



İSTATİSTİK

2. Tanımlayıcı İstatistikler

Tanımlayıcı İstatistikler ve Grafikle Gösterim

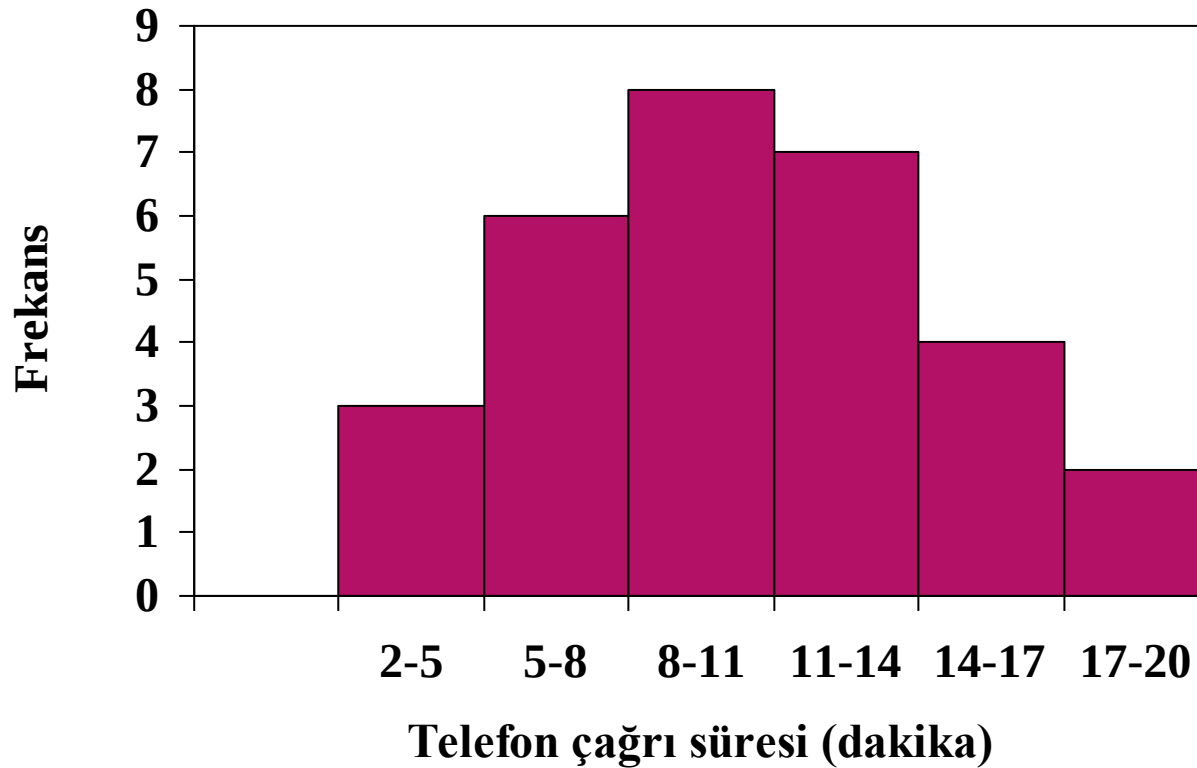
- ▶ Grafik ve bir ölçüde tablolar değişkenlerin görsel bir özetini verirler.
- ▶ İdeal olarak burada değişkenlerin merkezi (ortalama) değerlerinin ve değişkenlik (yayılma) durumu ve miktarının görülmesi gerekir.
- ▶ Gösterilen verinin ölçüm seviyesi kullanılabilecek grafik ve tablonun türü konusunda bir kısıt oluşturur

Frekans Dağılımları...

Sınıf	Aralık	Frekans
1	2 .. 5	3
2	5 .. 8	6
3	8 .. 11	8
4	11 .. 14	7
5	14 .. 17	4
6	17.. 20	2
	Toplam	30

Histogram

- Frekans tablosunun grafik gösterimidir

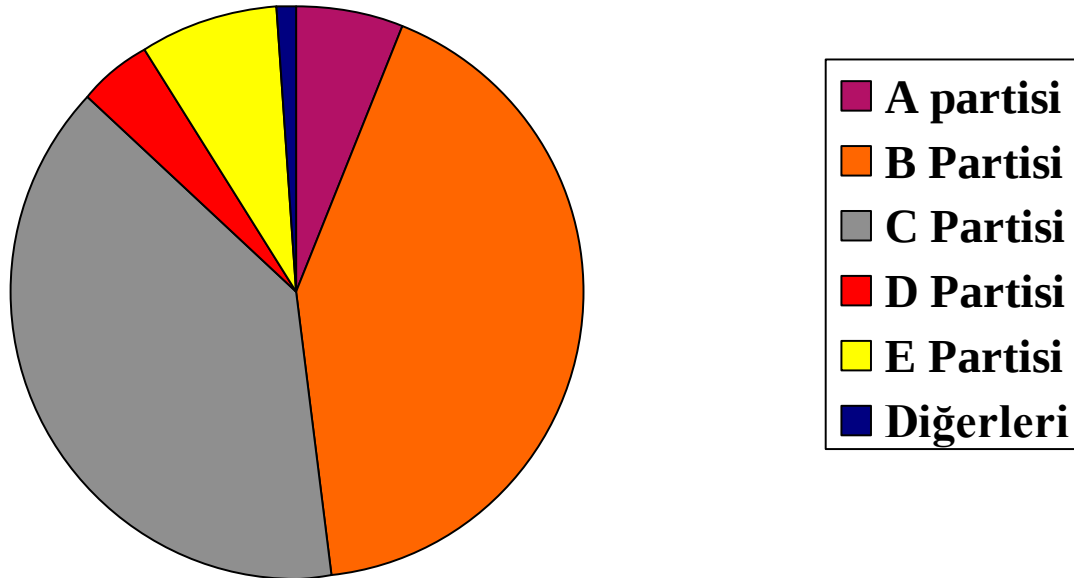


Pasta Grafiği

- ▶ Her dilimin bir kategoriye temsil ettiği, dilimlere ayrılmış bir dairedir.
- ▶ Her dilimin alanı o kategoriye karşılık gelen veri oranını temsil eder
- ▶ Amacımız her kategorinin görece büyüklüğünü göstermek olduğu durumlarda kullanılır

Pasta Grafiđi

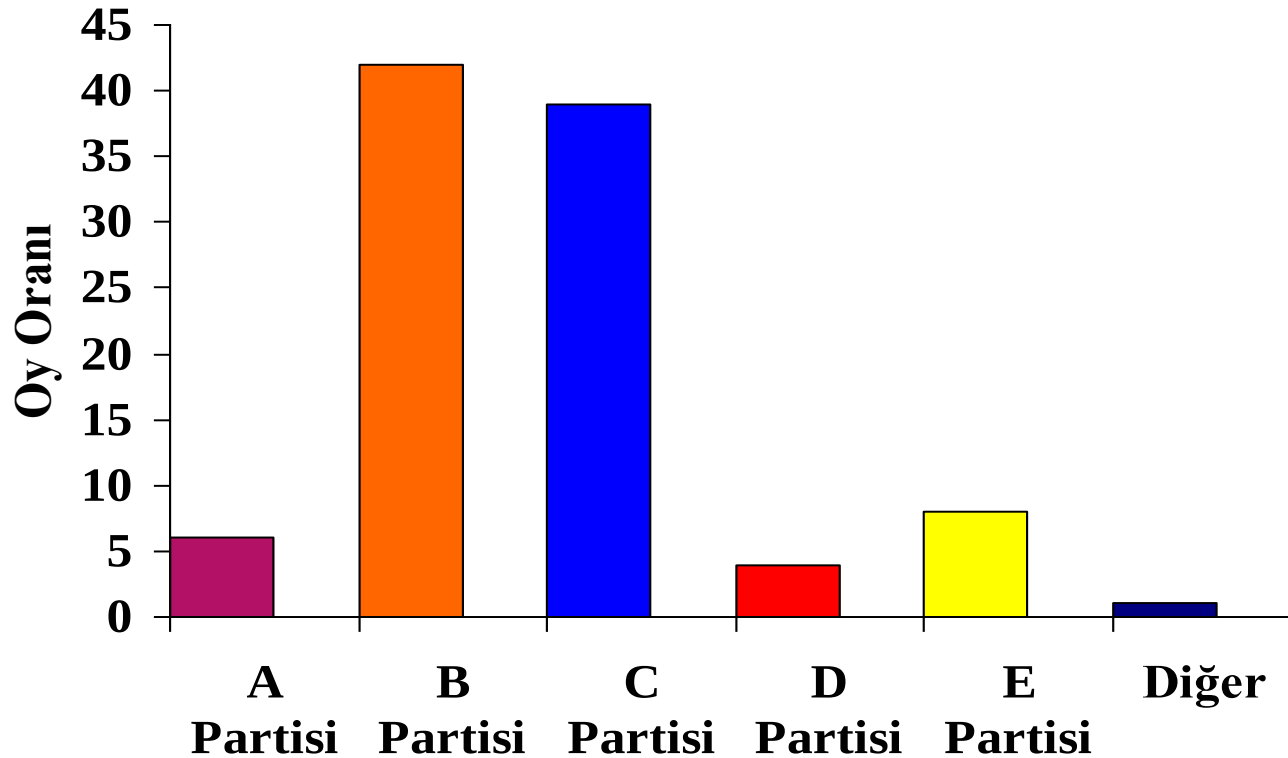
Oy oranları



Çubuk grafiği

- ▶ Her kategorinin frekansı dikey bir çubuk ile gösterilir. Bu ham veriler olabileceği gibi göreceli frekanslar da olabilir
- ▶ Farklı kategorileri karşılaştırmak için uygundur
- ▶ Çubuklar boşluklarla ayrılır

Çubuk grafiği...



Tanımlayıcı İstatistikler

- ▶ Büyük veri kümeleri çoğu zaman sadece birkaç sayı ile yeterince açıklanabilirler
 - ▶ Merkezi Eğilim Ölçüleri
 - ▶ Değişkenlik Ölçüleri

Merkezi Eğilim Ölçüleri

- ▶ Merkezi eğilim ölçüleri veri kümelerini açıklamada en sık kullanılan ölçülerdir. Bu sayılar bize veri kümesindeki ortalama, orta yada en sık görülen sayıyı verirler
- ▶ Bu bağlamda çeşitli ölçüler vardır. Burada en önemli 3 tanesinden bahsedeceğiz

Aritmetik Ortalama

- ▶ Genellikle kısaca ortalama olarak adlandırılır
- ▶ Bir örnekleme x üzeri çizgi ile gösterilir $\overline{x}$
- ▶ Ana kütle de ise μ ile gösterilir
- ▶ Ortalama bütün değerlerin toplamının değer sayısına bölünmesiyle bulunur. Örneğin n tane gözlemden oluşan bir örnekleme ortalamasının değeri

$$\overline{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Aritmetik Ortalama...

- ▶ Örneklem ortalaması genellikle ana kütle ortalamasını tahmin etmede kullanılır
- ▶ Ortalama her değeri toplama eklediğinden, uç değerlerden önemli ölçüde etkilenir. Ve dolayısıyla verileri temsil ederken yanıltıcı olabilir

Medyan - Ortanca

- ▶ Verinin bir modu yoksa, ve simetrik değilse o zaman medyan tercih edilir
- ▶ Medyan serideki veriler sıraya dizildiğinde ortadaki değerdir. Yani medyanın altında ve üstünde eşit sayıda gözlem bulunur
- ▶ Eğer tek sayıda değer varsa, medyan ortadaki değerdir. Değilse, ortadaki iki değer arasındadır ve bu iki sayının ortalaması alınarak bulunur
- ▶ Veriler sıraya konulduktan sonra, $\frac{n+1}{2}$ konumundaki değer ortancadır

Medyan...

- ▶ Örnek; 12, 34, 47, 32, 19, 17, 15
- ▶ 7 adet gözlem var, dolayısıyla;
- ▶ $\frac{n+1}{2} = \frac{7+1}{2} = 4$ 'üncü değer
- ▶ Sıraya konulduğunda; 12, 15, 17, 19, 32, 34, 47
- ▶ Medyan 19'dur

Medyan...

- ▶ Başka bir örnek;
- ▶ Veriler = 4, 6, 8, 9, 12, 16
- ▶ 6 adet gözlem var, dolayısıyla;
- ▶ $\frac{n+1}{2} = \frac{6+1}{2} = 3.5$ 'uncu değer
- ▶ = 3. ve 4. değerlerin ortalaması
- ▶ = 8 ve 9'un ortalaması = 8.5
- ▶ Medyan 8.5

Mod

- ▶ Mod bir veri kümesinde en sık rastlanan değer(ler) dir
- ▶ Özellikle ortalama ve medyanı hesaplamanın mümkün olmadığı kategorik verilerde faydalıdır.
- ▶ Bir dağılımın birden fazla modu olabilir

Değişkenlik - Yayılma Ölçüleri

- ▶ Bir dağılımın diğer önemli özelliği ise yayılma ölçüsüdür. Yani veri ne kadar değişiklik göstermektedir.
- ▶ Yaygın bir biçimde kullanılan 4 yayılma ölçüsü: Aralık, varyans, standart sapma ve değişim katsayısıdır.
- ▶ Bu bölümde ayrıca dördebölenler aralığına da bakacağız

Aralık

- ▶ Aralık bir veri kümesindeki en büyük değer ile en küçük değer arasındaki farktır
- ▶ **Aralık = $x_{\max} - x_{\min}$**
- ▶ Örneğin; “2016 yılında benzin fiyatları 4,30 TL ile 5,20 TL arasında değişmiştir”
- ▶ Aralık bu iki uç nokta arasında verilerin nasıl dağıldıkları konusunda bize bir bilgi sağlamaz. Örneğin dağılımın her iki ucunda da veya ortalarda da yığılmalar söz konusu olabilir ama biz bu bilgiyi aralıktan öğrenemeyiz.

Varyans

- ▶ Bütün verileri dikkate alan ve en çok kullanılan yayılma ölçüleri varyans ve standart sapmadır.
- ▶ Daha çok değişkenlik içeren bir veri kümesinin varyansı, nispeten daha homojen (daha az değişkenlik içeren) bir veri kümesinin varyansından büyük olur.
- ▶ Varyans, gözlemlerin ortalamadan sapmalarının kareleri toplamının gözlem sayısına bölümüyle bulunur.

Varyans...

- ▶ Şu sayıları ele alalım: 5, 17, 12, 10
- ▶ Bu verilerin ortalaması
- ▶ $(5+17+12+10)/4 = 11$
- ▶ Sapma her bir gözlemin ortalamaya olan uzaklığıdır.
- ▶ Bu veriler için sapmaları hesaplarsak
 - $5 - 11 = -6$
 - $10 - 11 = -1$
 - $12 - 11 = 1$
 - $17 - 11 = 6$

Varyans...

- Biz burada sapmaların karesi ile ilgileniyoruz. Dolayısıyla, sayıların karesini alırız

X_i	$X_i - \mu$	$(X_i - \mu)^2$
Sayı	Sapma	Sapmanın Karesi
5	-6	36
10	-1	1
12	1	1
17	6	36

Varyans...

- ▶ Karesi alınan sapmalar daha sonra toplanır ve gözlem sayısına bölünür. Böylece veri kümesinin sapması bulunur:
- ▶ $\text{Varyans} = (36 + 1 + 1 + 36) / 4$
- ▶ $= 18.5$
- ▶ Dolayısıyla varyans kareleri alınmış sapmaların ortalamasıdır

Varyans...

- Bir ana kütle için varyans σ^2 ile gösterilir ve formülü

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

- Bir örneklem için varyans s^2 ile gösterilir ve formülü

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Standart Sapma

- ▶ Standart sapma varyansın pozitif kareköküdür. Dolayısıyla ana kütle standart sapması σ , örneklem standart sapması ise s , yani;
- ▶ Standart sapma ortalama ile aynı birime sahiptir

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

Değişim katsayısı

- ▶ Değişim katsayısı göreceli bir değişkenlik ölçüsü olup herhangi bir birime sahip değildir ve genellikle yüzde olarak ifade edilir.
- ▶ Aynı birimlerle ölçülmemiş olan verileri karşılaştırmak veya birbirinden çok farklı ortalamalara sahip verileri kıyaslamak için kullanılır.
- ▶ Standart sapmayı ortalamaya bölerek hesaplanır

$$CV = \frac{s}{\bar{x}}$$

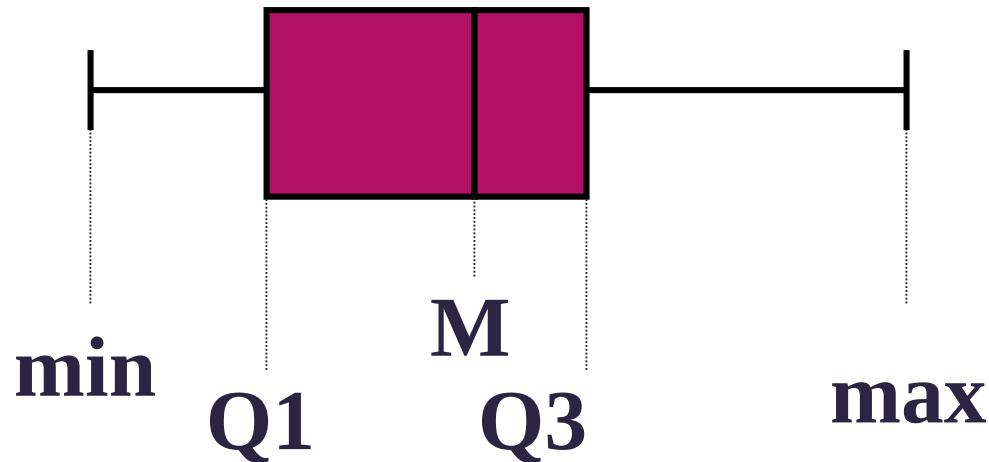
$$CV = \frac{\sigma}{\mu}$$

Dördebölenler

- ▶ Veri ile ilgili bilgilerimizi artırmanın bir diğer yolu da orta yarıya bakmaktır
- ▶ Hatırlayacağınız gibi Medyan veri kümesinin ortasındaki değerdir. Yani gözlenen değerlerin %50'si medyanın üstünde %50 ise altındadır.
- ▶ Dördebölenler de benzer şekilde hesaplanır.
- ▶ İlk dörde bölen başlangıçtan ilk çeyrek uzaklıktadır. Yani verilerin çeyrek bölümü ilk dördebölenden küçüktür.
- ▶ Üçüncü çeyrek ise veri başlangıcından 3 çeyrek uzaktadır. Yani verilerin dörtte üçü 3. dördebölenden küçüktür

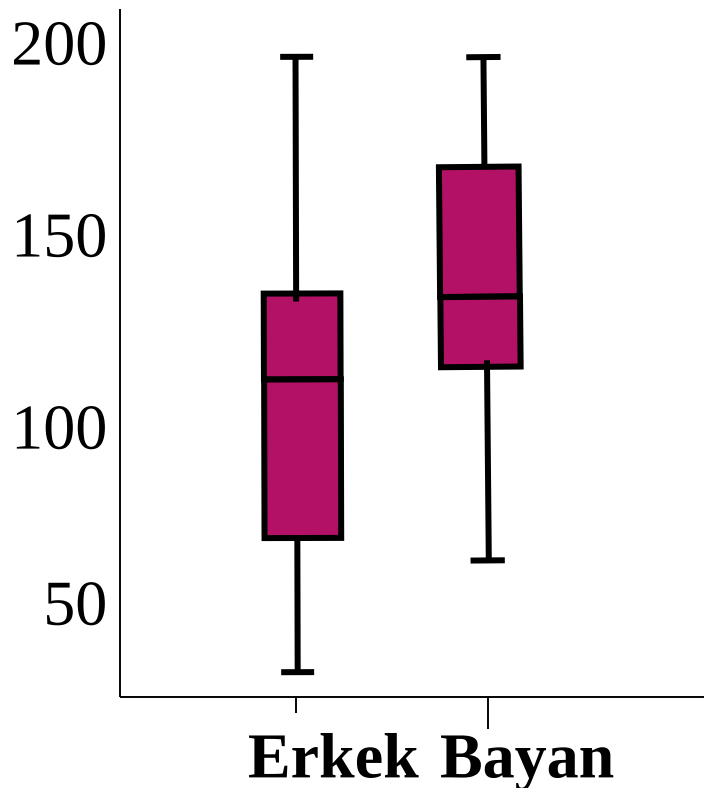
Kutu çizimleri

- Kutu grafiği yukarıdaki bilgilerin tümünü içeren güzel bir grafiktir
- Kutu çizimi kolayca verilerin aralığını, yayıklığını ve dağılımın merkezini görmemizi sağlar



Kutu çizimleri...

- Kutu Çizimleri birden fazla dağılımı karşılaştırmada oldukça faydalıdır.



Kutu çizimleri dikey veya yatay olarak çizilebilirler.