ulm university universität **uulm**

Dr. Tim Brereton
Lisa Handl

Winter Term 2014/15

Stochastics III

Problem Sheet 3

Deadline: November 26, 2014 at noon, before the practical

Exercise 1 (3 + 3 points)

a) Let $\boldsymbol{Z} = (Z_1, Z_2, Z_3)^\top \sim N(\boldsymbol{\mu}, I_3)$ with expectation vector $\boldsymbol{\mu} = (1, 7, -5)^\top$ and covariance matrix I_3. Determine the distribution of

$$\frac{1}{2}Z_1^2 + Z_2^2 + \frac{1}{2}Z_3^2 - Z_1 Z_3 \,.$$

b) Let $\boldsymbol{Z} = (Z_1, Z_2, Z_3)^\top \sim \mathrm{N}(\boldsymbol{\mu}, K)$ with expectation vector $\boldsymbol{\mu} = (1, -3, 2)^\top$ and covariance matrix

$$K = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Determine the distribution of

$$Z_1^2 + Z_2^2 + Z_3^2 + 2Z_1Z_2 - 2Z_1Z_3 - 2Z_2Z_3 \,.$$

Exercise 2 (4 points)

Consider the linear model $\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with $\mathbb{E}\boldsymbol{\varepsilon} = \boldsymbol{o}$, $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 I_n$ and $\mathrm{rk}(X) = m$. Let $\widehat{\boldsymbol{\beta}}$ be the least squares estimator for $\boldsymbol{\beta}$ and let $\widehat{\boldsymbol{\varepsilon}} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}$ with $\widehat{\boldsymbol{Y}} = X\widehat{\boldsymbol{\beta}}$ be the vector of residuals. Show that $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\varepsilon}}) = \mathbb{E}\left((\hat{\boldsymbol{\beta}} - \mathbb{E}\hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\epsilon}} - \mathbb{E}\hat{\boldsymbol{\epsilon}})^\top\right) = \boldsymbol{o}$.

Exercise 3 (2 + 3 + 2 points)

In the file *murders.txt* on the lecture website you find data on poverty (the percentage of people with incomes below \$5000), the unemployment rate (in percent), and the number of murders (per 1 000 000 inhabitants per year) in 20 different states. Download it (or copy it from the table below) and work through the following exercises.

a) Write out a linear model how the number of murders depends on poverty and the unemployment rate, and set up its design matrix in **R**. Make sure your model includes the possibility of a non-zero intercept.

b) Calculate the least squares estimate for the regression coefficients of your model **without using the command `lm` or similar**. You may use **R** as a calculator only. Explain what the estimated coefficients describe.

c) The regression function of this model is a function from $\mathbb{R}^2$ to $\mathbb{R}$. Plot it in 3D together with the data points. Use the commands `persp3d` and `points3d` from the `rgl` package.

poverty	unemployment	murders
16.5	6.2	11.2
20.5	6.4	13.4
26.3	9.3	40.7
16.5	5.3	5.3
19.2	7.3	24.8
16.5	5.9	12.7
20.2	6.4	20.9
21.3	7.6	35.7
17.2	4.9	8.7
14.3	6.4	9.6
18.1	6	14.5
23.1	7.4	26.9
19.1	5.8	15.7
24.7	8.6	36.2
18.6	6.5	18.1
24.9	8.3	28.9
17.9	6.7	14.9
22.4	8.6	25.8
20.2	8.4	21.7
16.9	6.7	25.7

Exercise 4 (3 + 2 + 2 points)

The production volume of the USA between 1932 and 1953 can be described with the Cobb-Douglas production function

$$Y_t = e^{\beta_1} \cdot K_t^{\beta_2} \cdot A_t^{\beta_3} \cdot \varepsilon_t, \quad t \in \{1, \dots, 22\} \tag{1}$$

with unknown constants $\beta_1, \beta_2, \beta_3 \in \mathbb{R}$. Here Y denotes the production (in billion dollars), K denotes the capital expenditure (in billion dollars), and A denotes the work input (in millions of workers). On the lecture website you will find a file *production.txt* containing the following data:

year	t	production	capital	work
1932	1	60.3	297.1	39.3
1933	2	58.2	290.1	39.6
1934	3	64.4	285.4	42.7
1935	4	75.4	287.8	44.2
1936	5	85.0	282.1	47.1
1937	6	92.7	300.3	48.2
1938	7	85.4	301.4	46.4
1939	8	92.3	305.6	47.8
1940	9	101.2	313.3	49.6
1941	10	113.3	327.4	54.1
1942	11	107.8	339.0	59.1
1943	12	105.2	347.1	64.9
1944	13	107.1	353.5	66.0
1945	14	108.8	354.1	64.4
1946	15	131.5	359.4	58.9
1947	16	130.9	359.3	59.3
1948	17	134.7	365.2	60.2
1949	18	129.1	363.2	58.7
1950	19	147.8	373.7	60.0
1951	20	152.1	386.0	63.8
1952	21	154.3	396.5	64.9
1953	22	159.9	408.0	66.0

Work through the following exercises using **R**.

a) Transform the modeling approach (1) into a suitable linear model and determine the least squares estimate for $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$. This time you may use `lm`.

b) Provide the estimated values $\widehat{Y}_t$ as well as the residuals $\widehat{\varepsilon}_t$.

c) Plot the estimated production volume Y from 1932 to 1953 together with the real values.

   Hint: You can add a polygonal track to an existing plot using the command `lines`.