İST 417 Lineer Modeller – 9. Hafta

Genel Lineer Model Örnekleri

Lineer modeller parametreye göre lineer olan modellerdir. Lineer modelin genel hali

$$Y = X\beta + \varepsilon$$

olarak ifade edilir. Burada,

Y: nx1 boyutlu yanıt değerlerinin bir vektörü (Y_{nx1})

X: nxp boyutlu tasarım matrisi (X_{nxp})

β: px1 boyutlu parametrelerin vektörü (β_{Px1})

ε: nx1 boyutlu hata terimlerinin vektörü (ε_{nx1})

olarak ifade edilir.

İstatistikte yaygın olarak kullanılan

i. Lineer regresyon modeli
ii. Varyans analizi modeli (ANOVA) vb.

gibi modeller $Y = X\beta + \varepsilon$ olarak ifade edilen genel lineer modelin örnekleri olarak verilebilir.

Not: Regresyon modelinde tasarım matrisi X tam ranklıdır. ANOVA modelinde ise tasarım matrisi X tam ranklı değildir.

Lineer Modelin Genel Sınıfları

Model 1: En küçük kareler modeli:

$$Y = X\beta + \varepsilon$$

Bu modelde hata terimleri ε hakkında herhangi bir varsayım yapılmaz.

Model 2: Gauss-Markov modeli:

$$Y = X\beta + \varepsilon$$

Bu modelde

$$E(\varepsilon) = 0, \;\; Cov(\varepsilon) = \sigma^2 I = \begin{bmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{bmatrix}$$

olduğu varsayılır.

Rastgele değişkenler birbirinden bağımsız çünkü kovaryansları 0.

Model 3: Aitken Modeli:

$$Y = X\beta + \varepsilon$$

Bu modelde

$$E(\varepsilon) = 0,\ \ Cov(\varepsilon) = \sigma^2 V_{nxn} = \sigma^2 \begin{bmatrix} V_{11} & \cdots & V_{1n} \\ \vdots & \ddots & \vdots \\ V_{n1} & \cdots & V_{nn} \end{bmatrix},\ V \text{ biliniyor.}$$

Gauss Markov Modeli

$$Y = X\beta + \varepsilon$$

modelini ele alırsak burada daha önce bahsedildiği gibi $E(\varepsilon) = 0,\ \ Cov(\varepsilon) = \sigma^2 I_{nxn}$ dir. Şimdi bu modelin özel durumlarını inceleyelim.

<u>Örnek:</u> Konum modeli

$y_i = \mu + \varepsilon_i,\ \ i = 1, \dots, n$

olarak tanımlanır.

$y_1, y_2, \dots, y_n$ ortalaması μ varyansı σ^2 olan bağımsız ve aynı dağılımlı (independent and identically distributed, iid) rastgele bir örneklem olsun. Bu durumda hata terimlerinin$\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ $E(\varepsilon_i) = 0$ ve $Var(\varepsilon_i) = \sigma^2$ olan iid rastgele değişkenler olduğu görülür. Bu nedenle, konum modeli Gauss-Markov (GM) modeli

$$Y = X\beta + \varepsilon$$

olarak yazılır. Burada,

$$y_{nx1} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},\quad X_{nx1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix},\quad \beta_{1x1} = \mu \ \text{ ve } \ \varepsilon_{nx1} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

olarak ifade edilir. Dolayısıyla, konum modeli

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \mu \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

şeklinde yazılır.

Burada,

$y_1 = \mu + \varepsilon_1$

$y_2 = \mu + \varepsilon_2$

$\vdots$

$y_n = \mu + \varepsilon_n$

$E(\varepsilon) = 0$ ve $Var(\varepsilon) = \sigma^2 I$ dır.

Not: $Cov(\varepsilon, \varepsilon) = Var(\varepsilon)$

Örnek: Basit Lineer Regresyon modeli

Basit lineer regresyon modeli

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1,2,\dots,n$$

olarak yazılır. Bu modelde bağımlı değişken y ile bağımsız değişken x arasında lineer bir ilişki vardır denir.

Bu modelde, $E(\varepsilon) = 0,\ \ Cov(\varepsilon) = \sigma^2 I$ dır.

Burada, $y_1 = \beta_0 + \beta_1 x_1 + \varepsilon_1$

$y_2 = \beta_0 + \beta_1 x_2 + \varepsilon_2$

$\vdots$

$y_n = \beta_0 + \beta_1 x_n + \varepsilon_n$

ve

$$\Rightarrow \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

olduğundan basit doğrusal regresyon modeli

$$Y = X\beta + \varepsilon$$

olarak ifade edilen GM modelidir.

Örnek: Çoklu Lineer Regresyon modeli

y bağımlı değişkeninin $x_1, x_2, \dots, x_k$ bağımsız değişkenleri ile aralarında lineer bir ilişki olduğunu varsayalım, bir başka deyişle

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1,2,\dots,n$$

olarak ifade edilsin. Bu modelde, $E(\varepsilon) = 0$, $Cov(\varepsilon) = \sigma^2 I$ dir. Çoklu lineer regresyon modeli

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + \varepsilon_1$$

$$y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_k x_{2k} + \varepsilon_2$$

$$\vdots$$

$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_k x_{nk} + \varepsilon_n$$

ve

$$\Rightarrow \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{1k} \\ 1 & x_{21} & x_{22} & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

olarak yazılabildiğinden, bu model

$$Y = X\beta + \varepsilon$$

şeklinde ifade edilen GM modelinin özel bir halidir.

Örnek: Bir-Yönlü ANOVA modeli

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, \dots, a, j = 1, \dots, n_i$$

olarak tanımlanır. Burada,

a: deneme sayısı

n_i: her bir denemedeki tekrar sayısı

dır.

Bir-yönlü ANOVA modeli

$$\begin{aligned} y_{11} &= \mu + \alpha_1 + \varepsilon_{11} \\ &\vdots \\ y_{1n_1} &= \mu + \alpha_1 + \varepsilon_{1n_1} \\ &\vdots \\ y_{a1} &= \mu + \alpha_a + \varepsilon_{a1} \\ &\vdots \\ y_{an_a} &= \mu + \alpha_a + \varepsilon_{an_a} \end{aligned}$$

ve

$$\Rightarrow \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \\ \vdots \\ \vdots \\ y_{a1} \\ y_{a2} \\ \vdots \\ y_{an_a} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & 0 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_a \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n_1} \\ \vdots \\ \vdots \\ \varepsilon_{a1} \\ \varepsilon_{a2} \\ \vdots \\ \varepsilon_{an_a} \end{bmatrix}$$

Not: X tasarım matrisi alternatif olarak

$$X = \begin{bmatrix} 1_{n_1} & 1_{n_1} & 0_{n_1} & \cdots & 0_{n_1} \\ 1_{n_2} & 0_{n_2} & 1_{n_2} & \cdots & 0_{n_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1_{n_a} & 0_{n_a} & 0_{n_a} & \cdots & 1_{n_a} \end{bmatrix}$$

şeklinde ifade edilir. Burada,

1_{n_i}: $n_i x1$ boyutlu 1 lerin vektörü

0_{n_i}: $n_i x1$ boyutlu 0 ların vektörü

olarak tanımlanır.

Sonuç olarak bir-yönlü ANOVA modelinin

$$Y = X\beta + \varepsilon$$

şeklinde ifade edilen GM modelinin özel bir hali olduğu görülür.

Not: $E(\varepsilon) = 0$ ve $Cov(\varepsilon) = \sigma^2 I$ dir.

Tasarım Matrisi X in Rankı:

Dikkat edilirse X matrisinin 1. sütunu diğer sütunlarının toplamına eşittir. Bu nedenle X matrisi tam ranklı değildir. Tam ranklı olmama ANOVA modelinin karakteristik bir özelliğidir. Bu örnekte rank(X)=a dır.