

AKT102 İSTATİSTİK

BÖLÜM 11
KORELASYON

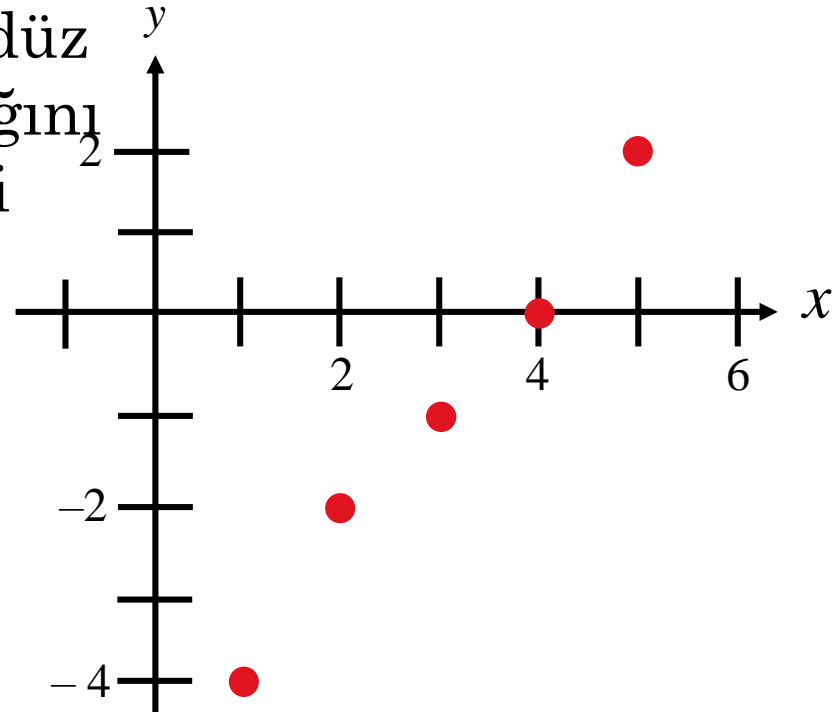
Korelasyon

Korelasyon iki deęişken arasındaki ilişkidir. Veriler, x 'in bağımsız (veya açıklayıcı) deęişken olduęu ve y 'nin bağımlı (veya yanıt) deęişken olduęu, sıralı (x, y) çiftleri ile temsil edilebilir.

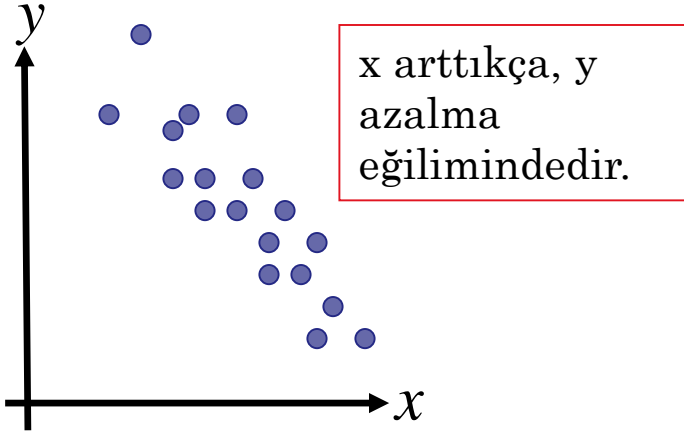
İki deęişken arasında doğrusal (düz çizgi) bir korelasyon olup olmadığını belirlemek için bir saçılım grafięi (scatter plot) kullanılabilir.

Örnek:

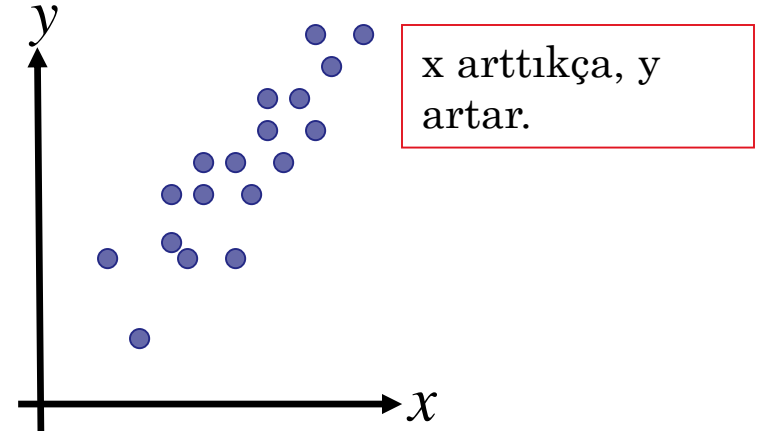
x	1	2	3	4	5
y	-4	-2	-1	0	2



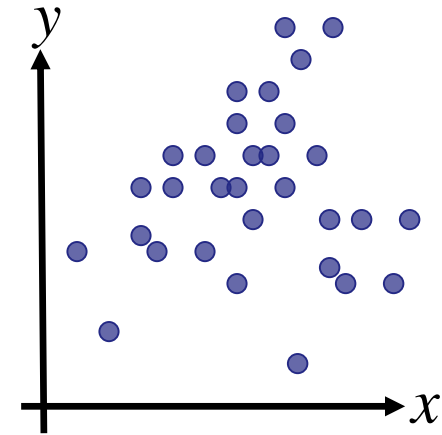
Liner (Doğrusal) Korelasyon



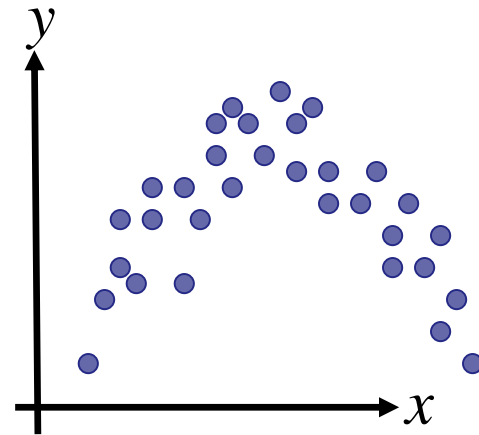
Negatif Liner Korelasyon



Positif Liner Korelasyon



Korelasyon Yok



Liner Olamayan Korelasyon

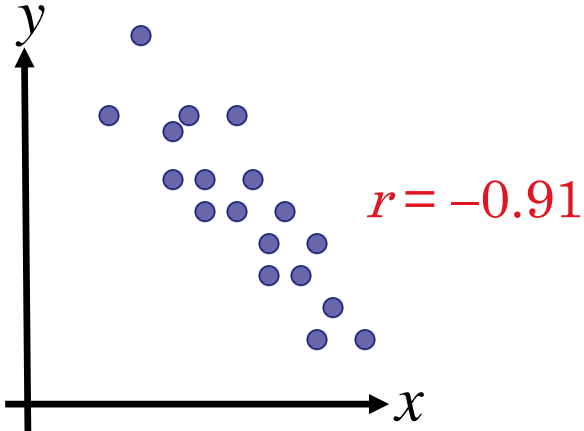
Korelasyon Katsayısı

Korelasyon katsayısı, iki deęişken arasındaki doğrusal bir ilişkinin gücünün ve yönünün bir ölçüsüdür. r sembolü, örneklem korelasyon katsayısını temsil eder. r için formül;

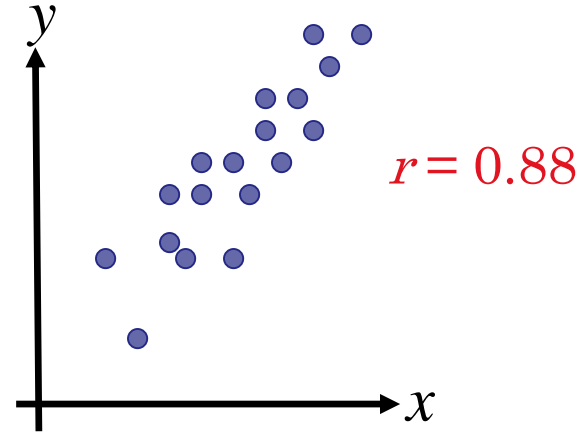
$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}.$$

Korelasyon katsayısının aralığı -1 ile 1'dir. Eğer x ve y güçlü bir pozitif doğrusal korelasyona sahipse r , 1'e yakındır. X ve y 'nin güçlü bir negatif doğrusal korelasyonu varsa, r , -1'e yakındır. Doğrusal bir korelasyon yoksa veya zayıf bir doğrusal korelasyon varsa r , 0'a yakındır.

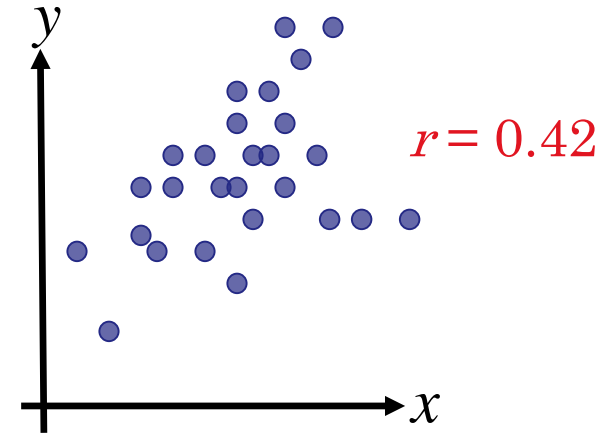
Linear Korelasyon



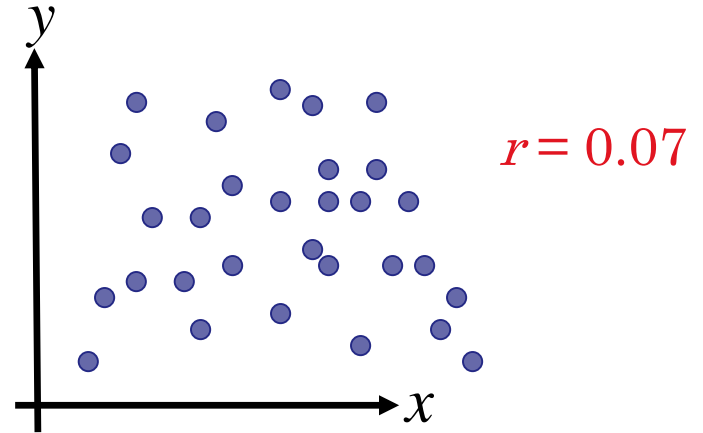
Güçlü negatif korelasyon



Güçlü pozitif korelasyon



Güçsüz pozitif korelasyon



Linear olmayan korelasyon

Korelasyon Katsayısının Hesaplanması

Korelasyon Katsayısının Hesaplanması

Açıklama

1. x - değerlerinin toplamını bulun
2. y - değerlerinin toplamını bulun
3. Her x değerini, karşılık gelen y değeriyle çarpın ve toplamı bulun.
4. Her x değerinin karesini alın ve toplamı bulun.
5. Her y değerinin karesini alın ve toplamı bulun..
6. Korelasyon katsayısını hesaplamak için bu beş toplamı kullanın.

Gösterim

$$\sum x$$

$$\sum y$$

$$\sum xy$$

$$\sum x^2$$

$$\sum y^2$$

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2} \sqrt{n\sum y^2 - (\sum y)^2}}$$

Korelasyon Katsayısı

Örnek:

Aşağıdaki veriler için korelasyon katsayısı r yi hesaplayın.

x	y	xy	x^2	y^2
1	-3	-3	1	9
2	-1	-2	4	1
3	0	0	9	0
4	1	4	16	1
5	2	10	25	4
$\Sigma x = 15$	$\Sigma y = -1$	$\Sigma xy = 9$	$\Sigma x^2 = 55$	$\Sigma y^2 = 15$

$$r = \frac{n \Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n \Sigma x^2 - (\Sigma x)^2} \sqrt{n \Sigma y^2 - (\Sigma y)^2}} = \frac{5(9) - (15)(-1)}{\sqrt{5(55) - 15^2} \sqrt{5(15) - (-1)^2}}$$
$$= \frac{60}{\sqrt{50} \sqrt{74}} \approx 0.986$$

x ve y arasında güçlü bir pozitif doğrusal ilişki vardır.

Korelasyon Katsayısı

Örnek:

Aşağıdaki veriler hafta sonu boyunca 12 farklı öğrencinin televizyon izlediği saat sayısını ve ertesi Pazartesi sınava giren her öğrencinin puanlarını göstermektedir.

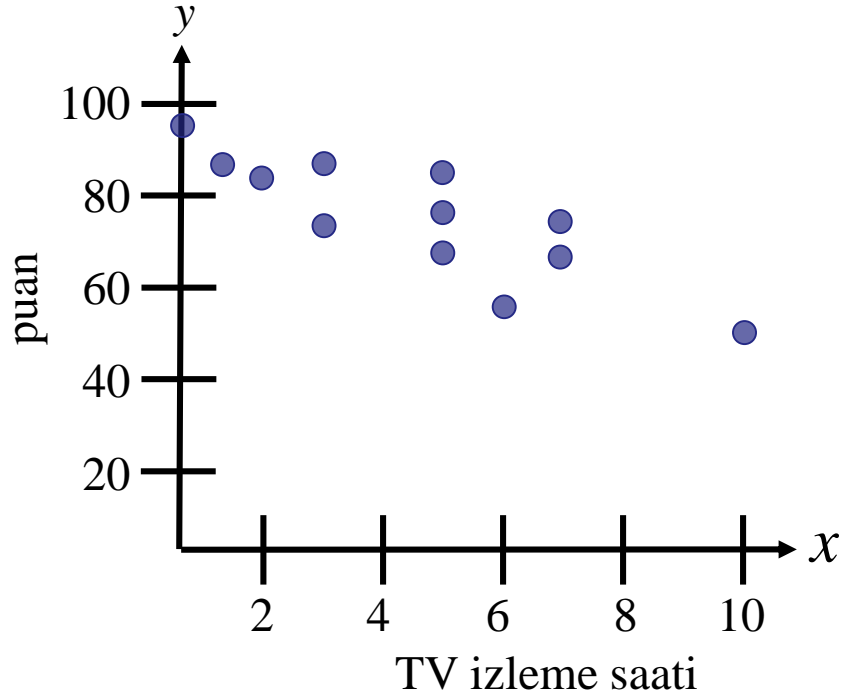
- Saçılım grafiğini görüntüleyin.
- Korelasyon katsayısını r hesaplayın.

saat, x	0	1	2	3	3	5	5	5	6	7	7	10
puan, y	96	85	82	74	95	68	76	84	58	65	75	50

Korelasyon Katsayısı

Örneğin devamı:

saat, x	0	1	2	3	3	5	5	5	6	7	7	10
puan, y	96	85	82	74	95	68	76	84	58	65	75	50



Korelasyon Katsayısı

Örneğin devamı:

saat x	0	1	2	3	3	5	5	5	6	7	7	10
puan, y	96	85	82	74	95	68	76	84	58	65	75	50
xy	0	85	164	222	285	340	380	420	348	455	525	500
x^2	0	1	4	9	9	25	25	25	36	49	49	100
y^2	9216	7225	6724	5476	9025	4624	5776	7056	3364	4225	5625	2500

$$\sum x = 54$$

$$\sum y = 908$$

$$\sum xy = 3724$$

$$\sum x^2 = 332$$

$$\sum y^2 = 70836$$

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}} = \frac{12(3724) - (54)(908)}{\sqrt{12(332) - 54^2} \sqrt{12(70836) - (908)^2}} \approx -0.831$$

Güçlü bir negatif doğrusal korelasyon var.

TV izlemek için harcanan saat arttıkça, puanlar düşme eğilimindedir.

Kitle Korelasyon Katsayısının Test Edilmesi

Örneklem korelasyon katsayısı (r) hesaplandıktan sonra, ρ popülasyon korelasyon katsayısının belirli bir anlamlılık düzeyinde anlamlı olduğuna karar vermek için yeterli kanıt olup olmadığını belirlememiz gerekir.

Bunu belirlemenin bir yolu Ek B'deki tablonun kullanılmasıdır.

Eğer $|r|$ kritik değerden daha büyükse, ρ korelasyon katsayısının anlamlı olduğuna karar vermek için yeterli kanıt vardır.

n	$\alpha = 0.05$	$\alpha = 0.01$
4	0.950	0.990
5	0.878	0.959
6	0.811	0.917
7	0.754	0.875

eğer $|r| > 0.811$ ise $n=6$ boyutundaki bir örneklem için ρ , % 5 anlamlılık düzeyinde anlamlıdır.

Kitle Korelasyon Katsayısının Test Edilmesi

Korelasyon Katsayısı ρ nun Bulunması

Açıklama

1. Örneklemdaki veri çiftlerinin sayısını belirleyin.
2. Önem düzeyini belirtin.
3. Kritik değeri bulun.
4. Korelasyonun anlamlı olup olmadığına karar verin.
5. Kararı yorumlayın.

Gösterim

n i belirleyin

α yı belirleyin

Ek B yi kullanın.

Eğer $|r| >$ kritik değer ise korelasyon anlamlıdır.

Aksi takdirde, korelasyonun anlamlı olduğunu destekleyen yeterli kanıt yoktur.

Kitle Korelasyon Katsayısının Test Edilmesi

Örnek:

Aşağıdaki veriler hafta sonu boyunca 12 farklı öğrencinin televizyon izlediği saat sayısını ve ertesi Pazartesi sınava giren her öğrencinin puanlarını göstermektedir.

Korelasyon katsayısı $r \approx -0.831$.

saat, x	0	1	2	3	3	5	5	5	6	7	7	10
puan, y	96	85	82	74	95	68	76	84	58	65	75	50

Korelasyon katsayısı $\alpha = 0.01$ de anlamlı mı?

Kitle Korelasyon Katsayısının Test Edilmesi

Örneği devamı:

Ek B:

$$r \approx -0.831$$

$$n = 12$$

$$\alpha = 0.01$$

n	$\alpha = 0.05$	$\alpha = 0.01$
4	0.950	0.990
5	0.878	0.959
6	0.811	0.917
// //		
10	0.632	0.765
11	0.602	0.735
12	0.576	0.708
13	0.553	0.684

$$|r| > 0.708$$

Kitle korelasyonu önemli olduğu için, haftasonu boyunca izlenen televizyon saatlerinin sayısı ile teste giren her öğrencinin notları arasında anlamlı bir doğrusal korelasyon olduğu sonucuna varmak için % 1 anlamlılık düzeyinde yeterli kanıt vardır.

ρ için Hipotez Testi

Örneklem korelasyon katsayısının r , popülasyon korelasyon katsayısının ρ belirli bir anlamlılık düzeyinde anlamlı olduğu sonucuna varmak için yeterli kanıt sağlayıp sağlamadığını belirlemek için hipotez testi de kullanılabilir.

Bir hipotez testi tek yönlü veya iki yönlü olabilir.

$$\begin{cases} H_0: \rho \geq 0 & (\text{anlamlı negatif korelasyon yok}) \\ H_a: \rho < 0 & (\text{anlamlı negatif korelasyon}) \end{cases} \quad \begin{array}{l} \text{Sol yönlü} \\ \text{(kuyruklu) test} \end{array}$$

$$\begin{cases} H_0: \rho \leq 0 & (\text{anlamlı pozitif korelasyon yok}) \\ H_a: \rho > 0 & (\text{anlamlı pozitif korelasyon}) \end{cases} \quad \begin{array}{l} \text{Sağ yönlü} \\ \text{(kuyruklu) test} \end{array}$$

$$\begin{cases} H_0: \rho = 0 & (\text{anlamlı korelasyon yok}) \\ H_a: \rho \neq 0 & (\text{anlamlı korelasyon}) \end{cases} \quad \text{İki yönlü test}$$

ρ için Hipotez Testi

Korelasyon Katsayısı için t Testi

İki değişken arasındaki korelasyonun anlamlı olup olmadığını test etmek için t testi kullanılabilir. Test istatistiği r ve standart test istatistiği;

$$t = \frac{r}{\sigma_r} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

$n - 2$ serbestlik dereceli t-dağılımı gösterir.

Bu bölümde, ρ için sadece yönlü hipotez testleri ele alınmıştır.

ρ için Hipotez Testi

t-Testinin Korelasyon Katsayısı ρ için kullanılması

Açıklama

1. Boş ve alternatif hipotezi açıklayın.
2. Anlamlılık düzeyini belirtin.
3. Serbestlik derecesini tanımlayın.
4. Kritik değerleri ve reddetme bölgelerini belirleyin.

Gösterim

H_0 ve H_a yı belirtin

α yı belirleyin

$$\text{s.d} = n - 2$$

Ek B tablosunu kullanın.

ρ için Hipotez Testi

t-Testinin Korelasyon Katsayısı ρ için kullanılması

Açıklama

5. Standartlaştırılmış test istatistiğini bulun.
6. Boş hipotezi reddetme veya reddetmeme konusunda bir karar verin.
7. Kararı yorumlayın

Gösterim

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

t reddetme bölgesinde ise H_0 'ı reddedin. Aksi takdirde H_0 'ı reddedemezsiniz.

ρ için Hipotez Testi

Örnek:

Aşağıdaki veriler hafta sonu boyunca 12 farklı öğrencinin televizyon izlediği saat sayısını ve ertesi Pazartesi sınava giren her öğrencinin puanlarını göstermektedir.

Korelasyon katsayısı $r \approx -0.831$.

saat, x	0	1	2	3	3	5	5	5	6	7	7	10
puan, y	96	85	82	74	95	68	76	84	58	65	75	50

Bu korelasyon katsayısının önemini $\alpha = 0.01$ 'de anlamlı mı?

ρ için Hipotez Testi

Örneğin devamı:

$H_0: \rho = 0$ (korelasyon yok) $H_a: \rho \neq 0$ (korelasyon anlamlı)

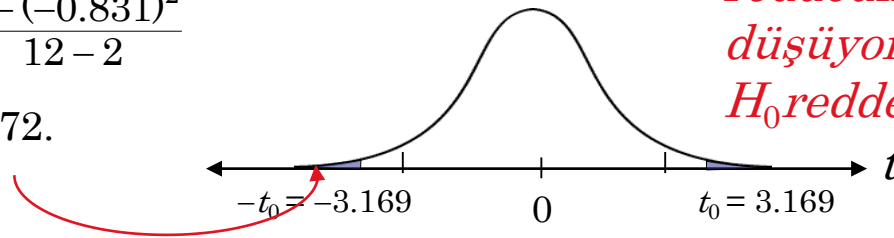
Anlam düzeyi $\alpha = 0.01$.

Serbestlik derecesi s.d = $12 - 2 = 10$.

Kritik değerler $-t_0 = -3.169$ ve $t_0 = 3.169$.

Standartlaştırılmış test istatistiği

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{-0.831}{\sqrt{\frac{1-(-0.831)^2}{12-2}}} \approx -4.72.$$



Test istatistiği reddedilme bölgesinde düşüyor, bu nedenle H_0 reddedildi.

% 1 anlamlılık düzeyinde, hafta sonu izlenen TV sayısı ile Pazartesi sabahı test puanları arasında anlamlı bir doğrusal ilişki olduğu sonucuna varmak için yeterli kanıt vardır.

Korelasyon ve Nedensellik

İki deęişkenin güçlü bir şekilde ilişkilendirildięi gerçeęi, başlı başına deęişkenler arasında sebep-sonuç ilişkisi anlamına gelmez.

İki deęişken arasında anlamlı bir ilişki varsa, aşağıdaki olasılıkları göz önünde bulundurmalısınız..

1. Deęişkenler arasında doğrudan sebep-sonuç ilişkisi var mı?
X, y'ye neden olur mu?
2. Deęişkenler arasında ters sebep-sonuç ilişkisi var mı?
Y x'e neden olur mu?
3. Deęişkenler arasındaki ilişkinin üçüncü bir deęişkenden veya başka birkaç deęişkenin birleşmesinden kaynaklanması olabilir mi?
4. İki deęişken arasındaki ilişkinin tesadüf olması olabilir mi?