



Bölüm 2

Verileri Düzenleme ve Görselleştirme



UZAKTAN EĞİTİM UYGULAMA VE ARAŞTIRMA MERKEZİ

telefon 0(312) 202 82 00 • eposta guzem@gazi.edu.tr • adres Gazi Üniversitesi Rektörlük Binası No:6/f

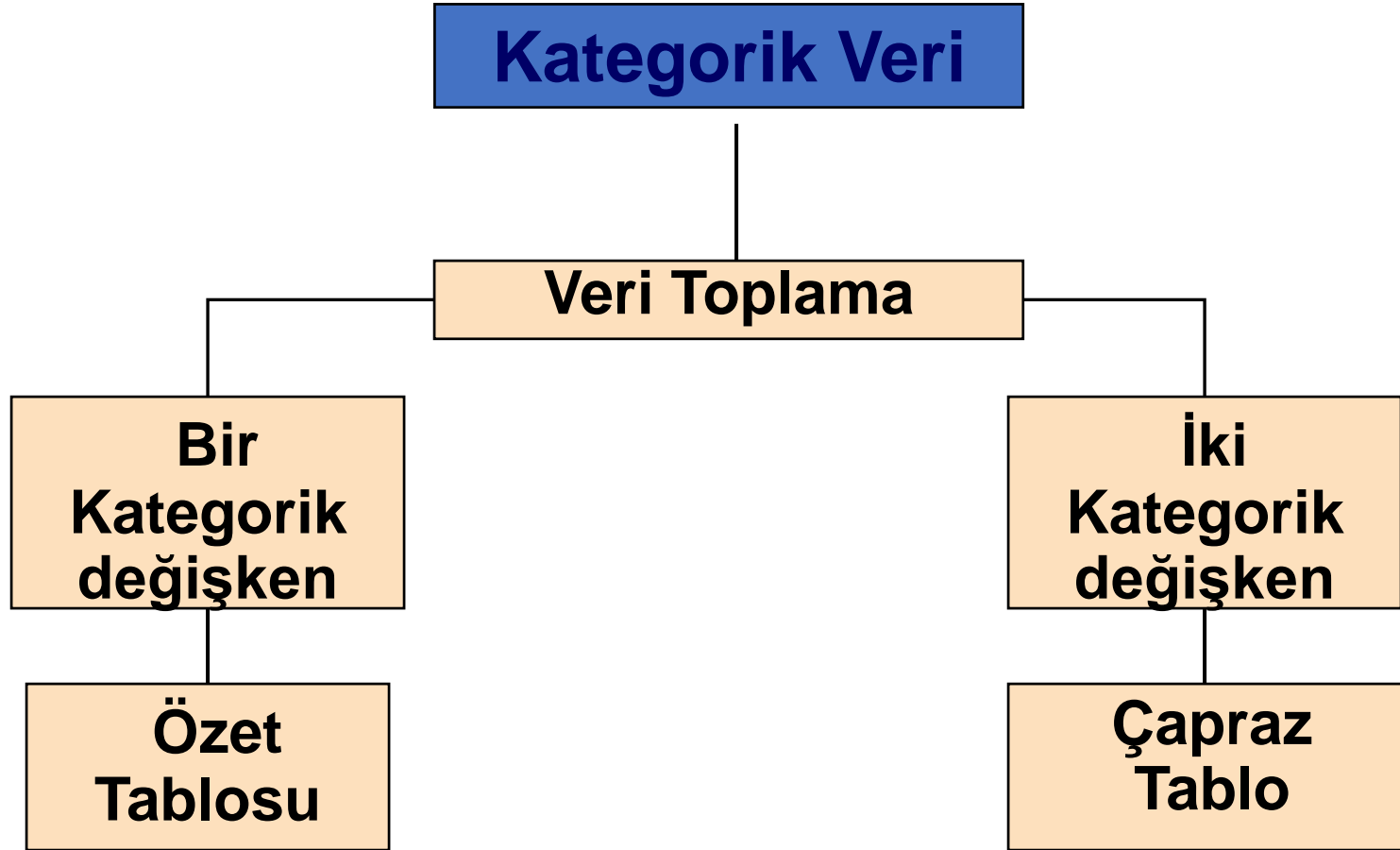
guzem.gazi.edu.tr • uzaktanegitim.gazi.edu.tr • lms.gazi.edu.tr

Öğrenme hedefleri

Bu bölümde şunlar öğrenilecektir:

- Kategorik veriler için Tablo ve Şemaların oluşturulması
- Sayısal veriler için Tablo ve Şemaların oluşturulması
- Grafikleri düzgün bir şekilde sunma ilkeleri
- Değişkenleri düzenlemek ve analiz etmek

Kategorik veriler tablolar kullanılarak düzenlenir



Kategorik verilerin düzenlenmesi: Özet tablo

- Bir **özet tablo**, kategoriler arasındaki farkları görebilmemiz için kategori kümesindeki öğelerin frekanslarını veya yüzdelerini belirler.

1000 Müşteriden Elde Edilen Özet Tablo

Bankacılık Tercihi?	Yüzde
ATM	16%
Otomatize veya canlı telefon	2%
Şubede araca servis hizmeti	17%
Şubede çalışan hizmeti	41%
İnternet	24%

Bir apraz tablo, iki veya daha fazla kategorik deęiřkeni dzenlemeye yardımcı olur

- İki veya daha fazla kategorik deęiřkenin yanıtları arasında var olabilecek iliřki modellerini incelemek iin kullanılır,
- Kategorik deęiřkenlerin tepkilerini birleřik olarak apraz tablolar veya tasnif eder,
- İki deęiřken iin birinci deęiřkene ait deęerler satırlar iinde ve ikinci deęiřken iin deęerler stunlarda bulunur.

Çapraz tablo- Örnek

- 400 faturalık rassal bir örnek oluşturuldu.
- Her fatura küçük, orta veya büyük miktarlı olarak sınıflandırıldı.
- Her bir fatura, ayrıca herhangi bir hata olup olmadığını tespit etmek için incelendi.
- Bu veriler daha sonra sağdaki çapraz tabloda düzenlendi.

Boyuta ve Hataların Varlığına Göre Sınıflandırılan Faturaların Sıklığını Gösteren Çapraz Tablo

	Hatasız	Hata Sayısı	Toplam
Küçük miktar	170	20	190
Orta miktar	100	40	140
Büyük miktar	65	5	70
Toplam	335	65	400

Genel toplamdaki yüzdeye dayalı çapraz tablo

	Hatasız	Hata Sayısı	Toplam
Küçük miktar	170	20	190
Orta miktar	100	40	140
Büyük miktar	65	5	70
Toplam	335	65	400

$$42.50\% = 170 / 400$$
$$25.00\% = 100 / 400$$
$$16.25\% = 65 / 400$$

	Hatasız	Hata Sayısı	Toplam
Küçük miktar	42.50%	5.00%	47.50%
Orta miktar	25.00%	10.00%	35.00%
Büyük miktar	16.25%	1.25%	17.50%
Toplam	83.75%	16.25%	100.0%

Örneklenen faturaların %83,75'inde hata bulunmuyor ve örneklenen faturaların %47,50'si küçük miktarlar içindir.

Satır toplamındaki yüzdeye dayalı çapraz tablo

	Hatasız	Hata sayısı	Toplam
Küçük miktar	170	20	190
Orta miktar	100	40	140
Büyük miktar	65	5	70
Toplam	335	65	400

Orta miktarlı faturaların, küçük miktarlı (% 10.53) veya büyük miktarlı (% 7.14) faturalardan daha büyük bir hata olasılığı mevcuttur (% 28.57).

$$89.47\% = 170 / 190$$
$$71.43\% = 100 / 140$$
$$92.86\% = 65 / 70$$

	Hatasız	Hata sayısı	Toplam
Küçük miktar	89.47%	10.53%	100.0%
Orta miktar	71.43%	28.57%	100.0%
Büyük miktar	92.86%	7.14%	100.0%
Toplam	83.75%	16.25%	100.0%

Sütun toplamındaki yüzdeye dayalı çapraz tablo

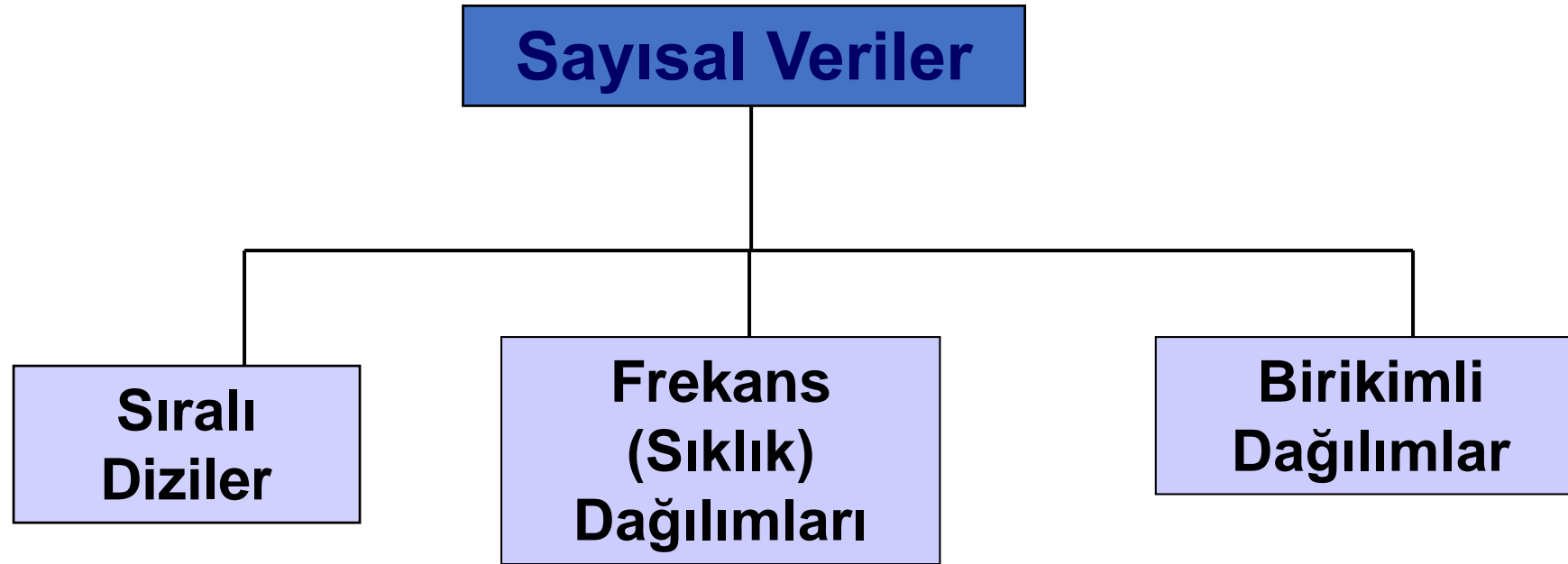
	Hatasız	Hata sayısı	Toplam
Küçük miktar	170	20	190
Orta miktar	100	40	140
Büyük miktar	65	5	70
Toplam	335	65	400

Hatalı faturaların orta miktarlılarda olması olasılığı % 61.54'tür.

$$50.75\% = 170 / 335$$
$$30.77\% = 20 / 65$$

	Hatasız	Hata sayısı	Toplam
Küçük miktar	50.75%	30.77%	47.50%
Orta miktar	29.85%	61.54%	35.00%
Büyük miktar	19.40%	7.69%	17.50%
Toplam	100.0%	100.0%	100.0%

Sayısal verilerin düzenlenmesi için kullanılan tablolar



Yığılanmış veya yığılanmamış biçim

- Analiz için sayısal değişkeninizi gruplayabilecek kategorik bir değişkene sahip olduğunuzdaki bir durumdur.
- Yığılanmış biçim, sayısal değişkeninizin bir sütunda olduğu, ikinci sütunun kategorik değişkenin değerini tanımladığı durumdur.
- Yığılanmamış biçim, her grubun sayısal değişkeninin değerleri (kategorik değişkenin belirli değeri) farklı sütunlarda olduğunda oluşur.

Yığınlanmış & Yığınlanmamış biçim örneği

Farklı Programlar
ve farklı analizler
için belirli bir
format gerekebilir

Yığınlanmış Biçim		Yığınlanmamış Biçim	
Öğrencilerin	Gündüz veya	Gündüz	Gece Öğrencilerinin
Yaşı	Gece Öğrencisi	Yaşı	Yaşı
16	G	16	18
19	G	19	23
22	G	22	18
18	Gc	17	28
23	Gc	19	19
17	G	25	32
19	G	17	19
25	G	20	33
18	Gc	27	
28	Gc	18	
17	G	20	
20	G	32	
27	G		
19	Gc		
32	Gc		
18	G		
20	G		
32	G		
19	Gc		
33	Gc		

Sayısal verilerin düzenlenmesi:

Sıralı diziler

- Bir **sıralı dizi**, en küçük değerden en büyük değere sıralı olarak verilen bir veri dizisidir.
- Aralık gösterir (en küçük değerden en büyük değere)
- Aykırı değer belirlemeye yardımcı olur (alışılmamış gözlemler)

Araştırılan Üniversite öğrencilerinin Yaşı	Gündüz Öğrencileri					
	16	17	17	18	18	18
	19	19	20	20	21	22
	22	25	27	32	38	42
	Gece Öğrencileri					
	18	18	19	19	20	21
	23	28	32	33	41	45

Sayısal verilerin düzenlenmesi: Frekans dağılımları

- **Frekans dağılımı**, verilerin sayısal olarak düzenlenmiş sınıflara göre düzenlendiği özet tablodur
- Tablo için uygun **sınıf gruplarının** *sayılarının* seçilmesine, sınıf gruplamasının uygun *genişliğini* belirlemeye ve çakışmayı önlemek için her bir sınıf gruplamasının *sınırlarının* belirlenmesine dikkat edilmelidir.
- Sınıfların sayısı verilerin içindeki değer sayısına bağlıdır. Çok sayıda değerle, genellikle daha fazla sınıf oluşacaktır. Genel olarak, bir frekans dağılımının en az 5 fakat en fazla 15 sınıfı içermesi gerekir.
- **Bir sınıf aralığının genişliğini** belirlemek için, verilerin **aralığını** (En Yüksek Değer - En Düşük Değer) istenen sınıf gruplamaların sayısına bölün.

Sayısal verilerin düzenlenmesi: Frekans dağılımları

Örnek: Bir yalıtım malzemesi üreticisi rastgele 20 kış gününü seçer ve günlük en yüksek sıcaklıkları kaydeder

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

Sayısal verilerin düzenlenmesi: Frekans dağılımları örneği

- Ham veriyi artan sırada sırala:
12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58
- Aralığı bul: **58 - 12 = 46**
- Sınıfların sayısını belirle: **5 (genellikle 5 ve 15 arasında)**
(Sturges kuralı kullanılabilir; sınıf sayısı = $1 + 3.2 * \log(n)$)
- Sınıf aralığını hesapla (genişlik): **10 (46/5 üste yuvarla)**
- Sınıf sınırlarını belirle (limitler):
 - Sınıf 1: 10 dan fazla 20 den aza
 - Sınıf 2: 20 den fazla 30 dan aza
 - Sınıf 3: 30 dan fazla 40 dan aza
 - Sınıf 4: 40 dan fazla 50 den aza
 - Sınıf 5: 50 den fazla 60 dan aza
- Sınıf orta noktalarını hesapla: **15, 25, 35, 45, 55**
- Gözlemleri say & sınıflara ata

Sayısal verilerin düzenlenmesi: Frekans dağılımları örneği

Sıralı dizide veriler:

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

Sınıf	Orta Değerler	Frekans
10 ile 20 arasında	15	3
20 ile 30 arasında	25	6
30 ile 40 arasında	35	5
40 ile 50 arasında	45	4
50 ile 60 arasında	55	2
Toplam		20

Sayısal verilerin düzenlenmesi: Bağlı & Yüzdelik frekans dağılımı örneği

Sıralı dizide veriler:

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

Sınıf	Frekans	Bağlı Frekans	Yüzdelik
10 ile 20 arasında	3	.15	15%
20 ile 30 arasında	6	.30	30%
30 ile 40 arasında	5	.25	25%
40 ile 50 arasında	4	.20	20%
50 ile 60 arasında	2	.10	10%
Toplam	20	1.00	100%

Sayısal verilerin düzenlenmesi : Birikimli frekans dağılımı örneği

Sıralı dizide veriler:

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

Sınıf	Frekans	Yüzdelik	Birikimli Frekans	Birikimli Yüzdelik
10 ile 20 arasında	3	15%	3	15%
20 ile 30 arasında	6	30%	9	45%
30 ile 40 arasında	5	25%	14	70%
40 ile 50 arasında	4	20%	18	90%
50 ile 60 arasında	2	10%	20	100%
Toplam	20	100	20	100%

Neden frekans dağılımı kullanılır?

- Ham veriyi daha kullanışlı bir şekle dönüştürür.
- Verilerin hızlı bir şekilde görsel olarak yorumlanmasına imkan verir.
- Verilerin yoğunlaştırıldığı / kümелendiği durumu da içermek üzere veri kümesinin temel özelliklerinin belirlenmesini sağlar.

Frekans dağılımları: Bazı ipuçları

- Aynı veri için farklı sınıf sınırları, farklı resimler (özellikle küçük veri setleri için) oluşturabilir.
- Farklı sınıf sınırları seçildiğinde veri konsantrasyonunda kaymalar ortaya çıkabilir.
- Veri kümesinin boyutu arttıkça, sınıf sınırlarının seçimindeki değişimlerin etkisi büyük ölçüde azalır.
- İki veya daha fazla grubu farklı örneklem boyutlarıyla karşılaştırırken, göreceli frekans ya da yüzdelik dağılım kullanılmalıdır.

Örnek

- Bir dolum tesisinde doldurulmakta olan bakliyat paketlerinden rastgele seçilen 40 paketin ağırlıkları kg olarak aşağıda verilmiştir.
- Sınıf sayısını 7 alarak frekans tablosunu oluşturunuz.
- Nispi (izafi) frekansları, birikimli ve birikimli nispi frekansları hesaplayınız.

2.3	4.1	3.6	4.5	3.1	3.8	3.2	2.8	3.3
1.6	3.1	3.2	3.7	3.1	4.7	3.9	2.6	4.2
3.2	3.6	2.8	3.4	3.8	3.1	3.3	3.0	3.6
4.3	3.1	4.2	1.8	3.4	4.7	3.9	3.4	2.7
3.6	3.2	4.4	3.5					

Örnek

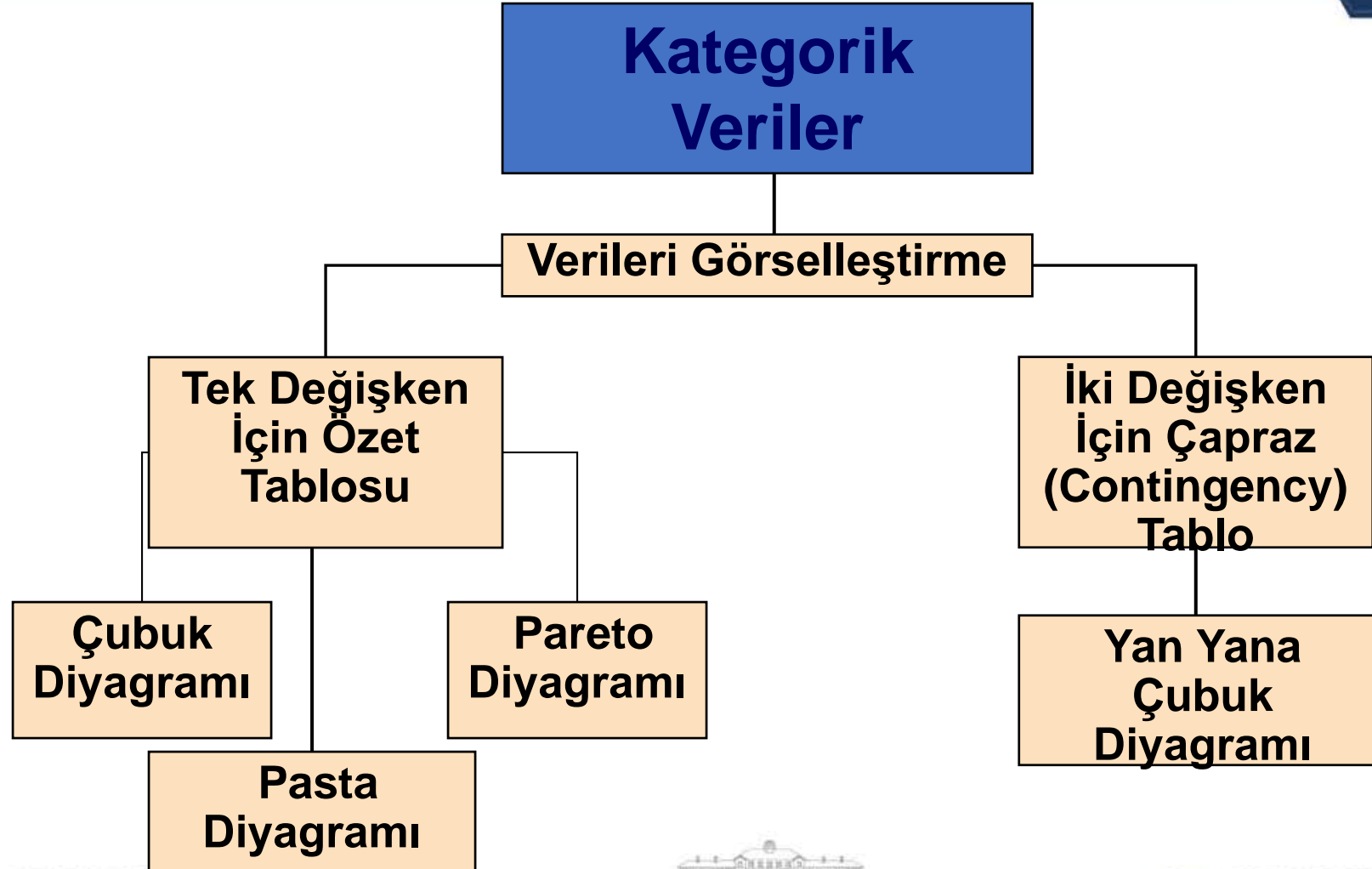
- DeğişimGenişliği= $4.7-1.6=3.1$
- SınıfAralığı= $3.1/7=0.44 \Rightarrow$ yaklaşık 0.5
- Veriler dikkate alınarak alt ve üst limitleri belirlemede kullanılacak yeterince küçük değer=0.1 olarak seçilebilir. (veriler onda birlik düzeyde olduğu için)
- Sınıf alt limitleri; $1.6-0.1=1.5$ $1.5+0.5=2.0$ $2.0+0.5=2.5$ 3.0 3.5 4.0 4.5
- Sınıf üst limitleri; $2.0-0.1=1.9$ $1.9+0.5=2.4$ $2.4+0.5=2.9$ 3.4 3.9 4.4 4.9
- Sınıf sınırları; $(2.0-1.9)/2=0.05$
 - birinci sınıf için $1.5-0.05=1.45$ ve $1.9+0.05=1.95$ (alt limitten çıkar üst limit ile topla)

Örnek

- Frekans tablosunun tarama sütununa bakıldığında verilerin hangi değerler etrafında yoğunlaştığı ve veri grubuna ait dağılımın şekli kabaca görülmektedir.

Sınıf Limitleri	Sınıf Sınırları	Sınıf Değer-leri	Tarama	Frekans	Nispi Frekans %	Sınıf sayıları	Biriki. Fre-kans	Birikiml i Nispi frekans
1.5-1.9	1.45-1.95	1.7	xx	2	5.0	1.95 denaz	2	5.0
2.0-2.4	1.95-2.45	2.2	x	1	2.5	2.45 denaz	3	7.5
2.5-2.9	2.45-2.95	2.7	xxxx	4	10.0	2.95 denaz	7	17.5
3.0-3.4	2.95-3.45	3.2	xxxxxxxxxxxxxxxxxxx	15	37.5	3.45 denaz	22	55.0
3.5-3.9	3.45-3.95	3.7	xxxxxxxxxxx	10	25.0	3.95 denaz	32	80.0
4.0-4.4	3.95-4.45	4.2	xxxxxx	5	12.5	4.45 denaz	37	92.5
4.5-4.9	4.45-4.95	4.7	xxx	3	7.5	4.95 denaz	40	100.0
T o p l a m				40	100.0			

Grafik göstergelerle kategorik verileri görselleştirme

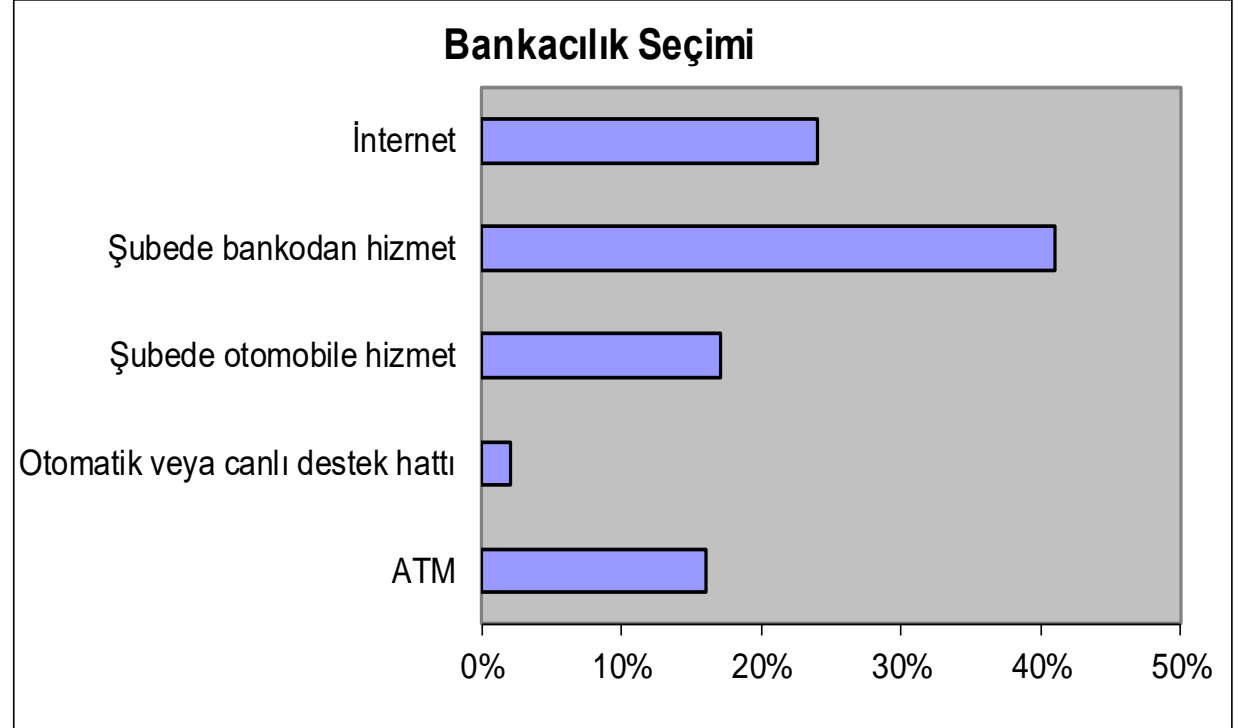


Kategorik verilerin görselleştirilmesi: Çubuk diyagramı



- Bir **çubuk diyagramda**, bir çubuk her bir kategoriye gösterir; bu çubuğun uzunluğu, değişkenin özet tablosundan gelen bir kategoriye giren değerlerin miktarını, sıklığını veya yüzdesini temsil eder.

Bankacılık Seçimi?	%
ATM	16%
Otomatik veya canlı destek hattı	2%
Şubede otomobile hizmet	17%
Şubede bankodan hizmet	41%
İnternet	24%

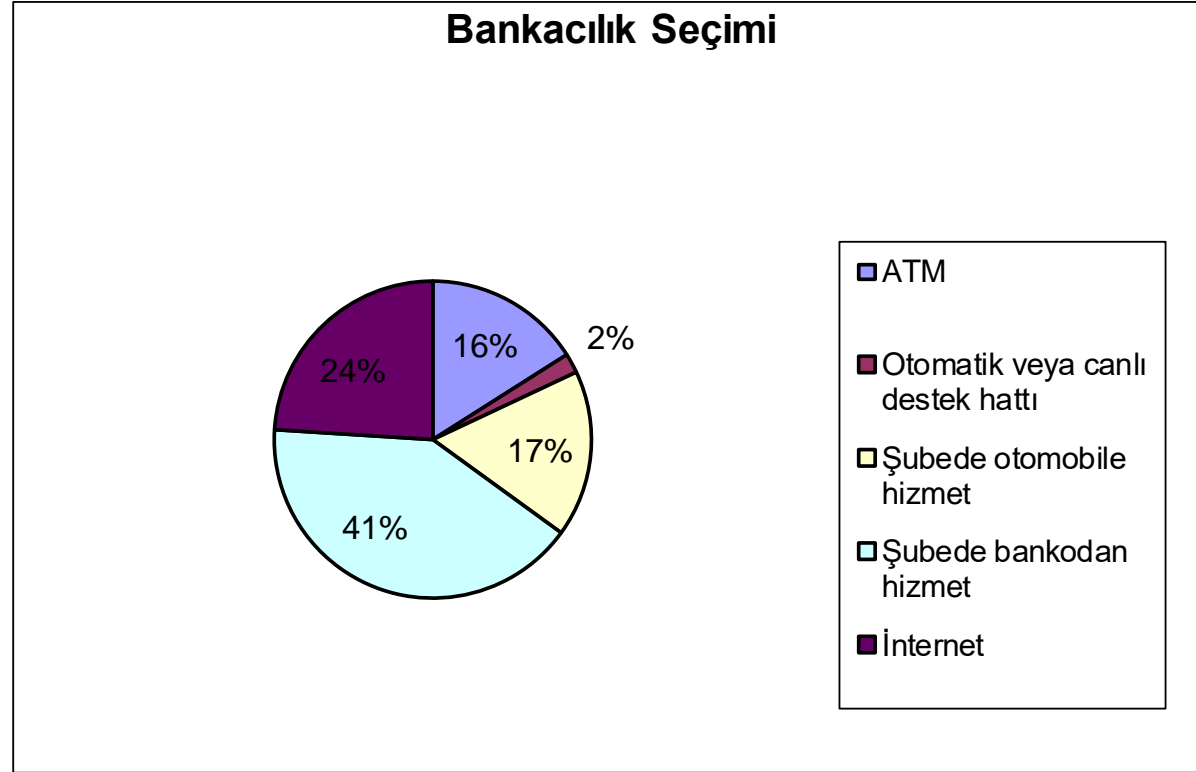


Kategorik verilerin görselleştirilmesi: Pasta diyagramı



- **Pasta diyagramı**, kategorileri temsil eden dilimlere ayrılmış bir dairedir. Pastanın her bir diliminin boyutu, her bir kategorinin yüzdesine göre değişir.

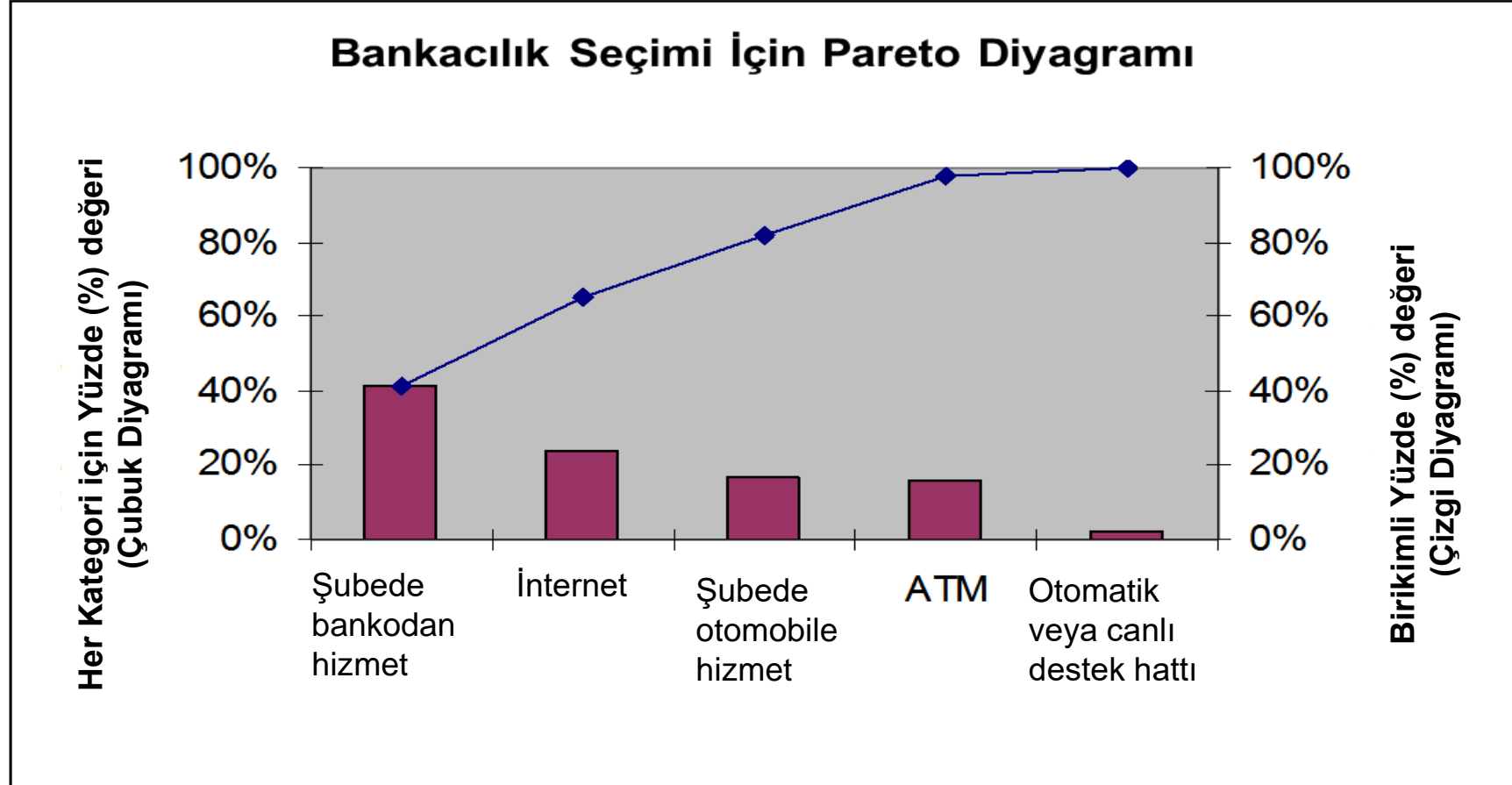
Bankacılık Seçimi?	%
ATM	16%
Otomatik veya canlı destek hattı	2%
Şubede otomobile hizmet	17%
Şubede bankodan hizmet	41%
İnternet	24%



Kategorik verilerin görselleştirilmesi: Pareto diyagramı

- Kategorik verileri tasvir etmek için kullanılır (nominal ölçek)
- Kategorilerin frekansın azalan sırasıyla gösterildiği dikey bir çubuk grafiği mevcuttur.
- Aynı grafikte kümülatif bir çokgen gösterilir.
- "Hayati az sayıdaki" 'ni "önemsiz çok" 'dan ayırmak için kullanılır.

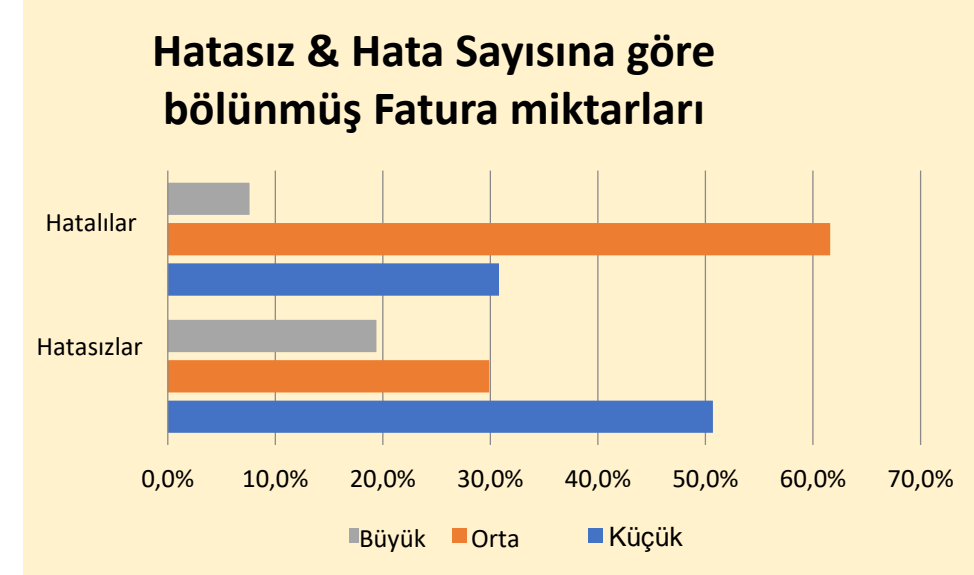
Kategorik verilerin görselleştirilmesi: Pareto diyagramı



Kategorik verilerin görselleştirilmesi: Yan yana çubuk diyagramı

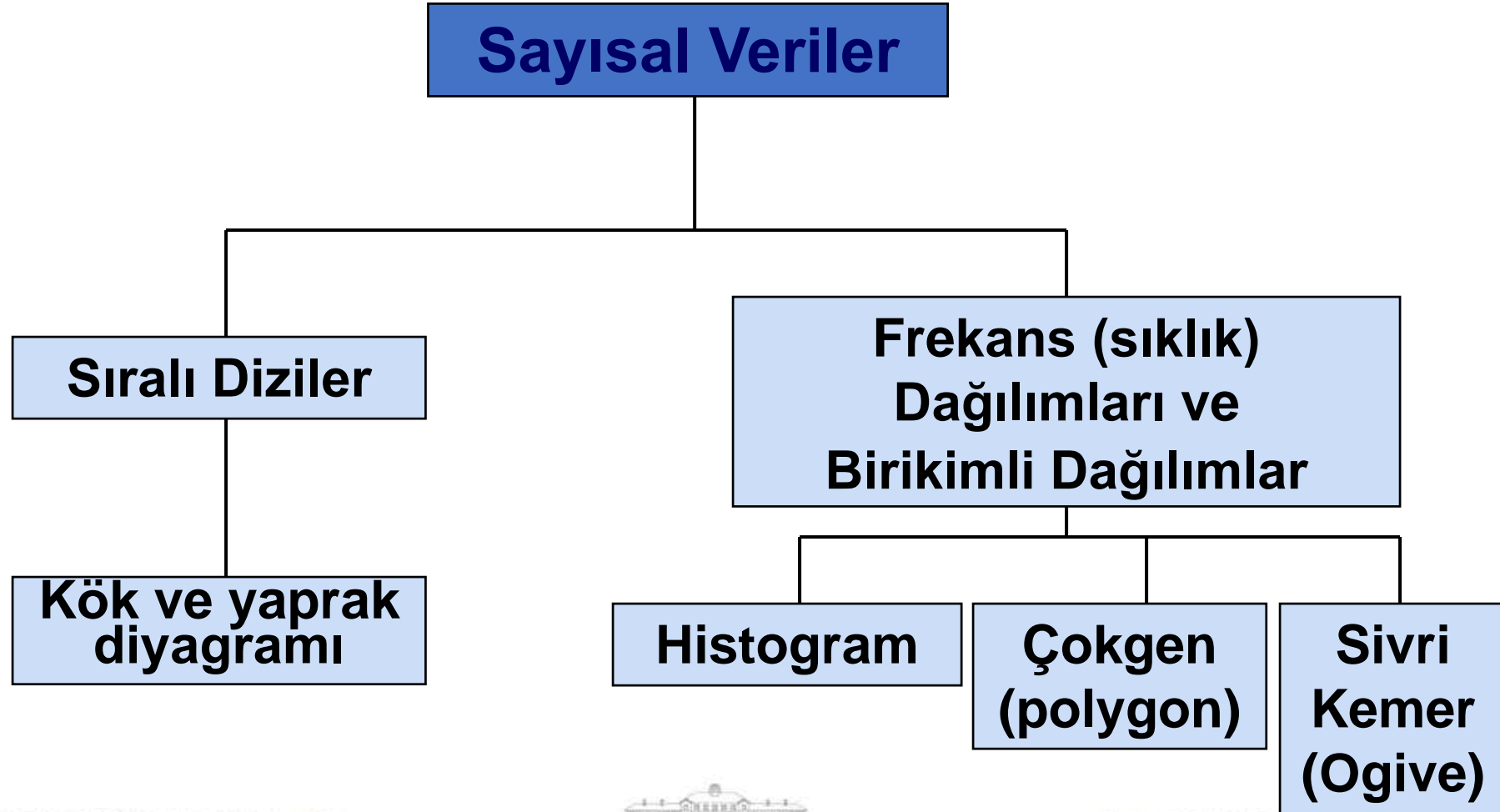
- Yan yana çubuk diyagram, bir çapraz (contingency) tablodaki verileri temsil eder.

	Hatasız	Hata Sayısı	Toplam
Küçük Miktar	50.75%	30.77%	47.50%
Orta Miktar	29.85%	61.54%	35.00%
Büyük Miktar	19.40%	7.69%	17.50%
Toplam	100.0%	100.0%	100.0%



Hatalı faturaların içerisinde olması en muhtemel olanları orta büyüklükte olanlardır.(30.77% ve 7.69%'e karşılık 61.54%)

Grafik gösterimleri kullanarak sayısal verileri görselleştirme



Sayısal verilerin düzenlenmesi: Kök ve Yaprak diyagramı

- Verilerin nasıl dağıtıldığını ve veri yoğunluğunun nerede olduğunu görmek için basit bir yoldur

YÖNTEM: Sıralanan veri serilerini
önde gelen rakamlara (**köklere**)
ve takip eden rakamlara (**yapraklara**)
ayır.

Sayısal verilerin düzenlenmesi: Kök ve Yaprak diyagramı



- **Kök-ve-yaprak diyagramı** her bir x_i gözlem değerinin en az iki haneli olduğu x_1, x_2, \dots, x_n veri setinden aydınlatıcı görsel sunum elde etmenin iyi bir yoldur. Kök-ve-yaprak diyagramı oluşturmak için aşağıdaki adımları kullanalım.

- (1) Her bir x_i sayısını iki parçaya ayıralım: bir ya da daha fazla en büyük haneden oluşan bir kök ve kalan hanelerden oluşan bir yaprak.
- (2) Yatay bir sütunda kök değerleri listelenir.
- (3) Her bir gözlem değerinin yaprak değeri ilgili kök değerinin yanına kaydedilir.
- (4) Kök ve yaprakların birimleri yazılır.

Sayısal verilerin düzenlenmesi: Kök ve Yaprak gösterimi



- Bir **kök ve yaprak gösterimi**, verileri gruplar halinde düzenler (kök olarak adlandırılır), böylece her bir gruptaki değerler (yapraklar) her bir satırda sağa doğru yayılır.

Üniversite Öğrencilerinin Yaşı

Araştırılan Üniversite öğrencilerinin Yaşı	Gündüz Öğrencileri					
	16	17	17	18	18	18
	19	19	20	20	21	22
	22	25	27	32	38	42
	Gece Öğrencileri					
	18	18	19	19	20	21
	23	28	32	33	41	45

Gündüz Öğrencileri		Gece Öğrencileri	
Kök	Yaprak	Kök	Yaprak
1	67788899	1	8899
2	0012257	2	0138
3	28	3	23
4	2	4	15

Örnek

- Örnek olması için eğer veriler bir yarı iletken devre levhaları üzerinde 0 ile 100 arasındaki yüzde kusurlu bilgisinden oluşuyorsa 76 değerini kök 7 ve yaprak 6 olarak bölebiliriz. Genel olarak gözlem sayısına göreli olarak az kök değerleri seçmeliyiz. Kök değerlerinin 7 ile 24 arasında seçilmesi önerilmektedir.

105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
160	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

Örnek

105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
160	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

Kök	Yaprak	Sıklık
7	6	1
8	7	1
9	7	1
10	5 1	2
11	5 8 0	3
12	1 0 3	3
13	4 1 3 5 3 5	6
14	2 9 5 8 3 1 6 9	8
15	4 7 1 3 4 0 8 8 6 8 0 8	12
16	3 0 7 3 0 5 0 8 7 9	10
17	8 5 4 4 1 6 2 1 0 6	10
18	0 3 6 1 4 1 0	7
19	9 6 0 9 3 4	6
20	7 1 0 8	4
21	8	1
22	1 8 9	3
23	7	1
24	5	1

Örnek

- Diyagramdaki son sütun her bir kök değeriyle ilişkili yaprak sayısının adet (frekans) değerini vermektedir. Diyagram incelendiğinde basınç mukavemetinin çok büyük bir kısmının 110 ve 200 psi aralığında olduğunu ve merkezin 150 ile 160 psi arasında bir yerde olduğu görülmektedir.
- Ayrıca mukavemet değerlerinin merkez etrafında yaklaşık olarak simetrik dağıldığı görülmektedir.
- Kök-ve-yaprak diyagramı orijinal veri tablosunda hemen göremeyeceğimiz verilerle ilgili bazı önemli bilgileri elde etmemize imkan sağlamaktadır.

Örnek

- Yandaki şekil basınç mukavemeti verileri için minitab programının sıralı kök-ve-yaprak diyagramını vermektedir.
- Paket programın her bir kök için yaprakları küçükten büyüğe sıralı olarak oluşturduğuna dikkat ediniz. Buna **sıralı kök-ve-yaprak diyagramı** da denir.
- Bilgisayar kök değerlerinin soluna o kökteki ve ondan önceki köklerdeki yaprakların sayısını da koymaktadır. Ortadaki 16. kök değerinde sütun bu kökteki gözlem değerini vermektedir.

1	7	6
2	8	7
3	9	7
5	10	15
8	11	058
11	12	013
17	13	133455
25	14	12356899
37	15	001344678888
(10)	16	0003357789
33	17	0112445668
23	18	0011346
16	19	034699
10	20	0178
6	21	8
5	22	189
2	23	7
1	24	5

Sayısal verilerin görselleştirilmesi: Histogram

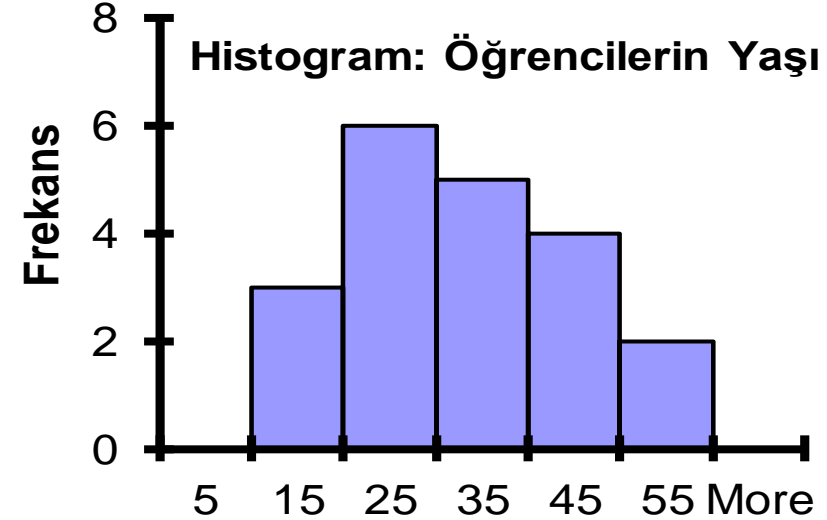
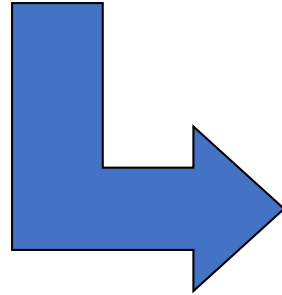
- **Histogram** frekans dağılımının görsel bir sunumudur.
- Bir frekans dağılımındaki verinin dikey bir çubuk diyagramı ile gösterimine denir.
- Bir histogramda bitişik çubuklar arasında boşluk yoktur.
- **Sınıf sınırları** (veya **sınıf orta noktaları**) yatay eksenle gösterilir.
- Dikey eksen, **frekans (sıklık)**, **göreceli frekans (göreceli sıklık)** veya **yüzde** değeridir.
- Çubukların yüksekliği frekansı (sıklığı) , göreceli frekansı (göreceli sıklığı) veya yüzdeyi temsil eder.
- Bir histogram aşağıdaki adımları takip ederek oluşturulabilir.

- (1) Her bir sınıf aralığının sınır değerlerini yatay eksenle gösteriniz.
- (2) Yatay eksenle sıklık ve göreceli sıklık değerlerini gösteriniz.
- (3) Her bir sınıf aralığının üzerinde yüksekliği ilgili sınıf aralığının sıklık (ya da göreceli sıklık) değerine eşit bir dikdörtgen çizin.

Sayısal verilerin görselleştirilmesi: Histogram

Sınıf	Frekans	Bağıl Frekans	Yüzdelik
10 ile 20 arasında	3	.15	15%
20 ile 30 arasında	6	.30	30%
30 ile 40 arasında	5	.25	25%
40 ile 50 arasında	4	.20	20%
50 ile 60 arasında	2	.10	10%
Toplam	20	1.00	100%

(Yüzde histogramında, sınıf başına gözlem yüzdesini göstermek için dikey eksen tanımlanmaktadır)



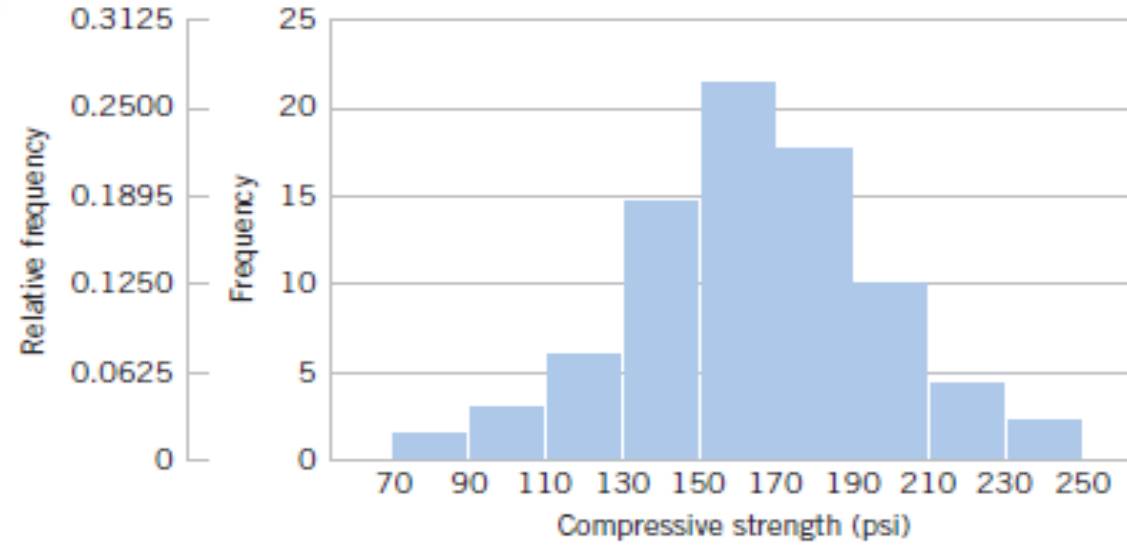
Örnek

- 80 adet alüminyum-lityum alaşım numunesine ait basınç mukavemeti verileri.

105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
160	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

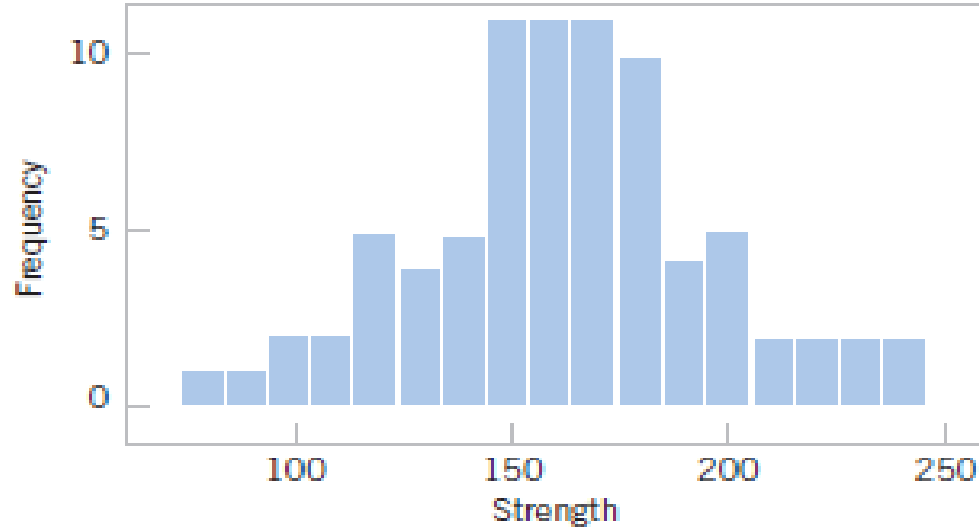
- Sonraki yansındaki şekil basınç mukavemeti verilerine ait bir histogramdır. Ölçümlerin dağılımının şekline ait bilgi vermekte ve bununla birlikte merkezi eğilim ve verilerdeki değişkenlik hakkında bilgi sağlamaktadır.

Örnek



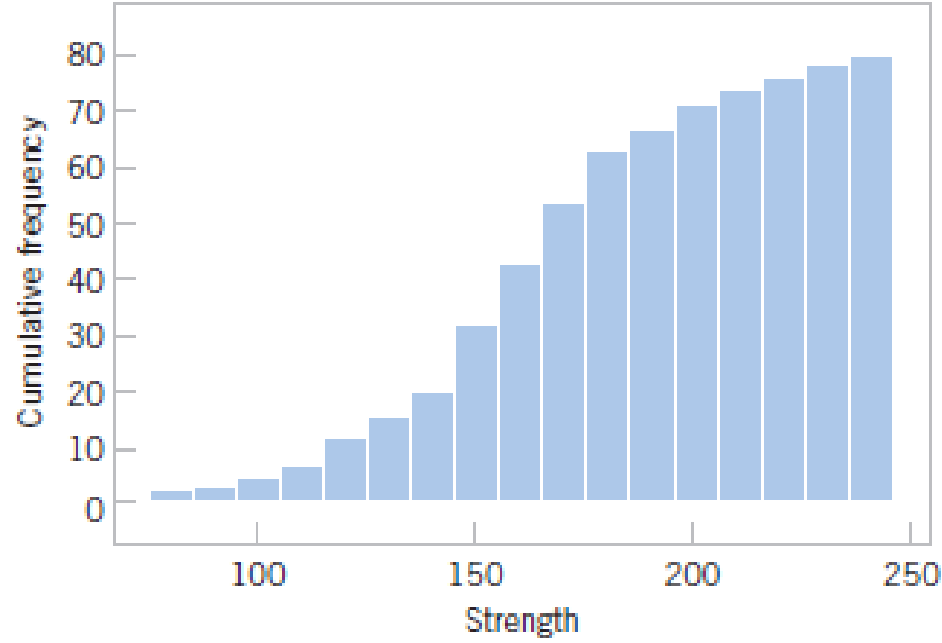
- Yukarıdaki şekilden mukavemet değerlerine ait ölçümlerin dağılımının simetrik ve çan şeklinde olduğu görülmektedir.
- Bu gösterim genellikle popülasyon için bir model olarak kullanılacak muhtemel olasılık dağılımlarının seçimi hakkında ipucu vermektedir. Örneğin normal dağılımın basınç mukavemeti verileri için uygun bir model olacağı sonucuna varabiliriz.
- (9 aralık mevcut)

Örnek



- Yukarıdaki aynı veriler için 17 sınıf aralığı kullanarak oluşturulmuş histogramı göstermektedir.
- Histogramların sınıf aralığı sayısına ve genişliğine hassas olduğunu görmekteyiz.
- Küçük veri setleri için eğer sınıf sayısı ve/veya sınıf genişliği değiştiğinde histogramların görünümü dramatik şekilde değişebilir.
- Histogramlar 75 ile 100 veya daha fazla gözlem içeren veri setleri için daha durağan yapıdadır.

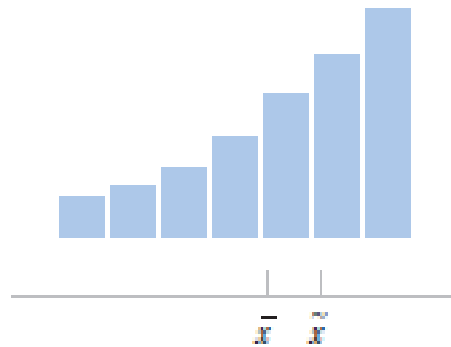
Örnek



- Yukarıdaki şekil, histogramdaki değişkenliği gösteren **birikimli sıklık grafiğidir**.
- Birikimli dağılımlar verilerin yorumlanmasında oldukça faydalıdır.
- Şekilden yaklaşık olarak verilerin %70'inin 200 psi değerinden küçük veya eşit olduğunu görebilmekteyiz.

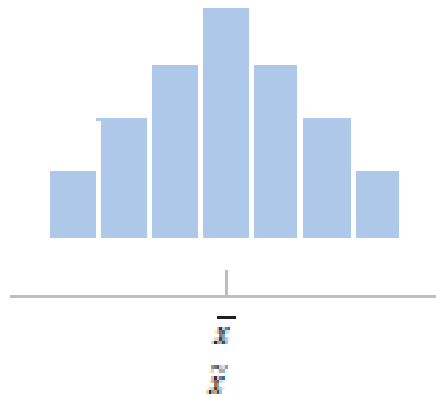
Sayısal verilerin görselleştirilmesi: Histogram

- Örneklem hacmi genişlediğinde histogram örneklemin seçildiği popülasyonun veya dağılımın şekli ile ilgili genel şekli hakkında güvenilir bir fikir verebilir.
- Aşağıdaki şekilde üç durum gösterilmektedir. Medyan (ortanca) \tilde{x} ile gösterilmektedir. Genellikle veriler şekil b'de olduğu gibi simetrikse ortalama ve medyan çakışmaktadır.



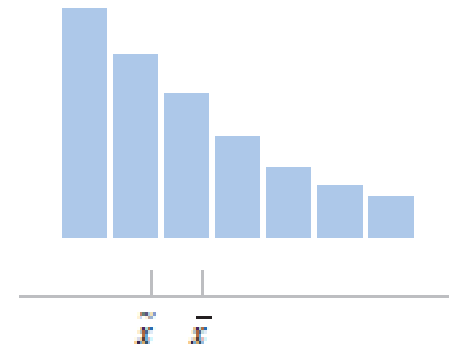
Negatif veya sola çarpık

(a)



Simetrik

(b)

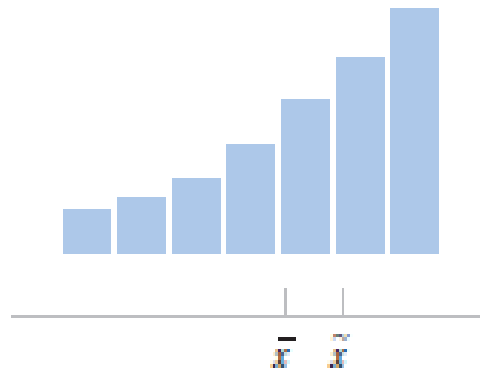


Pozitif veya sağa çarpık

(c)

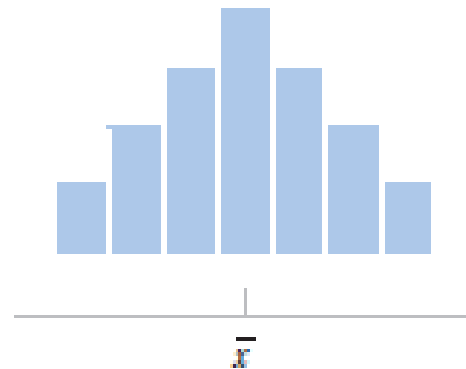
Sayısal verilerin görselleştirilmesi: Histogram

- Eğer ek olarak veri yalnızca bir mod (tepe) değerine sahipse (bu durumda veriler unimodal olarak isimlendirilir) ortalama, medyan ve mod değerleri çakışır.
- Eğer veriler şekil a ve c’de olduğu gibi *çarpıksa* (asimetrik ya da bir uca doğru uzun bir kuyruk) ortalama, medyan ve mod çakışmaz.
- Genellikle dağılım sağa çarpıksa $\text{mod} < \text{medyan} < \text{ortalama}$ olduğu görülür. Ancak dağılım sola çarpıksa genellikle $\text{mod} > \text{medyan} > \text{ortalama}$ olduğu görülür.



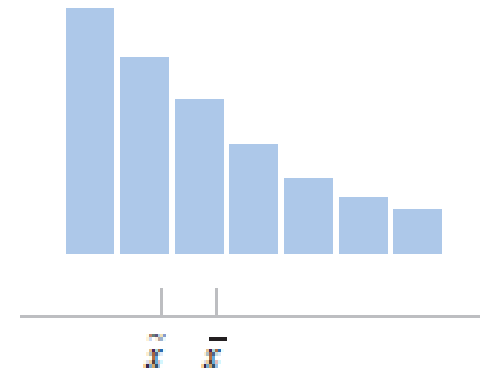
Negatif veya sola çarpık

(a)



Simetrik

(b)



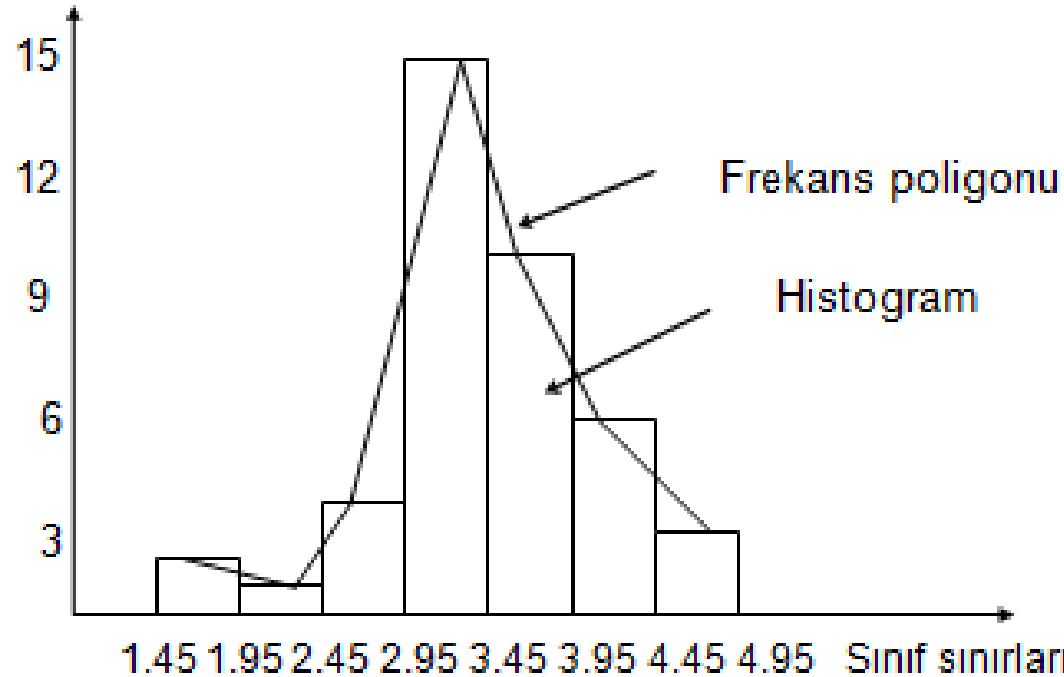
Pozitif veya sağa çarpık

(c)

Histogram ve frekans poligonları

- Histogramdaki dikdörtgenlerin orta noktalarının birleştirilmesiyle frekans poligonu elde edilir. Poligondaki kırık çizgilerin kaldırılmasıyla (bunun için gözlem sayısının yeterince artırılması gerekir) verilerin dağılım biçimi kabaca belirlenebilir.

Frekans



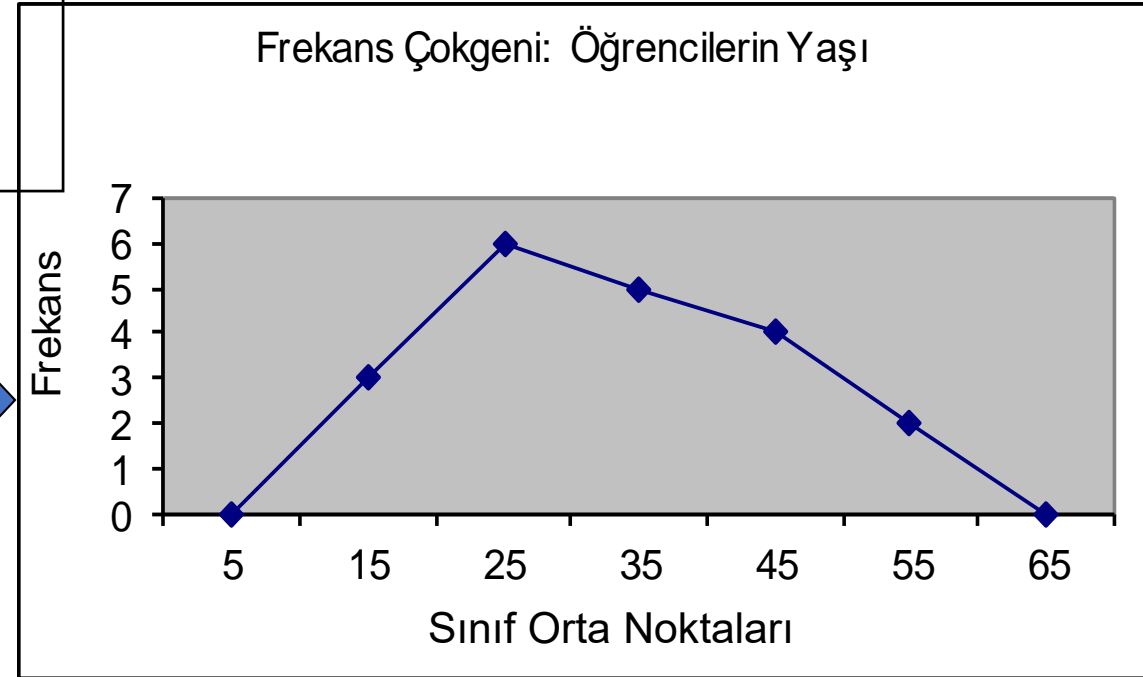
Sayısal verilerin görselleştirilmesi: Çokgen (Poligon)

- Bir **yüzde çokgen**, her sınıftaki veriyi temsil eden orta noktanın elde edilmesi ve bu şekilde orta noktaların ilgili sınıf yüzdelerine göre elde edilen sırasına göre bağlanmasıyla oluşturulur.
- **Birikimli yüzde çokgeni** veya **sivri kemer (ogive)**, X eksenini boyunca ilgilenilen değişkeni ve Y eksenini boyunca birikimli yüzdeleri görüntüler.
- Karşılaştırılacak iki veya daha fazla grup varsa kullanışlı bir gösterim yöntemidir.

Sayısal verilerin görselleştirilmesi: Frekans çokgeni (Polygon)

Sınıf	Sınıf Orta noktası	Frekans (Sıklık)
10 ile 20 arasında	15	3
20 ile 30 arasında	25	6
30 ile 40 arasında	35	5
40 ile 50 arasında	45	4
50 ile 60 arasında	55	2

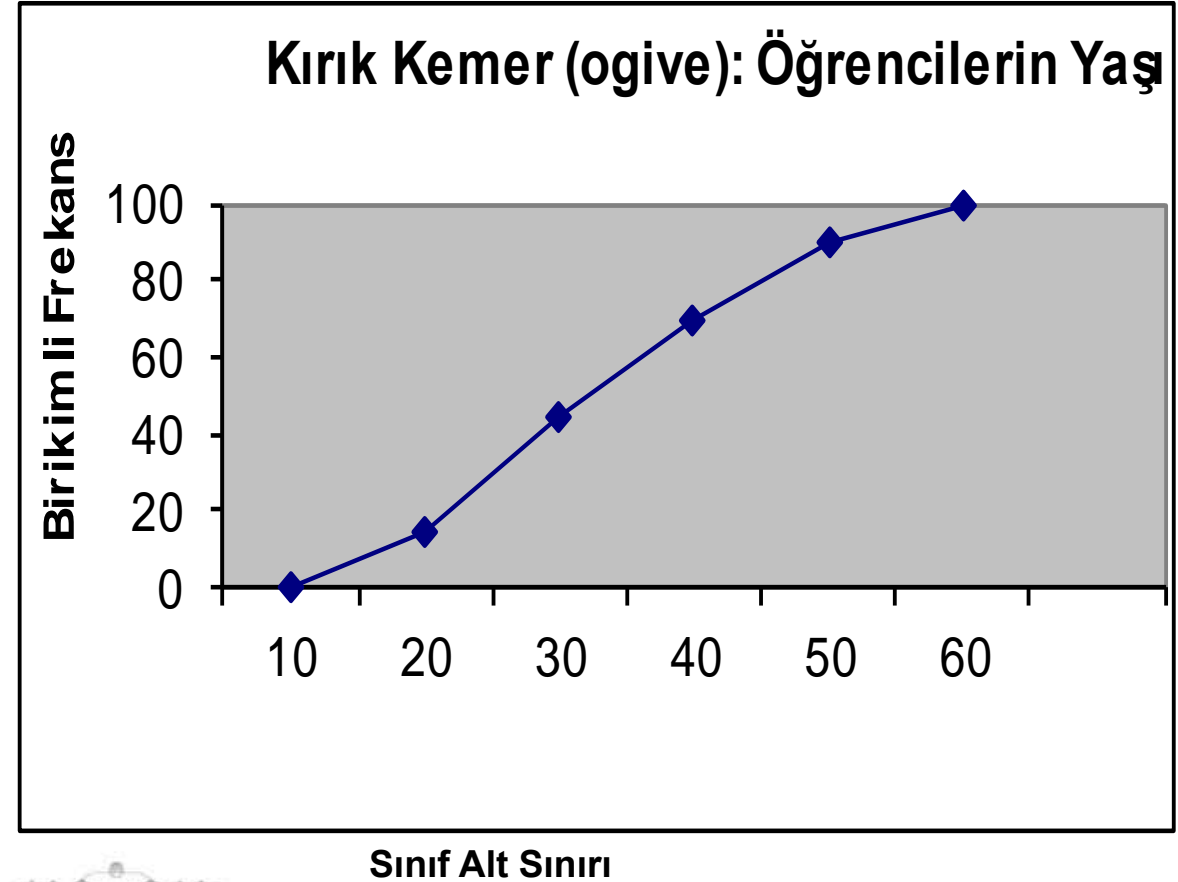
(Yüzde çokgende (poligonda), dikey eksen her bir sınıf için gözlem yüzdesini gösterecek şekilde tanımlanacaktır)



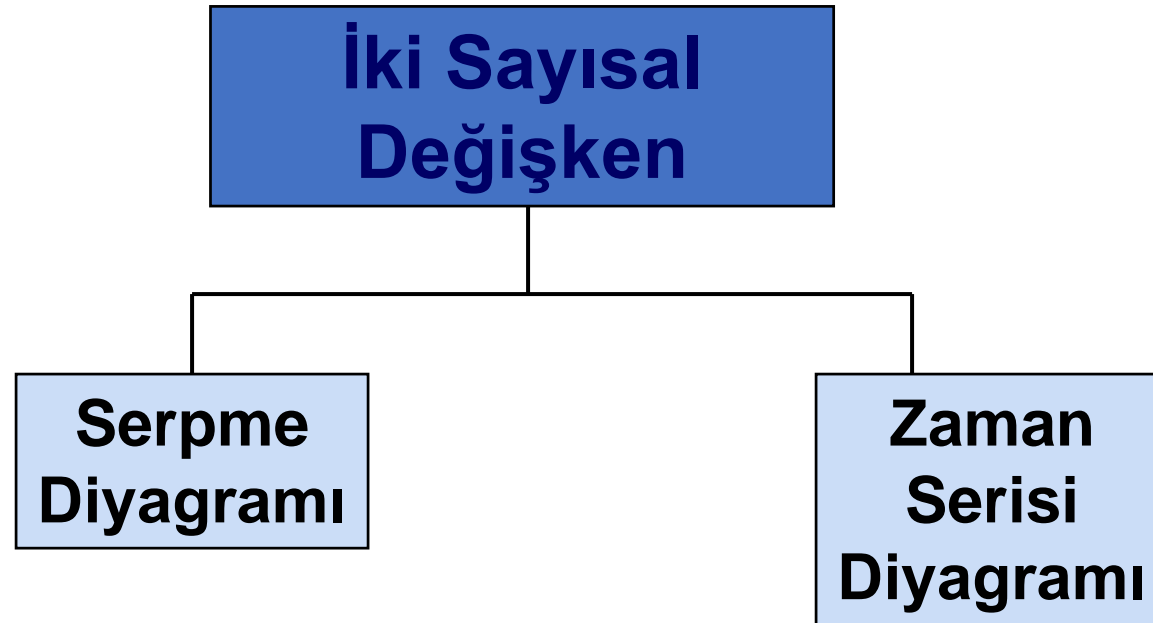
Sayısal verilerin görselleştirilmesi: Kırık kemer (Ogive) (birikimli % çokgen)

Sınıf	Sınıfın alt sınır değeri	Alt sınır değerinden düşük %'si
10 ile 20 arasında	10	15
20 ile 30 arasında	20	45
30 ile 40 arasında	30	70
40 ile 50 arasında	40	90
50 ile 60 arasında	50	100

(Bir kırık kemerde (ogive), her alt sınıf sınırından daha düşük yüzdeye sahip gözlemler, sınıf alt sınırlarına karşılık çizilir.



Grafik göstergeler kullanarak iki sayısal değişkeni görselleştirme

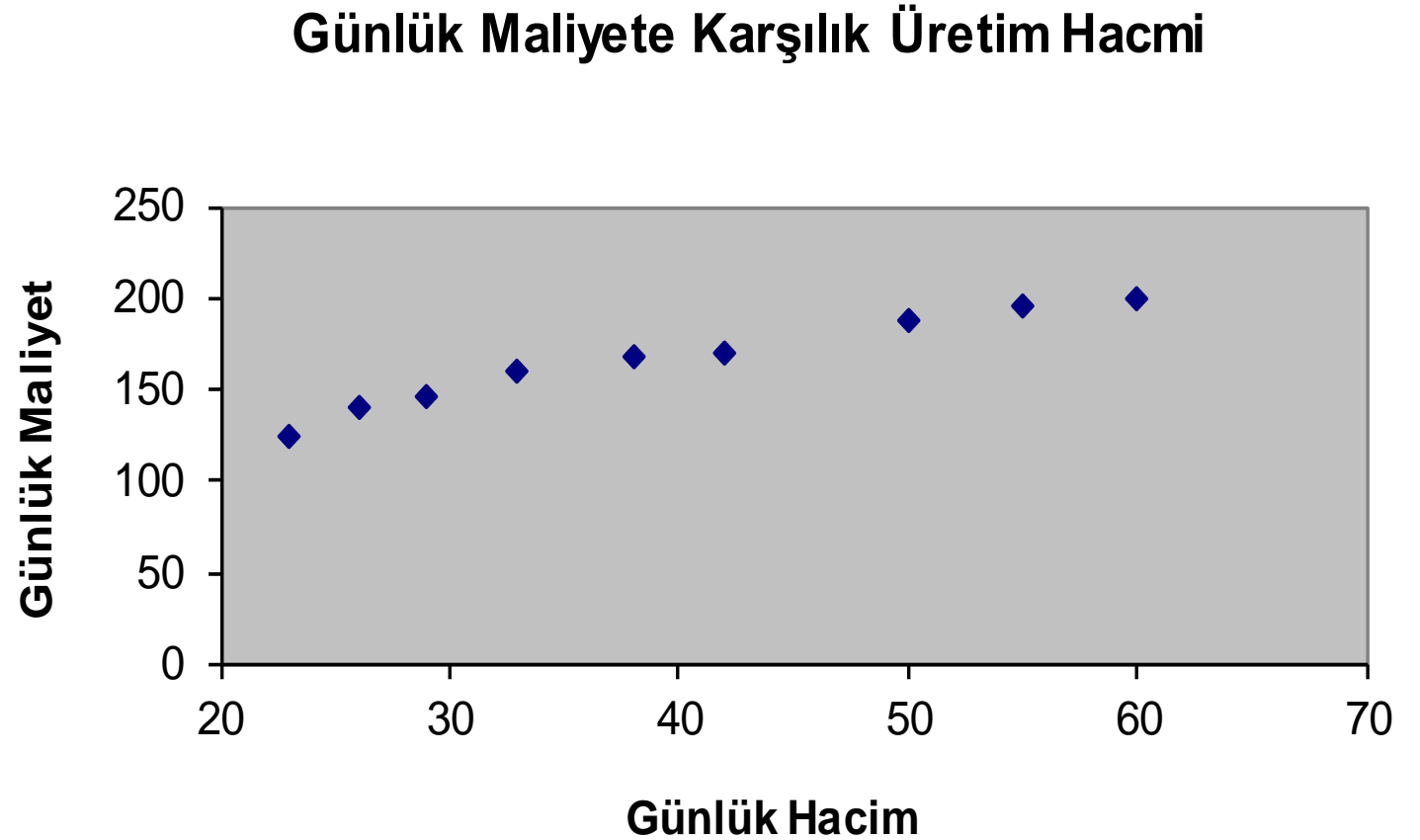


İki sayısal değişkenin görselleştirilmesi: Serpme diyagramı

- **Serpme diyagramları**, iki sayısal değişkenden alınan eşleştirilmiş gözlemlerden oluşan sayısal veriler için kullanılır.
- Bir değişken dikey eksende ölçülür ve diğer değişken yatay eksen üzerinde ölçülür.
- Serpme diyagramları, iki sayısal değişken arasındaki muhtemel ilişkileri incelemek için kullanılır.

Serpme diyagramı örneđi

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



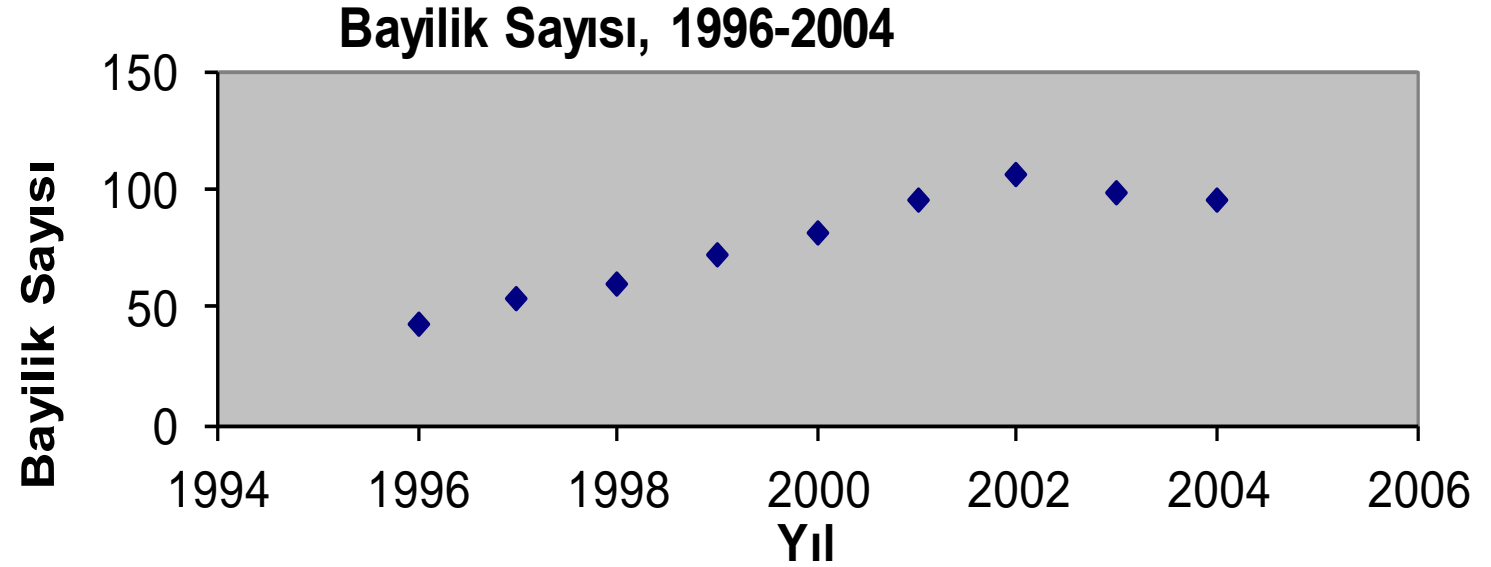
İki sayısal değişkenin görselleştirilmesi: Zaman serisi diyagramı

- **Zaman serisi diyagramı**, sayısal bir değişkenin zaman içindeki değerlerindeki desenleri incelemek için kullanılır
- Zaman Serisi Diyagramı:
 - Sayısal değişken dikey eksen ve zaman periyodu yatay eksen üzerinde ölçülür

Zaman serisi diyagramı örneđi



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



Görselleştirmenin geliştirilmesi için öneriler

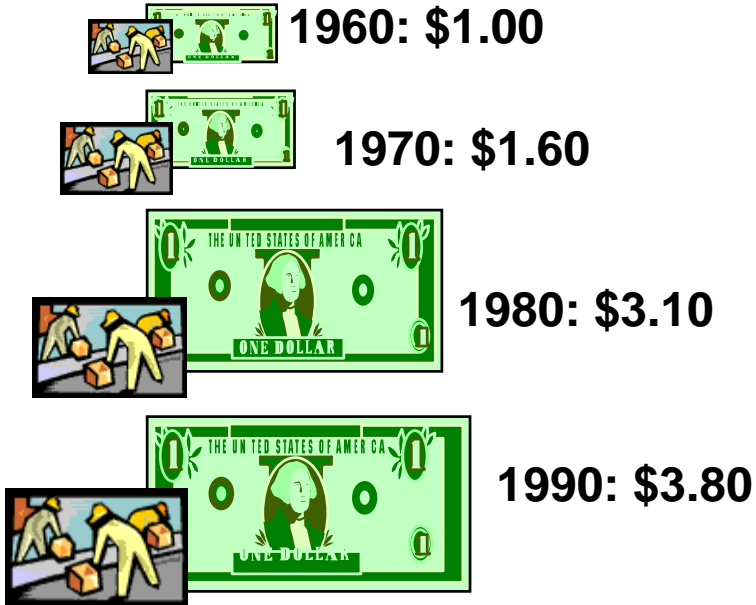
- Gereksiz çizimlerden kaçının
- Mümkün olan en basit görselleştirmeyi kullanın
- Bir başlık ekleyin
- Tüm eksenleri etiketleyin
- Grafik eksenler içeriyorsa, her bir eksen için bir ölçek ekleyin
- Dikey bir eksen için ölçeği sıfırdan başlatın
- Sabit bir ölçek kullanın

Grafik hataları: Gereksiz çizimler



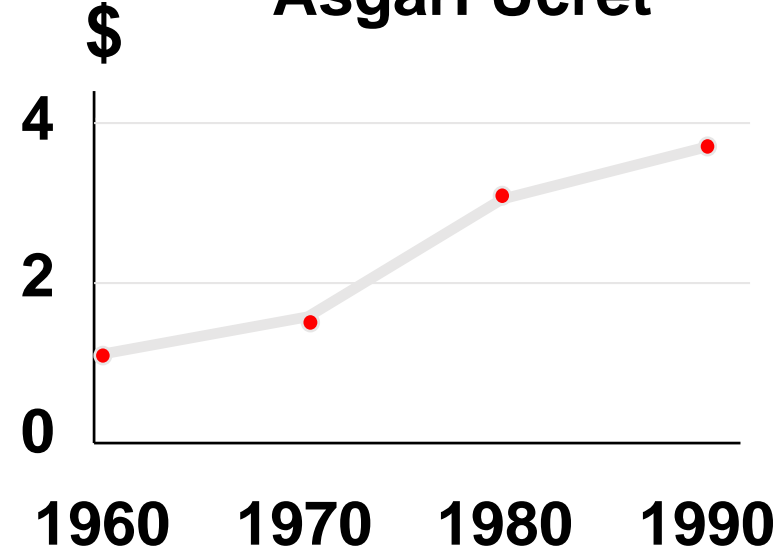
Kötü Gösterim

Asgari Ücret



İyi Gösterim

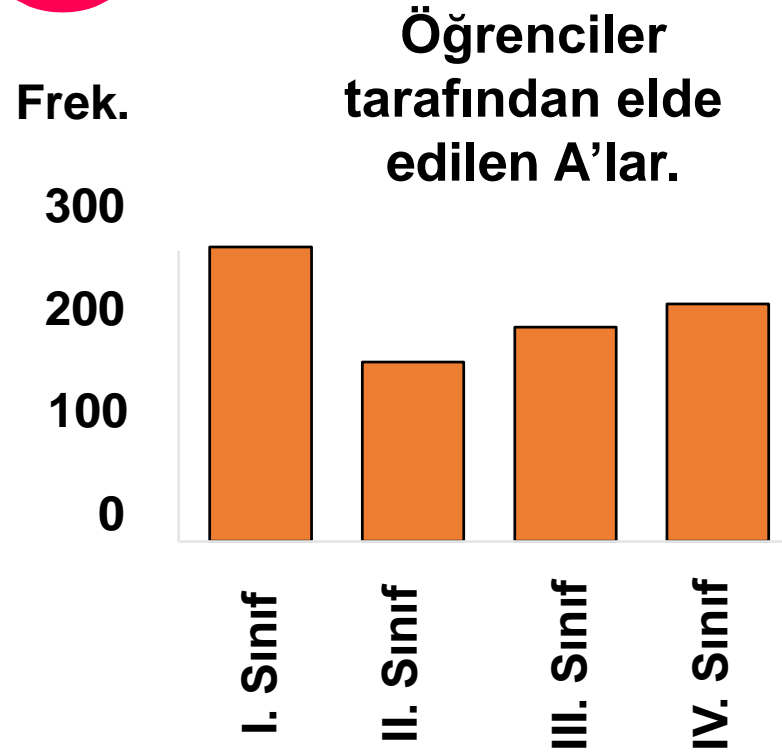
Asgari Ücret



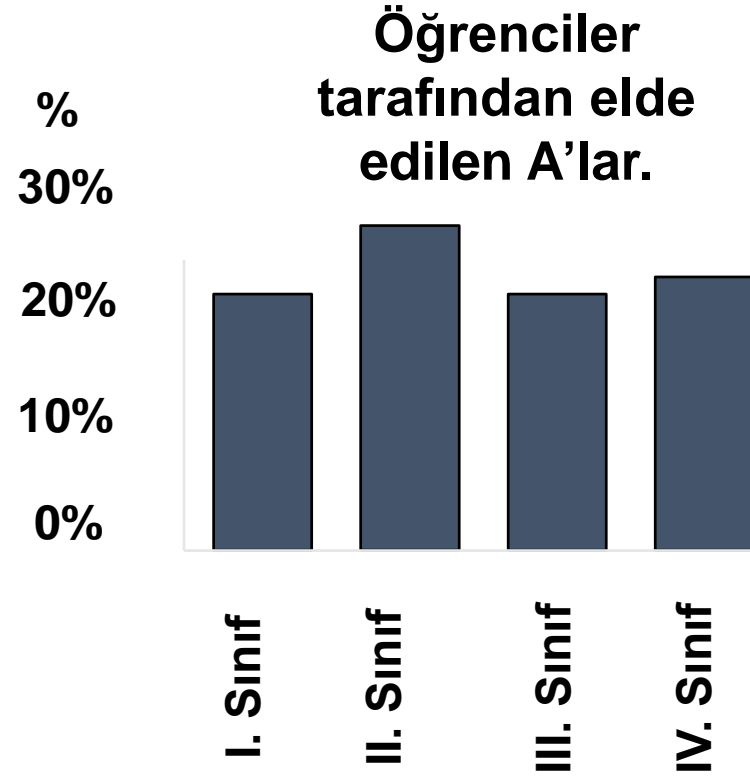
Grafik hataları: İlişkili matrahın olmaması



Kötü Gösterim



İyi Gösterim

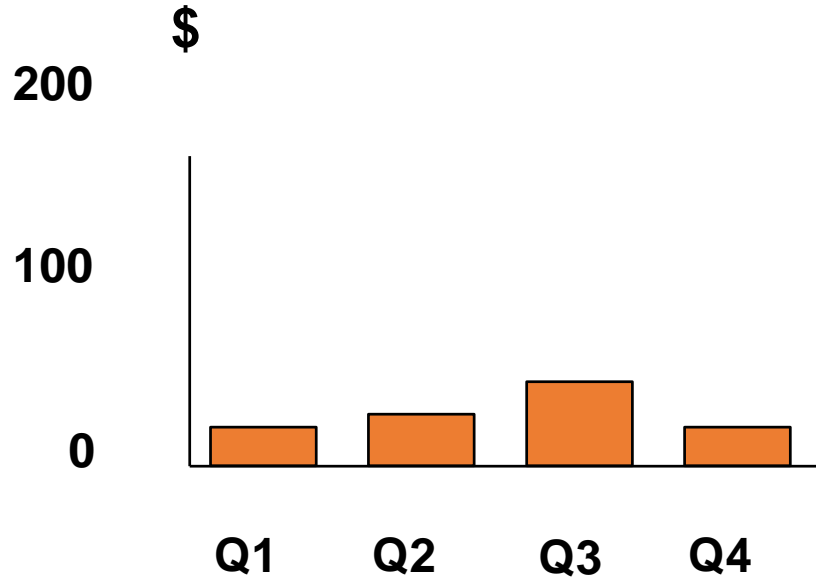


Grafik hataları: Dikey eksenin basıklaştırılması



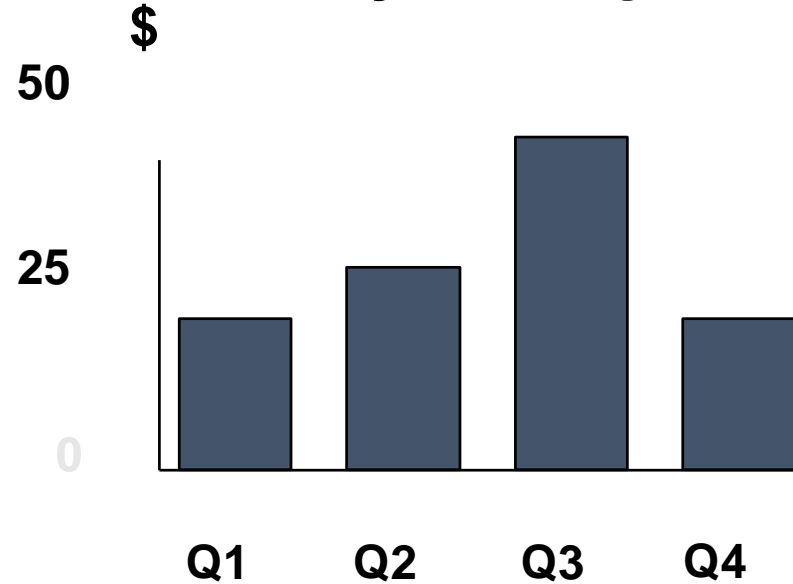
Kötü Gösterim

3 Aylık Satışlar



İyi Gösterim

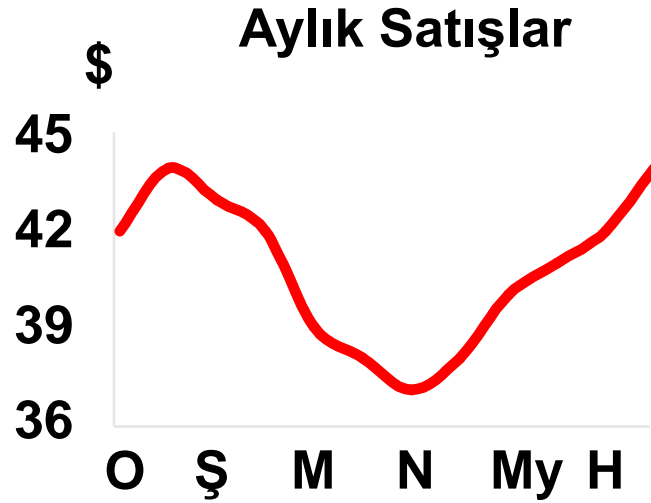
3 Aylık Satışlar



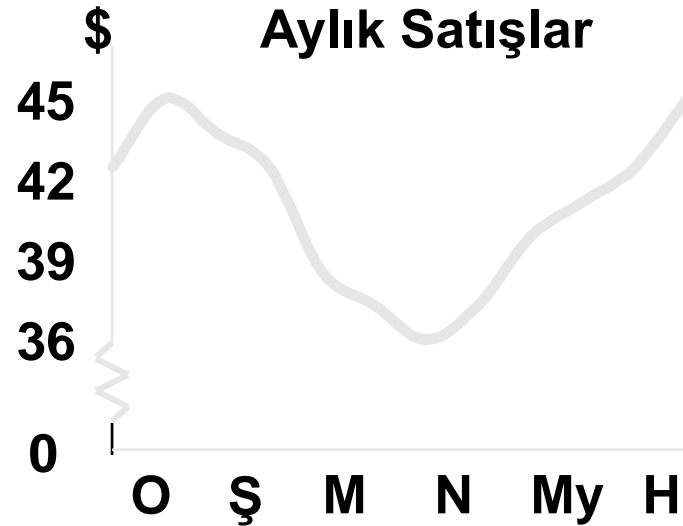
Grafik hataları: Dikey ekseninde sıfır noktasının olmaması



Kötü Gösterim



✓ İyi Gösterim



İlk altı ayın satış verilerini
görselleştirmektedir.

Excel'de yapılan hatalar

- Excel genellikle dikey eksenini 0'da başlamayan bir grafik oluşturur
- Excel, basit grafikleri 3-B diyagramlarına çevirme fırsatı sunar ve süreç içinde bozuk görüntüler oluşturabilir
- Excel ile seçenekler olarak sunulan alışılmadık diyagramlar çoğunlukla çarpık diyagramlar oluşturur

Excel pivot tablolarını kullanarak deęişkenleri organize etmek & görselleştirmek

Bir Pivot tablo:

- Deęişkenleri çok boyutlu bir özet tablosu olarak özetler
- Deęişkenlerin özetlenmesi ve biçimlendirilmesi düzeyinin interaktif olarak deęiştirilmesine imkan verir
- Belirlenen ölçütlere uyan veri alt bölümlerini özetlemek için verilerinizi etkileşimli olarak "dilimlemenize" izin verir.
- Daha basit tabloların ve diyagramların, çok boyutlu verilerdeki muhtemel kalıpları ve ilişkileri belirginleştirmekte başarısız olduęu durumlarda ilişkiyi ve kalıpları keşfetmek için kullanılabilir.

Emeklilik fonları verisi için iki değişkenli bir çapraz (contingency) tablo

Ortalama riskin, düşük veya yüksek riske göre çok daha fazla büyüme payı vardır

	A	B	C	D	E
1	Contingency Table of Fund Type and Risk				
2					
3		RISK ▼			
4	TYPE ▼	Low	Average	High	Grand Total
5	⊕ Growth	19.50%	35.53%	15.09%	70.13%
6	⊕ Value	11.64%	10.06%	8.18%	29.87%
7	Grand Total	31.13%	45.60%	23.27%	100.00%

Üç veya daha fazla kategorik değişkenin tepkilerini uyumlaştıran bir çok boyutlu çapraz tablo

	A	B	C	D	E
1	Contingency Table of Fund Type, Market Cap, and Risk				
2					
3		RISK ▼			
4	TYPE ▼	Low	Average	High	Grand Total
5	☐ Growth	19.50%	35.53%	15.09%	70.13%
6	Large	15.09%	14.78%	2.52%	32.39%
7	Mid-Cap	3.77%	13.84%	3.14%	20.75%
8	Small	0.63%	6.92%	9.43%	16.98%
9	☐ Value	11.64%	10.06%	8.18%	29.87%
10	Large	9.43%	7.86%	0.00%	17.30%
11	Mid-Cap	1.57%	1.57%	2.83%	5.97%
12	Small	0.63%	0.63%	5.35%	6.60%
13	Grand Total	31.13%	45.60%	23.27%	100.00%

•Büyüme fonlarının pazara bağlı risk ilişkisi örneği

•Değer fonu risk modeli büyüme fonlarınınkinden farklıdır.

Çok boyutlu çapraz (contingency) tablolar sayısal veriler içerebilir

Bu tablo, piyasa değeri birden düşmüş veya gözden saklanmış 10 yıllık ortalama getiriği göstermektedir

	A	B	C	D	E
1	Contingency Table of Fund Type, Market Cap, and Risk				
2					
3	Average of 10YrReturn%	RISK			
4	TYPE	Low	Average	High	Grand Total
5	+ Growth	4.12	5.07	4.72	4.73
6	+ Value	5.14	4.71	6.87	5.47
7	Grand Total	4.50	4.99	5.48	4.95

Bu risk seviyeleri ile düşük veya yüksek riskli değer fonları, büyüme fonlarına göre ortalama 10 yıl daha yüksek getiri elde ederler.

Aynı tablo genişletilmiş piyasa değeri ile daha karmaşık bir model göstermektedir

	A	B	C	D	E
1	Contingency Table of Fund Type, Market Cap, and Risk				
2					
3	Average of 10YrReturn%	RISK			
4	TYPE	Low	Average	High	Grand Total
5	<input checked="" type="checkbox"/> Growth	4.12	5.07	4.72	4.73
6	Large	3.69	3.65	1.26	3.48
7	Mid-Cap	5.62	6.04	5.77	5.92
8	Small	5.38	6.15	5.30	5.65
9	<input checked="" type="checkbox"/> Value	5.14	4.71	6.87	5.47
10	Large	4.52	4.13		4.34
11	Mid-Cap	6.62	6.27	5.52	6.01
12	Small	10.77	8.12	7.58	7.94
13	Grand Total	4.50	4.99	5.48	4.95

Geniş piyasa sermayelendirmesine sahip büyüme fonları en yoksul oyuncular olup büyüme fonu kategorisi için ortalamayı bastırmaktadır.

Bir hücreyi çift-tıklayarak altta yatan veriye odaklanılır & verinin gösterimi sağlanır

"Değer fonu ve yüksek risk" ortak yanıtının toplandığı hücrede çift tıklandığında, bu ölçütleri karşılayan tüm fonların ayrıntılarının görüntülendiği yeni bir çalışma sayfası oluşturulur

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Fund Number	Market Cap	Type	Assets	Turnover Ratio	Beta	SD	Risk	1YrReturn%	3YrReturn%	5YrReturn%	10YrReturn%	Expense Ratio	Star Rating
2	RF318	Small	Value	83.6	124	0.85	23.62	High	-3.77	19.06	2.16	9.13	1.6	Three
3	RF316	Small	Value	9	131	0.85	25.14	High	-1.34	20.13	-0.94	6.45	1.87	Two
4	RF315	Small	Value	22.3	127	0.68	24.86	High	4.63	20.9	-0.92	4.44	1.96	Five
5	RF314	Small	Value	1698.9	123	0.95	23.68	High	5.64	21.74	1.35	8.43	1.26	Three
23	RF229	Mid-Cap	Value	23.8	123	0.85	23.62	High	2.10	19.06	2.16	9.13	1.26	Three
24	RF228	Mid-Cap	Value	5926.3	95	1.38	25.91	High	-3.26	25.33	-1.41	6.41	0.6	Five
25	RF227	Mid-Cap	Value	1352.3	38	1.44	28.42	High	0.57	29.83	4.82	10.09	1.29	Four
26	RF226	Mid-Cap	Value	28	381	1.57	32.05	High	0.44	30.04	-2.87	2.03	1.54	Five
27	RF225	Mid-Cap	Value	18.3	26	1.46	28.97	High	0.81	35.01	2.73	4.46	2.07	Three

Pivot tablolar, dilimleyiciler & iş analitiği

- Pek çok analitik süreç birçok değişkenle başlar ve verileri filtreleme aracıyla keşfetmenize izin verir.
- Excel'de dilimleyicileri kullanmak bu filtreleme işlemi taklit etmenin bir yoludur.
- Dilimleyiciler, bir Pivot Tablosu ile ilişkili tüm değişkenleri filtrelemek için kullanılabilir.
- Dilimleyici panellerindeki düğmeleri tıklayarak verileri alt bölümlere ayırıp filtreleyebilir ve soruların yanıtlarını görsel olarak görebilirsiniz.

Bölüm özeti

Bu bölümde:

- Kategorik veriler için tablolar ve çizelgeler oluşturduk.
- Sayısal veriler için tablolar ve çizelgeler oluşturduk.
- Grafikleri düzgün bir şekilde sunmanın ilkelerini inceledik.
- Excel'de birçok değişkeni organize etmek ve analiz etmek için yöntemleri inceledik.