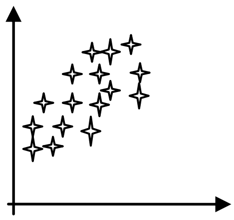
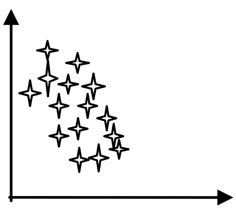
Korelasyon Çözümlemesi

Korelasyon katsayısı, iki ya da daha çok değişken arasındaki ilişkiyi gösterir. Korelasyon çözümlemesinin amacı, değişkenler arasındaki ilişkinin büyüklüğünü ve yönünü belirlemektir. İlişkinin büyüklüğünü belirlemede çeşitli ilişki katsayıları mevcuttur. Değişkenler arasındaki ilişkinin gücünü ölçmek için kullanılan bu ilişki katsayıları, analizin amacına, değerlendirilen değişkenlerin türüne ve sayısına göre farklılık göstermekle birlikte yaygın olarak kullanılan katsayılar Pearson Momentler Çarpım Korelasyonu, Spearman sıra farkları korelasyonu, Phi, Kendall Tau-a, Kendall Tau-b, Kendall Tau-c, Goodman ve Kruskal Gamma katsayılarıdır. Burada Pearson korelasyon katsayısına değinilecektir.

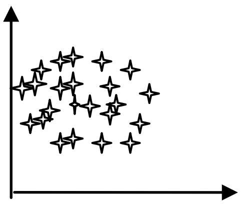
ρ : Pearson korelasyon katsayısı



İki değişkenin her ikisi de aynı yönde değişim gösterirse aralarındaki ilişki pozitiftir ve korelasyon katsayısının işareti (+) olur.



Değişkenlerden biri artarken diğeri azalıyorsa ilişki negatiftir ve korelasyon katsayısı (-) işaretlidir.



Değişkenler arasında negatif yada pozitif yönde birlikte bir değişim yoksa bu durumda korelasyon katsayısı sıfırdır.

$\rho = 1$ ise iki değişken arasında pozitif yönde tam bir ilişki vardır.

$\rho = -1$ ise iki değişken arasında negatif yönde tam bir ilişki vardır.

$R^2 \longrightarrow$ Belirtme katsayısı

$$R^2 = \frac{\text{RKT}}{\text{YOAKT}}$$

$r = \pm\sqrt{R^2} \longrightarrow$ örneklem korelasyon katsayısı (Pearson'ın korelasyon katsayısı ($\hat{\rho} = r$))

$-1 \leq \rho \leq 1, \qquad -1 \leq r \leq 1$

r'nin işareti regresyon katsayısı b_1'in işareti ile aynıdır.

$b_1 > 0 \quad \rightarrow \quad r_1 > 0$

$b_1 < 0 \quad \rightarrow \quad r_1 < 0$

$\sum(y - \hat{y}) = 0$ ise $r = 1$ ya da $r = -1$ olur

Eğer regresyon doğrusu $\bar{y}$ ile çakışan yatay doğru ise yani eğim sıfır olduğunda

$$\sum(y - \hat{y})^2 = \sum(y - \bar{y})^2$$

dolayısıyla $b_0 = \bar{y}$ olur. Buradan $r = 0$'dır.

$\rho = \frac{Cov(x,y)}{\sqrt{var(x)var(y)}} \quad ; \quad Cov(x, y) = E(xy) - E(x)E(y)$

Örneklem için,

$$r = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)\left(\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right)}}$$

Korelasyon Katsayısına İlişkin Hipotez Testi

Korelasyon katsayısının önemliliğini test ederken örneklem dağılımının bilinmesi gerekir. Örneklem korelasyon katsayısının (r) dağılımı, örneklem hacmi (n) ve kitle korelasyon katsayısına (ρ) bağlıdır. Örneklem korelasyon katsayısı r'nin dağılımı ρ'nun büyük değerleri ve örneklem hacminin küçük değerleri için normal dağılımdan farklıdır.(ρ küçük olduğunda dağılım normaldir.)

$\rho = 0$ olan bir kitleden çekilen rasgele örneklemin korelasyon katsayısının dağılımı $r \sim N(0, \sigma^2)$'dir. Ancak $\rho \neq 0$ olduğunda bu durum söz konusu değildir.

$\rho = 0$ ve $\rho \neq 0$ olmasına göre iki test yapılacaktır.

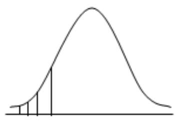
$\rho = 0$'ın Testi

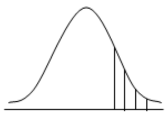
1. $H_0: \rho = 0$ ($H_0: \beta_1 = 0$ hipotezinin testine eşdeğer test)

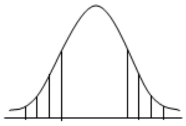
$H_1: \rho < 0,\ \rho > 0,\ \rho \neq 0$

2. $t_H = \frac{r-\rho}{S_r}$, $S_r = \sqrt{\frac{1-r^2}{n-2}}$

3.







$t_H < -t_{T(\alpha,n-2)}$ ise H_0 red

$t_H > t_{T(\alpha,n-2)}$ ise H_0 red

$t_H < -t_{T(\alpha/2,n-2)}$ ya da
$t_H > t_{T(\alpha/2,n-2)}$ ise H_0 red

$\rho \neq 0$'ın Testi

1. $H_0: \rho = \rho_0$

$H_1: \rho < \rho_0,\ \rho > \rho_0,\ \rho \neq \rho_0$

2. Fisher'in Z dönüşümü

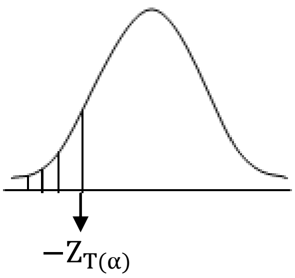
$$Z_r = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right), \qquad Z_r \sim N(\mu_r, \sigma_r^2)$$

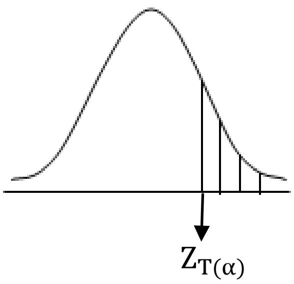
$$\mu_r = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right), \qquad \sigma_r = \sqrt{\frac{1}{n-3}}$$

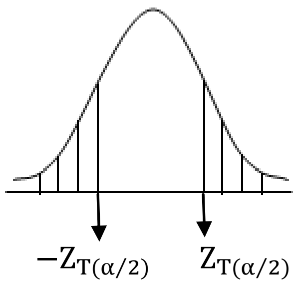
Test istatistiği

$$Z_H = \frac{Z_r - \mu_r}{\sigma_r}$$

3. $H_1: \rho < \rho_0$ $H_1: \rho > \rho_0$ $H_1: \rho \neq \rho_0$







$Z_H < -Z_{T(\alpha)} \rightarrow H_0$ red edilir

$Z_H > Z_{T(\alpha)} \rightarrow H_0$ red edilir

$Z_H < -Z_{T(\alpha/2)}$ yada
$Z_H > Z_{T(\alpha/2)} \rightarrow H_0$ red edilir

Korelasyon katsayısı ρ_0 değerinden farklıdır, biçiminde yorum yapılır.

Kitle Korelasyon Katsayısının Güven Aralığı

ρ kitle korelasyon katsayısının güven aralığını bulabilmek için önce μ_r için aralık bulunur sonra ρ'ya dönüştürülür.

$$P(Z_r - Z_{T(\alpha/2)}\sigma_r \leq \mu_r \leq Z_r + Z_{T(\alpha/2)}\sigma_r) = 1 - \alpha$$

μ_r yerine yazılırsa;

$$P\left(Z_r - Z_{T(\alpha/2)}\sigma_r \leq \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right) \leq Z_r + Z_{T(\alpha/2)}\sigma_r\right) = 1 - \alpha$$

olur.

μ_r için aralık bulunduktan sonra ρ için aralık bulunur.