Dr. Willi Mutschler
M.Sc. Manuel Stapper
Winter term 2017/2018

Westfälische Wilhelms-Universität Münster

Econometrics I

– Exercise Book –

Exercise 1

Let $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ be a matrix and let $a = (a_1, ..., a_n)'$, $x = (x_1, ..., x_n)'$, $y = (y_1, ..., y_n)'$ be vectors of length n.

a) Write the sum $\sum_{i=1}^{n} a_i \cdot x_i$ as a product of vectors.

> $\sum_{i=1}^{n} a_i \cdot x_i = a'x$

b) Write $c_i = \sum_{j=1}^{n} a_{ij} \cdot x_j$ in matrix form.

> $\sum_{j=1}^{n} a_{ij} \cdot x_j = A \cdot x$

c) Write $\sum_{i=1}^{n} y_i$ in matrix form. Hint: Use a vector of ones.

> Let $\mathbb{1}_n$ be a column vector of length n containing only ones.
> $\sum_{i=1}^{n} y_i = \mathbb{1}_n' \cdot y = y' \cdot \mathbb{1}_n$

d) Write $\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}$ in matrix form.

> $\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} = \mathbb{1}_n' \cdot A \cdot \mathbb{1}_n$

Exercise 2

Consider the following matrices

$$A = \begin{pmatrix} 1 & -1 & 2 \\ 3 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \qquad C = \begin{pmatrix} 2 & 0 & 3 \\ 7 & -1 & 2 \\ 0 & 1 & 1 \end{pmatrix}$$

$$D = \begin{pmatrix} 2 & 3 \\ -2 & 1 \\ 0 & 5 \end{pmatrix}, \qquad E = \begin{pmatrix} 0 & 2 & 1 \end{pmatrix}.$$

a) Calculate (if possible): AB, CD, EA, BE, ED.

$$AB = \begin{pmatrix} 5 \\ 6 \\ 4 \end{pmatrix}, \qquad CD = \begin{pmatrix} 4 & 21 \\ 16 & 30 \\ -2 & 6 \end{pmatrix} \qquad EA = \begin{pmatrix} 8 & 1 & 2 \end{pmatrix}$$

$$BE = \begin{pmatrix} 0 & 2 & 1 \\ 0 & 4 & 2 \\ 0 & 6 & 3 \end{pmatrix}, \qquad ED = \begin{pmatrix} -4 & 7 \end{pmatrix}$$

b) Verify that: $(A+C)' = A' + C'$ and $(AC)' = C'A'$.

$$\begin{aligned}
(A+C)' &= \left[\begin{pmatrix} 1 & -1 & 2 \\ 3 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix} + \begin{pmatrix} 2 & 0 & 3 \\ 7 & -1 & 2 \\ 0 & 1 & 1 \end{pmatrix}\right]' \\
&= \begin{pmatrix} 3 & -1 & 5 \\ 10 & -1 & 3 \\ 2 & 2 & 1 \end{pmatrix}' \\
&= \begin{pmatrix} 3 & 10 & 2 \\ -1 & -1 & 2 \\ 5 & 3 & 1 \end{pmatrix} \\
A' + C' &= \begin{pmatrix} 1 & 3 & 2 \\ -1 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix} + \begin{pmatrix} 2 & 7 & 0 \\ 0 & -1 & 1 \\ 3 & 2 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 3 & 10 & 2 \\ -1 & -1 & 2 \\ 5 & 3 & 1 \end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
(AC)' &= \left[\begin{pmatrix} 1 & -1 & 2 \\ 3 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 2 & 0 & 3 \\ 7 & -1 & 2 \\ 0 & 1 & 1 \end{pmatrix}\right]' \\
&= \begin{pmatrix} -5 & 6 & 11 \\ 3 & 1 & -1 \\ 3 & 10 & 8 \end{pmatrix} = C'A'
\end{aligned}$$

Exercise 3

Consider the following matrices

$$A = \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & -2 \\ 1 & -3 \end{pmatrix}.$$

a) Calculate the determinant of A, B, AB, A', $2A$ and $-B$.

$$\det(A) = 3, \qquad \det(B) = 2, \qquad \det(AB) = 6 = \det(A)\det(B),$$
$$\det(A') = 3 = \det(A), \qquad \det(2A) = 12 = 2^2\det(A),$$
$$\det(-B) = 2 = \det(B) = (-1)^2\det(B)$$

b) Calculate A^{-1} and B^{-1}.

$$A^{-1} = \frac{1}{3}\cdot\begin{pmatrix} 1 & 1 \\ -2 & 1 \end{pmatrix}, \qquad B^{-1} = \frac{1}{2}\cdot\begin{pmatrix} -3 & 2 \\ -1 & 0 \end{pmatrix}$$

c) Calculate $(AB)^{-1}$.

$$(AB)^{-1} = -\frac{1}{6}\begin{pmatrix} 7 & 1 \\ 1 & 1 \end{pmatrix}$$

d) Calculate $B^{-1}A^{-1}$ and compare with (b).

$$B^{-1}A^{-1} = -\frac{1}{6}\begin{pmatrix} 7 & 1 \\ 1 & 1 \end{pmatrix} = (AB)^{-1}$$

e) Are the matrices A and $C = A'A$ positive definite, negative definite or indefinite?

Approach 1:
The matrix A is positive definite, if $x'Ax > 0$ for every vector $x \neq 0$,
negative definite if $x'Ax > 0$
and indefinite, if there exist x, y for which $x'Ax > 0$ and $y'Ay < 0$.

$$\begin{aligned} x'Ax &= \begin{pmatrix} x_1 & x_2 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= x_1^2 + x_1x_2 + x_2^2 \end{aligned}$$

Case 1: $x_1x_2 \geq 0$: $x'Ax > 0$
Case 2: $x_1x_2 < 0$:
$x'Ax = x_1^2 + x_1x_2 + x_2^2 > x_1^2 + 2x_1x_2 + x_2^2 = (x_1 + x_2)^2 > 0$
$\Rightarrow x'Ax > 0$ for every vector $x \neq 0 \Rightarrow A$ is positive definite.

Approach 2:
A matrix A is positive definite if every principle minor is positive,
negative definite if odd principle minors are negative and even principle minors are positive
and indefinite if the determinant is not zero and the matrix is neither p.d. nor n.d.
For a matrix

$$C = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

the first principle minor is a and the second $\det(C) = ad - bc$.
The first principle minor of A is 1 and the second principle minor is 3, so A is positive definite.

$A'A$ is a so-called quadratic form and thus positive definite. This can be seen as follows:

$$x'A'Ax = (x'A')(Ax) = (Ax)' \underbrace{Ax}_{=:y} = y'y = y_1^2 + y_2^2 + ... + y_n^2 \geq 0$$

Both of the above approaches would of course give the same result.

f) Give an example of a negative definite and indefinite (2×2) matrix.

Since A is positive definite, $(-1) \cdot A$ is negative definite:
$x'Ax > 0 \Leftrightarrow (-1) \cdot x'Ax = x'(-1) \cdot Ax < 0$
The matrix $D = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ is indefinite, because there are two vectors $x = \begin{pmatrix} 1 & 0 \end{pmatrix}'$ and $y = \begin{pmatrix} 0 & 1 \end{pmatrix}'$ so that $x'Dx = 1 > 0$ and $y'Dy = -1 < 0$.

Exercise 4

In this exercise we take a look at an example to understand why the unbiased variance estimator has a factor $\frac{1}{T-1}$ instead of $\frac{1}{T}$. Imagine a box containing 3 coins with values 0, 2 and 4 denoted as x_1, x_2 and x_3 respectively. Please follow the steps:

- Calculate the *population sample*

$$\mu := \frac{x_1 + x_2 + x_3}{3}$$

  and the *population variance*

$$\sigma^2 = \frac{1}{3} \sum_{t=1}^{3} (x_i - \mu)^2 \ .$$

- We try to estimate the population variance by drawing a coin $T = 2$ times. Write down every combination of two values from the population. Hint: $(0, 2) \neq (2, 0)$
- For every of these 9 combinations, calculate the sample average and the sample variance with both factors $\frac{1}{T}$ and $\frac{1}{T-1}$.
- Calculate the average of these three key figures over all 9 samples.
- What do you notice?

Sample	Sample Mean	Sample variance ($^1/_T$)	Sample variance ($^1/_{T-1}$)
(0,0)	0	0	0
(0,2)	1	1	2
(0,4)	2	4	8
(2,0)	1	1	2
(2,2)	2	0	0
(2,4)	3	1	2
(4,0)	2	4	8
(4,2)	3	1	2
(4,4)	4	0	0
Average	2	$^4/_3$	$^8/_3$

Exercise 5

An investor wants to invest in two possible, independent stocks A and B. The return of those two stocks is denoted as R_A and R_B. To evaluate the stocks we assume

$$\mathrm{E}(R_A) = 2 \qquad \mathrm{E}(R_B) = 5 \qquad \mathrm{Var}(R_A) = 2 \qquad \mathrm{Var}(R_B) = 6$$

a) Calculate the expected return in case the investor splits his capital evenly.

Let p denote the proportion of capital invested in B. The return on the portfolio is then expressed as $(1-p) \cdot R_A + p \cdot R_B$.

$$\mathrm{E}\left[(1-p) \cdot R_A + p \cdot R_B\right] = (1-p) \cdot \mathrm{E}(R_A) + p \cdot \mathrm{E}(R_B)$$

For $p = \frac{1}{2}$ it yields $\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 5 = 3.5$

b) Calculate the variance of the above described portfolio. What do you notice?

$$\mathrm{Var}\left[(1-p) \cdot R_A + p \cdot R_B\right] = (1-p)^2 \cdot \mathrm{Var}(R_A) + p^2 \cdot \mathrm{Var}(R_B)$$

And for $p = \frac{1}{2}$: $\frac{1}{4} \cdot 2 + \frac{1}{4} \cdot 6 = 2$

c) The investor wants to limit his risk, so that the variance in the portfolio return does not exceed 3.5. What is the maximal return on such a portfolio? How much so you suggest him to invest in A and in B?

Using the above formula for the portfolio variance:

$$\begin{aligned}(1-p)^2 \cdot \mathrm{Var}(R_A) + p^2 \cdot \mathrm{Var}(R_B) &= (1-p)^2 \cdot 2 + p^2 \cdot 6 \\ &= 8p^2 - 4p + 2\end{aligned}$$

Solving $8p^2 - 4p + 2 = 3.5$ for p yields $p = \frac{3}{4}$.

Exercise 6

Let a be a column vector of length n and A an $(n \times n)$ matrix. For a scalar valued function $f(x)$ of a vector x,

$$\frac{\partial f(x)}{\partial x} = \begin{bmatrix} \partial f/\partial x_1 \\ \vdots \\ \partial f/\partial x_n \end{bmatrix}$$

is called the gradient vector of f. For the vector $y = Ax$, define

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \partial y_1/\partial x' \\ \partial y_2/\partial x' \\ \vdots \\ \partial y_n/\partial x' \end{bmatrix}.$$

a) Show the following equations:

i) $\frac{\partial a'x}{\partial x} = a$

$$\begin{aligned} \frac{\partial a'x}{\partial x} &= \frac{\partial}{\partial x} \sum_{i=1}^{n} a_i x_i \\ &= \begin{bmatrix} \frac{\partial}{\partial x_1} \sum_{i=1}^n a_i x_i \\ \vdots \\ \frac{\partial}{\partial x_n} \sum_{i=1}^n a_i x_i \end{bmatrix} \\ &= \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = a \end{aligned}$$

ii) $\frac{\partial Ax}{\partial x} = A$

$$\begin{aligned} \frac{\partial Ax}{\partial x} &= \begin{bmatrix} \frac{\partial}{\partial x'} A_{1\bullet} x \\ \vdots \\ \frac{\partial}{\partial x'} A_{n\bullet} x \end{bmatrix} \\ &\overset{i)}{=} \begin{bmatrix} A_{1\bullet} \\ \vdots \\ A_{n\bullet} \end{bmatrix} = A \end{aligned}$$

iii) $\frac{\partial x'Ax}{\partial x} = (A + A')x$

$$
\begin{aligned}
\frac{\partial x'Ax}{\partial x} &= \frac{\partial}{\partial x} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \\
&= \begin{bmatrix} \frac{\partial}{\partial x_1} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \\ \vdots \\ \frac{\partial}{\partial x_n} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \end{bmatrix} \\
&= \begin{bmatrix} \left(\sum_{i=2}^{n} a_{1i} x_i\right) + 2a_{11} x_1 + \left(\sum_{i=2}^{n} a_{i1} x_i\right) \\ \vdots \\ \left(\sum_{i=2}^{n} a_{ni} x_i\right) + 2a_{nn} x_1 + \left(\sum_{i=2}^{n} a_{in} x_i\right) \end{bmatrix} \\
&= \begin{bmatrix} \left(\sum_{i=1}^{n} a_{1i} x_i\right) + \left(\sum_{i=1}^{n} a_{i1} x_i\right) \\ \vdots \\ \left(\sum_{i=1}^{n} a_{ni} x_i\right) + \left(\sum_{i=1}^{n} a_{in} x_i\right) \end{bmatrix} \\
&= Ax + A'x = (A + A')x
\end{aligned}
$$

b) What does a) iii) imply for $\frac{\partial x'Ax}{\partial x}$ if A is symmetric?

If A is symmetric ($A' = A$), then $\frac{\partial x'Ax}{\partial x} = 2Ax$

Exercise 7

Consider the following linear regression:

$$
y_t = \alpha + \beta x_t + u_t \ , \qquad \text{for } t = 1, ..., T \ .
$$

Show that the OLS estimators $\hat{\alpha}$ and $\hat{\beta}$ are given by

$$
\hat{\beta} = \frac{\sum_{t=1}^{T} (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{T} (x_t - \bar{x})^2} \ , \qquad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}
$$

Minimize $\sum_{t=1}^{T}[y_t-(\alpha+\beta x_t)]^2$. Therefore derive w.r.t. α and β.

$$
\begin{aligned}
& \frac{\partial}{\partial \alpha}\sum_{t=1}^{T}(y_t-\alpha-\beta x_t)^2 \stackrel{!}{=} 0 \\
\Leftrightarrow \quad & \sum_{t=1}^{T} 2\cdot(y_t-\alpha-\beta x_t)\cdot(-1)=0 \\
\Leftrightarrow \quad & \sum_{t=1}^{T}(y_t-\alpha-\beta x_t)=0 \\
\Leftrightarrow \quad & \sum_{t=1}^{T} y_t-\sum_{t=1}^{T}\beta x_t=\sum_{t=1}^{T}\alpha \\
\Leftrightarrow \quad & \sum_{t=1}^{T} y_t-\beta\sum_{t=1}^{T} x_t=T\alpha \\
\Leftrightarrow \quad & \bar{y}-\beta\bar{x}=\alpha \qquad (\star)
\end{aligned}
$$

$$
\begin{aligned}
& \frac{\partial}{\partial \beta}\sum_{t=1}^{T}(y_t-\alpha-\beta x_t)^2 \stackrel{!}{=} 0 \\
\Leftrightarrow \quad & \sum_{t=1}^{T} 2\cdot(y_t-\alpha-\beta x_t)\cdot(-x_t)=0 \\
\Leftrightarrow \quad & \sum_{t=1}^{T}(y_t-\alpha-\beta x_t)\cdot x_t=0 \\
\Leftrightarrow \quad & \sum_{t=1}^{T} y_t x_t-\alpha\cdot\sum_{t=1}^{T} x_t=\beta\cdot\sum_{t=1}^{T} x_t^2 \\
\stackrel{(\star)}{\Leftrightarrow} \quad & \sum_{t=1}^{T} y_t x_t-(\bar{y}-\beta\bar{x})\cdot\sum_{t=1}^{T} x_t=\beta\cdot\sum_{t=1}^{T} x_t^2 \\
\vdots \quad & \\
\Leftrightarrow \quad & \frac{\left(\sum_{t=1}^{T} y_t x_t\right)-T\bar{y}\bar{x}}{\left(\sum_{t=1}^{T} x_t^2\right)-T(\bar{x})^2}=\beta \\
\Leftrightarrow \quad & \frac{\sum_{t=1}^{T}(x_t-\bar{x})(y_t-\bar{y})}{\sum_{t=1}^{T}(x_t-\bar{x})^2}=\beta
\end{aligned}
$$

Exercise 8

A supermarket chain plans to investigate the connection between marketing expenditure for a certain product and the number of items sold. $T = 10$ different but comparable markets are included in the study. For market t, marketing expenditure (in USD) and number of items sold are denoted as x_t and y_t respectively.
From this study, the following values are calculated:

$$\bar{x} = 1170 \qquad \sqrt{\frac{1}{10}\sum_{t=1}^{10}(x_t - \bar{x})^2} = 415$$

$$\sum_{t=1}^{10} y_t = 3500 \qquad \sum_{t=1}^{10} y_t^2 = 1'300'000 \qquad \sum_{t=1}^{10} x_t y_t = 4'443'000$$

a) Calculate the OLS-estimates $\hat{\alpha}$ and $\hat{\beta}$ of the regression model $y_t = \alpha + \beta x_t + u_t$.

$$\begin{aligned}
\hat{\beta} &= \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{T}(x_t - \bar{x})^2} = \frac{\left(\sum_{t=1}^{T} y_t x_t\right) - T\bar{x}\bar{y}}{\sum_{t=1}^{T}(x_t - \bar{x})^2} \\
&= \frac{\left(\frac{1}{T}\cdot\sum_{t=1}^{T} y_t x_t\right) - \bar{x}\bar{y}}{\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2} = \frac{\frac{1}{10}\cdot 4443000 - 1170\cdot 350}{415^2} \approx 0.202 \\
\hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x} \approx 350 - 0.202\cdot 1170 = 113.66
\end{aligned}$$

b) How is β interpreted from an economic point of view?

The assumed model and the estimate $\hat{\beta} = 0.202$ suggest that every dollar spent on marketing increases the number of articles sold by 0.202.

c) The manager of a market not included in the study wants to sell 500 products, what would you suggest him to spend on marketing according to the estimated regression?

We calculate the marketing expenditure x using the formula $\hat{y} = \hat{\alpha} + \hat{\beta}\cdot x$ with a desired $\hat{y} = 500$ and obtain $x = \frac{500-\hat{\alpha}}{\hat{\beta}} \approx 1912.35$

d) Calculate the coefficient of determination and interpret it.

The coefficient of determination can be expressed as squared correlation between x and y.

$$\begin{aligned}
\mathrm{Cor}(x,y) &= \frac{\frac{1}{10}\sum_{t=1}^{10}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\frac{1}{10}\sum_{t=1}^{10}(x_t - \bar{x})^2}\cdot\sqrt{\frac{1}{10}\sum_{t=1}^{10}(y_t - \bar{y})^2}} \\
&\dots \\
&= 0.96828 \\
R^2 &= 0.96828^2 \approx 0.9375
\end{aligned}$$

The coefficient of determination is close to 1, indicating that a linear dependence between marketing expenditures and number of items sold is plausible.

Exercise 9

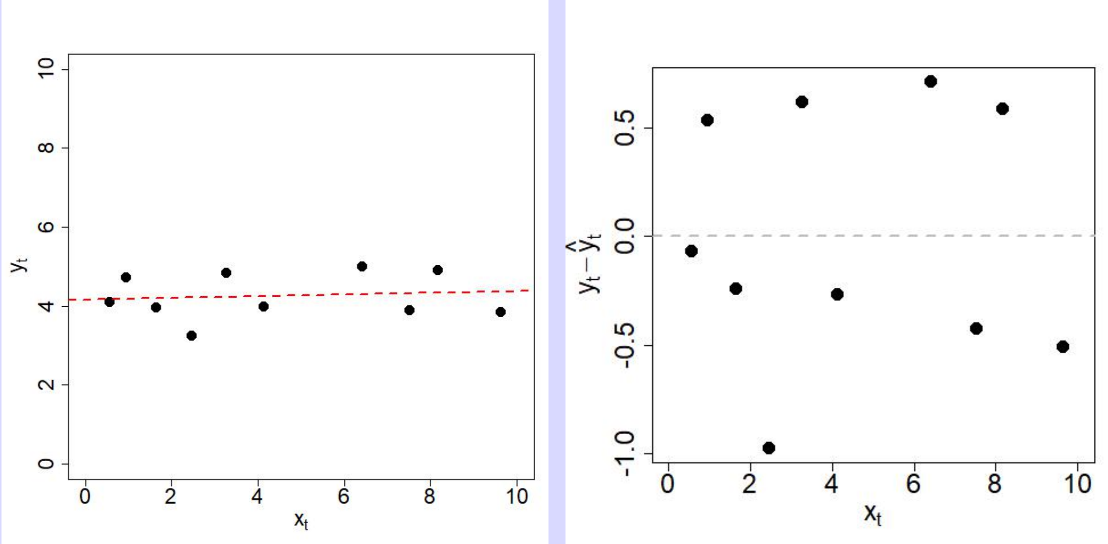
a) State the A-, B- and C-assumptions in your own words.

- **A1**: There is no relevant variable left out in the regression and every variable included is relevant.
- **A2**: There is a linear dependency between x_t and y_t.
- **A3**: The parameters are constant for each pair of observations.
- **B1**: $\mathrm{E}(u_t) = 0$. The deviations from observation y_t and fitted value $\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t$ spreads around zero.
- **B2**: The deviation $u_t = y_t - \hat{y}_t$ spreads evenly.
- **B3**: There is no dependency between the deviations $u_1, ..., u_T$: Knowledge about u_t does not give any information about u_s $(t \neq s)$.
- **B4**: The distribution of deviations u_t is normal.
- **C1**: The exogenous variables x_t can be controlled, they are not random.
- **C2**: The values of x_t are not the same for every pair of observations.

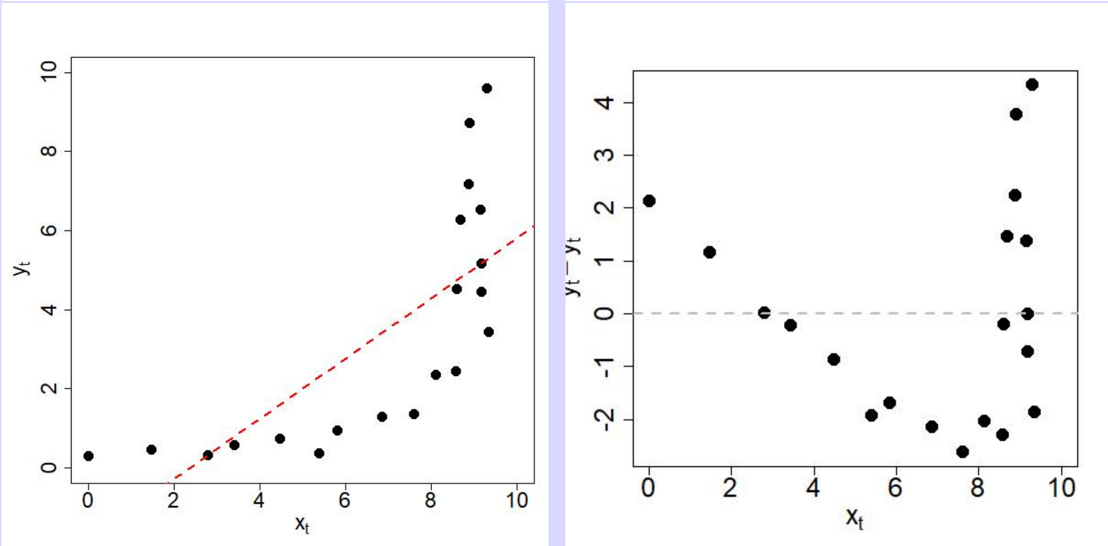
b) Give a counterexample of each assumption by drawing a picture if points (x_t, y_t), if possible. Otherwise describe the counterexample in your own words.

Below plots display counterexamples to A-, B- and C-assumptions. The observations (points) are drawn together with the regression line (red, dashed) on the left hand side. Residuals are drawn on the right hand side. C1 is the only assumption for which a counterexample could not be drawn. Examples for exogenous variables, which can not be controlled are easily found.

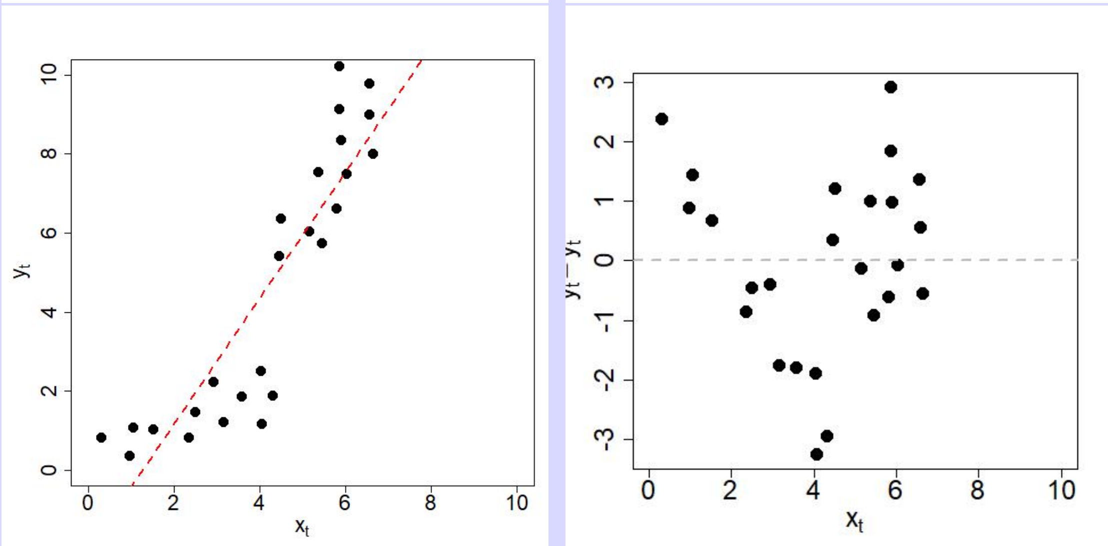
A1



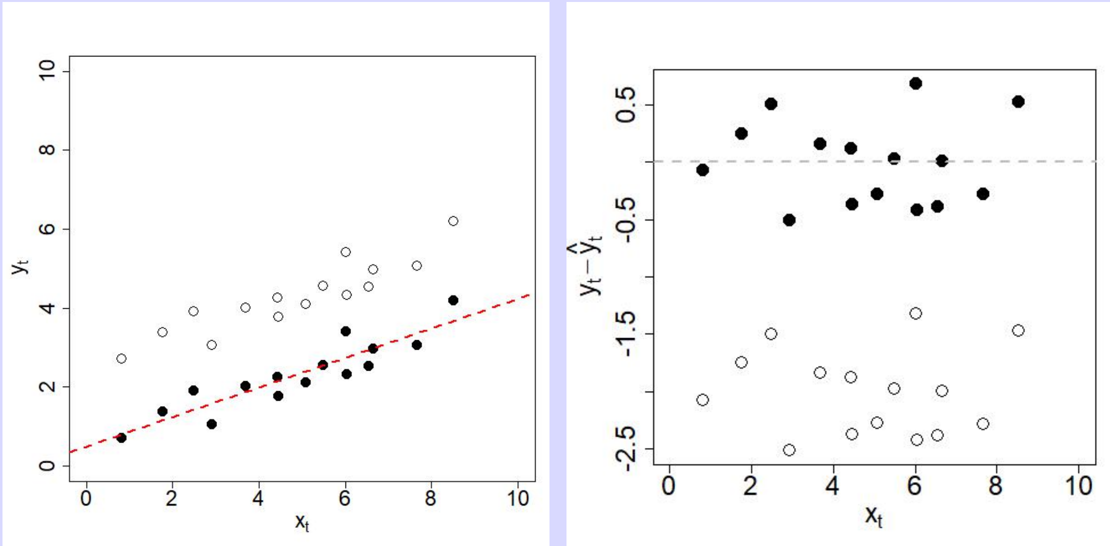
A2



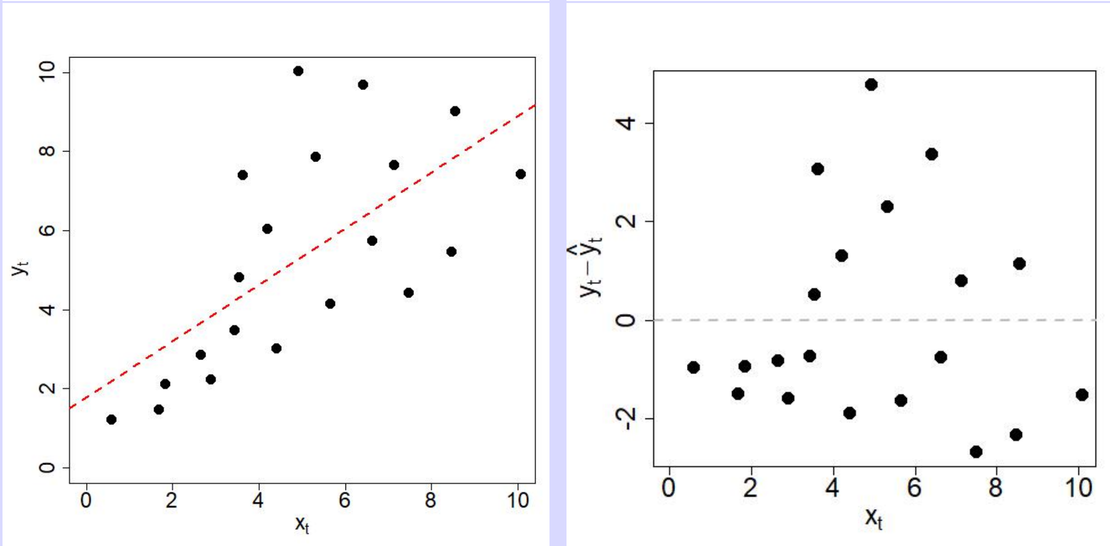
A3



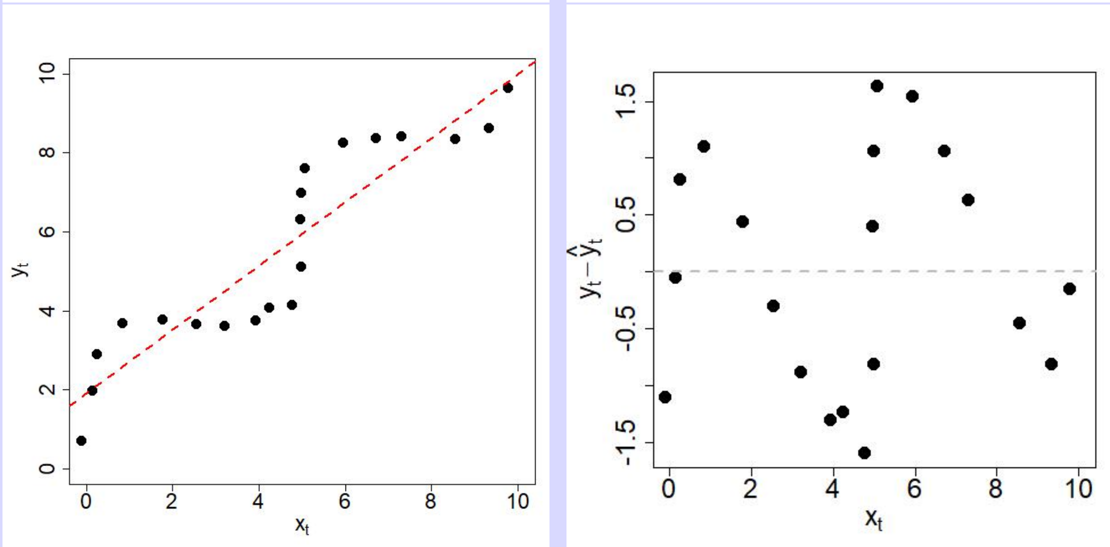
B1



B2

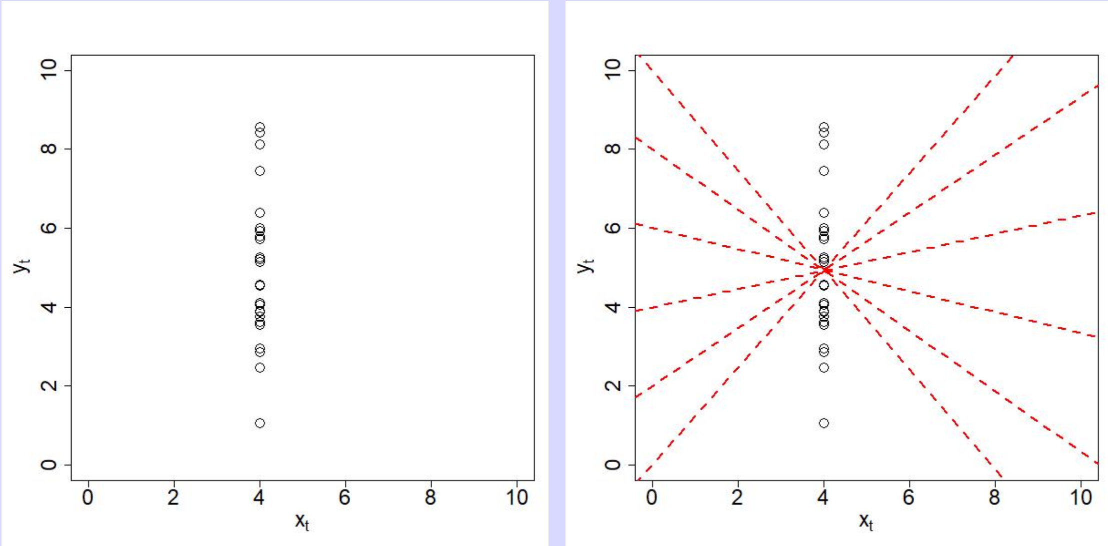


B3



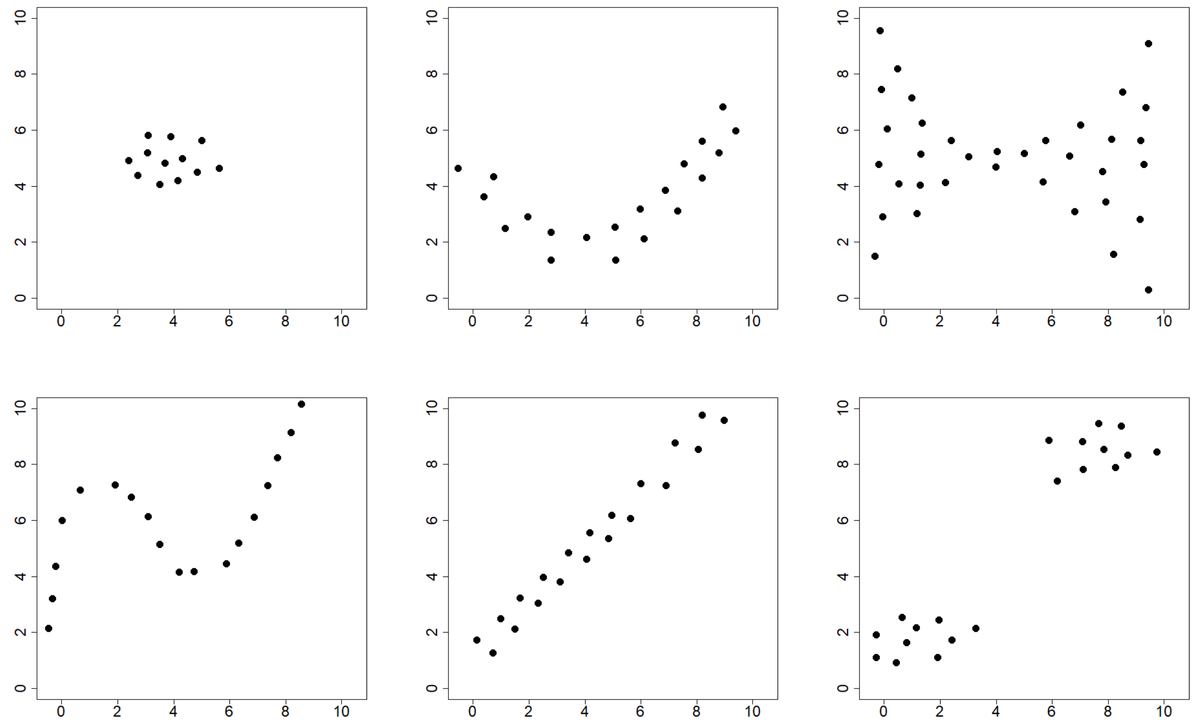
The assumption C2 is needed to ensure for example the uniqueness of the OLS estimate.

C2



Exercise 10

Take a look at the 6 scatterplots below.



a) Which of the A-, B- and C-assumptions are not satisfied in the individual cases?

A1 - x not relevant	A2 - Not linear A3 - Structural break B3 - Maybe autocorrelation	B2 - Heteroscedasticity
A2 - Not linear B3 - Positive autocorrelation	B3 - Neg. autocorrelation OR A3 - Structural break	A3 - Structural break

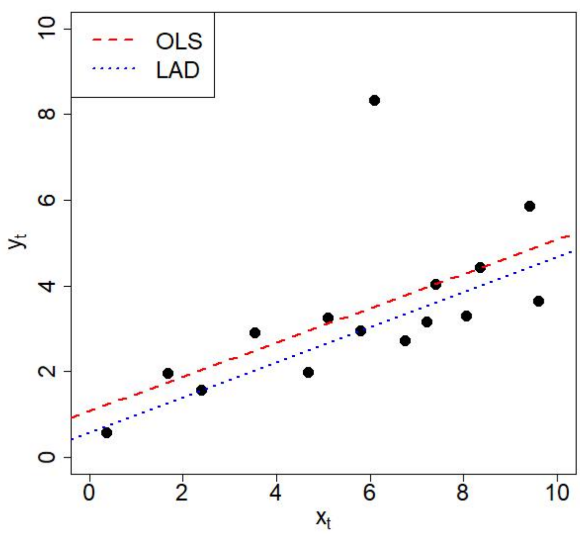
b) Which of the assumptions could be violated without being visible in a scatterplot?

- Stochastic exogenous variables
- Measurement errors and thus $\mathrm{E}(u_t) = 0$
- Relevant variables, that are left out? Workaround: Plot the residuals of models with and without the variable. If including the variable lowers the residuals significantly, it is probably relevant.

Exercise 11

The figure below shows the fitted values of two different regressions $y_t = \alpha + \beta x_t + u_t$. Besides the OLS regression line, the Least Absolute Deviation (LAD) regression line is shown. The LAD-estimators $\tilde{\alpha}$ and $\tilde{\beta}$ are calculated by minimizing

$$\sum_{t=1}^{15} |y_t - (\alpha + \beta x_t)| \; .$$



$$\sum_{t=1}^{15} y_t = 50.67 \qquad \sum_{t=1}^{15} y_t^2 = 219.53$$

$$\sum_{t=1}^{15} \hat{y}_t^2 = 188.77 \qquad \sum_{t=1}^{15} \tilde{y}_t = 44.16$$

$$\sum_{t=1}^{15} \tilde{y}_t^2 = 148.48$$

a) Calculate the coefficient of determination R^2 for both regressions using the above sums.

$$
\begin{aligned}
R^2_{\text{OLS}} &= \frac{\sum_{t=1}^T (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^T (y_t - \bar{y})^2} \\
&= \frac{\left(\sum_{t=1}^T \hat{y}_t^2\right) - T\bar{y}^2}{\left(\sum_{t=1}^T y_t^2\right) - T\bar{y}^2} \\
&= \frac{188.77 - 15 \cdot \left(\frac{50.67}{15}\right)^2}{219.53 - 15 \cdot \left(\frac{50.67}{15}\right)^2} \approx 0.3640 \\
R^2_{\text{LAD}} &= \frac{\sum_{t=1}^T (\tilde{y}_t - \bar{y})^2}{\sum_{t=1}^T (y_t - \bar{y})^2} \\
&= \frac{\left(\sum_{t=1}^T \tilde{y}_t^2\right) - 2\bar{y}\left(\sum_{t=1}^T \tilde{y}_t\right) + N\bar{y}^2}{\left(\sum_{t=1}^T y_t^2\right) - T\bar{y}^2} \\
&= \frac{148.48 - 2 \cdot \frac{50.67}{15} \cdot 44.16 + 15 \cdot \left(\frac{50.67}{15}\right)^2}{219.53 - 15 \cdot \left(\frac{50.67}{15}\right)^2} \approx 0.4403
\end{aligned}
$$

b) Calculate the correlation between y_t and x_t.

$$\text{Cor}(x, y) = \sqrt{R^2_{\text{OLS}}} \cdot \text{sign}(\hat{\beta}) = \sqrt{0.3640} = 0.6033$$

c) Interpret the above results.

The R^2 from LAD estimation is higher than the R^2 from OLS. The LAD-regression line suits the data better than the OLS-regression line. The reason for it is the outlier, which has a higher impact on the OLS estimation. LAD estimation is influences less and is called more robust against outliers.

Exercise 12

a) Consider the regression model

$$y_t = \alpha + \beta x_t + u_t \ .$$

Is it possible that the coefficient of determination is either negative or larger than 1? Justify your answer briefly.

The R^2 in this case is the squared correlation. Since the correlation only takes values between -1 and 1, it holds that $0 \leq R^2 \leq 1$.

b) Consider the regression model

$$y_t = \alpha x_t + u_t \ .$$

Is it possible that the coefficient of determination is negative or larger than 1? Justify your answer.

R^2 is never larger than 1, because $R^2 > 1$ would imply

$$\frac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{T}(y_t - \bar{y}_t)^2} < 0 \ .$$

Both sums are positive and therefore it holds that $R^2 \leq 1$.
It can be negative in this case due to the fast that the regression line is forced to go through the origin. A proof is left out.

Exercise 13

Consider the following regression:

$$y_t = \alpha x_t + u_t, \qquad \text{where } u_t \sim \mathcal{N}(0, \sigma^2)$$

for $t = 1, ..., T$. Assume validity if all A-, B- and C- assumptions.

a) Verify that the OLS estimator $\hat{\alpha}$ satisfies the following relationships

$$\hat{\alpha} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2}$$

and can be represented as

$$\hat{\alpha} = \alpha + \frac{\sum_{t=1}^{T} x_t u_t}{\sum_{t=1}^{T} x_t^2} \ .$$

$$\hat{\alpha} = \underset{\alpha \in \mathbb{R}}{\operatorname{argmin}} \sum_{t=1}^{T} (y_t - \alpha x_t)^2$$

$$\frac{\partial}{\partial \alpha} \sum_{t=1}^{T} (y_t - \alpha x_t)^2 = \sum_{t=1}^{T} 2 \cdot (y_t - \alpha x_t) \cdot (-x_t) \overset{!}{=} 0$$

$$\left(\sum_{t=1}^{T} x_t y_t\right) - \left(\sum_{t=1}^{T} \alpha x_t^2\right) = 0$$

$$\left(\sum_{t=1}^{T} x_t y_t\right) - \alpha \cdot \left(\sum_{t=1}^{T} x_t^2\right) = 0$$

$$\begin{aligned}
\alpha &= \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2} = \frac{\sum_{t=1}^{T} x_t(\alpha x_t + u_t)}{\sum_{t=1}^{T} x_t^2} \\
&= \frac{\left(\sum_{t=1}^{T} \alpha x_t^2\right) + \left(\sum_{t=1}^{T} x_t u_t\right)}{\sum_{t=1}^{T} x_t^2} \\
&= \frac{\alpha \sum_{t=1}^{T} x_t^2}{\sum_{t=1}^{T} x_t^2} + \frac{\sum_{t=1}^{T} x_t u_t}{\sum_{t=1}^{T} x_t^2} = \alpha + \frac{\sum_{t=1}^{T} x_t u_t}{\sum_{t=1}^{T} x_t^2}
\end{aligned}$$

b) Calculate the ML-estimator of α.

First we realize that y_t given x_t is normally distributed with expectation αx_t and variance σ. Therefore the likelihood is given by

$$\mathcal{L} = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left\{-\frac{1}{2}\frac{(y_t - \alpha x_t)^2}{\sigma^2}\right\}$$

and the log-likelihood

$$\ell = -\sum_{t=1}^{T} \left[\log\left(\sqrt{2\pi}\sigma\right) + \frac{1}{2}\frac{(y_t - \alpha x_t)^2}{\sigma^2}\right] .$$

To find $\hat{\alpha}$, we need the derivation of ℓ

$$\frac{\partial \ell}{\partial \alpha} = -\sum_{t=1}^{T} \frac{x_t y_t - \alpha x_t^2}{\sigma^2} .$$

$\hat{\alpha}$ is then the root of the above derivation, which is simply

$$\hat{\alpha} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2} .$$

c) Verify that the OLS estimator $\hat{\alpha}$ is unbiased for α.

To show that $\hat{\alpha}$ is unbiased, we use that x_t is not stochastic and the general rules for expectations.

$$\begin{aligned}
\mathrm{E}(\hat{\alpha}) &= \mathrm{E}\left(\alpha + \frac{\sum_{t=1}^{T} x_t u_t}{\sum_{t=1}^{T} x_t^2}\right) = \alpha + \mathrm{E}\left(\frac{\sum_{t=1}^{T} x_t u_t}{\sum_{t=1}^{T} x_t^2}\right) \\
&= \alpha + \frac{\mathrm{E}\left(\sum_{t=1}^{T} x_t u_t\right)}{\sum_{t=1}^{T} x_t^2} = \alpha + \frac{\sum_{t=1}^{T} \mathrm{E}\left(x_t u_t\right)}{\sum_{t=1}^{T} x_t^2} \\
&= \alpha + \frac{\sum_{t=1}^{T} x_t \, \mathrm{E}\left(u_t\right)}{\sum_{t=1}^{T} x_t^2} = \alpha + \frac{\sum_{t=1}^{T} x_t \cdot 0}{\sum_{t=1}^{T} x_t^2} = \alpha
\end{aligned}$$

d) Find the distribution of $\hat{\alpha}$.

From the alternative representation of $\hat{\alpha}$ in a), we see that $\hat{\alpha}$ is a linear transformation of independent normally distributed random variables $(u_1, ..., u_T)$. Therefore we know that $\hat{\alpha}$ is also normally distributed. The expectation has already been derived in c), the variance is found using that u_t and u_s are independent for $t \neq s$:

$$\begin{aligned}
\operatorname{Var}(\hat{\alpha}) &= \operatorname{Var}\left(\alpha + \frac{\sum_{t=1}^T x_t u_t}{\sum_{t=1}^T x_t^2}\right) = \operatorname{Var}\left(\frac{\sum_{t=1}^T x_t u_t}{\sum_{t=1}^T x_t^2}\right) \\
&= \frac{\sum_{t=1}^T \operatorname{Var}(x_t u_t)}{\left(\sum_{t=1}^T x_t^2\right)^2} = \frac{\sum_{t=1}^T x_t^2 \operatorname{Var}(u_t)}{\left(\sum_{t=1}^T x_t^2\right)^2} \\
&= \frac{\sum_{t=1}^T x_t^2 \sigma^2}{\left(\sum_{t=1}^T x_t^2\right)^2} = \frac{\sigma^2 \sum_{t=1}^T x_t^2}{\left(\sum_{t=1}^T x_t^2\right)^2} = \frac{\sigma^2}{\sum_{t=1}^T x_t^2}
\end{aligned}$$

So, finally

$$\hat{\alpha} \sim \mathcal{N}\left(\alpha, \frac{\sigma^2}{\sum_{t=1}^T x_t^2}\right)$$

Exercise 14

Consider the simple regression model with binary regressor, i.e.

$$y_t = \alpha + \beta x_t + u_t, \qquad x_t \in \{0, 1\}, \quad y_t \in \mathbb{R}, \quad t = 1, ..., T.$$

Let $\bar{y}_A$ and $\bar{y}_B$ denote the mean of y_t for those observations with $x_t = 1$ and $x_t = 0$ respectively. Further, let T_A denote the number of observations with $x_t = 1$ and correspondingly $T_B = T - T_A$.

a) Show that the OLS estimators can be expressed as

$$\hat{\alpha} = \bar{y}_B \qquad \hat{\beta} = \bar{y}_A - \bar{y}_B \ .$$

To derive the statement, we use three equations:

i) From exercise 7 we know that $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \Leftrightarrow \bar{y} = \hat{\alpha} + \hat{\beta}\bar{x}$.

ii) The global mean $\bar{y}$ can be written as $\bar{y} = \frac{T_A}{T}\bar{y}_A + \frac{T_B}{T}\bar{y}_B = \bar{x}\bar{y}_A + (1-\bar{x})\bar{y}_B$.

iii) Further we know that the OLS estimators solve $\frac{\partial}{\partial\beta}\sum_{t=1}^{T}(y_t - \alpha - \beta x_t)^2 = 0$, so that

$$
\begin{aligned}
& \frac{\partial}{\partial\beta}\sum_{t=1}^{T}(y_t - \alpha - \beta x_t)^2 = 0 \\
\Leftrightarrow \quad & \sum_{t=1}^{T}(y_t - \alpha - \beta x_t)\cdot x_t = 0 \\
\Leftrightarrow \quad & \sum_{t=1}^{T} y_t x_t - \alpha\sum_{t=1}^{T} x_t - \beta\sum_{t=1}^{T} x_t^2 = 0 \\
\Leftrightarrow \quad & \sum_{t=1}^{T} y_t x_t - T\alpha\bar{x} - \beta\sum_{t=1}^{T} x_t = 0 \\
\Leftrightarrow \quad & \bar{y}_A\sum_{t=1}^{T} x_t = T\alpha\bar{x} + \beta\sum_{t=1}^{T} x_t \\
\Leftrightarrow \quad & \bar{y}_A T\bar{x} = \alpha T\bar{x} + \beta T\bar{x} \\
\Rightarrow \quad & \bar{y}_A = \hat{\alpha} + \hat{\beta}
\end{aligned}
$$

Equating i) and iii) and using ii) in the first step, we see that

$$
\begin{aligned}
& \bar{x}\bar{y}_A + (1-\bar{x})\bar{y}_B = \hat{\alpha} + \hat{\beta}\bar{x} \\
\Leftrightarrow \quad & \bar{x}\bar{y}_A + \bar{y}_B - \bar{x}\bar{y}_B = \hat{\alpha} + (\bar{y}_A - \hat{\alpha})\bar{x} \\
\Leftrightarrow \quad & \bar{y}_B - \bar{x}\bar{y}_B = \hat{\alpha} - \hat{\alpha}\bar{x} \\
\Leftrightarrow \quad & \bar{y}_B(1-\bar{x}) = \hat{\alpha}(1-\bar{x}) \\
\Leftrightarrow \quad & \bar{y}_B = \hat{\alpha}
\end{aligned}
$$

Then using iii), we have

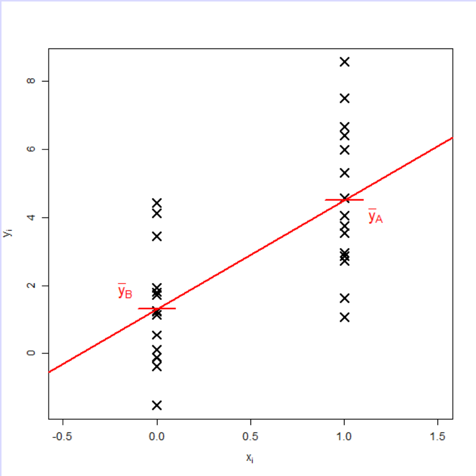
$$\bar{y}_A = \hat{\alpha} + \hat{\beta} \quad \Leftrightarrow \quad \bar{y}_A = \bar{y}_B + \hat{\beta} \quad \Leftrightarrow \quad \hat{\beta} = \bar{y}_A - \bar{y}_B$$

To clarify why these findings are plausible, we take a look at a sample $z_1, ..., z_T$. If we're looking for a value μ that minimizes the sum of squared distances to the observations, we see by derivation

$$
\begin{aligned}
\frac{\partial}{\partial \mu} \sum_{t=1}^{T} (z_t - \mu)^2 = \sum_{t=1}^{T} 2 \cdot (z_t - \mu) \cdot (-1) &\overset{!}{=} 0 \\
\Leftrightarrow \qquad \sum_{t=1}^{T} z_t - \mu &= 0 \\
\Leftrightarrow \qquad \mu &= \frac{1}{T} \sum_{t=1}^{T} z_t
\end{aligned}
$$

Performing this in the two groups separately shows why the regression line has to go through the two group means, which defined the regression line.

b) Sketch the above findings in a plot.



c) What would the OLS estimators be if it wasn't $x_t \in \{0, 1\}$, but $x_t \in \{-1, 1\}$?

Using a sketch, we see that:

$$
\hat{\beta} = \frac{\bar{y}_A - \bar{y}_B}{2} \qquad \text{and} \qquad \hat{\alpha} = \frac{\bar{y}_A + \bar{y}_B}{2}
$$

For $\hat{\beta}$ we simply calculate the slope of the regression line. For $\hat{\alpha}$ we calculate the mean of the two means $\bar{y}_A$ and $\bar{y}_B$.

d) Show that $S_{xx} = \frac{T_A T_B}{T}$ and $(T-2)\hat{\sigma}^2 = S_{yy}^A + S_{yy}^B$, where

$$
S_{yy}^A = \sum_A (y_t - \bar{y}_A)^2 \qquad \text{and} \qquad S_{yy}^B = \sum_A (y_t - \bar{y}_B)^2 .
$$

For the first part, we use $\bar{x} = T_A/T$ and $x_t \in \{0,1\}$ so that $x_t = x_t^2$:

$$\begin{aligned} S_{xx} &= \sum_{t=1}^{n} (x_t - \bar{x})^2 = \sum_{t=1}^{T} x_t^2 - T\bar{x}^2 \\ &= \sum_{t=1}^{T} x_t - T\bar{x}^2 = T\bar{x}(1-\bar{x}) = \frac{T_A T_B}{T} \end{aligned}$$

The second part is shown by splitting the sum. $\sum_A$ shall denote the sum over all indices t where $x_t = 1$.

$$(T-2)\hat{\sigma}^2 = \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 = \sum_{A} (y_t - \hat{y}_t)^2 + \sum_{B} (y_t - \hat{y}_t)^2$$

In the sums, the term $\hat{y}_t$ is constant, i.e. $\hat{y}_t = \hat{\alpha} + \hat{\beta} = \bar{y}_A$ for group A and $\hat{y}_t = \hat{\alpha} = \bar{y}_B$ for group B.

$$\Rightarrow \quad (T-2)\hat{\sigma}^2 = \sum_{A} (y_t - \bar{y}_A)^2 + \sum_{B} (y_t - \bar{y}_B)^2 = S_{yy}^A + S_{yy}^B$$

Exercise 15

A supermarket chain is interested in the effect of bonus scheme on the profit. The 35 stores in the suburbs offer the bonus scheme, while the 25 stores in the city do not use it. The following table summarizes the results of a study:

	No bonus scheme	Bonus scheme
Number of Stores	25	35
Mean Return	117.58	122.46
Sum of squared Returns	347686	527799

a) What model would be suitable to investigate the effect of the bonus scheme on the return?

To investigate the effect of a bonus scheme the model

$$y_t = \alpha + \beta x_t + u_t$$

would be a good choice, where $x_t = 1$ if store t offers the bonus scheme, $x_t = 0$ if not and y_t the return of store t.

b) Estimate the parameters of the above suggested model and interpret them.

Using the above exercise and the same notation, we see that

$$\hat{\alpha} = \bar{y}_B = 117.58 \qquad \text{and} \qquad \hat{\beta} = \bar{y}_A - \bar{y}_B = 4.88$$

The parameter β expresses the effect of the bonus scheme on the return.

c) Calculate 95% confidence intervals for the above estimates. (table of quantiles attached)

The calculate the confidence intervals, we need the following values

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{347686.25 \cdot 117.58^2 + 527799 - 122.46^2}{58} \approx 85.93 \\
S_{xx} &= \frac{T_A T_B}{T} = {}^{175}\!/_{12} \\
\widehat{\mathrm{Var}}(\hat{\alpha}) &= \hat{\sigma}^2 \left(\frac{1}{T} + \frac{\bar{x}^2}{S_{xx}} \right) \approx 3.44 \\
\widehat{\mathrm{Var}}(\hat{\beta}) &= \frac{\hat{\sigma}^2}{S_{xx}} \approx 5.90 \\
t_{58;0.975} &\approx 2
\end{aligned}$$

$$\Rightarrow \qquad \begin{aligned}
\mathrm{CI}_\alpha &= 117.58 \pm \sqrt{3.44} \cdot 2 = [113.87; 121.29] \\
\mathrm{CI}_\beta &= 4.88 \pm \sqrt{5.9} \cdot 2 = [0.02; 9.74]
\end{aligned}$$

d) The CEO of the chain asks, if you suggest to introduce the bonus scheme to the stores in the city based on the data. What do you reply?

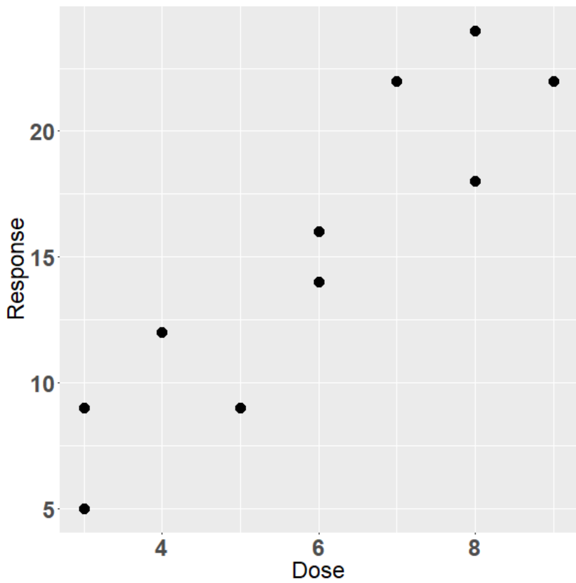
The confidence intervals suggest that there is a significant difference in the return of the stores. Under the assumptions, we would suggest to introduce the bonus scheme in city stores. Assumption A1 is probably not fulfilled, since there are variables which influence the return but are not included in the model.
Another shortcoming of this model is that the measured effect of the bonus scheme could also stem from the stores location. A model with only a dummy variable z_t (with $z_t = 1$ if store is in the suburbs) yields the same estimates.

Exercise 16

The following table displays the results of a study with allergy patients. We investigate the effect of medication (Dosage x_t) on the time of relief y_t by assuming the model $y_t = \alpha + \beta x_t + u_t$.

	x_t	y_t	$x_t y_t$	$\hat{y}_t$
	3	9	27	7.15
	3	5	15	7.15
	4	12	48	9.89
	5	9	45	12.63
	6	14	84	15.37
	6	16	96	15.37
	7	22	154	18.11
	8	18	144	20.86
	8	24	192	20.86
	9	22	198	23.60
Σ	59	151	1003	151
SSQ[a]	389	2651	–	2587.35

[a] SSQ: Sum of squared values - $\text{SSQ}(x) = \sum x_t^2$
Not the same as S_{xx}



a) Are the assumptions satisfied?

The only assumption that may not be fulfilled is the model choice, because a model without intercept possibly suits the reality better (no medication = no relief).

b) Estimate the parameters α and β by OLS.

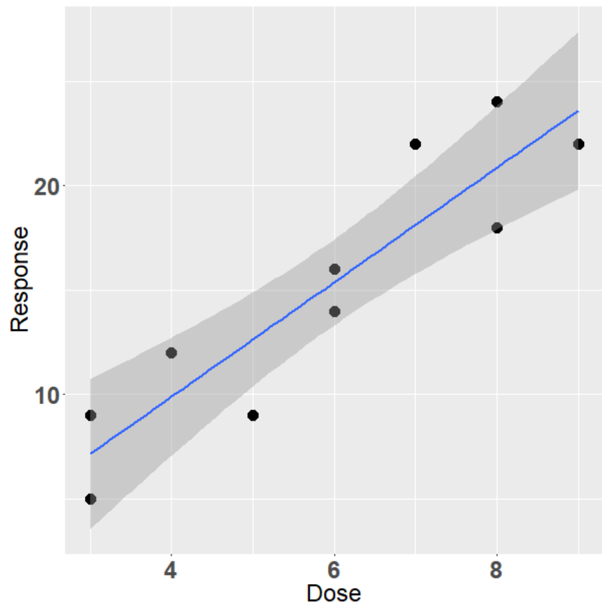
$$\hat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t - T\bar{x}\bar{y}}{\sum_{t=1}^{T} x_t^2 - T\bar{x}^2} = \frac{1003 - \frac{59 \cdot 151}{10}}{389 \frac{59^2}{10}} = 2.74$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = \frac{151}{10} - 2.74\frac{59}{10} = -1.066$$

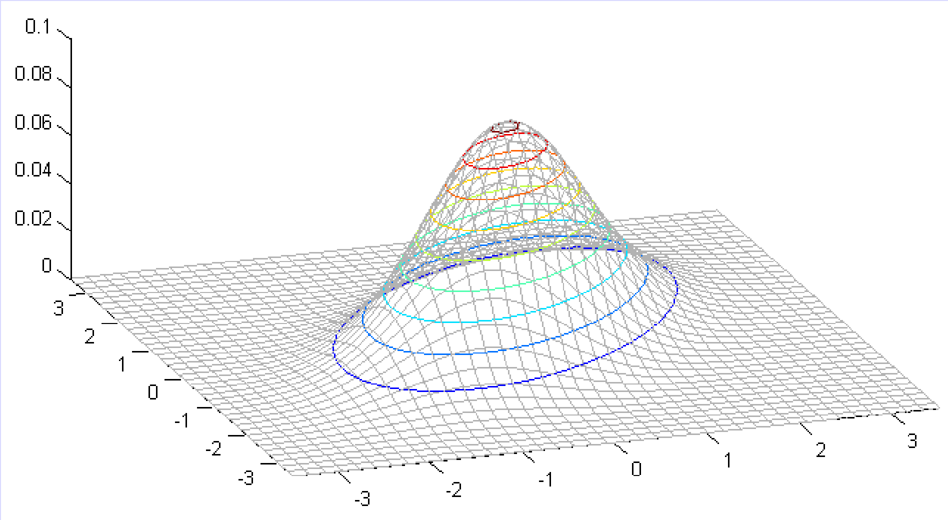
c) Calculate the 95% confidence intervals for $\hat{\alpha}$ and $\hat{\beta}$ and interpret.

$$\hat{\sigma}^2 = \frac{1}{T-2}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2 = \frac{1}{T-2}\left(\sum_{t=1}^{T} y_t^2 - 2\sum_{t=1}^{T} y_t\hat{y}_t + \sum_{t=1}^{T}\hat{y}_t^2\right)$$
$$= \frac{1}{T-2}\left(\sum_{t=1}^{T} y_t^2 - 2\sum_{t=1}^{T}\hat{y}_t^2 + \sum_{t=1}^{T}\hat{y}_t^2\right) = \frac{1}{T-2}\left(\sum_{t=1}^{T} y_t^2 - \sum_{t=1}^{T}\hat{y}_t^2\right)$$
$$= \frac{1}{8}(2651 - 2587.35) \approx 7.96$$
$$S_{xx} = \sum_{t=1}^{T} x_t^2 - T\bar{x}^2 = 389 - 59^2/10 = 40.9$$
$$\widehat{\text{Var}}(\hat{\alpha}) = \hat{\sigma}^2\left(\frac{1}{T} + \frac{\bar{x}^2}{S_{xx}}\right) = 7.96 \cdot \left(\frac{1}{10} + \frac{34.81}{40.9}\right) = 7.57$$
$$\widehat{\text{Var}}(\hat{\beta}) = \frac{\hat{\sigma}^2}{S_{xx}} = \frac{7.96}{40.9} \approx 0.19$$
$$CI_\alpha = [-7.42; 5.29]$$
$$CI_\beta = [1.73; 3.75]$$

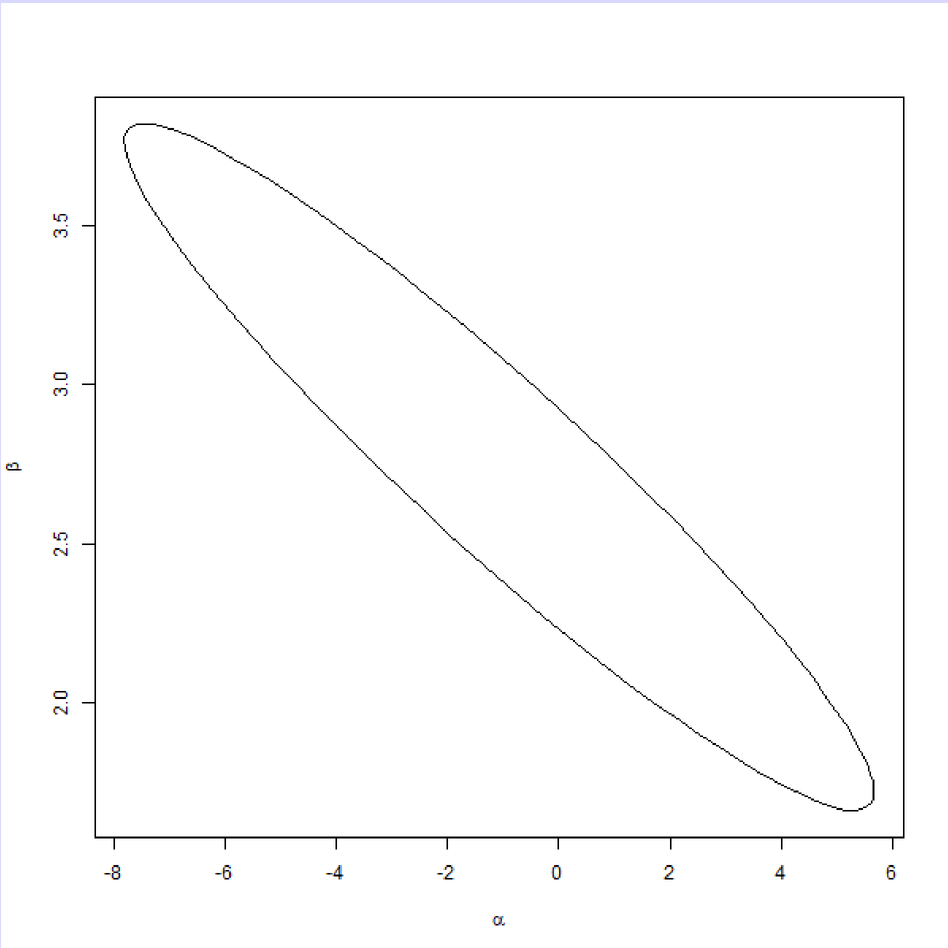
d) The following figure shows the 95% confidence interval around the regression line. Try to think of a reason why the this confidence band is wider at the edges.



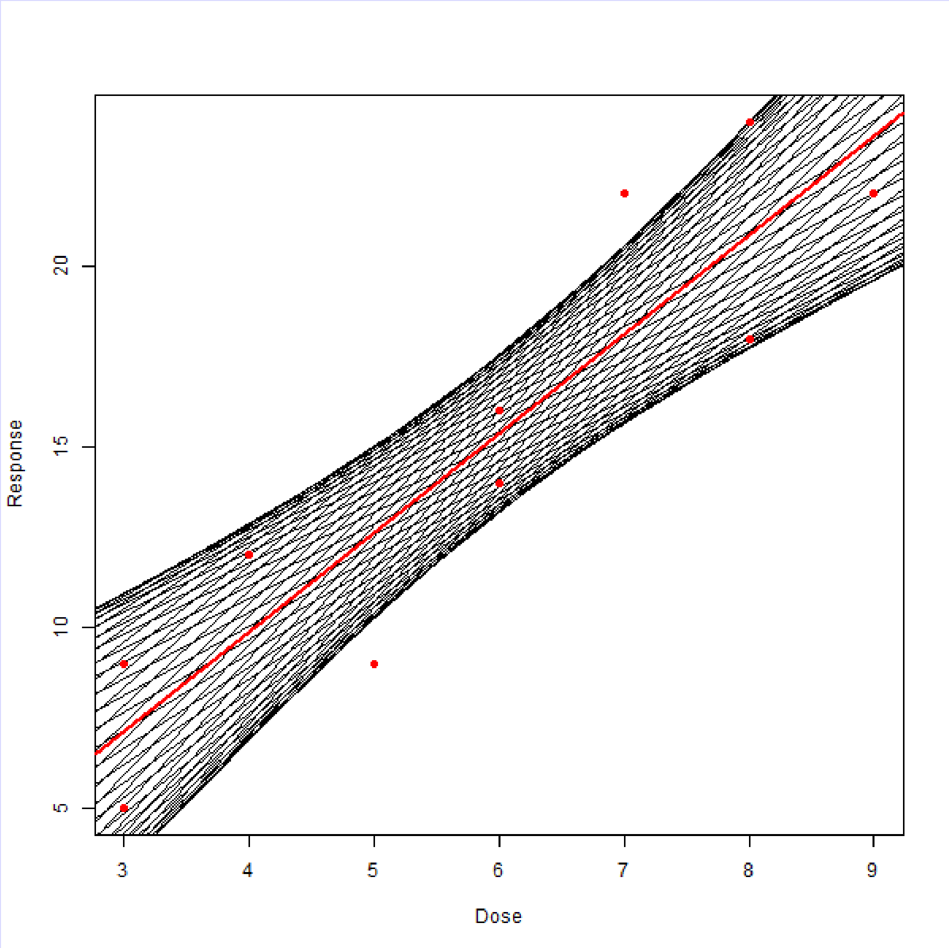
The intuition behind it is that there is less information at the edges and therefore a higher uncertainty. For clarification, we look at how the confidence interval around the regression line is done.
The parameter estimates $\hat{\alpha}$ and $\hat{\beta}$ are jointly normally distributed with expectation α and β and covariance matrix as presented in the lecture. The density looks similar to the one displayed below:



We define the confidence interval around the regression line as the region where the regression lines of the 95% "most plausible"parameters are. These parameters are displayed for our exercise below:



Just to illustrate, we take a few parameter combination from the border of this ellipse and plot the corresponding regression line:

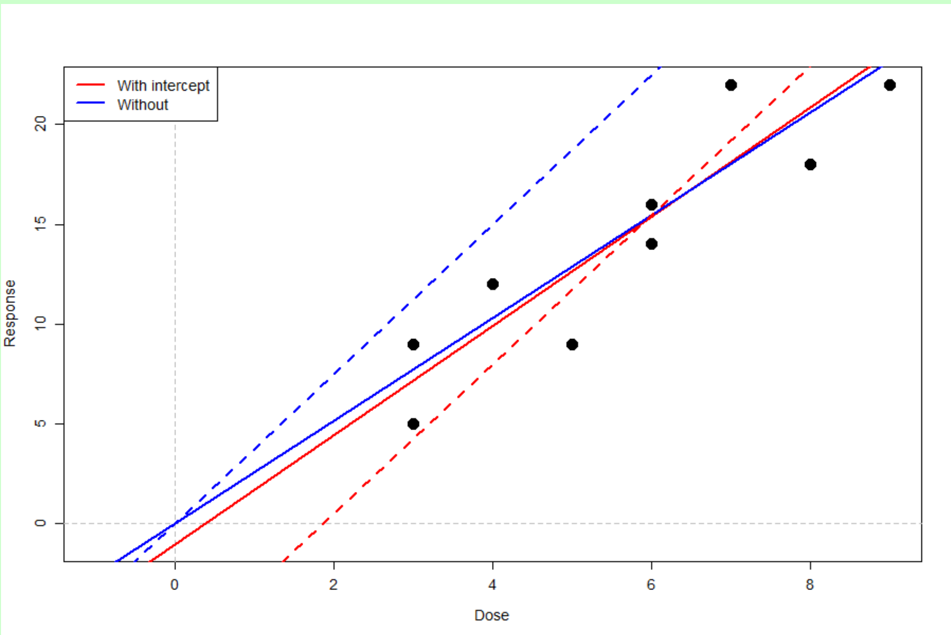


The interpretation is similar to the one of confidence intervals for parameters: The tube covers the true regression line with a probability of 95%.

In class we also discussed the model without intercept $y_t = \beta x_t + u_t$. In the above model, the confidence interval for α covers the zero, which indicates that it might not be relevant.
Exercise 13 gives a formula for the estimator and its variance to calculate the confidence for β with:

$$
\begin{aligned}
\hat{\beta} &= \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2} = \frac{1003}{389} (\approx 2.58) \\
\hat{\sigma}^2 &= \frac{1}{9} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 = \frac{1}{9} \sum_{t=1}^{T} (y_t - \hat{\beta} x_i)^2 \\
&= \frac{1}{9} \left[\sum_{t=1}^{T} y_t^2 - 2\hat{\beta} \sum_{t=1}^{T} x_t y_t + \hat{\beta}^2 \sum_{t=1}^{T} x_t^2 \right] = \frac{8410}{1167} (\approx 7.21) \\
\widehat{\mathrm{Var}}(\hat{\beta}) &= \frac{\hat{\sigma}^2}{\sum_{t=1}^{T} x_t^2} = \frac{8410}{1167 \cdot 389} (\approx 0.018526) \\
CI_\beta &= \hat{\beta} \pm \sqrt{\widehat{\mathrm{Var}}(\hat{\beta})} \cdot t_{9;0.975} = [\mathbf{2.27; 2.89}]
\end{aligned}
$$

In comparison to that: the confidence interval for β in the model with intercept was $[1.73; 3.75]$. One sees that the estimates are similar, but the confidence interval for β in the model with intercept is much wider than for the model without intercept. This difference is plausible by understanding what a model without intercept does: it forces the regression line to go through the origin. This way a different slope of the regression line can not be compensated by a changing intercept. In the figure below, the regression lines of models with and without intercept are displayed. Also added are 1) a regression line with intercept and value for $\hat{\beta} = 3.75$ (upper limit of CI) and corresponding intercept $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}\bar{y} - 3.75\bar{x}$ (red, dashed) 2) A regression line without intercept with the <u>same slope</u> $\hat{\beta} = 3.75$ (blue, dashed)
As you can see, the line from 1) suits the data not really well, but definitely better than line 2).



Exercise 17

Consider the regression model $y_t = \alpha + \beta x_t + u_t$. Show that $\hat{\sigma}^2 = \frac{1}{T-2}\sum_{t=1}^{T} \hat{u}_t^2$ is an unbiased estimator of the variance of the error terms by following the steps:

Step 1) Show that $y_t - \bar{y} = \beta(x_t - \bar{x}) + (u_t - \bar{u})$

Note: As we discussed in course, every sum in this exercise will have the index $t = 1$ to T.
From the model we see that $y_t = \alpha + \beta x_t + u_t$. Summing up and dividing by T gives

$$\bar{y} = \frac{1}{T}\sum y_t = \frac{1}{T}\sum(\alpha + \beta x_t + u_t) = \alpha + \beta\bar{x} + \bar{u}$$

So that

$$\begin{aligned} y_t - \bar{y} &= \alpha + \beta x_t + u_t - \alpha - \beta\bar{x} - \bar{u} \\ &= \beta(x_t - \bar{x}) + (u_t - \bar{u}) \end{aligned}$$

Step 2) Show that $\hat{\beta} = \beta + \frac{S_{xu}}{S_{xx}}$.

$$\begin{aligned} \hat{\beta} &= \frac{S_{xy}}{S_{xx}} && \text{(Def. of } \hat{\beta}\text{)} \\ &= \frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sum(x_t - \bar{x})^2} && \text{(Def. of } S_{xy}, S_{xx}\text{)} \\ &= \frac{\sum(x_t - \bar{x})(\beta(x_t - \bar{x}) + (u_t - \bar{u}))}{\sum(x_t - \bar{x})^2} && \text{(Step 1)} \\ &= \beta\frac{\sum(x_t - \bar{x})^2}{\sum(x_t - \bar{x})^2} + \frac{\sum(x_t - \bar{x})(u_t - \bar{u})}{\sum(x_t - \bar{x})^2} \\ &= \beta + \frac{S_{xu}}{S_{xx}} \end{aligned}$$

Step 3) Show that $\hat{u}_t = (y_t - \bar{y}) - \hat{\beta}(x_t - \bar{x}) = -(\hat{\beta} - \beta)(x_t - \bar{x}) + u_t - \bar{u}$

$$\begin{aligned} & \hat{u}_t = y_t - \hat{\alpha} - \hat{\beta}x_t \\ \Rightarrow\quad & \bar{\hat{u}} = \frac{1}{T}\sum \hat{u}_t = \bar{y} - \hat{\alpha} - \hat{\beta}\bar{x} && \text{(= 0 , because sum of residuals is 0)} \\ \Rightarrow\quad & \hat{u}_t = \hat{u}_t - \bar{\hat{u}} \\ & \quad = y_t - \hat{\alpha} - \hat{\beta}x_t - \bar{y} + \hat{\alpha} + \hat{\beta}\bar{x} \\ & \quad = (y_t - \bar{y}) - \hat{\beta}(x_t - \bar{x}_t) \\ & \quad = \beta(x_t - \bar{x}) + (u_t - \bar{u}) - \hat{\beta}(x_t - \bar{x}) && \text{(Step 1)} \\ & \quad = (\beta - \hat{\beta})(x_t - \bar{x}) + (u_t - \bar{u}) \\ & \quad = -(\hat{\beta} - \beta)(x_t - \bar{x}) + (u_t - \bar{u}) \end{aligned}$$

Step 4) Conclude that $\sum_{t=1}^{T} \hat{u}_t^2 = S_{uu} - (\hat{\beta} - \beta)^2 S_{xx}$

$$\begin{aligned}
\hat{u}_t^2 =& \left[-(\hat{\beta} - \beta)(x_t - \bar{x}) + (u_t - \bar{u})\right]^2 && \text{(Step 3)} \\
=& (\hat{\beta} - \beta)^2 (x_t - \bar{x})^2 - 2(\hat{\beta} - \beta)(x_t - \bar{x})(u_t - \bar{u}) + (u_t - \bar{u})^2 \\
\Rightarrow \sum \hat{u}_t^2 =& (\hat{\beta} - \beta)^2 \sum (x_t - \bar{x})^2 - 2(\hat{\beta} - \beta) \sum (x_t - \bar{x})(u_t - \bar{u}) \\
&+ \sum (u_t - \bar{u})^2 \\
=& (\hat{\beta} - \beta)^2 S_{xx} - 2(\hat{\beta} - \beta) S_{xu} + S_{uu} \\
=& S_{uu} - (\hat{\beta} - \beta)^2 S_{xx}
\end{aligned}$$

In the last equation, it is used that $\hat{\beta} = \beta + \frac{S_{xu}}{S_{xx}}$ (Step 2) so that $S_{xu} = (\hat{\beta} - \beta) S_{xx}$.

Step 5) Show that $\mathrm{E}(S_{uu}) = (T-1)\sigma^2$

For this step, we first see that

- $\mathrm{Var}(\bar{u}) = \mathrm{E}(\bar{u}^2) - \underbrace{\mathrm{E}(\bar{u})^2}_{=0} = \mathrm{E}(\bar{u}^2)$ (because of assumption $\mathrm{E}(u_t) = 0$)
- $\mathrm{Var}(u_t) = \mathrm{E}(u_t^2) - \underbrace{\mathrm{E}(u_t)^2}_{=0} = \mathrm{E}(u_t^2) = \sigma^2$
- $\mathrm{Var}(\sum u_t) = \sum \mathrm{Var}(u_t) = T\sigma^2$ (because u_t independent)
- $\mathrm{Var}(\bar{u}_t) = \mathrm{Var}\left(\frac{1}{T} \sum u_t\right) = \frac{1}{T^2} \mathrm{Var}\left(\sum u_t\right) = \frac{\sigma^2}{T}$

$$\begin{aligned}
\mathrm{E}(S_{uu}) =& \mathrm{E}\left(\sum (u_t - \bar{u})^2\right) = \mathrm{E}\left(\left(\sum u_t^2\right) - T\bar{u}^2\right) \\
=& \mathrm{E}\left(\sum u_t^2\right) - T\mathrm{E}(\bar{u}^2) \\
=& \sum \mathrm{E}(u_t^2) - T\frac{\sigma^2}{T} \\
=& T\sigma^2 - \sigma^2 = (T-1)\sigma^2
\end{aligned}$$

Step 6) Show that $\mathrm{E}\left((\hat{\beta} - \beta)^2\right) S_{xx} = \sigma^2$ which finishes the proof

We know that $\hat{\beta} \sim \mathcal{N}\left(\beta, \frac{\sigma^2}{S_{xx}}\right)$. Although β in not explicitly known, it is a constant and not stochastic. Therefore, the variance of $(\hat{\beta} - \beta)$ is the same as the variance of $\hat{\beta}$, i.e. $\frac{\sigma^2}{S_{xx}}$.
The expectation of $(\hat{\beta} - \beta)$ is 0, so that $\mathrm{Var}\left((\hat{\beta} - \beta)\right) = \mathrm{E}\left((\hat{\beta} - \beta)^2\right)$ and finally $\mathrm{E}\left((\hat{\beta} - \beta)^2\right) S_{xx} = \mathrm{Var}\left(\hat{\beta}\right) S_{xx} = \sigma^2$.
The expectation of $\sum \hat{u}_t^2$ is found by combining steps 4, 5 and 6, which completes the proof.

Exercise 18

This exercise deals with multicollinearity. As a reminder, consider a regression model $y_t = \beta_0 + \beta_1 x_{1,t} + ... + \beta_k x_{k,t} + u_t$. Multicollinearity exists if there is a vector $\gamma = (\gamma_0, ..., \gamma_k)' \neq (0, ..., 0)'$ so that $\gamma_0 + \gamma_1 x_{1,t} + ... + \gamma_k x_{k,t} = 0$ for every $t = 1, ..., T$.

a) Are the following statements true or false?

(i) In case of multicollinearity, at least one exogenous variable can be expressed as a linear combination of the remaining ones.

Correct - that is just a more literal definition of multicollinearity.

(ii) If y_t can be expressed as linear combination of the exogenous variables $x_{1,t}, ..., x_{k,t}$, the OLS estimator can not be calculated.

Not correct - the linear regression aims to express y_t as a linear combination of exogenous variables best possible.

(iii) To check if multicollinearity is in the data, it suffices to calculate the pairwise correlation coefficients of exogenous variables.

Not correct - One variable may be weakly correlated with the single other variables, but highly correlated with the combination of other variables.

(iv) In case two exogenous variables are perfectly correlated, i.e. $x_{1,t} = \gamma_0 + \gamma_1 x_{2,t}$, the total impact of those variables on y_t can be estimated.
If yes, can the impact be separated on the two variables?
If no, can you think of a way to solve the problem?

The impact of both variables together can be estimated in case of multicollinearity, but the influence can not be separated. If two variables were not perfectly correlated, but highly correlated, the separation would be estimated with a high uncertainty.

(v) In case of multicollinearity, the determinant of $X'X$ is 0.

Correct

b) Decide if there is multicollinearity in the cases below:

(i) Model: $y_t = \beta_0 + \beta_1 x_t + u_t$
$S_{xx} = \sum_{t=1}^{T} (x_i - \bar{x})^2 = 0$

If $S_{xx} = 0$, then $\mathrm{Var}(x) = 0$, so we see that there is no variation in x-values. That means $x_1 = x_2 = ... = x_T$ and a perfect correlation between x_t and the intercept. Another way to see the multicollinearity is by calculating $X'X$. Therefore we denote the value of all x-observations as a constant c. The data matrix X is then

$$\begin{pmatrix} 1 & c \\ 1 & c \\ \vdots & \vdots \\ 1 & c \end{pmatrix}$$

and

$$X'X = \begin{pmatrix} T & \sum x_t \\ \sum x_t & \sum x_t^2 \end{pmatrix} = \begin{pmatrix} T & Tc \\ Tc & Tc^2 \end{pmatrix}$$

. Calculating the determinant yields

$$\det(X'X) = T \cdot (Tc^2) - (Tc) \cdot (Tc) = T^2c^2 - T^2c^2 = 0$$

(ii) Model: $y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + u_t$

t	$x_{1,t}$	$x_{2,t}$	$x_{3,t}$
1	2	1	12
2	2	1	12
3	3	3	9
4	5	1	9
5	2	3	10
6	5	2	8
7	6	3	6
8	4	6	5
9	3	1	11
10	1	3	11
$\sum_{t=1}^{10} x_{\bullet t}$	33	24	93
$\sum_{t=1}^{10} x_{\bullet t}^2$	133	80	917

$$\sum_{t=1}^{10} x_{1,t}x_{2,t} = 82$$

$$\sum_{t=1}^{10} x_{1,t}x_{3,t} = 280$$

$$\sum_{t=1}^{10} x_{2,t}x_{3,t} = 198$$

Calculating the pairwise correlations would result in absolute values up to 0.75, which already indicates multicollinearity. With far less effort, we see that the sum of x-values for every t is 15. Therefore we can express $x_{1,t}$ as $x_{1,t} = 15 - x_{2,t} - x_{3,t}$ and see by above definition that there is multicollinearity in the x-values.

Exercise 19

Consider the restricted regression

$$y_t = \alpha + u_t \ , \ t = 1, ..., T \ ,$$

for which the A-, B- and C-assumption are satisfied.

a) Derive the ML estimator for α.

We so similar steps to the ML-estimator derived in the lecture. First we see that y_t is normally distributed with mean α and variance σ^2. The log-likelihood is then

$$\begin{aligned}\ln(\mathcal{L}) &= \sum_{t=1}^{T} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(y_t-\alpha)^2}{\sigma^2}\right\}\right) \\ &= -\frac{1}{2}\sum_{t=1}^{T}\left(\underbrace{\ln(2\pi)}_{\text{const.}} + \ln(\sigma^2) + \frac{(y_t-\alpha)^2}{\sigma^2}\right)\end{aligned}$$

To find the ML-estimate we derive $\ln(\mathcal{L})$ with respect to α and find the root.

$$\begin{aligned} \frac{\partial \ln(\mathcal{L})}{\partial \alpha} &\overset{!}{=} 0 \\ \Leftrightarrow \frac{1}{2}\sum_{t=1}^{T}\frac{2(y_t-\alpha)}{\sigma^2} &= 0 \\ \Leftrightarrow \quad \sum_{t=1}^{T}(y_t-\alpha) &= 0 \\ \Leftrightarrow \qquad\qquad \alpha &= \frac{1}{T}\sum_{t=1}^{T} y_t = \bar{y}\end{aligned}$$

So we know that if we set $\alpha = \bar{y}$, the derivation of $\ln(\mathcal{L})$ w.r.t α is zero. Hence we see that $\bar{y}$ is the ML estimator for α.

b) Calculate the ML estimator for the error variance given the sample

y_i	4	6	7	4
x_i	2	4	5	3

First we need to derive the ML estimator for σ^2 with similar steps as above:

$$
\begin{aligned}
\frac{\partial \ln(\mathcal{L})}{\partial \sigma^2} &\overset{!}{=} 0 \\
\Leftrightarrow -\frac{1}{2}\sum_{t=1}^{T}\frac{1}{\sigma^2} + (-1)\cdot\frac{(y_t-\alpha)^2}{\sigma^4} &= 0 \\
\Leftrightarrow \qquad \sum_{t=1}^{T}\left(\sigma^2-(y_t-\alpha)^2\right) &= 0 \\
\Leftrightarrow \qquad \sigma^2 &= \sum_{t=1}^{T}(y_t-\alpha)^2
\end{aligned}
$$

Here we see that the expression for σ^2 depends on α. The term above for $\hat{\alpha}_{ML}$ did not depend on σ^2. This shows that the ML-estimate for α does not depend on the error variance. For estimating σ^2 by ML, we need the ML-estimate of α to plug it in above formula. The estimated error variance is then simply the mean of squared residuals:

$$
\hat{\sigma}^2_{\text{ML}} = \frac{1}{T}\sum_{t=1}^{T}(y_t-\hat{\alpha}_{\text{ML}})^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2
$$

For the given data, $\hat{\alpha}_{\text{ML}}$ is $^{21}/_{4}$ and $\hat{\sigma}^2 = {}^{21}/_{16}$

Exercise 20

Reconsider exercise 8.

a) Test if β is significantly different from 0 at a 5% level and interpret the outcome.

b) Is the intercept α significantly greater than 1? (Choose a 5% level of significance)
Hint: $\text{Var}(\hat{\alpha}) = \sigma^2\left[\frac{1}{T}+\frac{\bar{x}^2}{\sum_{t=1}^{T}(x_t-\bar{x})^2}\right]$

Not discussed in class.

Exercise 21

It is known that the IQ is normally distributed with mean $\mu = 100$ and standard deviation $\sigma = 15$. Four students want to find out if students have a significantly higher IQ. Therefore they do an IQ test, in which they scored 120 on average.

a) Formulate the hypotheses for above question.

We denote the expected score of an IQ test a student does as μ_S. The students want to test if a students IQ is larger than 100. To prove it with a statistical test and a given uncertainty (significance level) we formulate it as the alternative hypothesis H_1. The null hypothesis is the opposite of H_1, so that we test:

$$
H_0 : \mu_s \leq 100 \text{ vs. } H_1 : \mu_S > 100
$$

b) We want to use the average score of 120 as a test statistics. Find the critical value for testing with a 5% level of significance.
Hint: Use the 95% quantile of the $\mathcal{N}(0,1)$ distribution 1.645.

The test statistics is used as an indicator for the question: How plausible is the null hypothesis? In out case a larger value of the average IQ score from the four students indicates that the null hypothesis might be wrong. We set the level of significance to 5%, so that we commit a type-I error with a probability of 5%. That means: Given that H_0 is true, it is rejected with a probability of 5%. That's why we need to find out the 5% extremest values that speak against H_0. In our case here we need a critical value so that - given that H_0 is true - the test statistics only takes larger values with a probability of 5%, which is called 95% quantile. Under H_0, the test statistics is normally distributed with mean 100 and standard deviation 15/4. The 95% Quantile of this distribution is found using the quantile of the $\mathcal{N}(0,1)$ distribution as: $100 + 1.645 \cdot \frac{15}{\sqrt{4}} = 112.3375$. For an average score of the four students lower than 112.3375, H_0 is not rejected, an average larger than that allows to reject H_0.

c) Below figure shows the densities of the average score if H_0 was true (black) and the estimated density from the sample (red). Complete the plot by adding

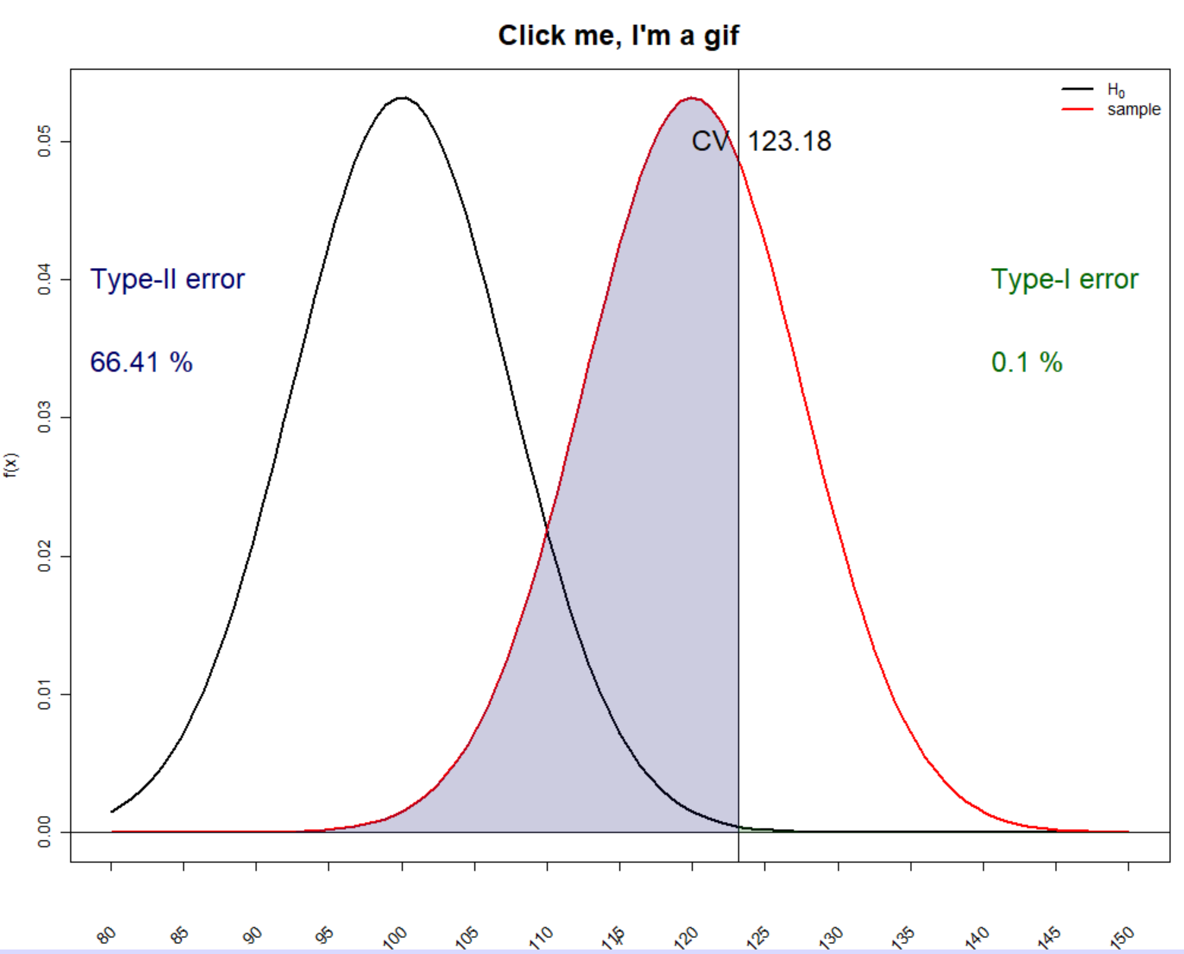
(i) The critical value

(ii) The type I error

(iii) The type II error if the IQ of students is really 120 on average

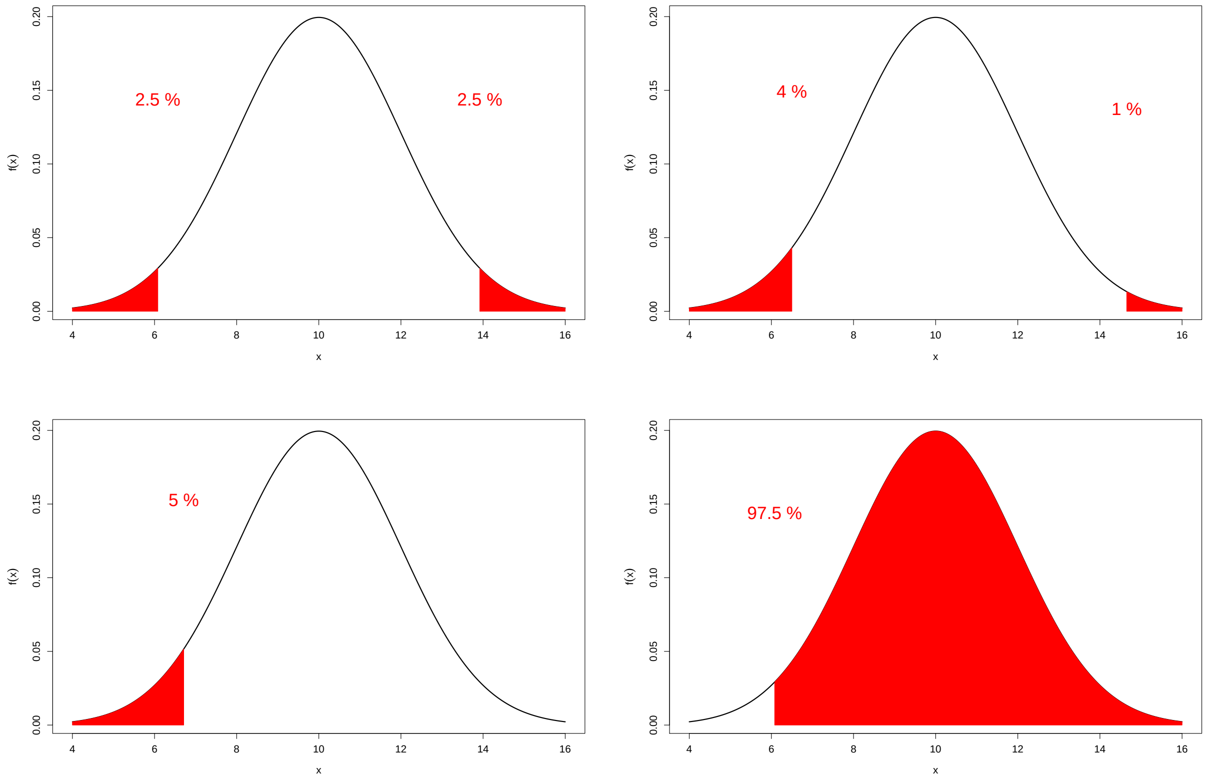
d) Sketch what happens with type I and type II error if the level of significance is increased. Use below figure again.

Below animation shows type-I and type-II errors and critical value for different levels of significance. One chooses the probability of a type-I error as level of significance. The plots shall demonstrate why we don't set this error probability as small as possible: Decreasing the type-I error probability increases the probability for a type-II error.



Exercise 22

The following plots show possible density functions of test statistics and critical values. If it belongs to a valid test, give the hypotheses and the level of significance.



The red areas show the regions where the null hypothesis is rejected. This is also the "direction" of the alternative hypothesis. The probability of a test statistic being in that region under the null hypothesis is the level of significance. For the 4 plots starting in the first row that means:

<u>Plot 1:</u> $H_0 : \mu = 10$ at a 5% level

<u>Plot 2:</u> $H_0 : \mu = 10$ at a 5% level as well. Critical areas need not be symmetric

<u>Plot 3:</u> $H_0 : \mu \geq 10$ at a 5% level

<u>Plot 4:</u> According to above way of proceeding: $H_0 : \mu = 10$ at a 97.5% level, but such a high level of significance is not useful.

Exercise 23

For the annual turnover of a large store group, the validity of the following linear regression model is assumed:

$$y_t = \alpha + \beta_1 x_{1,t} + \beta_2 x_{2,t} + u_t$$

with y_t: annual turnover (of store t), $x_{1,t}$: sales area (of store t), $x_{2,t}$: average frequency of people passing by (of store t), u_t: error term (of store t). Based on the observation values collected in $\boldsymbol{y}$ and $\boldsymbol{X}$ one obtains

$$\boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} 20 & 200 & 0 \\ 200 & 2014.4 & 16 \\ 0 & 16 & 40 \end{pmatrix}, \qquad (\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 12.55 & -1.25 & 0.5 \\ -1.25 & 0.125 & -0.05 \\ 0.5 & -0.05 & 0.045 \end{pmatrix},$$

$$\boldsymbol{X}'\boldsymbol{y} = \begin{pmatrix} 946 \\ 9532 \\ 85 \end{pmatrix}, \qquad \hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} = 90.28125 .$$

a) How many stores were involved in the investigation?

The regressor matrix X for the given model has structure

$$X'X = \begin{pmatrix} T & \sum_{t=1}^{T} x_{1,t} & \sum_{t=1}^{T} x_{2,t} \\ \sum_{t=1}^{T} x_{1,t} & \sum_{t=1}^{T} x_{1,t}^2 & \sum_{t=1}^{T} x_{1,t} \cdot x_{2,t} \\ \sum_{t=1}^{T} x_{2,t} & \sum_{t=1}^{T} x_{1,t}x_{2,t} & \sum_{t=1}^{T} x_{2,t}^2 \end{pmatrix}$$

Therefore 15 stores are involved in the investigation.

b) Find the OLS estimators of α, β_1 and β_2.

The OLS esimators $(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2)'$ are calculated as

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X'X)^{-1}X'y = \begin{pmatrix} -0.2 \\ 4.75 \\ 0.225 \end{pmatrix}$$

c) Test the null hypothesis that coefficients associated with the regressors $x_{1,t}$ and $x_{2,t}$ are equal at a 5% level of significance.

We're testing the hypothesis $H_0 : r'\beta = q$ with $r' = (0, 1, -1)$ and $q = 0$. Using the formulas in the script, we need the estimated standard error of $r'\hat{\beta}$: $\hat{\sigma}\sqrt{r'(X'X)^{-1}r}$ with $\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-K-1}$. In this case

$$\begin{aligned} \widehat{\text{se}}(r'\hat{\beta}) &= \sqrt{\frac{90.28125}{17} \begin{pmatrix} 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 12.55 & -1.25 & 0.5 \\ -1.25 & 0.125 & -0.05 \\ 0.5 & -0.05 & 0.045 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}} \\ &= 1.197447 \end{aligned}$$

Then the test statistic is $\frac{r'\hat{\beta}-q}{\widehat{\text{se}}(r'\hat{\beta})} = 3.778874$. The critical values are the 2.5% and 97.5% quantile of a t distribution with 17 degrees of freedom, so -2.11 and 2.11. That means that the null hypothesis is rejected at the 5% level.

Exercise 24

Consider a dataset with 500 observations of the variables y, x_1, x_2, x_3 and the specification

$$y_t = \alpha + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + u_t$$

under the validity of A-, B- and C-assumptions. The following values are given

$$\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} = 2912.778 \ , \qquad \hat{\boldsymbol{\beta}} = \begin{pmatrix} 10.85 \\ 2.14 \\ -0.28 \\ 9.53 \end{pmatrix}$$

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 0.0229 & -0.0020 & -0.0005 & -0.0060 \\ -0.0020 & 0.0020 & -0.0001 & 0.0001 \\ -0.0005 & -0.0001 & 0.0031 & -0.0019 \\ -0.0060 & 0.0001 & -0.0019 & 0.0033 \end{pmatrix} .$$

a) Test $H_0 : \beta_2 = 3$ vs. $H_1 : \beta_2 \neq 3$.

First we calculate $\hat{\sigma}$ as

$$\sqrt{\frac{\hat{u}'\hat{u}}{T-K-1}} = \sqrt{5.872536} = 2.423332 \ .$$

Next we estimate the standard error of $\hat{\beta}_2$. We use the formula $\text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$. The variance of $\hat{\beta}_2$ (3rd element in parameter vector) is σ^2 multiplied with the element of $(X'X)^{-1}$ in the 3rd row and 3rd column.

$$\widehat{\text{se}}(\hat{\beta}_2) = \sqrt{\frac{2912.778}{500-3-1} \cdot 0.0031} = 0.1349254$$

That results in a test statistic

$$\frac{-0.28-3}{0.1349254} = -24.30973$$

. The critical values are 2.5% and 97.5% quantile of a t-distribution with 496 degrees of freedom (± 1.96). Therefore, H_0 is rejected.

After your preliminary analysis you find an additional observed variable x_4 and may want to include it in your specification.

$$y_t = \alpha + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \beta_4 x_{4,t} + u_t \ .$$

The values for the new specification are given by

$$\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} = 117.138 \ , \qquad \hat{\boldsymbol{\beta}} = \begin{pmatrix} 1.061 & 1.965 & 2.967 & 4.040 & 4.986 \end{pmatrix}'$$

$$
(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 0.0572 & -0.0014 & -0.0119 & 0.0132 & -0.0175 \\ -0.0014 & 0.0020 & -0.0003 & 0.0004 & -0.0003 \\ -0.0119 & -0.0003 & 0.0069 & -0.0083 & 0.0058 \\ 0.0132 & 0.0004 & -0.0083 & 0.0141 & -0.0098 \\ -0.0175 & -0.0003 & 0.0058 & -0.0098 & 0.0089 \end{pmatrix}.
$$

b) Test $H_0 : \beta_2 = 3$ vs. $H_1 : \beta_2 \neq 3$.

To test H_0 we do the same steps as in a). The results are summarized below:

$$
\hat{\sigma} = 0.4865 \qquad \widehat{\text{se}}(\hat{\beta}_2) = 0.0404 \qquad \text{tStat} = -0.8167
$$

The critical values now stem from a t distribution with 495 degrees of freedom, which are also rounded to $\pm$ 1.96. H_0 is not rejected.

c) How would you explain the difference in the results from a) and b)?

- Model fits better measured ar residual variance
- additional variable seems to have an impact on y
- estimates of other parameters change by adding x_4
- "Omitted variable bias"?

d) Can you be sure that your result from b) is valid?

The second model is surely better than the first one. The decision if a fifth variable shall be added to the model, can not be seen from the data given.

NOTE:
There was a typo in this exercise: Testing $\beta_3 = 3$ would not be useful in this situation. For those who want to check results:

a) $\hat{\sigma} = 2.4233 \qquad \widehat{\text{se}}(\beta_3) = 0.1392 \qquad \text{tStat} = 49.9076$
$\Rightarrow$ Reject H_0

b) $\hat{\sigma} = 0.4865 \qquad \widehat{\text{se}}(\beta_3) = 0.0578 \qquad \text{tStat} = 18.0042$
$\Rightarrow$ Reject H_0

Exercise 25

Consider a regression model $y_t = \alpha + \beta x_t + u_t$. The parameters are estimated by OLS and the following sums are obtained:

$$
\sum_{t=1}^{15} y_t = 63.8808 \qquad \sum_{t=1}^{15} y_t^2 = 289.9335 \qquad \sum_{t=1}^{15} \hat{y}_t^2 = 272.1864
$$

a) Calculate the adjusted coefficient of determination

$$\bar{R}^2 = 1 - \frac{\frac{1}{T-k-1}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{\frac{1}{T-1}\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

For this we use that $\sum_{t=1}^{T} y_t\hat{y}_t = \sum_{t=1}^{T} \hat{y}_t^2$

$$\begin{aligned}
\sum_{t=1}^{T}(y_t - \hat{y}_t)^2 &= \sum_{t=1}^{T} y_t^2 - 2\cdot\sum_{t=1}^{T} y_t\hat{y}_t + \sum_{t=1}^{T}\hat{y}_t^2 \\
&= \sum_{t=1}^{T} y_t^2 - \sum_{t=1}^{T}\hat{y}_t^2 \\
&= 189.9335 - 272.1864 = 17.7471 \\
\sum_{t=1}^{T}(y_t - \bar{y})^2 &= \sum_{t=1}^{T} y_t^2 - T\cdot(\bar{y})^2 \\
&= 289.9335 - 15\cdot\left(\frac{63.8808}{15}\right) = 17.88306
\end{aligned}$$

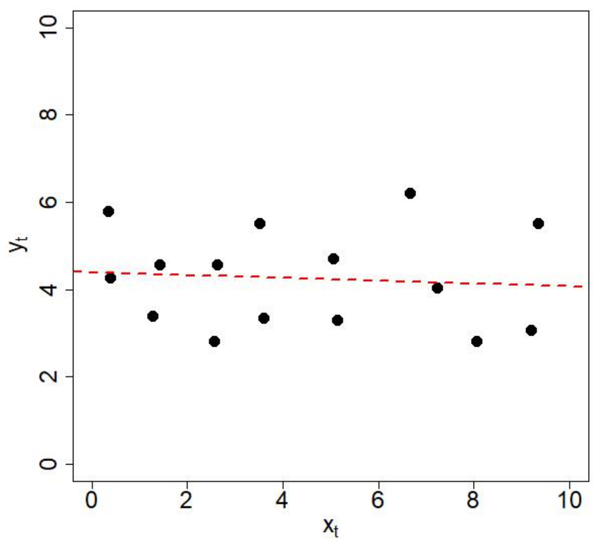
$\Rightarrow \bar{R}^2 = 1 - \frac{17.7471/13}{17.88306/14} = -0.069$

The adjusted coefficient of determination is slightly negative. The normal R^2 does not take negative values, but the adjusted R^2 can. From the alternative formula

$$\bar{R}^2 = 1 - (1-R)\frac{T-1}{N-K-1}$$

it can be seen that the adjusted R^2 is negative if $R^2 < \frac{K}{T-1}$.

b) The observations are displayed in the figure below. How do you explain the results in a)?



The results in a) are caused by the fact that x does not have a high impact on y. Therefore one does not gain a lot of information about y_t by knowing x_t.

Exercise 26

Consider the multivariate linear regression model $y = X\beta + u$. The exogenous variables summarized in matrix X were assumed to be non-stochastic. That is not plausible in many economic applications. For this exercise we replace the assumption "X is not stochastic" by the two assumptions $\mathrm{E}(u|X) = 0$ and $\mathrm{Var}(u|X) = \sigma^2 I_T$, where I_T is the identity matrix of dimension $T \times T$.
For the following proofs you can use the *law of iterated expectation* and the *law of total variance*. For two random variables X and Y they are defied as:

$$\begin{aligned}\mathrm{E}(X) &= \mathrm{E}\left[\mathrm{E}\left(X|Y\right)\right]\\ \mathrm{Var}(X) &= \mathrm{Var}\left[\mathrm{E}\left(X|Y\right)\right] + \mathrm{E}\left[\mathrm{Var}\left(X|Y\right)\right]\end{aligned}$$

a) Show that the OLS estimator $\hat{\beta} = (X'X)^{-1}X'y$ is unbiased for β.

$$\mathrm{E}(\hat{\beta}) = \mathrm{E}\left[\mathrm{E}(\hat{\beta}|X)\right] = \mathrm{E}(\beta) = \beta$$

b) Derive the variance of the OLS estimator $\hat{\beta}$.

$$\begin{aligned}\mathrm{Var}(\hat{\beta}) &= \mathrm{Var}\left[\mathrm{E}(\hat{\beta}|X)\right] + \mathrm{E}\left[\mathrm{Var}(\hat{\beta}|X)\right] = \mathrm{Var}\left[\beta\right] + \mathrm{E}\left[\sigma^2(X'X)^{-1}\right]\\ &= 0 + \sigma^2 \cdot \mathrm{E}\left((X'X)^{-1}\right) = \sigma^2 \cdot \mathrm{E}\left((X'X)^{-1}\right)\end{aligned}$$

To estimate $\mathrm{Var}(\hat{\beta})$ we estimate σ^2 as usual with $\hat{\sigma}^2$. E((X'X)$^{-1}$) is estimated with the observed $(X'X)^{-1}$.
Conclusion: The OLS estimator is still unbiased even with stochastic regressors. The Variance of estimates is also estimated in the same way.
Without proof: The OLS estimator is still **B**est**L**inear**U**nbiased**E**stimator.

Exercise 27

Consider the multivariate linear regression model $y = X\beta + u$ with $K = 3$ exogenous variables and an intercept. How would you test the following hypotheses? Translate the hypotheses to the notation $R\,\beta(\geq / = / \leq)\,q$ and sketch the steps to find the test decision if possible.

To formulate the hypotheses as $R\beta = q$, we set up the matrix R, which has as many columns as parameters and as many rows as equal signs in the hypotheses.

a) $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

$$R = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad q = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \qquad \Rightarrow \qquad \text{F-test}$$

b) $H_0 : \beta_1 = \beta_2 = \beta_3$

$$R = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \qquad q = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \Rightarrow \qquad \text{F-test}$$

c) $H_0 : \beta_1 = 2\beta_2 = 3\beta_3$

$$R = \begin{pmatrix} 0 & 1 & -2 & 0 \\ 0 & 1 & 0 & -3 \end{pmatrix} \qquad q = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \Rightarrow \qquad \text{F-test}$$

d) $H_0 : \frac{\beta_0}{\beta_1} = 2$

$\frac{\beta_0}{\beta_1} = 2 \Leftrightarrow \beta_0 - 2\beta_1 = 0$

$$R = \begin{pmatrix} 1 & -2 & 0 & 0 \end{pmatrix} \qquad q = 0 \qquad \Rightarrow \qquad \text{F- or t-test}$$

e) $H_0 : \beta_1 \geq \beta_2$

$$R = \begin{pmatrix} 0 & 1 & -1 & 0 \end{pmatrix} \qquad q = 0 \qquad \Rightarrow \qquad \text{t-test for } H_0 : R\beta \geq 0$$

Exercise 28

To estimate the consumption function for the years 1962 to 2001, the regression model

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + u_t$$

is considered with

$$\begin{aligned} y_t &: \text{Consumption} \\ x_{1,t} &: \text{Income} \\ X_{2,t} &: \text{Interest rate.} \end{aligned}$$

The following results are obtained:

$$X'X = \begin{pmatrix} 40 & 0 & 10 \\ 0 & 100 & 10 \\ 10 & 10 & 20 \end{pmatrix}$$

The OLS estimators were first calculated without the variable $x_{2,t}$.

a) On what condition are $\hat{\beta}_0$ and $\hat{\beta}_1$ unbiased?

First we derive the formula for omited variable bias. Therefore we assume that the true regression model is $y = X\beta + u$. The matrix X is then split into variables that stay in the reduced model X_a and variables that are omitted X_b so that $X = \begin{pmatrix} X_a & X_b \end{pmatrix}$. The regression model can then be written as $y = X_a\beta_a + X_b\beta_b + u$.
We only estimate the reduced model $y = X_a\beta_a + \tilde{u}$. The OLS estimator is

$$\hat{\beta}_a = (X_a'X_a)^{-1}X_a'y$$

We plug in what we know about y so that

$$\begin{aligned} \hat{\beta}_a &= (X_a'X_a)^{-1}X_a'y \\ &= (X_a'X_a)^{-1}X_a'(X_a\beta_a + X_b\beta_b + u) \\ &= \underbrace{(X_a'X_a)^{-1}X_a'X_a}_{=I}\beta_a + (X_a'X_a)^{-1}X_a'X_b\beta_b + (X_a'X_a)^{-1}X_a'u \\ &= \beta_a + (X_a'X_a)^{-1}X_a'X_b\beta_b + (X_a'X_a)^{-1}X_a'u \end{aligned}$$

To calculate the omitted variable bias $\mathrm{E}(\hat{\beta}_a) - \beta_a$, we calculate the expectation

$$\begin{aligned} \mathrm{E}(\hat{\beta}_a) &= \mathrm{E}(\beta_a + (X_a'X_a)^{-1}X_a'X_b\beta_b + (X_a'X_a)^{-1}X_a'u) \\ &= \mathrm{E}\left[\underbrace{\beta_a}_{\text{not stochastic}}\right] + \mathrm{E}\left[\underbrace{(X_a'X_a)^{-1}X_a'X_b\beta_b}_{\text{not stochastic}}\right] + \mathrm{E}\left[\underbrace{(X_a'X_a)^{-1}X_a'}_{\text{not stochastic}}u\right] \\ &= \beta_a + (X_a'X_a)^{-1}X_a'X_b\beta_b + (X_a'X_a)^{-1}X_a'\underbrace{\mathrm{E}(u)}_{=0} \\ &= \beta_a + (X_a'X_a)^{-1}X_a'X_b\beta_b \end{aligned}$$

Then the omitted variable bias is

$$E(\hat{\beta}_a) - \beta_a = \beta_a + (X_a'X_a)^{-1}X_a'X_b\beta_b - \beta_a = (X_a'X_a)^{-1}X_a'X_b\beta_b \ .$$

It can be seen that the omitted variable bias is zero, whenever $X_a'X_b$ contains only zeros, or if β_b is zero. The first case implies that every omitted variable is not correlated with any variable left in the reduced model. The latter case implies that the omitted variables do not have an impact on y.

b) Is there an omitted variable bias?

For the given situation we can look at the matrix $X'X$:

$$X'X = \begin{pmatrix} X_a' \\ X_b' \end{pmatrix} \cdot \begin{pmatrix} X_a & X_b \end{pmatrix} = \begin{pmatrix} X_a'X_a & X_a'X_b \\ X_b'X_a & X_b'X_b \end{pmatrix}$$

The first two variables stay in the reduced model and the third variable is omitted. Therefore, we know that $X_a \in \mathbb{R}^{T\times 2}, X_b \in \mathbb{R}^{T\times 1}$.

$$\Rightarrow X'X = \left(\begin{array}{cc|c} 40 & 0 & 10 \\ 0 & 100 & 10 \\ \hline 10 & 10 & 20 \end{array}\right) = \left(\begin{array}{c|c} X_a'X_a & X_a'X_b \\ \hline X_b'X_a & X_b'X_b \end{array}\right)$$

In this case, $X_a'X_b$ does not contain only zeros, so there is an omitted variable bias, whenever the true parameter β_2 is not zero.

c) Calculate the bias $\mathrm{E}(\hat{\beta}_0) - \beta_0$ and $\mathrm{E}(\hat{\beta}_1) - \beta_1$ by assuming $\beta_2 = -0.1$.

By using the above notation, we see that

$$\begin{aligned} \mathrm{E}(\hat{\beta}_a) - \beta_a &= (X_a'X_a)^{-1} X_a'X_b\beta_b \\ &= \begin{pmatrix} 40 & 0 \\ 0 & 100 \end{pmatrix}^{-1} \begin{pmatrix} 10 \\ 10 \end{pmatrix} \cdot (-0.1) \\ &= \begin{pmatrix} 1/40 & 0 \\ 0 & 1/100 \end{pmatrix} \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1/40 \\ -1/100 \end{pmatrix} \end{aligned}$$

d) Assuming that there is no omitted variable bias. What consequences would if expect when adding $x_{2,t}$ to the model?

There are two possible scenarios. The variable $x_{2,t}$ can be uncorrelated to the other exogenous variables and still have an impact on y. Then the model would be better including $x_{2,t}$. On the other side, $x_{2,t}$ can be irrelevant for y. Then the standard errors would increase. Including irrelevant variables to the model increases the uncertainty in the estimation of all parameters.

Exercise 29

A friend had asked you to do a regression for his master's thesis. For the model

$$y_t = \alpha + \beta x_t + u_t \qquad t = 1, ..., T$$

you estimated the parameters given the following information

$$\sum x_t = 19.3 \qquad \sum x_t^2 = 47.16 \qquad \sum x_t y_t = 654.4$$

$$\sum y_t = 272.48 \qquad \sum y_t^2 = 9223.33 \qquad \bar{x} = 1.93$$

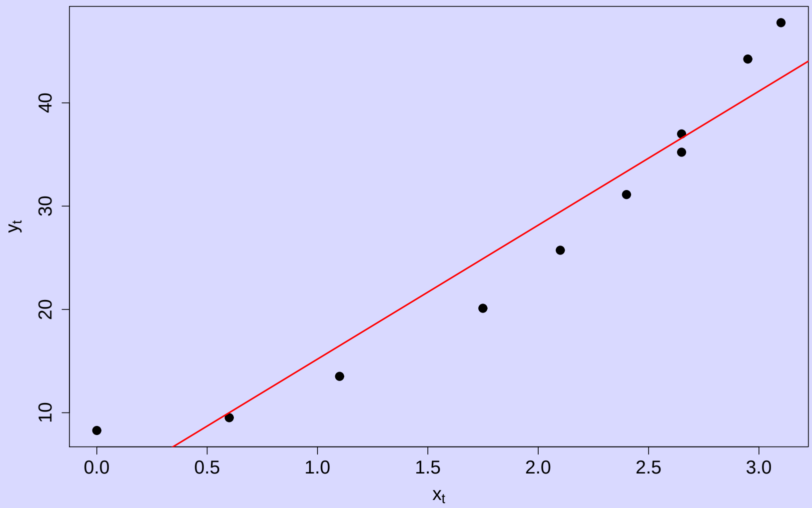
After speaking to his professor, your friends tells you about two problems. First, he omitted a relevant variable and, second, x_t has a quadratic impact on y_t.

a) Your friend asks you why you did not tell him about these problems. Was it possible for you to recognize them without further information about the theoretical background of the data?

There was no way to see the problems given only the above sums. From them you could calculate the correlation between x_t and y_t and the coefficient of determination

$$\rightarrow \qquad \text{Cov}(x_t, y_t) = 0.9625 \qquad R^2 = 0.9254$$

Both indicate a linear dependency between the variables. The omitted, relevant variable can also not be seem from the data. In a plot, first problem would have been visible:



b) Which assumptions are violated by the described problems and what are the consequences?

The first problem of an omitted variable implies that assumption A1 is violated. The consequence is - as proven before - a bias in the estimation of the reduced model.
The problem of an undetected quadratic impact shows that assumption A2 is violated. The model is specified incorrectly and may lead to biased estimates, wrong prognoses, ...
Also there is autocorrelation in the error terms. By looking at the plot in a), one can see that the residuals at the edges are positive and in the center negative.

c) Consider the model $y_t = \alpha + \beta_1 x_{1,t} + \beta_2 x_{2,t} + u_t$ where $x_{2,t} = x_t^2$. Assume that the A-, B- and C-assumptions hold. Calculate the OLS esimator.
First calculations provide the following results.

See ex. d)

$$X = \begin{pmatrix} 1 & 7.02 & 0.10 \\ 1 & 5.76 & 0.30 \\ 1 & 9.61 & 0.39 \\ 1 & 1.21 & 0.41 \\ 1 & 4.41 & 0.26 \\ 1 & 0.00 & 0.11 \\ 1 & 8.70 & 0.55 \\ 1 & 7.02 & 0.63 \\ 1 & 0.36 & 0.36 \\ 1 & 3.06 & 0.27 \end{pmatrix} \qquad y = \begin{pmatrix} 35.22 \\ 31.12 \\ 47.76 \\ 13.52 \\ 25.73 \\ 8.28 \\ 44.24 \\ 36.99 \\ 9.52 \\ 20.11 \end{pmatrix}$$

$$X'X = \begin{pmatrix} 10.00 & 47.15 & 3.38 \\ 47.15 & 330.19 & 17.98 \\ 3.38 & 17.98 & 1.40 \end{pmatrix} \quad (X'X)^{-1} = \begin{pmatrix} 0.59 & -0.02 & -1.13 \\ -0.02 & 0.01 & -0.09 \\ -1.13 & -0.09 & 4.54 \end{pmatrix}$$

$$X'y = \begin{pmatrix} 272.49 \\ 1724.82 \\ 101.12 \end{pmatrix} \qquad My = \begin{pmatrix} -0.58 \\ -0.22 \\ 0.67 \\ 0.15 \\ -0.06 \\ 0.68 \\ 0.32 \\ -0.41 \\ -0.28 \\ -0.28 \end{pmatrix}$$

d) Your friend believes that $\beta_1 = \beta_2$ and, in addition, $\alpha = \beta_1 + \beta_2$. Test his hypothesis with one of the tests discussed in the lecture.

First is is noted, that the model contains the omitted variable ($x_{1,t}$) and the variable from the reduces, first model, but squared ($x_{2,t}$).
In class we discussed the consequences of rounded values. We will first derive the solution to this exercise and then look at different outcomes depending on when and if values were rounded.
The OLS estimator is simply $\hat{\beta} = (X'X)^{-1}X'y$. To test the hypotheses (simultaneously), we set up R and q

$$R = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 1 & 1 \end{pmatrix} \quad q = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

There we can see that $L = 2$. It can be tested with an F test and test statistic

$$
\begin{aligned}
F &= \frac{(R\hat{\beta} - q)' \left[R(X'X)^{-1}R'\right]^{-1} (R\hat{\beta} - q)/L}{\hat{u}'\hat{u}/(T-K-1)} \\
&= \frac{(R\hat{\beta} - q)' \left[R(X'X)^{-1}R'\right]^{-1} (R\hat{\beta} - q)/L}{\hat{\sigma}^2} \\
&= (R\hat{\beta} - q)' \left[\hat{\sigma}^2 R(X'X)^{-1}R'\right]^{-1} (R\hat{\beta} - q)/L \\
&= (R\hat{\beta} - q)' \left[R\widehat{\mathrm{Var}}(\hat{\beta})R'\right]^{-1} (R\hat{\beta} - q)
\end{aligned}
$$

$\hat{\sigma}^2$ is calculated as $\hat{u}'\hat{u}/(T-K-1) = (My)'My/(T-K-1)$.
Then after calculating F, it is compared to the $1-\alpha$ quantile of an F-distribution with L and $T-K-1$ degrees of freedom (4.7374) and H_0 is rejected if F exceeds this value. Below different approaches are summarized: Using the exact values and rounding in each step to 2,3 or 4 digits.

	Exact	2 Digits	3 Digits	4 Digits
$\hat{\alpha}$	7.263110	11.00	7.871	7.3070
$\hat{\beta}_1$	4.021619	2.70	4.282	4.0175
$\hat{\beta}_2$	3.029453	3.75	3.021	2.9797
$\hat{\sigma}^2$	0.250028	66.84	6.093	0.2506
F	3.892633	0.34	0.160	3.6208
Decision	H_0 not rejected	H_0 not rejected	H_0 not rejected	H_0 not rejected

You can see that the estimates and test statistics are quite different in the scenarios. In addition to that, here is another version, how I would've done it:

$$
\hat{\beta} = \begin{pmatrix} 0.59 & -0.02 & -1.13 \\ -0.02 & 0.01 & -0.09 \\ -1.13 & -0.09 & 4.54 \end{pmatrix} \begin{pmatrix} 272.49 \\ 1724.82 \\ 101.12 \end{pmatrix} = \begin{pmatrix} 12.0071 \\ 2.6976 \\ -4.0627 \end{pmatrix}
$$

$$
R\hat{\beta} = \begin{pmatrix} 6.7603 \\ -13.3722 \end{pmatrix} \qquad \hat{\sigma}^2 = \frac{(-0.58)^2 + (-0.22)^2 + ...}{7} = \frac{1.7495}{7} = 0.2499286 \approx 0.25
$$

$$
R(X'X)^{-1}R' = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0.59 & -0.02 & -1.13 \\ -0.02 & 0.01 & -0.09 \\ -1.13 & -0.09 & 4.54 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 4.73 & -5.64 \\ -5.64 & 7.26 \end{pmatrix}
$$

$$
\begin{aligned}
\left[R(X'X)^{-1}R'\right]^{-1} &= \begin{pmatrix} 4.73 & -5.64 \\ -5.64 & 7.26 \end{pmatrix}^{-1} = \frac{1}{4.73 \cdot 7.26 - 5.64^2} \begin{pmatrix} 7.26 & 5.26 \\ 5.64 & 4.73 \end{pmatrix} \\
&\approx \begin{pmatrix} 2.8693 & 2.2291 \\ 2.2291 & 1.8694 \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
F &= \frac{\begin{pmatrix} 6.7603 & -13.3722 \end{pmatrix} \begin{pmatrix} 2.8693 & 2.2291 \\ 2.2291 & 1.8694 \end{pmatrix} \begin{pmatrix} 6.7603 \\ -13.3722 \end{pmatrix} /2}{0.25} \\
&= 4 \cdot 62.38824 = 249.553
\end{aligned}
$$

This test statistic is a lot larger that the ones before. The only difference in the calculation was when and where it was rounded...

Exercise 30

At the football world cup 2010 in South Africa, the atmosphere at match t (y_t) was measured by a rating system. To model it, the regression model

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

is assumed, where x_t is the number of vuvuzelas (in 10 thousand). The following results were obtained from the study

$$X'X = \begin{pmatrix} 64 & 288.128 \\ 288.128 & 1842.3374 \end{pmatrix} \qquad (X'X)^{-1} = \begin{pmatrix} 0.0528 & -0.0083 \\ -0.0083 & 0.0018 \end{pmatrix}$$

$$X'y = \begin{pmatrix} 6080 \\ 27496 \end{pmatrix}$$

a) Calculate the OLS estimator.

To estimate the coefficients we use the formula $\hat{\beta} = (X'X)^{-1}X'y$. Since the given values in $(X'X)^{-1}$ are rounded to 4 decimal places, and produce implausible results, we use the exact inverse:

$$\begin{aligned}(X'X)^{-1} &= \frac{1}{64 \cdot 1842.3374 - 288.128^2} \begin{pmatrix} 1842.3374 & -288.128 \\ -288.128 & 64 \end{pmatrix} \\ &= \begin{pmatrix} 0.052801369 & -0.008257745 \\ -0.008257745 & 0.001834239 \end{pmatrix}\end{aligned}$$

Then the OLS estimates are

$$\hat{\beta} = \begin{pmatrix} 93.9773608 \\ 0.2271522 \end{pmatrix} \qquad \hat{\beta}_{\text{rounded}} = \begin{pmatrix} 92.8072 \\ -0.9712 \end{pmatrix}$$

b) Calculate the coefficient of determination under the assumption $y'y = 577700$.

Here we use the formula $R^2 = \frac{y'X\hat{\beta} - T\bar{y}^2}{y'y - T\bar{y}^2}$. First we see that $y'X\hat{\beta} = (\hat{\beta}'X'y)' = 577628.1$, so we plug in the values:

$$\begin{aligned}R^2 &= \frac{577628.1 - 64 \cdot \frac{6080^2}{64^2}}{577700 - 64 \cdot \frac{6080^2}{64^2}} \\ &= 0.2813053\end{aligned}$$

Using the same formula but with the rounded $(X'X)^{-1}$ matrix, you would get $R^2_{\text{rounded}} = -400.3634$.

It would give the same results if you had used:

$$\begin{aligned}R^2 &= 1 - \frac{S_{\hat{u}\hat{u}}}{S_{yy}} \\ S_{\hat{u}\hat{u}} &= y'y - \hat{\beta}'X'y \;.\end{aligned}$$

c) Give a prognosis for the atmosphere rating if $x_0 = 3$.

The prognosis $\hat{y}_0$ for a value x_0 is simply

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_0 = 94.65882 \ .$$

Again, if you used the rounded matrix, it yields $\widehat{y_{0,\text{rounded}}} = 89.8936$.

d) Conduct a 95% prognosis interval for y_0 with $x_0 = 3$. Thereby assume an error variance $\hat{\sigma}^2 = 1$.

To construct the prognosis interval for $\hat{y}_0$, we first look at he formula

$$\text{PI}_{\hat{y}_0} = \hat{y}_0 \pm t_{1-\frac{\alpha}{2};T-2} \cdot \widehat{\text{se}}(\hat{y}_0 - y_0) \ .$$

where

$$\begin{aligned}
\widehat{\text{se}}(\hat{y}_0 - y_0) &= \sqrt{\widehat{\text{Var}}(\hat{y}_0 - y_0)} \\
\widehat{\text{Var}}(\hat{y}_0 - y_0) &= \hat{\sigma}^2 \left[1 + \frac{1}{T} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right] \ .
\end{aligned}$$

We gather the values needed

$$\begin{aligned}
\hat{\sigma}^2 &= 1 \\
\frac{1}{T} &= \frac{1}{64} \\
(x_0 - \bar{x})^2 &= \left(3 - \frac{288.128}{64} \right)^2 = (3 - 4.502)^2 \\
&= 2.256004 \\
S_{xx} &= \sum_{t=1}^{T} (x_t - \bar{x})^2 = \sum_{t=1}^{T} x_t^2 - T\bar{x}^2 \\
&= 1842.3374 - 64 \cdot 4.502^2 = 545.1851 \\
\widehat{\text{Var}}(\hat{y}_0 - y_0) &= 1 + \frac{1}{64} + \frac{2.256004}{545.1851} = 1.019763 \\
\widehat{\text{se}}(\hat{y}_0 - y_0) &= \sqrt{1.019763} = 1.009833 \\
t_{1-\frac{\alpha}{2};62} &= 1.999
\end{aligned}$$

So that finally the prognosis interval is

$$[92.64016; 96.67748] \ .$$

Calculating with rounded values, $\hat{y}_0$ is the only value that changes. It would give

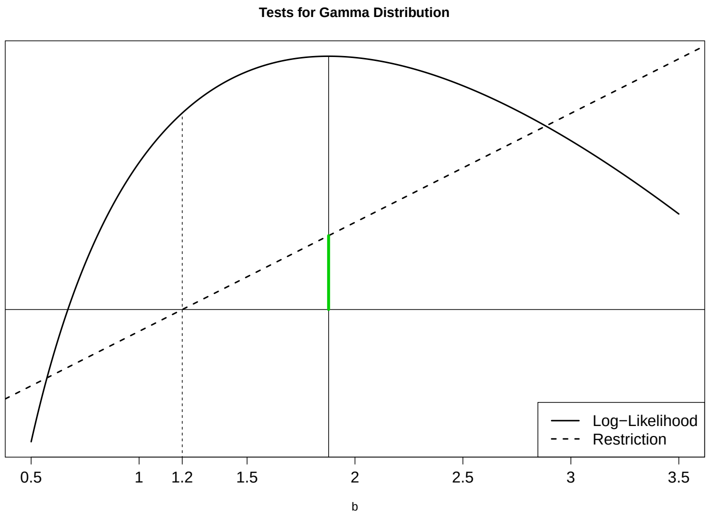
$$[87.87494; 91.91226] \ .$$

Exercise 31

Explain the idea behind Wald-, LR- and LM-test by summarizing the idea in two sentences and completing the images.

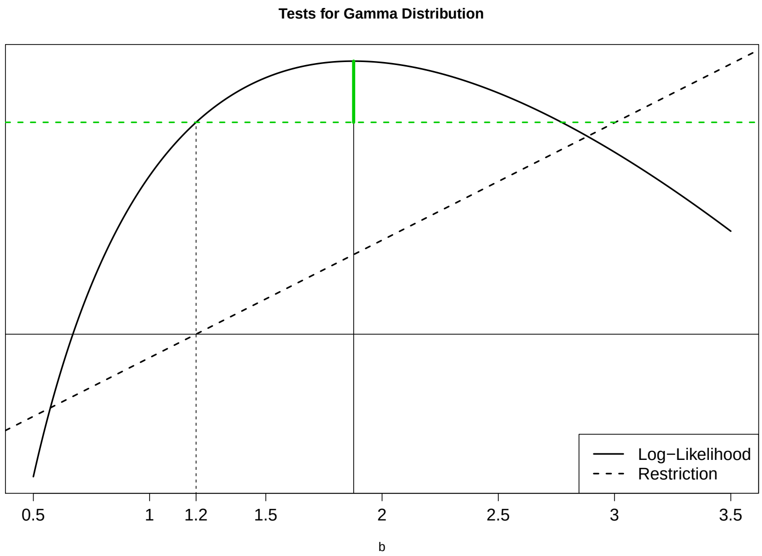
a) Wald-test

Idea: Calculate the ML-estimator and $R\hat{\beta}_{\text{ML}} - q$ to test $H_0 = R\beta = q$. The more it differs from zero, the more unlikely is the null hypothesis.



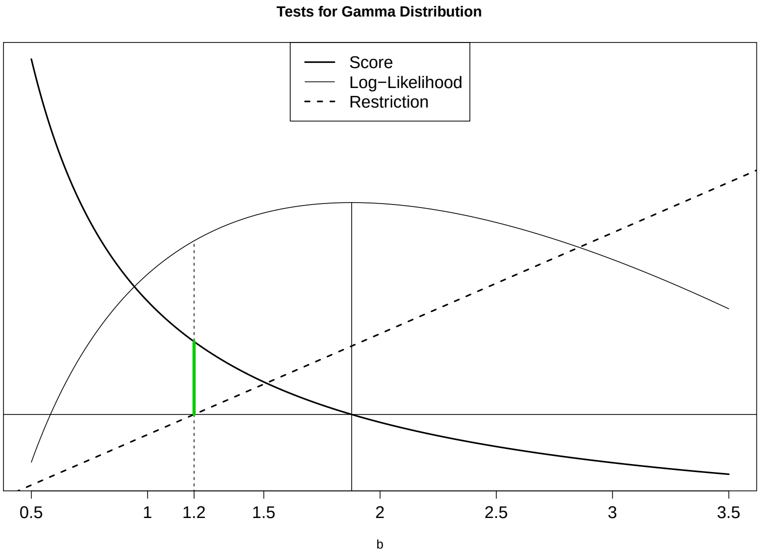
b) LR-test

Idea: Is likelihood of restriction smaller than the one without restrictions?



c) LM-test

Is the slope of likelihood different from zero at restriction?



Exercise 32

The *International Monetary Found* hires you to estimate the credit demand function for the private sector

$$y_t = \alpha + \beta x_t + u_t \qquad t = 1, ..., T$$

based on a sample of $T = 40$ countries. Thereby y_t gives the total of private credits in country t (in billion USD) and x_t the average nominal interest rate in country t (in percent). Further you get the following values:

$$\sum_{t=1}^{T} x_t = 250; \quad \sum_{t=1}^{T} x_t^2 = 1600; \quad \sum_{t=1}^{T} y_t = 200; \quad \sum_{t=1}^{T} y_t^2 = 1015.78; \quad \sum_{t=1}^{T} x_t y_t = 1227.5$$

a) Estimate the coefficients α and β by OLS and interpret.

Here we use $\hat{\beta} = \frac{S_{xy}}{S_{xx}}$ and $S_{xy} = \left(\sum_{t=1}^{T} x_t y_t\right) - T\bar{x}\bar{y}$. So the estimates are

$$\hat{\beta} = \frac{1227.5 - 40 \cdot \frac{250 \cdot 200}{40^2}}{1600 - 40 \cdot \frac{250^2}{40^2}} = -0.6$$
$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = \frac{200}{40} + 0.6 \cdot \frac{250}{40} = 8.75$$

We keep in mind that $S_{xy} = -22.5$ and $S_{xx} = 37.5$.
The estimates indicate that raising the interest rate by 1% (e.g. 3% $\rightarrow$ 4%) would cause the total private credits sum to go down by 0.6 billion USD (everything else unchanged).

b) Calculate and interpret the coefficient of determination R^2.

The easiest way here is to use

$$R^2 = \left(\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}\right)^2 = \frac{\frac{S_{xy}}{S_{xx}}S_{xy}}{S_{yy}} = \frac{\hat{\beta}S_{xy}}{S_{yy}} \ .$$

With $S_{yy} = 15.78$ we calculate

$$R^2 = \frac{0.6 \cdot 22.5}{15.78} = 0.8555 \ .$$

c) Estimate the error variance $\sigma^2 = \text{Var}(u_t)$.

From the lecture we know that

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{T-2}\sum_{t=1}^{T}\hat{u}^2 = \sum_{t=1}^{T}(\hat{u} - \underbrace{\bar{\hat{u}}}_{=0})^2 = \frac{1}{T-2}S_{\hat{u}\hat{u}} \\
R^2 &= 1 - \frac{S_{\hat{u}\hat{u}}}{S_{yy}} \ .
\end{aligned}$$

The estimated error variance is then

$$\hat{\sigma}^2 = \frac{1}{T-2}S_{yy}(1-R^2) = \frac{1}{38} \cdot 15.78 \cdot (1 - 0.8555) = 0.06 \ .$$

d) Test if the impact of nominal interest rate on the total of private credits is significantly different from zero.

The null hypothesis to test is $H_0 = \beta = 0$ vs. $H_1 : \beta \neq 0$. This can be done by the t-test with test statistic

$$\frac{\hat{\beta}}{\widehat{\text{se}}(\hat{\beta})}$$

The standard error is simply $\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{0.06}{37.5}} = 0.04$. Using this, the test statistic $\frac{-0.6}{0.04} = -15$ is compared with the quantiles $\pm t_{1-\frac{\alpha}{2};T-2} = \pm 2.024$. The test statistic is smaller than the lower quantile, so that the null hypothesis is rejected.

e) Give a prognosis for a country with a nominal interest rate of $x_0 = 3$ (percent) and construct the 95% prognosis interval.

For the prognosis we use the estimated parameters: $\hat{y}_0 = \hat{\alpha} + \hat{\beta}x_0 = 6.95$. The values are summarized:

$$\begin{aligned}
\widehat{\text{Var}}(\hat{y}_0 - y_0) &= \hat{\sigma}^2\left[1 + \frac{1}{T}\frac{(3-6.25)^2}{37.5}\right] &&= 0.0784 \\
\widehat{\text{se}}(\hat{y}_0 - y_0) &= \sqrt{0.0784} = 0.28 \\
t_{0.975;38} &= 2.024
\end{aligned}$$

So that the prognosis interval is finally

$$[6.38328; 7.51672] \ .$$

Exercise 33

Consider a multivariate linear regression model $y = X\beta + u$. Show that the estimator $\hat{\beta}^* = (X'BX)^{-1}X'By$ is unbiased. The matrix B is not stochastic, of suitable dimension and chosen so that the inverse $X'BX$ exists.

$$
\begin{aligned}
\mathrm{E}\left(\hat{\beta}^*\right) &= \mathrm{E}\Big((X'BX)^{-1}X'By\Big) \\
&= \mathrm{E}\Big((X'BX)^{-1}X'B(X\beta + u)\Big) \\
&= \mathrm{E}\left(\underbrace{(X'BX)^{-1}X'BX}_{=I}\,\beta\right) + \mathrm{E}\left(\underbrace{(X'BX)^{-1}X'B}_{\text{not stoch.}}\,u\right) \\
&= \beta + \underbrace{(X'BX)^{-1}X'\,\underbrace{\mathrm{E}(u)}_{=0}}_{=0} = \beta
\end{aligned}
$$

df	90%	92.5%	95%	97.5%	99%	**df**	90%	92.5%	95%	97.5%	99%
1	3.078	4.165	6.314	12.706	31.821	**51**	1.298	1.462	1.675	2.008	2.402
2	1.886	2.282	2.920	4.303	6.965	**52**	1.298	1.461	1.675	2.007	2.400
3	1.638	1.924	2.353	3.182	4.541	**53**	1.298	1.461	1.674	2.006	2.399
4	1.533	1.778	2.132	2.776	3.747	**54**	1.297	1.460	1.674	2.005	2.397
5	1.476	1.699	2.015	2.571	3.365	**55**	1.297	1.460	1.673	2.004	2.396
6	1.440	1.650	1.943	2.447	3.143	**56**	1.297	1.460	1.673	2.003	2.395
7	1.415	1.617	1.895	2.365	2.998	**57**	1.297	1.459	1.672	2.002	2.394
8	1.397	1.592	1.860	2.306	2.896	**58**	1.296	1.459	1.672	2.002	2.392
9	1.383	1.574	1.833	2.262	2.821	**59**	1.296	1.459	1.671	2.001	2.391
10	1.372	1.559	1.812	2.228	2.764	**60**	1.296	1.458	1.671	2.000	2.390
11	1.363	1.548	1.796	2.201	2.718	**61**	1.296	1.458	1.670	2.000	2.389
12	1.356	1.538	1.782	2.179	2.681	**62**	1.295	1.458	1.670	1.999	2.388
13	1.350	1.530	1.771	2.160	2.650	**63**	1.295	1.457	1.669	1.998	2.387
14	1.345	1.523	1.761	2.145	2.624	**64**	1.295	1.457	1.669	1.998	2.386
15	1.341	1.517	1.753	2.131	2.602	**65**	1.295	1.457	1.669	1.997	2.385
16	1.337	1.512	1.746	2.120	2.583	**66**	1.295	1.456	1.668	1.997	2.384
17	1.333	1.508	1.740	2.110	2.567	**67**	1.294	1.456	1.668	1.996	2.383
18	1.330	1.504	1.734	2.101	2.552	**68**	1.294	1.456	1.668	1.995	2.382
19	1.328	1.500	1.729	2.093	2.539	**69**	1.294	1.456	1.667	1.995	2.382
20	1.325	1.497	1.725	2.086	2.528	**70**	1.294	1.456	1.667	1.994	2.381
21	1.323	1.494	1.721	2.080	2.518	**71**	1.294	1.455	1.667	1.994	2.380
22	1.321	1.492	1.717	2.074	2.508	**72**	1.293	1.455	1.666	1.993	2.379
23	1.319	1.489	1.714	2.069	2.500	**73**	1.293	1.455	1.666	1.993	2.379
24	1.318	1.487	1.711	2.064	2.492	**74**	1.293	1.455	1.666	1.993	2.378
25	1.316	1.485	1.708	2.060	2.485	**75**	1.293	1.454	1.665	1.992	2.377
26	1.315	1.483	1.706	2.056	2.479	**76**	1.293	1.454	1.665	1.992	2.376
27	1.314	1.482	1.703	2.052	2.473	**77**	1.293	1.454	1.665	1.991	2.376
28	1.313	1.480	1.701	2.048	2.467	**78**	1.292	1.454	1.665	1.991	2.375
29	1.311	1.479	1.699	2.045	2.462	**79**	1.292	1.454	1.664	1.990	2.374
30	1.310	1.477	1.697	2.042	2.457	**80**	1.292	1.453	1.664	1.990	2.374
31	1.309	1.476	1.696	2.040	2.453	**81**	1.292	1.453	1.664	1.990	2.373
32	1.309	1.475	1.694	2.037	2.449	**82**	1.292	1.453	1.664	1.989	2.373
33	1.308	1.474	1.692	2.035	2.445	**83**	1.292	1.453	1.663	1.989	2.372
34	1.307	1.473	1.691	2.032	2.441	**84**	1.292	1.453	1.663	1.989	2.372
35	1.306	1.472	1.690	2.030	2.438	**85**	1.292	1.453	1.663	1.988	2.371
36	1.306	1.471	1.688	2.028	2.434	**86**	1.291	1.453	1.663	1.988	2.370
37	1.305	1.470	1.687	2.026	2.431	**87**	1.291	1.452	1.663	1.988	2.370
38	1.304	1.469	1.686	2.024	2.429	**88**	1.291	1.452	1.662	1.987	2.369
39	1.304	1.468	1.685	2.023	2.426	**89**	1.291	1.452	1.662	1.987	2.369
40	1.303	1.468	1.684	2.021	2.423	**90**	1.291	1.452	1.662	1.987	2.368
41	1.303	1.467	1.683	2.020	2.421	**91**	1.291	1.452	1.662	1.986	2.368
42	1.302	1.466	1.682	2.018	2.418	**92**	1.291	1.452	1.662	1.986	2.368
43	1.302	1.466	1.681	2.017	2.416	**93**	1.291	1.452	1.661	1.986	2.367
44	1.301	1.465	1.680	2.015	2.414	**94**	1.291	1.451	1.661	1.986	2.367
45	1.301	1.465	1.679	2.014	2.412	**95**	1.291	1.451	1.661	1.985	2.366
46	1.300	1.464	1.679	2.013	2.410	**96**	1.290	1.451	1.661	1.985	2.366
47	1.300	1.463	1.678	2.012	2.408	**97**	1.290	1.451	1.661	1.985	2.365
48	1.299	1.463	1.677	2.011	2.407	**98**	1.290	1.451	1.661	1.984	2.365
49	1.299	1.462	1.677	2.010	2.405	**99**	1.290	1.451	1.660	1.984	2.365
50	1.299	1.462	1.676	2.009	2.403	**100**	1.290	1.451	1.660	1.984	2.364

Tabelle 1: Quantiles of t-distribution