Applied Data Analysis

Exercise Sheet 4 - Solutions

Exercise 14

The derivations in parts (a) and (b) are based on Theorems I.4.32, I.4.39 and I.4.40, I.5.21.

(a)

General linear hypothesis testing

(i) Choosing $\boldsymbol{c} := (1, -1, 0, \dots, 0)' \in \mathbb{R}^p$, we obtain:

(1) $$H_0 : \beta_1 = \beta_2 \iff H_0 : \beta_1 - \beta_2 = 0 \iff H_0 : \boldsymbol{c}'\boldsymbol{\beta} = 0$$

(ii) Choosing

$$K := \begin{pmatrix} 1 & -1 & 0 & & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & & \vdots \\ 0 & & \cdots & 0 & 1 & -1 \end{pmatrix} \in \mathbb{R}^{(p-1)\times p}$$

we obtain:

(2) $$H_0 : \beta_1 = \ldots = \beta_p \iff$$
$$H_0 : (\beta_1 - \beta_2, \beta_2 - \beta_3, \dots, \beta_{p-1} - \beta_p) = (0, \dots, 0) \iff$$
$$H_0 : K\boldsymbol{\beta} = \mathbf{0}$$

Note that the given matrix is not uniquely determined. Another possible matrix to test the same null hypothesis is e.g. given by

$$\widetilde{K} = \begin{pmatrix} 1 & -1 & 0 & & \cdots & 0 \\ 1 & 0 & -1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & & \vdots \\ 1 & & \cdots & 0 & 0 & -1 \end{pmatrix} \in \mathbb{R}^{(p-1)\times p}$$

since it (also) holds:

$$H_0 : \beta_1 = \ldots = \beta_p \iff$$
$$H_0 : (\beta_1 - \beta_2, \beta_1 - \beta_3, \dots, \beta_1 - \beta_p) = (0, \dots, 0) \iff$$
$$H_0 : \widetilde{K}\boldsymbol{\beta} = \mathbf{0}$$

Nested model testing

(i) We add the first and the second column of the design matrix

$$B = (b_{ij})_{1 \le i \le n\,,\, 1 \le j \le p} \in \mathbb{R}^{n \times p}$$

and remove one parameter from the parameter vector $\boldsymbol{\beta}$ to form model $\mathcal{M}_1$. The more complex model $\mathcal{M}_2$ is the model with the unchanged design matrix B. Then we can test $\mathcal{M}_1$ against $\mathcal{M}_2$.

$\mathcal{M}_2$ corresponds to the given linear model

$$Y_i = \sum_{j=1}^{p} b_{ij}\,\beta_j + \varepsilon_i \ , \ i \in \{1, \ldots, n\} \ ,$$

and the nested model $\mathcal{M}_1$ (corresponding to $H_0 : \beta_1 = \beta_2$) can be formulated as

$$Y_i = \beta_1 \,(b_{i1} + b_{i2}) + \sum_{j=3}^{p} b_{ij}\,\beta_j + \varepsilon_i \ , \ i \in \{1, \ldots, n\} \ .$$

(ii) We add all columns of the design matrix B and proceed like in (i).

Remark

In an ANOVA, the global null hypothesis is usually stated as *no factor has any effect*, which can be formulated as in (ii). Notice that for reduction of complexity in the formulation, here we have used a model without intercept. The analogous null hypothesis for models with intercept (respectively, the effect of a factor compared to the overall mean) can be stated completely analogous. Only the intercept term should not be tested.

After rejecting this global null hypothesis, one conducts an analysis using so called *post-hoc* tests, to determine which effect caused the rejection of the hypothesis (see part (c)).

(b) Let $\alpha \in (0,1)$.

(i) Using the notations of theorem I.5.21 with $m := p-1$ we consider (generally) the null hypothis

(3) $$H_0 \,:\, K\boldsymbol{\beta} \;=\; \boldsymbol{\delta}$$

for some matrix $K \in \mathbb{R}^{q\times p}$ with $\operatorname{rank}(K) = q \le p$ and for some $\boldsymbol{\delta} \in \mathbb{R}^q$.

Then, applying Theorem I.5.21, based on the corresponding F-test, H_0 in (3) is rejected at significance level α if

(4) $$F \;=\; \frac{(K\widehat{\boldsymbol{\beta}} - \boldsymbol{\delta})'\,K\,(\boldsymbol{B}'\boldsymbol{B})^{-1}\,K'\,(K\widehat{\boldsymbol{\beta}} - \boldsymbol{\delta})/q}{\mathrm{SSE}/(n-p)} \;>\; F_{1-\alpha}(q, n-p)$$

The corresponding confidence interval for $K\boldsymbol{\beta}$ of confidence level $1-\alpha$ is given by the acceptance region of this F-test, i.e. by

(5) $$F \;=\; \frac{(K\widehat{\boldsymbol{\beta}} - \boldsymbol{\delta})'\,K\,(\boldsymbol{B}'\boldsymbol{B})^{-1}\,K'\,(K\widehat{\boldsymbol{\beta}} - \boldsymbol{\delta})/q}{\mathrm{SSE}/(n-p)} \;\le\; F_{1-\alpha}(q, n-p)$$

Thus, (generally) by (5), a confidence region for $K\boldsymbol{\beta}$ of confidence level $1-\alpha$ based on the F-distribution is given by

(6) $$\mathcal{K}_F \;=\; \left\{ \boldsymbol{\delta} \in \mathbb{R}^q \;\middle|\; (K\widehat{\boldsymbol{\beta}} - \boldsymbol{\delta})'\,K\,(\boldsymbol{B}'\boldsymbol{B})^{-1}\,K'\,(K\widehat{\boldsymbol{\beta}} - \boldsymbol{\delta}) \;\le\; F_{1-\alpha}(q, n-p)\,\frac{q\,\mathrm{SSE}}{n-p} \right\}$$

Now, for $K := c' = (1,-1,0,\dots,0) \in \mathbb{R}^{1\times p}$ (i.e. $K\boldsymbol{\beta} = \beta_1 - \beta_2$), $\mathcal{K}_F$ yields the same confidence interval for $\beta_1 - \beta_2$ that will be derived in (b), (ii) (using $Z \sim t(n-p) \iff Z^2 \sim F(1, n-p)$).

But, unfortunately, in general $\mathcal{K}_F$ will yield a set which only is abstractly given by the corresponding inequality.

(ii) Based on the t-distribution, it is possible to explicitly construct a confidence interval for $\beta_1 - \beta_2$.

For that purpose, again, we choose $\boldsymbol{c} := (1,-1,0,\dots,0)' \in \mathbb{R}^p$ and

(7) $$v \;:=\; \big((B'B)^{-1}\big)_{11} + \big((B'B)^{-1}\big)_{22} - 2\,\big((B'B)^{-1}\big)_{12} \overset{\text{Def. }\boldsymbol{c}}{=} \boldsymbol{c}'\,(B'B)^{-1}\,\boldsymbol{c}$$

with $\big((B'B)^{-1}\big)_{ij}$ denoting the entry in row i and column j of the matrix $(B'B)^{-1}$ for $i,j \in \{1,\dots,p\}$. Then, we obtain by Theorem I.4.32:

(8) $$\frac{(\widehat{\beta}_1 - \widehat{\beta}_2) - (\beta_1 - \beta_2)}{\sqrt{v\,||Y - B\widehat{\boldsymbol{\beta}}||^2/(n-p)}} \;\underset{(7)}{\overset{\text{Def. }\boldsymbol{c}}{=}}\; \frac{\boldsymbol{c}'\,(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sqrt{\boldsymbol{c}'\,(B'B)^{-1}\,\boldsymbol{c}\;||Y - B\widehat{\boldsymbol{\beta}}||^2/(n-p)}} \;\sim\; t(n-p)$$

Then, it holds (using the symmetry of the t-distribution):

(9) $$\begin{aligned} 1-\alpha \;&=\; P_{\boldsymbol{\beta}}\left(-\,t_{1-\frac{\alpha}{2}}(n-p) \;\le\; \frac{(\widehat{\beta}_1 - \widehat{\beta}_2) - (\beta_1 - \beta_2)}{\sqrt{v\,||Y - B\widehat{\boldsymbol{\beta}}||^2/(n-p)}} \;\le\; t_{1-\frac{\alpha}{2}}(n-p) \right) \\ &=\; P_{\boldsymbol{\beta}}\Bigg(\widehat{\beta}_1 - \widehat{\beta}_2 - t_{1-\frac{\alpha}{2}}(n-p)\,\sqrt{\frac{v\,||Y - B\widehat{\boldsymbol{\beta}}||^2}{n-p}} \;\le\; \beta_1 - \beta_2 \;\le \\ &\qquad\qquad \widehat{\beta}_1 - \widehat{\beta}_2 + t_{1-\frac{\alpha}{2}}(n-p)\,\sqrt{\frac{v\,||Y - B\widehat{\boldsymbol{\beta}}||^2}{n-p}} \Bigg) \end{aligned}$$

Hence, by (9), a confidence interval for $\beta_1 - \beta_2$ of confidence level $1-\alpha$ based on the t-distribution is given by

(10) $$\mathcal{K}_t \;=\; \left[\widehat{\beta}_1 - \widehat{\beta}_2 - t_{1-\frac{\alpha}{2}}(n-p)\sqrt{\frac{v\,||Y - B\,\widehat{\boldsymbol{\beta}}||^2}{n-p}}\,,\; \widehat{\beta}_1 - \widehat{\beta}_2 + t_{1-\frac{\alpha}{2}}(n-p)\sqrt{\frac{v\,||Y - B\,\widehat{\boldsymbol{\beta}}||^2}{n-p}}\right]$$

Remark

This confidence interval $\mathcal{K}_t$ corresponds to the null hypothesis H_0 in (1) of part (a) and we reject this null hypothesis H_0 if $0 \notin \mathcal{K}_t$.

(c) Let $\alpha \in (0,1)$ and let each single hypothesis $H_0^{(j,k)}$, $(j,k) \in M$, be tested on significance level
$$\frac{\alpha}{m} \quad \text{with} \quad m := \binom{p}{2} = |M|$$
Then, using the notations given by the setting of part (c), we obtain by applying the first Bonferroni inequality (sub-additivity of probability measures):

$$P_{\boldsymbol{\beta}}\Bigg(\underbrace{\bigcup_{(j,k)\,\in\, I_0(\boldsymbol{\beta})} A^{(j,k)}}_{\substack{\text{At least one null hypothis } H_0^{(j,k)} \\ \text{is rejected although it is true}}} \Bigg) \overset{\text{Bonf.}}{\leq} \sum_{(j,k)\,\in\, I_0(\boldsymbol{\beta})} \underbrace{P_{\boldsymbol{\beta}}\big(A^{(j,k)}\big)}_{\leq \frac{\alpha}{m}} \overset{\text{Def. } I_0(\boldsymbol{\beta})}{\leq} \sum_{(j,k)\,\in\, I_0(\boldsymbol{\beta})} \frac{\alpha}{m}$$

$$\overset{I_0(\boldsymbol{\beta}) \subseteq M}{\leq} \sum_{(j,k)\,\in\, M} \frac{\alpha}{m} \;=\; m\,\frac{\alpha}{m} \;=\; \alpha$$

Exercise 15

(a) First, for the model of simple linear regression, it holds (as already stated for Example I.4.6 on slide 101 of the Lecture):

$$(1) \qquad (B' B)^{-1} \;=\; \frac{1}{(n-1)\, s_{\boldsymbol{xx}}} \begin{pmatrix} \overline{x}^2 + \frac{n-1}{n}\, s_{\boldsymbol{xx}} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix}$$

(Instead of using the statement of the Lecture, one could easily calculate the product $B' B$ and then derive $(B' B)^{-1}$ by the usual rule for inverting a 2×2-matrix.)

Second, by the definition of B (given on slide 67 of the Lecture), we get:

$$(2) \qquad B' \boldsymbol{Y} \;=\; \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \;=\; \begin{pmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} x_i\, Y_i \end{pmatrix}$$

(1) and (2) yield:

$$(3) \qquad \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{pmatrix} \overset{\text{Def.}}{=} \widehat{\boldsymbol{\beta}} \overset{\text{I.5.4}}{=} (B' B)^{-1} B' \boldsymbol{Y}$$

$$\overset{(1),(2)}{=} \frac{1}{(n-1)\, s_{\boldsymbol{xx}}} \begin{pmatrix} \overline{x}^2 + \frac{n-1}{n}\, s_{\boldsymbol{xx}} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix} \begin{pmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} x_i\, Y_i \end{pmatrix}$$

$$= \frac{1}{(n-1)\, s_{\boldsymbol{xx}}} \begin{pmatrix} n\, \overline{x}^2\, \overline{Y} + (n-1)\, s_{\boldsymbol{xx}}\, \overline{Y} \;-\; \overline{x} \sum_{i=1}^{n} x_i\, Y_i \\ -n\, \overline{x}\, \overline{Y} \;+\; \sum_{i=1}^{n} x_i\, Y_i \end{pmatrix}$$

Further, it holds (usual formula for simplifying covariances or variances):

$$(4) \qquad (n-1)\, s_{\boldsymbol{xY}} \overset{\text{Def.}}{=} \sum_{i=1}^{n} (x_i - \overline{x})\, (Y_i - \overline{Y})$$

$$= \sum_{i=1}^{n} x_i\, Y_i \;-\; \underbrace{\overline{x} \sum_{i=1}^{n} Y_i}_{=\, n\, \overline{x}\, \overline{Y}} \;-\; \underbrace{\overline{Y} \sum_{i=1}^{n} x_i}_{=\, n\, \overline{x}\, \overline{Y}} \;+ n\, \overline{x}\, \overline{Y} \;=\; \sum_{i=1}^{n} x_i\, Y_i \;-\; n\, \overline{x}\, \overline{Y}$$

Thus, we obtain:

$$(5) \qquad \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{pmatrix} \overset{(3),(4)}{=} \frac{1}{(n-1)\, s_{\boldsymbol{xx}}} \begin{pmatrix} n\, \overline{x}^2\, \overline{Y} \;+\; (n-1)\, s_{\boldsymbol{xx}}\, \overline{Y} \;-\; (n-1)\, \overline{x}\, s_{\boldsymbol{xY}} \;-\; n\, \overline{x}^2\, \overline{Y} \\ -n\, \overline{x}\, \overline{Y} \;+\; (n-1)\, s_{\boldsymbol{xY}} + n\, \overline{x}\, \overline{Y} \end{pmatrix}$$

$$= \frac{1}{(n-1)\, s_{\boldsymbol{xx}}} \begin{pmatrix} (n-1)\, s_{\boldsymbol{xx}}\, \overline{Y} \;-\; (n-1)\, \overline{x}\, s_{\boldsymbol{xY}} \\ (n-1)\, s_{\boldsymbol{xY}} \end{pmatrix} \;=\; \begin{pmatrix} \overline{Y} \;-\; \dfrac{s_{\boldsymbol{xY}}}{s_{\boldsymbol{xx}}}\, \overline{x} \\ \dfrac{s_{\boldsymbol{xY}}}{s_{\boldsymbol{xx}}} \end{pmatrix}$$

(b) Applying the variance decomposition (for the simple linear regression) and using the representations of $\widehat{\beta}_0$ and $\widehat{\beta}_1$ derived in (a), we obtain:

$$
\begin{aligned}
(6)\quad \mathrm{SSE} &\overset{\text{I.5.11,(1)}}{=} \mathrm{SST} - \mathrm{SSR} \overset{\text{Def.}}{=} \sum_{i=1}^{n} \big(Y_i - \overline{Y}\big)^2 - \sum_{i=1}^{n} \big(\widehat{\beta}_0 + \widehat{\beta}_1 \, x_i - \overline{Y}\big)^2 \\
&\underset{\text{(a)}}{\overset{\text{Def. } s_{\boldsymbol{YY}}}{=}} (n-1)\, s_{\boldsymbol{YY}} - \sum_{i=1}^{n} \big(\overline{Y} - \widehat{\beta}_1 \, \overline{x} + \widehat{\beta}_1 \, x_i - \overline{Y}\big)^2 \\
&= (n-1)\, s_{\boldsymbol{YY}} - \widehat{\beta}_1^{\,2} \sum_{i=1}^{n} (x_i - \overline{x})^2 \overset{\text{Def. } s_{\boldsymbol{xx}}}{=} (n-1)\, s_{\boldsymbol{YY}} - \widehat{\beta}_1^{\,2} \, (n-1)\, s_{\boldsymbol{xx}} \\
&\overset{\text{(a)}}{=} (n-1) \left(s_{\boldsymbol{YY}} - \frac{s_{\boldsymbol{xY}}^2}{s_{\boldsymbol{xx}}^2} \, s_{\boldsymbol{xx}} \right) = (n-1)\, s_{\boldsymbol{YY}} \left(1 - \frac{s_{\boldsymbol{xY}}^2}{s_{\boldsymbol{xx}} \, s_{\boldsymbol{YY}}} \right) \\
&\overset{\text{Def. } r_{\boldsymbol{xY}}}{=} (n-1)\, s_{\boldsymbol{YY}} \big(1 - r_{\boldsymbol{xY}}^2\big)
\end{aligned}
$$

(c) Again, applying the variance decomposition (for the simple linear regression) and the representation of SSE derived in (b), we obtain:

$$
\begin{aligned}
(7)\qquad R^2 &\overset{\text{I.5.11,(1)}}{=} \frac{\mathrm{SST} - \mathrm{SSE}}{\mathrm{SST}} = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}} \\
&\underset{\text{Def. SST}, s_{\boldsymbol{YY}}}{\overset{(6)}{=}} 1 - \frac{(n-1)\, s_{\boldsymbol{YY}} \big(1 - r_{\boldsymbol{xY}}^2\big)}{(n-1)\, s_{\boldsymbol{YY}}} = 1 - \big(1 - r_{\boldsymbol{xY}}^2\big) = r_{\boldsymbol{xY}}^2
\end{aligned}
$$

(as denoted in I.5.44 for the simple linear regression).

Exercise 16

Let $\mu_i := \mu + \alpha_i$, $i \in \{1, \ldots, p\}$. Then, by the model assumptions in Def. I.6.15, it holds for $\boldsymbol{Y} = (Y_{11}, \ldots, Y_{1n_1}, \ldots, Y_{11}, \ldots, Y_{1n_1})'$:

$$(1) \qquad \mathrm{E}(\boldsymbol{Y}) \;=\; \mathrm{E}(B_* \boldsymbol{\beta}_* + \boldsymbol{\varepsilon}) \overset{\text{Lin.}}{=} B_* \boldsymbol{\beta}_* + \underbrace{\mathrm{E}(\boldsymbol{\varepsilon})}_{=\mathbf{0}} \underset{B_*, \boldsymbol{\beta}_*}{\overset{\text{Def.}}{=}} \begin{pmatrix} (\mu+\alpha_1)\, \mathbb{1}_{n_1} \\ \vdots \\ (\mu+\alpha_p)\, \mathbb{1}_{n_p} \end{pmatrix} \underset{\mu_1,\ldots,\mu_p}{\overset{\text{Def.}}{=}} \begin{pmatrix} \mu_1\, \mathbb{1}_{n_1} \\ \vdots \\ \mu_p\, \mathbb{1}_{n_p} \end{pmatrix}$$

$$(2) \qquad \mathrm{Cov}(\boldsymbol{Y}) \;=\; \mathrm{Cov}(B_* \boldsymbol{\beta}_* + \boldsymbol{\varepsilon}) \;=\; \mathrm{Cov}(\boldsymbol{\varepsilon}) \overset{\text{Ass.}}{=} \sigma^2 I_n$$

This implies:

$$(3) \qquad \mathrm{E}\big(\overline{Y}_{i\bullet}\big) \overset{\text{Def.}}{=} \mathrm{E}\left(\frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}\right) \overset{\text{Lin.}}{=} \frac{1}{n_i}\sum_{j=1}^{n_i} \mathrm{E}\big(Y_{ij}\big) \overset{(1)}{=} \frac{1}{n_i}\sum_{j=1}^{n_i} \mu_i \;=\; \mu_i \;,\; i \in \{1,\ldots,p\}$$

$$(4) \qquad \mathrm{E}\big(\overline{Y}_{\bullet\bullet}\big) \overset{\text{Def.}}{=} \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i} Y_{ij}\right) \overset{\text{Def.}}{=} \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{p} n_i\, \overline{Y}_{i\bullet}\right) \overset{\text{Lin.}}{=} \frac{1}{n}\sum_{i=1}^{p} n_i\, \mathrm{E}\big(\overline{Y}_{i\bullet}\big)$$

$$\overset{(3)}{=} \frac{1}{n}\sum_{i=1}^{p} n_i\, \mu_i \underset{\mu_1,\ldots,\mu_p}{\overset{\text{Def.}}{=}} \left(\frac{1}{n}\sum_{i=1}^{p} n_i\right)\mu + \frac{1}{n}\underbrace{\sum_{i=1}^{p} n_i\, \alpha_i}_{=0} \underset{\text{side cond.}}{\overset{\text{Def. } n}{=}} \frac{n}{n}\,\mu \;=\; \mu$$

$$(5) \quad \mathrm{Var}\big(\overline{Y}_{i\bullet}\big) \overset{\text{Def.}}{=} \mathrm{Var}\left(\frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}\right) \overset{\text{Indep.}}{=} \frac{1}{n_i^2}\sum_{j=1}^{n_i} \mathrm{Var}\big(Y_{ij}\big) \overset{(2)}{=} \frac{1}{n_i^2}\sum_{j=1}^{n_i} \sigma^2 \;=\; \frac{\sigma^2}{n_i} \;, \quad i \in \{1,\ldots,p\}$$

$$(6) \quad \mathrm{Var}\big(\overline{Y}_{\bullet\bullet}\big) \overset{\text{Def.}}{=} \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i} Y_{ij}\right) \overset{\text{Indep.}}{=} \frac{1}{n^2}\sum_{i=1}^{p}\sum_{j=1}^{n_i} \mathrm{Var}\big(Y_{ij}\big) \overset{(2)}{=} \frac{\sigma^2}{n^2}\sum_{i=1}^{p} n_i \overset{\text{Def. } n}{=} \frac{\sigma^2}{n}$$

Further, applying these expressions, we obtain:

$$(7) \qquad \mathrm{E}\big(\overline{Y}_{i\bullet}^{\,2}\big) \;=\; \mathrm{Var}\big(\overline{Y}_{i\bullet}\big) + \Big(\mathrm{E}\big(\overline{Y}_{i\bullet}\big)\Big)^2 \overset{(3),(5)}{=} \frac{\sigma^2}{n_i} + \mu_i^2 \;,\; i \in \{1,\ldots,p\}$$

$$(8) \qquad \mathrm{E}\big(\overline{Y}_{\bullet\bullet}^{\,2}\big) \;=\; \mathrm{Var}\big(\overline{Y}_{\bullet\bullet}\big) + \Big(\mathrm{E}\big(\overline{Y}_{\bullet\bullet}\big)\Big)^2 \overset{(4),(6)}{=} \frac{\sigma^2}{n} + \mu^2 \;,\; i \in \{1,\ldots,p\}$$

$$(9) \qquad \mathrm{E}\big(\overline{Y}_{i\bullet}\,\overline{Y}_{\bullet\bullet}\big) \overset{\text{cf. (4)}}{=} \mathrm{E}\left(\frac{1}{n}\sum_{j=1}^{p} n_j\, \overline{Y}_{i\bullet}\,\overline{Y}_{j\bullet}\right) \overset{\text{Lin.}}{=} \frac{1}{n}\left(n_i\, \mathrm{E}\big(\overline{Y}_{i\bullet}^{\,2}\big) + \sum_{\substack{j=1 \\ j\neq i}}^{p} n_j\, \mathrm{E}\big(\overline{Y}_{i\bullet}\,\overline{Y}_{j\bullet}\big)\right)$$

$$\overset{\text{Indep.}}{=} \frac{1}{n}\left(n_i\, \mathrm{E}\big(\overline{Y}_{i\bullet}^{\,2}\big) + \sum_{\substack{j=1 \\ j\neq i}}^{p} n_j\, \mathrm{E}\big(\overline{Y}_{i\bullet}\big)\,\mathrm{E}\big(\overline{Y}_{j\bullet}\big)\right) \overset{(3),(7)}{=} \frac{1}{n}\left(\sigma^2 + n_i\,\mu_i^2 + \sum_{\substack{j=1 \\ j\neq i}}^{p} n_j\,\mu_i\,\mu_j\right)$$

$$= \frac{\sigma^2}{n} + \frac{\mu_i}{n}\sum_{j=1}^{p} n_j\,\mu_j \overset{\text{cf. (4)}}{=} \frac{\sigma^2}{n} + \mu_i\,\mu$$

Thus, we finally obtain:

$$(10)\qquad \mathrm{E}(\mathrm{SSR}) \overset{\text{Def.}}{=} \mathrm{E}\left(\sum_{i=1}^{p} n_i\,(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2\right) = \mathrm{E}\left(\sum_{i=1}^{p} n_i\left(\overline{Y}_{i\bullet}^{\,2} - 2\,\overline{Y}_{i\bullet}\,\overline{Y}_{\bullet\bullet} + \overline{Y}_{\bullet\bullet}^{\,2}\right)\right)$$

$$\overset{\text{Lin.}}{=} \sum_{i=1}^{p} n_i\left(\mathrm{E}\big(\overline{Y}_{i\bullet}^{\,2}\big) - 2\,\mathrm{E}\big(\overline{Y}_{i\bullet}\,\overline{Y}_{\bullet\bullet}\big) + \mathrm{E}\big(\overline{Y}_{\bullet\bullet}^{\,2}\big)\right)$$

$$\overset{(7),(8),(9)}{=} \sum_{i=1}^{p} n_i\left(\frac{\sigma^2}{n_i} + \mu_i^2 - 2\,\frac{\sigma^2}{n} - 2\,\mu_i\mu + \frac{\sigma^2}{n} + \mu^2\right)$$

$$= p\,\sigma^2 - \frac{\sigma^2}{n}\underbrace{\sum_{i=1}^{p} n_i}_{=\,n} + \sum_{i=1}^{p} n_i\,(\underbrace{\mu_i - \mu}_{=\,\alpha_i})^2 \underset{\mu_1,\ldots,\mu_p}{\overset{\text{Def. }n}{=}} (p-1)\,\sigma^2 + \sum_{i=1}^{p} n_i\,\alpha_i^2$$

Remarks to Exercise 16

(i) For the proof of the representation of $\mathrm{E}(\mathrm{SSR})$ shown in Exercise 16, the normal assumption for the error term $\boldsymbol{\varepsilon}$ was *not* needed.

(ii) The representation of $\mathrm{E}(\mathrm{SSR})$ shown in Exercise 16 implies for $p > 1$:

$$(*)\qquad \mathrm{E}\left(\frac{\mathrm{SSR}}{p-1}\right) = \sigma^2 + \frac{1}{p-1}\sum_{i=1}^{p} n_i\,\alpha_i^2 \geq \sigma^2$$

with equality if and only if the following null hypothesis is fulfilled:

$$H_0 : \alpha_1 = \ldots = \alpha_p = 0$$

For $1 < p < n$, under H_0,

$$\frac{\mathrm{SSR}}{p-1} \quad \text{as well as} \quad \frac{\mathrm{SSE}}{n-p}$$

are both unbiased estimators of the unknown variance σ^2.

Thus, under H_0, we expect vaulues of the ANOVA test statistic

$$F = \frac{\mathrm{SSR}/(p-1)}{\mathrm{SSE}/(n-p)}$$

near 1, whereas, according to $(*)$, values of F significantly larger than 1 correspond to the alternative hypothesis H_1.

(iii) The result of Exercise 16 is in accordance with statement (5) of Theorem I.6.22

$$\frac{\mathrm{SSR}}{\sigma^2} \sim \chi^2(p-1)$$

for $p > 1$, implying (by the properties of a χ^2-distribution):

$$\mathrm{E}\left(\frac{\mathrm{SSR}}{\sigma^2}\right) = p - 1$$

In this connection, notice that the result of Theorem I.6.22, (5) – with the **central** χ^2-distribution $\chi^2(p-1)$ – only holds under H_0!

Exercise 17

(a) (i) According to Example II.2.5, for the class

$$\{\mathcal{N}(\mu,\sigma^2)\,\big|\,\mu\in\mathbb{R}\,,\ \sigma>0\}$$

we set $\vartheta := \mu$, $\phi := \sigma$ and define the functions $b:\mathbb{R}\longrightarrow\mathbb{R}$, $a:(0,\infty)\longrightarrow\mathbb{R}$ and $c:\mathbb{R}\times(0,\infty)\longrightarrow\mathbb{R}$ by

$$(1)\qquad b(\vartheta) \;=\; b(\mu) \;:=\; \frac{\mu^2}{2} \;=\; \frac{\vartheta^2}{2}\ ,\quad \vartheta=\mu\in\mathbb{R}$$

$$(2)\qquad a(\phi) \;=\; a(\sigma) \;:=\; \sigma^2 \;=\; \phi^2\ ,\quad \phi=\sigma\in(0,\infty)$$

$$(3)\qquad c(y,\phi) \;=\; c(y,\sigma) \;:=\; -\frac{y^2}{2\,\sigma^2} - \frac{\ln(2\,\pi\,\sigma^2)}{2} \;=\; -\frac{y^2}{2\,\phi^2} - \frac{\ln(2\,\pi\,\phi^2)}{2}\ ,\qquad (y,\phi)=(y,\sigma)\in\mathbb{R}\times(0,\infty)$$

Now, let $\vartheta=\mu\in\mathbb{R}$ and $\phi=\sigma\in(0,\infty)$. Then, by I.2.1, the (usual) probability density function $\varphi_{\mu,\sigma^2}:\mathbb{R}\longrightarrow[0,\infty)$ of $Y\sim\mathcal{N}(\mu,\sigma^2)$ is given by

$$
\begin{aligned}
(4)\qquad \varphi_{\mu,\sigma^2}(y) \;&=\; \frac{1}{\sqrt{2\,\pi\,\sigma^2}}\,\exp\left(-\frac{(y-\mu)^2}{2\,\sigma^2}\right)\\
&=\; \exp\Big(\ln\big((2\,\pi\,\sigma^2)^{-1/2}\big)\Big)\,\exp\left(-\frac{y^2-2\,y\,\mu+\mu^2}{2\,\sigma^2}\right)\\
&=\; \exp\left(\frac{y\,\mu-\frac{\mu^2}{2}}{\sigma^2} - \frac{y^2}{2\,\sigma^2} - \frac{\ln(2\,\pi\,\sigma^2)}{2}\right)\\
&\underset{(1),(2),(3)}{\overset{\mu=\vartheta,\sigma=\phi}{=}}\; \exp\left(\frac{y\,\vartheta-b(\vartheta)}{a(\phi)}+c(y,\phi)\right) \;=:\; f(y;\vartheta,\phi)\ ,\quad y\in\mathbb{R}\,,
\end{aligned}
$$

with a representation of $f(\cdot\,;\vartheta,\phi)$ as given in Definition II.2.3.

(b) (i) Again, let $Y\sim\mathcal{N}(\mu,\sigma^2)$ with $\vartheta:=\mu\in\mathbb{R}$ and $\phi:=\sigma\in(0,\infty)$. Then, applying Prop. II.2.4 and the results of (a),(i), we obtain:

$$(5)\qquad \mathrm{E}(Y) \;\overset{\text{II.2.4}}{=}\; b'(\vartheta) \;\overset{(1)}{=}\; \frac{d}{d\vartheta}\,\frac{\vartheta^2}{2} \;=\; \vartheta \;=\; \mu$$

$$(6)\qquad \mathrm{Var}(Y) \;\overset{\text{II.2.4}}{=}\; b''(\vartheta)\,a(\phi) \;\overset{(2),(5)}{=}\; \left(\frac{d}{d\vartheta}\,\vartheta\right)\sigma^2 \;=\; \sigma^2$$

and, thus by (4) and (5), the expressions for $\mathrm{E}(Y)$ and $\mathrm{Var}(Y)$ already well known for $Y\sim\mathcal{N}(\mu,\sigma^2)$.

(a) (ii) According to Example II.2.5, for the class

$$\{\mathcal{P}(\mu) \,|\, \mu \in (0,\infty)\}$$

we set $\vartheta := \ln(\mu)$, $\phi := \mu$ and define the functions $b : \mathbb{R} \longrightarrow \mathbb{R}$, $a : (0,\infty) \longrightarrow \mathbb{R}$ and $c : N_0 \times (0,\infty) \longrightarrow \mathbb{R}$ by

$$(7) \qquad b(\vartheta) := e^{\vartheta} \overset{\text{Def. } \vartheta}{=} e^{\ln(\mu)} = \mu \;,\quad \vartheta = \ln(\mu) \in \mathbb{R}$$

$$(8) \qquad a(\phi) := 1 \;,\quad \phi = \mu \in (0,\infty)$$

$$(9) \qquad c(y,\phi) := -\ln(y!) \;,\quad (y,\phi) = (y,\mu) \in \mathbb{N}_0 \times (0,\infty)$$

Now, let $\phi := \mu \in (0,\infty)$ and $\vartheta := \ln(\mu)$. Then, by foundations of probability theory, the probability mass function $p_\mu : \mathbb{N}_0 \longrightarrow [0,1]$ of $Y \sim \mathcal{P}(\mu)$ is given by

$$
\begin{aligned}
(10) \quad p_\mu(y) &= \frac{\mu^y}{y!}\, e^{-\mu} = \exp\left(\ln\left(\frac{\mu^y}{y!}\right)\right) \exp(-\mu) \\
&= \exp\big(y \ln(\mu) - \ln(y!) - \mu\big) \overset{\vartheta = \ln(\mu)}{=} \exp\big(y\,\vartheta - e^{\vartheta} - \ln(y!)\big) \\
&\overset{(7),(8),(9)}{=} \exp\left(\frac{y\,\vartheta - b(\vartheta)}{a(\phi)} + c(y,\phi)\right) =: f(y;\vartheta,\phi) \;,\quad y \in \mathbb{R} \;,
\end{aligned}
$$

with a representation of $f(\cdot\,;\vartheta,\phi)$ as given in Definition II.2.3

(b) (ii) Again, let $Y \sim \mathcal{P}(\mu)$ with $\phi := \mu \in (0,\infty)$ and $\vartheta := \ln(\mu)$. Then, applying Prop. II.2.4 and the results of (a),(ii), we obtain:

$$(11) \qquad \mathrm{E}(Y) \overset{\text{II.2.4}}{=} b'(\vartheta) \overset{(7)}{=} \frac{d}{d\vartheta}\, e^{\vartheta} = e^{\vartheta} \overset{\text{Def. } \vartheta}{=} \mu$$

$$(12) \qquad \mathrm{Var}(Y) \overset{\text{II.2.4}}{=} b''(\vartheta)\, a(\phi) \overset{(8),(11)}{=} \frac{d}{d\vartheta}\, e^{\vartheta} = e^{\vartheta} \overset{\text{Def. } \vartheta}{=} \mu$$

and, thus by (11) and (12), the expressions for $\mathrm{E}(Y)$ and $\mathrm{Var}(Y)$ already well known for $Y \sim \mathcal{P}(\mu)$.