

KORELASYON

Korelasyon katsayısı, iki deęişken arasındaki doğrusal ilişkinin yönünü ve derecesini belirtir. Söz konusu deęişkenler oranlama ya da eşit aralıklı düzeyde ölçülmüş olmalıdır.

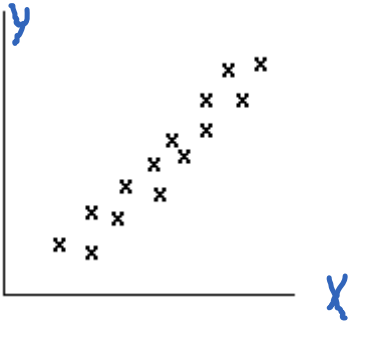
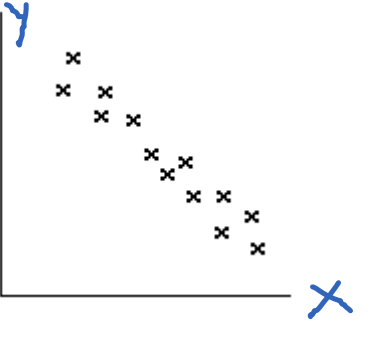
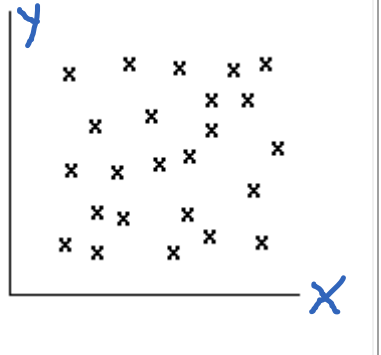
Örneğin sınava hazırlanma süresi ile sınavdan alınan puan arasındaki ilişki, reklam harcaması ile satış miktarı arasındaki ilişki, insanların boy uzunlukları ile ağırlıkları arasındaki ilişki korelasyon katsayısı ile belirlenebilir.

Pearson korelasyon katsayısı olarak da bilinen korelasyon katsayısı, her iki deęişkenin de sürekli olması ve normal dağılım göstermesi durumunda hesaplanır.

Eğer deęişkenler normal dağılım göstermiyorsa, Sperman Brown sıra farkları korelasyon katsayısı kullanılır.

Bir deęişkenin deęerleri artarken dięer deęişkenin deęerleri düzenli olarak artıyor veya azalıyorsa iki deęişken arasında doğrusal bir ilişki vardır denir. Her ikisi de artarsa aynı yönde (pozitif), biri artarken dięeri azalıyorsa ters yönde (negatif) ilişkiden bahsedilir.

Deęişkenlerin serpilme diyagramına bakılarak, korelasyon hakkında ön bilgi edinilebilir.

		
Aynı yönde (pozitif) ilişki, Pozitif korelasyon	Ters yönde (negatif) ilişki, Negatif korelasyon	İlişki yok, Korelasyon yok

Pearson korelasyon katsayısı aşağıdaki formül ile hesaplanır:

$$r_{XY} = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{n})(\sum Y^2 - \frac{(\sum Y)^2}{n})}}$$

$-1 \leq r_{XY} \leq 1$ dir. Yani, Korelasyon katsayısının -1 ile +1 arasında değer alır.

$r_{XY} = 0$ ise değişkenler arasında ilişki yoktur.

$r_{XY} > 0$ ise değişkenler arasında aynı yönde (pozitif) ilişki vardır.

$r_{XY} < 0$ ise değişkenler arasında ters yönde (negatif) ilişki vardır.

r_{XY} 'nin aldığı değer;

sıfıra yaklaştıkça ilişkinin derecesi azalmaktadır,

+1 'e yaklaştıkça ilişkinin derecesi aynı yönde artmaktadır (kuvvetlenmektedir),

- 1 'e yaklaştıkça ilişkinin derecesi ters yönde artmaktadır (kuvvetlenmektedir).

Örnek:

Gübre miktarı	Verim
30	150
20	128
42	140
50	200
45	180
65	210

Yukarıdaki bilgilere göre değişkenler arasında ilişki olup olmadığını, varsa nasıl ve ne kadarlık bir ilişki olduğunu belirleyiniz.

Çözüm. Değişkenlerin dağılımı hakkında bilgi yok. Normal dağıldıklarını varsayalım.

Değişkenler nicel.

Pearson korelasyon katsayısı ile ilişkiye bakabiliriz.

$$r_{XY} = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{n})(\sum Y^2 - \frac{(\sum Y)^2}{n})}}$$

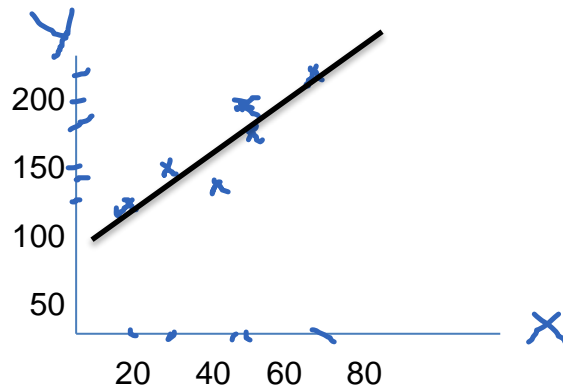
Hesaplama için gerekli bilgiler aşağıdaki tablodadır.

Gübre miktarı (X)	Verim (Y)	X ²	Y ²	XY
30	150	900	22500	4500
20	128	400	16384	2560
42	140	1764	19600	5880
50	200	2500	40000	10000
45	180	2025	32400	8100
65	210	4225	44100	13650
$\sum X = 257$	$\sum Y = 1008$	$\sum X^2 = 11814$	$\sum Y^2 = 174984$	$\sum XY = 44690$

$$r_{XY} = \frac{44690 - \frac{(257)(1008)}{6}}{\sqrt{(11814 - \frac{(257)^2}{6})(174984 - \frac{(1008)^2}{6})}} = 0,89$$

bulunur.

Değişkenler arasında doğrusal, aynı yönde(pozitif), kuvvetli bir ilişki vardır. Bu ilişkin derecesi yaklaşık %89 dur.



Korelasyon katsayısının önem kontrolü:

Hesaplanmış olan korelasyon katsayısının tesadüfi mi yoksa gerçek bir ilişkiyi mi gösterdiğinin belirlenmesi için denetlenmesi gerekir. Bunun için hipotez testi yapılır. Kurulacak hipotezler aşağıdaki gibi olmalıdır.

$H_0 : \rho=0$ (doğrusal ilişki yoktur);

$H_1 : \rho > 0$ (doğrusal ilişki vardır)

Burada ρ , kitle korelasyon katsayısını belirtmektedir.

Spearman Sıra Korelasyon Katsayısı

Doğrusal korelasyonda ilişkisi araştırılan değişkenlerin nicel olması ve normal dağılması gerekir. Bu varsayımlar sağlanmadığında Spearman Sıra Korelasyon Katsayısı kullanılır.

Sıra korelasyon katsayısının hesaplanırken önce gözlem değerleri büyükten küçüğe (veya küçükten büyüğe) doğru sıralanır ve bu sıralamaya göre sıra numarası verilir. Aynı olan gözlemlerin sıra numaraları verilirken sıra no ortalaması alınır. Formülü aşağıdaki biçimdedir.

$$r_s = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)}$$

D_i : Değişkenlerinin sıra numaraları arasındaki fark

n : Toplam gözlem sayısı

Örnek. 7 öğrencinin boy uzunlukları(X) ve ağırlıkları(Y) aşağıda gösterilmiştir. Veriler normal dağılım göstermediğine göre, boy uzunlukları ile ağırlıklar arasındaki ilişkiyi hesaplayıp yorumlayınız.

Veriler normal dağılım göstermediği için Pearson korelasyon katsayısı hesaplanmaz.

İlişki Sperman sıra korelasyonu ile belirlenmelidir.

Boy uzunluğu(X) (cm)	Ağırlık(Y) (kg)	X sıra no	Y sıra no	D_i	D_i^2
190	130	2	1	1	1
180	72	4	6	-2	4
175	77	6	5	1	1
198	102	1	2	-1	1
186	88	3	3	0	0
169	66	7	7	0	0
177	80	5	4	1	1

$$r_s = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)} = 1 - \frac{6(8)}{7(7^2 - 1)} = 0,86$$

Öğrencilerin boy uzunlukları ile ağırlıkları arasında aynı yönde kuvvetli bir ilişki vardır.