# Online self assessment
# for applicants of the master program
# Econometrics

University Duisburg-Essen | Ruhr-University Bochum | TU Dortmund University

Version 2021

(1) Upon completing the self assessment you will receive a certificate.

- Disable any pop-up blocker in your browser in order to obtain your certificate of attendance afterwards.
- Download the certificate.
- Fill in the certificate.
- Sign the certificate.
- Submit the certificate with your application.

While we require you to **do** the self-assessment, we do not consider the answers you actually gave when assessing your application.

(2) This is not a knowledge test. There will be no score, grade or points at the end. And your answers are **not** used to evaluate your application.

(3) The self assessment is designed to:

- illustrate the mathematical focus of the degree program.
- give you the opportunity to judge whether the degree program is of interest to you.
- give you the chance to gauge whether you have the background to successfully complete the program.

**Read the next page carefully!!!**

## Information on self assessment - read carefully!

- Completion of the **self** assessment is **required** to apply for the master program Econometrics.

  **After completing the self assessment, you obtain a certificate, which you have to: (i) fill in, (ii) sign, and (iii) submit during the application process.**

- This is **not** an entry test! It is a **self** assessment!

- You are not expected to be able to answer all the questions.

- You should assess whether:

  (i) you have a sufficient mathematical background that you believe that with the corresponding effort you will be able to answer such questions.

  (ii) you are motivated to learn how to answer such questions.

## Structure of the self assessment - read carefully!

- You will be given multiple choice questions from the areas of *mathematics, statistics, econometrics* and *economics* along with a set of possible answers.

- Do **not answer** the actual question. You should **assess** your (expected) ability to answer the question!

- To gauge your own assessment of your ability to answer the questions you should choose one of the following responses for each question:

  (A) I know the correct answer(s).

  (B) I could work out the correct answer(s) given enough time.

  (C) I understand the question, but I have no idea how to work out the correct answer(s).

  (D) I don't understand the question.

- To provide your self assessment, please use the input mask available on the 'EvaSys' website:

  `https://evaluation.tu-dortmund.de/evasys/online.php?p=38AHR`

  The questions in the online form are numbered as below.

- You should not interrupt filling in the online form. You need to provide an answer to every question in the online form.

## Purpose of the self assessment - read carefully!

- We want to emphasize to you the mathematical focus of the degree program.

- Having gone through all the questions of the self assessment you should ask yourself:

  - (Motivation) Do I want to learn about these things? Do I find understanding these things interesting?
  - (Background) Do I think I have a background to be able to acquire the skills needed to answer these questions?

  If you do not answer these last two points with yes, then the Econometrics master might not be the right program for you.

# 1 Mathematics

- The symbol ln denotes the natural logarithm, that is, with base $e$.
- The symbols $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ denote the sets of integers, rational numbers, real numbers, and complex numbers, respectively.

## 1.1 Calculus

**Question 1.** For $a \in \mathbb{R}$, which statements do hold for the one-sided limit $\lim_{x \to a+} \dfrac{\ln(x-a)}{\ln(e^x - e^a)}$ ?

a) The limit exists.

b) Its value is 0.

c) Its value is 1.

d) Its value is $e^a$.

**Question 2.** Which statements do hold for the definite integral $\int_0^{\pi} e^{\cos x} \sin x \, dx$ ?

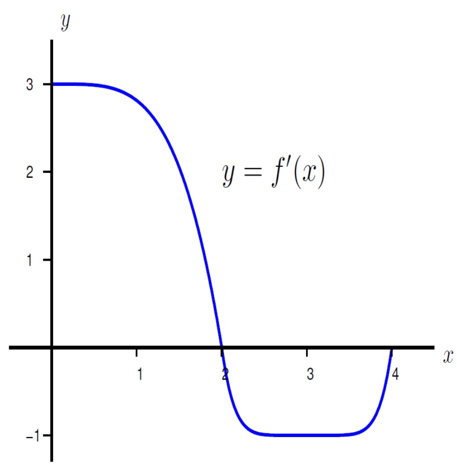
a) The antiderivative is not explicitly calculable.

b) The definite integral has a finite value.

c) Its value is $e - \dfrac{1}{e}$.

d) Its value is 0.

**Question 3.** The following figure shows the graph of the derivative $f'$ of a function $f$, where $f$ is continuous on the interval $[0,4]$ and differentiable on the interval $(0,4)$. Which of the following statements give the correct ordering of the function values $f(0)$, $f(2)$, and $f(4)$?




a) $f(0) < f(4)$

b) $f(4) < f(2)$

c) $f(2) \leq f(0)$

d) $f(4) = f(2)$

**Question 4.** Let $f$ be the function defined by the series $f(x) = \sum_{n=1}^{\infty} \frac{x^n}{n}$ for all $x$ such that $-1 < x < 1$. Which statements do hold?

a) The series converges absolutely.

b) The derivative is $f'(x) = \sum_{n=1}^{\infty} x^n$.

c) The derivative is $f'(x) = \sum_{n=0}^{\infty} x^n$.

d) The derivative equals $f'(x) = \frac{1}{1-x}$.

## 1.2 Linear Algebra

**Question 5.** Consider the following system of linear equations

$$\begin{array}{rcrcrcl} 3x & + & 2y & + & z & = & 0 \\ x & + & y & + & z & = & 0 \\ x & & & - & z & = & 0 \end{array}$$

with solutions of the form $(x, y, z)$ where $x, y, z$ are real numbers. Which of the following statements are correct?

a) The system is consistent.

b) The sum of any two solutions is a solution.

c) The system has a unique solution.

d) The system has infinitely many solutions.

**Question 6.** Which are eigenvalues of the matrix

$$\begin{pmatrix} 3 & 2 & 5 \\ 0 & 2 & 3 \\ 0 & 1 & 4 \end{pmatrix} \quad ?$$

a) 2

b) 3

c) 5

d) 0

## 1.3 Analytic Geometry

**Question 7.** Consider the solid in $xyz$-space, which contains all points $(x, y, z)$ whose $z$-coordinate satisfies

$$0 \leq z \leq 4 - x^2 - y^2$$

. Which statements do hold?

a) The solid is a sphere.

b) The solid is a pyramid.

c) Its volume is $8\pi$.

d) Its volume is $\dfrac{16\pi}{3}$.

**Question 8.** Consider the function $g$ defined by $g(x, y) = e^y(y - x^2)$ for all real $x, y$. Which of the following terms are needed to represent the length of the gradient $\nabla g(1, -1)$ ?

a) $\sqrt{10}$

b) $\sqrt{5}$

c) $e$

d) $\pi$

**Question 9.** A circular helix in $xyz$-space has the following parametric equations, where $\theta \in \mathbb{R}$.

$$\begin{aligned} x(\theta) &= 4\cos\theta \\ y(\theta) &= 4\sin\theta \\ z(\theta) &= 3\theta \end{aligned}$$

Let $L(\theta)$ be the arclength of the helix from the point $P(\theta) = (x(\theta), y(\theta), z(\theta))$ to the point $P(0) = (4, 0, 0)$, and let $D(\theta)$ be the distance between $P(\theta)$ and the origin $(0, 0, 0)$. Let $L(\theta) = 10$. Which statements do hold?

a) $\theta = 4$

b) $\theta = 2$

c) To calculate the value of $D$ for a given $\theta$, $x(\theta)$ and $y(\theta)$ have to be evaluated explicitely.

d) $D(\theta) = \sqrt{52}$

### 1.4 Differential Equations

**Question 10.** Let $y : \mathbb{R} \to \mathbb{R}$ be the real-valued function defined on the real line, which is the solution of the initial value problem

$$y' = -xy + x, \qquad y(0) = 2.$$

Which statements are correct?

a) The problem is not uniquely solvable.

b) The solution $y(x)$ contains an exponential function.

c) $\lim_{x\to\infty} y(x) = 1$

d) $\lim_{x\to\infty} y(x) = 0$

# 2 Statistics

## 2.1 Descriptive Statistics

**Question 11.** Which of the following sets have an arithmetic mean of 100, but a median smaller than 100?

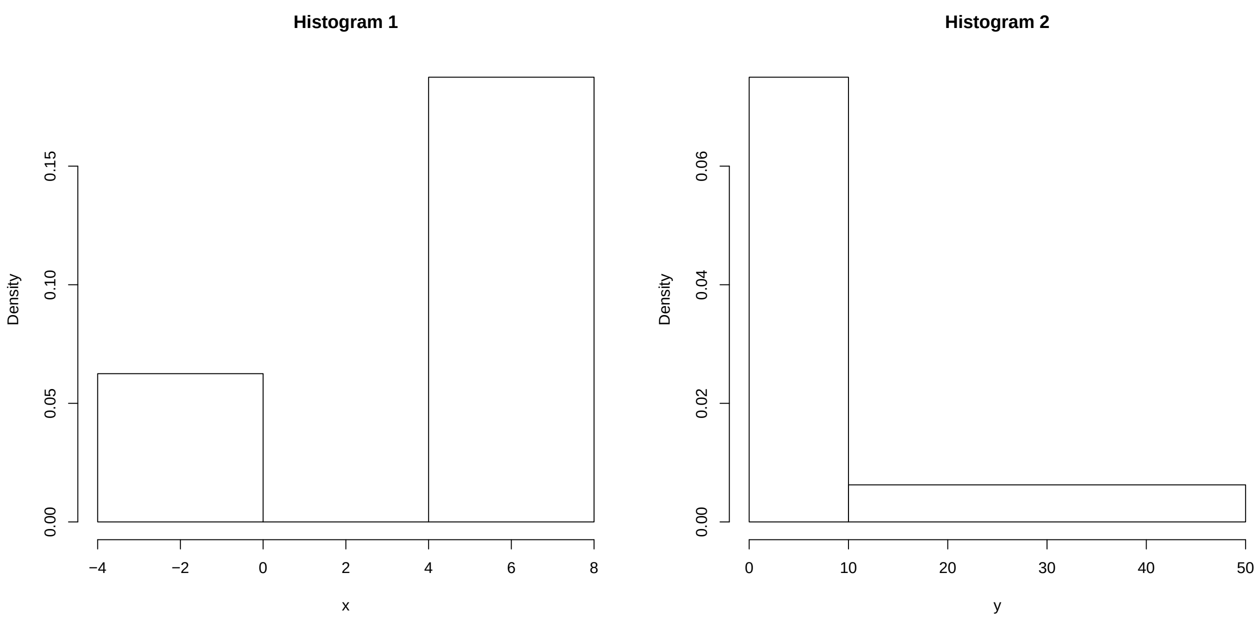
a) {80, 100, 120}

b) {80, 80, 140}

c) {0, 50, 150}

d) {60, 120, 120}

**Question 12.** Can the data sets $x$ and $y$ used to make the following histograms be identical? Which of these answers are correct?




a) No, because the right one is calculated from positive data only.

b) Yes, the right one includes all possible data from which the left one may be calculated.

c) No, the right one must be calculated with at least one value greater than 8.

d) No, the left one can not have been calculated with a value of 10 or more.

**Question 13.** Calculate estimates of the standard deviations $s_x$, $s_y$ of the samples $x = (5, 9, 7)$ and $y = (-1, 2, 5)$ as well as the Pearson coefficient of correlation $r_{xy}$ of $x$ and $y$. Which of the following answers are correct?

a) $s_x = 4$, $s_y = 9$

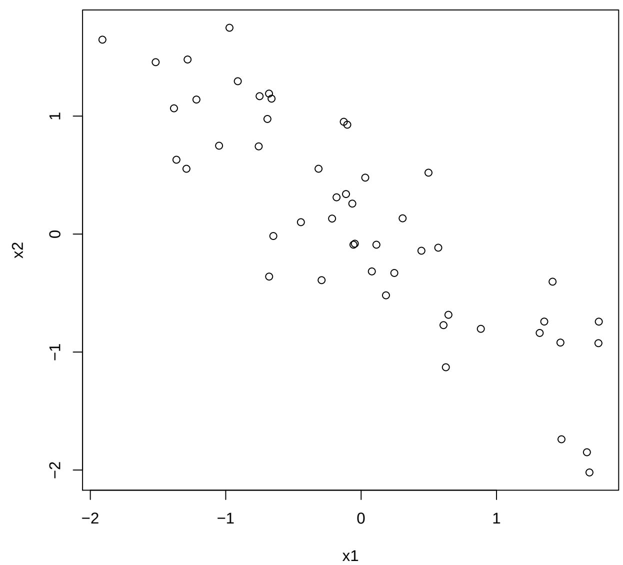
b) $r_{xy} = 0$

c) $s_x = 2$, $s_y = 3$

d) $r_{xy} = \frac{1}{2}$

e) $r_{xy} = \frac{1}{4}$

**Question 14.** Consider the following scatter plot.




The coefficient of correlation of the two variables ...

a) is negative.

b) is positive.

c) should have an absolute value greater than 0.4.

d) should be close to zero.

**Question 15.** Let the coefficient of correlation of two variables $X$ and $Y$ be larger than zero. What will be the effect on it, if the data of $X$ are multiplied by the factor of 2?

a) The effect depends on the data of $X$.

b) It depends on $Y$.

c) The coefficient will be doubled.

d) It will be increased fourfold.

## 2.2 Probability

**Question 16.** There are 8 socks in your drawer: 4 black and 4 red. You take 3 of them with you in the dark. Which statements are correct?

a) It is sure that you get at least two socks (a pair) of the same colour.

b) It is sure that you get a pair of reds.

c) The probability to get 3 of the same colour is $\frac{1}{8}$.

d) The probability to get 3 of the same colour is $\frac{1}{7}$.

**Question 17.** In the sports injuries unit of a hospital, 40% of the patients are rugby players, 20% are swimmers and the remaining 40% play soccer. For a rugby player, the probability to be released on the first day is 10%; for a swimmer, it is 20%; for a soccer player, it is 80%. Which of the following statements are correct?

a) 40% of all patients are released on the first day.

b) Given a patient is released on the first day, the probability of her/him being a soccer player is 80%.

c) 80% of the non-swimmers have to stay for more than one day.

**Question 18.** Let $X$ be a random variable with probability density function

$$f(x) = \begin{cases} \frac{1}{9}x^2, & x \in [0,3], \\ 0, & \text{else.} \end{cases}$$

Which of the following statements are correct?

a) The expected value of $X$ is $\frac{9}{4}$.

b) The probability of $X < 1$ is $\frac{1}{27}$.

c) The probability of $X \in [0, 0.5]$ is $\frac{1}{54}$.

d) The probability of $X = 1$ is zero.

### 2.3 Inference and Linear Models

**Question 19.** Let $X$ be a random variable defined by the density function

$$f(x) = \begin{cases} \dfrac{\alpha\beta^\alpha}{x^{\alpha+1}} & \text{, if } x \geq \beta \\ 0 & \text{, else} \end{cases}$$

with parameters $\alpha > 0$ and $\beta > 0$. We observe a sample $\{3, 4, 8\}$. Which of the following statements are correct?

a) The expected value of $X$ exists for all combinations of $\alpha$ and $\beta$.

b) The expected value does only depend on $\alpha$, but not on $\beta$.

c) Given the sample, $\beta$ can not be larger than 3.

d) If we assume $\beta = 2$, the estimation of the expected value from the sample mean leads to an estimate of $\dfrac{5}{3}$ for $\alpha$.

e) $\dfrac{5}{3}$ is also the maximum likelihood estimate of $\alpha$ in this case.

**Question 20.** We are interested in significant differences (level $\alpha = 0.05$) between the expected values $\mu_1$ and $\mu_2$ of two populations. Which of the following statements on statistical tests are correct?

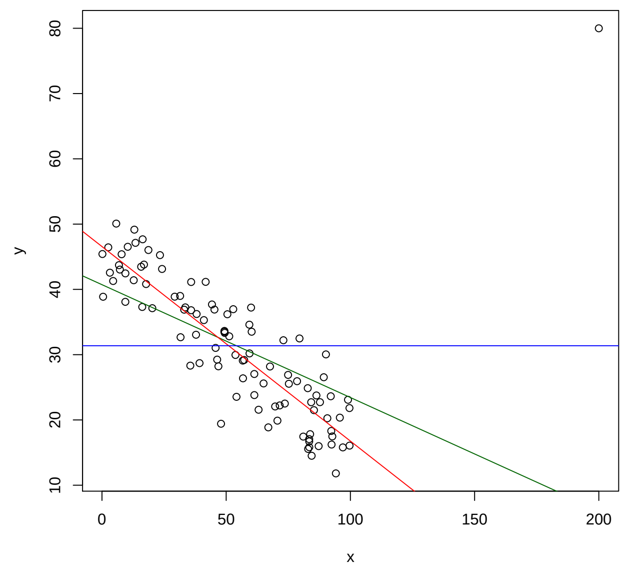
a) We will formulate the null hypothesis as $\mu_1 = \mu_2$.

b) A t-test can always be applied in this situation.

c) A $p$-value is the probability that the null hypothesis is correct, given the observed data.

d) If we obtain a $p$-value of 0.04, we will reject (level $\alpha = 0.05$) the null hypothesis.

**Question 21.** One of the lines in the following scatter plot is the regression line fitted to the data. Which of the statements are correct?




a) The red and green line have the right direction, and, hence, one of them could be the regression line.

b) The blue line seems to represent the mean value of the data with respect to $y$ and thus could be the regression line.

c) The point in the top right corner has a strong influence on the regression line.

d) Leaving aside the point in the corner, the red line seems to fit better to the rest of the data.

**Question 22.** You have performed a linear regression analysis to explore sunflowers' growth (in meters per month) depending on the watering (in litres per day). You have estimated the regression coefficient to be $\hat{\beta} = 1.6$. What can you conclude?

a) There is a significant correlation between watering and growth.

b) An average sunflower grows 1.6 meters per month.

c) If you give it an additional litre of water per day, there will be an additional average growth of 1.6 meters per month.

d) According to the model assumptions, an additional litre of water per day will result in additional 19.2 meters of growth after one year.

e) You should consider further influencing quantities.

## 2.4 R Programming

**Question 23.** Which of the following `R` commands evaluates to `TRUE`?

a) `5 >= 5`

b) `TRUE & FALSE | FALSE & TRUE`

c) `FALSE & FALSE & FALSE | TRUE`

d) `!(((TRUE > FALSE) > TRUE) & !TRUE)`

**Question 24.** Consider the following `R` code chunk:

```
x <- 0
while(x < 4) {
    x <- sample(1:3, 1)
    print(x)
}
```

It is not a good idea to run these lines because...

a) `x` is an invalid argument to `print()`.

b) the condition `x < 4` is never violated.

c) the function `sample()` does not exist.

d) `x` is initialised with the wrong type.

**Question 25.** Which of the following code lines return `TRUE`?

a) `max(c(2, 3, 4, NA, 1, 5)) == NA`

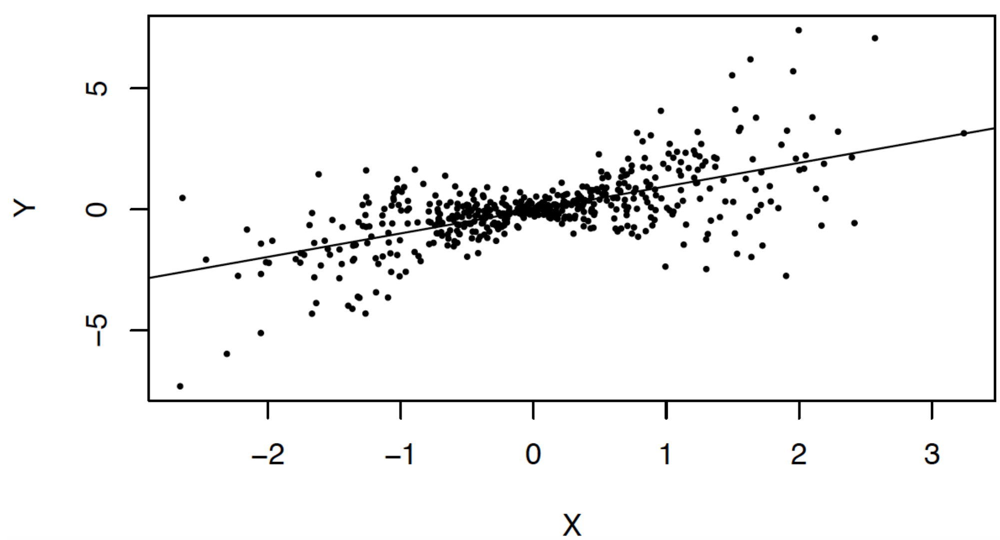
b) `max(c(2, 3, 4, NA, 1, 5), na.rm = TRUE) == 5`

c) `typeof(sum(c(1, 2, 3, 4, NA))) == "double"`

d) `typeof(sum(1:4)) == "integer"`

e) `typeof(sum(c(1L, 2L, 3L, 4L, NA_real_), na.rm = TRUE)) == "integer"`

**Question 26.** Given the data for the vectors `X` and `Y`. Which functions in `R` may have been used to generate the following plot?



a) `lm()`

b) `points()`

c) `abline()`

d) `integrate()`

**Question 27.** Consider the following `R` code chunk and output and note that `NA` appears in the output of `lm()`.

```
X1 <- rnorm(1e2)
X2 <- X1 + 3
Y <- X1 + X2 + rnorm(1e2)
lm(Y ~ X1 + X2)

Call:
lm(formula = Y ~ X1 + X2)

Coefficients:
(Intercept)           X1           X2
      2.979        2.019           NA
```

Which of the following statements are correct?

a) The regressors `X1` and `X2` are perfectly correlated.

b) The estimated model suffers from omitted variable bias.

c) `lm()` excludes `X2` from the regression so that there is a least squares solution.

d) `NA` indicates that the model fit to the data is perfect.

**Question 28.** Consider the function `f()` defined below.

```
f <- function(mod, ...) {
+   if(class(mod) == 'lm') {
+     sqrt(
+       diag(
+         sandwich::vcovHC(mod, ...)
+       )
+     )
+   } else {
+     warning('input not an object of class lm')
+   }
+ }
```

Which of the following statements are correct?

a) `f()` returns standard errors for the linear model object `mod`.

b) `f()` reports heteroscedasticity robust standard errors by default.

c) `f()` accepts additional arguments to `vcovHC()`.

d) The `sandwich` package is required for running `f()`.

# 3 Econometrics

## 3.1 Simple linear regression

**Question 29.** Suppose you have data on the individual wages and working locations of employees, as well as population data on the working locations. You wish to learn whether moving to a larger city would result in higher earnings for an average worker. From previous research, it is known that more skilled workers are more likely to work in larger cities. You now estimate the regression model

$$Y_{i,c} = \beta_0 + \beta_1 pop_c + u_{i,c},$$

where $Y_{i,c}$ is the wage of employee $i$ in city $c$ and $pop_c$ is the population in city $c$.
Which statements about the OLS estimator of $\beta_1$ are true?

a) $\mathbb{E}(\widehat{\beta}_1) > \beta_1$.

b) $\mathbb{E}(\widehat{\beta}_1) < \beta_1$.

c) $\mathbb{E}(\widehat{\beta}_1) = \beta_1$ and $\widehat{\beta}_1$ is efficient.

d) $\widehat{\beta}_1$ is BLUE.

**Question 30.** Consider the simple regression model

$$Y_i = \beta_0 + \beta_1 X_i + u_i.$$

Which statements about omitted variable bias (OVB) are true?

a) OVB will always be present as long as $R^2 \neq 1$.

b) OVB exists if an omitted variable $Z$ is correlated with $X$ but is not a determinant of $Y$.

c) OVB exists if the data come from a randomized controlled experiment.

d) OVB exists if the omitted variable $Z$ is correlated with $X$ and is a determinant of $Y$.

**Question 31.** Consider the regression model

$$\log Y_i = \beta_0 + \beta_1 \log X_i + u_i.$$

Which of the following statements about the interpretation of $\beta_1$ are (approximately) correct and which are not?

a) If the logarithmic $X$ value increases by one unit, the logarithmic $Y$ value increases by $\beta_1$ units.

b) If $X$ increases by 1%, then $Y$ increases by $\beta_1$%.

c) $\beta_1$ indicates a percentage change in $Y$ and therefore lies between 0 and 1.

d) If the logarithmic $X$ value increases by 1%, the logarithmic $Y$ value increases by $\beta_1$%.

**Question 32.** Consider the regression model

$$Y_i = \beta_0 + \beta_1 X_i + u_i.$$

Assume that the units of measurement are changed for both $X$ and $Y$. The new variables are $Y^* = aY$ and $X^* = bX$.
Denote by $\widehat{\beta}_1$ the OLS slope estimate from a regression of $Y$ on $X$ and $\widehat{\beta}_1^*$ the estimate from regression of $Y^*$ on $X^*$. Which statements about the effect of this change in measurements are true?

a) $\widehat{\beta}_1^* = a + b\widehat{\beta}_1$.

b) $X^*$ is now an endogeneous regressor

c) $\widehat{\beta}_1^* = \frac{a}{b}\widehat{\beta}_1$.

d) $\widehat{\beta}_1^* = \frac{a^2}{b^2}\widehat{\beta}_1$.

**Question 33.** Which of the following statements about the 95% confidence interval for $\beta_1$, the unknown slope coefficient in a simple regression model, are true?

a) A 95% confidence interval is a random interval that contains the true parameter $\beta_1$ in 95% of all cases.

b) The true parameter $\beta_1$ is a random variable that lies within the 95% confidence interval with 95% probability.

c) You perform a linear regression and the 95% confidence interval for $\beta_1$ is $[0.4, 0.5]$. Then, $P[0.4 < \beta_1 < 0.5] = 95\%$.

d) Assuming the true parameter is 0.45, you draw a sample and construct a 95% confidence interval. Ex ante, that is before you draw your sample, you can safely say that this interval will contain the true value with 95% probability.

## 3.2 Principle of Econometrics

**Question 34.** A colleague wants to determine the effect of education on wage using data on $n$ employees. He tells you that this can be done with the famous Mincer regression model. In it's simplest form this model is given by

$$\log(w_i) = \beta_0 + \beta_1 edu_i + \beta_2 exp_i + \beta_3 exp_i^2 + u_i, \quad i = 1, \ldots, n,$$

where $w_i$ is the wage, $edu_i$ is the total years of schooling, and $exp_i$ measures the years of potential work experience by subtracting $edu_i$ from the individual's age $age_i$, that is

$$exp_i = age_i - edu_i.$$

Which of the following statements are true?

a) Additional factors such as age, gender, ethnicity, employment sector, and characteristics of the local economy should **all** be included to ensure that our model does not suffer from the omitted variable bias invalidating the least squares estimate.

b) We should add as many additional explanatory factors as possible to mitigate the possibility of the omitted variable bias. However, we need to make sure that the regressors are not perfectly multicollinear. Thus adding $age_i$ for instance is not admissible.

c) Unbiased estimation of the effect of education in Mincer equations by least squares is near impossible as there are certain unobserved factors such as individual ability that affect wages and schooling simultaneously.

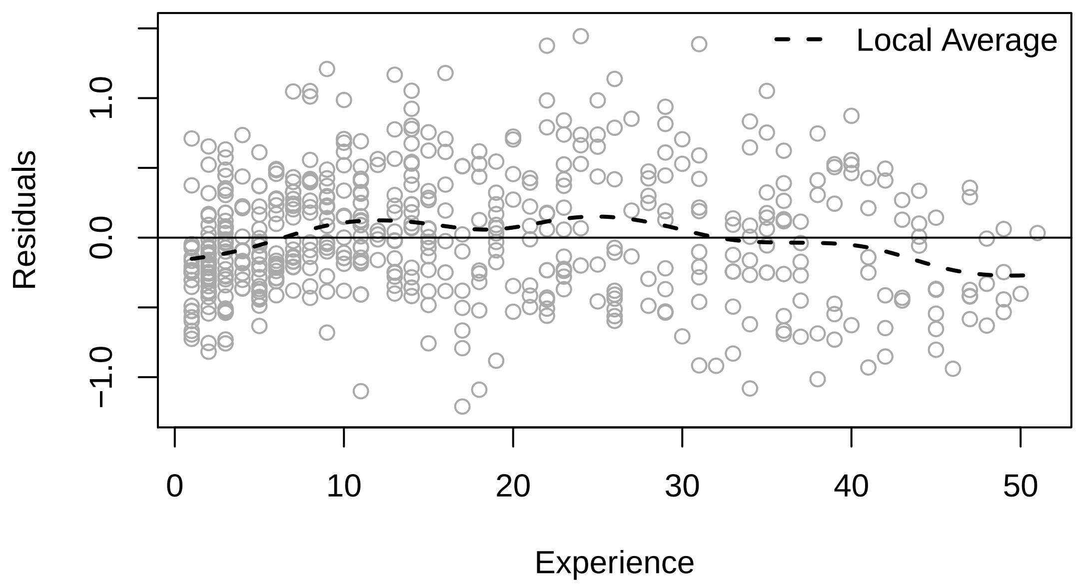
d) In order to unbiasedly estimate the effect of education on wages in large samples, we need to resort to procedures such as instrumental variable regression that allow for unobserved factors that are correlated with education.

e) Even though $\log(w_i) = 1$ when $w_i = 0$ we should be careful of using unemployed individuals in the Mincer regresssion model as unemployed and employed individuals may differ in many unobserved determinants of wage.

**Question 35.** You are asked to determine the effect of education on wage. You are provided with data related to wages ($w$), years of education ($edu$), and years of previous work experience ($exp$) for $n$ individuals. You estimate the linear regression model

$$w_i = \beta_0 + \beta_1 edu_i + \beta_2 exp_i + u_i, \quad i = 1, \ldots, n$$

using the least squares estimator. Subsequently, you wish to investigate the validity of the standard least squares assumptions. Whilst doing so you create the following diagnostic plot, which plots the residuals of the fit against the regressor $exp$ along with a local average of the residuals.




Which of the following statements are true?

a) By construction the sum of the residuals is zero.

b) The plot indicates that there may be unexplained structure in the residuals.

c) Any unexplained structure in the residuals with respect to $exp_i$ will not bias the estimator for $\beta_1$, the effect of $edu_i$.

d) The diagnostic plot supports the inclusion of $exp_i^2$ as an additional regressor.

e) As $edu_i$ and $exp_i^2$ are probably correlated, the estimator for $\beta_1$, the effect of education, will not be unbiased.

**Question 36.** Suppose you are in a situation with stochastic regressors. In particular, assume you have data on $n$ individual units. For each unit you observe the vector $(y_i, x_i, w_i) \in \mathbb{R}^3$, where $y_i$ is the dependent variable and $(x_i, w_i)$ are the two stochastic regressors. The vectors $(y_i, x_i, w_i)_{i=1}^n$ are i.i.d. meaning that the vector for all individuals $i$ and $j$ follow the same distribution, i.e. they are subject to the same data generating mechanism and are independent of each other. The data generating process is given by

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 w_i + u_i$$

and all the standard least squares assumptions with stochastic regressors hold in this model.
Suppose that instead of estimating the correct model you erroneously estimate the "shorter" model

$$y_i = \beta_0 + \beta_1 x_i + v_i$$

by least squares.
Which of the following statements are true?

a) The resulting estimator for $\beta_1$ is always unbiased.

b) The resulting estimator for $\beta_1$ is unbiased if $\beta_2 = 0$.

c) The resulting estimator for $\beta_1$ is unbiased if the regressor $x_i$ and the error in the shorter model $v_i$ are uncorrelated.

d) The resulting estimator for $\beta_1$ is **only** unbiased if $\beta_2 = 0$ and the regressor $x_i$ and the error in the shorter model $v_i$ are uncorrelated.

**Question 37.** Consider the linear regression model

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + u_i, \quad i = 1, \ldots, n$$

with $\mathbb{E}[u_i] = 0$ for all $i = 1, \ldots, n$. The model equation can also be written as

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + u_i, \quad i = 1, \ldots, n$$

with $\boldsymbol{\beta} := (\beta_0, \ldots, \beta_p)'$ and $\mathbf{x}_i' := (1, x_{1i}, \ldots, x_{pi})$ or in the matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{u},$$

with

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_n' \end{pmatrix} = \begin{pmatrix} 1 & x_{1i} & \cdots & x_{pi} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{pn} \end{pmatrix} \text{ and } \mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}.$$

The least squares estimator $\hat{\boldsymbol{\beta}}$ is defined as the minimizer of the sum of squared distances

$$S(\mathbf{b}) := \sum_{i=1}^{n} (y_i - \mathbf{x}_i'\mathbf{b})^2 = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) = ||\mathbf{y} - \mathbf{X}\mathbf{b}||_2^2,$$

i.e. $\hat{\boldsymbol{\beta}} = \arg\min_{\mathbf{b} \in \mathbb{R}^{p+1}} S(\mathbf{b})$. Assuming that the regressors are not perfectly multicollinear, which of the following statements are true?

a) $\hat{\boldsymbol{\beta}} = (\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i')^{-1} \sum_{i=1}^{n} \mathbf{x}_i y_i$.

b) $\hat{\boldsymbol{\beta}} = \sum_{i=1}^{n} \mathbf{x}_i y_i (\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i')^{-1}$.

c) $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i')^{-1} \sum_{i=1}^{n} \mathbf{x}_i u_i$.

d) $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, where $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$ and $\mathbf{y} = (y_1, \ldots, y_n)'$.

e) If the regressors are perfectly multicollinear then $\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' = \mathbf{X}'\mathbf{X}$ is not invertible and there is **no unique solution** to $(\mathbf{X}'\mathbf{X})\mathbf{b} = \mathbf{X}'\mathbf{Y}$, the so called normal equations.

**Question 38.** Assume that you have data that was generated by

$$y_i = \beta_0 + \beta_1 x_i + u_i, \quad i = 1, \ldots, n,$$

and all the standard assumptions for least squares estimation under stochastic regressors are fulfilled. Which of the following statements are true?

a) If the dependent variable $y_i$ suffers from additive random nonsystematic measurement error this leads to the least squares estimator being biased.

b) Measurement error in the regressor $x_i$ does not result in biasedness of the least squares estimator.

c) Random i.i.d. measurement error in the dependent variable $y_i$ always necessitates the use of heteroscedastic robust standard errors.

**Question 39.** You are given the task to estimate the production function of a factory. You are given data on the output $Q$ measured in units produced as well as on the two main production inputs $(K, L)$ measured in input units. The manager tells you that both production factors are important, so that a sensible data generating process may be assumed to be of the form

$$Q_i = f(K_i, L_i) + u_i, \quad i = 1, \ldots, n$$

with $u_i, i = 1, \ldots, n$ an error term capturing the random fluctuation in production output. You consult your micro lecture notes and find the following production functions with two inputs.

(1) Cobb-Douglas: $Q = \beta_0 K^{\beta_1} L^{\beta_2}$

(2) Linear: $Q = \beta_0 + \beta_1 K + \beta_2 L$

(3) Quadratic: $Q = \beta_0 + \beta_1 K^2 + \beta_2 K + \beta_3 L^2 + \beta_4 L$

(4) Leontief: $Q = \min\{\beta_1 K, \beta_2 L\}$

(5) Log-linear: $\log(Q) = \beta_0 + \beta_1 K + \beta_2 L$

(6) Linear-log: $Q = \beta_0 + \beta_1 \log(K) + \beta_2 \log(L)$

Which of the following statements are true?

a) Only the linear production function (2) leads to a data generating process that can be modelled by a linear regression model.

b) The Cobb-Douglas (1) and Log-Linear (5) production functions can both be used as the basis for a linear regression model with dependent variable $\log(Q_i)$.

c) The exponents in the Cobb-Couglas function (1) can be interpreted as elasticites. For instance $\beta_1$ provides the percentage change in output given a percentage in the input $K$.

d) Only the linear production technology (2) results in a ceteris paribus change of the dependent variable ($Q$ or $\log(Q)$) following a change in an input that does not depend on the level of the input.

e) The marginal effect of changing an input in the quadratic production (3) is linear in the corresponding input level.

f) The coefficients in the Log-linear (5) and Linear-log (6) specifications can be interpreted as semi-elasticites: A percentage change in the logarithmic variable leads to or is caused by ceteris paribus change in the nonlogarithmic variable.

g) The input factors in the Leontief production technology (4) are always used in a fixed proportion.

**Question 40.** You are given data $(y_i, x_i)_{i=1,\dots,n}$ on $n$ units, where the deterministic regressor $x_i$ does not take the same value for all individuals. The data have been generated by the process

$$y_i = \beta_0 + \beta_1 x_i + u_i = \mathbf{x}_i' \boldsymbol{\beta} + u_i, \tag{1}$$

where the errors $u_i$ are mean zero, uncorrelated and homoscedastic. That is for $\mathbf{u} = (u_1, \dots, u_n)'$ we have $\mathbb{E}[\mathbf{u}] = \mathbf{0}$ and $\mathbb{E}[\mathbf{u}\mathbf{u}'] = \sigma^2 \mathbf{I}_n$ for some constant $\sigma^2 \in (0, \infty)$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix.
The Gauss Markov Theorem states that the least squares estimator $\hat{\boldsymbol{\beta}}$ in this setting is the **B**est **l**inear **u**nbiased **e**stimator (**BLUE**). Which of the following statements are true?

a) In addition to the assumptions on the error vector $\mathbf{u}$ stated above, the Gauss-Markov Theorem requires that the Gram matrix $\mathbf{X}'\mathbf{X}$ with $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$ is invertible.

b) Invertibility of the Gram matrix $\mathbf{X}'\mathbf{X}$ is implicitly guaranteed in the setting of the question.

c) Invertibility of the Gram matrix $\mathbf{X}'\mathbf{X}$ is equivalent to the regressors not being perfectly multi-collinear.

d) The term *linear* in the expression **BLUE** refers to any estimator of the parameters in the linear model.

e) Informally the **BLUE** property of the least squares estimator states that of all the possible linear unbiased estimators the least squares estimator is the one with the "smallest" variance-covariance matrix.

f) We can use the Gauss-Markov Theorem to compare the least squares estimator to the scaled least squares estimator $\tilde{\boldsymbol{\beta}} = a\hat{\boldsymbol{\beta}}$ for some $0 < a < 1$.

g) We can use the Gauss-Markov Theorem to compare the least squares estimator to the shifted least squares estimator $\check{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} + a$ for some $a \neq 0$.

h) If the form of heteroscedasticity is known, we can appeal to the Gauss-Markov Theorem to show that the generalized least squares estimator is BLUE.

**Question 41.** For $n = 145$ electricity plants you are given annual data on the total cost $TC_i$ and the output produced $Q_i$. Furthermore, you have data on prices of the three main production factors: labour $L$ with a price/wage rate of $p_i^L$, capital $C$ with a rental price of $p_i^C$ and fuel $F$ with a price of $p_i^F$. Based on the assumption of a Cobb-Douglas production function it is possible to show that the cost function is also Cobb-Douglas. Thus, you decide to estimate a model for the transformed cost function, for instance

$$\log(TC_i/p_i^C) = \beta_0 + \beta_1 \log(Q_i) + \beta_2 \log(p_i^L/p_i^C) + \beta_3 \log(p_i^F/p_i^C) + u_i. \tag{2}$$

Let $\hat{u}_i, i = 1, \dots, 145$ denote the residuals.
You also decide to estimate an extension of the above model, which can also be interpreted as a restricted translog cost function. The second model you estimate is given by

$$\log(TC_i/p_i^C) = \gamma_0 + \gamma_1 \log(Q_i) + \gamma_2 (\log(Q_i))^2 + \gamma_3 \log(p_i^L/p_i^C) + \gamma_4 \log(p_i^F/p_i^C) + v_i. \tag{3}$$

The residuals are denoted by $\hat{v}_i, i = 1, \dots, n$. Plots of the residuals versus the log of output for both models are given in Figure 1.
Which of the following statements are true?

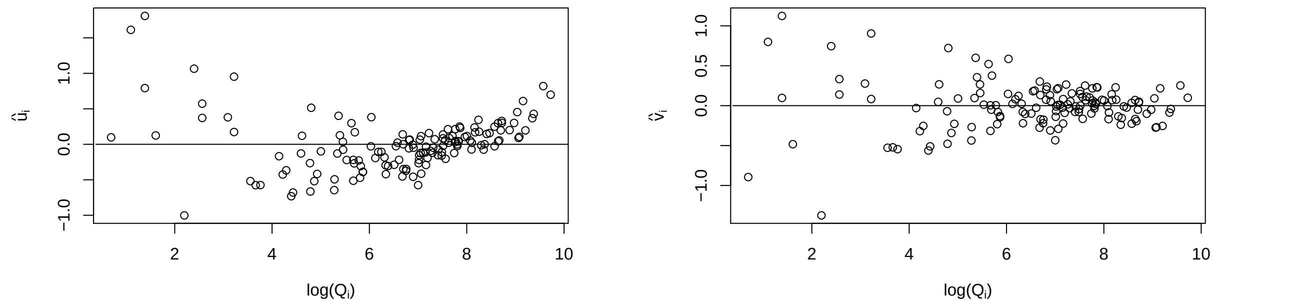



Figure 1: Plot of the residuals against $\log(Q_i)$. Left: Residuals $\hat{u}_i$ from model in (2). Right: Residuals $\hat{v}_i$ from model in (3).

a) The quadratic structure of the residuals $\hat{u}_i$ with respect to $\log(Q_i)$ suggests that $(\log(Q_i))^2$ should be added as an additional regressor in (2).

b) The variability in the residuals $\hat{v}_i$ depends on the value of $\log(Q_i)$, thus it seems inadequate to assume homoscedastic errors $v_i$ in (3).

c) The presence of heteroscedasticity changes the least squares point estimates.

d) Under the presence of heteroscedasticity we cannot use classical standard errors to perform inference as they are inconsistent.

e) In the presence of heteroscedasticity (or autocorrelation) invalidity of standard parameter significance tests can be fixed in large samples by using an appropriate estimator for the standard errors.

f) The right hand panel in Figure 1 suggests modelling the variance of the errors $v_i$ as a decreasing function of $\log(Q_i)$.

**Question 42.** The linear regression model with $k$ regressors in matrix form is given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{u},$$

with $\mathbf{X}$ the $n \times k$ regressor matrix. Assume that $\mathbf{X}'\mathbf{X}$ is a $(k \times k)$ invertible matrix, $\mathbb{E}[\mathbf{u}] = \mathbf{0}$ and $\mathbb{E}[\mathbf{u}\mathbf{u}'] = \boldsymbol{\Sigma}$ is a positiv definit matrix. In this model define the following matrices

$$\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \quad \text{ and } \mathbf{M} = \mathbf{I}_n - \mathbf{P}$$

Which of the following statements are true?

a) $\mathbf{P}$ often called the *hat matrix* can be used to generate the vector of fitted values as $\hat{\mathbf{y}} = \mathbf{P}\mathbf{y}$.

b) $\mathbf{M}$ often called the *residual maker matrix* can be used to generate the vector of residuals as $\hat{\mathbf{u}} = \mathbf{M}\mathbf{y}$.

c) The matrices $\mathbf{P}$ and $\mathbf{M}$ are orthogonal projection matrices.

d) The matrix $\mathbf{P}$ is symmetric and idempotent, i.e. it holds that $\mathbf{P} = \mathbf{P}'$ and $\mathbf{P}^2 = \mathbf{P}\mathbf{P} = \mathbf{P}$.

e) The decomposition $\mathbf{y} = \mathbf{P}\mathbf{y} + \mathbf{M}\mathbf{y}$ corresponds to an orthogonal decomposition that is $\mathbf{y}'\mathbf{P}'\mathbf{M}\mathbf{y} = 0$.

f) The estimator $\hat{\boldsymbol{\Sigma}} = \frac{1}{n-k}\hat{\mathbf{u}}\hat{\mathbf{u}}'$, with $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ the residual vector, is a good estimator for the variance covariance matrix $\boldsymbol{\Sigma}$.

g) In the case of homoscedasticity $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_n$, the error variance $\sigma^2$ can be unbiasedly estimated by $\hat{\sigma}^2 = \frac{1}{n-k}\hat{\mathbf{u}}'\hat{\mathbf{u}}$, where $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ is the residual vector.

### 3.3 Time Series Analysis

**Question 43.** Which of the following are typical characteristics of financial asset return time-series?

a) Distributions of asset returns are typically thin-tailed.

b) Asset return time series are not weakly stationary.

c) Asset return time series are highly correlated.

d) Asset return time series have no trend.

**Question 44.** Financial market data are often characterized by so-called volatility clusters, i.e. large absolute returns are often followed by large absolute returns and small absolute returns are often followed by small absolute returns.
Denote $R_t$ the return of some asset in period $t$. Which of the following assumptions are likely violated in the simple AR(1) model $R_t = \beta_1 R_{t-1} + u_t$?

a) $\mathbb{E}(u_t|R_{t-1}) = 0$.

b) $\beta_1 > u_t$.

c) Conditional homoskedasticity of the $u_t$.

d) $\mathbb{E}(R_t^4) < \infty$.

**Question 45.** Consider the model

$$Y_t = \beta Y_{t-1} + u_t, \quad t = 1, \ldots, T.$$

with $|\beta| < 1$ and $\mathbb{E}(u_t^4) < \infty$. A fellow student states that *all* of the following assumptions are required for the OLS estimator of $\beta$ to be consistent. Which of the following statements are true?

a) $\mathbb{E}(u_t|Y_{t-1}) = 0$.

b) $\text{Var}(u_t) = \sigma^2 \;\; \forall \;\; i = 1, \ldots, n$.

c) The $u_t$ are independent.

d) $u_t \sim \mathcal{N}(0, \sigma^2)$.

**Question 46.** Consider the following time series model with i.i.d. disturbances $u_t$ having zero mean and unit variance

$$Y_t = 0.2 + 0.4Y_{t-1} + u_t$$

. Which of the following statements are true?

a) The unconditional mean of $Y_t$ is $1/3$.

b) $Y_t$ is an autoregressive process of order 1.

c) $\mathbb{E}(Y_t|Y_{t-1}) = 0.4Y_{t-1}$.

d) The unconditional variance of $Y_t$ is $\approx 1.19$.

**Question 47.** Which of the following conditions are necessary for a stochastic process $Y_t$ to be classifiable as a weakly stationary process?

a) The mean must be constant.

b) The variance of the process must constant.

c) $Y_t$ must have constant autocovariances.

d) The probability distribution of $Y_t$ must be the same for every $t$.

**Question 48.** Consider the MA(3) process

$$Y_t = \beta_0 + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3}$$

where $\varepsilon_t$ is a white noise process with variance $\sigma^2$. Which of the following statements are true?

a) The process $Y_t$ has a zero mean.

b) The autocorrelation function is zero at lag 5.

c) The process $Y_t$ has variance $\sigma^2$.

d) The autocorrelation function of $Y_t$ will have a value of one at lag 0.

**Question 49.** Consider the autoregressive process

$$Y_t = 2Y_{t-1} + u_t$$

with characteristic equation $1 - 2z = 0$. Which of the following statements are true?

a) $Y_t$ is a non-stationary process because all roots of the characteristic equation lie inside the unit circle.

b) $Y_t$ is a stationary process because all roots of the characteristic equation lie on the unit circle.

c) $Y_t$ is a stationary process because all roots of the characteristic equation lie outside the unit circle.

d) $Y_t$ is a stationary process because all roots of the characteristic equation are less than one in absolute value.

# 4 Economics

## 4.1 Microeconomics

**Question 50.** A household can choose between goods A and B. It has an income of 100 Euro. Good A costs 5 Euro and good B costs 10 Euro. Its utility-function is given by $A^{0.5}B^{0.5}$. What statements hold for the optimal consumption bundle?

a) For the optimal consumption bundle it holds that $A = 5$.

b) For the optimal consumption bundle it holds that $A = 10$.

c) For the optimal consumption bundle it holds that $B = 5$.

d) For the optimal consumption bundle it holds that $B = 10$.

**Question 51.** In general which statements are true about the optimal consumption bundle?

a) The marginal utility per dollar spent must be the same for all goods and services in the consumption bundle.

b) The optimal consumption bundle is the consumption bundle that maximizes a consumer total utility given his or her budget constraint.

c) The optimal consumption bundle is the consumption bundle that maximizes a consumers total utility independent on his or her budget constraint.

d) As long as the consumer spends all of his income he maximizes his utility.

**Question 52.** Demand and supply are given by

$$\text{Demand} = -1.5p + 10$$
$$\text{Supply} = 0.5p + 4$$

What statements hold for the equilibrium price and quantity?

a) The equilibrium price is $p = 3$.

b) The equilibrium price is $p = 3.5$.

c) The equilibrium quantity is $q = 5$.

d) The equilibrium quantity is $q = 5.5$.

**Question 53.** As in the previous question let demand and supply be given by

$$\text{Demand} = -1.5p + 10$$
$$\text{Supply} = 0.5p + 4$$

Would a price of 6 Euro clear the market?

a) No, because this price is the equilibrium price.

b) Yes, because this price is below the equilibrium price.

c) Yes, because this price is the equilibrium price.

d) No, because this price is above the equilibrium price.

**Question 54.** Two members of a criminal gang are arrested and imprisoned. Each prisoner is in solitary confinement with no means of communicating with the other. The prosecutors lack sufficient evidence to convict the pair on the principal charge, but they have enough to convict both on a lesser charge. Simultaneously, the prosecutors offer each prisoner a bargain. Each prisoner is given the opportunity either to betray the other by testifying that the other committed the crime, or to cooperate with the other by remaining silent. The possible outcomes are:

- If A and B each betray the other, each of them serves two years in prison
- If A betrays B but B remains silent, A will be set free and B will serve three years in prison (and vice versa)
- If A and B both remain silent, both of them will serve only one year in prison (on the lesser charge).

What is the Nash-equilibrium?

a) Both betray each other.

b) Both remain silent.

c) One is silent and the other betrays.

d) There is no Nash-equilibrium.

## 4.2 Macroeconomics

**Question 55.** Which relationship does the Phillips curve describe?

a) A trade off between an expansive monetary policy and unemployment.

b) A trade off between inflation and unemployment.

c) A positive correlation between inflation and unemployment.

d) A trade off between inflation and expansive monetary policy.

**Question 56.** If the Consumer Price Index rises from 101 to 104, what does this imply for the real income of households?

a) The real income increases.

b) The real income decreases.

c) The real income does not change.

d) The nominal income decreases.

**Question 57.** What are the consequences of an increased capital stock in the Solow-Swan Model?

a) It increases production growth in the long run.

b) It decreases production growth in the long run.

c) It increases production growth in the short run.

d) It decreases production growth in the short run.

**Question 58.** What effect does a positive supply shock have in macroeconomics?

a) The price increases and the quantity increases.

b) The price decreases and the quantity increases.

c) The price increases and the quantity decreases.

d) The price decreases and the quantity decreases.