6.5 Basit Doğrusal Regresyonda Hipotez Testleri

6.5.1 $\boldsymbol{\beta_1}$ İçin Hipotez Testi:

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

1. Hipotez kurulur.

$$H_0: \beta_1 = \beta_{1.0}$$
$$H_1: \beta_1 \neq \beta_{1.0}$$

2. Test istatistiği hesaplanır.

$\varepsilon_J \sim N(0, \sigma^2)$ olduğu biliniyor buna göre;

$y_j = \beta_0 + \beta_1 x_j + \varepsilon_j$

$y_j \sim N(\beta_0 + \beta_1 x_j, \sigma^2)$

$b_1 \sim N(\beta_1, \frac{\sigma^2}{\sum (x_j - \bar{x})^2})$

H_0 hipotezinin doğruluğu altında test istatistiği

$t_H = \frac{b_1 - \beta_{1.0}}{S_{b_1}}$

$S_{b_1} \rightarrow b_1$' in standart hatası

$S_{b_1} = \sqrt{\frac{RAKO}{XOAKT}}$

3. Karar verilir ve yorum yapılır.

$|t_H| > t_T$ ise H_0 red edilir.

$t_T = t_T\,(\alpha/2\,, n-2)$

6.5.2 β_0 İçin Hipotez Testi:

1. Hipotez kurulur.

$H_0: \beta_0 = \beta_{0.0}$
$H_1: \beta_0 \neq \beta_{0.0}$

2. Test istatistiği hesaplanır.

$$t_H = \frac{b_0 - \beta_{0.0}}{S_{b_0}} \qquad , \qquad S_{b_0} = \sqrt{\mathrm{RAKO}\left[\frac{1}{n} + \frac{\bar{x}^2}{\mathrm{XOAKT}}\right]}$$

3. Karar verilir ve yorum yapılır.

$|t_H| > t_T\ (^{\alpha}/_{2}\ , n-2)$ ise H_0 red edilir.

6.5.3 Regresyon Doğrusunun Anlamlılık Testi

1. β_1' in sıfıra eşit olup olmadığının testidir.

$H_0: \beta_1 = 0$ (Regresyon doğrusu önemsizdir.)

$H_1: \beta_1 \neq 0$ (Regresyon doğrusu önemlidir.)

2.

$$t_H = \frac{b_1}{S_{b_1}} \ , \qquad S_{b_1} = \sqrt{\frac{\mathrm{RAKO}}{\mathrm{XOAKT}}}$$

3. $|t_H| > t_T\ (^{\alpha}/_{2}\ , n-2)$ ise H_0 red edilir.

H_0 hipotezi red edilirse regresyon doğrusunun anlamlı olduğu söylenebilir. H_0 hipotezi red edilemezse regresyon doğrusu anlamsızdır. İki değişken arasında doğrusal bir ilişki olmadığı söylenir. Bu hipotez F testi ile de yapılabilir.

1. "Deneysel noktaların doğrusal regresyona uyumu önemsizdir" ya da "deneysel noktalar regresyon doğrusu ile gösterilemez" şeklinde yorumlanabilen H_0 hipotezi $\beta_1 = 0$ olarak kurulur.

$H_0: \beta_1 = 0$
$H_1: \beta_1 \neq 0$

2. Bu hipotezi test etmek amacıyla varyans analizi tablosu yapılır.

Varyans Analizi Tablosu

Değişim Kaynakları(DK)	**Serbestlik Derecesi(sd)**	**Kareler Toplamı (KT)**	**Kareler Ortalaması (KO)**	**Test**
Regresyon	1	RKT=b_1XOAKT	RKO=RKT/1	$F_H = \frac{RKO}{RAKO}$
Regresyondan Ayrılış	n-2	RAKT=YOAKT-RKT	RAKO=RAKT/n-2	
Toplam	n-1	YOAKT		

3. $F_H > F_T(\alpha, 1, n-2)$ ise H_0 hipotezi red edilir.

$t^2 = F_{(1,n-2)}$ dir.

$$t_H = t = \frac{b_1}{S_{b_1}} = \frac{b_1}{\sqrt{RAKO/XOAKT}}$$

$$t^2 = \frac{b_1^2}{RAKO/XOAKT} = \frac{b_1^2 XOAKT}{RAKO} = \frac{\left(\frac{XYOAÇT}{XOAKT}\right)^2 XOAKT}{RAKO} = \frac{\left(\frac{XYOÇAT}{XOAKT}\right)XOAKT}{RAKO}$$

$$= \frac{b_1 XYOAÇT}{RAKO} = \frac{RKT/1}{RAKT/n-2} = F$$

Örnek 6.2 12 kadına ilişkin yaş ve sistolik kan basıncı arasındaki doğrusal regresyon denklemi;

$$\hat{y} = 80.778 + 1.138x$$

olarak bulunmuştu.

a) $H_0: \beta_0 = 0$ hipotezini %95 güven düzeyinde test ediniz.
b) $H_0: \beta_1 = 0$ hipotezini %95 güven düzeyinde hem t hem de F testi ile test ediniz.

$\sum x_j = 628$,$\sum x_j^2 = 34416$,$\sum y_j = 1684$,$\sum y_j^2 = 238822$,$\sum x_j y_j = 89894$

a)
1. $H_0: \beta_0 = 0$
$H_1: \beta_0 \neq 0$

2. $t_H = \frac{b_0 - \beta_{0.0}}{s_{b_0}} = \frac{b_0}{s_{b_0}}$

$$s_{b_0} = \sqrt{\text{RAKO}\left[\frac{1}{n} + \frac{\bar{x}^2}{\text{XOAKT}}\right]}$$

$$\text{XOAKT} = \sum x_j^2 - \frac{\left(\sum x_j\right)^2}{n} = 1550.67$$

$$\text{YOAKT} = \sum y_j^2 - \frac{\left(\sum y_j\right)^2}{n} = 2500.667$$

$$\text{XYOAÇT} = \sum x_j y_j - \frac{\sum x_j \sum y_j}{n} = 1764.67$$

$$\text{RKT} = b_1 \text{XYOAÇT} = 2008.194$$

$$\text{RAKT} = \text{YOAKT} - \text{RKT} = 2500.667 - 2008.194 = 492.473 \Longrightarrow \text{RAKO} = \frac{492.473}{10}$$

$$S_{b_0} = \sqrt{\text{RAKO}\left[\frac{1}{n} + \frac{\bar{x}^2}{\text{XOAKT}}\right]} = \sqrt{\frac{492.473}{10}\left(\frac{1}{12} + \frac{(52.33)^2}{1550.67}\right)} = \sqrt{91.07} = 9.54$$

$$t_H = \frac{b_0}{S_{b_0}} = \frac{80.778}{9.54} = 8.467$$

3. $t_T(0.025\,, \text{sd} = 10) = 2.228$

$t_H > t_T$ olduğundan H_0 red edilir. Yani, β_0 katsayısı önemlidir.

b) **1.** $H_0: \beta_1 = 0$
$H_1: \beta_1 \neq 0$

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

$$S_{b_1} = \sqrt{\frac{\text{RAKO}}{\text{XOAKT}}} = \sqrt{\frac{492.473/10}{1550.67}} = 0.173$$

$$t_H = \frac{b_1}{S_{b_1}} = \frac{1.138}{0.173} = 6.578$$

3. $t_T(0.025\,, \text{sd} = n - 2 = 10) = 2.228$

$t_H > t_T$ 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	2008.194	2008.194	$F_H = \dfrac{2008.194}{49.2473} = 40.778$
Regresyondan ayrılış	n-2=10	492.473	49.2473	
Toplam	n-1=11	83		

3. $F_T(0.05\,, sd1 = 1, sd2 = n - 2 = 10) = 4.96$

$F_H > F_T$ olduğundan H_0 red edilir. Yani, regresyon doğrusu önemlidir. Kadınlarda sistolik kan basıncı ile yaş arasında pozitif bir ilişkiden söz edilebilir ($b_1 > 0$ olduğu için).

6.6. Basit Doğrusal Regresyonda Aralık Tahmini

$\varepsilon_j \sim N(0, \sigma^2) \longrightarrow$ Bağımsız ve normal dağılıma sahip

$\frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}} \sim t_{n-2} \quad , \quad \frac{\hat{\beta}_0 - \beta_0}{s_{\hat{\beta}_0}} \sim t_{n-2}$

6.6.1. β_1'in Güven Aralığı

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

$t_T = t_T(\alpha/2\,, sd = n - 2)$

6.6.2. β_0'ın Güven Aralığı

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

$t_T = t_T(\alpha/2\,, sd = n - 2)$

6.6.3. Bilinen bir x_0 değerine karşılık y değerinin ortalaması için güven aralığı verilebilir. x_0 bilindiğinde y değerinin ortalamasını tanımlayan ifade $E(Y|x_0)$ ile gösterilir.

$E(Y|x_0)$' ın güven aralığı;

$P\left(\hat{y}_0 - t_T s_{\hat{y}_0} \le E(Y|x_0) \le \hat{y}_0 + t_T s_{\hat{y}_0}\right) = 1-\alpha$

$\hat{y}_0 = b_0 + b_1 x_0$' nın hesaplanan değerdir.

$$S_{\hat{y}_0} = \sqrt{RAKO\left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{XOAKT}\right]}$$

$t_T = t_T(\alpha/2\,, sd = n-2)$

6.6.4. Bilinen bir x_0 değerine karşılık y'nin yeni ya da gelecekteki değerini tahmin etmek, regresyon çözümlemesinde önemlidir.X Rasgele değişkeninin x_0 gibi bir değeri verildiğinde Y Rasgele değişkeninin y_0 gibi bir özel değeri için aralık tahmini verilebilir.

$(y_0 - \hat{y}_0)$ rasgele değişkeni ortalaması sıfır, standart sapması $S_{y_0-\hat{y}_0}$ olan normal dağılıma sahiptir.

$$S_{y_0-\hat{y}_0} = \sqrt{RAKO\left[1 + \frac{1}{n} + \frac{(x_o-\bar{x})^2}{XOAKT}\right]}$$

y_0' ın güven aralığı;

$P\left(\hat{y}_0 - t_T S_{y_0-\hat{y}_0} \le y_0 \le \hat{y}_0 + t_T S_{y_0-\hat{y}_0}\right) = 1-\alpha$

$\hat{y}_0 = b_0 + b_1 x_0$

$t_T = t_T(\alpha/2\,, sd = n-2)$

Örnek 6.3. 12 kadına ait yaş ve sistolik kan basıncı ile ilgili verilere ilişkin

a) β_1'in güven aralığını tahmin ediniz.
b) β_0'ın güven aralığını tahmin ediniz.
c) $x_o = 45$ değeri için y değerinin ortalaması için güven aralığını tahmin ediniz.
d) $x_o = 40$ için y_0'ın güven aralığını tahmin ediniz.

a) β_1 için;

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

$(1.138 - 2.228(0.173) \le \beta_1 \le 1.138 + 2.228(0.173))$

$\beta_1 \epsilon [0.7526; 1.5234]$

b) β_0 için;

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

$(80.778 - 2.228(9.54) \leq \beta_0 \leq (80.778 + 2.228(9.54))$

$\beta_0 \epsilon [59.523; 102.033]$

c) $x_o = 45$ için $\hat{y} = 80.778 + 1.138(45) = 131.988$

$$S_{\hat{y}_0} = \sqrt{\mathrm{RAKO}\left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\mathrm{XOAKT}}\right]}$$

$$= \left[\frac{492.473}{10}\left(\frac{1}{12} + \frac{(45 - 52.33)^2}{1550.67}\right)\right]^{1/2} = 2.41$$

$P(\hat{y}_0 - t_T S_{\hat{y}_0} \leq E(Y|x_0) \leq \hat{y}_0 + t_T S_{\hat{y}_0}) = 1 - \alpha$

$(131.988 - 2.228(2.41) \leq E(Y|x_0) \leq 131.988 + 2.228(2.41))$

$E(Y|x_0 = 45) \epsilon [126.618; 137.357]$

Bu aralığın yaşı 45 olan bir kadının ortalama sistolik kan basıncını içeren aralıklardan biri olması olasılığı %95'tir.

d) $x_o = 40$ için $\hat{y} = 80.778 + 1.138(40) = 126.299$

$$S_{y_0 - \hat{y}_0} = \sqrt{\mathrm{RAKO}\left[1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{\mathrm{XOAKT}}\right]}$$

$$= \left[\frac{492.473}{10}\left(1 + \frac{1}{12} + \frac{(45 - 52.33)^2}{1550.67}\right)\right]^{1/2} = 7.42$$

$P(\hat{y}_0 - t_T S_{y_0 - \hat{y}_0} \leq y_0 \leq \hat{y}_0 + t_T S_{y_0 - \hat{y}_0}) = 1 - \alpha$

$(126.299 - 2.228(7.42) \leq y_0 \leq 126.299 + 2.228(7.42))$

$y_0 \epsilon [109.767; 142.831]$

Bu aralığın yaşı 40 olan bir kadının sistolik kan basıncını içeren aralıklardan biri olması olasılığı %95'tir.

Örnek 6.4. 16 bebeğin yaş ve vücut ağırlıkları aşağıda verilmiştir:

Yaş(ay) (x) :	2	2	3	4	3	5	6	5	7	8	8	7	9	10	10	9
Ağırlık(kg): (y)	4.2	3.6	4.4	5.2	4.0	5.8	5.7	6.0	7.0	7.2	7.0	6.9	7.5	8.0	7.8	7.9

a) Serpilme diyagramını çiziniz.

b) $Y = \beta_0 + \beta_1 x + \varepsilon$ doğrusal regresyon denklemini tahmin ediniz.

c) $x = 15$ ay için çocuğun ağırlığını tahmin ediniz.

d) Bağımsız değişkenin bağımlı değişkendeki değişimin yüzde kaçını açıkladığını ifade ediniz.

e) Regresyon katsayılarının önem kontrolünü yapınız.(β_1 için t ve F ile ($\alpha = 0.05$))

f) β_0 ve β_1 için %95 güven düzeyinde güven aralıklarını oluşturunuz.

g) 12 aylığa karşı gelen ortalama ve öngörülen ağırlık için %95 güven sınırlarını bulunuz.

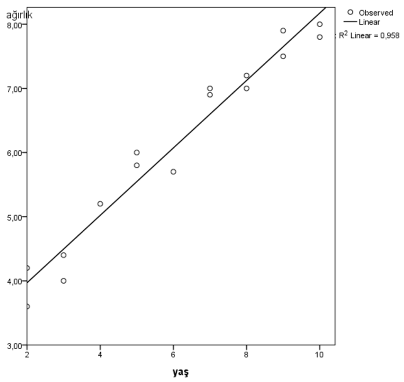
Çözüm:

$\sum x_j = 98, \quad \sum x_j^2 = 716, \quad \bar{x} = 6.125$

$\sum y_j = 98.2, \quad \sum y_j^2 = 636.08, \quad \bar{y} = 6.138$

$\sum x_j y_j = 662.3$

a)



b) $\text{YOAKT} = \sum y_j^2 - \frac{\left(\sum y_j\right)^2}{n} = 636.08 - \frac{(98.2)^2}{16} = 33.378$

$\text{XOAKT} = \sum x_j^2 - \frac{\left(\sum x_j\right)^2}{n} = 716 - \frac{(98)^2}{16} = 115.750$

$\text{XYOAÇT} = \sum x_j y_j - \frac{\sum x_j \sum y_j}{n} = 662.3 - \frac{(98)(98.2)}{16} = 60.825$

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

$$b_1 = \frac{XYOAÇT}{XOAKT} == \frac{60.825}{115.750} = 0.525$$

$$b_0 = \bar{y} - b_1\bar{x} = 6.138 - 0.525(6.125) = 2.922$$

$$\hat{y} = b_0 + b_1x$$

$$\hat{y} = 2.922 + 0.525x$$

c) $\hat{y} = 2.922 + 0.525(15) = 10.797$

d) $R^2 = \frac{RKT}{YOAKT} = \frac{b_1XYOAÇT}{YOAKT} = \frac{31.933}{33.378} = 0.957 \cong 0.96$

Bağımsız değişken, bağımlı değişkendeki değişimin %96'sını açıklamaktadır.

$$RKT = b_1XYOAÇT = (0.525)(60.825) = 31.933$$

$$RAKT = YOAKT - RKT = 33.378 - 31.933 = 1.445$$

$$RAKO = \frac{1.445}{n-2} = \frac{1.445}{14} = 0.103$$

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

2. $t_H = \frac{b_0}{s_{b_0}}$

$$s_{b_0} = \sqrt{RAKO\left[\frac{1}{n} + \frac{\bar{x}^2}{XOAKT}\right]} = \sqrt{0.103\left(\frac{1}{16} + \frac{(6.125)^2}{115.750}\right)} = \sqrt{0.0398} = 0.1996$$

$$t_H = \frac{2.922}{0.1996} = 14.639$$

3. $t_{T(\alpha/2=0.025\,,n-2=14)} = 2.145$

$t_H = 14.639 > t_T = 2.145$ olduğundan H_0 red edilir. β_0 katsayısı önemlidir.

1. $H_0: \beta_1 = 0$
$H_1: \beta_1 \neq 0$

2. $t_H = \frac{b_1}{s_{b_1}}$, $\quad s_{b_1} = \sqrt{\frac{RAKO}{XOAKT}} = \sqrt{\frac{0.103}{115.750}} = 0.0298$

$$t_H = \frac{0.525}{0.0298} = 17.617$$

3. $t_{T(\alpha/2=0.025\,,n-2=14)} = 2.145$

$t_H = 17.617 > t_T = 2.145$ olduğundan H_0 red edilir. β_1 katsayısı anlamlıdır. Regresyon doğrusu anlamlıdır.

F Testi

$H_0: \beta_1 = 0$
$H_1: \beta_1 \neq 0$

DK	Sd	KT	KO	Test
Regresyon	1	31.933	31.933	$F_H = \dfrac{31.933}{0.103} = 310.029$
Regresyondan ayrılış	n-2=14	1.445	0.103	
Toplam	n-1=15	33.378	-	

$F_{T(0.05, sd_1=1,\, sd_2=14)} = 4.60$

$F_H = 310.029 > F_T = 4.60$ olduğundan H_0 red edilir. Yani β_1 katsayısı önemli ve regresyon doğrusu anlamlıdır.

f) β_0 için güven aralığı;

$P\left(b_0 - t_{T(\alpha/2,n-2)} S_{b_0} \leq \beta_0 \leq b_0 + t_{T(\alpha/2,n-2)} S_{b_0}\right) = 1 - \alpha$

$$b_0 \pm t_{T(0.025,14)} S_{b_0}) = 2.922 \mp 2.145(0.1996) \begin{array}{l} \nearrow 2.494 \\ \longrightarrow 3.350 \end{array}$$

$\beta_0: [2.494; 3.350]$ Bu aralığın β_0'ı içeren aralıklardan biri olması olasılığı %95'dir.

β_1 için güven aralığı;

$P\left(b_1 - t_{T(\alpha/2,n-2)} S_{b_1} \leq \beta_1 \leq b_1 + t_{T(\alpha/2,n-2)} S_{b_1}\right) = 1 - \alpha$

$$b_1 \pm t_{T(0.025,14)} S_{b_1}) = 0.525 \mp 2.145(0.0298) \begin{array}{l} \nearrow 0.461 \\ \longrightarrow 0.589 \end{array}$$

$\beta_1: [0.461; 0.589]$ Bu aralığın β_1'i içeren aralıklardan biri olması olasılığı %95'dir.

g) $x_o = 12$ için $\hat{y}_0 = 2.922 + 0.525(12) = 9.222$

$$s_{\hat{y}_0} = \sqrt{\text{RAKO}\left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\text{XOAKT}}\right]} = \sqrt{0.103\left(\frac{1}{16} + \frac{(12-6.125)^2}{115.750}\right)} = 0.193$$

$$P\left(\hat{y}_0 - t_{T(\alpha/2,n-2)}s_{\hat{y}_0} \leq E(y|x_0) \leq \hat{y}_0 + t_{T(\alpha/2,n-2)}s_{\hat{y}_0}\right) = 1 - \alpha$$

$$\hat{y}_0 \mp t_{T(0.025,14)}s_{\hat{y}_0} = 9.222 \mp 2.145(0.193) \begin{array}{l} \nearrow 8.808 \\ \rightarrow 9.636 \end{array}$$

$E(y|x_0 = 12)$: [8.808; 9.636] Bu aralığın 12 aylık olan bebeklerde ortalama ağırlığı içeren aralıklardan biri olması olasılığı %95'tir.

$x_o = 12, \quad \hat{y}_0 = 9.222$

$$s_{y_0-\hat{y}_0} = \sqrt{\text{RAKO}\left[1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{\text{XOAKT}}\right]} = \sqrt{0.103\left(1 + \frac{1}{16} + \frac{(12-6.125)^2}{115.750}\right)} = 0.374$$

$$P\left(\hat{y}_0 - t_{T(\alpha/2,n-2)}s_{y_0-\hat{y}_0} \leq y_0 \leq \hat{y}_0 + t_{T(\alpha/2,n-2)}s_{y_0-\hat{y}_0}\right) = 1 - \alpha$$

$$\hat{y}_0 \mp t_{T(0.025,14)}s_{y_0-\hat{y}_0} = 9.222 \mp 2.145(0.374) \begin{array}{l} \nearrow 8.420 \\ \rightarrow 10.024 \end{array}$$

y_0: [8.420; 10.024] Bu aralığın 12 aylık olan bebeklerde öngörülen ağırlığı içeren aralıklardan biri olması olasılığı %95'tir.