En Küçük Kareler Kestiricilerinin Özellikleri ve Uydurulan Regresyon Modeli

Denklem (1.6) ve (1.7)'den $\hat{\beta}_0$ ve $\hat{\beta}_1$'nın y_i gözlemlerinin **doğrusal birleşimleri** olduğu görülmektedir.

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \sum_{i=1}^{n} c_i y_i \quad , \quad i = 1,2,...,n \quad , \quad c_i = {}^{(x_i - \bar{x})}\!\big/_{S_{xx}}$$

olarak yazılabilir.

- $\hat{\beta}_0$ ve $\hat{\beta}_1$, en küçük kareler kestiricileri, β_0 ve β_1 model parametrelerinin **yansız** kestiricileridir.

  $E(\varepsilon_i) = 0$ olarak varsayıldığından,

$$E(\hat{\beta}_1) = E(\sum_{i=1}^{n} c_i y_i) = \sum_{i=1}^{n} c_i \; E(y_i)$$

$$= \sum_{i=1}^{n} c_i \; (\beta_0 + \beta_1 x_i) = \beta_0 \sum_{i=1}^{n} c_i + \beta_1 \sum_{i=1}^{n} c_i x_i$$

Burada, $\sum_{i=1}^{n} c_i = \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{S_{xx}} = 0$ ve $\sum_{i=1}^{n} c_i x_i = \sum_{i=1}^{n} \frac{(x_i - \bar{x}) x_i}{S_{xx}} = 1$ olduğundan ;

$$E(\hat{\beta}_1) = \beta_1$$

olur. Diğer bir deyişle, eğer modelin doğru olduğu ($E(y_i) = \beta_0 + \beta_1 x_i$) kabul edilirse o zaman $\hat{\beta}_1$, β_1'in **yansız** kestiricisidir.

$$E(\hat{\beta}_0) = E(\sum_{i=1}^{n} d_i y_i) = \sum_{i=1}^{n} d_i \; E(y_i)$$

$$= \sum_{i=1}^{n} d_i \; (\beta_0 + \beta_1 x_i) = \beta_0 \sum_{i=1}^{n} d_i + \beta_1 \sum_{i=1}^{n} d_i x_i$$

Burada $\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} \left(\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{S_{xx}} \right) = 1$ ve $\sum_{i=1}^{n} d_i x_i = \sum_{i=1}^{n} \left[\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{S_{xx}} \right] x_i = 0$

olduğundan;

$$E(\hat{\beta}_0) = \beta_0$$

olur.

Eğer modelin doğru olduğu ($E(y_i) = \beta_0 + \beta_1 x_i$) kabul edilirse bu durumda $\hat{\beta}_0$, β_0'in **yansız** kestiricisidir.

- $\hat{\beta}_1$'nın varyansı,

$$Var(\hat{\beta}_1) = Var\left(\sum_{I=1}^{n} c_i y_i\right) = \sum_{i=1}^{n} c_i^2 Var(y_i) \qquad \textbf{(1.13)}$$

y_i gözlemleri ilişkisiz olduğundan toplamın varyansı, sadece varyansların toplamıdır. $Var(y_i) = \sigma^2$ olup,

$$Var(\hat{\beta}_1) = \sigma^2 \sum_{i=1}^{n} c_i^2 = \frac{\sigma^2 \sum_{i=1}^{n} (x_i - \overline{x})^2}{S_{xx}^{\ 2}} = \frac{\sigma^2}{S_{xx}} \qquad \textbf{(1.14)}$$

$\hat{\beta}_0$'nın varyansı,

$$\begin{aligned} Var(\hat{\beta}_0) &= Var(\overline{y} - \hat{\beta}_1 \overline{x}) \\ &= Var(\overline{y}) + \overline{x}^2 Var(\hat{\beta}_1) - 2\,\overline{x}\, Cov(\overline{y}, \hat{\beta}_1) \end{aligned}$$

$Var(\overline{y}) = \dfrac{\sigma^2}{n}$ ve $Cov\,(\overline{y}, \hat{\beta}_1) = 0$ olup,

$$Var(\hat{\beta}_0) = Var(\overline{y}) + \overline{x}^2 Var(\hat{\beta}_1) = \sigma^2 \left(\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}\right) \qquad \textbf{(1.15)}$$

olur.

- **GAUSS-MARKOV TEOREMİ**

$E(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2$ varsayımına ve ilişkisiz hatalara sahip (1.1) denklemindeki regresyon modeli için en küçük kareler kestiricilerinin, y_i'nin **doğrusal** birleşimleri olan diğer bütün yansız kestiricilerle karşılaştırıldığında **yansız** olduklarını ve **minimum varyansa** sahip olduklarını belirten bir teoremdir.

Bu teorem, en küçük kareler kestiricilerinin **"en iyi doğrusal yansız kestiriciler (Best Linear Unbiased Estimator, BLUE)"** olduğunu kanıtlamaktadır.

❖ **En küçük kareler uyumunun daha başka yararlı özellikleri de vardır :**

1) β_0 kesim noktasını içeren herhangi bir regresyon modelindeki artıkların toplamı her zaman sıfırdır. (Yuvarlama hataları toplamı etkileyebilir.)

$$\sum_{i=1}^{n}(y_i - \hat{y}_i) = \sum_{i=1}^{n} e_i = 0$$

2) y_i gözlenen değerlerinin toplamı, $\hat{y}_i$ değerlerinin toplamına eşittir :

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$$

3) En küçük kareler regresyon doğrusu her zaman verilerin **merkezinden** $\left[(\bar{y},\bar{x})\ noktası\right]$ geçer.

4) Bağımsız değişkenin karşılık gelen değeriyle ağırlıklandırılmış artıkların toplamı her zaman sıfırdır :

$$\sum_{i=1}^{n} x_i e_i = 0$$

5) $\hat{y}_i$ kestirim değerleri ile ağırlıklandırılmış artıkların toplamı her zaman sıfırdır.

$$\sum_{i=1}^{n} \hat{y}_i e_i = 0$$