Gesamtpunktzahl **10 von 10 Punkte**

Let $\boldsymbol{X} = (X_1, X_2)' \sim N_2(\boldsymbol{\mu}, \Sigma)$ with

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 25 & -3 \\ -3 & 1 \end{pmatrix}.$$

For the following tasks, give your answers by filling the results into the blanks. If necessary, numerical values have to be rounded to three decimal places.

(a) Calculate the correlation $\rho = \operatorname{Corr}(X_1, X_2) = \frac{\operatorname{Cov}(X_1, X_2)}{\sqrt{\operatorname{Var}(X_1)\operatorname{Var}(X_2)}}$. **1 von 1 Punkt**

1 von 1 Punkt

Give the value of ρ.

-0.6 ✓

(b) According to the Lecture or to the Theoretical Exercises, it holds $P^{X_1 \mid X_2 = 1} = N(\nu, \tau^2)$ with $\nu \in \mathbb{R}$, $\tau > 0$ and P denoting the underlying probability distribution. **4 von 4 Punkten**

2 von 2 Punkten

(i) Give the value of $\mathrm{E}(X_1 \mid X_2 = 1)$.

-1 ✓

2 von 2 Punkten

(ii) Give the value of $\operatorname{Var}(X_1 \mid X_2 = 1)$.

16 ✓

(c) Check, wether there exists some $c \in \mathbb{R}$ with $(1, c)\,(\boldsymbol{X} - \boldsymbol{\mu}) \sim N(0, 1)$. **2 von 2 Punkten**

Marginals & Conditionals

For a vector $\mathbf{x} \in \mathbb{R}^p$ and $\emptyset \neq K \subseteq \{1, \ldots, p\}$, let $\mathbf{x}_K = (x_i)_{i \in K}$.

I.2.7 Theorem (parameters and marginals of a multivariate normal distribution)

Let $\mathbf{X} \sim N_p(\mu, \Sigma)$ with $\mu \in \mathbb{R}^p, \Sigma \in \mathbb{R}^{p \times p}_{\geqslant 0}$ and $\emptyset \neq K \subseteq \{1, \ldots, p\}$ and $\Sigma_{K,K} = \mathrm{Cov}\,(\mathbf{X}_K)$. Then:

1. $E\mathbf{X} = \mu$
2. $\mathrm{Cov}\,(\mathbf{X}) = \Sigma$
3. $\mathbf{X}_K \sim N(\mu_K, \Sigma_{K,K})$ ('**marginals of normals are normal**')

I.2.8 Theorem (conditionals of a multivariate normal distribution)

Let $\mathbf{X} \sim N_p(\mu, \Sigma)$ with $\mu \in \mathbb{R}^p, \Sigma \in \mathbb{R}^{p \times p}_{>0}$ and $\emptyset \neq K, L \subseteq \{1, \ldots, p\}$, $K \cap L = \emptyset$, $k = |K|$. Further, let $\Sigma_{K,L} = \mathrm{Cov}\,(\mathbf{X}_K, \mathbf{X}_L)$ and $\Sigma_{KK|L} = \Sigma_{K,K} - \Sigma_{K,L}\Sigma_{L,L}^{-1}\Sigma_{K,L}'$. Then:

1. $\mathbf{X}_K \mid \mathbf{X}_L = \mathbf{x}_L \sim N_k(\mu_K + \Sigma_{K,L}\Sigma_{L,L}^{-1}(\mathbf{x}_L - \mu_L), \Sigma_{KK|L})$

   ('**conditionals of normals are normal**')
2. $E(\mathbf{X}_K \mid \mathbf{X}_L = \mathbf{x}_L) = \mu_K + \Sigma_{K,L}\Sigma_{L,L}^{-1}(\mathbf{x}_L - \mu_L)$

   The matrix $\Sigma_{K,L}\Sigma_{L,L}^{-1}$ is called **regression matrix**.
3. $\mathrm{Cov}\,(\mathbf{X}_K \mid \mathbf{X}_L = \mathbf{x}_L) = \Sigma_{KK|L}$,

(c) Check, wether there exists some $c \in \mathbb{R}$ with $(1, c)\,(\boldsymbol{X} - \boldsymbol{\mu}) \sim N(0, 1)$.

2 von 2 Punkten

2 von 2 Punkten

If there exists such an $c \in \mathbb{R}$, then give its value. If such an $c \in \mathbb{R}$ does not exist, then type "NA" (without quotation marks) into the blank.

NA ✓

(d) It ist known that $\boldsymbol{Y} = \Sigma^{-1/2}\boldsymbol{X} \sim N_2(\Sigma^{-1/2}\boldsymbol{\mu}\,,\, I_2)$ and furthermore, it holds $\boldsymbol{Y}'\boldsymbol{Y} \sim \chi^2(2, \delta)$, i.e. $\boldsymbol{Y}'\boldsymbol{Y}$ has a non-central χ^2-distribution with 2 degrees of freedom and non-centrality parameter $\delta \geq 0$ (see the Lecture or the Theoretical Exercises).

3 von 3 Punkten

3 von 3 Punkten

Give the value of δ with three decimal places. **Hint:** In order to derive the wanted result, calculation of the matrix $\Sigma^{-1/2}$ is **not** necessary.

0.125 ✓

Nächste
Aufgabe →

☐ **Nachkorrekturantrag anlegen?**

Bitte beachte, dass dieses Dokument zu einem Teil deiner Prüfungsakte wird!
- Sei höflich und freundlich.
- Beschreibe deine Begründung so präzise wie möglich.

Orthogonal projectors & quadratic forms

I.3.4 Lemma

Let $Q \in \mathbb{R}^{p \times p}$ be an orthogonal projector. Then, an eigenvalue λ of Q satisfies $\lambda \in \{0, 1\}$. Furthermore, $\text{rank}(Q) = \text{trace}(Q)$.

I.3.5 Theorem

Let $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, I_p)$ be a random vector and Q be an orthogonal projector. Then,

$$\mathbf{Y}'Q\mathbf{Y} \sim \chi^2\Big(\text{rank}(Q), \frac{1}{2}\boldsymbol{\mu}'Q\boldsymbol{\mu}\Big).$$

Gesamtpunktzahl **10 von 10 Punkte**

Let $\boldsymbol{X} = (X_1, X_2, X_3, X_4, X_5)' \sim N_5(\boldsymbol{0}, I_5)$. Furthermore, as defined in the Lecture or in the Exercises, let $E_5 = I_5 - \frac{1}{5}\mathbb{1}_{5\times 5}$.

For the following tasks, give your answers by filling the results into the blanks. If necessary, numerical values have to be rounded to three decimal places.

(a) Calculate the expectation $\mathrm{E}(\boldsymbol{X}' E_5 \boldsymbol{X})$. **Hint:** Lemma I.3.3. **2 von 2 Punkten**

2 von 2 Punkten

Give the value of $\mathrm{E}(\boldsymbol{X}' E_5 \boldsymbol{X})$.

4 ✓

(b) Determine $a \in \mathbb{R}$ such that $\left(a\, I_5 - \mathbb{1}_{5\times 5}\right)\boldsymbol{X}$ and $\overline{X} = \frac{1}{5}\sum_{i=1}^{5} X_i$ are stochastically independent. **Hint:** Use the representation $\overline{X} = B\,\boldsymbol{X}$ with a suitable matrix $B \in \mathbb{R}^{1\times 5}$. **4 von 4 Punkten**

4 von 4 Punkten

Give the value of a.

5 ✓

(c) Check, if there exists $c \in \mathbb{R}$ such that $c\,\boldsymbol{X}'\mathbb{1}_{5\times 5}\boldsymbol{X}$ and $\sum_{i=1}^{5}(X_i - \overline{X})^2 = \boldsymbol{X}' E_5 \boldsymbol{X}$ are stochastically independent. **4 von 4 Punkten**

4 von 4 Punkten

If there exists a unique $c \in \mathbb{R}$, then give its value. If such an $c \in \mathbb{R}$ does not exist, then type "NA" (without quotation marks) into the blank. If the stochastical independence is fulfilled for all $c \in \mathbb{R}$, then type "R" (without quotation marks) into the blank.

R ✓

I.2.11 Theorem

Let $\mathbf{X} \sim N_p(\mu, \Sigma)$ with $\mu \in \mathbb{R}^p, \Sigma \in \mathbb{R}^{p \times p}_{\geqslant 0}$ and $a \in \mathbb{R}^k, B \in \mathbb{R}^{k \times p}$, $1 \leqslant k \leqslant p$. Then:

$$\mathbf{Y} = a + B\mathbf{X} \sim N_k(a + B\mu, B\Sigma B').$$

In particular, we get for $\Sigma \in \mathbb{R}^{p \times p}_{>0}$ and $\Sigma^{-1/2} = (\Sigma^{1/2})^{-1}$

$$\mathbf{Y} = \Sigma^{-1/2}(\mathbf{X} - \mu) \sim N_p(0, I_p).$$

I.2.12 Remark

- The transformation
$$\mathbf{Y} = \Sigma^{-1/2}(\mathbf{X} - \mu)$$
considered in Theorem I.2.11 is called **Mahalanobis transformation**.
- Considering the Euclidean norm $\|\mathbf{Y}\|$ of $\mathbf{Y}$, we get
$$\|\mathbf{Y}\|^2 = \mathbf{Y}'\mathbf{Y} = (\Sigma^{-1/2}(\mathbf{X} - \mu))'\Sigma^{-1/2}(\mathbf{X} - \mu) = (\mathbf{X} - \mu)'\Sigma^{-1}(\mathbf{X} - \mu) = \|\mathbf{X} - \mu\|^2_\Sigma, \text{ say.}$$
$\|\mathbf{X}\|^2_\Sigma$ is called **Mahalanobis norm** of $\mathbf{X}$. For random vectors $\mathbf{X}$ and $\mathbf{Y}$ of the same dimension p, $\|\mathbf{X} - \mathbf{Y}\|^2_\Sigma$ is called **Mahalanobis distance** of $\mathbf{X}$ and $\mathbf{Y}$.

Orthogonal projectors & quadratic forms

I.3.4 Lemma

Let $Q \in \mathbb{R}^{p \times p}$ be an orthogonal projector. Then, an eigenvalue λ of Q satisfies $\lambda \in \{0, 1\}$. Furthermore, $\operatorname{rank}(Q) = \operatorname{trace}(Q)$.

I.3.5 Theorem

Let $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, I_p)$ be a random vector and Q be an orthogonal projector. Then,

$$\mathbf{Y}'Q\mathbf{Y} \sim \chi^2\Big(\operatorname{rank}(Q), \frac{1}{2}\boldsymbol{\mu}'Q\boldsymbol{\mu}\Big).$$

I.3.1 Definition

Let $\mathbf{Y}$ be a random vector and $A \in \mathbb{R}^{p\times p}$ be a symmetric matrix. Then $\mathbf{Y}'A\mathbf{Y}$ is called **quadratic form**.

I.3.2 Remark

- For a random vector $\mathbf{Y} \in \mathbb{R}^n$, $\sum_{i=1}^n Y_i^2$ is a quadratic form (for $A = I_n$): $\sum_{i=1}^n Y_i^2 = \mathbf{Y}'\mathbf{Y} = \mathbf{Y}'I_n\mathbf{Y}$.
- Symmetry of A in Definition I.3.1 is not required, since, from $\mathbf{x}'A\mathbf{x} = \mathbf{x}'A'\mathbf{x}$, we have for any $A \in \mathbb{R}^{p\times p}$

$$\mathbf{x}'A\mathbf{x} = \frac{1}{2}(\mathbf{x}'A\mathbf{x} + \mathbf{x}'A'\mathbf{x}) = \mathbf{x}'A_*\mathbf{x}$$

with the symmetric matrix $A_* = \frac{1}{2}(A + A')$. Therefore, without loss of generality, quadratic forms discussed in the following are based on symmetric matrices.

I.3.3 Lemma

Let $\mathbf{Y}$ be a random vector with $E\mathbf{Y} = \boldsymbol{\mu}$ and $\mathrm{Cov}\,(\mathbf{Y}) = \Sigma$ and $A \in \mathbb{R}^{p\times p}$. Then,

$$E\mathbf{Y}'A\mathbf{Y} = \mathrm{trace}(A\Sigma) + \boldsymbol{\mu}'A\boldsymbol{\mu}.$$

Gesamtpunktzahl **8 von 10 Punkte**

Please provide numbers in the requested precision within each question. The use of different precision is evaluated as wrong.

Load the data **Mental.dat**, which can be found below, into your R workspace. This data shows the results of a study with a random sample of adults in Alachua County (Florida) that investigated the relationship between certain mental health indices and several explanatory variables. The sample size is $n = 40$. Here, the explanatory variables are X_1 which is a life events score and X_2 which is the socioeconomic status (SES). The life events score is a composite measure of both the number and severity of life events the subject experienced within the past three years, such as death in the family or losing a job. It ranged from 3 to 97 in the sample. The SES is a composite index based on occupation, income and education. It ranged from 0 to 100 in the sample. The response variable Y is an index of mental impairment, which incorporates various dimensions of psychatric symptoms, including aspects of anxiety and depression. For the given observations, this measure is ranged from 17 to 41.

Mental

(a) Compute the mean values $\bar{X}_1$ and $\bar{X}_2$, for the following two cases, based each time only on the sub-sample including those observations $x_{1,i}$ for X_1 and $x_{2,i}$ for X_2, respectively, $i \in \{1, \ldots, n\}$, satisfying

(i) $y_i \geq 25$,
(ii) $y_i \in (30, 35]$.

Give the corresponding proportion of samples that satisfy the requirements (i) and (ii) respectively. **(requested precision: 1 digit)**

Answers for (a) (i) **1.5 von 1.5 Punkten**

0.5 von 0.5 Punkten

value for $\bar{X}_1$

49.1

0.5 von 0.5 Punkten

value for $\bar{X}_2$

53.4

Give the corresponding proportion of samples that satisfy the requirements (i) and (ii) respectively. **(requested precision: 1 digit)**

Answers for (a) (i) — 1.5 von 1.5 Punkten

value for $\bar{X}_1$ — 0.5 von 0.5 Punkten

49.1 ✓

value for $\bar{X}_2$ — 0.5 von 0.5 Punkten

53.4 ✓

proportion of samples — 0.5 von 0.5 Punkten

0.7 ✓

Answers for (a) (ii) — 1.5 von 1.5 Punkten

value for $\bar{X}_1$ — 0.5 von 0.5 Punkten

53.9 ✓

value for $\bar{X}_2$ — 0.5 von 0.5 Punkten

35.2 ✓

proportion of samples — 0.5 von 0.5 Punkten

0.2 ✓

(b) Fit a normal linear model using Y as response variable and X_1, X_2 as explanatory variables. What are the standard error values for the least squared estimates of the regression coefficients? **(requested precision: 3 digits)**

Answers for (b) **2 von 2 Punkten**

1 von 1 Punkt

Standard error value for the least squares estimate of the regression coefficient of X_1

0.033 ✓

1 von 1 Punkt

Standard error value for the least squares estimate of the regression coefficient of X_2

0.029 ✓

(c) Fit a normal linear model using Y as response variable and X_1, X_2 as explanatory variables where you restrict the samples on a sub-sample including those $i \in \{1, \ldots, n\}$ satisfying $x_{2,i} \geq 70$. What are the standard error values for the least squared estimates of the regression coefficients? **(requested precision: 3 digits)**

Answers for (c) **2 von 2 Punkten**

1 von 1 Punkt

Standard error value for the least squares estimate of the regression coefficient of X_1

0.054 ✓

1 von 1 Punkt

Standard error value for the least squares estimate of the regression coefficient of X_2

0.164 ✓

(d) Compute the correlation between X_1 and X_2. **(requested precision: 3 digits)**

Answers for (c) **2 von 2 Punkten**

1 von 1 Punkt

Standard error value for the least squares estimate of the regression coefficient of X_1

0.054 ✓

1 von 1 Punkt

Standard error value for the least squares estimate of the regression coefficient of X_2

0.164 ✓

(d) Compute the correlation between X_1 and X_2. **(requested precision: 3 digits)**

Answers for (d) **1 von 1 Punkt**

1 von 1 Punkt

correlation

0.123 ✓

(e) Compute the value of the sum $\sum_{i=1}^{n}(x_{1,i} \cdot x_{2,i})^{1/i^2}$, where $x_{1,i}$ and $x_{2,i}$, $i = 1, \ldots, n$, are the sample values of X_1 and X_2, respectively. **(requested precision: 3 digits)**

Answers for (e) **0 von 2 Punkten**

2 Punkte

value of the sum

3912.821 ✗ 5586.052

Gesamtpunktzahl **7 von 10 Punkte**

Please provide numbers in the requested precision within each question. The use of different precision is evaluated as wrong.

Load the data set `Boston` from the `MASS` library into your data frame and inform yourself about this data set using `library(MASS); ?Boston`.

(a) What is the value of the sample size n and how many variables does this data set contain? **(requested precision: whole numbers)**

Answers for (a) — 1 von 1 Punkt

sample size — 0.5 von 0.5 Punkten

506 ✓

number of variables — 0.5 von 0.5 Punkten

14 ✓

(b) Using `medv` as response variable, fit the full normal linear model. Choose the variable name with the highest and lowest regression coefficient estimate, excluding the intercept.

Answers for (b) — 0 von 1 Punkt

variable with the highest regression coefficient estimate — 0.5 Punkte

rm ✗ nox

(b) Using `medv` as response variable, fit the full normal linear model. Choose the variable name with the highest and lowest regression coefficient estimate, excluding the intercept.

Answers for (b) **0 von 1 Punkt**

0.5 Punkte

variable with the highest regression coefficient estimate

rm ✗ nox

0.5 Punkte

variable with the lowest regression coefficient estimate

nox ✗ rm

(c) Fit a normal linear model using `medv` as response variable and `lstat` and `rm` as predictor variables. Give the resulting error sum of squares (SSE) for this model. **(requested precision: 2 digits)**

Answers for (c) **2 von 2 Punkten**

2 von 2 Punkten

value for SSE

15439.31 ✓

(d) For the linear model fitted in (c), predict the value of the response variable for a case having values of the explanatory variables `lstat` = 10 and `rm` = 6. **(requested precision: 2 digits)**

Answers for (d) **1 von 1 Punkt**

1 von 1 Punkt

predicted value

(c) Fit a normal linear model using `medv` as response variable and `lstat` and `rm` as predictor variables. Give the resulting error sum of squares (SSE) for this model. **(requested precision: 2 digits)**

Answers for (c) **2 von 2 Punkten**

2 von 2 Punkten

value for SSE

15439.31 ✓

(d) For the linear model fitted in (c), predict the value of the response variable for a case having values of the explanatory variables `lstat` = 10 and `rm` = 6. **(requested precision: 2 digits)**

Answers for (d) **1 von 1 Punkt**

1 von 1 Punkt

predicted value

22.79 ✓

(e) Now we turn our view to a model of quadratic regression. Fit a normal linear model using `medv` as response variable and `lstat` and `lstat` ^{2} as explanatory variables. Give the value of the estimated regression coefficient for the explanatory variable `lstat` ^{2}. **(requested precision: 2 digits)**

Answers for (e) **0 von 2 Punkten**

2 Punkte

value of the estimated regression coefficient for the explanatory variable `lstat` ^{2}

64.23 ✗ 0.04

(f) Compare the model of (e) with the simple linear regression model using `medv` as response variable and `lstat` as explanatory variable. Based on the residual standard error of the models, which model would you prefer? **(requested precision: 2 digits)**

(e) Now we turn our view to a model of quadratic regression. Fit a normal linear model using `medv` as response variable and `lstat` and `lstat` ^{2} as explanatory variables. Give the value of the estimated regression coefficient for the explanatory variable `lstat` ^{2}. **(requested precision: 2 digits)**

Answers for (e) **0 von 2 Punkten**

2 Punkte

value of the estimated regression coefficient for the explanatory variable `lstat` ^{2}

64.23 ✗ 0.04

(f) Compare the model of (e) with the simple linear regression model using `medv` as response variable and `lstat` as explanatory variable. Based on the residual standard error of the models, which model would you prefer? **(requested precision: 2 digits)**

Answers for (f) **3 von 3 Punkten**

1 von 1 Punkt

Residual standard error for model in (e)

5.52 ✓

1 von 1 Punkt

Residual standard error for simple linear regression

6.22 ✓

1 von 1 Punkt

Which model would you prefer?

model of (e) ✓