LECTURE - 36

Agenda:

1. Some properties of least squares estimates in the multivariate linear model

In the previous lecture, we considered the multivariate linear model to study the relationship between a characteristic Y and related characteristics $X_1, X_2, ---, X_p$.

Data: n independent vectors

$(Y_1, X_{11}, X_{12}, ---, X_{1p}), (Y_2, X_{21}, X_{22}, ---, X_{2p}),$

$---, (Y_n, X_{n1}, X_{n2}, ---, X_{np})$

Model: $Y_i = \beta_0 + \beta_1 X_{i1} + --- + \beta_p X_{ip} + \varepsilon_i,$

where $\varepsilon_1, \varepsilon_2, ---, \varepsilon_n$ are independent errors with a common distribution with $E[\varepsilon] = 0$ and $V(\varepsilon) = \sigma^2$.

Let us recall some notation from the previous lecture, which makes presentation easier.

$$\underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix},$$

Dimension $= (p+1) \times 1$ Dimension $= n \times (p+1)$

$$\underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

Dimension $= n \times 1$

A natural way to estimate $\underline{\beta}$ is by minimizing the error sum of squares

$$S(\underline{\beta}) = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_{p-1} x_{i,p-1} - \beta_p x_{ip})^2$$

$$= (\underline{Y} - X\underline{\beta})^T (\underline{Y} - X\underline{\beta}).$$

We derived that if $\hat{\underline{\beta}} = \arg\min_{\underline{\beta}} S(\underline{\beta})$, then

$$\boxed{\hat{\underline{\beta}} = (X^T X)^{-1} X^T \underline{Y}}.$$

Today, we will study some properties of this estimator.

PROPERTY 1: $E[\hat{\beta}_i] = \beta_i$ for every $i = 0, 1, \ldots, p$, i.e. each component of $\hat{\underline{\beta}}$ is an unbiased estimator of the corresponding parameter.

Let $SSE \triangleq (\underline{Y} - X\hat{\underline{\beta}})^T(\underline{Y} - X\hat{\underline{\beta}})$ denote the estimated error sum of squares.

PROPERTY 2: $\hat{\sigma}^2 = \dfrac{SSE}{n-p-1}$ is an unbiased estimate of the error variance σ^2.

PROPERTY 3: $Cov(\hat{\beta}_i, \hat{\beta}_j) = (X^TX)^{-1}_{i+1,j+1}\sigma^2$ i.e. the $(i+1, j+1)^{th}$ entry of $(X^TX)^{-1}$.

If we make the additional assumption that $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are Normal, then the following properties hold.

PROPERTY 4: $\hat{\beta}_i$ is Normal$\left(\beta_i, (X^TX)^{-1}_{i+1,i+1}\right)$ for every $i = 1, 2, \ldots, n$.
Infact, for every vector $\underline{a}$ (dimension $= (p+1) \times 1$),

$$a^T\hat{\beta} = \sum_{i=0}^{p} a_i \hat{\beta}_i \text{ is Normal}\left(\underline{a}^T\underline{\beta}, \underline{a}^T(X^TX)^{-1}\underline{a}\right).$$

PROPERTY 5: $\dfrac{(n-p-1)\hat{\sigma}^2}{\sigma^2}$ has a χ^2 distribution

with $n-p-1$ degrees of freedom.

These properties are useful in deriving confidence intervals for the parameters. Suppose we want a $(1-\alpha)$-confidence interval for the parameter β_i.

Step 1: Since $\hat{\beta}_i$ is Normal$\left(\beta_i, (X^TX)^{-1}_{i+1,i+1}\sigma^2\right)$, it

follows that $\dfrac{\hat{\beta}_i - \beta_i}{\sqrt{(X^TX)^{-1}_{i+1,i+1}}\,\sigma}$ is Normal(0,1). However,

this depends on σ and cannot be used as a pivotal quantity. Note that

$$\frac{\hat{\beta}_i - \beta_i}{\sqrt{(X^TX)^{-1}_{i+1,i+1}}\,\hat{\sigma}}$$ has a t-distribution with

$n-p-1$ degrees of freedom, and can be used as a pivotal quantity. Hence, a $(1-\alpha)$-confidence interval for β_i is given by

$$\cancel{\left[\hat{\beta}_i - t_{n-p-1,1-\frac{\alpha}{2}}\,\hat{\sigma}\sqrt{(X^TX)^{-1}_{i+1,i+1}},\ \hat{\beta}_i + t_{n-p-1,1-\frac{\alpha}{2}}\,\hat{\sigma}\sqrt{(X^TX)^{-1}}\right]}$$

$$\left[\hat{\beta}_i - t_{n-p-1,1-\frac{\alpha}{2}}\,\hat{\sigma}\sqrt{[(X^TX)^{-1}]_{i+1,i+1}},\ \hat{\beta}_i + t_{n-p-1,1-\frac{\alpha}{2}}\,\hat{\sigma}\sqrt{[(X^TX)^{-1}]_{i+1,i+1}}\right]$$