Chapter 12

- Prediction. New value for the covariates is

$$\mathbf{x'}_0 = (1, x_{10}, x_{20}, ..., x_{k0}).$$

A $(1-\alpha)$ confidence interval for the mean response $E(Y_0|\mathbf{x}_0)$ is

$$\left[\hat{Y}_0 - t_{\alpha/2,n-k-1} S\sqrt{\mathbf{x'}_0(\mathbf{X'X})^{-1}\mathbf{x}_0}, \hat{Y}_0 + t_{\alpha/2,n-k-1} S\sqrt{\mathbf{x'}_0(\mathbf{X'X})^{-1}\mathbf{x}_0}\right],$$

where

$$\hat{Y}_0 = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j x_{j0}.$$

A $(1-\alpha)$ prediction interval for the new observation Y_0 is

$$\left[\hat{Y}_0 - t_{\alpha/2,n-k-1} S\sqrt{1+\mathbf{x'}_0(\mathbf{X'X})^{-1}\mathbf{x}_0}, \hat{Y}_0 + t_{\alpha/2,n-k-1} S\sqrt{1+\mathbf{x'}_0(\mathbf{X'X})^{-1}\mathbf{x}_0}\right].$$

- Coefficient of determination

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}.$$

Adjusted coefficient of determination

$$R^2_{\text{adj}} = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)}.$$

- C_p statistic

$$C_p = p + \frac{(s^2 - \hat{\sigma}^2)(n-p)}{\hat{\sigma}^2},$$

where p is the number of model parameters, s^2 is the mean square error for the candidate model, $\hat{\sigma}^2$ is the mean square error from the most complete model. One chooses the model with minimal C_p.

Chapter 13

- One-way ANOVA. Assumptions: k samples of sizes $n_1, n_2, ..., n_k$ (from k populations) are independent and normally distributed with means $\mu_1, \mu_2, ..., \mu_k$ and common variance σ^2.

Hypothesis

$$H_0 : \mu_1 = \mu_2 = ... = \mu_k$$

$$H_1 : \mu_i \neq \mu_j \text{ for at least one pair } (i,j)$$

- Model. Data

$$Y_{ij} = \mu_i + \epsilon_{ij},\ i = 1, ..., k,\ j = 1, ..., n_i,\ \epsilon_{ij} \sim N(0, \sigma^2).$$

An alternative form

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij},\ \mu = \frac{1}{k}\sum_{i=1}^{k}\mu_i$$

(α_i is called the effect of the i-th treatment)

$$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_k = 0$$

$$H_1 : \alpha_i \neq 0 \text{ for at least one } i$$

- Means and sums of squares:

$$\bar{Y}_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij},\ \bar{Y}_{..} = \frac{1}{N}\sum_{i=1}^{k} n_i \bar{Y}_{i.}\ \left(N = \sum_{i=1}^{k} n_i\right),$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{..})^2$$

(total sum of squares),

$$SSA = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{Y}_{i.} - \bar{Y}_{..})^2$$

(treatment sum of squares),

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i.})^2$$

(error sum of squares).

- Properties:

1) $SST = SSA + SSE$,
2) SSA and SSE are independent,
3) $SSE/\sigma^2 \sim \chi^2_{N-k}$,
4) $SSA/\sigma^2 \sim \chi^2_{k-1}$ under H_0.

- Test statistic

$$F = \frac{SSA/(k-1)}{SSE/(N-k)}.$$

Under H_0 F has the Fisher distribution with $k-1$ and $N-k$ degrees of freedom. Test: if $F \geq f_{\alpha,k-1,N-k}$, then H_0 is rejected.