**Chapter 10**

• Goodness-of-fit test. $k$ cells, $e_i$ are expected frequencies, $o_i$ are observed frequencies, $i = 1, ..., k$. The test statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}.$$

The critical region: $\chi^2 \geq \chi^2_{\alpha,k-1}$.

• Test for independence (categorical data: $r \times c$ contingency table). The test statistic

$$\chi^2 = \sum_{j=1}^{c} \sum_{i=1}^{r} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}.$$

The critical region: $\chi^2 \geq \chi^2_{\alpha,(r-1)(c-1)}$.

The same test statistic and critical region are used for testing homogeneity.

**Chapter 11**

• Simple linear regression

$$Y = \beta_0 + \beta_1 x + \epsilon,$$

$x$ is regressor (predictor, independent variable, explanatory variable), $Y$ is response (dependent variable), $\epsilon$ is error (random variable), $E\epsilon = 0$; $\beta_0$ (intercept) and $\beta_1$ (slope) are parameters of interest.

• Data: $(x_1, Y_1), (x_2, Y_2), ..., (x_n, Y_n)$,

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

where $\epsilon_1, \epsilon_2, ..., \epsilon_n$ are independent normally distributed, $E\epsilon_i = 0$, $\text{Var}(\epsilon_i) = \sigma^2$.

If $\hat{\beta}_0$ and $\hat{\beta}_1$ are some estimators of $\beta_0$ and $\beta_1$, then

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

are fitted values, and the differences

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

are residuals.

A visual analysis of the residuals gives useful preliminary information about data and model. On the basis of this analysis, the model can be corrected.

• Least squares estimators: minimization of

$$\sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i)^2.$$

The estimators are

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) Y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2},$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}.$$

- Properties of the estimators. $\hat{\beta}_1$ and $\hat{\beta}_0$ are normally distributed.

$$E\hat{\beta}_1 = \beta_1,\ E\hat{\beta}_0 = \beta_0,$$

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

$$\mathrm{Var}(\hat{\beta}_0) = \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \sigma^2.$$