İST 417 Lineer Modeller – 13. Hafta

$\boldsymbol{\beta_k}$ için Hipotez Testi:

$H_0: \beta_k = 0$

$H_1: \beta_k \neq 0$ hipotezini test etmek için

$t = \frac{\widehat{\beta_k} - 0}{\sqrt{Var(\widehat{\beta_k})}}$ istatistiği kullanılır.

Eğer, $|t| > t_{\left(1-\frac{\alpha}{2}\right)}(n-p)$ $\Rightarrow$ H_0 reddedilir

ya da $t > t_{\left(1-\frac{\alpha}{2}\right)}(n-p)$ $veya$ $t < -t_{\left(1-\frac{\alpha}{2}\right)}(n-p)$ $\Rightarrow$ H_0 reddedilir

Not: $H_0: \beta_k = 0$ hipotezini alternatif olarak genel lineer test yaklaşımı kullanarak da test edebiliriz.

Aşağıda verilen hipotezi

$H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$

bu yaklaşımı kullanarak test edelim.

$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$ Tam model (full model)

$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$ İndirgenmiş model (Reduced model)

$SSE(F) = SSE(X_1, X_2, X_3) \rightarrow df_F = n - 4$

$SSE(R) = SSE(X_1, X_2) \rightarrow df_R = n - 3$

$$F = \frac{\frac{SSE(R) - SSE(F)}{df_R - df_F}}{\frac{SSE(F)}{df_F}} = \frac{\frac{SSE(X_1, X_2) - SSE(X_1, X_2, X_3)}{n - 3 - n + 4}}{\frac{SSE(X_1, X_2, X_3)}{n - 4}} = \frac{\frac{SSR(X_3|X_1, X_2)}{1}}{\frac{SSE(X_1, X_2, X_3)}{n - 4}}$$

$$= \frac{MSR(X_3|X_1, X_2)}{MSE(X_1, X_2, X_3)} \sim F_{(1-\alpha)}(1, n - 4)$$

Aynı model için

$H_0: \beta_2 = \beta_3 = 0$

$H_1: En\ az\ biri\ sıfırdan\ farklıdır.$

$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$ Tam model (full model)

$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i$ İndirgenmiş model (Reduced model)

$$F = \frac{\frac{SSE(X_1) - SSE(X_1, X_2, X_3)}{n - 2 - n + 4}}{\frac{SSE(X_1, X_2, X_3)}{n - 4}} = \frac{\frac{SSR(X_2, X_3|X_1)}{2}}{\frac{SSE(X_1, X_2, X_3)}{n - 4}}$$

$$= \frac{MSR(X_2, X_3|X_1)}{MSE(X_1, X_2, X_3)} \sim F_{(1-\alpha)}(2, n - 4)$$

Benzer şekilde

$H_0: \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$

$H_1: En\ az\ biri\ sıfırdan\ farklıdır.$

$$F = \frac{\frac{SSR(X_1, X_2, \ldots, X_{p-1})}{p - 1}}{\frac{SSE(X_1, X_2, \ldots, X_{p-1})}{n - p}} = \frac{MSR}{MSE} \sim F_{(1-\alpha)}(p - 1, n - p)$$

$H_0: \beta_k = 0$

$H_1: \beta_k \neq 0$

$$F = \frac{\frac{SSR\left(X_k \middle| X_1, X_2, \ldots, X_{k-1}, X_{k+1}, \ldots X_{p-1}\right)}{1}}{\frac{SSE\left(X_1, \ldots, X_{p-1}\right)}{n - p}}$$

$$= \frac{MSR\left(X_k \middle| X_1, X_2, \ldots, X_{k-1}, X_{k+1}, \ldots X_{p-1}\right)}{MSE} \sim F_{(1-\alpha)}(1, n - p)$$

Ya da denk olarak

$$t = \frac{\widehat{\beta_k} - 0}{\sqrt{Var(\widehat{\beta_k})}} \sim t_{(n-p)}$$

$H_0: \beta_q = \beta_{q+1} = \cdots = \beta_{p-1} = 0$

$H_1: En\ az\ biri\ sıfırdan\ farklıdır.$

$$F = \frac{\frac{SSR(X_q, \dots, X_{p-1} | X_1, \dots, X_{q-1})}{p-q}}{\frac{SSE(X_1, X_2, \dots, X_{p-1})}{n-p}}$$

$$= \frac{MSR(X_q, \dots, X_{p-1} | X_1, \dots, X_{q-1})}{MSE(X_1, X_2, \dots, X_{p-1})} \sim F_{(1-\alpha)}(p-q, n-p)$$

Modelimizin,

$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$

olduğunu varsayalım.

$H_0: \beta_1 = \beta_2 (= \beta_c)$

$H_1: \beta_1 \neq \beta_2$

hipotezini test ettiğimizi düşünelim. İndirgnmiş model

$y_i = \beta_0 + \beta_c x_{i1} + \beta_c x_{i2} + \beta_3 x_{i3} + \varepsilon_i = \beta_0 + \beta_c (x_{i1} + x_{i2}) + \beta_3 x_{i3} + \varepsilon_i$

olur.

$$F = \frac{\frac{SSE(R) - SSE(F)}{n - 3 - (n-4)}}{\frac{SSE(F)}{n-4}} = \frac{\frac{SSE(R) - SSE(F)}{1}}{\frac{SSE(F)}{n-4}} \sim F_{(1-\alpha)}(1, n-4)$$

Örnek: Aşağıdaki veri seti için

x_{i1}	x_{i2}	x_{i3}	y
38.5	27.2	12.3	18.5
41.7	34.1	23.1	21.7
59.2	29.8	19.6	29.2
⋮	⋮	⋮	⋮
44.6	32.6	23.6	24.6
47.5	37.5	17.5	27.5
41.2	31.2	11.4	31.2

a. X_3 değişkeninin modelde olup olmaması gerektiğini α=0.01 anlam düzeyinde sınayınız.
b. X_2 ve X_3 değişkenlerinin modelde olup olmaması gerektiğini α=0.05 anlam düzeyinde sınayınız.

Aşağıda ilgili regresyon denklemleri ve ANOVA tabloları verilmiştir.

$$\hat{y} = 17.4636 + 0.1422x_1$$

Kaynak	SS	df	MS
Regresyon	19.07	1	19.07
Hata	481.49	18	26.75
Toplam	500.56	19	

$$\hat{y} = 17.9553 + 0.1459x_1 - 0.0204x_2$$

Kaynak	SS	df	MS
Regresyon	19.21	2	9.605
Hata	481.35	17	28.315
Toplam	500.56	19	

$\hat{y} = 13.2908 + 0.1450x_1 + 0.0315x_2 + 0.1703x_3$

Kaynak	SS	df	MS
Regresyon	30.01	3	10.003
Hata	470.55	16	29.409
Toplam	500.56	19	

a. $H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$

$$F^* = \frac{\frac{SSR(X_3|X_1,X_2)}{1}}{\frac{SSE(X_1,X_2,X_3)}{16}}$$

$$SR(X_3|X_1,X_2) = SSR(X_1,X_2,X_3) - SSR(X_1,X_2) = 30.01 - 19.21 = 10.8$$

veya

$$SR(X_3|X_1,X_2) = SSE(X_1,X_2) - SSE(X_1,X_2,X_3) = 481.35 - 470.55 = 10.8$$

$$F^* = \frac{\frac{10.8}{1}}{\frac{470.55}{16}} = 0.367$$

$\alpha = 0.01$ anlam düzeyinde *F(0.99;1,16)=8.53≥F*=0.367*. H_0 reddedilemez. X_3 modelden çıkarılabilir.

b. $H_0: \beta_2 = \beta_3 = 0$

$H_1: En\ az\ biri\ sıfırdan\ farklıdır.$

$$F^* = \frac{SSE(X_1) - SSE(X_1,X_2,X_3)}{(n-2)-(n-4)} \div \frac{SSE(X_1,X_2,X_3)}{n-4}$$

$$= \frac{481.49 - 470.55}{18 - 16} \div \frac{470.55}{16} = 0.186$$

$\alpha = 0.05$ anlam düzeyinde *F(0.95;2,16)=3.63≥F*=0.186*. H_0 reddedilemez. X_2 ve X_3 modelden çıkarılabilir.