

Başlarken

Nicel Veri Analizi

Teori ve Uygulama



Prof. Dr. Ünal H. ÖZDEN

Analitik Yaklaşımlarda Olası Hataların En Aza İndirilmesi İçin

TTDGA Süreci İzlenmelidir.

- Bir sorunu çözmek ya da bir hedefe ulaşmak için çalışmak istenilen verileri veya sorunu **Tanımlamak**
- Uygun kaynaklardan veri **Toplamak**
- Toplanan verileri **Düzenlemek**
- Grafiklerle verileri **Görselleştirmek**
- Sonuçları **Analiz** etmek ve yorumlamak

TTDGA Sayesinde

- Sorunu özetler ve veriler görselleştirilebilir
- Bu verilerden sonuçlara ulaşılabilir
- İş faaliyetleri hakkında güvenilir tahminler yapılabilir
- İş süreçlerini iyileştirilebilir

Temel Kavramlar

DEĞİŞKEN

Bir ögenin ya da birimin özelliklerinin her biri.

VERİ

Bir değişken ile ilişkili öge ya da birim değerlerinin kümesi.

İSTATİSTİK

Karar vermede verileri yararlı bilgilere dönüştürmeye yardımcı yöntemler bütünü. Ham verilerden bilgi üretme.

Bugünün iş dünyasında verilerden (sayılardan) kaçış yok.

- Bugünün dijital dünyasında ileri çalışmalar için herhangi bir olguya ait giderek artan miktarda veri toplanmakta, saklanmakta ve kullanılmaktadır.
- Her yerde veri kelimesini duymaktayız.
- Veriler dünya hakkındaki gerçekleri göstererek farklı alanlarda bilgiler sunar.

Karar Vermede Sayılar Faydalıdır

- Yapılan araştırma sonucu gençlerin %54'ü A marka ürünü, %24'ü B marka ürünü ve %22'si C marka ürünü tercih etmektedir.
- 3000'den fazla öğrenciye yapılan bir araştırmada öğrencilerin %51'i boş zamanlarını sosyalleşme, eğlence ve diğer faaliyetlerle, %19'u sınıf / laboratuvar çalışmalarına katılarak ve %7'i de kitap okuyarak geçirmektedir.

Analitik Yaklaşımlar Olmadan Olmaz

- Araştırmadaki sayıların yararlı bilgiler olup olmadığını belirlemek
- Belirsizlik altında karar vermek
- Sınıflama ve kümeleme
- Nedensellik iddialarını doğrulamak (Değişimin ve farklılaşmanın sebebi)
- Büyük miktarda verilerin ortaya koyduğu kalıpları görmek
- Tahmin

Doğru Veriler Analitik Yaklaşımlar İçin Çok Önemlidir



İstatistik nedir?

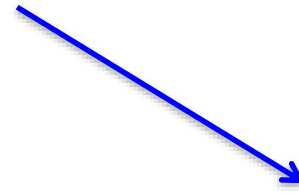
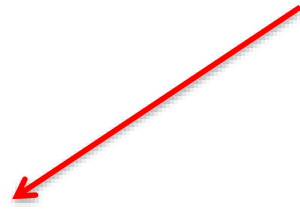
- Üç çeşit yalan vardır:
 - Yalan
 - Kuyruklu yalan
 - İstatistik

Benjamin Disraeli

İstatistiğin Konusu Olan Olaylar

İstatistik olaylarla ilgilenir.

Olayları ikiye ayırmak mümkündür.



Toplu olaylar

Bir çok faktör tarafından etkilenen olaylardır. İstatistiğin konusu kapsamındadır.
(Enflasyon, başarıya etki eden faktörler...)

Tekil olaylar

Tek bir faktör tarafından etkilenen olaylardır. İstatistiğin konusu kapsamında değildir. Belirli şartlar birleştiğinde daima aynı sonucu verir. (Kimyasal olaylar...)

İstatistiğin İki Dalı

İstatistik

Karar vermede verileri yararlı bilgilere dönüştürmeye yardımcı yöntemler bütünü. Ham verilerden bilgi üretme.



Tanımsal İstatistik

Verileri toplama, düzenleme, görselleştirme, analiz etme ve yorumlamadan oluşan süreci kapsar.



Çıkarımsal İstatistik

Küçük bir gruptan (örnekten) toplanan verileri kullanarak daha büyük bir grup (anakitle) hakkında sonuçlara varmak için kullanılır.

Tanımsal İstatistik

- Verilerin toplanması

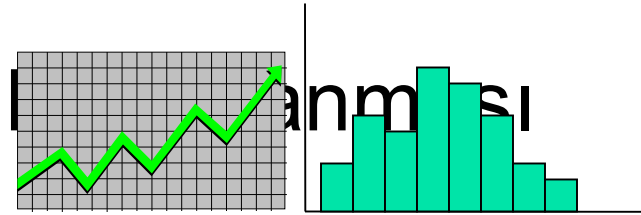
- Ör. Anket

- Verilerin düzenlenmesi ve sunulması

- Ör. Tablolar ve grafikler

- Karakteristik değerlerin bulunması

- Ör. Örnek ortalaması =



$$\frac{\sum X_i}{n}$$

Çıkarımsal İstatistik

- Tahmin
 - Ör. Anakitle ağırlık ortalamasının örnek ortalamasından yararlanarak tahmin edilmesi.
- Hipotez testleri
 - Ör. Anakitle ortalama ağırlığının 75 kg olduğu iddasının testi.



Örnekten hesaplanan sonuçlara göre anakitle hakkında karar verilir.

İstatistiğin Kullanım Alanları

- İşletmelerde; insan kaynakları, finansal analiz, Pazar arařtırmaları, tedarik zinciri gibi...
- Psikoloji
- Sosyoloji
- Ekonomi
- Tıp
- Biyoloji
- Fizik
- Mühendislik
- vs.

İş analizi: İstatistik Değişen yüzü

- Öngörülemeyen ilişkileri açığa çıkarmak için verileri analiz etmek ve keşfetmek için istatistiksel yöntemleri kullanın.
- Bir sistemin strateji, planlama ve operasyonlarını etkileyen optimizasyon modellerini geliştirmek için yönetim bilimi yöntemlerini kullanın.
- Verileri toplamak ve tüm yönleri ile verileri işlemek için bilgi sistemlerini kullanın. Aksi takdirde verileri etkin bir şekilde incelemek zor olur.

“Big Data” ve İş Analizi

- “Big Data” hala bulanık bir kavramdır.
- Yüksek hacimli verilerin otomatik olarak toplanabilmesi nedeniyle çok büyük veri setleri hızlı oranda artmaktadır.
- Eski istatistiksel tekniklerde büyük verilerin kullanımı pratik değildir.

İstatistikler, İşinizin Önemli Bir kısmıdır

- İşiniz giderek artan veri odaklı, analitik beceri gerektirir.
- Çalışmalar iş analitiğini uygulayan kuruluşların verimlilik, inovasyon ve rekabet gücünün arttığını göstermektedir.
- Hal Varian, Google Inc. Baş Ekonomist, "Önümüzdeki 10 yıl içinde en iyi meslek istatistik ve işletme analistliği olacak. Ve ben şaka yapmıyorum."

Software (Bilg. Paket Programı) ve İstatistik

- Software, istatistiksel yöntemleri uygularken hesaplamalarda size yardımcı olacak programlardır.
- Microsoft Excel ile istatistiksel veri analizi yapabilirsiniz.
- Bir çok istatistik paket programı vardır. En bilinenleri;
 - SPSS
 - Minitab
 - R
 - Eviews
 - SAS

Verilerin Tanımlanması ve Toplanması

Nicel Veri Analizi

Teori ve Uygulama



Prof. Dr. Ünal H. ÖZDEN

Temel Kavramlar

Değişken: Herhangi bir ögenin veya birimin herhangi bir özelliği

Veri (Data): Herhangi bir değişkenin birimlerine ilişkin değerler kümesi

İstatistik: Karar vermede verilerden yararlanarak, yararlı bilgiler üretmeye yardımcı yöntemler bütünü veya ham verilerden bilgi üretme süreci

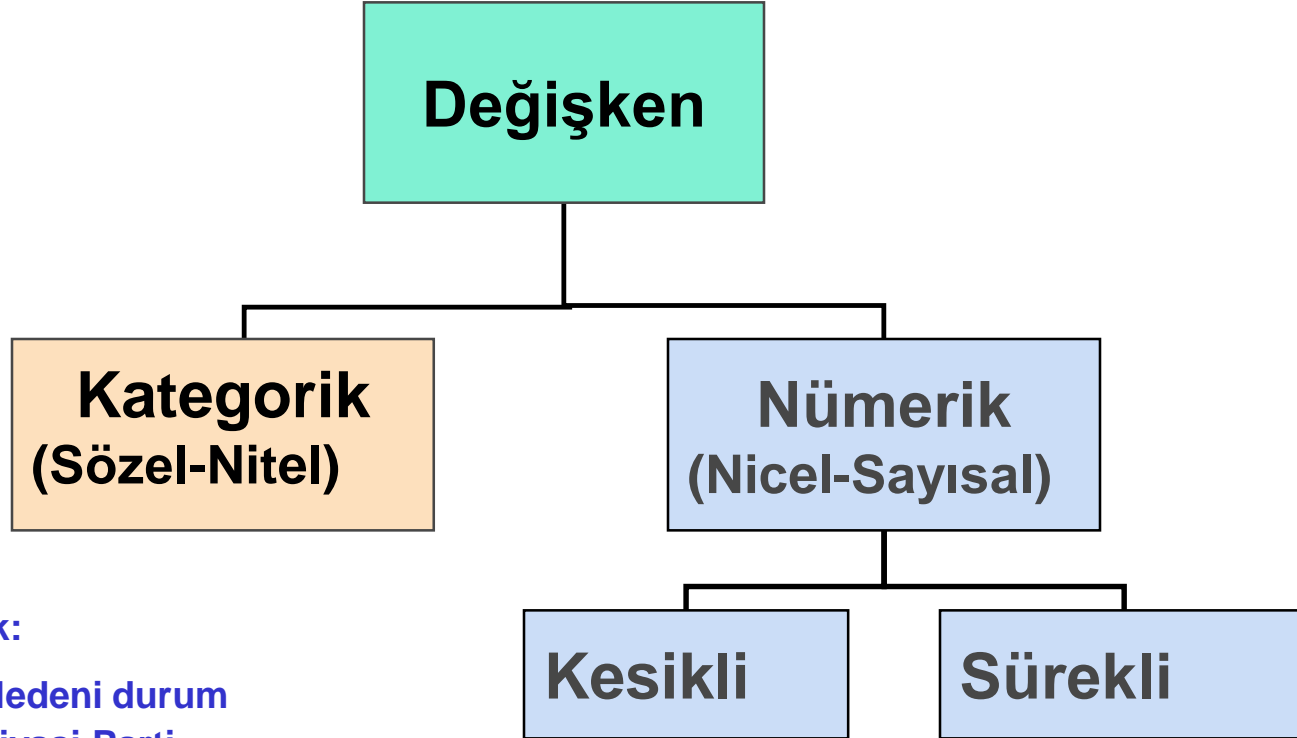
Değişken Türleri

- **Kategorik** (*Kalitatif-Nesnel-Sözel*) değişkenler verileri yalnızca kavramsal olan değişkenlerdir. Sayılabılırlardır. (kadın –erkek; evet-hayır gibi...)
- **Sayısal** (*Kantitatif-Nicel*) değişkenler verileri sayılar olan değişkenlerdir ve ölçülebilir ve sayılabılırlardır. (boy, kilo, yaş vs. gibi).
 - **Kesikli değişken**
 - **Sürekli değişken**

Tanımlar karışıklık ve hatalardan kaçınmak için çok önemlidir.

- Tanımlar ortak bir anlayışla açık ve kesin anlam sağlayan bir ifadedir.
- Tanımlar olmadığında iletişimsizlik ve hatalar oluşur.
- Tanım(lar) TCOVA ve tanımlama adımının önemli bir parçasıdır

Değişken Türleri



Örnek:

- Medeni durum
 - Siyasi Parti
 - Göz rengi
- (Kategorilerin tanımlanması ile ortaya çıkar)

Örnek:

- Çocuk sayısı
 - Satılan araba sayısı
- (Sayım işleminde ortaya çıkar)

Örnek:

- Ağırlık
 - Voltaj
- (Ölçüm işleminde ortaya çıkar)

Ölçüm Düzeyleri ve Ölçekler

Oransal



Aralıklı



Sıralı



İsimsel

Bu ölçme düzeyi, aralıklı ölçme düzeyinin bütün özelliklerine sahiptir. Aralıklı ölçme düzeyinden farklı olarak; oransal ölçekte sıfır gerçek yokluğu ifade eder ve iki sayı arasında oransal ilişki vardır.

Örnek:

Boy, yaş, haftalık tüketilen gıda miktarı...

Bütün sıralı veri türlerini kapsar, değerler arasındaki uzaklık sabit büyüklüktedir, sayılar arasında oransal ilişki yoktur ve sıfırın gerçek bir yokluğu ifade etmez.

Hava sıcaklığı, standartlaştırılmış sınav skoru...

Veriler farklı sıralı kategorilere göre sınıflandırılır. Nominal ölçme düzeyi ile sıralı ölçme düzeyi arasındaki temel farklılık, sıralı ölçme düzeyi sınıfları arasında '... den daha iyi' ilişkisinin olmasıdır.

Hizmet kalite puanı, ürün memnuniyeti, akademik ünvan, S & P derecelendirmesi, Öğrenci bağıl notu (harf olarak)...

Nominal ölçekte veriler için hiçbir sıralama yoktur. Veriler farklı kategorilere göre sınıflandırılır.

Medeni durum, araba markası, facebook profili sahipliği, yatırım türü...

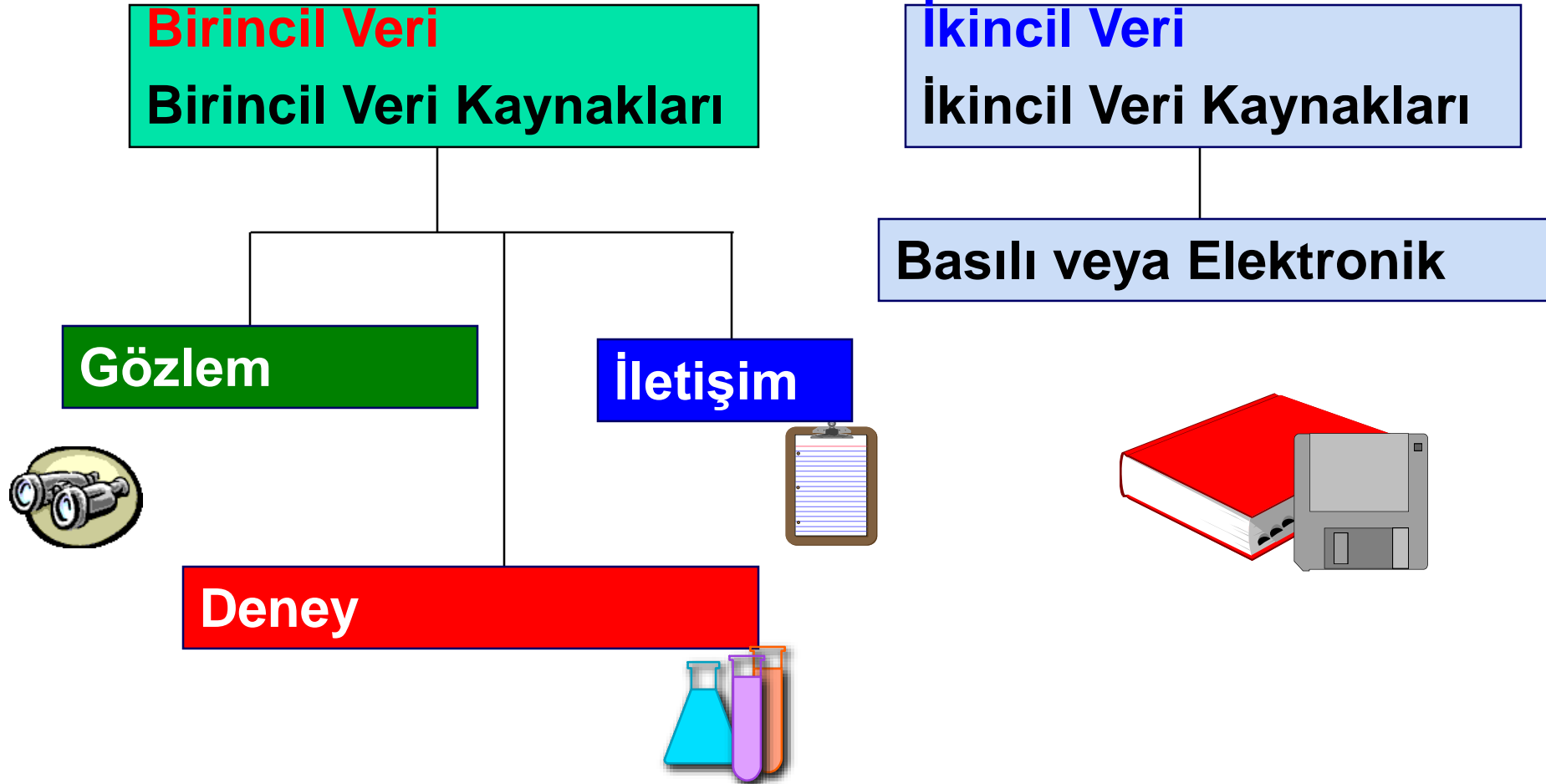
Dođru Veri Toplama İstatistiksel Analiz İin ok Kritiktir

- Yanlı(yanıltıcı), kusurlu veya diđer hatalı trlerdeki verilerden kaınmak
- Dođru olmayan verilerden elde edilen sonular Őüpheli ya da hatalı olacaktır.
- En iyi istatistiksel yntemler bile, veri kusurlu olduđunda hatalı sonulara ve kararlara neden olacaktır.

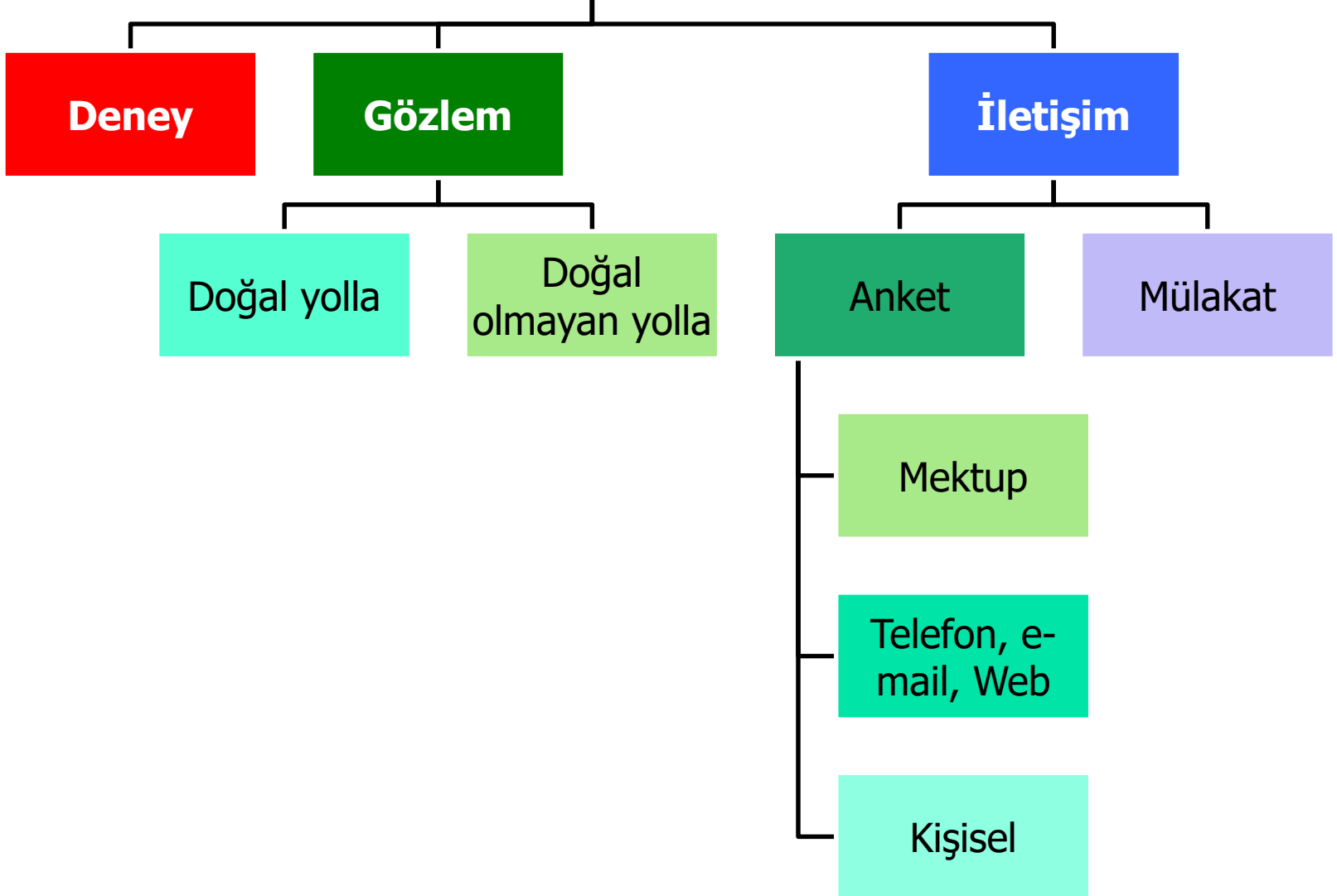
Veri Kaynakları

- **Birincil Veriler:** Veri analizi yapacak kişi/kişiler tarafından toplanmış veriler
 - Siyasetle ilgili anketlerden elde edilen veriler
 - Deneylerden elde edilen veriler
 - Gözlemlerden elde edilen veriler
- **İkincil Veriler:** Veri analizi yapacak kişi(ler)den farklı kişiler tarafından toplanmış veriler
 - Nüfus sayımı verileri
 - İnternet veya basılı yayınlardaki yer alan veriler

Veri Toplama



Birincil Veri Kaynakları

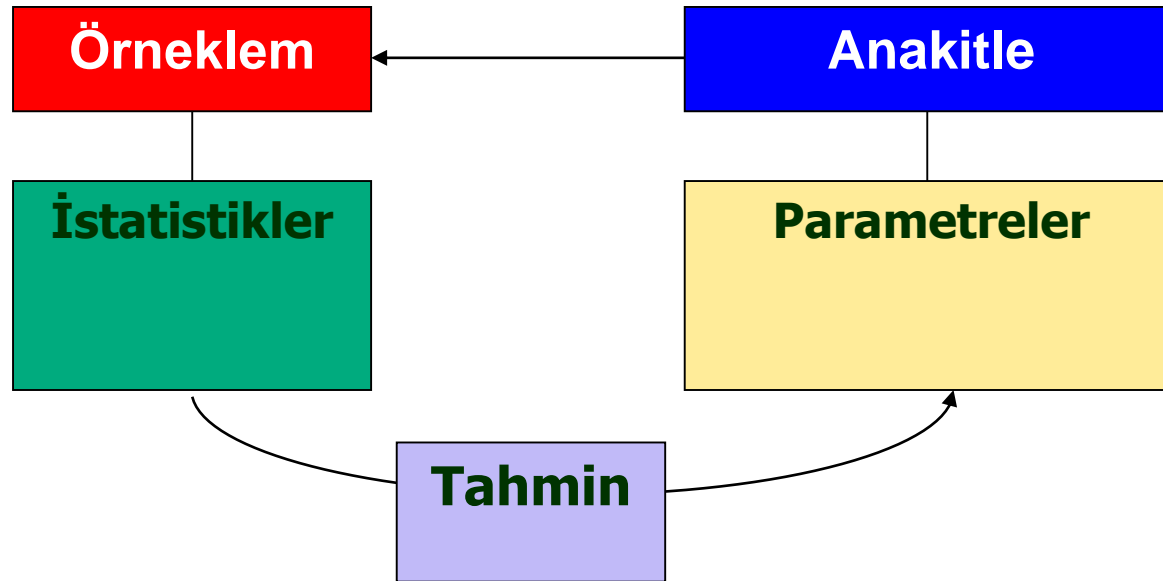


Tanımlar

- **Anakitle:** Araştırmaya konu olan birimlerin oluşturduğu kümeye denir.
- **Örnekleme:** Belli kurallara göre, belli bir anakitleden seçilmiş ve seçildiği anakitleyi temsil yeterliliği olan alt kümedir. (temsil gücü ve yeterlilik)
- **Parametre:** Anakitleyi tanımlamak için hesaplanan karakteristik değerler
- **İstatistik:** Örnekten hesaplanan karakteristik değerler
- **Tamsayım:** Anakitleyi oluşturan birimlerin tamamının sayılması
- **Örnekleme:** Bir araştırmanın konusunu oluşturan anakitlenin bütün özelliklerini yansıtan bir parçasının seçilmesi ve seçilen bu örneklemden yararlanarak hesaplanan karakteristik değerlerden (istatistik) yararlanarak anakitle karakteristik değerlerinin (parametre) tahmin edilmesi
- **Birim:** Anakitleyi oluşturan en küçük parça. Birim tekil olmak zorunda değildir.
- **Karakteristik Değer (tanımlayıcı ölçüler):** Herhangi bir verinin veya değişkenin özelliklerini tanımlamak için hesaplanan değerlerdir (aritmetik ortalama, mod, medyan, standart sapma vb...)

Parametre ve İstatistik

- Parametre
- İstatistik



Anakitle ve Örneklem

Anakitle



Araştırmaya konu olan birimlerin oluşturduğu kümedir

Örneklem



Bir anakitleden seçilmiş ve seçildiği anakitleyi temsil yeterliliği olan alt kümedir.

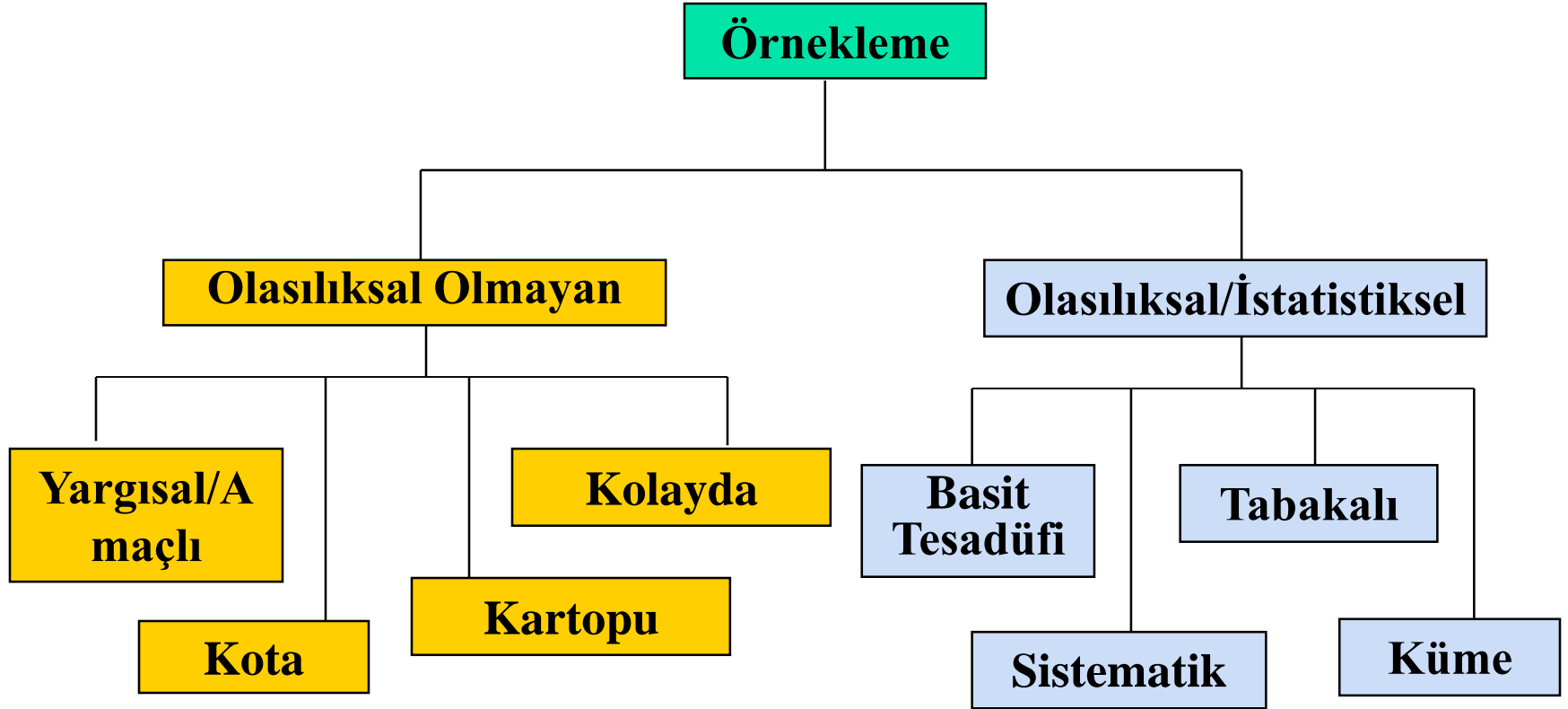
Niçin Örneklememe?

- Anakütleye ulaşılammaması
- Zaman
- Maliyet
- Kolaylık

Veri Kaynakları ve Veriler ile İlgili Bilinmesi Gerekenler

- Veri kaynağı yapılandırılmış veya yapılandırılmamış mı?
 - Yapılandırılmamış/Düzenlenmiş
 - Yapılandırılmış/Düzenlenmemiş
- Elektronik veriler hangi formatta yer almaktadır?
- Veriler nasıl kodlanmış?
 - Veriler kodlanmış mı?
 - Kodlanmış verilerin tekrar orijinal hale dönüştürülmesi gerekir mi?
- Veri temizlemesi yapılmış mı?
 - Veri yanlışlıklar, Kayıp veriler, Uç değerler...
 - Tanımlanamayan veriler vs.

Örnekleme Yöntemleri



Olasılıksal Olmayan Örnekleme Yöntemleri

Olasılıksal olmayan örnekleme, birimlerin seçiminde keyfi seçim yönteminin uygulandığı örnekleme yöntemleridir.

Kolayda (Gelişigüzel) Örnekleme: Kolayca ulaşılabilir birimleri seçmek suretiyle bir örnek oluşturulmaya çalışılır. Örneklemede birimlerinin seçimi görüşmeci tarafından doğru zamanda doğru yerde bulunan birimler, gönüllü katılımcılar arasından yapılır. Herhangi bir fakülteye gidip saptanacak sayıda rastlanan öğrenciyi örnekleme alma

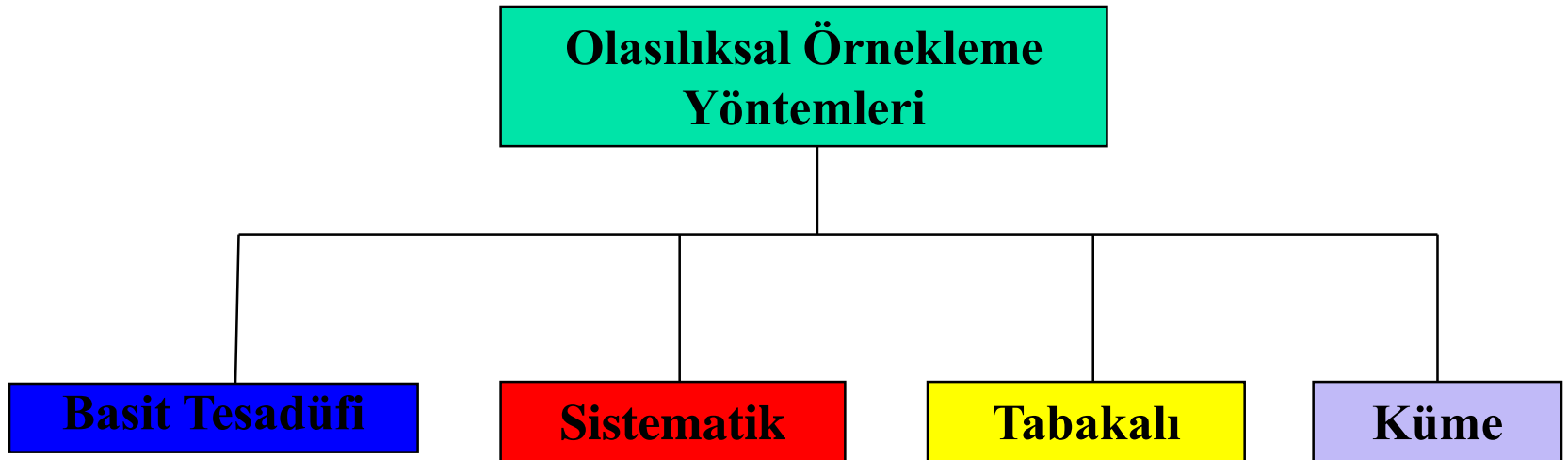
Yargısal Örnekleme: Birimlerin seçiminin araştırmacının amacına, arzu, düşünce ve deneyimlerine dayanarak yapılmasıdır. Meslek hastalıklarıyla ilgili yapılacak bir araştırmada örneklemin, meslek hastalıklarının tüm anakitle içinden değil, özellikle belli bir hizmet süresini aşmış ya da belli bir yaş sınırının üstündekiler arasından seçmesi gibi.

Kota Örnekleme: Bu yöntemde tabakalı örnekleme yönteminde olduğu gibi anakitle alt tabakalara ayrılır. Her alt tabakanın temsili için kota konulur. Bu kota belirlenen tabakanın anakütleye oranına göre belirlenir. Kota örneklemede örneğe girecek elemanlar tesadüfen değil araştırmacını kendi isteğine göre belirlenir.

Kartopu Örnekleme: Anakitleye ulaşmak mümkün olmadığında, ulaşabilen ilk birim belirlenir. Bu birimden elde edilen bilgilerle diğer birimlere ve bu şekilde zincirleme olarak anakitleyi temsil eden örneğe ulaşılmaya çalışılır.

Olasılıksal Örnekleme Yöntemleri

- Olasılık örnekleme yöntemlerinde, birimler bilinen olasılıklara bağlı olarak seçilir.



Basit Tesadüfi Örnekleme

- Anakitlede yer alan her bir birimin örneklem kümesine girme şansı var ve bu şanslar eşit
- Seçimler iadeli olarak yapılabilir.
- Birimler tesadüfi sayılar tablosu veya bilgisayar yardımı ile çekilebilir.
- Anakütle incelenen konu açısından HOMOJEN yapıda olduğunda iyi sonuç verir
- Anakitleyi oluşturan birimlere birer numara verilir ve rasgele bu numaralar çekilir.

Basit Tesadüfi Örnekleme

850 birimlik
anakitleden
örneklem seçimi

<u>Birim adı</u>	<u>Birim no</u>
Bev R.	001
Ulan X.	002
.	.
.	.
.	.
.	.
Joann P.	849
Paul F.	850

Rasgele Sayılar Tablosu

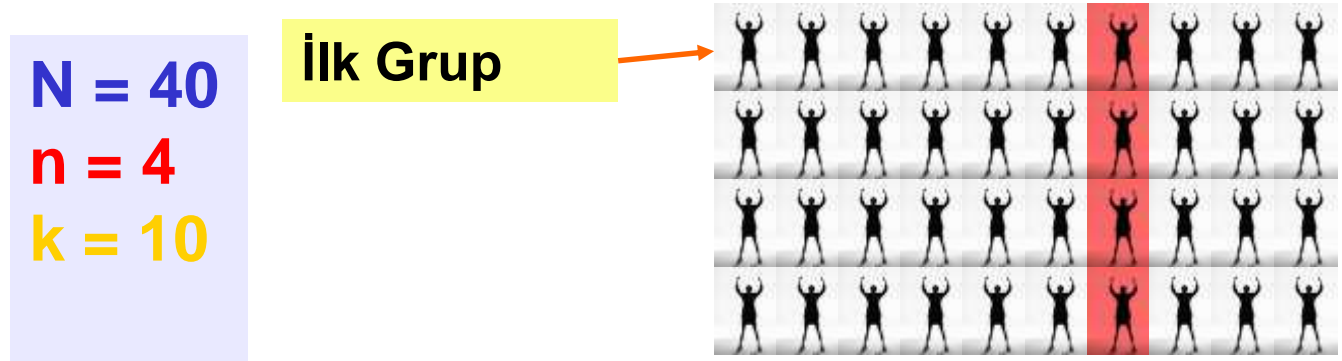
49280 88924 35779 00283 81163 07275
11100 02340 12860 74697 96644 89439
09893 23997 20048 49420 88872 08401

Örneklem seçilen ilk 5 birim

Birim 492
Birim 808
Birim 892 -- iptal böyle bir gözlem yok
Birim 435
Birim 779
Birim 002

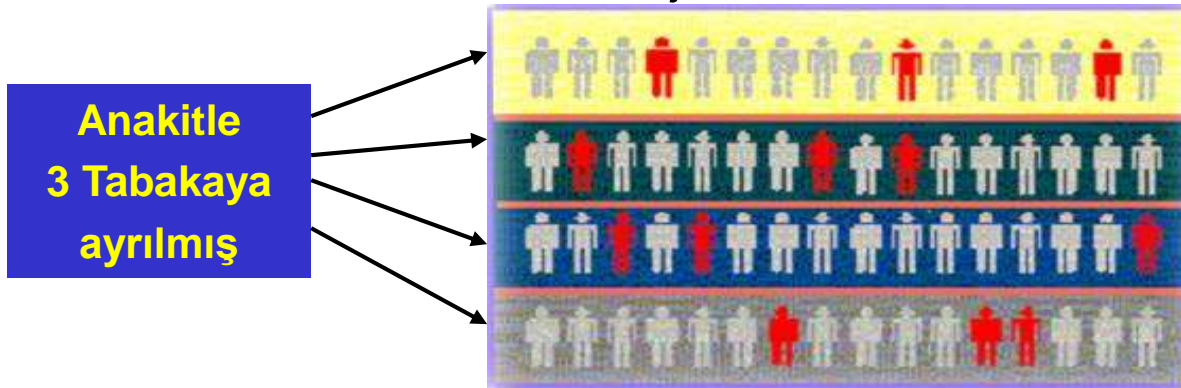
Sistematik Örnekleme

- Anakitle birimlerini kurala göre numaralandırılır (1...N) ve örneklem büyüklüğünü (n) belirlenir
- Örnekleme oranı k'yı ($k=N/n$) hesaplanır ve anakitle sıra numarasına göre her biri k birimden oluşan n gruba ayrılır.
- 1 ile k arasında rasgele bir rakam (s) seçilir.
- Her gruptaki s'inci sıradaki birim örneklem kümesine dahil edilir.



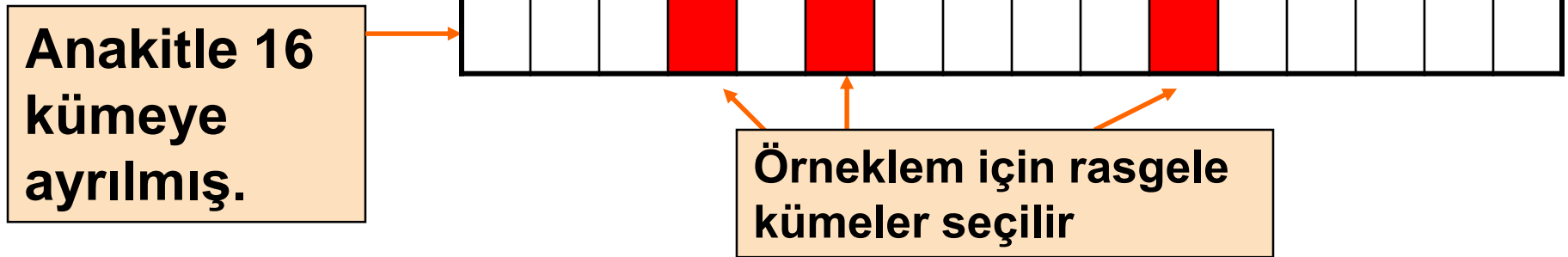
Tabakalı Örneklem

- Homojen olmayan anakitle birimleri, karakteristik özelliklerine göre tabaka denilen homojen alt gruplara ayrıştırılır
- Her tabakadan anakitle içindeki oranına bağlı olarak basit tesadüfi örnekleme yöntemi ile birimler seçilir
- Bu tabakalardan seçilen birimler birleştirilerek örneklem oluşturulur
- Çok yaygın kullanılan bu teknikte tabakalar kendi içinde homojen birbirleri arasında heterojendir.



Küme Örnekleme

- Anakitle, anakitleyi temsil eden birden fazla “küme”ye bölünür
- Kümeler arasından basit tesadüfi örneklem ile rasgele seçim yapılır
- Seçilen küme içindeki tüm birimler örneklem içinde yer alır veya seçilen kümelerdeki birimler başka bir örneklem tekniğinde kullanılabilir
- Kümeler kendi içinde heterojen, kümeler arasında homojendir.



Örnekleme Yöntemlerinin Karşılaştırılması

- Basit tesadüfi örnekleme ve sistematik örneklem
 - Kullanımı kolay
 - Anakitle özelliklerini için temsil sorunu yaşanabilir.
- Tabakalı örnekleme
 - Anakitleyi oluşturan ve farklı karakteristiklere sahip tüm birimlerin temsil edilmesini sağlar.
- Küme örnekleme
 - Daha düşük maliyetlidir.
 - Daha az etkindir. Etkinliğin ve temsiliyetin diğerleri kadar olabilmesi için daha yüksek örneklem büyüklüğüne ihtiyaç vardır.

Hata Türleri

- **Kapsama hatası veya seçim yanlılığı**
 - Bazı gruplar çerçeve dışında kalması nedeniyle seçilme şansları yoktur
- **Cevaplamama hatası**
 - Cevap vermeyen insanlar cevap verenlerden farklı kanaate sahip olabilir
- **Örnekleme hatası**
 - Herzaman var olur ve örneklemden örnekleme değişkenlik gösterir.
- **Ölçme hatası**
 - Yanlış ve yönlendirici hazırlanmış sorular nedeniyle yanlı cevaplar olacaktır.

Hata Türleri

(continued)

- Kapsam Hatası



Listeden dışlama

- Cevaplama Hatası



Cevaplamayanları takip

- Örnekleme Hatası



Örneklemden
örnekleme tesadüflük farkı

- Ölçme Hatası



Kötü ve yönlendirici sorular

Verilerin Düzenlenmesi ve Görselleştirilmesi

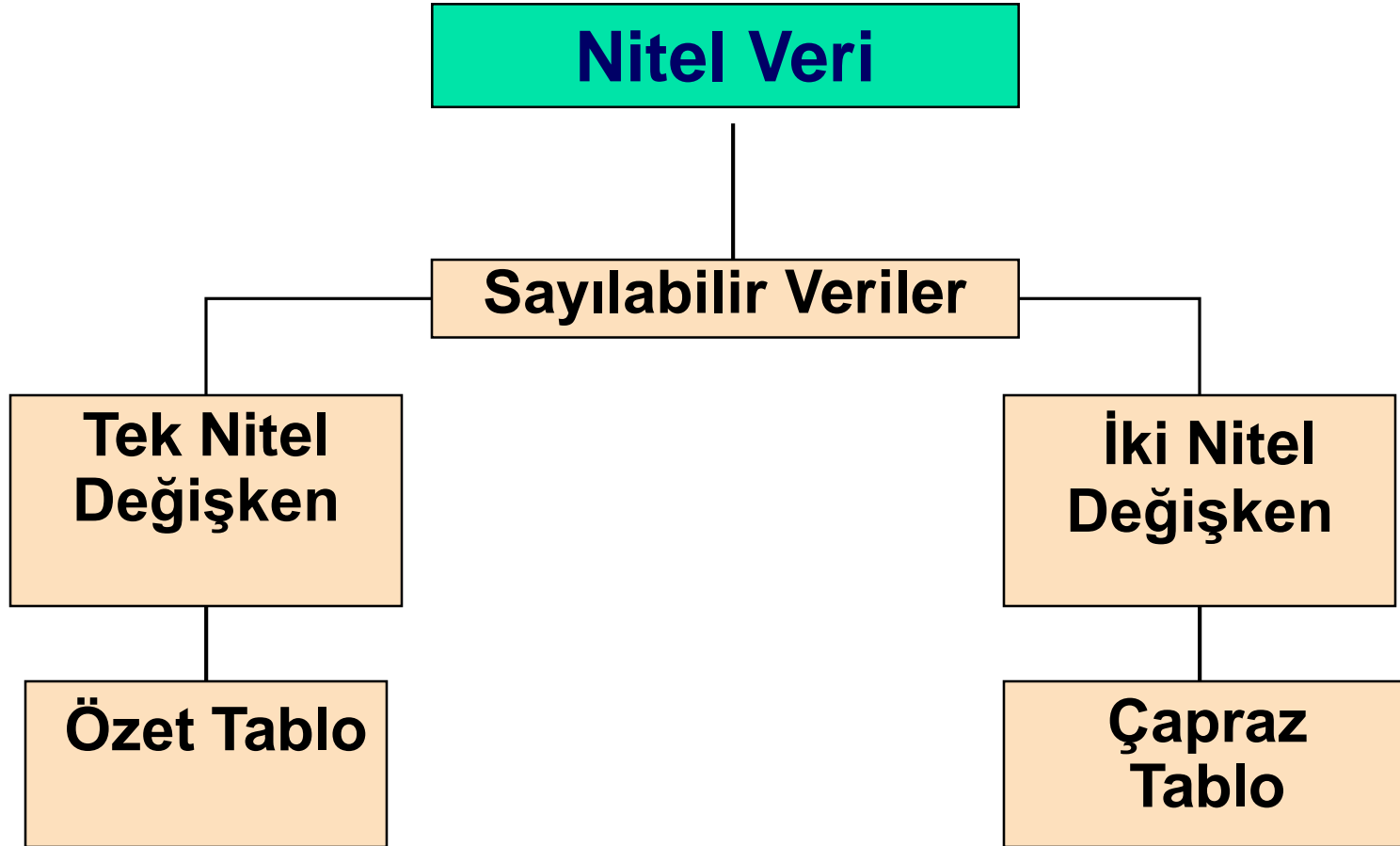
Nicel Veri Analizi

Teori ve Uygulama



Prof. Dr. Ünal H. ÖZDEN

Nitel Verilerin Tablolar Yardımı ile Düzenlenmesi



Nitel Verilerin Düzenlenmesi: Özet Tablo

- Özet tablo, kategoriler arasındaki farkları görebilmemiz için bir gruptaki öğelerin frekanslarını veya yüzdelerini belirtir.

Gençlerin Online Alışveriş Yapmalarının Temel Nedenleri

Online Alışverişi Tercih Etme Nedeni?	Percent
Uygun fiyat	37%
Kalabalık kaçınmak veya zaman harcamak istememek	29%
Kolyalık	18%
Daha iyi seçim yapabilmek	13%
Diğer	3%

Kaynak: "Main Reason Young Adults Shop Online?"
USA Today, December 5, 2012, p. 1A.

Çapraz Tablo Birden Fazla Nitel Değişkenin Düzenlenmesinde Kullanılır

- İki veya daha fazla kategorik değişkenin grupları arasında var olabilecek kalıpları (bağımlılıkları) incelemek için kullanılır.
- Çapraz tablolar ile nitel değişkenlerin grupları arasındaki eşleşmeler gösterilir.
- Bir değişkenin grupları satırlarda, diğer değişkenin grupları sütunlarda yer alır.

Örnek Çapraz Tablo

- Rasgele çekilen 400 Fatura.
- Her bir fatura «küçük miktarlı», orta miktarlı ve yüksek miktarlı diye kategorize edilmiştir.
- Her bir fatrunun hatalı olup olmadığı kontrol edilmiştir.
- Bu veriler çapraz tablo ile düzenlenmiştir.

Hatalı Olup Olmamaile Fatura Miktarına İlişkin Çapraz Tablo

	Hatasız	Hatalı	Toplam
Küçük Miktarlı	170	20	190
Orta Miktarlı	100	40	140
Yüksek Miktarlı	65	5	70
Toplam	335	65	400

Genel Toplama Göre Yüzdesele Değerlerin Yer Aldığı Çapraz Tablolar

	Hatasız	Hatalı	Toplam
Küçük Tutarlı	170	20	190
Orta Tutarlı	100	40	140
Yüksek Tutarlı	65	5	70
Toplam	335	65	400

$$\begin{aligned} \%42,50 &= 170 / 400 \\ \%25,00 &= 100 / 400 \\ \%16,25 &= 65 / 400 \end{aligned}$$

	Hatasız	Hatalı	Toplam
Küçük Tutarlı	%42.50	%5.00	%47.50
Orta Tutarlı	%25.00	%10.00	%35.00
Yüksek Tutarlı	%16.25	%1.25	%17.50
Toplam	%83.75	%16.25	%100.0

Örneklemedeki faturaların %83,75'i hatasızdır ve örneklemedeki faturaların %47,50'si küçük tutarlıdır.

Satır Toplamlarına Göre Yüzdesele Değerlerin Yer Aldığı Çapraz Tablolar

	Hatasız	Hatalı	Toplam
Küçük Tutarlı	170	20	190
Orta Tutarlı	100	40	140
Yüksek Tutarlı	65	5	70
Toplam	335	65	400

$$89,47\% = 170 / 190$$

$$71,43\% = 100 / 140$$

$$92,86\% = 65 / 70$$

	Hatasız	Hatalı	Toplam
Küçük Tutarlı	%89,47	%10,53	%100,0
Orta Tutarlı	%71,43	%28,57	%100,0
Yüksek Tutarlı	%92,86	%7,14	%100,0
Toplam	%83,75	%16,25	%100,0

Orta tutarlı faturalarda (% 28,57), küçük (% 10,53) veya büyük (% 7,14) faturalardan daha çok hata yapma ihtimali vardır.

Sütun Toplamlarına Göre Yüzdesele Deęerlerin Yer Aldığı Çapraz Tablolar

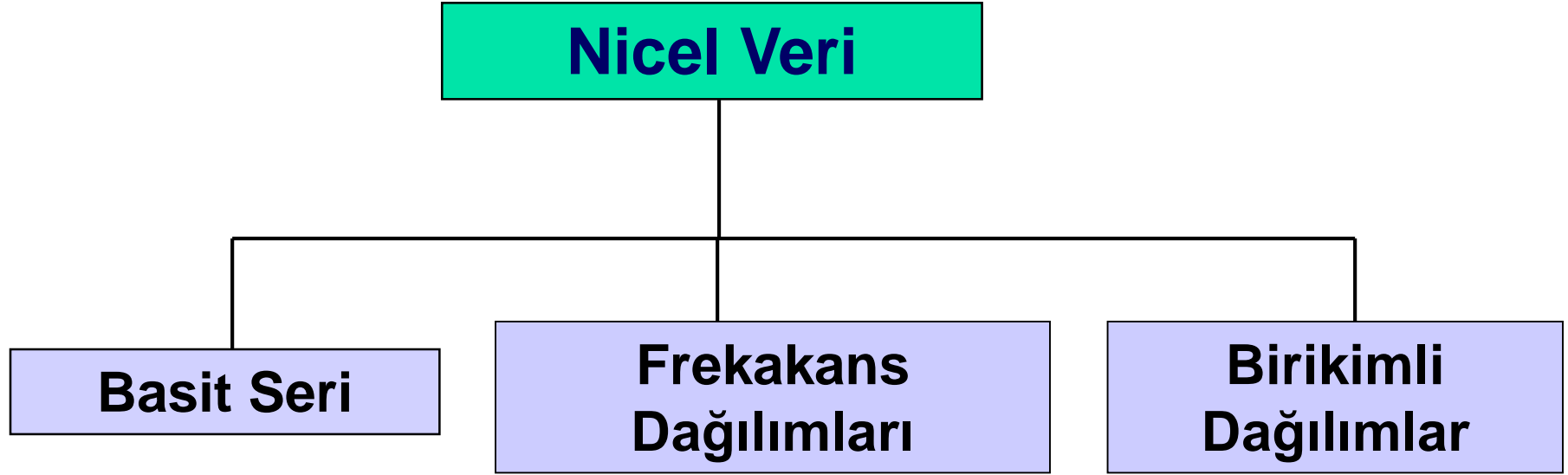
	Hatasız	Hatalı	Toplam
Küçük Tutarlı	170	20	190
Orta Tutarlı	100	40	140
Yüksek Tutarlı	65	5	70
Toplam	335	65	400

$$\begin{aligned} \%50,75 &= 170 / 335 \\ \%30,77 &= 20 / 65 \end{aligned}$$

	Hatasız	Hatalı	Toplam
Küçük Tutarlı	%50,75	%30,77	%47,50
Orta Tutarlı	%29,85	%61,54	%35,00
Yüksek Tutarlı	%19,40	%7,69	%17,50
Toplam	%100,0	%100,0	%100,0

Hatalı faturaların orta büyüklükte olma ihtimali %61,54'tür.

Nicel Verilerin Düzenlenmesi(nde Kullanılan Tablolar)



Nicel Verilerin Düzenlenmesi: Basit Seri

- Basit seri, en küçük değerden en büyük değere doğru verilerin sıralanması ile elde edilir.
- Değişim aralığını gösterir (minimum değer ve maksimum değer)
- Aykırı (uç) değerleri belirlemeye yardımcı olabilir (olağandışı gözlemler)

Öğrencilerin Yaşları	Birinci Öğretim Öğrencileri					
	16	17	17	18	18	18
19	19	20	20	21	22	
22	25	27	32	38	42	
İkinci Öğretim Öğrencileri						
18	18	19	19	20	21	
23	28	32	33	41	45	

Nicel Verilerin Düzenlenmesi: Frekans Serisi

- Frekans seileri 2 sütun halinde düzenlenir.
- Sütunların ilkinde gözlem değerleri (X_i), ikincisinde gözlemlerin tekrarlanma sayıları (sıklık) (f_i) yer alır.

Birinci Öğretim Öğrenci Yaşları (X_i)	Öğrenci Sayıları Frekanslar (f_i)
16	1
17	2
18	3
19	2
20	2
21	1
22	2
25	1
27	1
32	1
38	1
42	1
Toplam	18

Nicel Verilerin Düzenlenmesi: Frekans Serisi

X_i	f_i	Oransal Frekanslar	Artan Birikimli	Azalan Birikimli	Artan Oransal Birikimli	Azalan Oransal Birikimli
16	1					
17	2					
18	3					
19	2					
20	2					
21	1					
22	2					
25	1					
27	1					
32	1					
38	1					
42	1					
Toplam	18					

Nicel Verilerin Düzenlenmesi: Sınıflı (Gruplandırılmış) Seriler

- Sınıflı seriler, verilerin sayısal olarak düzenlenmiş sınıflara yerleştirildiği özet bir tablodur.
- Sınıflı seriler için uygun sınıf sayısı ve sınıf aralığını belirlemeye ve çakışmayı önlemek için her sınıfın sınırlarını belirlemeye dikkat edilmelidir.
- Sınıfların sayısı, verilerdeki değerlerin sayısına bağlıdır. Çok sayıda değer olduğunda daha fazla sınıf sayısı olur. Genelde, bir Sınıflı seri en az 5, en fazla 15 sınıfa sahip olmalıdır. Bu da sınıf aralığının enişliğine bağlıdır.
- Bir sınıf aralığının genişliğini belirlemek için, değişim (En yüksek değer – En düşük değer) aralığı istenen sınıf sayısına bölünür.

Nicel Verilerin Düzenlenmesi: Örnek Sınıflı Seri

Örnek: Yalıtım üreticisi, 20 kış gününü rastgele seçmiş ve günlük hava sıcaklığı değerlerini aşağıdaki gibi belirlemiştir.

24, 35, 17, 21, 24, 37, 26, 46, 58, 30, 32, 13, 12, 38, 41, 43, 44, 27, 53, 27

Nicel Verilerin Düzenlenmesi: Örnek Sınıflı Seri

- Ham verileri küçükten büyüğe doğru sıralarak basit seri olarak düzenleyiniz:
12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58
- Değişim aralığını bulunuz: **58 - 12 = 46**
- Sınıf sayısını belirleyiniz: **5 (genellikle 5 ile 15)**
- Sınıf aralığını (genişliği) hesaplayınız: **10 (46/5 üste yuvarla)**
- Sınıf alt ve üst değerlerini (limitlerini) belirleyiniz:
 - Sınıf 1: **10 - 20**
 - Sınıf 2: **20 - 30**
 - Sınıf 3: **30 - 40**
 - Sınıf 4: **40 - 50**
 - Sınıf 5: **50 - 60**
- Sınıf orta değerlerini hesaplayınız: **15, 25, 35, 45, 55**
- Gözlem değerlerini sayınız ve ilgili sınıflara atayınız.

Nicel Verilerin Düzenlenmesi: Örnek Sınıflı Seri

Basit seri olarak düzenlenmiş veriler:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

Sınıflar	Sınıf Ortası	Frekanslar
10 - 20	15	3
20 - 30	25	6
30 - 40	35	5
40 - 50	45	4
50 - 60	55	2
Total		20

Nicel Verilerin Düzenlenmesi: Örnek Oransal & Yüzde Frekanslar

Basit Seri:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

Sınıflar	Frekans	Oransal Frekans	Oransal (Yüzde gösterim %)
10 - 20	3	0,15	%15
20 - 30	6	0,30	%30
30 - 40	5	0,25	%25
40 - 50	4	0,20	%20
50 - 60	2	0,10	%10
Toplam	20	1,00	%100

Nicel Verilerin Düzenlenmesi: Örnek Birikimli Frekanslar

Basit Seri:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

Sınıflar	Frekanslar	Oransal Frekanslar	Artan Birikimli Frekanslar	Artan Oransal Frekanslar
10 - 20	3	%15	3	%15
20 - 30	6	%30	9	%45
30 - 40	5	%25	14	%70
40 - 50	4	%20	18	%90
50 - 60	2	%10	20	%100
Toplam	20	100	20	%100

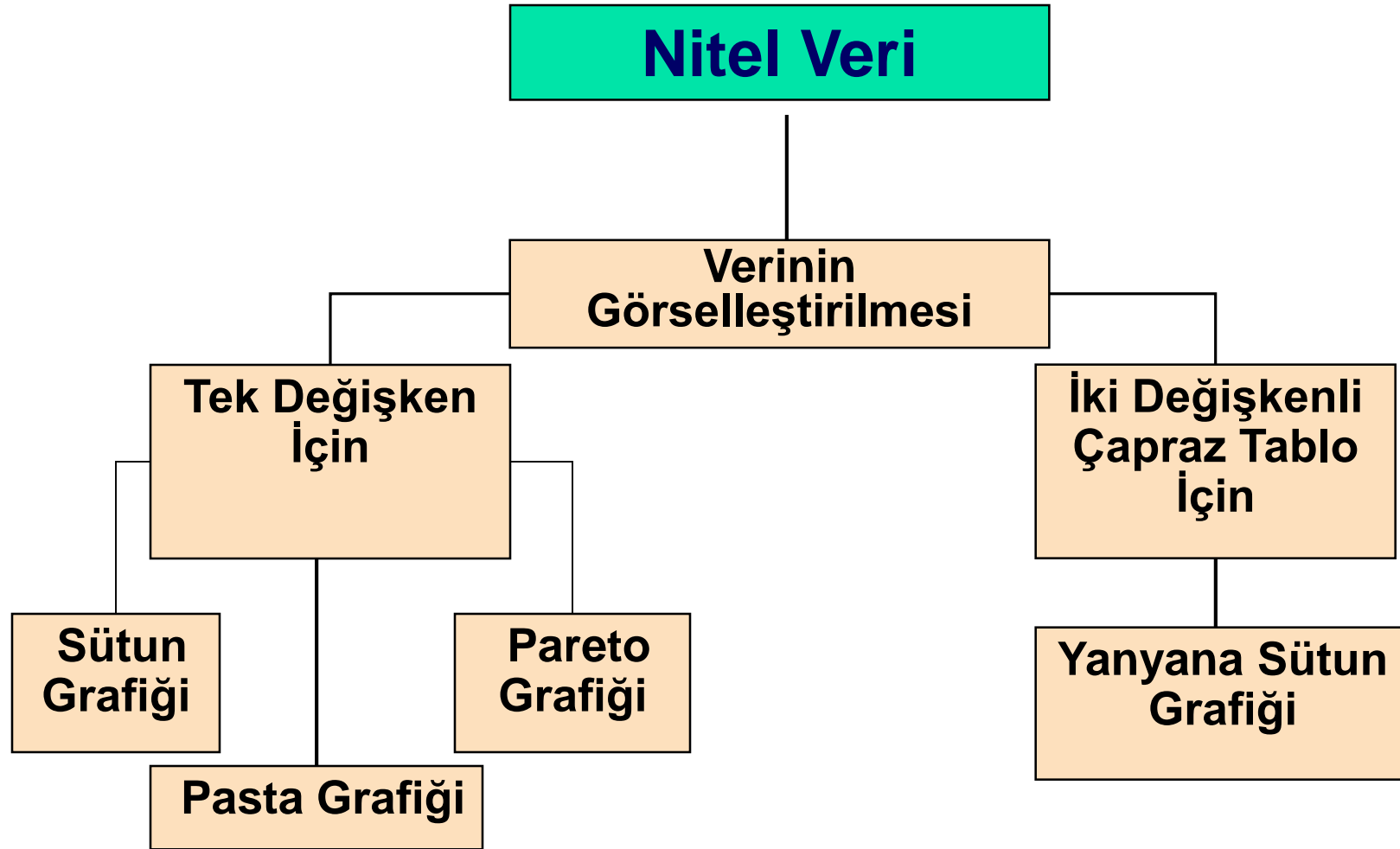
Frekans Dağılımları Niçin Kullanılır?

- Ham veriler daha kullanışlı bir forma dönüştürülmüş olur.
- Veriler daha hızlı görsel olarak gösterilmesini sağlar.
- Verilerin kümelenmiş olduğu yerler ve bu kümelenmelerin karakteristiklerinin belirlenmesine olanak tanır.

Sınıflı Seriler: Bazı İpuçları

- Aynı veriler için farklı sınıf aralıkları, farklı dağılımlara neden olabilir. (özellikle daha küçük veri kümeleri için)
- Farklı sınıf aralıkları seçildiğinde verilerin merkezi eğilimlerinde kaymalara neden olabilir.
- Veri kümesinin boyutu arttıkça, sınıf aralıklarının seçimindeki değişikliklerin etkisi büyük ölçüde azalabilir.
- Farklı örnek büyüklüklerine sahip iki veya daha fazla grubu karşılaştırırken, oransal frekansların kullanılması daha uygun olabilir.

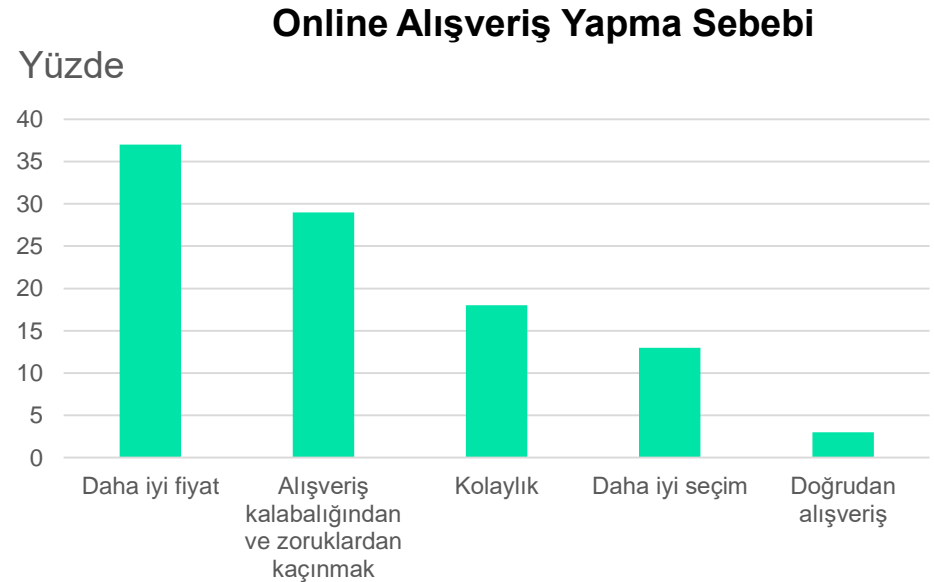
Grafik Yardımıyla Nitel Verilerin Görselleştirilmesi



Nitel Verilerin Görselleştirilmesi: Sütun Grafik

- Çubuk grafik, nitel bir değişkene ait kategorileri bir dizi çubuk olarak görselleştirir. Her çubuğun uzunluğu, her bir kategorinin sıklığını veya yüzdesini temsil eder. Her çubuk boşluk denilen bir alanla ayrılır.

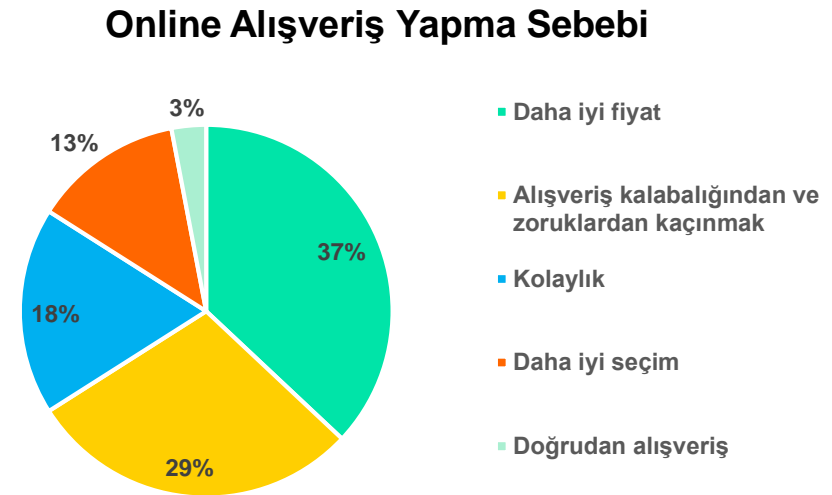
Online Alışveriş Yapma Sebebi	Yüzde
Daha iyi fiyat	37
Alışveriş kalabalığından kaçınmak	29
Kolaylık	18
Daha iyi seçim	13
Doğrudan alışveriş	3



Nitel Verilerin Görselleştirilmesi: Pasta Grafiği

Pasta grafiği, kategorileri temsil eden dilimlere ayrılmış bir çemberdir. Pastanın her diliminin büyüklüğü, her kategorideki yüzdeye göre değişir.

Online Alışveriş Yapma Sebebi	Yüzde
Daha iyi fiyat	37
Alışveriş kalabalığından kaçınmak	29
Kolaylık	18
Daha iyi seçim	13
Doğrudan alışveriş	3



Nitel Verilerin Görselleştirilmesi: Pareto Grafiği

- Kategorik verileri tasvir etmek için kullanılır (nominal ölçek)
- Kategorilerin frekansları büyükten küçüğe doğru sıralanarak dikey çubuklarla gösterilir.
- Artan birikimli frekanslara ait frekans eğrisi (çokgen) aynı grafikte gösterilir.
- Önemli sebepleri, nispeten daha önemsiz sebeplerden ayırmak için kullanılır.

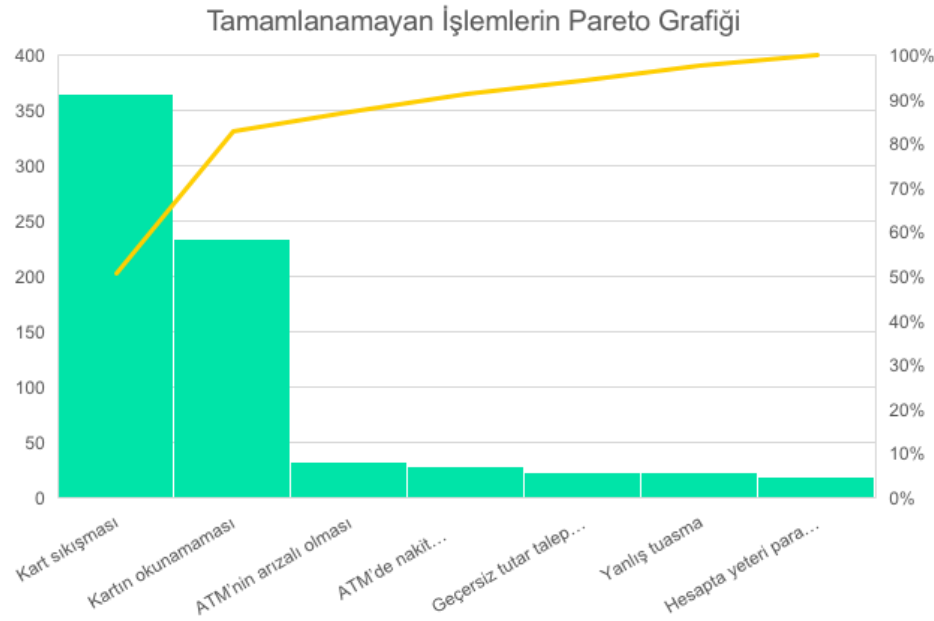
Nitel Verilerin Görselleştirilmesi: Pareto Grafiği

Tamamlanamayan ATM İşlemleri Sebeplerinin Sıralı Özet Tablosu

Nedenler	Frekans	Yüzde	Oransal Birikimli
Kart sıkışması	365	%50.41	%50.41
Kartın okunamaması	234	%32.32	%82.73
ATM'nin arızalı olması	32	%4.42	%87.15
ATM'de nakit bulunmaması	28	%3.87	%91.02
Geçersiz tutar talep etme	23	%3.18	%94.20
Yanlış tuasma	23	%3.18	%97.38
Hesapta yeteri para olmaması	19	%2.62	%100.00
Toplam	724	%100.00	

Kaynak: Data extracted from A. Bhalla, "Don't Misuse the Pareto Principle," *Six Sigma Forum Magazine*, May 2009, pp. 15–18.

Nitel Verilerin Görselleştirilmesi: Pareto Grafiği



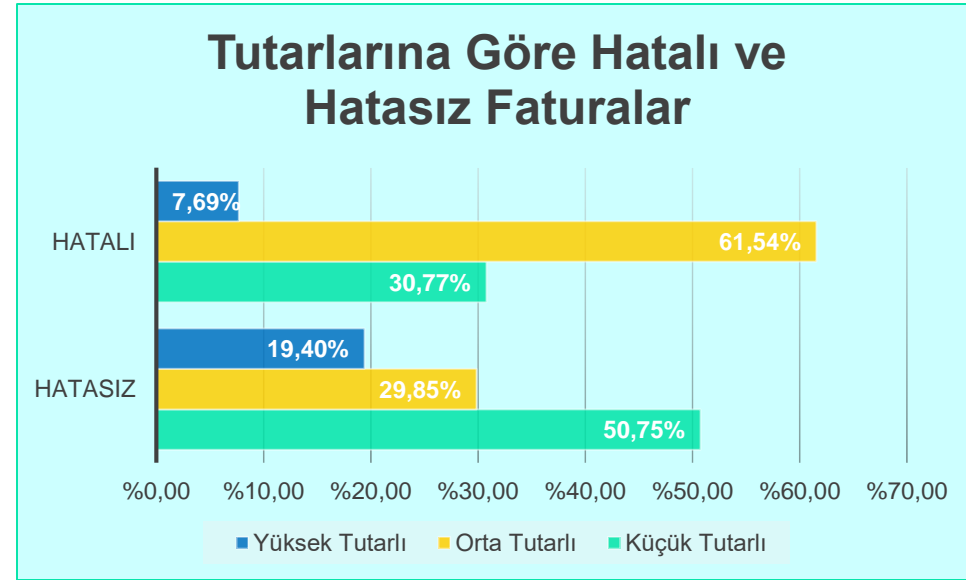
Yüzde

Nispi Olarak
Önemli Sebepler

Nitel Verilerin Görselleştirilmesi: Yan-Yana Sütun Grafikler

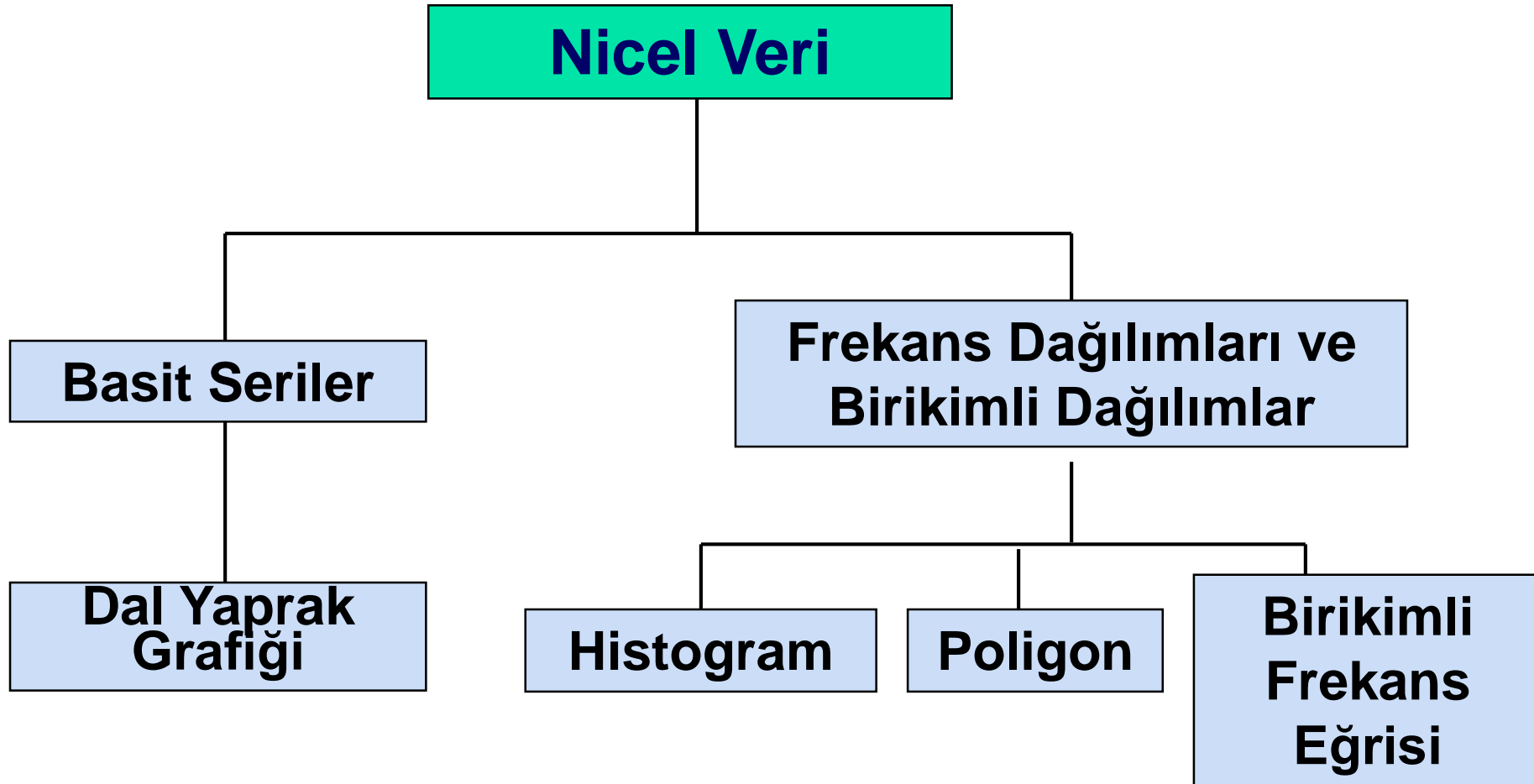
- Yan yana sütun grafik, bir çapraz tablodaki verileri temsil eder. (Olasılık veya yüzde değerler yerine frekanslar da kullanılabilir).

	Hatasız	Hatalı	Toplam
Küçük Tutarlı	%50,75	%30,77	47,50%
Orta Tutarlı	%29,85	%61,54	%35,00
Yüksek Tutarlı	%19,40	%7,69	17,50%
Toplam	%100,0	%100,0	%100,0



Orta tutarlı faturaların hatalı olma olasılığı daha yüksektir. (%61,54%; 30,77% ve 7,69%)

Nicel Verilerin Grafik Olarak Görselleştirilmesi



Dal-Yaprak Gösterimi

- Verilerin nasıl dağıldığını ve veri yoğunluğunun nerede olduğunu görmenin basit bir yoludur.

Yöntem: Sıralanan veriler ayrıştırılır. Öndeki basamaklara “dal”, sondaki basamaklara “yaprak” denir.

Nicele Verilerin Görselleştirilmesi: Dal-Yaprak Grafiği

- Bir dal ve yaprak grafiği örneği aşağıda almaktadır.

Öğrenci Yaşları	Birinci Öğretim Öğrencileri					
	16	17	17	18	18	18
19	19	20	20	21	22	
22	25	27	32	38	42	
İkinci Öğretim Öğrencileri						
18	18	19	19	20	21	
23	28	32	33	41	45	

Öğrenci Yaşları

Birinci Öğretim		İkinci Öğretim	
Dal	Yaprak	Dal	Yaprak
1	67788899	1	8899
2	0012257	2	0138
3	28	3	23
4	2	4	15

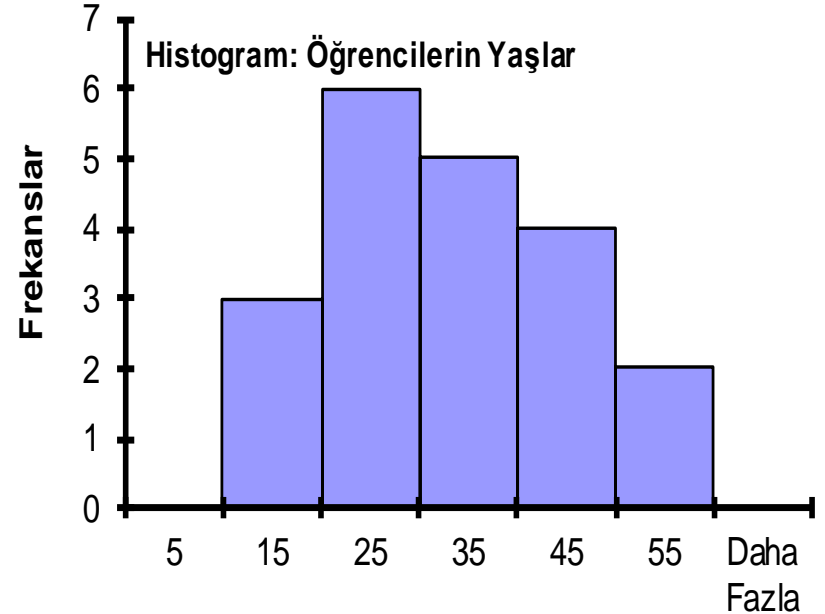
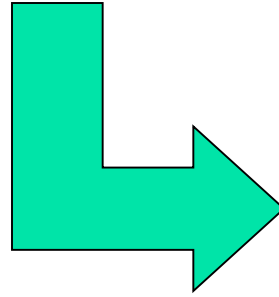
Nicel Verilerin Görselleştirilmesi: Histogram

- Verilerin (sınıflı serilerin) frekans dağılımındaki dikey çubuk grafiğine histogram denir.
- Bir histogramda bitişik çubuklar arasında boşluk yoktur.
- Sınıf sınırları (sınıf alt-üst değerleri veya sınıf orta noktaları) yatay ekseninde gösterilir.
- Dikey ekseninde; frekans, oransal frekanslar veya yüzde değerleri yer alır.
- Çubukların yüksekliği frekansın, oransal frekansın veya yüzdeninbüyüklüğünü temsil eder.

Nicel Verilerin Görselleştirilmesi: Histogram

Sınıflar	Frekanslar	Oransal Frekanslar	Yüzde
10 - 20	3	0,15	15
20 - 30	6	0,30	30
30 - 40	5	0,25	25
40 - 50	4	0,20	20
50 - 60	2	0,10	10
Total	20	1,00	100

(Dikey ekseninde yüzde dağılımlara da yer verilebilir)

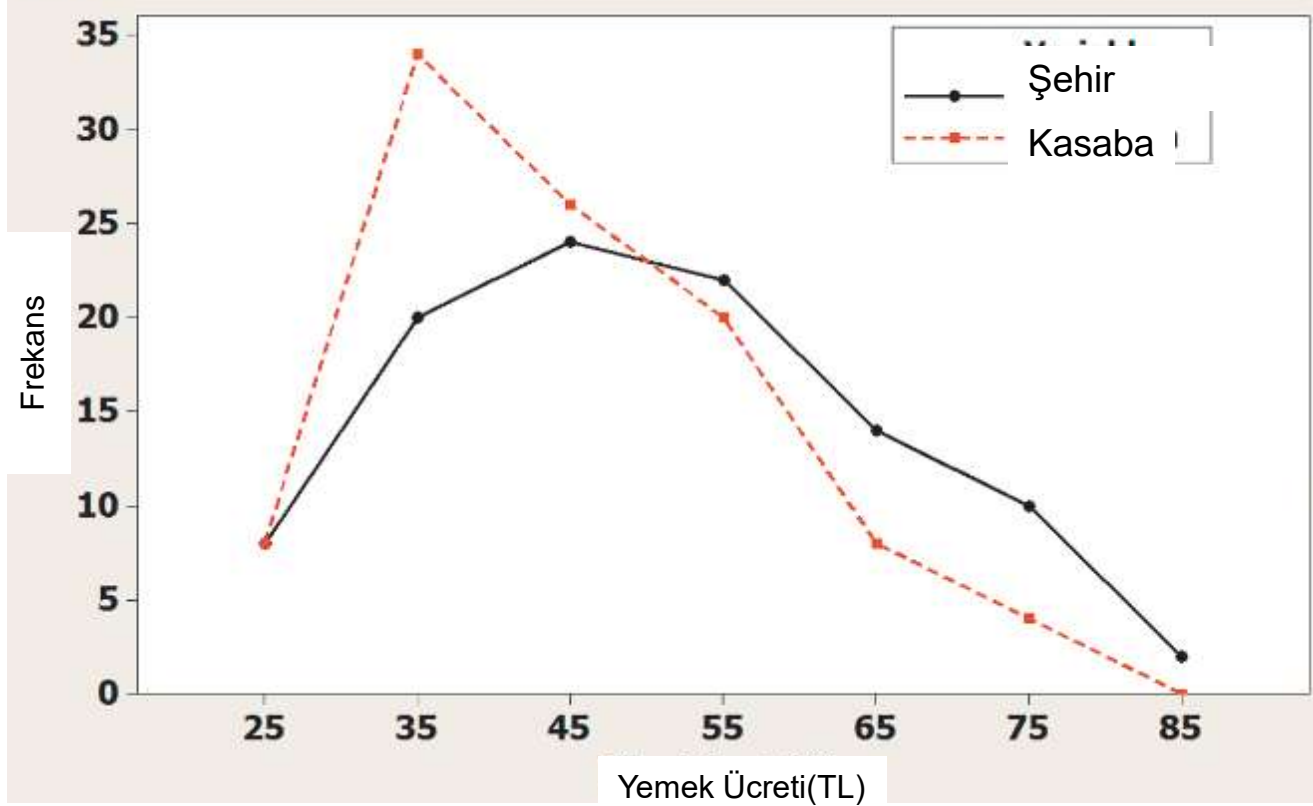


Nicel Verilerin Görselleştirilmesi: Poligon

- Her sınıfın orta noktasının bu sınıftaki verileri temsil etmesi ve daha sonra orta noktaların sırasını ilgili sınıf yüzdelerine bağlamak suretiyle bir yüzde çokgen oluşur.
- Kümülatif yüzde çokgen veya ogive, X eksenini boyunca ilgilenilen değişkeni ve Y eksenini boyunca birikmiş yüzdeleri görüntüler.
- Karşılaştırılacak iki veya daha fazla grup olduğunda kullanışlıdır.

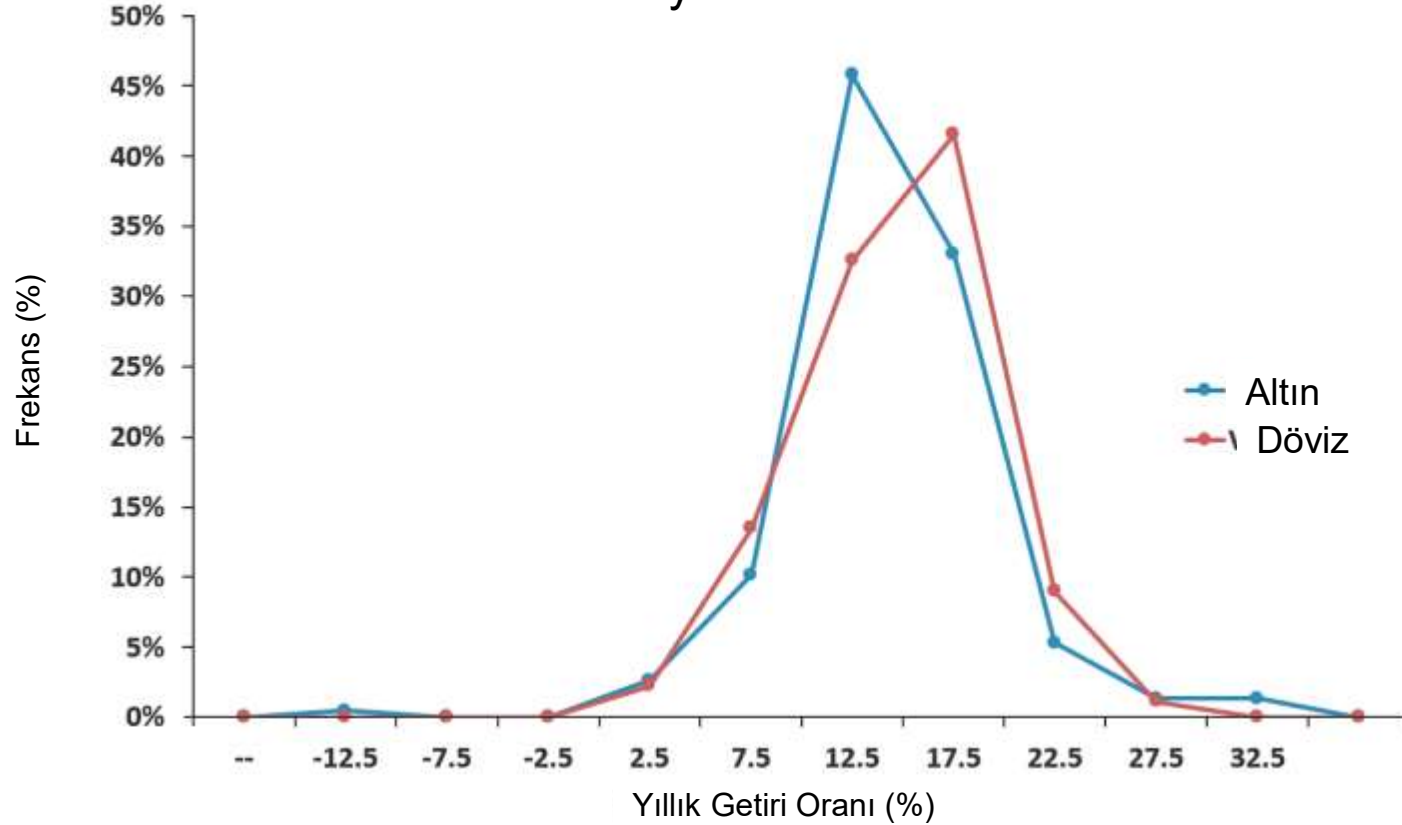
Nicel Verilerin Görselleştirilmesi: Frekans Poligonu

Şehir ve kasabalardaki yemek fiyatlarının frekans poligonu

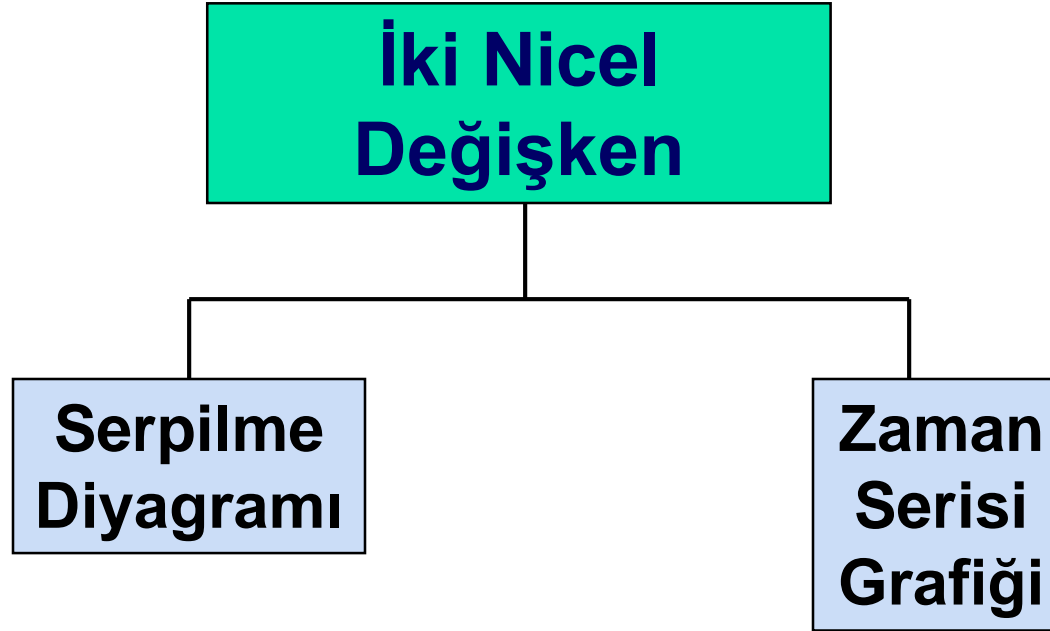


Nicel Verilerin Görselleştirilmesi: Oransal Poligon

Altın ve döviz için bir yıllık getiri oranı yüzdesi



İki Nicel Değişkenin Grafik Olarak Görselleştirilmesi

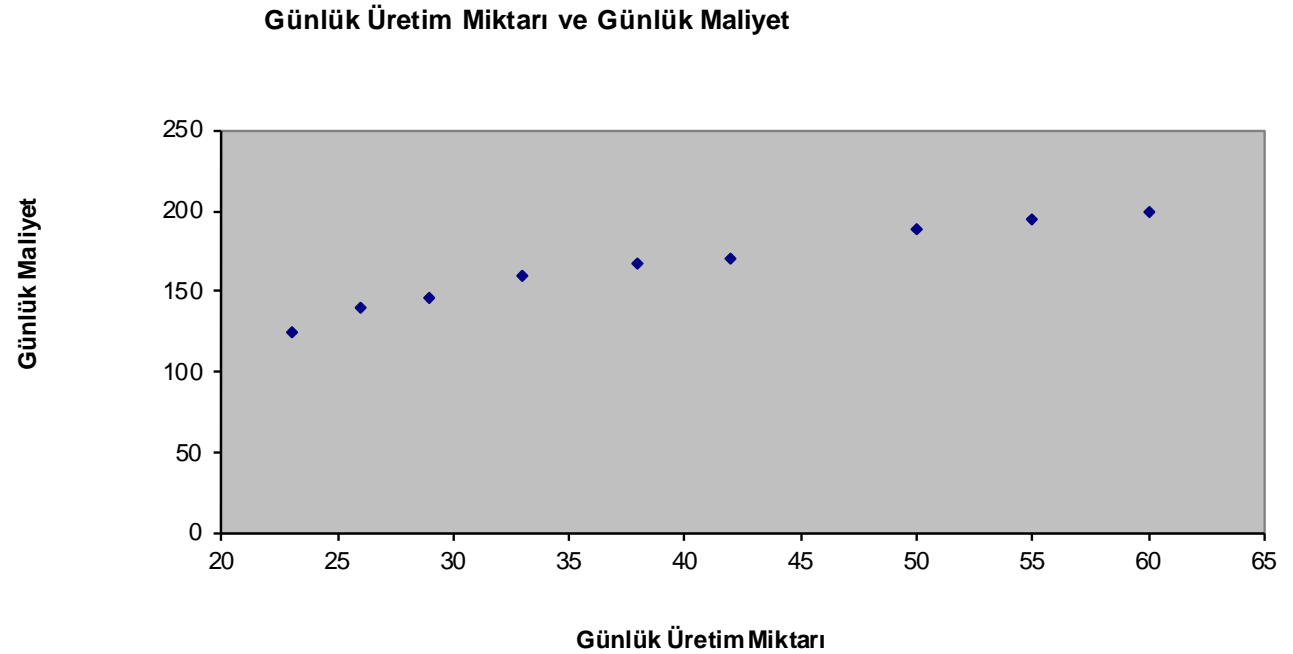


İki Nicel Verinin Görselleştirilmesi: Serpilme Diyamı

- İki nicel deęişkenden alınan eşleřtirilmiř gözlemlerden oluřan sayısal verilerin görselleřtirilmesinde serpilme diyagramı grafikleri kullanılır.
- Bir deęişken dikey eksene ve dięer deęişken yatay eksene yerleřtirilir.
- Serpilme diyagramı grafikleri, iki nicel deęişken arasındaki olası iliřkileri incelemek için de kullanılır.

Örnek Serpilme Diyagramı

Günlük Üretim Miktarı	Günlük Maliyet
23	125
26	140
29	146
33	160
38	167
42	170
50	188
55	195
60	200

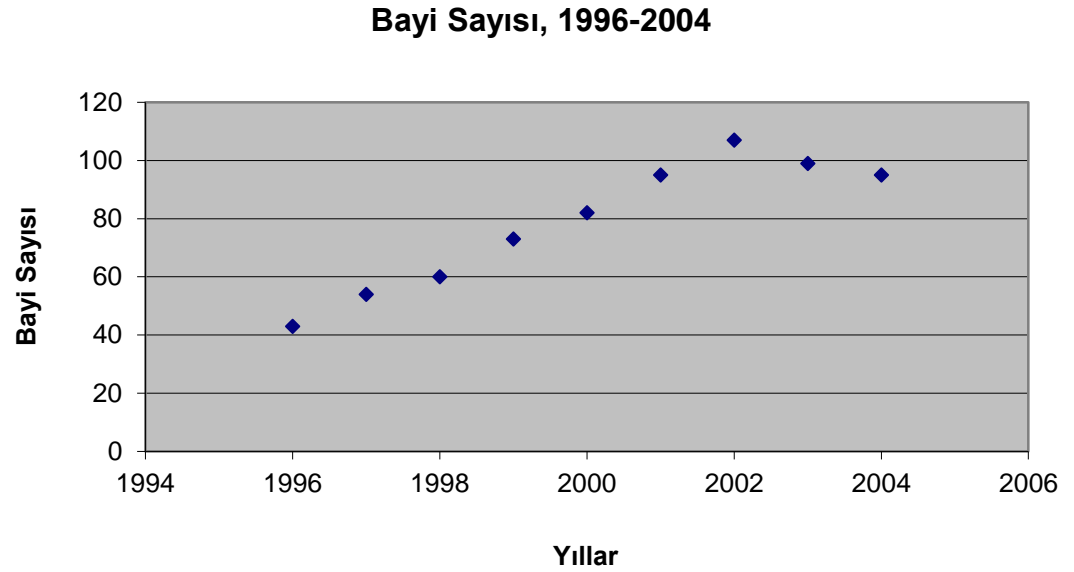


İki Nicel Değişkenin Görselleştirilmesi: Zaman Serisi Grafiği

- Zaman Serileri grafiği, zaman durak noktalarına göre nicel bir değişkenin aldığı değerlerin görselleştirilmesidir..
- Zaman Serisi Grafiği:
- Nicel değişken dikey eksene ve zaman durak noktaları yatay eksene yerleştirilir.

Örnek Zaman Serisi Grafiği

Yıllar	Bayi Sayısı
1996	43
1997	54
1998	60
1999	73
2000	82
2001	95
2002	107
2003	99
2004	95



Birden Çok Kategorik Değişkenlerin Düzenlenmesi: Çok Boyutlu Çapraz Tablo

- Üç veya daha fazla kategorik değişkenin olduğu durumlarda da çapraz tablo oluşturulabilir.
- Excel'de, 2'den daha fazla kategorik değişkenler için Pivot Tablo oluşturulabilir.

Excel Pivot Tablo ile Değişkenlerin Düzenlenmesi ve Görselleştirilmesi

Bir pivot tablo:

- Değişkenleri çok boyutlu bir özet tablo olarak özetler.
- Değişkenlerin özetlenmesi ve değişkenlerin etkileşimli olarak biçimlendirilmesine izin verir.
- Belirtilen ölçütleri karşılayan verilerin alt kümelerini özetlemek için verilerinizi etkileşimli olarak “sınıflamanızı” sağlar.
- Çok boyutlu verilerdeki olası kalıpları ve ilişkileri keşfetmek için daha basit tabloların ve grafiklerin kullanılabilmesine olanak sağlar.

Verileri İyi Bir Şekilde Görselleştirmek İçin

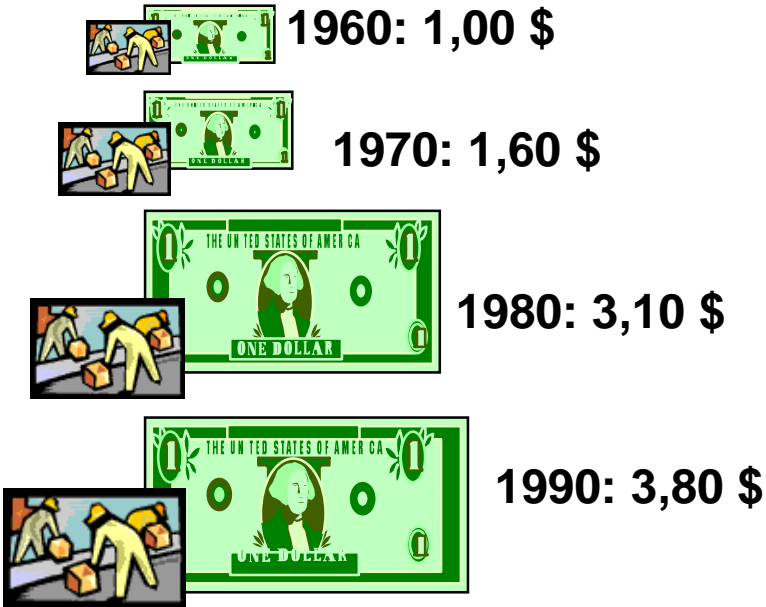
- Hatalı grafik kullanmaktan kaçınılmalı
- Mümkün olan en basit görselleştirmeyi kullanılmalı
- Grafiklere başlık eklenmeli
- Tüm eksenleri etiketlenmeli
- Grafik eksen içeriyorsa her eksen için bir ölçek eklenmeli
- Dikey eksen için ölçeğe sıfırdan başlanmalı
- Sabit bir ölçek kullanılmalı

Hatalı Grafik: Uygun Gösterim Değil



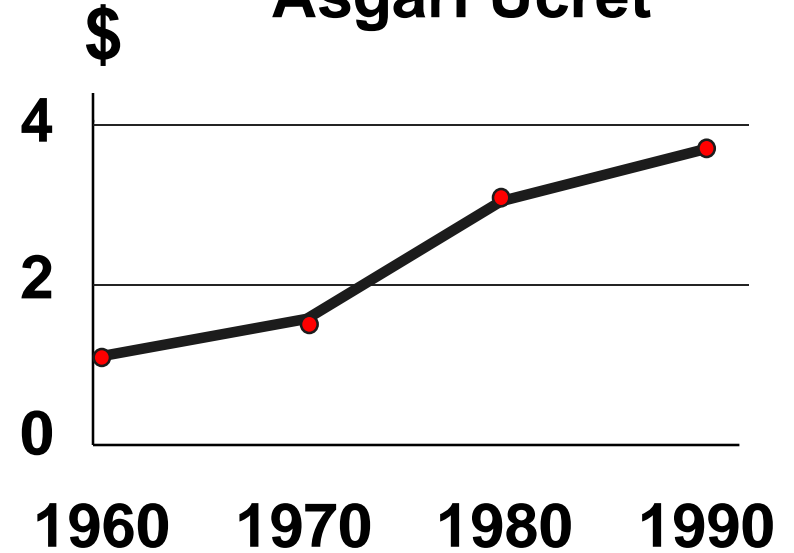
Uygun Değil

Asgari Ücret



Uygun

Asgari Ücret



Hatalı Grafik: Oransal Gösterim Yok



Uygun Değil

100 alan
öğrenciler

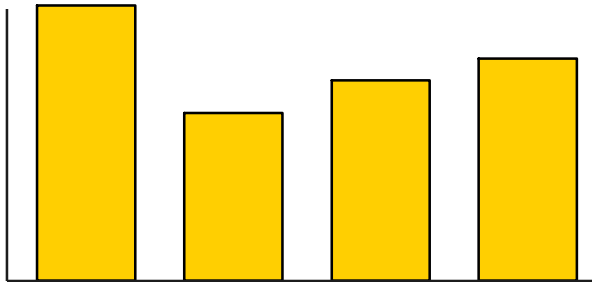
Freq.

300

200

100

0



A

B

C

D



Uygun

100 alan
öğrenciler

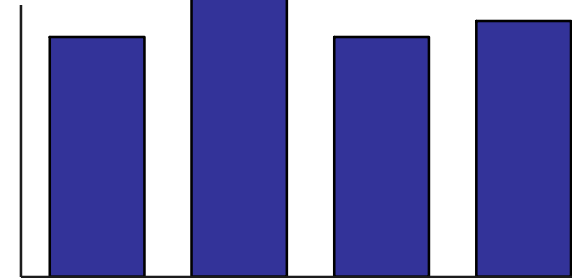
%

%30

%20

%10

%0



A

B

C

D

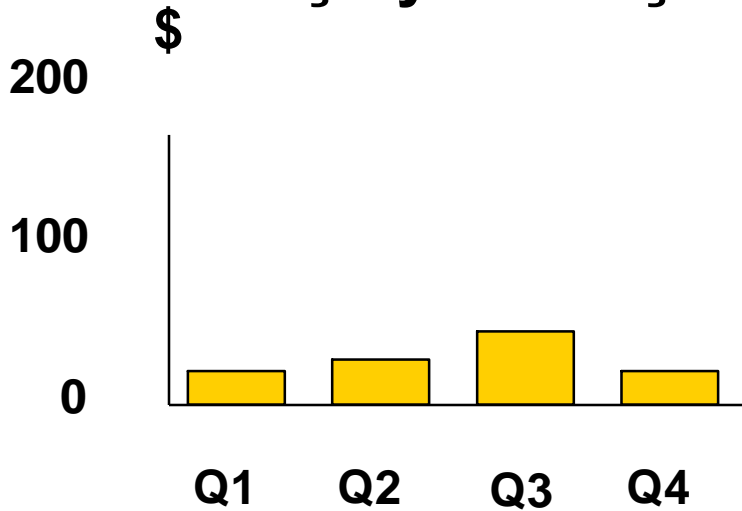
A = A Şubesi, B = B Şubesi, C = C Şubesi, D = D Şubesi

Hatalı Grafik: Sıkıştırılmış Dikey Eksen



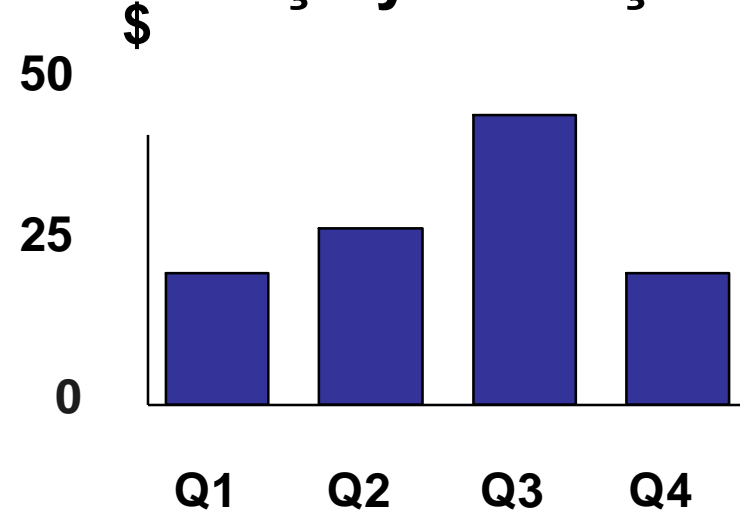
Uygun Değil

Üç Aylık Satışlar



Uygun

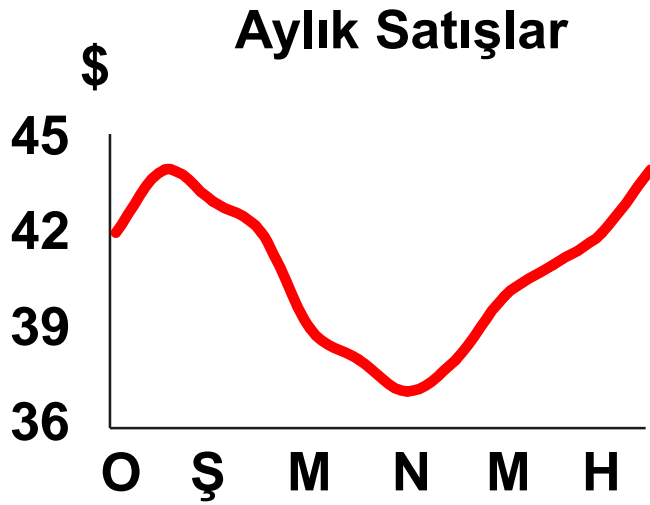
Üç Aylık Satışlar



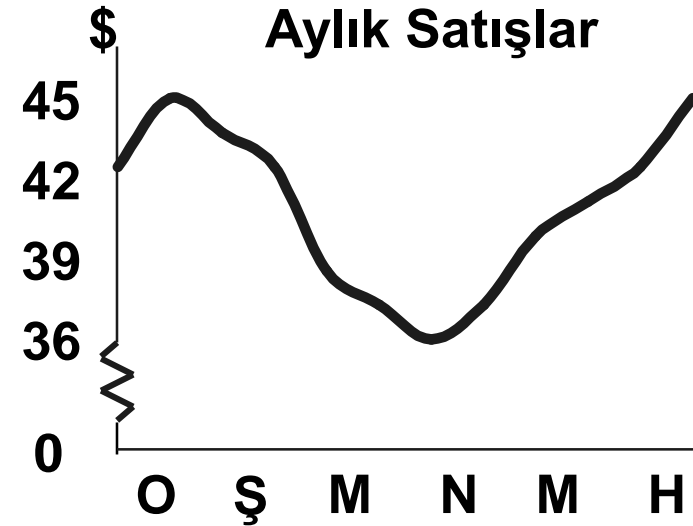
Hatalı Grafik: Dikey Eksende Sıfır Noktası Yok



Uygun Değil



Uygun



İlk altı aylık satış gelirleri grafiği

Tanımlayıcı Karakteristik Değerler

Nicel Veri Analizi

Teori ve Uygulama



Prof. Dr. Ünal H. ÖZDEN

Merkezi Eğilim Ölçüleri: Aritmetik Ortalama

- Aritmetik ortalama (genellikle “ortalama” olarak adlandırılır), en yaygın kullanılan merkezi eğilim ölçüsüdür.
 - Örneklem için :

Aritmetik ortalama

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

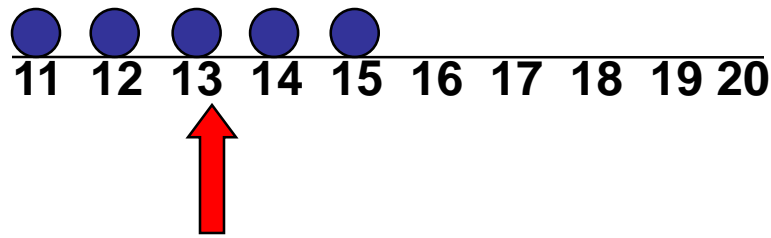
i' inci X değeri

Örneklem büyüklüğü,
gözlem sayısı

Gözlem değerleri

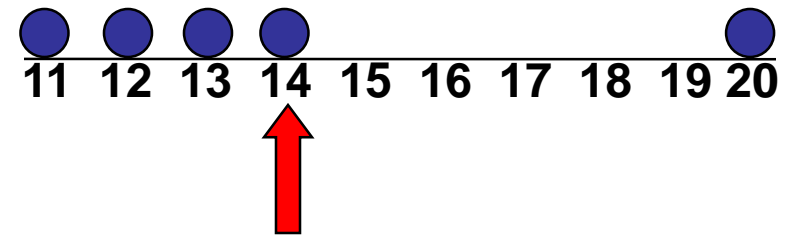
Merkezi Eğilim Ölçüleri: Aritmetik Ortalama

- Merkezi eğilimin en yaygın ölçüsü
- Ortalama = değerler toplamı gözlem sayısına bölünür
- Aşırı (uç) değerlerden etkilenir (aykırı değerler)



A.O. = 13

$$\frac{11+12+13+14+15}{5} = \frac{65}{5} = 13$$

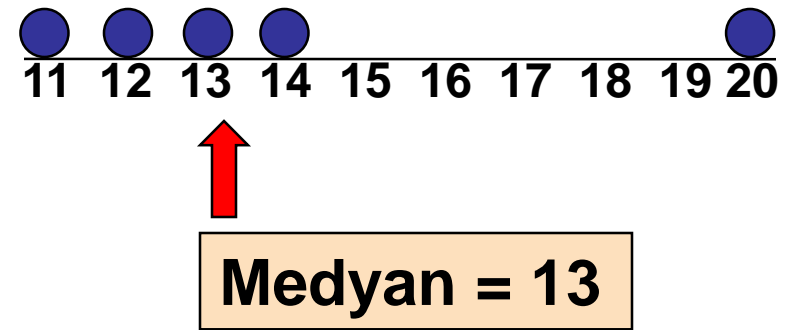
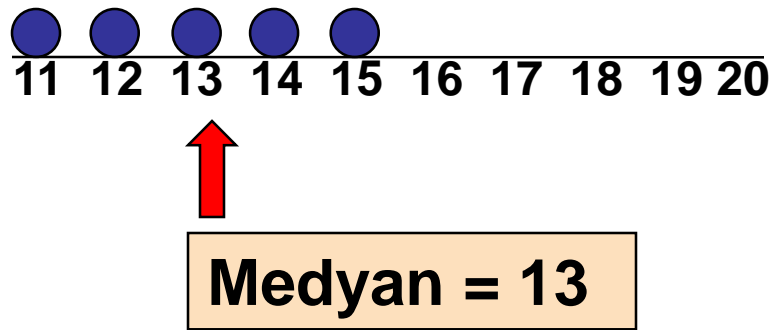


A.O. = 14

$$\frac{11+12+13+14+20}{5} = \frac{70}{5} = 14$$

Merkezi Eğilim Ölçüleri: Medyan

- Medyan sıralanmış verilerde seriyi iki eşit parçaya bölen değerdir. (Seri değerlerinin %50'si medyandan büyük, %50'si medyandan küçüktür.)



- A.O.'ya göre uç değerlerden daha az etkilenir.

Merkezi Eğilim Ölçüleri: Medyan

- Medyan değerinin yeri veriler küçükten büyüğe doğru sıralandığında:

$$\frac{n+1}{2}$$

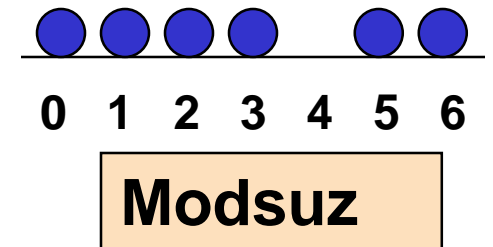
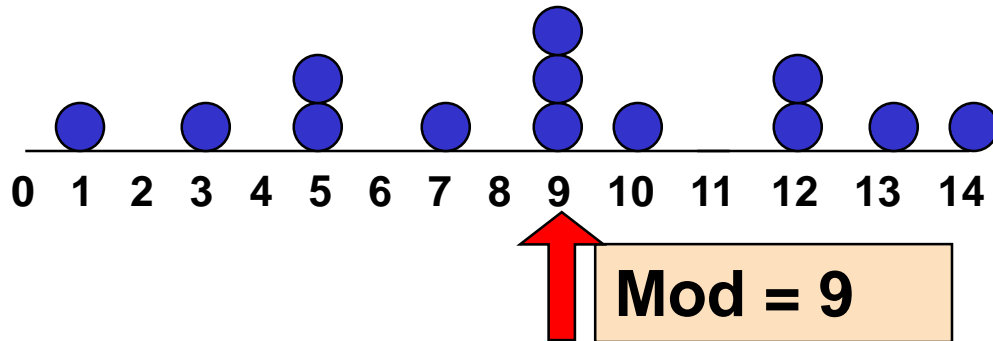
formülü ile belirlenir.

- Gözlem sayısı tek ise medyan, ortada yer alan değerdir.
- Gözlem sayısı çift ise medyan ortada yer alan iki değer ortalamasıdır.

$\frac{n+1}{2}$ ifadesinin sonucu medyan değildir, bu ifadenin sonucu sıralanmış verilerde medyanın kaçınıcı sıradaki değer olduğunu gösterir.

Merkezi Eğilim Ölçüleri: Mod

- En sık görülen (en çok tekrarlanan) değer
- Aşırı değerlerden etkilenmez
- Sayısal veya kategorik (nominal) veriler için kullanılır
- Bir seri «mod»suz veya birden çok «mod»lu olabilir.



Merkezi Eğilim Ölçüleri: Örnek

Ev Fiyatları (\$):

2,000,000

500,000

300,000

100,000

100,000

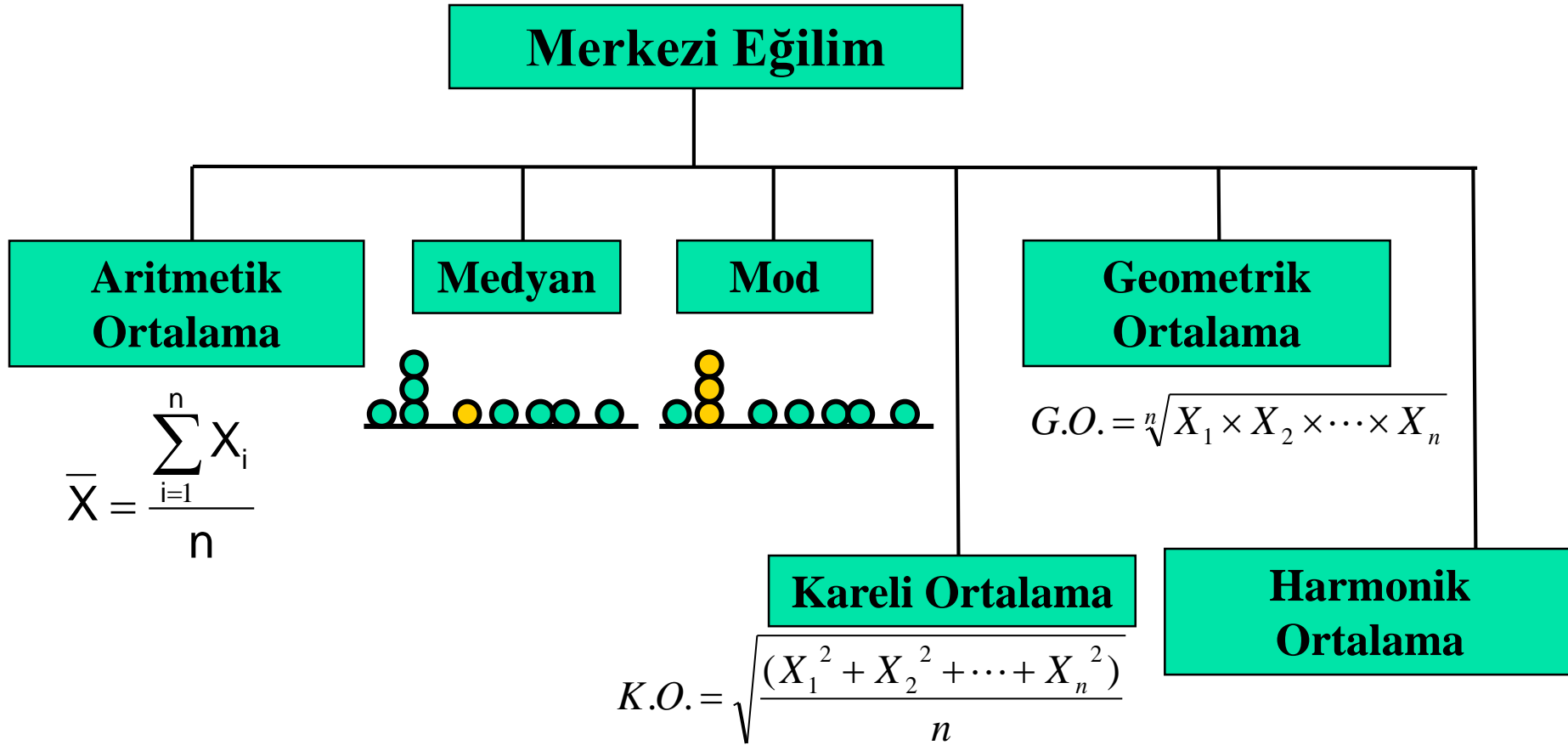
Toplam 3,000,000\$

- **A.O:** $(3,000,000/5)$
= **600,000\$**
- **Medyan:** Seri orta değeri
= **300,000\$**
- **Mode:** En çok tekrarlanan
= **100,000\$**

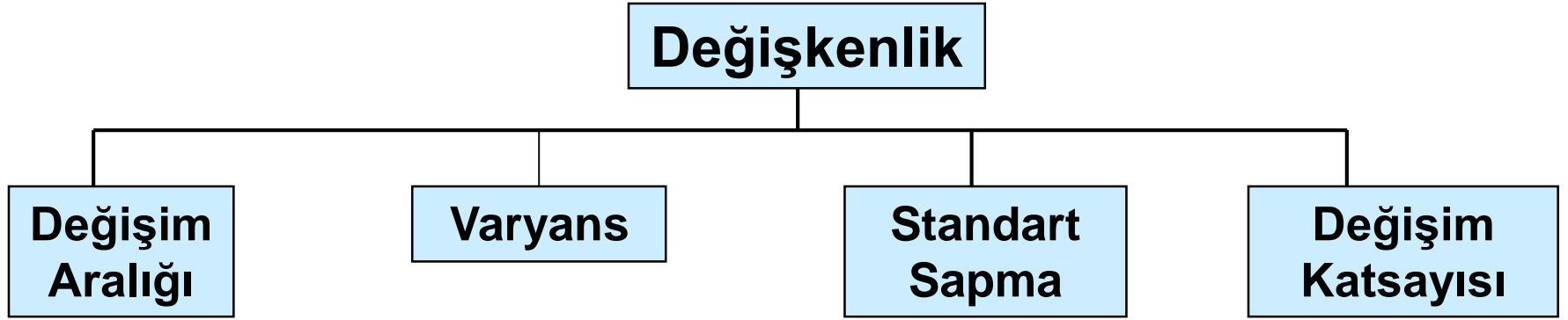
Merkezi Eğilim Ölçüleri: Hangisi Seçilmeli?

- A. O. Uç değerlerden çok etkilenir. En çok kullanılan merkezi eğilim ölçüsüdür.
- Medyan uç değerlere duyarlı değildir.
- Genelde raporlarda üç merkezi eğilim ölçüsü de verilmesi daha uygundur.

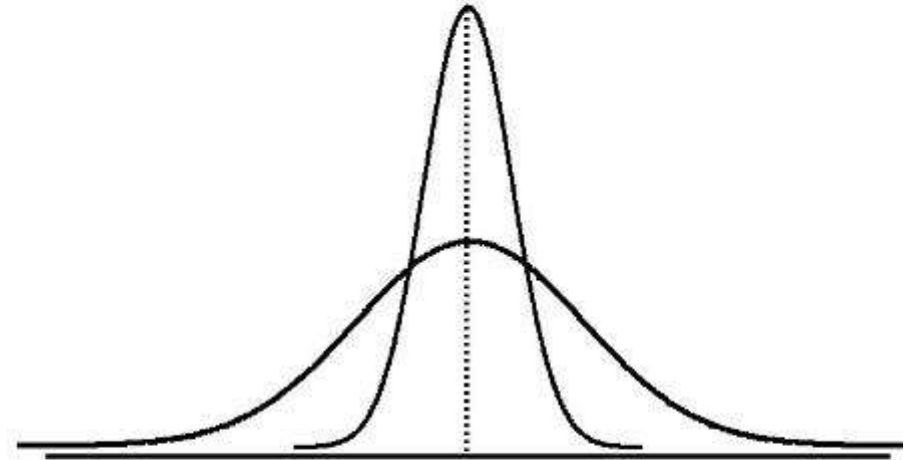
Merkezi Eğilim Ölçüleri: Özet



Değişkenlik Ölçüleri



- Değişkenlik ölçüleri, veri değerlerinin yayılması veya değişkenliği veya dağılımı hakkında bilgi verir.



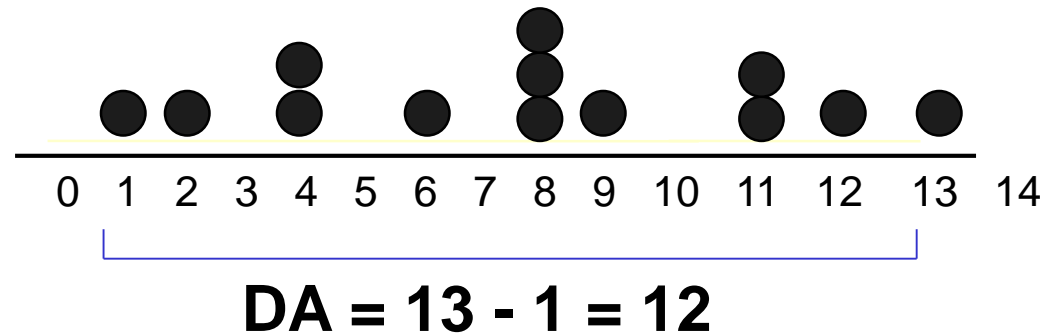
Aynı merkezi eğilim,
Farklı değişkenlik

Değişkenlik Ölçüleri: Değişim Aralığı

- En basit değişim ölçüsüdür.
- En büyük ve en küçük değerler arasındaki farka eşittir:

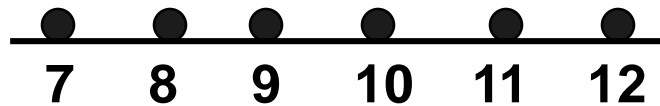
$$\text{Değişim Aralığı (DA)} = X_{\max} - X_{\min}$$

Örnek:

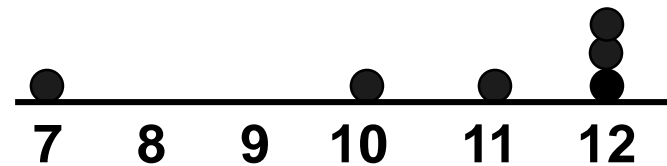


Değişkenlik Ölçüleri: Değişim Aralığı neden yanıtıcı olabilir?

- Verilerin nasıl dağıldığını dikkate almaz



$$\text{D.A.} = 12 - 7 = 5$$



$$\text{D.A.} = 12 - 7 = 5$$

- Aykırı (uç) değerlere duyarlı

1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,4,5

$$\text{D.A.} = 5 - 1 = 4$$

1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,4,120

$$\text{D.A.} = 120 - 1 = 119$$

Değişkenlik Ölçüleri: Örneklem Varyansı

- Değişken değerlerinin ortalamadan ortalama (yaklaşık) kare sapması
- Örnek varyansı:

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

\bar{X} = aritmetik ortalama

n = örneklem büyüklüğü

X_i = X değişkeninin i 'inci değeri

Değişkenlik Ölçüleri: Örneklem Standart Sapması

- En yaygın kullanılan değişkenlik ölçüsü
- Aritmetik ortalamaya göre değişkenliği gösterir.
- Varyansın kareköküdür.
- Orijinal verilerle aynı ölçü birimine sahiptir.

- Örneklemin standart sapması:

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

Değişkenlik Ölçüleri: Standart Sapması

Standart sapmanın hesaplanmasının adımları

1. Gözlem değerlerinin aritmetik ortalmadan farklarının hesaplanması.
2. Her bir farkın karesinin hesaplanması.
3. Farkların karesinin toplamının alınması.
4. Bu toplamın $n-1$ 'e bölünmesi (örneklem için)(ana kitle için N 'e bölünmesi).
5. Örneklem standart sapmasını elde etmek için örneklem varyansın karekökünün alınması.

Değişkenlik Ölçüsü: Örneklem Standart Sapma Örneği

Örneklem

Veri (X_i):

10 12 14 15 17 18 18 24

$n = 8$

A.O. = $\bar{X} = 16$

$$S = \sqrt{\frac{(10 - \bar{X})^2 + (12 - \bar{X})^2 + (14 - \bar{X})^2 + \dots + (24 - \bar{X})^2}{n - 1}}$$

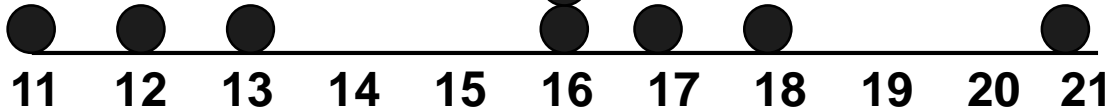
$$= \sqrt{\frac{(10 - 16)^2 + (12 - 16)^2 + (14 - 16)^2 + \dots + (24 - 16)^2}{8 - 1}}$$

$$= \sqrt{\frac{130}{7}} = 4.3095 \rightarrow$$

Gözlem değerlerinin aritmetik ortalama etrafında "ortalama" dağılımı

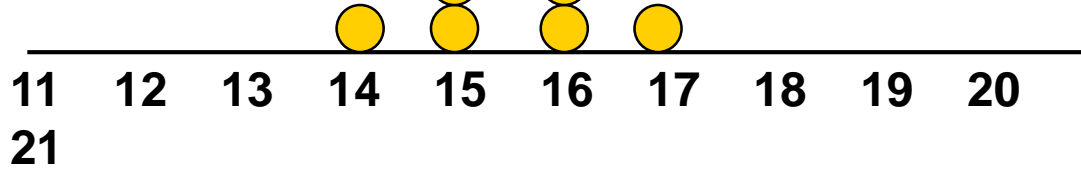
Değişkenlik Ölçüsü: Standart sapmaların karşılaştırılması

Veri A



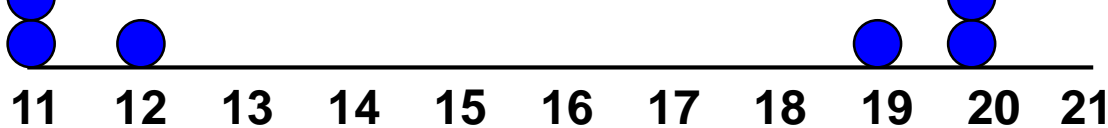
$$A.O. = 15.5$$
$$S = 3.338$$

Veri B



$$A.O. = 15.5$$
$$S = 0.926$$

Veri C

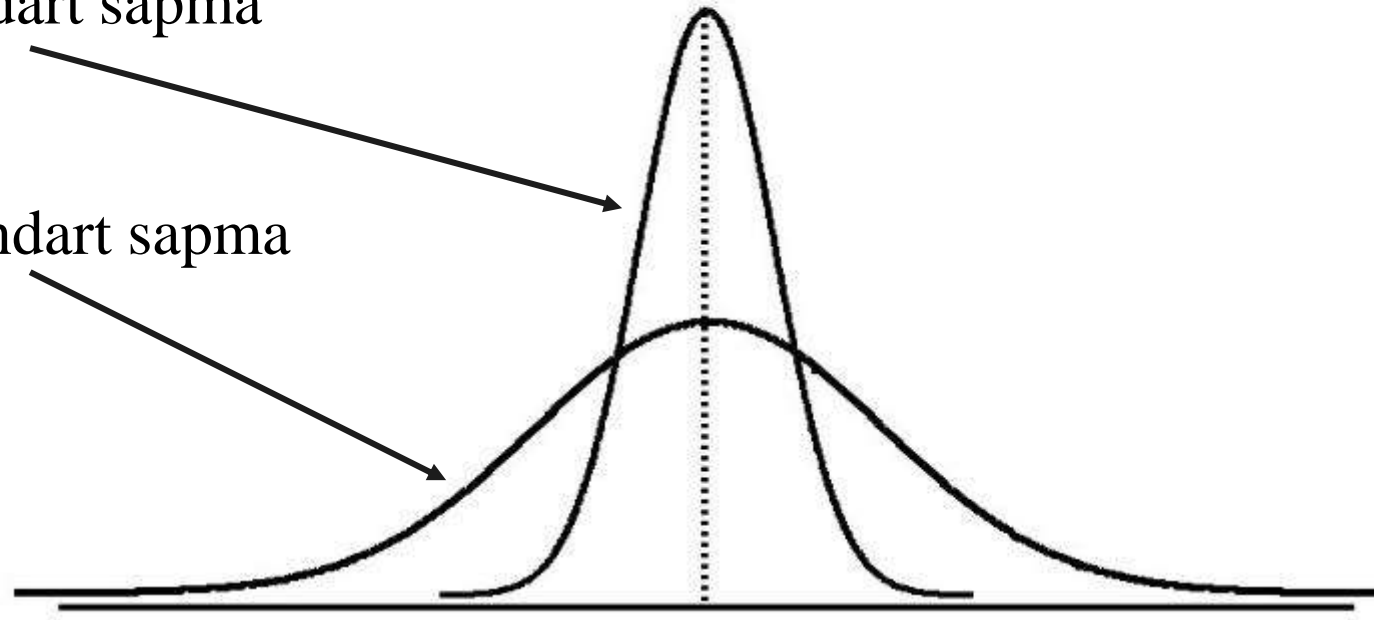


$$A.O. = 15.5$$
$$S = 4.567$$

Değişkenlik Ölçüsü: Standart sapmaların karşılaştırılması

Düşük standart sapma

Yüksek standart sapma



Değişkenlik Ölçüleri: Değişim Katsayısı

- Değişkenliği nisbi olarak ölçer.
- Değişkenliği %'desel olarak gösterir.
- Aritmetik ortalamaya göre nisbi değişkenliği ifade eder.
- Ölçü biriminden etkilenmez.
- Farklı ölçü birimleri ile ölçülen iki veya daha fazla veri kümesinin değişkenliğini karşılaştırmak için kullanılabilir.

$$CV = \left(\frac{S}{\bar{X}} \right) \cdot 100\%$$

Değişkenlik Ölçüleri: Değişim Katsayısı

■ A Hisse Senedi:

- Geçen yılın fiyat ortalaması = 50 \$
- Standart sapma = 5\$

$$CV_A = \left(\frac{S}{\bar{X}} \right) \cdot 100\% = \frac{\$5}{\$50} \cdot 100\% = 10\%$$

■ B Hisse Senedi:

- Geçen yılın fiyat ortalaması = 100\$
- Standart sapma = 5\$

$$CV_B = \left(\frac{S}{\bar{X}} \right) \cdot 100\% = \frac{\$5}{\$100} \cdot 100\% = 5\%$$

Her iki hisse senedi de aynı standart sapmaya sahip, ancak B hisse senedi fiyatına göre daha az değişken

Değişkenlik Ölçüleri: Değişim Katsayısı

■ Hisse Senedi A:

- Geçen yılın fiyat ortalaması = 50 \$
- Standart sapma = 5 \$

$$CV_A = \left(\frac{S}{\bar{X}} \right) \cdot 100\% = \frac{\$5}{\$50} \cdot 100\% = 10\%$$

■ Hisse Senedi C:

- Geçen yılın fiyat ortalaması = 8 \$
- Standart sapma = 2 \$

$$CV_C = \left(\frac{S}{\bar{X}} \right) \cdot 100\% = \frac{\$2}{\$8} \cdot 100\% = 25\%$$

C hisse senedi daha küçük bir standart sapmaya sahiptir, ancak daha yüksek bir değişim katsayısına sahiptir.

Değişkenlik Ölçülerinin Özellikleri

- Veriler ne kadar çok yayılırsa, aralık, varyans ve standart sapma o kadar büyük olur.
- Veriler ne kadar yoğunlaşırsa, aralık, varyans ve standart sapma o kadar küçük olur.
- Değişken değerlerinin hepsi aynıysa (değişiklik yoktur), tüm değişkenlik ölçüsü değerleri sıfır olur.
- Değişkenlik ölçüsü değerlerinin hiçbiri negatif çıkmaz.

Uç Değerleri Bulma : Z-Değerleri

- Bir veri değerinin Z-değerini hesaplamak için, değişken değerinden aritmetik ortalama çıkarılır ve standart sapmaya bölünür.
- Z-değeri, bir verinin ortalamadan standart sapmasının değeridir.
- Bir verinin Z-değeri -3.0'dan küçükse veya +3.0'dan büyükse aşırı bir değer olarak kabul edilir.
- Z-değerinin mutlak değeri ne kadar büyükse, veri değerinin ortalamadan uzaklığı da o kadar büyük olur.

$$Z = \frac{X - \bar{X}}{S}$$

Uç Değerleri Bulma : Z-Değerleri

- SAT sınavında ortalama matematik skorunun, 100'lük bir standart sapma ile 490 olduğunu varsayalım.
- 620'lik bir test puanı için Z-skorunu

$$Z = \frac{X - \bar{X}}{S} = \frac{620 - 490}{100} = \frac{130}{100} = 1.3$$

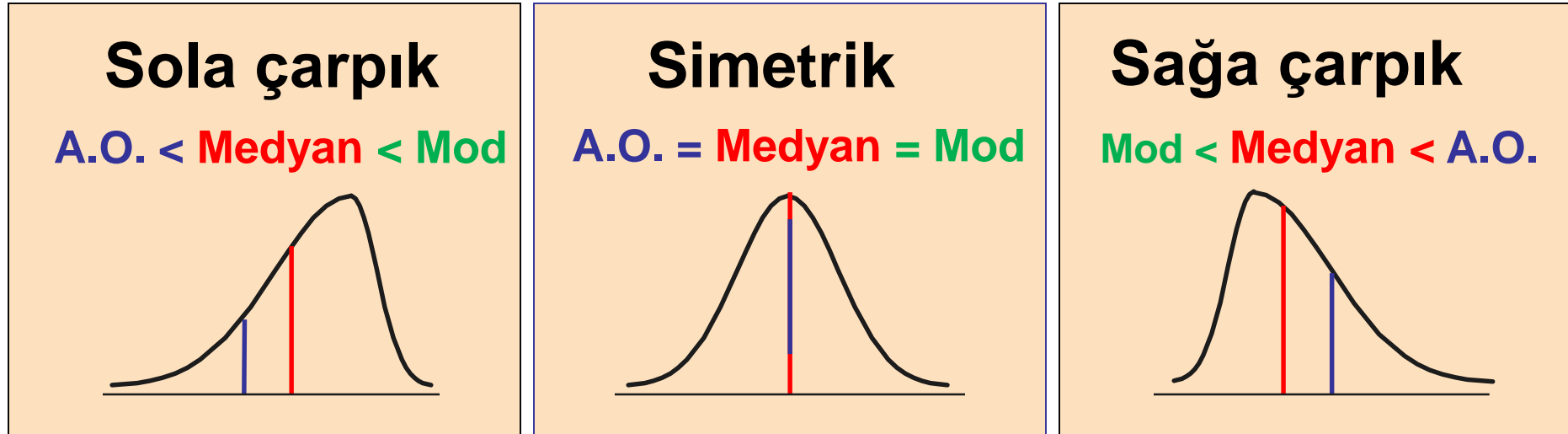
620 puan ortalamasının 1.3 standart sapma üzerindedir ve aykırı değer sayılmayacaktır.

Dağılım Ölçüleri (Şekli)

- Verilerin nasıl dağıldığını tanımlar
- İki önemli dağılım ölçüsü vardır:
 - Çarpıklık
 - Verilerin simetrik dağılıma derecesini ölçer
 - Basıklık
 - Kurtosis, dağılım eğrisinin zirvesini etkiler – yani dağılımın tepe noktasının merkezine yaklaşırken ne kadar keskin yükseldiğini gösterir.

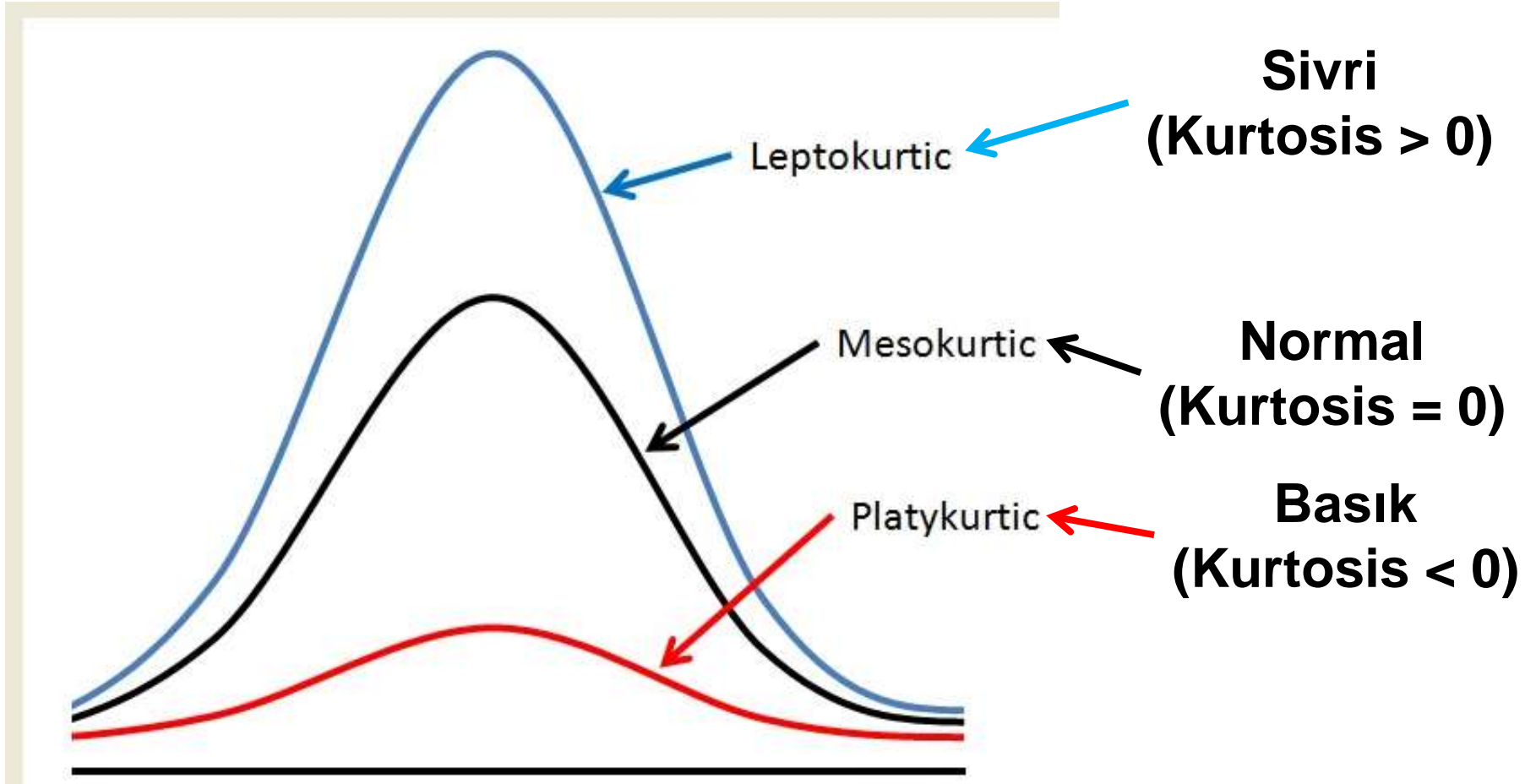
Dağılımın Şekli: Çarpıklık (Skewness)

- Verilerin simetrik olma derecesini ölçer



Dağılımın Şekli: Basıklık (Kurtosis)

- Dağılımın tepe noktasına (merkezinine) yaklaştıkça nasıl keskinleştiğini ölçer.

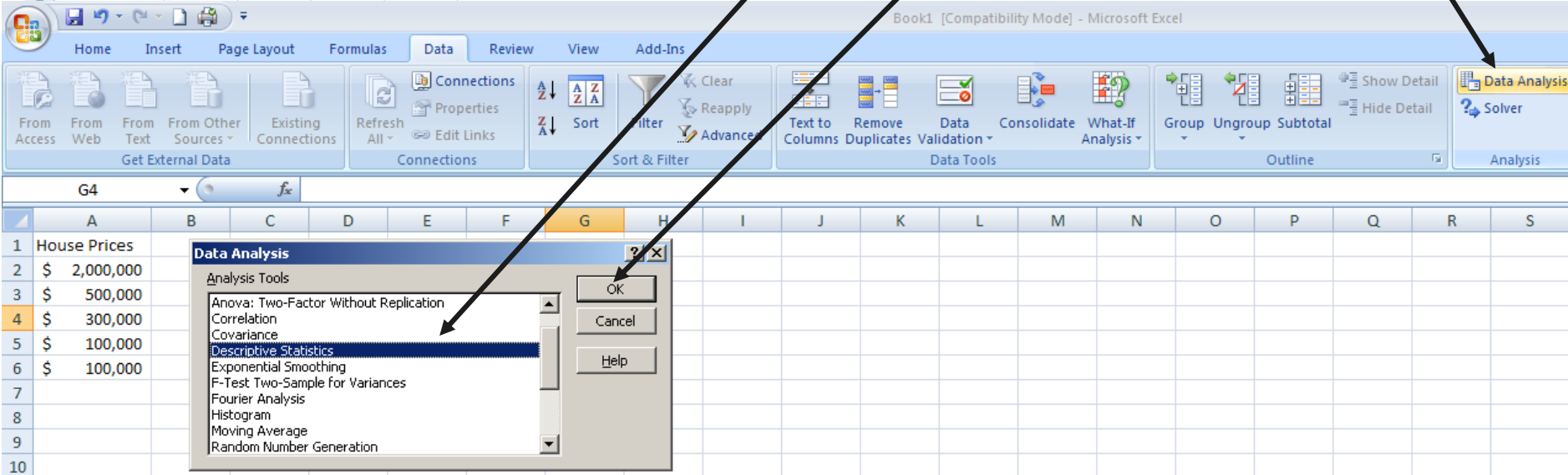
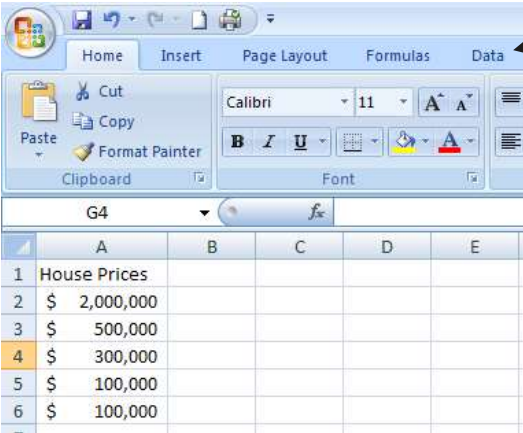


Microsoft Excel'de Tanımsal İstatistik Fonksiyonları

Ev Fiyatları		Tanımsal İstatistikler		
\$ 2.000.000		Mean (Aritmetik Ortalama)	\$ 600.000	=AVERAGE(A2:A6)
\$ 500.000		Standard Error (S. Hata)	\$ 357.770,88	=D6/SQRT(D14)
\$ 300.000		Median (Medyan)	\$ 300.000	=MEDIAN(A2:A6)
\$ 100.000		Mode (Mod)	\$ 100.000,00	=MODE(A2:A6)
\$ 100.000		Standard Deviation (S. Sapma)	\$ 800.000	=STDEV(A2:A6)
		Sample Variance (Örneklem Varyansı)	640.000.000.000	=VAR(A2:A6)
		Kurtosis (Basıklık)	4,1301	=KURT(A2:A6)
		Skewness (Çarpıklık)	2,0068	=SKEW(A2:A6)
		Range (Değişim Aralığı)	\$ 1.900.000	=D12 - D11
		Minimum (Enküçük Değer)	\$ 100.000	=MIN(A2:A6)
		Maximum (Enbüyük Değer)	\$ 2.000.000	=MAX(A2:A6)
		Sum (Serinin Toplam Değeri)	\$ 3.000.000	=SUM(A2:A6)
		Count (Gözlem Sayısı, N veya n)	5	=COUNT(A2:A6)

Microsoft Excel Data Analysis Tool'da Tanımsal İstatistik Değerlerinin Hesaplanması

1. Data (Veri)
2. Data Analysis (Veri Analizi)
3. Descriptive Statistics (Tanımsal İstatistik)
4. OK.



Microsoft Excel Data Analysis Tool'da Tanımsal İstatistik Değerlerinin Hesaplanması

5. Verilerin yer aldığı hücre aralığını girin

6. "Summary Statistics"i işaretleyin

7. OK

The screenshot shows the 'Descriptive Statistics' dialog box in Microsoft Excel. The dialog box is open over a spreadsheet with data in column A. The 'Input Range' is set to '\$A\$2:\$A\$6'. The 'Summary statistics' checkbox is checked. The 'OK' button is highlighted. The spreadsheet data is as follows:

	A	B	C	D	E	F	G	H
1	House Prices							
2	\$ 2,000,000							
3	\$ 500,000							
4	\$ 300,000							
5	\$ 100,000							
6	\$ 100,000							
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								

Microsoft Excel Data Analysis Tool'da Tanımsal İstatistik Değerlerinin Hesaplanması-Çıktılar

Microsoft Excel
descriptive statistics output,
using the house price data:

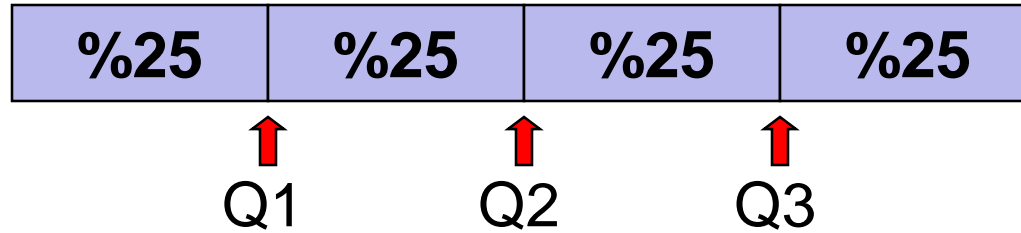
Ev Fiyatları:

\$2,000,000
500,000
300,000
100,000
100,000

<i>Ev Fiyatları</i>	
Mean	600000
Standard Error	357770,8764
Median	300000
Mode	100000
Standard Deviation	800000
Sample Variance	640.000.000.000
Kurtosis	4,1301
Skewness	2,0068
Range	1900000
Minimum	100000
Maximum	2000000
Sum	3000000
Count	5

Kartil (Dörde Bölenler) Ölçüleri

- Kartiller sıralanmış verileri eşit sayıda değerden oluşan 4 segmente böler.



- Birinci kartil Q_1 , en küçük %25 sıradaki değerdir.
- İkinci kartil Q_2 , medyana eşittir ve verilerin %50'si medyan değerinden büyük, %50 medyandan küçük değerlere sahiptir..
- Üçüncü kartil Q_3 , Verilerin yalnızca %25'i üçüncü kartilden büyük değerlere sahiptir.

Kartillerin Hesaplanması

Verileri sıralayınız...

Birinci kartil: $Q_1 = (n+1)/4$ sırdaki değer

İkinci kartil: $Q_2 = (n+1)/2$ sıradaki değer

Üçüncü kartil: $Q_3 = 3(n+1)/4$ sıradaki değer

n: gözlem sayısı

Kartil Hesaplama Kuralları

- Sonuç tam sayı ise, o sıradaki değer kartil değeridir.
- Eğer sonuç ondalıklı bir sayı ise (ör. 5.5, vb.), o 5. ve 6. sıradaki sayıların ortalaması kartildir.
- Sonuç, tam sayı veya kesirli bir sayı değilse, kartili bulmak için sonucu en yakın tam sayıya yuvarlayın ve o sıradaki değer kartildir.

Kartiller Örnek

Örnek: 11 12 13 16 16 17 18 21 22



(n = 9)

$Q_1 (9+1)/4 = 2.5$ sıradaki

2 ve 3'üncü sıradaki değerlerin ortalaması 1. kartildir.

$$Q_1 = 12.5$$

Q_1 ve Q_3 merkezi bir yer belirtmez.

Q_2 = median, merkezi eğilim ölçüsüdür.

Kartiller Örnek

Örnek: 11 12 13 16 16 17 18 21 22

(n = 9)

Q_1 $(9+1)/4 = 2.5$ sıradaki değer, 2'inci ve 3'üncü sıradaki değerlerin ortalaması

$$Q_1 = (12+13)/2 = 12.5$$

Q_2 $(9+1)/2 = 5.$ sıradaki değer,

$$Q_2 = \text{medyan} = 16$$

Q_3 $3(9+1)/4 = 7.5$ sıradaki değer, 7'inci ve 8'inci sıradaki değerlerin ortalaması

$$Q_3 = (18+21)/2 = 19.5$$

Q_1 ve Q_3 merkezi bir yer belirtmez.

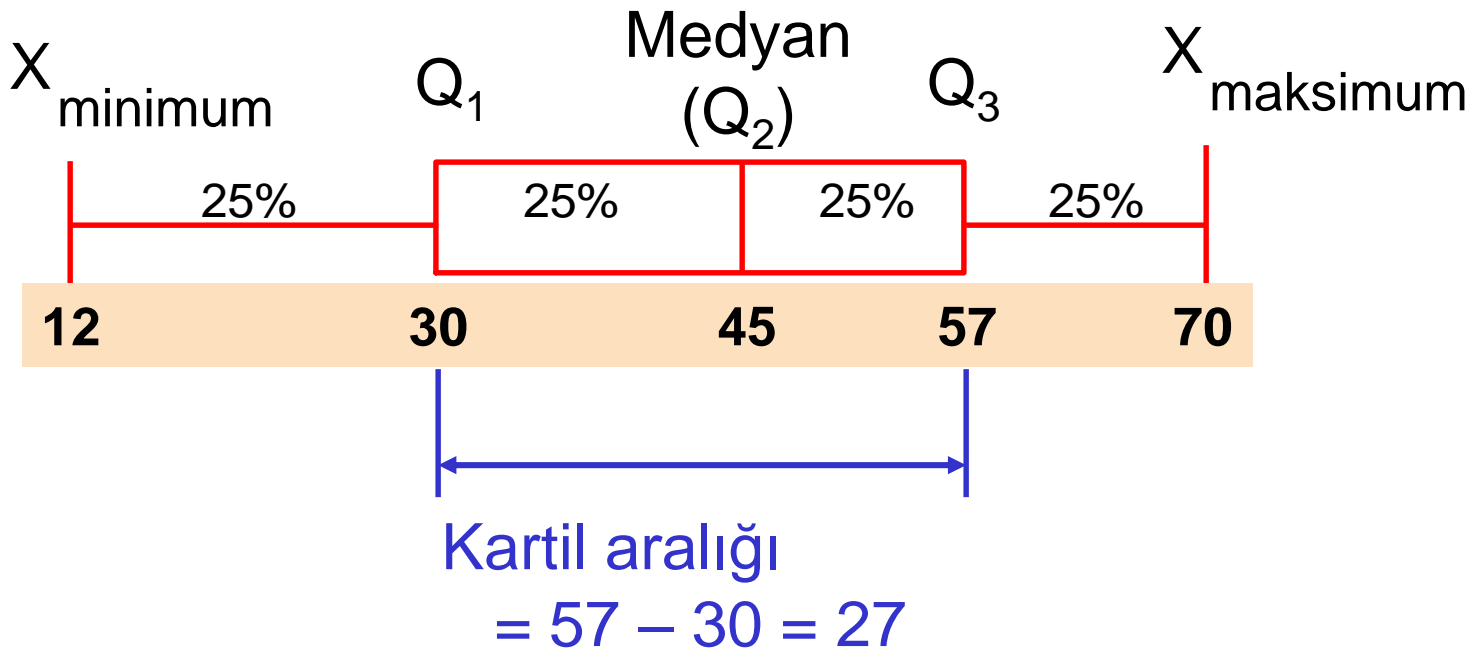
$Q_2 = \text{median}$, merkezi eğilim ölçüsüdür.

Kartiller: Kartil Aralığı

- Kartil aralığı $Q3 - Q1$ 'dir ve verilerin ortada yer alan % 50'sinin dağılımını ölçer.
- Kartil aralığı, aynı zamanda verilerin ortadaki % 50'sini kapsadığı için midspread olarak da adlandırılmaktadır.
- Kartil aralığı, aykırı veya aşırı değerlerden etkilenmeyen bir değişkenlik ölçüsüdür.
- Aykırı değerlerden etkilenmeyen $Q1$, $Q3$ ve kartil aralığı gibi ölçülere dayanıklı ölçüler de denir.

Kartil Aralığı Hesaplama

Örnek:



Beş Özet Ölçü

Veriyi tanımlamaya yarayan 5 özet değer:

- $X_{\text{Enküçük}}$
- (Q_1)
- Medyan (Q_2)
- (Q_3)
- $X_{\text{Enbüyük}}$

Beş Özet Ölçü ile Dağılım Şekli Arasındaki İlişki

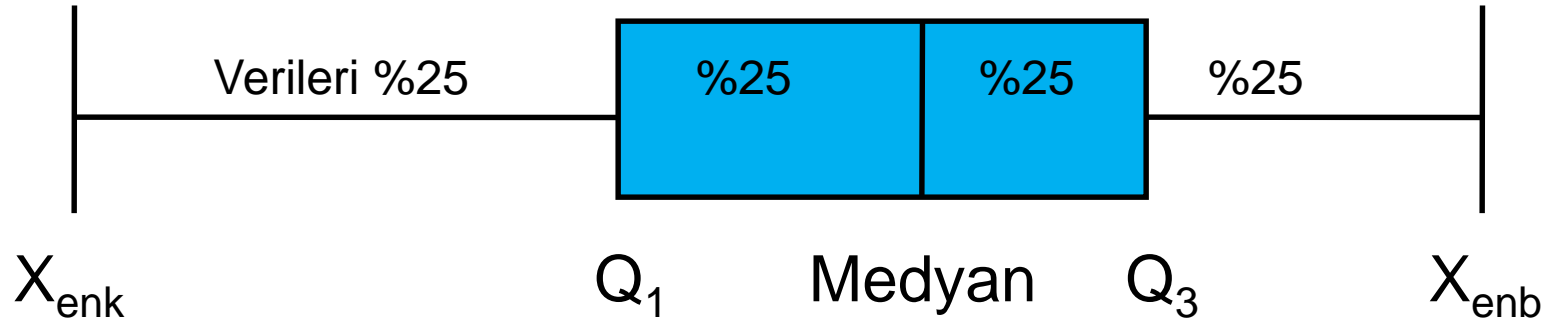
Sola Çarpık	Simetrik	Sağa Çarpık
Medyan $- X_{enk}$ > $X_{enb} - \text{Medyan}$	Medyan $- X_{enk}$ \approx $X_{enb} - \text{Medyan}$	Medyan $- X_{enk}$ < $X_{enb} - \text{Medyan}$
$Q_1 - X_{enk}$ > $X_{enb} - Q_3$	$Q_1 - X_{enk}$ \approx $X_{enb} - Q_3$	$Q_1 - X_{enk}$ < $X_{enb} - Q_3$
Medyan $- Q_1$ > $Q_3 - \text{Medyan}$	Medyan $- Q_1$ \approx $Q_3 - \text{Medyan}$	Median $- Q_1$ < $Q_3 - \text{Medyan}$

Beş Özet Ölçü Değeri ve Boxplot (Kutu Grafiği)

- **Boxplot:** Verilerin beş özet ölçü değerine dayanan grafiğine boxplot denir.

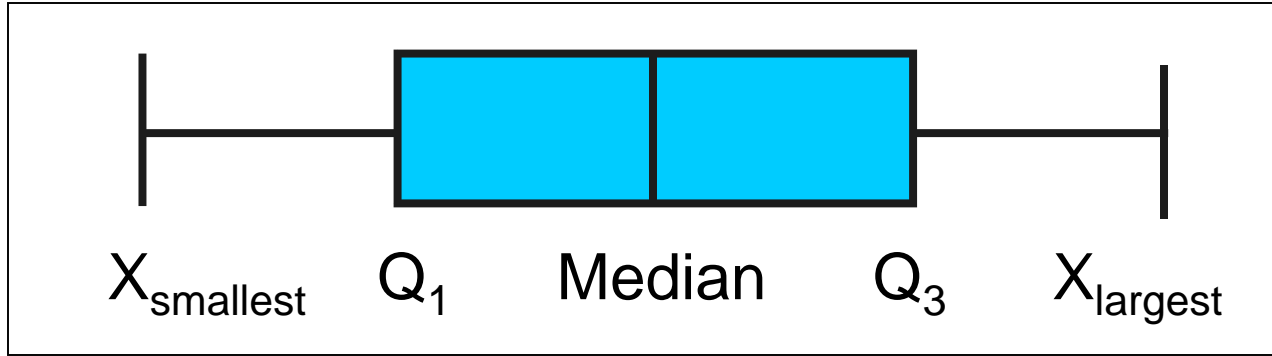


Örnek:



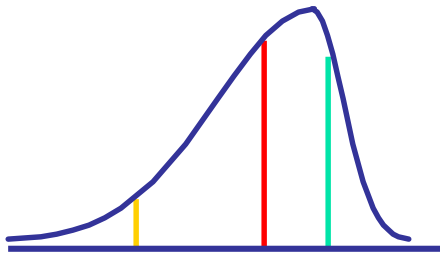
Beş Özet Ölçü Değeri ve Boxplot (Kutu Grafiği)

- Veriler simetrik dağılıyorsa, boxplot aşağıdaki gibi görünecektir.



Dağılım Biçimleri ve Kutu Grafiği (Boxplot)

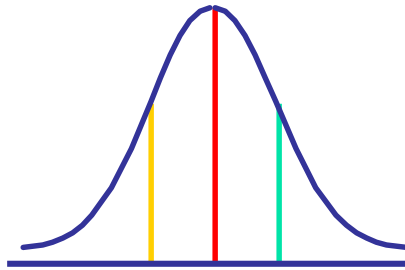
Sola Çarpık



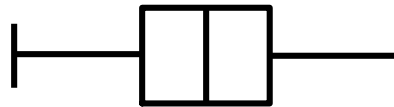
Q_1 Q_2 Q_3



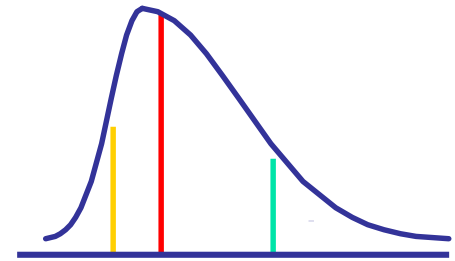
Simetrik



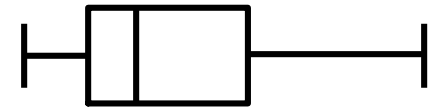
Q_1 Q_2 Q_3



Sağa Çarpık

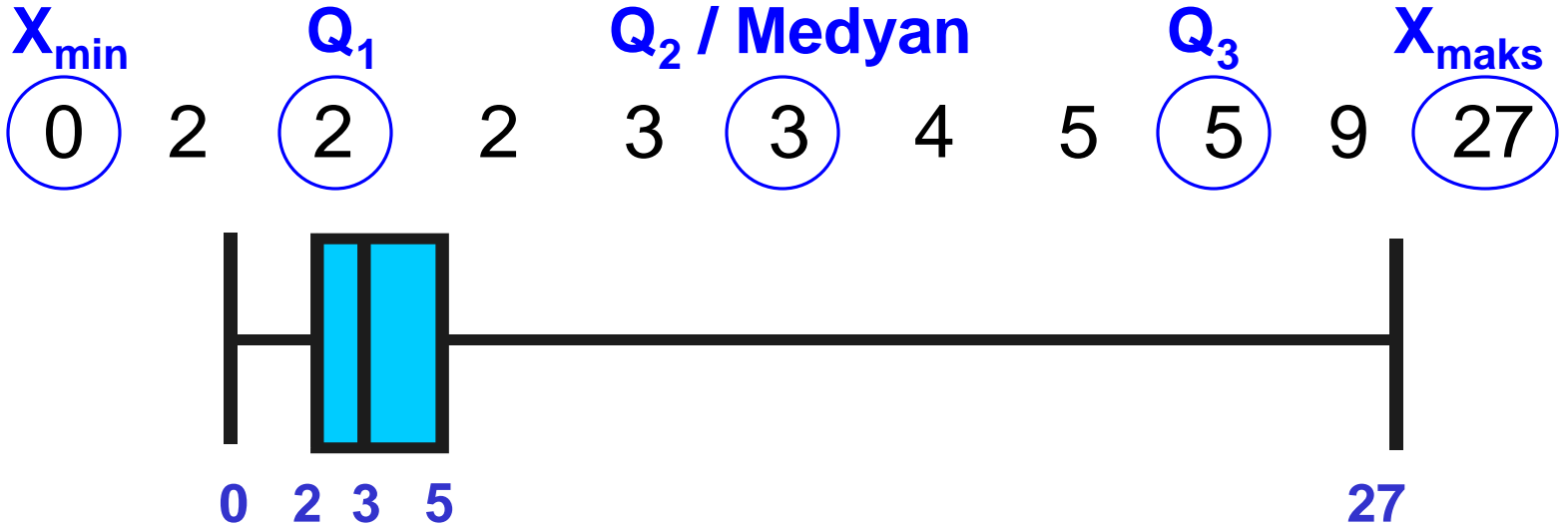


Q_1 Q_2 Q_3



Boxplot Örneđi

- Aşağıdaki verilere göre kutu grafiđi:



- Seri sađa çarpıktır

Anakitle Tanımlayıcı Ölçüleri

- Buraya kadar örneklem için tanımlayıcı ölçüler (karakterisitik değerler) açıklanmıştır. Örneklem için hesaplanan tanımlayıcı ölçülere istatistik denir ve latin harfleri ile gösterilir.
- Anakitleyi tanımlayıcı ölçülere **parametre** denir ve yunan harflari ile gösterilir.
- Önemli anakitle parametreleri aritmetik ortalama, varyans ve standart sapmadır.

Anakitle Tanımlayıcı Ölçüleri: Ortalama (μ)

- Anakitle ortalaması, anakitle toplam değerinin anakitle birim sayısına (N) bölümüne eşittir.

$$\mu = \frac{\sum_{i=1}^N X_i}{N} = \frac{X_1 + X_2 + \dots + X_N}{N}$$

μ = anakitle ortalaması

N = anakitle birim sayısı

X_i = X değişkeninin i'inci değeri

Anakitle Tanımlayıcı Ölçüleri: Varyans (σ^2)

- Aritmetik ortalamadan sapmaların karelerinin ortalaması
- Anakitle varyansı:

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$$

μ = anakitle ortalaması

N = anakitle büyüklüğü (birim sayısı)

X_i = X değişkeninin i 'inci değeri

Anakitle Tanımlayıcı Ölçüleri: Standart Sapma (σ)

- En yaygın kullanılan değişkenlik ölçüsü
- Aritmetik ortalamadan sapmaları gösterir
- Anakitle varyansının kareköküdür
- Orijinal verilerle aynı ölçü birimine sahiptir

- Anakitle standart sapması:

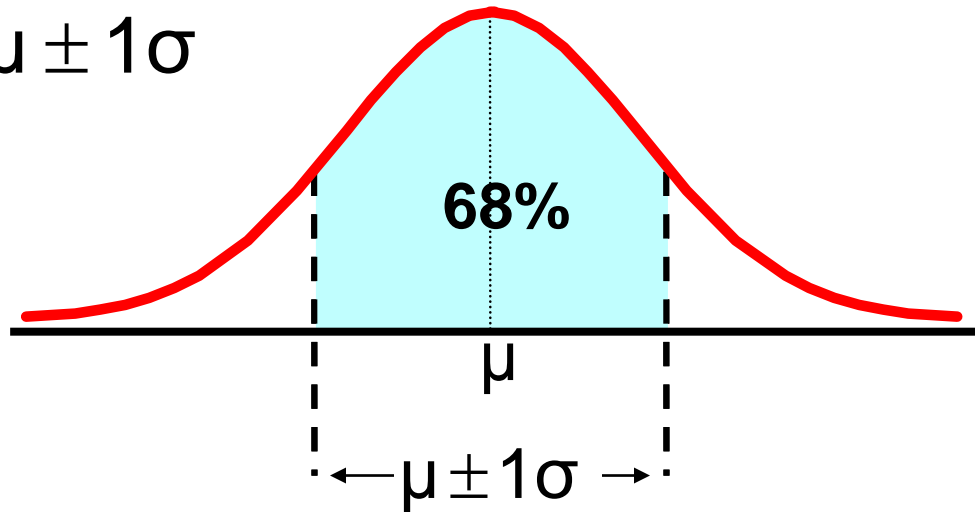
$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}}$$

Örnekleme İstatistikleri ve Anakitle Parametreleri

Ölçü	Anakitle Parametre	Örnekleme İstatistik
Aritmetik Ortalama	μ	\bar{X}
Varyans	σ^2	S^2
Standard Sapma	σ	S

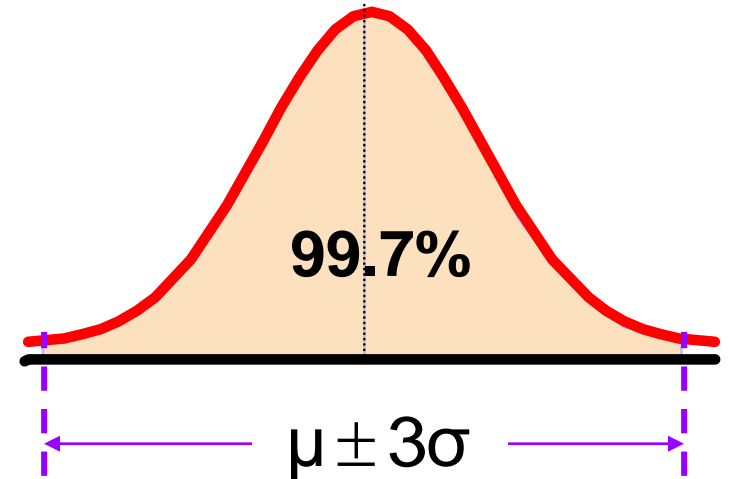
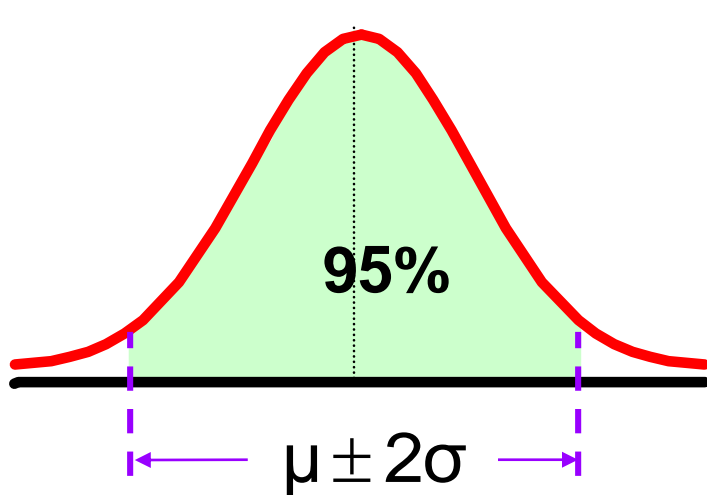
Ampirik Kural

- Bir verilerin yayılımı (dağılımı) çan eğrisine dağılımına benzer.
- Çan eğrisi şeklinde dağılan bir veri setinde verilerin yaklaşık %68'i aritmetik ortalamanın ± 1 standard sapma uzaklığındaki aralıkta yer alır. $\mu \pm 1\sigma$



Ampirik Kural

- Çan eğrisi şeklindeki bir dağılımda verilerin yaklaşık% 95'i, ortalamanın ± 2 standart sapma uzaklığındaki aralıkta yer alır. ($\mu \pm 2\sigma$).
- Çan eğrisi şeklindeki bir dağılımda verilerin yaklaşık% 99,7'si, ortalamanın ± 3 standart sapma uzaklığındaki aralıkta yer alır. ($\mu \pm 3\sigma$).



Ampirik Kuralın Kullanma

- Matematik puanlarının ortalamasının 500 ve standart sapmasının 90 olduğunu ve çan eğirisi biçiminde dağıldığını varsayalım. Buna göre;
 - Tüm test katılımcılarının %68'i 410 ile 590 (500 ± 90) arasında puan almıştır.
 - Tüm test katılımcılarının %95'i 320 ile 680 (500 ± 180) arasında puan almıştır.
 - Tüm test katılımcılarının %99,7'si 230 ile 770 arasında (500 ± 270) almıştır.

Chebyshev Kuralı

- Verilerin nasıl dağıldığına bakılmaksızın, değerlerin en az $\%(1 - 1 / k^2) \times 100$ 'ü, ortalamadan $\pm k$ standart sapma içerisinde kalacaktır ($k > 1$ için)
 - Örnek:

En az	İçinde
$(1 - 1/2^2) \times 100\% = 75\%$	$k=2 \quad (\mu \pm 2\sigma)$
$(1 - 1/3^2) \times 100\% = 88.89\%$	$k=3 \quad (\mu \pm 3\sigma)$

İki Nicel Değişken Arasındaki İlişkinin İki Ölçüsü

- Serpilme diyagramları, iki sayısal değişken arasındaki ilişkiyi görsel olarak inceleyebilmemize olanak sağlar ve bu ilişkilerin iki nicel ölçümü söz konusudur:
- Kovaryans (birlikte değişim)
- Korelasyon Katsayısı (ilişki katsayısı)

Kovaryans

- Kovaryans, iki sayısal değişken (X & Y) arasındaki doğrusal ilişkinin (birlikte değişiminin) gücünü ölçer.
- Örneklem Kovaryansı:

$$kov(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$$

- Sadece ilişkinin gücünü ölçer.
- Nedensellik ile ilgili bilgi vermez.

Kovaryansın yorumlanması

- İki değişken arasındaki **kovaryans**:

$\text{kov}(X, Y) > 0 \rightarrow X$ ve Y aynı yönde değişiyor

$\text{kov}(X, Y) < 0 \rightarrow X$ ve Y negatif (ters) yönde değişiyor

$\text{kov}(X, Y) = 0 \rightarrow X$ ve Y bağımsız

- Kovaryansın eksik yanı:

- İlişkinin gücünü kovaryansın büyüklüğünden belirlemek mümkün değildir.

Korelasyon Katsayısı

- İki nicel değişken arasındaki doğrusal ilişkinin gücünü ölçer.
- Örneklem için korelasyon katsayısı:

$$r = \frac{\text{cov}(X, Y)}{S_X S_Y}$$

$$\text{kov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$

$$S_X = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

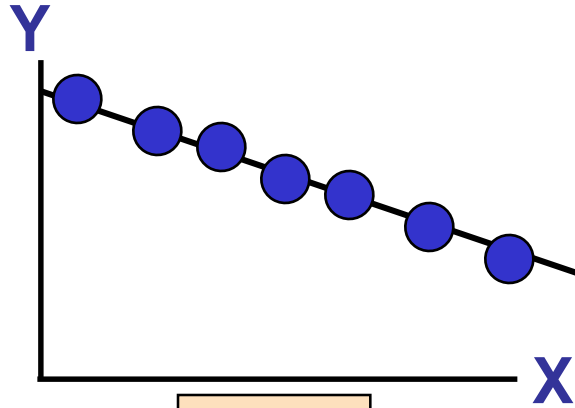
$$S_Y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}}$$

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$$

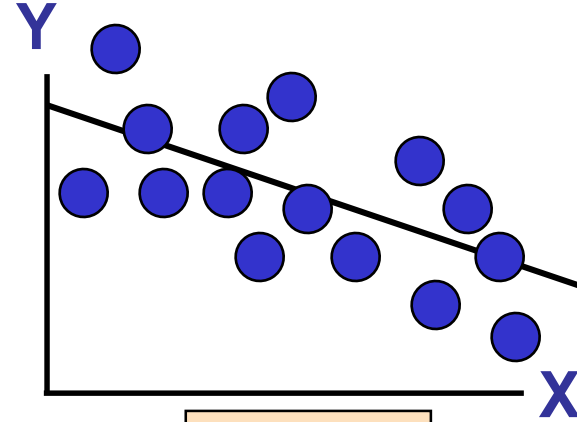
Korelasyon Katsayısının Özellikleri

- Anakitle doğrusal korelasyon katsayısı = ρ
- Örneklem için doğrusal korelasyon katsayısı = r
- ρ ve r aşağıdaki özelliklere sahiptir:
 - Ölçü birimi yoktur.
 - -1 ve $+1$ arasında değerler alır.
 - -1 'e yaklaştıkça, negatif doğrusal ilişki daha güçlü hale gelir.
 - $+1$ 'e yaklaştıkça, pozitif doğrusal ilişki daha güçlü hale gelir.
 - 0 'a yaklaştıkça, doğrusal ilişki zayıflar.

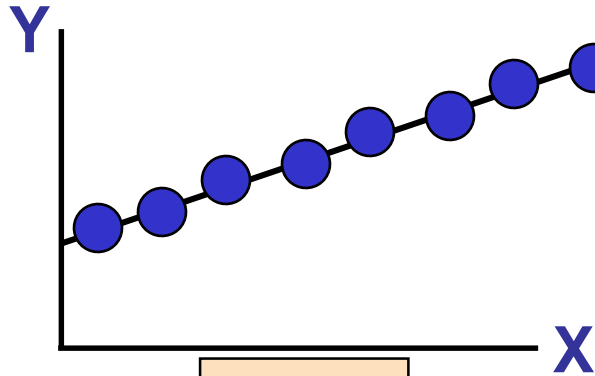
Farklı Korelasyonlara İlişkin Veriler Serpilme Diyagramı



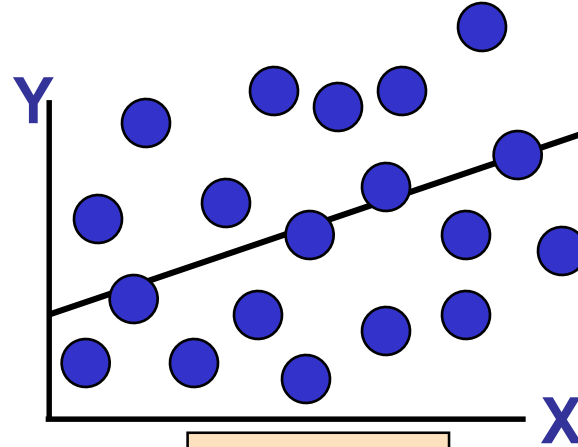
$$r = -1$$



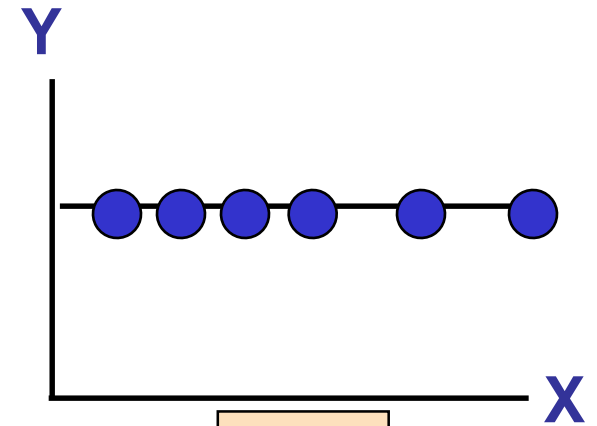
$$r = -0,6$$



$$r = +1$$



$$r = +0,3$$

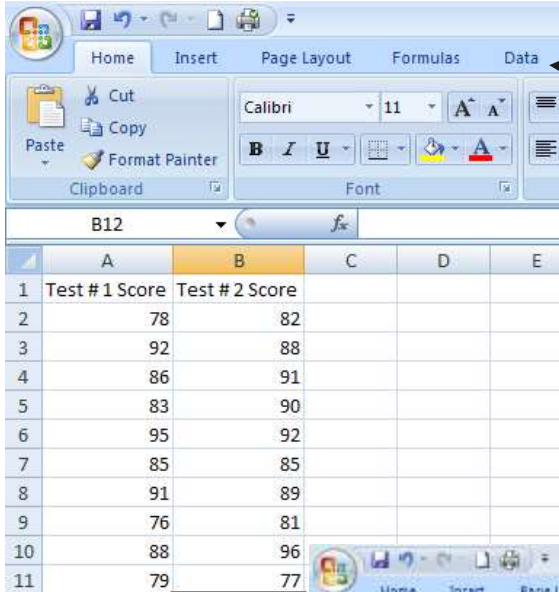


$$r = 0$$

«Microsoft Excel Function»ı Kullanarak Korelasyon Hesaplama

Test1 Puanı	Test2 Puanı		<u>Korelasyon Katsayısı</u>
78	82		0,7332 =CORREL(A2:A11,B2:B11)
92	88		
86	91		
83	90		
95	92		
85	85		
91	89		
76	81		
88	96		
79	77		

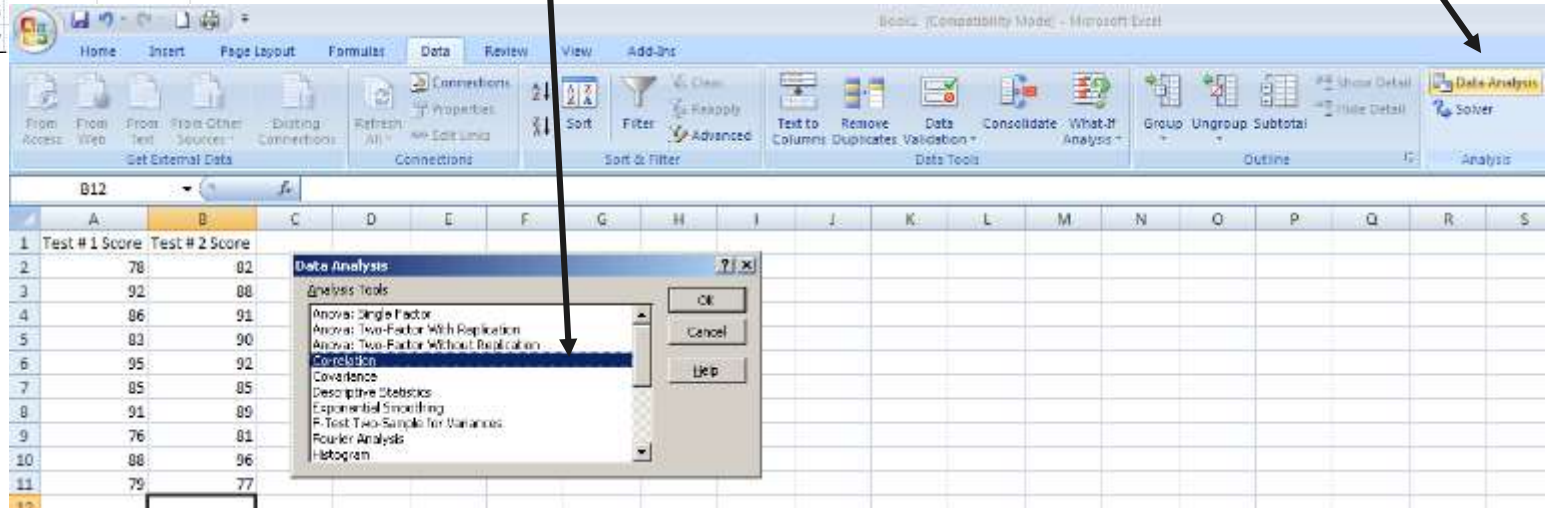
«Microsoft Excel Data Analysis Tool»u Kullanarak Korelasyon Katsayısı Hesaplama



The screenshot shows the Microsoft Excel interface with the 'Data' menu selected. The table below contains the data used for the correlation analysis.

	A	B	C	D	E
1	Test # 1 Score	Test # 2 Score			
2	78	82			
3	92	88			
4	86	91			
5	83	90			
6	95	92			
7	85	85			
8	91	89			
9	76	81			
10	88	96			
11	79	77			

1. «Data» menüsünü seç
2. «Data Analysis» seç
3. «Correlation»ı seç ve «OK» tıkla



The screenshot shows the Microsoft Excel interface with the 'Data Analysis' toolpak installed. The 'Data Analysis' dialog box is open, and 'Correlation' is selected in the list of analysis tools. The 'OK' button is highlighted.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
1	Test # 1 Score	Test # 2 Score																		
2	78	82																		
3	92	88																		
4	86	91																		
5	83	90																		
6	95	92																		
7	85	85																		
8	91	89																		
9	76	81																		
10	88	96																		
11	79	77																		

Microsoft Excel'de Koralsyon Hesaplama

	A	B
1	Test # 1 Score	Test # 2 Score
2	78	82
3	92	88
4	86	91
5	83	90
6	95	92
7	85	85
8	91	89
9	76	81
10	88	96
11	79	77

Correlation

Input Range:

Grouped By: Columns Rows

Labels in First Row

Output options:

Output Range:

New Worksheet Ply:

New Workbook

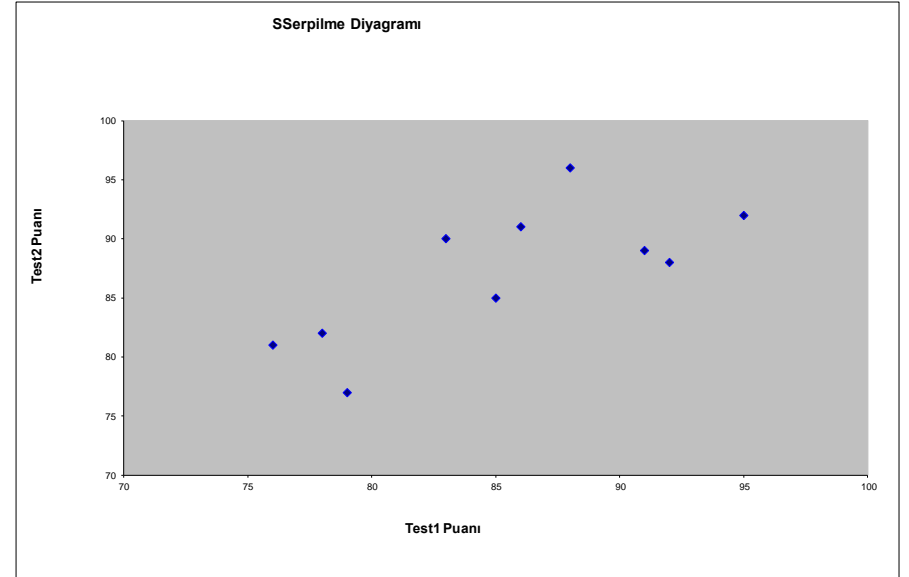
OK Cancel Help

4. Verilerin olduđu aralıđı belirle ve gerekli opsiyonu işaretleyerek «OK»ye tıkla.

	A	B	C
1		Test # 1 Score	Test # 2 Score
2	Test # 1 Score	1	
3	Test # 2 Score	0.733243705	1
4			

Korelasyon katsayısını Microsoft Excel' hesaplanması

- $r = 0,733$
- Test1 puanı ile Test2 puanı arasında pozitif yönlü korelasyon vardır.
- Test1'den yüksek puan alan öğrenciler genel olarak Test2'den de yüksek puan almıştır.



Sayısal tanımlayıcı ölçülerde tuzaklar

- Veri analizi objektiftir.
 - Veri kümesinin önemli yönlerini en iyi tanımlayan ve belirten özet ölçüler rapor edilmelidir.
- Veri yorumlama subjektiftir.
 - Adil, tarafsız ve açık bir şekilde yapılmalıdır.

Sayısal tanımlayıcı ölçüler:

- Hem olumlu hem de olumsuz sonuçları belgelenmelidir.
- Adil, tarafsız ve tarafsız bir şekilde sunulmalıdır.
- Gerçekleri çarpıtmak için uygunsuz özet ölçüler kullanılmamalıdır.



Doğrusal Regresyon

- Doğrusal trend doğrusunun tahmininde kullanılır.
- Bağımlı değişkeni Y ile zaman X arasındaki ilişkinin doğrusal bir fonksiyon olduğunu varsayar...

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

- En küçük kareler yöntemiyle tahminlenir.
 - Hata kareleri toplamını en küçükler

Denkleminde sabit

β_0

a

Denkleminde eğim

β_1

b

Trend denklemi

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

$$\hat{Y}_i = a + bX_i$$

$$Y_i = a + bX_i + \hat{e}_i$$

Doğrusal Regresyon

- Bağımlı değişken: Y_i
- Bağımsız değişkenler: X_1, X_2, X_3

$$Y_i = a + b_1X_1 + b_2X_2 + b_3X_3$$

Bağımlı
değişken

Sabit

Katsayılar

Bağımsız değişkenler

Doğrusal Regresyon Örneği

- Talep birden çok değişkenle ilişkili
- Örnekler
 - Reklam giderleri
 - Satış elemanı sayısı
 - Nüfus artışı
 - Enflasyon hızı
 - v.b.

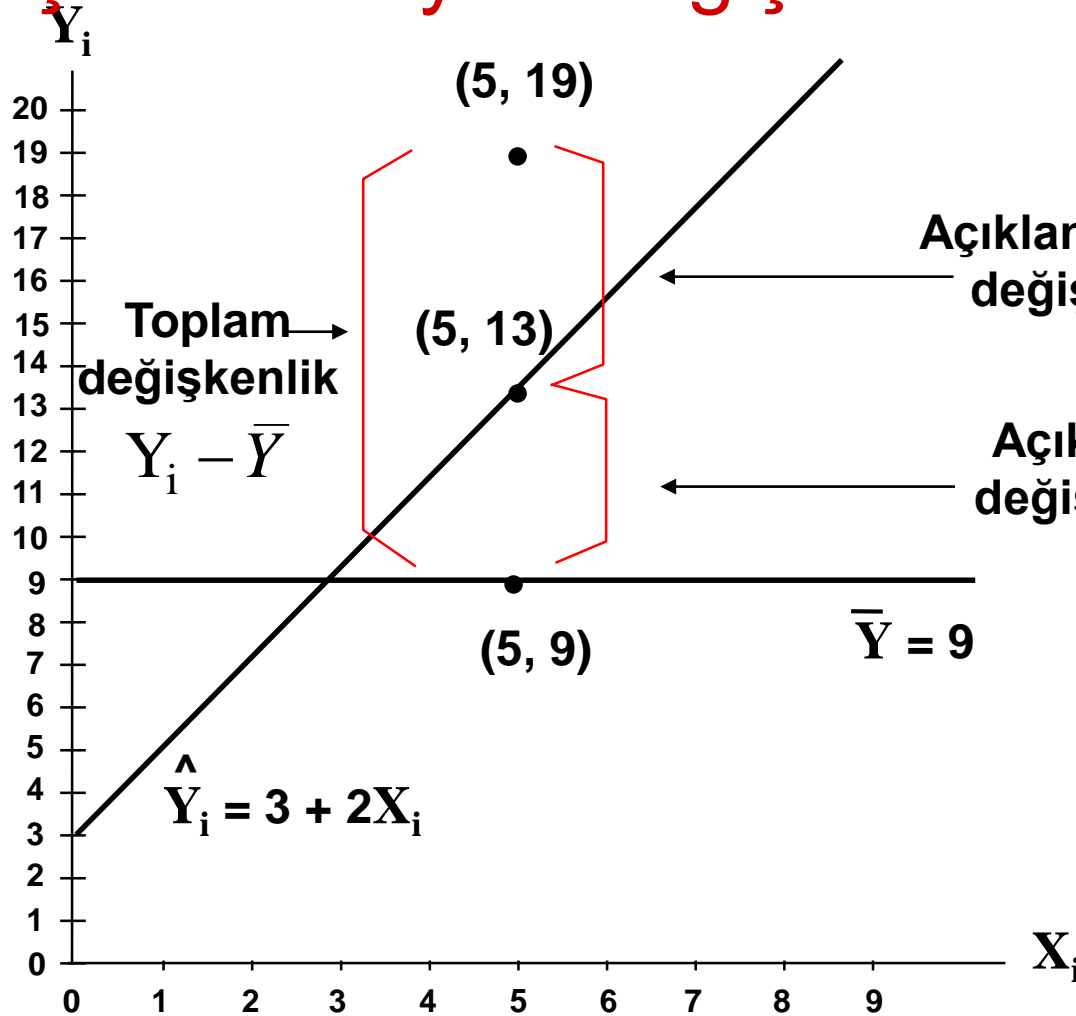
Doğrusal Regresyon

- Doğrusal regresyon, iki değişken arasındaki ilişkinin bir doğru ile modelleneceği esasına dayanır.
- Öngörülecek bağımlı değişken Y , diğer değişkene (X -bağımsız değişken) bir doğru şeklinde ilişkilendirilir.
- İki değişken arasındaki ilişki:
 $Y_i = a + b \cdot X_i$ a ve b , doğrudan sapmaları –hataların kareleri toplamını- minimum yapacak şekilde seçilir
- a = doğrunun Y 'yi kestiği yer
- b = doğrunun eğimi

Sabitlerin Açıklaması

- Eğim(b)
 - X 'teki her 1 birim artış için Y 'de b ile öngörülen değişimler
 - $b = 2$ ise reklam harcamalarındaki (X) her 1 birim artış için satışların (Y) 2 kat artması beklenir.
- Y -eksen kesimi (a)
 - $X = 0$ iken Y 'nin ortalama değeri
 - $a = 4$ ise reklam harcamaları (X) 0 olduğunda, ortalama satışların (Y) 4 olması beklenir.

Toplam Değişkenlik, Açıklanan Değişkenlik, Açıklanamayan Değişkenlik ve Regresyon



**(Toplam değişkenlik) =
(Açıklanan değişkenlik) +
(Açıklanamayan değişkenlik)**

$$Y_i - \bar{Y} = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i)$$

**(Genel kareler toplamı) =
(Regresyon kareler
toplamı) + (Artık kareler
toplamı)**

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2$$

Doğrusal Regresyon

Regresyon Denklemi: $\hat{Y}_i = a + bX_i$

Eğim:
$$b = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2}$$

Y-eksen kesmesi: $a = \bar{Y} - b\bar{X}$