Regresyon-Korelasyon Örnek

Örnek: Ciğerlerinden rahatsız olan kişilerden 10 tanesi rasgele seçilmiş ve sigara içtikleri yıl sayısı ve ciğer rahatsızlığının derecesine ilişkin bilgileri aşağıda verilmiştir.

Sigara içilen yıl:	25 36 22 15 48 39 42 31 28 33
Rahatsızlık derecesi:	55 60 50 30 75 70 70 55 30 35

a) Bağımlı ve bağımsız değişkeni belirleyerek sigara içmek ve ciğerlerinden rahatsız olmak arasındaki ilişkiyi ifade eden regresyon denklemini bulunuz. 19 yıl sigara içen bir kimsenin rahatsızlığının derecesini tahmin ediniz.
b) $\alpha = 0.05$ anlam düzeyinde regresyon katsayılarının önem kontrolünü yapınız.
c) β_0 ve β_1 için %95 güven düzeyinde güven aralıklarını oluşturunuz.
d) Sigara içmek ile ciğerlerinden rahatsız olma arasındaki ilgi miktarını bularak, bu iki değişken arasındaki ilgi miktarının %75 olduğu hipotezini $\alpha = 0.05$ anlam düzeyinde test ediniz.
e) Önem kontrolünü $\rho = 0$ durumu için yapınız.
f) Kitle korelasyon katsayısının %95'lik güven sınırlarını belirleyiniz.
g) Bağımsız değişkenin bağımlı değişkendeki değişimin yüzde kaçını açıkladığını ifade ederek yorumlayınız.

$$\sum x_j = 319 \quad , \sum x_j^2 = 11053 \quad , \bar{x} = 31.9$$
$$\sum y_j = 530 \quad , \sum y_j^2 = 30600 \quad , \bar{y} = 53$$
$$\sum x_j y_j = 18055$$

a) x: sigara içilen yıl (bağımsız)
y: rahatsızlık derecesi (bağımlı)

$\hat{y} = b_0 + b_1 x$

$$b_1 = \frac{\sum x_j y_j - \frac{\sum x_j \sum y_j}{n}}{\sum x_j^2 - \frac{(\sum x_j)^2}{n}} = \frac{18055 - \frac{(319)(530)}{10}}{11053 - \frac{(319)^2}{10}} = \frac{1148}{876.9} = 1.309$$

$b_0 = \bar{y} - b_1\bar{x} = 53 - 1.309(31.9) = 11.243$

$\hat{y} = 11.243 + 1.309x$

$\hat{y} = 11.243 + 1.309(19) = 36.114$

b) **1)** $H_0: \beta_0 = 0$
$H_1: \beta_0 \neq 0$

2) $t_H = \frac{b_0 - \beta_{0.0}}{S_{b_0}} = \frac{b_0}{S_{b_0}}$

$$S_{b_0} = \sqrt{HKO\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum x_j^2 - \frac{\left(\sum x_j\right)^2}{n}}\right]}$$

$$\sum x_j^2 - \frac{\left(\sum x_j\right)^2}{n} = 876.9$$

$$\sum y_j^2 - \frac{\left(\sum y_j\right)^2}{n} = 30600 - \frac{(530)^2}{10} = 2510$$

$$\sum x_j y_j - \frac{\sum x_j \sum y_j}{n} = 1148$$

$$RKT = 1.309 \times 1148 = 1502.732$$

$$HKT = GnKT - RKT = 2510 - 1502.732 = 1007.268$$

$$HKO = \frac{1007.268}{8} = 125.909$$

$$S_{b_0} = \sqrt{HKO\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum x_j^2 - \frac{\left(\sum x_j\right)^2}{n}}\right]}$$

$$= \sqrt{125.909\left(\frac{1}{10} + \frac{(31.9)^2}{876.9}\right)} = \sqrt{158.704} = 12.598$$

$$t_H = \frac{b_0}{S_{b_0}} = \frac{11.243}{12.598} = 0.892$$

$$t_T(0.025\,, sd = 8) = 2.306$$

3) $t_H < t_T$ olduğundan H_0 red edilemez. Yani, β_0 katsayısı önemli değildir.

1. $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

2. $t_H = \frac{b_1}{S_{b_1}}$

$$S_{b_1} = \sqrt{\frac{HKO}{\sum x_j^2 - \frac{\left(\sum x_j\right)^2}{n}}} = \sqrt{\frac{125.909}{876.9}} = 0.144$$

$$t_H = \frac{b_1}{S_{b_1}} = \frac{1.309}{0.144} = 9.09$$

3. $t_T(0.025\,, sd = n - 2 = 8) = 2.306$

$t_H > t_T$ olduğundan H_0 red edilir. Regresyon katsayısı önemlidir. Eğim önemli olduğu için regresyon doğrusunun anlamlı olduğu söylenebilir.

F testi ile,

1) $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

2) Varyans Analizi Tablosu

DK	Sd	KT	KO	Test
Regresyon	1	1502.732	1502.732	$F_H = \dfrac{1502.732}{125.909} = 11.935$
Regresyondan ayrılış	n-2=8	1007.268	125.909	
Toplam	n-1=9	2510		

3) $F_T(0.05\,, sd1 = 1, sd2 = n - 2 = 8) = 5.32$

$F_H > F_T$ olduğundan H_0 red edilir. Yani, regresyon doğrusu önemlidir.

c) β_1 için;

$P\left(b_1 - t_T S_{b_1} \leq \beta_1 \leq b_1 + t_T S_{b_1}\right) = 1 - \alpha$

$\beta_1 \epsilon (1.309 - 2.306(0.144); 1.309 + 2.306(0.144))$

$\beta_1 \epsilon [0.9769; 1.6411]$

β_0 için;

$P\left(b_0 - t_T S_{b_0} \leq \beta_0 \leq b_0 + t_T S_{b_0}\right) = 1 - \alpha$

$\beta_0 \epsilon (11.243 - 2.306(12.598); (11.243 + 2.306(12.598))$

$\beta_0 \epsilon [-17.808; 40.294]$

d) $H_0: \rho = 0.75$

$H_1: \rho \neq 0.75$

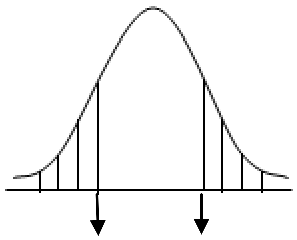
$$GnKT = \sum y_j^2 - \frac{\left(\sum y_j\right)^2}{n} = 30600 - \frac{(530)^2}{10} = 2510$$

$$r = \frac{1148}{\sqrt{(876.9)(2510)}} = 0.774$$

$$Z_H = \frac{Z_r - \mu_r}{\sigma_r} = \frac{1.030 - 0.973}{0.378} = 0.151$$

$$Z_r = \frac{1}{2}\ln\left(\frac{1 + r}{1 - r}\right) = \frac{1}{2}\ln\left(\frac{1 + 0.774}{1 - 0.774}\right) = 1.030$$

$$\mu_r = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right) = \frac{1}{2}\ln\left(\frac{1+0.75}{1-0.75}\right) = 0.973 \qquad , \qquad \sigma_r = \sqrt{\frac{1}{n-3}} = \sqrt{\frac{1}{7}} = 0.378$$



$-Z_{T(0.025)} = -1.96 \quad Z_{T(0.025)} = 1.96$

$Z_H < Z_T$ olduğundan H_0 red edilemez. Buradan ilgi miktarının 0.75 olduğu söylenebilir.

e) $H_0: \rho = 0$

$H_1: \ \rho \neq 0$

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

$S_r = \sqrt{\frac{1-(0.774)^2}{10-2}} = 0.05012 \quad , \ t_H = \frac{0.774}{0.05012} = 15.443$

$t_{T\left(\frac{\alpha}{2}=0.025, n-2=8\right)} = 2.306$

$t_H > t_T$ olduğunda H_0 red edilir.

Örneklemin geldiği kitlenin korelasyon katsayısı sıfırdan farklıdır.

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

$\mu_r: 1.030 \pm (1.96)(0.378)$

$\mu_r: (0.289; 1.771)$

Buradan;

$(0.28 \leq \rho \leq 0.95)$

bulunur.

g) $R^2 = \frac{RKT}{GnKT} = \frac{(1.309)(1148)}{2510} \cong 0.60$

Bağımsız değişken bağımlı değişkenin %60'ını açıklamaktadır. Bu yeterince yüksek değildir. Bağımlı değişkeni daha iyi açıklayabilmek için başka bağımsız değişkenlere ihtiyaç duyulmaktadır.