 91, 241
berk (robust) 225, 332, 337, 343
berk kestirim yöntemleri (robust estimation methods) 22

berk ortalamalar (robust averages) 169-181
 beşebölenler (quantiles) 179, 246
 beşli özet (five-point-summary) 280
 biçim (shape) 167, 223, 243
 birikimli sıklık (cumulative frequency) 32, 176, 178, 280
 birikimli göreceli sıklık (cumulative relative frequency) 32, 178
 birinci el veri (primary data) 66
 Bowley çarpıklık ölçüsü (Bowley measure of skewness) 241, 287
 bölün (quantile) 170, 179-181, 369
 bölünler ortası (mid-quantiles) 300
 büyük sayılar yasası (law of large numbers) 21

Ç

çan dağılımı (bell distribution) 244
 çan eğrisi (bell curve) 244
 çarpıklık (skewness) 133, 167, 339-343
 çarpıklık katsayısı (ÇK) (coefficient of skewness) 37, 244
 çarpıklık-basıklık ölçüsü (measure of skewness and kurtosis) 244
 Çebişev eşitsizliği (Chebyshev inequality) 234
 çizelgeler (tables) 107-116
 çizimler (charts, graphs) 116-147
 çoklu bağlanım (multiple regression) 99
 çoklu varyans çözümlemesi (multivariate analysis of variance) 99, 336
 çubuk çizimi (bar chart) 121

D

dal-yaprak (stem-and-leaf) 8, 55, 133, 271-280
 değişken (variable) 25
 değişkenlik (variation) 167, 223,

değişkenlik katsayısı (coefficient of variation) 35, 37, 237
 deney tasarımı (experimental design) 7, 22, 96-99
 dışadüşen (outlier) 273-275, 284-288
 DİE yayınları (SIS publications) 67-70
 dirençli doğru (resistant line) 349-357
 dizgisel hata (systematic error) 88
 doğrusal dönüştürme (linear transformation) 34, 296, 297
 doğrusal ilişki (linear relationship) 337
 dönüştürme (transformation) 35, 142, 296-312
 dönüştürme merdiveni (transformation ladder) 8, 299, 358
 dördebölenler (quartiles) 33, 179, 280, 324
 dördebölenler aralığı (DAr) (interquartile range - IQR) 226, 285, 294
 dördebölenler arası ortalama (inter-quartile mean) 210, 337
 dördebölenler ortalaması (trimean) 170, 180-182, 210, 300
 dördebölenler ortası (midpoint of quartiles) 182, 300
 dördebölenler sapması (DS) (deviation of quartiles) 227
 DPT yayınları (SPO publications) 70
 duvar çizimi (area chart) 127
 düşük-yüksek çizimi (stock chart) 138
 duyarlı ortalamalar (sensitive averages) 182-209
 düzey (level) 167

E

egemen değer, bkz ensıkdeğer
 eksi iki terimli dağılımı (negative binomial distribution) 20
 ekonometri (econometrics) 22
 en küçük kareler yöntemi (method of least squares) 21, 343

ensıkdeğer (mode) 30, 169, 170-175,
178 240

en yüksek olabilirlik kestirimcileri
(maximum likelihood estimators)
210, 337

eşik (treshhold) 324

eşlenik veri (paired data) 203

etkin basamak (significant digit) 57

etkin kestirimci (efficient estimator) 84

etkin olmayan kestirimci (inefficient
estimator) 80

etkinlik (efficiency) 77-80

etmen çözümlemesi (factor analysis) 22
EVIEWWS, 13

F

F dağılımı (*F* distribution) 22

F sınaması (*F* test) 35

faktor analizi *bkz* etmen çözümlemesi

fark (difference) 39

frekans, *bkz* sıklık

G

Gauss dağılımı (Gauss distribution) 21

geçici kalıntılar (temporary residuals)
332

geçici ortak değer (temporary common
value) 333

geçişlilik (transitivity) 28, 31

gelir dağılımı (income distribution)
245-267

gelir dağılımında eşitsizlik ölçüleri
(measures of skewness in income
distribution) 245-267

genel eğilim (trend) 107

genişletilmiş ortanca (*G-Ortc*)
(expanded median) 210

geometrik ortalama (geometric mean)
35, 109, 193-203

Gini yoğunlaşma katsayısı (Gini
coefficient of concentration) 22,
244, 247-253

görelî çarpıklık (relative skewness)
295

görelî fark (relative difference) 238

görelî hata (relative error) 59

görelî sıklık (relative frequency) 30,
173

görelî varyans (relative variance) 237

gözlem birimi (observation unit) 25

gözlem değeri (observation value) 22,
25

güven aralığı (confidence interval) 22,
107

H

harmonik ortalama, *bkz* ters ortalama

hata terimi (error term) 22, 351

Hazine Müsteşarlığı yayınları
(Treasury publications) 70

hipotez testi, *bkz* önsav sınaması

histogram (histogram) 132

İ

iki boyutlu çizelgeler (two dimensional
tables) 323

ikinci el veri (secondary data) 66

iki terimli dağılımı (binomial distribu-
tion) 20

ilişki (correlation) 22, 37, 106, 140,
167, 323-365

ilişki katsayısı (correlation coefficient)
35

indeks (index) 22

istatistiğin kötüye kullanımı (abuse of
statistics) 10-12

istatistik yöntemlerinin evrimi
(evolution of statistical methods)
20-23

işlevsel gelir dağılımı (functional
income distribution) 245

J

Jarque -Bera çarpıklık - basıklık katsa-
yısı (Jarque-Berra coefficient of
skewness and kurtosis) 244

K

kalıntı (residual) 330, 351
kalite denetimi (quality control) 22, 88
kantil, *bkz* bölen
kapsayıcılık (exhaustiveness) 28, 29
kartil *bkz* dördebölen
karar kuramı (decision theory) 20
kareli ortalama (quadratic mean) 35, 169, 208-209
karşı önsav (alternative hypothesis) 9
karşılaştırma değerleri (comparison values) 337
kartil, *bkz* dördebölen
katmanlı rassal örnekleme (stratified random sampling) 88-93
katsayısal olmayan sınamalar (nonparametric tests) 22, 31, 41
katsayısal sınamalar (parametric tests) 32, 41
kentil, *bkz* beşebölen
kesikli değişken (discrete variable) 38
kesme (truncating) 51-60
kestirimci (estimator) 9, 77, 233
kırpık ortalama (trimmed mean) 209-211, 307
ki-kare sınaması (chi-square test) 22, 106
ki-kare dağılımı (chi-square distribution) 245
kişisel gelir dağılımı (individual income distribution) 245
kişisel olasılık (personal probability) 22
Kolmogorov-Smirnov sınaması (Kolmogorov-Smirnov test) 245
kombinasyon (combination) 21
koni çizimi (cone graph) 138
korelasyon *bkz* ilişki
kuadratik ortalama, *bkz* kareli ortalama
kutu-çizim (box-plot) 9, 191, 246, 283
kümeli rassal örnekleme (clustered random sampling) 95-96

L

Lilliefors sınaması (Lilliefors test) 245
LIMDEP, 13
logaritmali dönüştürme (logarithmic transformation) 142
Lorenz eğrisi (Lorenz curve) 246

M

Maliye Bakanlığı yayınları (Ministry of Finance publications) 71
Markov zinciri (Markov chain) 22
matematiksel beklenti (mathematical expectation) 20
medyan, *bkz* ortanca
Merkez Bankası yayınları (Central Bank publications) 71
merkezi limit teoremi (central limit theorem) 22
MICRO TSP, 13
MINITAB 13, 270, 365
M-kestirimcileri, *bkz* en yüksek olabilirlik kestirimcileri
mod, *bkz* ensıkdeğer
model (model) 22
Monte Carlo benzetimi (Monte Carlo simulation) 225

N

nicel değişken (quantitative variable) 25, 35, 39, 172
nitel değişken (qualitative variable) 26, 35
nokta-çizim (dot-plot) 137, 186
normal dağılım (normal distribution) 21, 201, 235, 244, 286
nüfus sayımı (census) 17-19

O

olağandışı değerler (extreme values) 167, 226, 324
olasılığın belirtileri (probability axioms) 22
olasılık (probability) 20
onabölenler (deciles) 179, 246

onabölenler aralığı (interdecile range) 227

oran (ratio) 29, 30, 39

orta öbek (middle group) 343

oran ölçeği (ratio scale) 31-39, 172, 323, 340

orta yarı ortalaması, *bkz* dördebölenler ortalaması

ortak değer (common value) 329

ortalama düzey (average level) 167

ortalama mutlak sapma (*OMS*) (mean absolute deviation - *MAD*) 228

ortanca (*Ortc*) (median) 32, 169, 176-179, 279

ortanca işleme (median polish) 111, 324, 329-339

Osmanlılarda sayım (19)

oynaklık (volatility) 167, 223

Ö

öbek aralığı (group range) 113-115, 174, 201

ölçekler (scales) 29-41

ölçme (measurement) 26

ölçme düzeyi (measurement level) 27

ön çözümleme (initial analysis) 270

önsav sınaması (hypothesis testing) 9, 22, 245

örnekleme (sampling) 7, 11, 25, 75-99

örtüşmezlik (mutually exclusiveness) 32

özet nokta (summary point) 348-363

P

pasta çizimi (pie chart) 120

pay (share) 29

PcGive, 13

permutasyon (permutation) 21

Phstat 13

piramit çizimi (pyramid chart) 138

R

rassal olmayan örnekleme (nonrandom sampling) 77, 82

rassal örnekleme (random sampling) 77-96

rassal sayılar (random numbers) 82

rassallık (randomness) 97

regresyon, *bkz* bağlanım

risk (risk) 225

S

sağa çarpık (skewed towards right) 239

sapkın (biased) 77, 88

SAS, 13, 270

satır etkisi (row effect) 330

serbestlik derecesi (degrees of freedom) 235

serpilme çizimi (scatter diagram, scatter plot) 137, 140-142, 340

Shapiro-Wilks sınaması (Shapiro-Wilks test) 245

SHAZAM, 13

sıfır dolayındaki beklemler (moments around origin) 241

sıfır önsavı (null hypothesis) 9

sıklık (frequency) 19, 104-125

sıklık dağılımı (frequency distribution) 114

sıklık eğrisi (frequency curve) 137

sınama (testing) 9

sıra ilişki katsayısı (rank korelation coefficient) 33

sıra ölçeği (ordinal scale) 27, 31-33, 35, 39, 40, 41, 104, 114, 124, 297, 323, 340

sigorta matematikçileri (insurance mathematicians) 18, 21

simgeli çizelgeler (coded tables) 324

simit-çizim (doughnut chart) 138

sivri (leptokurtic) 167, 223, 244, 269, 271

sola çarpık (skewed towards left) 239

sorgulayıcı veri çözümlemesi (exploratory data analysis) 7, 22, 169, 269-313

SPSS 13, 210, 270
 standart sapma (standard deviation) 35,
 191, 228, 229-237
 standartlaştırma (standardization) 294-
 296
 STATA 13
 sürekli değişken (continuous variable)
 38, 323
 sürekli olasılık dağılımı (continuous
 probability distribution) 137
 sütun etkisi (column effect) 330
 SYSTAT 13

şans oyunları (games of chance) 20

T

t dapılımı (*t* distribution) 22
t sınaması (*t* test) 35, 98, 306
 tahmin edici *bkz* kestirimci
 tamsayım (census) 61
 tartılı ortalama (weighted average) 40,
 185
 tartılı geometrik ortalama (weighted
 geometric mean) 200
 temsilci (representative) 168, 194
 tepe değer, *bkz* ensıkdeğer
 tepe sayısı (number of peaks) 167, 223
 tepki değeri (response value) 330
 ters ortalama (harmonic mean) 35, 169,
 204-208
 tesadüfi sayılar *bkz* rassal sayılar
 toplamalı model (additive model) 330
 toplanma oranı (centralization ratio)
 253
 TreePlan 13
 Tukey menteşeleri (Tukey's hinges)
 180, 281
 Türkiye İstatistik Günü 1

U

uçuşan şerit çizimi (band chart) 128
 uzak dışadüşen (far outlier) 285
 uzak eşik (far threshold) 285

uzak alt dışadüşen (lower far outlier)
 324

uzak üst dışadüşen (upper far outlier)
 324

Ü

üst dışadüşen (upper outlier) 286
 üst dördebölen (upper quartile) 180,
 226, 227, 240, 282, 285
 üst eşik (upper threshold) 325
 üst öbek (upper group) 343
 üstel dönüştürme (exponential trans-
 formation) 296, 298

V

varyans (variance) 191, 229-238,
 varyans çözümlemesi (analysis of
 variance) 22, 37, 99, 306
 veri derleme (data collection) 7
 veri derlemenin evrimi (evolution of
 data collection) 17-19
 veri mürekkebi oranı (data ink ratio)
 119
 veri sunumu (data presentation) 103-
 107
 veri yoğunluğu (data concentration)
 119

VISUAL STATISTICS 13

Y

yabancı veri kaynakları (foreign
 data sources 72-73
 yanlış kestirimci (biased estimator) 77,
 80, 88
 yansız kestirimci (unbiased estimator)
 79, 80, 233
 yansızlık (unbiasedness) 77, 80
 yarı duyarlı ortalamalar (quasi sensitive
 averages) 209-211
 yayıklık (spread) 167, 209, 306, 309,
 310, 312
 yayılım (variation) 167, 209, 306, 309,
 310, 312

- yerli veri kaynakları (domestic data sources) 67-72
- yuvarlatma (rounding) 51-60
- yuvarlatma hatası (rounding error) 58-60
- yuvarlatma kuralları (rules of rounding) 53-58
- yüzde (percentage) 29, 30
- yüzebölenler (centiles) 179, 246

A

- abuse of statistics (istatistiğin kötüye kullanımı) 10-12
- additive model (toplamalı model) 330
- alternative hypothesis (karşı önsav) 9
- analysis of variance (varyans çözümlemesi) 22, 37, 99, 306
- area chart (duvar çizimi) 127
- arithmetic mean (aritmetik ortalama) 30, 33, 35, 148, 184-192
- average level (ortalama düzey) 167

B

- band chart (uçuşan şerit çizimi) 128
- bar chart (çubuk çizimi) 121
- bell curve (çan eğrisi) 244
- bell distribution (çan dağılımı) 244
- biased (sapkın) 77, 88
- biased estimator (yanlı kestirimci) 77, 80, 88
- binomial distribution (iki terimli dağılımı) 20
- Bowley measure of skewness (Bowley çarpıklık ölçüsü) 241, 287
- box-plot (kutu-çizim) 9, 191, 246, 283

C

- census (nüfus sayımı, tamsayım) 17-19, 61
- center of gravity (ağırlık merkezi) 190
- centiles (yüzebölenler) 179, 246
- central limit theorem (merkezi limit teoremi) 22
- centralization ratio (toplanma oranı) 253

- charts (çizimler) 116-147

- Chebyshev inequality (Çebişev eşitsizliği) 234
- chi-square distribution (ki-kare dağılımı) 245
- chi-square test (ki-kare sınaması) 22, 106
- clustered random sampling (kümeli rassal örnekleme) 95-96
- coded tables (simgeli çizelgeler) 324
- coefficient of kurtosis (basıklık katsayısı) 37, 242
- coefficient of skewness (çarpıklık katsayısı) 37, 244
- coefficient of variation (değişkenlik katsayısı) 35, 37, 237
- column effect (sütun etkisi) 330
- combination (kombinasyon) 21
- common value (ortak değer) 329
- comparison values (karşılaştırma değerleri) 337
- cone graph (koni çizimi) 138
- confidence interval (güven aralığı) 22, 107
- continuous variable (sürekli değişken) 38, 323
- correlation (ilişki) 22, 37, 106, 140, 167, 323-365
- correlation coefficient (ilişki katsayısı) 35
- cumulative frequency (birikimli sıklık) 32, 176, 178, 280

cumulative relative frequency (birikimli göreceli sıklık) 32, 178

D

data collection (veri derleme) 7

data concentration (veri yoğunluğu) 119

data ink ratio (veri mürekkebi oranı) 119

data presentation (veri sunumu) 103-107

deciles (onabölenler) 179, 246

decision theory (karar kuramı) 20

degrees of freedom (serbestlik derecesi) 235

dependent variable (bağımlı değişken) 7, 96, 340

deviation of quartiles (dördebölenler sapması) 227

discrete variable (kesikli değişken) 38

domestic data sources (yerli veri kaynakları) 67-72

dot-plot (nokta-çizim) 137, 186

doughnut chart (simit-çizim) 138

E

econometrics (ekonometri) 22

efficiency (etkinlik) 77-80

efficient estimator (etkin kestirimci) 84

error term (hata terimi) 22, 351

estimator (kestirimci) 9, 77, 233

EVIEWES, 13

evolution of data collection (veri derlemenin evrimi) 17-19

evolution of statistical methods (istatistik yöntemlerinin evrimi) 20-23

exhaustiveness (kapsayıcılık) 28, 29

expanded median (genişletilmiş ortanca) 210

expected value (beklenen değer) 77, 79, 91, 241

experimental design (deney tasarımı) 7, 22, 96-99,

exploratory data analysis (sorgulayıcı veri çözümlemesi) 7, 22, 169, 269-313

exponential transformation (üstel dönüştürme) 296, 298

extreme values (olağandışı değerler) 167, 226, 234

F

F distribution (*F* dağılımı) 22

F test (*F* sınaması) 35

factor analysis (etmen çözümlemesi) 22

far outlier (uzak dışadüşen) 285

far threshold (uzak eşik) 285

five-point-summary (beşli özet) 280

foreign data sources (yabancı veri kaynakları) 72-73

frequency curve (sıklık eğrisi) 137

frequency distribution (sıklık dağılımı) 114

frequency (sıklık) 19, 104-125

functional income distribution (işlevsel gelir dağılımı) 245

G

games of chance (şans oyunları) 20

Gauss distribution (Gauss dağılımı) 21

geometric mean (geometrik ortalama) 35, 109, 193-203

Gini coefficient of concentration (Gini yoğunlaşma katsayısı) 22, 244, 247-253

graphs (çizimler) 116-147

group range (öbek aralığı) 113-115, 174, 201

H

harmonic mean (ters ortalama) 35, 160, 204-208

histogram (histogram) 132

hypothesis testing (önsav sınaması) 9, 22, 245

İ

income distribution (gelir dağılımı) 245-267

independent variable (bağımsız değişken) 7, 96, 106, 340

index (indeks) 22

individual income distribution (kişisel gelir dağılımı) 245

inefficient estimator (etkin olmayan kestirimci) 80

initial analysis (ön çözümleme) 270

insurance mathematicians (sigorta matematikçileri) 18, 21

interdecile range (onabölenler aralığı) 227

interquartile mean (dördebölenler arası ortalama) 210, 337

interquartile range (*IQR*) (dördebölenler aralığı) 226, 285, 294

interval scale (aralık ölçeği) 33-36, 172

J

Jarque-Berra coefficient of skewness and Kurtosis (Jarque –Bera çarpıklık – basıklık katsayısı) 244

K

Kolmogorov-Smirnov test (Kolmogorov-Smirnov sınaması) 245

L

law of large numbers (büyük sayılar yasası) 21

leptokurtic (sivri) 167, 223, 244, 269, 271

level (düzey) 167

Lilliefors test (Lilliefors sınaması) 245

LIMDEP, 13

linear relationship (doğrusal ilişki) 337

linear transformation (doğrusal dönüş-
türme) 34, 296, 297

logarithmic transformation (logaritmali
dönüştürme) 142

Lorenz curve (Lorenz eğrisi) 246

lower far outlier (uzak alt dışadüşen) 324

lower group (alt öbek) 343

lower outlier (alt dışadüşen) 286, 324

lower quartile (alt dördebölen) 179, 226, 227, 240, 282, 324

lower threshold (alt eşik) 285, 324

M

Markov chain (Markov zinciri) 22

mathematical expectation (matematik-
sel beklenti) 20

maximum likelihood estimators (en
yüksek olabilirlik kestirimcileri) 210, 337

mean absolute deviation (*MAD*) (orta-
lama mutlak sapma) 209, 228

measure of kurtosis based on moments
(beklemlere dayalı basıklık katsayısı) 244

measure of skewness and kurtosis (çar-
pıklık-basıklık ölçüsü) 244

measure of skewness based on
moments (beklemlere dayalı çarpık-
lık katsayısı) 243

measures of kurtosis (basıklık ölçüleri) 244

measures of inequality in income
distribution (gelir dağılımında çar-
pıklık ölçüleri) 245-267

measurement level (ölçme düzeyi) 27

median polish (ortanca işleme) 111, 324, 329-339

median (ortanca) 32, 169, 176-179, 279

method of least squares (en küçük kare-
ler yöntemi) 21, 343

middle group (orta öbek) 343

midpoint of quartiles (dördebölenler
ortası) 182, 300

mid-quantiles (bölenler ortası) 300
 MICRO TSP,13
 MINITAB 13, 270, 365
 mode (ensikdeğer) 30, 169, 170-175, 178, 240
 model (model) 22
 moment (beklem) 241
 moments around mean (AO dolayındaki beklemler) 241
 moments around origin (sıfır dolayındaki beklemler) 241
 Monte Carlo simulation (Monte Carlo benzetimi) 225
 multiple regression (çoklu bağlanım) 99
 multistage random sampling (aşamalı rassal örnekleme) 93-95
 multivariate analysis of variance (çoklu varyans çözümlemesi) 99, 336
 mutually exclusiveness (örtüşmezlik) 32
N
 negative binomial distribution (eksi iki terimli dağılımı) 20
 nominal scale (ad ölçeği) 27-30, 104, 108, 114, 124, 172, 323, 340
 nonparametric tests (katsayısal olmayan sınamalar) 22, 31, 41
 nonrandom sampling (rassal olmayan örnekleme) 77, 82
 normal distribution (normal dağılım) 21, 201, 235, 244, 286
 number of peaks (tepe sayısı) 167, 223
 null hypothesis (sıfır önsavı) 9
O
 observation unit (gözlem birimi) 25
 observation value (gözlem değeri) 22, 25
 open ended class interval (açık uçlu öbek) 115

ordinal scale (sıra ölçeği) 27, 31-33, 35, 39, 40, 41, 104, 114, 124, 297, 323, 340
 other effects (başka etkiler) 330
 outlier (dışadüşen) 273-275, 284-288
P
 paired data (eşlenik veri) 203
 parametric tests (katsayısal sınamalar) 32, 41
 PcGive, 13
 perentage (yüzde) 29, 30
 permutation (permutasyon) 21
 personal probability (kişisel olasılık) 22
 Phstat 13
 pie chart (pasta çizimi) 120
 platykurtic (basık) 136, 167, 170, 223, 224, 271
 population (anakütle) 25, 75-92, 226-233, 238
 population mean (anakütle ortalaması) 75-92
 primary data (birinci el veri) 66
 probability (olasılık) 20
 probabilitiy axioms (olasılığın belitleri) 22
 pyramid chart (piramit çizimi) 138
Q
 quadratic mean (kareli ortalama) 35, 169, 208-209
 qualitative variable (nitel değişken) 26, 35
 quality control (kalite denetimi) 22, 88
 quantile (bölen) 170, 179-181, 369
 quantitative variable (nicel değişken) 25, 35, 39, 172
 quartiles (dördebölenler) 33, 179, 280, 324
 quasi sensitive averages (yarı duyarlı ortalamalar) 209-211
 quintiles (beşebölenler) 179, 246

R

random numbers (rassal sayılar) 82
 random sampling (rassal örnekleme) 77-96
 randomness (rassallık) 97
 range (aralık) 225
 rank korelation coefficient (sıra ilişki katsayısı) 33
 ratio (oran) 29, 30, 39
 ratio scale (oran ölçeği) 31-39, 172, 323, 340
 regression coefficient (bağlanım katsayısı) 24
 regression (bağlanım) 22, 37, 99, 191
 relative difference (görelî fark) 238
 relative error (görelî hata) 59
 relative frequency (görelî sıklık) 30, 173
 relative skewness (görelî çarpıklık) 295
 relative variance (görelî varyans) 237
 representative (temsilci) 168, 194
 residual (kalıntı) 330, 351
 resistant line (dirençli doğru) 349-357
 response value (tepki değeri) 330
 risk (risk) 225
 robust averages (berk ortalamalar) 169-181
 robust estimation methods (berk kestirim yöntemleri) 22
 robust (berk) 225, 332, 337, 343
 rounding (yuvarlatma) 51-60
 rounding error (yuvarlatma hatası) 58-60
 row effect (satır etkisi) 330
 rules of rounding (yuvarlatma kuralları) 53-58

S

sampling (örnekleme) 7, 11, 25, 75-99
 SAS, 13, 270
 scales (ölçekler) 29-41

scatter diagram (serpilme çizimi) 137, 140-142, 340
 scatter plot, *see* scatter diagram
 secondary data (ikinci el veri) 66
 sensitive averages (duyarlı ortalamalar) 182-209
 shape (biçim) 167, 223, 243
 Shapiro-Wilks test (Shapiro-Wilks sınaması) 245
 share (pay) 29
 SHAZAM, 13
 significant digit (etkin basamak) 57
 skewed towards left (sola çarpık) 239
 skewed towards right (sağa çarpık) 239
 skewness (çarpıklık) 133, 167, 339-343
 spread (yayıklık) 167, 209, 306, 309, 310, 312
 SPSS 13, 210, 270
 standard deviation (standart sapma) 35, 191, 228, 229-237
 standardization (standartlaştırma) 294-296
 STATA 13
 stem-and-leaf (dal-yaprak) 8, 55, 133, 271-280
 step (adım) 85, 324
 stock chart (düşük-yüksek çizimi) 138
 stratified random sampling (katmanlı rassal örnekleme) 88-93
 summary point (özet nokta) 348-363
 symmetry (bakışım) 28, 131, 133, 167, 239, 260-275, 294
 SYSTAT 13
 systematic error (dizgisel hata) 88
 systematic random sampling (atlamalı rassal örnekleme) 87-89

T

t distribution (*t* dağılımı) 22
t test (*t* sınaması) 35, 98, 306
 tables (çizelgeler) 107-116

temporary common value (geçici ortak değer) 333
temporary residuals (geçici kalıntılar) 332
testing (sınama) 9
threshold (eşik) 324
transformation (dönüştürme) 35, 142, 296-312
transformation ladder (dönüştürme merdiveni) 8, 299, 358
transitivity (geçişlilik) 28, 31
TreePlan 13
trend (genel eğilim) 107
trimean (dördebölenler ortalaması) 170, 180-182, 210, 300
trimmed mean (kırpık ortalama) 200-211, 307
truncating (kesme) 51-60
Tukey's hinges (Tukey menteşeleri) 180, 281
two dimensional tables (iki boyutlu çizelgeler) 223

U

unbiased estimator (yansız kestirimci) 79, 80, 233
unbiasedness (yansızlık) 77, 80
upper far outlier (uzak üst dışadüşen) 324
upper group (üst öbek) 343
upper outlier (üst dışadüşen) 286
upper quartile (üst dördebölen) 180, 226, 227, 240, 282, 285
upper threshold (üst eşik) 325

V

variable (değişken) 25
variance (varyans) 191, 229-237
variation (değişkenlik, yayılım) 167, 209, 223, 306, 309, 310, 312

VISUAL STATISTICS 13

volatility (oynaklık) 167, 223

W

weighted average (tartılı ortalama) 40, 185
weighted geometric mean (tartılı geometrik ortalama) 240

İstatistik formüllerinin ezberlenmesinden çok kavramların anlaşılması amaçlanan bu kitapta; karmaşık gibi görünen sayı kümelerinin arkasında yatan gizli yapıların nasıl ortaya çıkarılacağı açıklanmaya çalışıldı.

Verilen örnekler, araştırma soruları ya Türkiye'deki yaşamın içinden ya da bu ülkede yaşayan insanları ilgilendiren konulardan seçildi. Bebek ölüm oranlarından vergi oranlarına, fiyatlardan futbol karşılaşmalarına, TV izlenme oranlarından boşanma oranlarına, gelir dağılımından not dağılımına, Irak savaşından ortalama ömüre, kadınların yönetimde görev almasından üretimin dışarıya bağımlılığına, savunma bütçesinden içki tüketimine kadar değişen bir yelpaze, sayılar kullanılarak incelendi.

LİTERATÜR: YAYINCILIK, DAĞITIM,
PAZARLAMA, SANAYİ VE TİCARET LTD. ŞTİ
İSTİKLAL CADDESİ, EMGEN HAN,
NO: 47 KAT: 4 TR-34433
BEYOĞLU, İSTANBUL, TÜRKİYE
T 0(212) 292 4120
F 0(212) 245 5987
E literatur@literatur.com.tr
www.literatur.com.tr

ISBN 978-975-04-0283-8



LİTERATÜR®