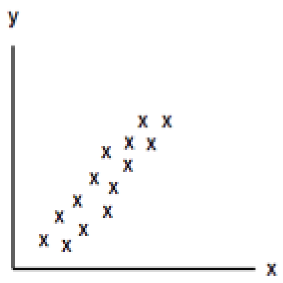
BÖLÜM 14

Korelasyon:

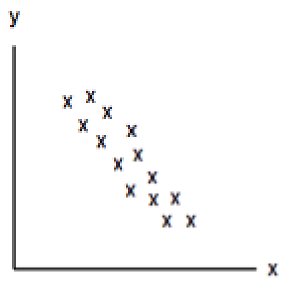
İki değişken arasındaki ilişki miktarı korelasyon katsayısı ile tespit edilir. Korelasyon çözümlemesi ise değişkenler arasındaki ilişkinin derecesini ve yönünü belirler. İki değişken arasındaki korelasyon katsayısına ‘basit korelasyon katsayısı’, ikiden çok değişken arasında korelasyon katsayısına ise ‘kısmi korelasyon katsayısı’ denir.

Kitle korelasyon katsayısı: δ

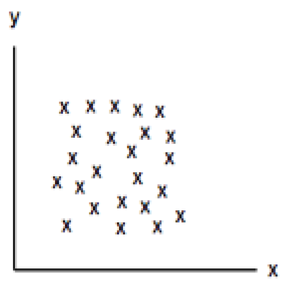
Örneklem korelasyon katsayısı: r



Her iki değişkende aynı yönde değişim gösterir. Aralarındaki değişim pozitiftir. Korelasyon katsayısı pozitiftir.



İki değişken arasında negatif yönde bir ilişki vardır. Değişkenlerden biri artarken diğeri azalacaktır. Korelasyon katsayısı negatiftir.



İlişki yok

Örneklem için korelasyon katsayısı

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sqrt{(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2)(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2)}}$$

$r = 1$ ise iki değişken arasında pozitif yönde tam ilişki vardır.

$r = -1$ ise iki değişken arasında negatif yönde tam ilişki vardır.

$-1 \leq r \leq 1$

Belirtme katsayısı $R^2 = \dfrac{SSR}{SST}$

$r = \mp\sqrt{R^2}$

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

$b_1 > 0 \ ise \ r > 0$

$b_1 < 0 \ ise \ r < 0$

dir.

Örnek: Bebeğin boy uzunluğuna annenin karın çevresinin etkili olduğu düşünülüyor. İki değişken arasında ilişki olup olmadığını saptamak için 20 bebek seçiliyor ve şu veri tablosu elde ediliyor:

X (Annenin karın çevresi):

102, 99, 106, 96, 100, 84, 104, 97, 106, 100, 92, 95, 109, 106, 100, 99, 90, 110, 116, 104

Y (Bebeğin boyu):

50, 50, 51, 49, 49, 47, 48, 46, 47, 52, 50, 46, 50, 53, 50, 52, 47, 52, 50, 51

a) Tahmini regresyon modelini bulunuz.

$$\sum_{i=1}^{20} X_i = 2015 \qquad \sum_{i=1}^{20} X_i^2 = 204053$$

$$\sum_{i=1}^{20} Y_i = 990 \qquad \sum_{i=1}^{20} Y_i^2 = 49088$$

$$\sum_{i=1}^{20} X_i Y_i = 99884$$

$$\bar{X} = 100.75, \ \bar{Y} = 49.5$$

$$\hat{Y} = b_0 + b_1 X \qquad b_1 = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}$$

$$b_1 = \frac{99884 - 20(100.75)(49.5)}{204053 - 20(100.75)^2} = \frac{141.5}{104175} = 0.135829$$

$$b_0 = \bar{Y} - b_1\bar{X} = 49.5 - 0.135829(100.75) = 35.8152$$

$$\hat{Y} = 35.8152 + 0.135829X$$

b) Annenin karın çevresi 90 cm olduğunda bebeğin tahmini boyu kaç cm 'dir?

$X = 90cm$

$\hat{Y} = 35.8152 + 0.135829(90) = 48.05cm$

c) Modelin uygunluğunu test ediniz.

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

Değişimin Kaynağı	Serbestlik Derecesi	Kareler Toplamı	Kareler Ortalaması	Test İstatistiği
Regresyon	1	$SSR = 19.2198$	$MSR = 19.2198$	$F_t = \dfrac{MSR}{MSE} = 5.424$
Hata	$n-2 \Rightarrow 18$	$SSE = 63.78$	$MSE = 3.54$	
Toplam	$n-1 \Rightarrow 19$	$SST = 83$		

$$* \ SSR = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2} = 19.2198$$

$$* \ SST = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2 = 83$$

$$* \ SSE = 63.78$$

Karar:

$F_{\alpha:k:(n-1)} = F_{\alpha:1:18} = 4.41 \qquad (\alpha : 0.05 \ olsun)$

$F_t > F_{0.05:1:18}$ olduğundan dolayı H_0 hipotezi red edilir. Model denklemine uygundur.

*Regresyon katsayısı önemlidir. Gözlemlerin model denklemine uyumu önemlidir.

d) Bağımsız değişken, bağımlı değişkenin % kaçını açıklar? (Belirtme katsayısı nedir?)

$$R^2 = \frac{SSR}{SST} = 0.231$$

e) Korelasyon katsayısını hesaplayınız.

$$r = \pm\sqrt{R^2} = \pm\sqrt{0.231} = 0.480$$

veya;

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sqrt{(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2)(\sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2)}} = \frac{99884 - 20(100.75)(49.5)}{\sqrt{(204053 - 20(100.75)^2)(49088 - 20(49.5)^2)}} = 0.480$$

Ki-Kare Çözümlemesi (χ^2)

Khi-kare istatistiği kullanılarak iki ya da daha çok değişkenin bağımsızlığının testi yapılır. İncelenen olaylar bazı durumlarda rakamla belirtilemez. Ancak niteliklerine göre sınıflanabilirler. (Örneğin; evlilik durumu, politik tercih, sağlık durumu vb.)

Örneğin; bir doktor hastalığı önleyici tedbirler ile hastalığa karşı direnmeyi, bir politikacı eğitim seviyesi ile oy tercihlerini, bir trafik mühendisi araba kazaları ile yol durumlarını incelemek isteyebilir. Bu değişkenlerin bağımsızlıklarının test edilmesinde Khi-kare çözümlenmesi kullanılır.

$$f_{ij} \rightarrow i.satır\ j.sütundaki\ gözlenen\ frekans$$

$$f_{ij}' \rightarrow i.satır\ j.sütundaki\ beklenen\ frekans$$

$$f_{ij}' = \frac{n_i \times n_j}{n}$$

$$\sum_{i=1}^{R} n_i. = \sum_{i=1}^{C} n_{.j} = n$$

$$\sum_{i=1}^{R} f_{ij} = \sum_{i=1}^{R} f_{ij}' = n._j$$

y \ x	x_1	x_2	$\cdots$	x_C	$Toplam$
y_1	f_{11}	f_{12}	$\cdots$	f_{1C}	$n_{1.}$
y_2	f_{21}	f_{22}	$\cdots$	f_{2C}	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
y_R	f_{R1}	f_{R2}	$\cdots$	f_{RC}	$n_{R.}$
$Toplam$	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.C}$	n

$$\sum_{i=1}^{R}\sum_{i=1}^{C} f_{ij} = \sum_{i=1}^{R}\sum_{i=1}^{C} f_{ij}' = n$$

H_0 : Değişkenler bağımsızdır. Gruplar arası fark yoktur.

H_1 : Değişkenler bağımsız değildir. Gruplar arası fark vardır.

Test İstatistiği:

$$\chi_t^2 = \sum_{i=1}^{R}\sum_{j=1}^{C}\frac{(f_{ij} - f_{ij}')^2}{f_{ij}'}$$

$$\chi^2_{tablo} = \chi^2_{\alpha:(R-1)(C-1)}$$

$\chi_t^2 > \chi^2_{\alpha;(R-1)(C-1)}$ olduğundan H_0 hipotezi red edilir.

Örnek: Bir ürünün Pazar araştırılmasında ürünün beğenilip beğenilmemesinin cinsiyete göre dağılımı aşağıdaki gibidir.

Beğeni	**Kadın**	**Erkek**	**Toplam**
Beğendim	90	40	130
Beğenmedim	35	110	145
Toplam	125	150	275

Rastgele seçilen 275 kişi üzerinden yapılan araştırmaya göre ürün beğenisinin cinsiyetle alakası nedir? ($\alpha : 0.05$)

H_0 : Cinsiyetten bağımsız

H_1 : Cinsiyetten bağımsız değil

$$f_{11}' = \frac{130 \times 125}{275} = 59.1$$

$$f_{12}' = 70.9$$

$$f_{21}' = 65.9$$

$$f_{22}' = 56.19$$

$$\chi_t^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(f_{ij} - f_{ij}'\right)^2}{f_{ij}'} = \frac{(90-59.1)^2}{59.1} + \frac{(40-70.9)^2}{70.9} + \frac{(35-65.9)^2}{65.9} + \frac{(110-79.1)^2}{79.1}$$

$$= 56.19$$

Karar:

$$\chi^2_{\alpha:(R-1)(C-1)} = \chi^2_{0.05:(2-1)(2-1)}$$

$$= \chi^2_{0.05:1} = 3.84$$

$\chi_t^2 > \chi^2_{0.05:1} = 3.84$ olduğu için H_0 hipotezi red edilir.

KAYNAKLAR

1. Uygulamalı İstatistik (1994)

Ayşen APAYDIN , Alaettin KUTSAL, Cemal ATAKAN

2. Olasılık ve İstatistik Problemler ve Çözümleri ile (2008)

Prof. Dr. Semra ERBAŞ

3. Olasılık ve İstatistik (2006)

Prof. Dr. Fikri Akdeniz

4. Olasılık ve İstatistiğe Giriş I-II (2011)

Prof. Dr. Fikri Öztürk

5. Fikri Öztürk web sitesi

http://80.251.40.59/science.ankara.edu.tr/ozturk/index.html