## Chapter 10

• Two samples $X_1, ..., X_n$ and $Y_1, ..., Y_m$ (from normal populations or $n$ and $m$ are large enough); $\mathrm{Var}(X_i) = \sigma_X^2$, $\mathrm{Var}(Y_i) = \sigma_Y^2$. Testing $H_0 : \sigma_X^2 = \sigma_Y^2$. Alternatives a) $H_1 : \sigma_X^2 > \sigma_Y^2$, b) $H_1 : \sigma_X^2 < \sigma_Y^2$, c) $H_1 : \sigma_X^2 \neq \sigma_Y^2$. The level of significance $\alpha$.

The test statistic

$$F = \frac{S_X^2}{S_Y^2}.$$

Critical region: a) $F \geq f_{\alpha,n-1,m-1}$, b) $F \leq f_{1-\alpha,n-1,m-1}$, a) $F \geq f_{\alpha/2,n-1,m-1}$ or $F \leq f_{1-\alpha/2,n-1,m-1}$

• $P$-value is the minimal level of significance such that $H_0$ is rejected for the observed data.

• 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, $E\epsilon_i = 0$, $\mathrm{Var}(\epsilon_i) = \sigma^2$.