

## Regresyon Analizi Nedir?

Regresyon analizi, bir veya daha fazla bağımsız değişken (açıklayıcı değişkenler) ile bir bağımlı değişken (tahmin edilen değişken) arasındaki ilişkiyi inceleyen istatistiksel bir yöntemdir. Bu analiz, veriler arasındaki ilişkileri modellemek ve bağımsız değişkenlerin bağımlı değişken üzerindeki etkilerini anlamak için kullanılır.

## Basit Doğrusal Regresyon Modeli Nedir?

Basit doğrusal regresyon modeli, yalnızca bir bağımsız değişken içeren regresyon analizidir. Model, bağımlı değişken  $YYY$  ile bağımsız değişken  $XXX$  arasındaki doğrusal ilişkiyi şu formülle ifade eder:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Burada:

- $\beta_0$  kesme noktası (intercept),
- $\beta_1$  eğim (slope),
- $\epsilon$  hata terimidir (modelin açıklayamadığı rastgele sapma).

Basit doğrusal regresyon modelinin varsayımları:

1. **Doğrusallık:** Bağımlı ve bağımsız değişkenler arasında doğrusal bir ilişki olmalıdır.
2. **Homoskedastisite:** Hataların varyansı sabit olmalıdır.
3. **Bağımsızlık:** Gözlemler birbirinden bağımsız olmalıdır.
4. **Normal Dağılım:** Hata terimleri normal dağılıma sahip olmalıdır.

## 1. Adım: Regresyon Denklemine Belirleme

Basit doğrusal regresyon denklemi:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Verilen veri setine göre:

$X$	$Y$
1	2
2	4
3	5
4	4
5	5

## 2. Adım: Ortalama Değerleri Hesaplama

Öncelikle,  $X$  ve  $Y$  için ortalamaları hesaplayalım:

$$\bar{X} = \frac{1 + 2 + 3 + 4 + 5}{5} = 3$$

$$\bar{Y} = \frac{2 + 4 + 5 + 4 + 5}{5} = 4$$

### 3. Adım: Katsayıların Hesaplanması

a) Eğim Katsayısı ( $\beta_1$ )

Eğim katsayısı formülü:

$$\beta_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

İlgili hesaplamalar:

$X_i$	$Y_i$	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$
1	2	-2	-2	4	4
2	4	-1	0	0	1
3	5	0	1	0	0
4	4	1	0	0	1
5	5	2	1	2	4

Toplamlar:

$$\sum(X_i - \bar{X})(Y_i - \bar{Y}) = 4 + 0 + 0 + 0 + 2 = 6$$

$$\sum(X_i - \bar{X})^2 = 4 + 1 + 0 + 1 + 4 = 10$$

Eğim katsayısı:

$$\beta_1 = \frac{6}{10} = 0.6$$

b) Kesme Noktası ( $\beta_0$ )

Kesme noktası formülü:

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}$$

$$\beta_0 = 4 - 0.6 \times 3 = 2.2$$

Regresyon denklemi:

$$\hat{Y} = 2.2 + 0.6X$$

#### 4. Adım: Tahminler ve Artıklar (Residuals)

Tahmin edilen ( $\hat{Y}$ ) ve artıklar ( $e = Y - \hat{Y}$ ) değerlerini hesaplayalım:

$X$	$Y$	$\hat{Y}$ (Tahmin)	$e = Y - \hat{Y}$
1	2	$2.2 + 0.6 \times 1 = 2.8$	$2 - 2.8 = -0.8$
2	4	$2.2 + 0.6 \times 2 = 3.4$	$4 - 3.4 = 0.6$
3	5	$2.2 + 0.6 \times 3 = 4.0$	$5 - 4.0 = 1.0$
4	4	$2.2 + 0.6 \times 4 = 4.6$	$4 - 4.6 = -0.6$
5	5	$2.2 + 0.6 \times 5 = 5.2$	$5 - 5.2 = -0.2$

## 5. Adım: T-istatistikleri Hesaplama

### a) Standart Hata

Önce hata kareler toplamını (SSE) hesaplayalım:

$$\begin{aligned}SSE &= \sum (e^2) = (-0.8)^2 + (0.6)^2 + (1.0)^2 + (-0.6)^2 + (-0.2)^2 \\SSE &= 0.64 + 0.36 + 1.00 + 0.36 + 0.04 = 2.4\end{aligned}$$

Hata varyansı ( $s^2$ ):

$$s^2 = \frac{SSE}{n - 2} = \frac{2.4}{5 - 2} = 0.8$$

Bağımsız değişkenlerin standart hatası:

$$SE(\beta_1) = \sqrt{\frac{s^2}{\sum (X_i - \bar{X})^2}} = \sqrt{\frac{0.8}{10}} = 0.283$$

Kesme noktasının standart hatası:

$$SE(\beta_0) = \sqrt{s^2 \left( \frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right)} = \sqrt{0.8 \left( \frac{1}{5} + \frac{3^2}{10} \right)} = 0.938$$

### b) t-istatistikleri

t-istatistik formülü:

$$t = \frac{\beta}{SE(\beta)}$$

- $\beta_0 = 2.2, SE(\beta_0) = 0.938,$

$$t = \frac{2.2}{0.938} = 2.345$$

- $\beta_1 = 0.6, SE(\beta_1) = 0.283,$

$$t = \frac{0.6}{0.283} = 2.121$$

## 6. Adım: F-Testi Hesaplama

F-istatistik formülü:

$$F = \frac{SSR/k}{SSE/(n - k - 1)}$$

*SSR* (Regresyon Kareler Toplamı):

$$\begin{aligned} SSR &= \sum (\hat{Y} - \bar{Y})^2 = (2.8 - 4)^2 + (3.4 - 4)^2 + (4 - 4)^2 + (4.6 - 4)^2 + (5.2 - 4)^2 \\ &= 1.44 + 0.36 + 0 + 0.36 + 1.44 = 3.6 \end{aligned}$$

F-istatistiği:

$$F = \frac{3.6/1}{2.4/3} = \frac{3.6}{0.8} = 4.5$$

## Sonuç

- Regresyon Denklemi:  $\hat{Y} = 2.2 + 0.6X$
- Katsayılar:  $\beta_0 = 2.2, \beta_1 = 0.6$
- t-istatistikleri: Kesme noktası: 2.345, Eğim: 2.121
- F-istatistiği: 4.5

## İktisadi Bir Örnek: Tüketim ve Gelir Arasındaki İlişki

Bir ekonomide hane halkının geliri ( $X$ ) ve buna karşılık gelen tüketim harcamaları ( $Y$ ) arasındaki ilişkiyi inceleyelim. 5 hane halkı için veri setimiz şu şekilde olsun:

Hanehalkı Geliri ( $X$ , Bin TL)	Tüketim Harcaması ( $Y$ , Bin TL)
20	15
30	18
40	20
50	24
60	27

Bu veri setini kullanarak basit doğrusal regresyon modelinin aşamalarını gerçekleştirelim.

### 1. Adım: Ortalama Değerlerin Hesaplanması

Öncelikle  $X$  ve  $Y$  ortalamalarını bulalım:

$$\bar{X} = \frac{20 + 30 + 40 + 50 + 60}{5} = 40$$

$$\bar{Y} = \frac{15 + 18 + 20 + 24 + 27}{5} = 20.8$$

## 2. Adım: Katsayıların Hesaplanması

### a) Eğim Katsayısı ( $\beta_1$ )

Eğim katsayısının formülü:

$$\beta_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

İlgili hesaplamalar:

$X_i$	$Y_i$	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$
20	15	-20	-5.8	116	400
30	18	-10	-2.8	28	100
40	20	0	-0.8	0	0
50	24	10	3.2	32	100
60	27	20	6.2	124	400

Toplamlar:

$$\sum(X_i - \bar{X})(Y_i - \bar{Y}) = 116 + 28 + 0 + 32 + 124 = 300$$

$$\sum(X_i - \bar{X})^2 = 400 + 100 + 0 + 100 + 400 = 1000$$

Eğim katsayısı:

$$\beta_1 = \frac{300}{1000} = 0.3$$



## b) Kesme Noktası ( $\beta_0$ )

Kesme noktası formülü:

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}$$

$$\beta_0 = 20.8 - 0.3 \times 40 = 20.8 - 12 = 8.8$$

Regresyon denklemi:

$$\hat{Y} = 8.8 + 0.3X$$

## 3. Adım: Tahminler ve Artıklar (Residuals)

Tahmin edilen ( $\hat{Y}$ ) ve artıklar ( $e = Y - \hat{Y}$ ):

$X$	$Y$	$\hat{Y}$ (Tahmin)	$e = Y - \hat{Y}$
20	15	$8.8 + 0.3 \times 20 = 14.8$	$15 - 14.8 = 0.2$
30	18	$8.8 + 0.3 \times 30 = 17.8$	$18 - 17.8 = 0.2$
40	20	$8.8 + 0.3 \times 40 = 20.8$	$20 - 20.8 = -0.8$
50	24	$8.8 + 0.3 \times 50 = 23.8$	$24 - 23.8 = 0.2$
60	27	$8.8 + 0.3 \times 60 = 26.8$	$27 - 26.8 = 0.2$

#### 4. Adım: Standart Hata ve t-istatistikleri Hesaplanması

a) Hata Kareler Toplamı (SSE)

$$SSE = \sum (e^2) = (0.2)^2 + (0.2)^2 + (-0.8)^2 + (0.2)^2 + (0.2)^2$$

$$SSE = 0.04 + 0.04 + 0.64 + 0.04 + 0.04 = 0.8$$

Hata varyansı ( $s^2$ ):

$$s^2 = \frac{SSE}{n - 2} = \frac{0.8}{5 - 2} = 0.267$$

Standart hata:

- $SE(\beta_1) = \sqrt{\frac{s^2}{\sum(X_i - \bar{X})^2}} = \sqrt{\frac{0.267}{1000}} = 0.0163$
- $SE(\beta_0) = \sqrt{s^2 \left( \frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i - \bar{X})^2} \right)} = 1.981$

b) t-istatistikleri

$$t_{\beta_0} = \frac{\beta_0}{SE(\beta_0)} = \frac{8.8}{1.981} \approx 4.44$$

$$t_{\beta_1} = \frac{\beta_1}{SE(\beta_1)} = \frac{0.3}{0.0163} \approx 18.4$$

## 5. Adım: F-Testi Hesaplaması

Regresyon kareler toplamı ( $SSR$ ):

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

Hesaplamalar:

$$SSR = (14.8 - 20.8)^2 + (17.8 - 20.8)^2 + (20.8 - 20.8)^2 + (23.8 - 20.8)^2 + (26.8 - 20.8)^2$$

$$SSR = 36$$

F-istatistiği:

$$F = \frac{SSR/k}{SSE/(n - k - 1)} = \frac{36/1}{0.8/3} = 135$$

## 6. Adım: Modelin ve Katsayıların İstatistiksel Anlamlılığı

- Katsayıların Anlamlılığı:
  - $t$ -istatistikleri p-değerleri açısından değerlendirildiğinde, genellikle  $t > 2$  olması %95 güvenle anlamlı olduğunu gösterir.
  - Burada  $t_{\beta_0} = 4.44$  ve  $t_{\beta_1} = 18.4$  olduğu için katsayılar istatistiksel olarak anlamlıdır.
- Modelin Anlamlılığı:
  - F-istatistiği 135, çok yüksek ve modelin genel olarak anlamlı olduğunu gösterir ( $F > 4$  genelde anlamlıdır).

Sonuç: Gelir ve tüketim arasındaki doğrusal ilişki istatistiksel olarak anlamlıdır.