Chapter 11

- $(1-\alpha)$ prediction interval for Y_0:

$$\left[\hat{Y}_0 - t_{\alpha/2,n-2} S\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}, \hat{Y}_0 + t_{\alpha/2,n-2} S\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}\right].$$

Chapter 12

- Multiple linear regression model

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + +\beta_k x_{ki} + \epsilon_i,\ \ i = 1, 2, ..., n;\ \ n > k,$$

where $\epsilon_1, \epsilon_2, ..., \epsilon_n$ are independent, $\epsilon_i \sim N(0, \sigma^2)$. In matrix form

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon,$$

where

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix},\ \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{bmatrix},\ \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix},\ \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

- Least squares estimators $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$ of $\beta_0, \beta_1, ..., \beta_k$ are solutions of the normal equations

$$(\mathbf{X}'\mathbf{X})\beta = \mathbf{X}'\mathbf{Y}.$$

In particular, if the matrix $\mathbf{X}'\mathbf{X}$ is nonsingular, then

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{Y}.$$

The estimators $\hat{\beta}$ are normally distributed and have the following parameters:

$$E\hat{\beta} = \beta,\ \ \mathrm{Cov}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

- An unbiased estimator of σ^2 is

$$S^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2.$$

- Let $\hat{\mathbf{Y}}$ and $\mathbf{e}$ be fitted values and residuals, that is

$$\hat{\mathbf{Y}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix},\ \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix},$$

where

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_1 x_{2i} + ... + \hat{\beta}_1 x_{ki}$$

and

$$e_i = Y_i - \hat{Y}_i.$$

• ANOVA in multiple regression. Sums of squares:

$$SSR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$

(regression sum of squares),

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

(error sum of squares),

$$SST = \sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2$$

(total sum of squares).

Properties: 1) SSR and SSE are independent, 2) $SST = SSR + SSE$.

The hypothesis $H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$ is tested versus the alternative H_1 that at least one of β-s is not 0 (H_0 means that the regression is not significant, H_1 – significant). The test statistik is

$$F = \frac{SSR/k}{SSE/(n-k-1)}.$$

H_0 is rejected if $F \geq f_{\alpha,k,n-k-1}$. The rejection means that at least one regressor is important.

• Inference about β_j is based on

$$\frac{\hat{\beta}_j - \beta_j}{S\sqrt{c_{jj}}} \sim t_{n-k-1},$$

where c_{jj} is the j-th diagonal element of the matrix $(\mathbf{X}'\mathbf{X})^{-1}$, and is performed in the usual way.