

Aufgabe 1**8 Punkte**

Consider a linear model $\mathbf{Y} = \mathbf{B}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ according to Definition I.4.2 with $n \geq p$ and design matrix $\mathbf{B} \in \mathbb{R}^{n \times p}$ having orthonormal columns, i.e. $\mathbf{B}'\mathbf{B} = \mathbf{I}_p$. Further, let $\mathbf{y} \in \mathbb{R}^n$ be a realization of \mathbf{Y} .

Then, the usual objective function for that linear model $\psi : \mathbb{R}^p \rightarrow [0, \infty)$ defined by

$$\psi(\boldsymbol{\beta}) := \|\mathbf{y} - \mathbf{B}\boldsymbol{\beta}\|_2^2, \quad \boldsymbol{\beta} \in \mathbb{R}^p,$$

with $\|\cdot\|_2$ denoting the corresponding Euclidean norm, leads to the (uniquely determined) *least-squares-estimate* $\widehat{\boldsymbol{\beta}}^{\text{LS}}$ of $\boldsymbol{\beta}$.

As another objective function for that linear model, consider $\tilde{\psi} : \mathbb{R}^p \rightarrow [0, \infty)$ defined by

$$\tilde{\psi}(\boldsymbol{\beta}) := \|\mathbf{y} - \mathbf{B}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_2^2, \quad \boldsymbol{\beta} \in \mathbb{R}^p,$$

with some $\lambda \geq 0$. Let the (also uniquely determined) solution of the corresponding minimization problem

$$\tilde{\psi}(\boldsymbol{\beta}) \longrightarrow \min_{\boldsymbol{\beta} \in \mathbb{R}^p},$$

known as *ridge-estimate* of $\boldsymbol{\beta}$, be denoted by $\widehat{\boldsymbol{\beta}}^{\text{Ridge}}$. Here, we use the same notations for the estimates as for the corresponding estimators, i.e. the corresponding random variables.

Hint: The objective function $\tilde{\psi}$ can be represented as follows:

$$\tilde{\psi}(\boldsymbol{\beta}) = \|\tilde{\mathbf{y}} - \widetilde{\mathbf{B}}\boldsymbol{\beta}\|_2^2, \quad \boldsymbol{\beta} \in \mathbb{R}^p,$$

1 2 3 4

ABGABE

ÜBERSICHT



(b) In this part, assume $p = 2$ and that the variance parameter of the considered linear model is given by $\sigma^2 = 9$. Further assume that the two estimators fulfill the following equation:

$$\hat{\beta}^{\text{Ridge}} = c(\lambda) \hat{\beta}^{\text{LS}}$$

with $c(\lambda) = \frac{2}{3}$ for some suitably chosen $\lambda \geq 0$.

Calculate the entries of the covariance matrix $\text{Cov}(\hat{\beta}^{\text{Ridge}}) = \begin{pmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{pmatrix}$.

3 Punkte

Give the value of c_{11} .

Zahl

1 Punkt

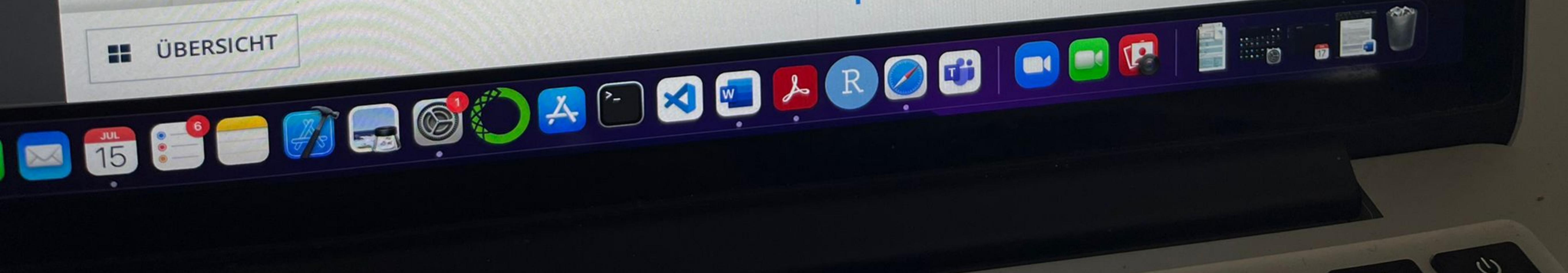
Give the value of c_{12} .

Zahl

1 2 3 4

ABGABE

ÜBERSICHT





Calculate the entries of the covariance matrix $\text{Cov}(\hat{\beta}^{\text{Ridge}}) = \begin{pmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{pmatrix}$.

3 Punkte

Give the value of c_{11} .

1 Punkt

Zahl

1 Punkt

Give the value of c_{12} .

Zahl

1 Punkt

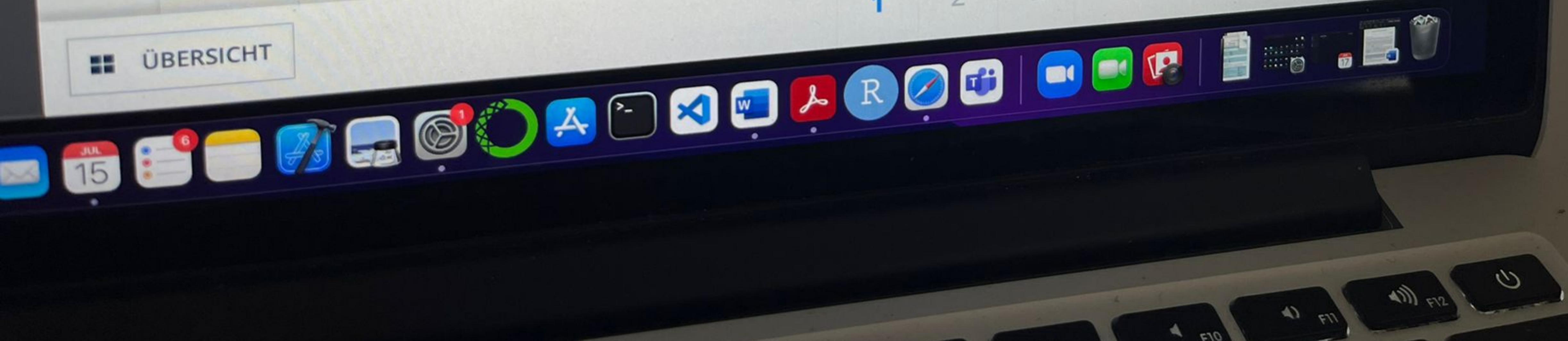
Give the value of c_{22} .

Zahl

ÜBERSICHT

1 2 3 4

ABGABE



Aufgabe 2

Let X_1, \dots, X_n be stochastically independent and identically distributed continuous random variables, each having a density function $f(\bullet; \alpha, \beta) : \mathbb{R} \rightarrow [0, \infty)$ defined by

$$(*) \quad f(x; \alpha, \beta) := \begin{cases} \frac{1}{\alpha} \exp\left(-\frac{x-\beta}{\alpha}\right) & , \quad x \geq \beta, \\ 0 & , \quad x < \beta, \end{cases}$$

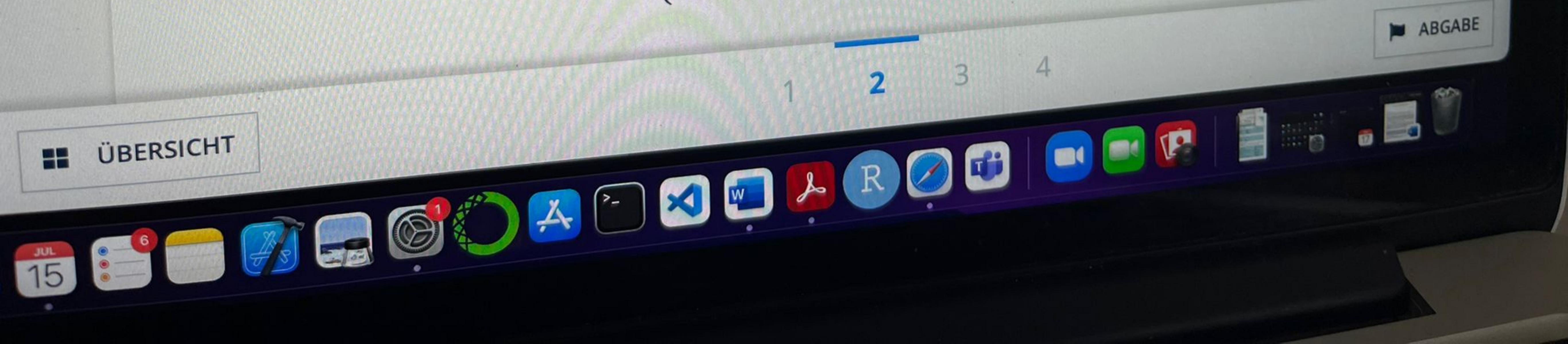
with parameters $\alpha > 0$ and $\beta > 0$.

In the following tasks, round your derived numerical solutions to three decimal places, if necessary.

(a) For a given realization $\mathbf{x} = (x_1, \dots, x_n)' \in (0, \infty)^n$ of $\mathbf{X} = (X_1, \dots, X_n)'$, the corresponding (full) likelihood function $L(\bullet | \mathbf{x}) : (0, \infty)^2 \rightarrow [0, \infty)$ is given by

$$(**) \quad L(\alpha, \beta | \mathbf{x}) = \begin{cases} \alpha^{-n} \exp\left(-\frac{n}{\alpha}(\bar{x} - \beta)\right), & \alpha \in (0, \infty), \beta \in (0, x_{(1)}], \\ 0, & \alpha \in (0, \infty), \beta \in (x_{(1)}, \infty), \end{cases}$$

ÜBERSICHT



where $x_{(1)} := \min\{x_1, \dots, x_n\}$ denotes the minimum and $\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i$ denotes the mean of the given realization $\mathbf{x} = (x_1, \dots, x_n)'$.

In this part, consider the profile likelihood approach for estimating the parameter α . Let $\hat{\beta}(\alpha)$ denote the maximum likelihood estimate of the nuisance parameter β for fixed $\alpha \in (0, \infty)$.

Suppose that the following sample of size $n = 4$ has been observed:

$$x_1 = 4, x_2 = 8, x_3 = 2, x_4 = 6.$$

- (i)** Based on the observed sample, calculate the corresponding estimate $\hat{\beta}(\alpha)$ of β for $\alpha = 5$ 4 Punkte

4 Punkte

Give the value of $\hat{\beta}(\alpha)$.

Zahl

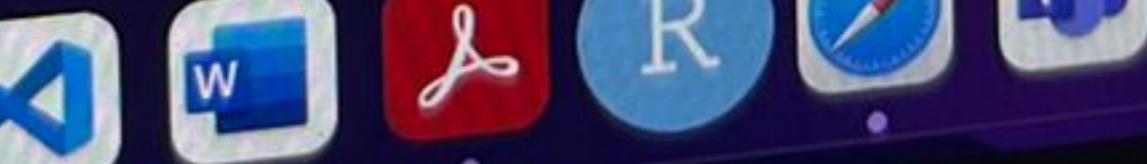
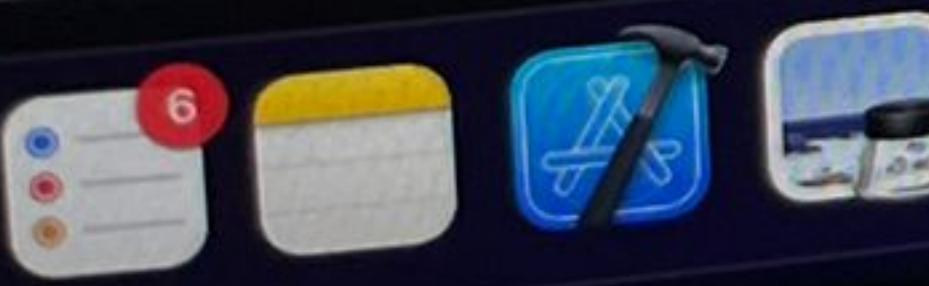
ABGABE

- (ii)** Based on the observed sample, calculate the corresponding profile maximum likelihood estimate $\hat{\alpha}^{\text{profile}}$ of α . 4 Punkte

1 2 3 4

ÜBERSICHT

JUL
15



with $\tilde{B} := \begin{pmatrix} B \\ \sqrt{\lambda} I_p \end{pmatrix} \in \mathbb{R}^{(n+p) \times p}$ and $\tilde{y} := \begin{pmatrix} y \\ 0_p \end{pmatrix} \in \mathbb{R}^{n+p}$.

(a) Under the given assumptions, show that the following equation holds:

$$\hat{\beta}^{\text{Ridge}} = c(\lambda) \hat{\beta}^{\text{LS}}$$

with some factor $c(\lambda) \in \mathbb{R}$, which (only) depends on λ .

:= Calculate the factor $c(\lambda)$ for $\lambda = \frac{1}{4}$.

5 Punkte

5 Punkte

Give the value of $c(\lambda)$ with a precision of **one decimal place**.

Zahl

(b) In this part, assume $p = 2$ and that the variance parameter of the considered linear model is given by $\sigma^2 = 9$. Further

1 2 3 4

ABGABE

ÜBERSICHT

Here, $E(Y | X = x)$ denotes the expectation with respect to the distribution $P^{Y|X=x}$ for $x \in (0, \infty)$.

For $\alpha = 5$ and $\beta = 1$, calculate the expectation $E(Y)$.

4 Punkte

Give the value of $E(Y)$.

4 Punkte

Nächste
Aufgabe →

Vorherige
← Aufgabe

ABGABE

Alle Antworten wurden gespeichert!

1 2 3 4

ÜBERSICHT



R Task 1

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

Consider the following contingency table describing the attitude of Germans to genetically modified (GM) foods according to their income (low/middle/high). Let X denote the attitude (for GM foods/against GM foods) and Y the income (low/middle/high). This contingency table is a realization of a random contingency tables with independent cell entries $N_{ij} \sim \mathcal{P}(\mu_{ij})$ for $i = 1, 2$ and $j = 1, 2, 3$.

	High income	Middle income	Low income	Total
For GM foods	320	280	258	858
Against GM foods	150	166	112	428
Total	470	446	370	1286

Read the contingency table above in R in a `data.frame` format, appropriate for fitting log-linear models with the `glm` function.

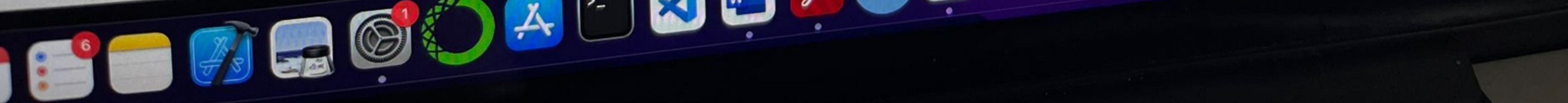
(a) Provide the marginal probabilities π_{+1} and π_{2+} . (requested precision: 2 digits)

1 Punkt

ABGABE

1 2 3 4

ÜBERSICHT



(a) Provide the marginal probabilities π_{+1} and π_{2+} . (requested precision: 2 digits)

1 Punkt

 π_{+1}

0.5 Punkte

Zahl

 π_{2+}

0.5 Punkte

Zahl

(b) Fit the log-linear model of independence that estimates the expected cell frequencies μ_{ij} , $i = 1, 2, j = 1, 2, 3$, under the hypothesis of independence of the classification variables X and Y . Provide the corresponding Pearsonian residual for the case of responders having medium income and being against GM foods. Provide the value of the Pearson's χ^2 goodness of fit statistic. (requested precision: 2 digits)

2 Punkte

1 Punkt

Pearsonian residual for the case of responders having medium income and being against GM foods

ABGABE

1 2 3 4

BERSICHT

Pearsonian residual for the case of responders having medium income and being against GM foods

1 Punkt

Zahl

Pearsons χ^2 statistic

1 Punkt

Zahl

(c) For the model fitted in (b), provide the value of the likelihood ratio statistic G^2 . Further, provide the corresponding asymptotic p -value. Based on G^2 and at significance level $\alpha = 0.05$, does the independence model adequately describe this dataset? (requested precision: 2 digits)

2 Punkte

1 Punkt

value of likelihood ratio statistic G^2

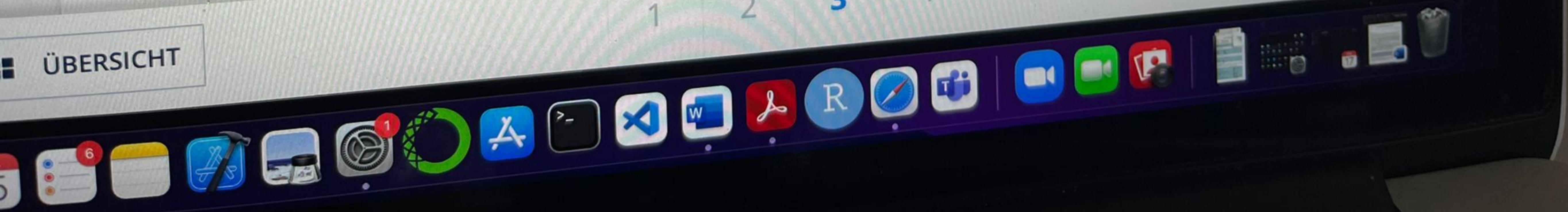
Zahl

0.5 Punkte

ABGABE

1 2 3 4

ÜBERSICHT



(ii) Based on the observed sample, calculate the corresponding profile maximum likelihood estimate $\hat{\alpha}^{\text{profile}}$ of α .

4 Punkte

Give the value of $\hat{\alpha}^{\text{profile}}$

4 Punkte

Zahl

(b) In this part, let X be a continuous random variable with corresponding density function $f(\bullet; \alpha, \beta)$ given by (*) for some $\alpha, \beta > 0$. Further, let Y be a discrete random variable taking values in \mathbb{N}_0 with the conditional distribution of Y under the condition $X = x$, $x \in (0, \infty)$, being a Poisson distribution with parameter x . i.e.,

$$P^{Y|X=x} = P(x), \quad x \in (0, \infty),$$

with P denoting the underlying probability distribution.

Then, by conditioning on X , the expectation $E(Y)$ of Y can be determined as follows:

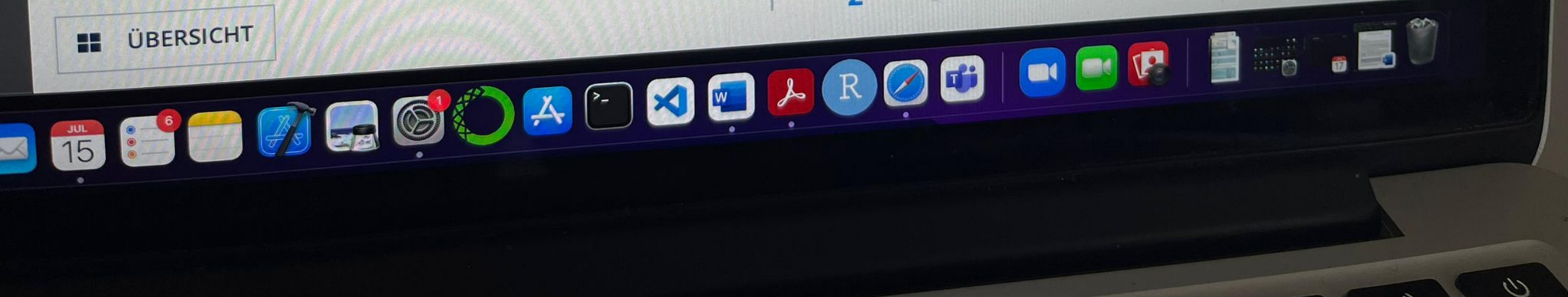
$$E(Y) = E(E(Y|X)) = \int_{-\infty}^{\infty} E(Y|X=x) f(x; \alpha, \beta) dx.$$

Here, $E(Y|X=x)$ denotes the expectation with respect to the distribution $P^{Y|X=x}$ for $x \in (0, \infty)$.

1 2 3 4

ABGABE

ÜBERSICHT



1 Punkt

(e) Starting with the model in (d), select an appropriate nested model based on AIC using backward selection. Give the resulting value for AIC of this selected model. (requested precision: 2 digits)

AIC

1 Punkt

Zahl

(f) For the model selected in (e), give the 95% asymptotic (Wald) confidence interval for the intercept (requested precision: 3 digits)

2 Punkte

lower bound

1 Punkt

Zahl

upper bound

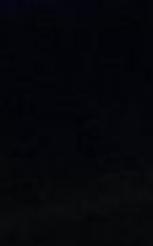
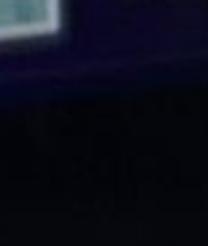
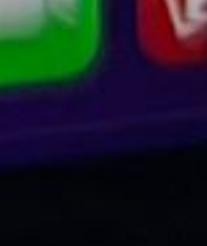
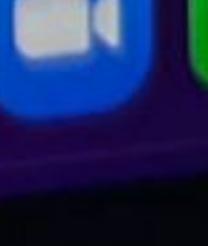
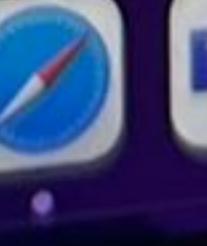
1 Punkt

Zahl

ABGABE

1 2 3 4

ÜBERSICHT

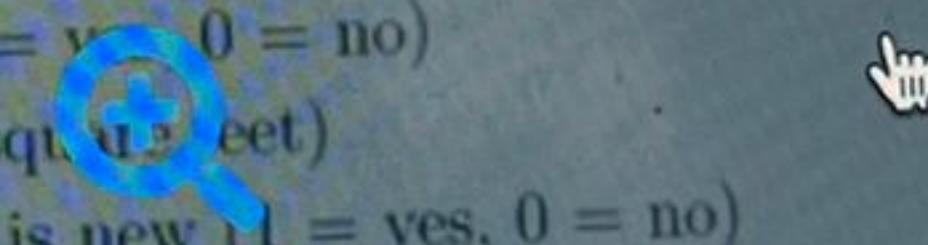


R Task 2

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

The file **Houses.txt**, which can be downloaded by clicking on the button "Houses", contains data with information of 100 home sales in Gainesville, Florida. Find the details of the dataset below.

attribute	description & properties
case	index
price	whether the selling price of the house (in thousands of dollars) is ≥ 200 or not (1 = yes, 0 = no)
size	size of the house (square feet)
new	whether the house is new (1 = yes, 0 = no)
taxes	tax bill (dollars)
bedrooms	number of bedrooms
baths	number of bathrooms

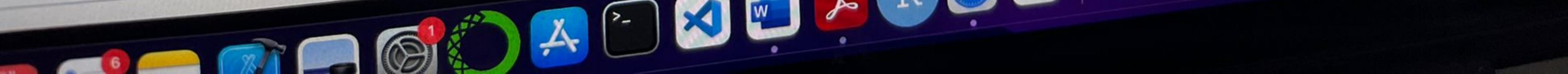


Houses

1 2 3 4

ABGABE

ÜBERSICHT



(a) Load the data file into your workspace and transform the attribute `new` into a factor variable. Obtain the number of houses in the dataset for which the tax bills are greater than 3000 dollars and the number of bedrooms are at least equal to 2. 1 Punkt

value

Zahl

(b) Fit a GLM using the canonical link from predicting the value of `price` using the variables `size`, `new`, `taxes`, `bedrooms`, `baths` as explanatory variables. Provide the resulting value for the estimated coefficient of the attribute `bedrooms`. (requested precision: 3 digits) 1 Punkt

value for estimated coefficient of attribute `bedrooms`

Zahl

(c) Let β_4 denote the coefficient of the attribute `taxes` in the model fitted in (b) above. Test at significance level $\alpha = 0.01$ the hypotheses $H_0 : \beta_4 = 0$ versus $H_1 : \beta_4 \neq 0$. Base your test on

ABGABE

1 2 3 4

ÜBERSICHT



(c) Let β_4 denote the coefficient of the attribute **taxes** in the model fitted in (b) above. Test at significance level $\alpha = 0.01$ the hypotheses $H_0 : \beta_4 = 0$ versus $H_1 : \beta_4 \neq 0$. Base your test on the Wald's test statistic and provide the *p*-values and your decision. (requested precision: 3 digits)

1 Punkt

0.5 Punkte

p-value

Zahl

0.5 Punkte

decision

Bitte auswählen ▾

(d) Based on the model fitted in (b), predict the probability of having a price ≥ 200 thousands dollars for a house that is new, has a size of 1000 square feet, a tax bill of 1200 dollars, 2 bedrooms and 2 baths. (requested precision: 3 digits)

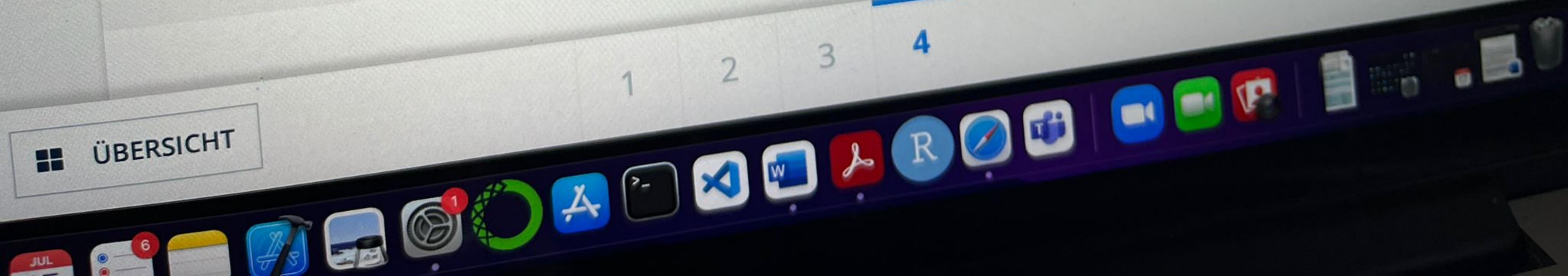
3 Punkte

3 Punkte

ABGABE

1 2 3 4

ÜBERSICHT



asymptotic p -value01:28:22 

0.5 Punkte

Zahl

Based on G^2 and at significance level $\alpha = 0.05$, does the independence model adequately describe this dataset? 0.5 Punkte

Bitte auswählen ▾

- ☰ (d) Fit the saturated log-linear model. Give the resulting estimate for λ . (requested precision: 2 digits) 2 Punkte

2 Punkte

estimate for λ

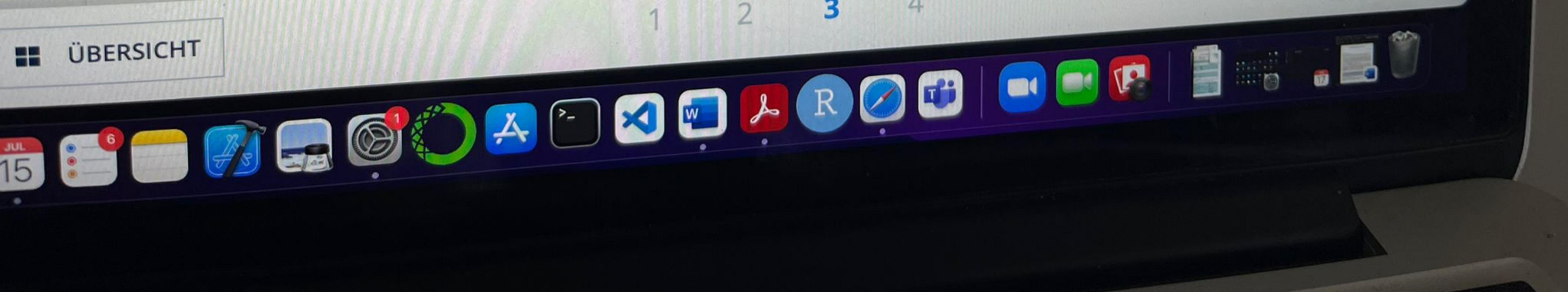
Zahl

- ☰ (e) Starting with the model in (d), select an appropriate nested model based on AIC using backward selection. Give the resulting value for AIC of this selected model. (requested precision: 2 digits) 1 Punkt

1 2 3 4

ABGABE

ÜBERSICHT



Zahl

(d) Obtain the area under the curve (AUC) for the ROC curve of the model in (b). **(requested precision: 2 digits)**

2 Punkte

AUC

2 Punkte

Zahl

Vorherige
← Aufgabe

Alle Antworten wurden gespeichert!

ABGABE

1 2 3 4

- (d) Based on the model fitted in (b), predict the probability of having a price ≥ 200 thousands dollars for a house that is new, has a size of 1000 square feet, a tax bill of 1200 dollars, 2 bedrooms and 2 baths. (requested precision: 3 digits)

3 Punkte

predicted probability

3 Punkte

Zahl

- (e) For the model fitted in (b), compute 95% profile likelihood confidence interval for β_6 (coefficient of baths). (requested precision: 3 digits)*

2 Punkte

lower bound

1 Punkt

Zahl

1 Punkt

upper bound

Zahl

ABGABE

1 2 3 4

ÜBERSICHT

