Chair of Behavioral Research Methods
Technical University of Munich

Mock (!) Exam

Empirical Research in Management and Economics (WI000258)

General information:

- Format: For each questions there are 4 possible answer options, and **only one option** is correct.
- This mock exam has 35 questions (**note: the *real* exam will have 55 questions!**).
- **NOTE: In the real exam, only answers entered on the answer sheet (last page) will be graded.**
- Closely follow the instructions regarding the answer sheet given at the end of this document.
- For the real exam, you must use a document-proof pen and are only permitted to use a nonprogrammable pocket calculator. For notes, you can only use blank paper provided by the examiners. You can request additional blank papers.
- During the real exam, any aids containing course content or enabling access to such materials are not permitted (scripts, lecture notes, smartphones, etc.) and must be kept out of reach and sight.
- Please enter your **name** and your **matriculation number** here **and on the answer sheet** (starting with a zero, e.g., 01234567):

Name: _____

Matriculation number: _____

Good luck! ☺

Information regarding the answer sheet (final page)

• Enter your name and signature.

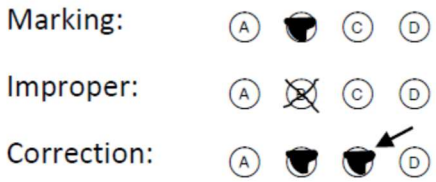
• Enter your student (matriculation) number using the bubble format. Columns designate the digit at the respective position. Fill in the bubble that corresponds to the digit at that position (e.g., the left-most column is the first digit of your student number, hence 0 is pre-filled.)

• Enter your answer for each question using the bubble format. Fill in the bubble that corresponds to your final answer.

• Closely follow the instructions from to the example.

- Completely fill the box that indicates your final answer (no crosses, etc.) to ensure that the answer is registered.
- You receive 0 points if an incorrect box is checked, no box is checked, or more than one box is checked.
- To correct an answer, also fill in the correct bubble and draw a little arrow (or something to the effect) to indicate which bubble should count as an answer.

<u>Example:</u>



Only answers entered on the answer sheet (last page) will be graded. It is not sufficient to indicate your answers in the question boxes below only.

1	**Which of the following statements is false?**	Answer (2 points)
a	In a left-skewed distribution the bigger part of the data is typically located to the right of the arithmetic mean.	
b	In a left-skewed distribution the arithmetic mean is typically smaller than the mode.	
c	In a symmetric distribution the arithmetic mean coincides with the median and mode.	
d	In a left-skewed distribution the bigger part of the data is typically located to the left of the arithmetic mean.	

2	**You collect data on the number of hamburgers sold and the price of hamburgers in 9 different diners between 6pm and 7pm on a Monday evening.**	Answer (2 points)

Diner index	Number of hamburgers sold	Hamburger price [€]
1	10	3.00
2	11	3.50
3	11	2.90
4	15	2.40
5	16	3.10
6	18	2.80
7	19	3.30
8	21	2.20
9	23	1.90

What is the median price of sold hamburgers [€]?

a	2.79	
b	2.70	
c	2.90	
d	2.80	

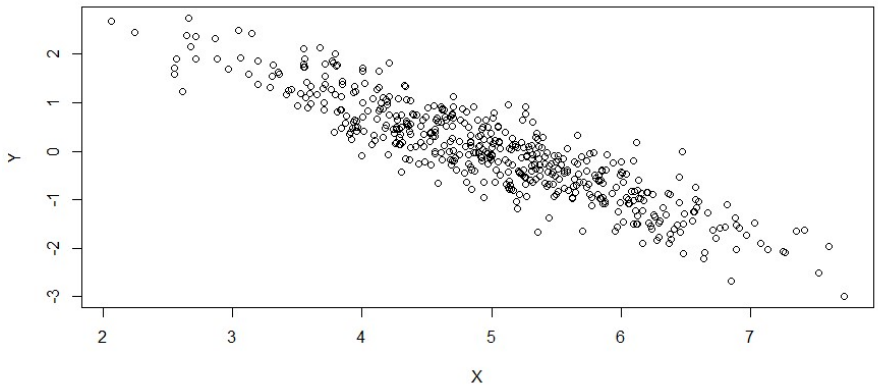
3	**You collect a sample of the maximum temperatures in Munich for one week.** **Please calculate the arithmetic mean temperature and the respective standard deviation (SD) of the week. Tick the correct answer.**	Answer (2 points)

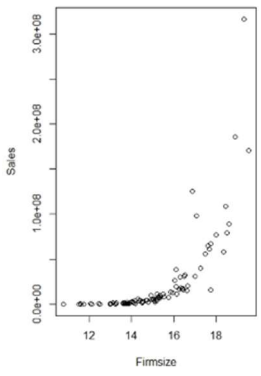
Day	**MO**	**TU**	**WE**	**TH**	**FR**	**SA**	**SU**
Temperature in °C	**23.2**	**27**	**23**	**30**	**28**	**20**	**21**

a	Mean: 25.3 SD: 14.2	
b	Mean: 23.2 SD: 3.5	
c	Mean: 24.6 SD: 3.8	
d	Mean: 24.6 SD:14.2	

5	**Which of the following statements regarding scientific hypotheses is correct? If all statements (a)-(c) are false, please tick (d).**	Answer (2 points)
a	They must be falsifiable.	
b	They must be verifiable.	
c	They should refer to an individual situation or event.	
d	All statements (a)-(c) are false.	

6	**Which of the following statements regarding different types of correlation coefficients is correct? If all three statements (a)-(c) are false, please tick (d).**	Answer (2 points)
a	In the presence of outliers in the data it is preferable to use the Spearman rank correlation.	
b	The Spearman rank correlation coefficient should always be used if your data is on a metric level of measurement.	
c	While the Pearson correlation allows one to test for causality between variables, this is not possible with the Spearman rank correlation.	
d	All statements (a)-(c) are false.	

7	**Look at the scatter plot below displaying Y against X. What statement regarding the relationship between Y and X is correct? If all of the three statements (a)-(c) are false, please tick (d).**  	Answer (2 points)
a	There is likely no correlation between the two variables.	
b	Y and X have a strong negative correlation. Therefore, the Pearson correlation coefficient between the two variables is likely to be < -1	
c	The Pearson correlation coefficient between the two variables is likely positive	
d	All statements (a)-(c) are false.	

8	**Look at the scatter plot below displaying Sales [€] against Firmsize [Million €]. Which statement regarding the relationship between Sales and Firmsize is correct? If all of the three statements (a)-(c) are false, tick (d).**  	Answer (2 points)
a	There is likely no correlation between the two variables.	
b	The Pearson correlation coefficient between the two variables is likely negative.	
c	The Pearson correlation coefficient between the two variables is likely positive.	

d	All statements (a)-(c) are false.	

9	**Which of the following statements regarding a dependent-samples t-test (two-sided) is correct? If all of the three statements (a)-(c) are false, please tick (d).**	Answer (2 points)
a	If the critical t-value is larger than the absolute value of the empirically observed value of the test statistic, you will reject the null hypothesis.	
b	If the absolute value of the empirically observed test statistic is larger than the critical t-value, you will reject the null hypothesis.	
c	If the test statistic is larger than zero, you will always reject the null hypothesis.	
d	All statements (a)-(c) are false.	

10	**You work for a car manufacturer and want to reevaluate the nitrogen oxide emissions of your Diesel cars. For that purpose, you measure the emissions of 500 cars. The mean of the nitrogen oxide emissions in this sample is 157 mg/km with a variance of 25 (mg/km)^2. The goal is to achieve emissions of 150 mg/km. You decide to conduct a t-test to check whether the emissions in your sample are significantly different from your reference value (μ).** **What is the empirically observed value of the test statistic? (Note: t = $\frac{\bar{X}-\mu}{s/\sqrt{n}}$)**	Answer (2 points)
a	31.30	
b	700	
c	6.26	
d	702.13	

11	**You estimate two regression models. The ratio of the explained sum of squares over the total sum of squares is smaller for the first than for the second model. Which statement regarding the relationship of the R^2s of the two models is correct?**	Answer (2 points)
a	The R^2s of the two models are identical.	
b	We cannot make any statement on the relationship of the R^2s based on the given information.	

c	The R^2 of the second model is larger than that of the first model.	
d	The R^2 of the second model is smaller than that of the first model.	

12	**You create a variable Y from a variable X by calculating Y = 0.5*X. What statement regarding the correlation between X and Y is correct?**	Answer (2 points)
a	The Pearson correlation coefficient will be equal to 0.5.	
b	The Pearson correlation coefficient will be equal to -1.	
c	The Spearman rank correlation coefficient will be equal to the Pearson correlation coefficient.	
d	The Spearman rank correlation should be used since this relationship is not linear.	

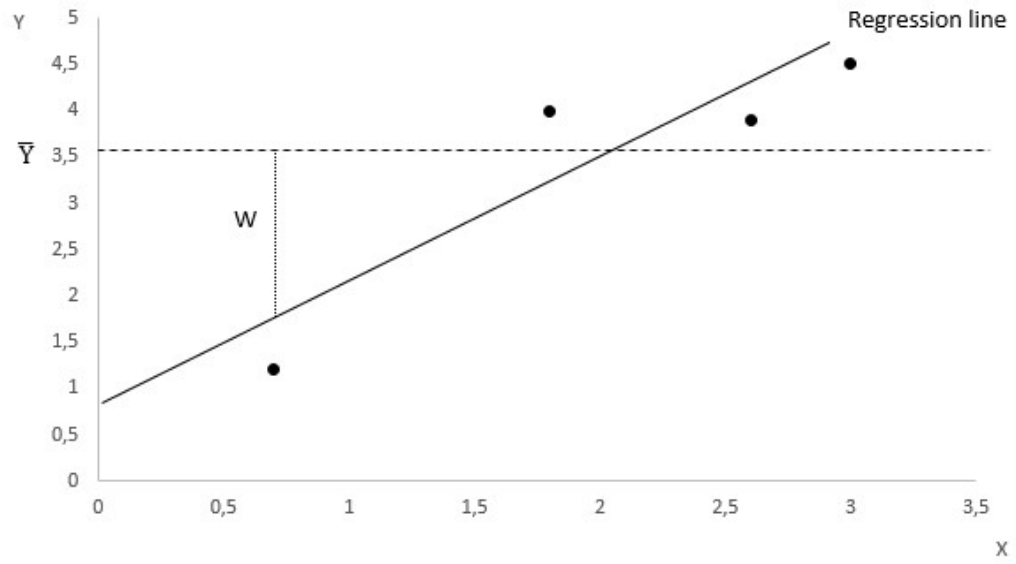
13	**If T is a test statistic and t_{obs} is the empirically observed value of the test statistic, then the p-value for the latter can be expressed as follows:**	Answer (2 points)
a	$P(T \geq 0)$	
b	$P(H_0)$	
c	$P(T \geq t_{obs} \mid H_0)$	
d	$P(T \geq t_{obs} \mid H_1)$	

14	**Suppose a population can be divided into 3 subgroups, which differ substantially in the relevant characteristics and their group size. What will happen if a stratified random sample is collected?**	Answer (2 points)
a	Not all members of the population are equally likely to be sampled.	
b	All members of only one subgroup will be sampled.	
c	All subgroups will be equally represented in the sample.	
d	All members within a subgroup are equally likely to be sampled.	

15	**In a regression model with only one independent variable an R^2 value close to one indicates that …**	Answer (2 points)
a	… the explained sum of squares exceeds the total sum of squares.	
b	… there is an almost perfect fit and the observed data points are close to the estimated regression line.	
c	… the share of the variation in the dependent variable that is explained by the independent variable is very low.	
d	… the sum of squares of the residuals is larger than the explained sum of squares.	

16	**In a multiple linear regression model variables x1 and x2 are used to predict the variable y. You obtain the following descriptive statistics, correlations, and regression outputs. What is the standardized regression weight for x2?** Descriptives N Mean SD SE y 100 1762.647 278.665 27.867 x1 100 999.382 74.230 7.423 x2 100 99.754 30.577 3.058 Pearson's Correlations Variable x1 x2 y 1. x1 Pearson's r — p-value — 2. x2 Pearson's r 0.129 — p-value 0.202 — 3. y Pearson's r 0.556 −0.736 — p-value < .001 < .001 — Coefficients Model Unstandardized Standard Error Standardized t p H_0 (Intercept) 1762.647 27.867 63.253 < .001 H_1 (Intercept) 26.813 63.185 0.424 0.672 x1 2.484 0.063 [redacted] 39.327 < .001 x2 −7.488 0.153 [redacted] −48.827 < .001	Answer (2 points)
1	-0.821	
2	-0.082	
3	-0.736	
4	-0.245	

17	**What is a typical consequence of decreasing the significance level (lower α) for a given statistical test?**	Answer (2 points)
a	An increase in the probability of a Type I error (false positive).	
b	A decrease in the probability of a Type II error (false negative).	
c	A decrease in statistical power.	
d	An increase in effect size.	

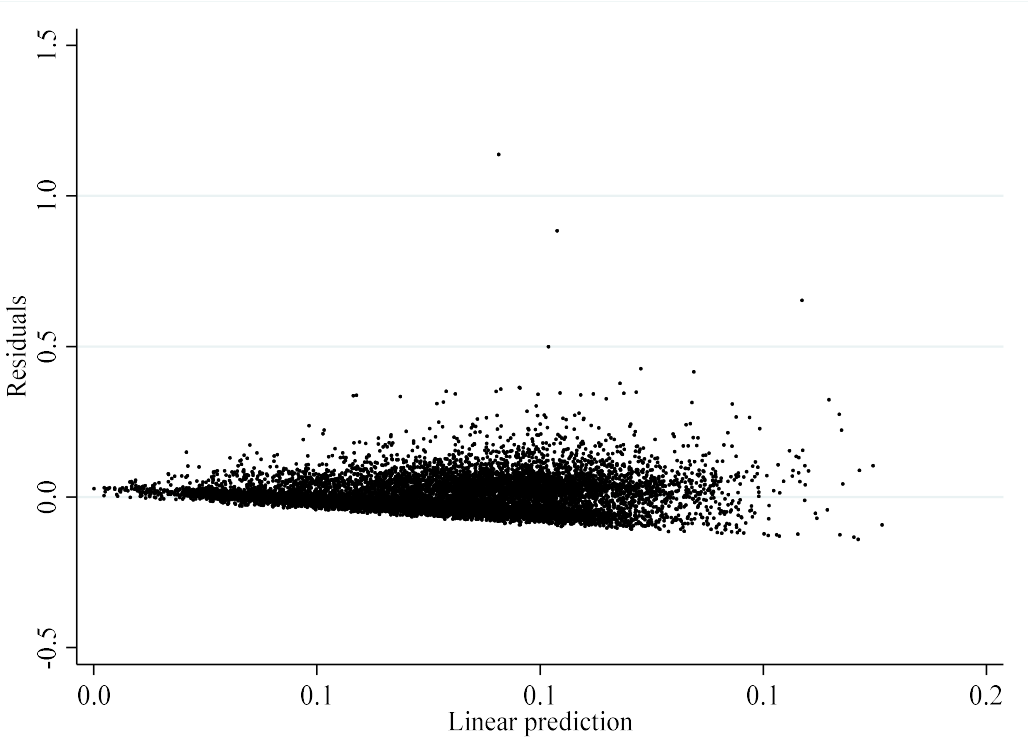
18	**Which of the following statements regarding the figure below is correct (dots represent observed values)? If all statements (a)-(c) are correct, please tick (d).**  	Answer (2 points)
a	The majority of the predicted values is larger than the observed values for the dependent variable.	
b	For x = 3, the predicted value of the regression model exceeds the mean ($\overline{Y}$)	
c	The distance W between the mean ($\overline{Y}$) and the regression line indicates a deviation of Y from its mean that is explained by the model.	
d	All statements (a)-(c) are correct.	

19	**Which statement concerning heteroscedasticity in regression analysis is correct? If all statements (a)-(c) are correct, please tick (d).**	Answer (2 points)
a	Heteroscedasticity implies that the t-test of regression coefficients becomes more reliable.	
b	Homoscedasticity distorts the standard errors of regression coefficients.	
c	Heteroscedasticity means that the variance of the error term is not constant across observations.	
d	All statements (a)-(c) are correct.	

20	**What is the purpose of calculating standardized regression coefficients?**	Answer (2 points)
a	To reduce heteroscedasticity.	
b	To be able to translate the regression results for a factor analysis.	
c	To compare the importance of different independent variables in predicting the dependent variable.	
d	To evaluate the standard errors of the regression coefficients.	

21	**If your regression model is affected by heteroscedasticity, you should …**	Answer (2 points)
a	… use robust standard errors.	
b	… center the dependent variable.	
c	… transform the error term with ungeneralized least squares.	
d	… use unweighted least squares for estimation.	

22	**Which of the following indices for model evaluation takes the number of observations into account?**	Answer (2 points)
a	R^2	
b	BIC	
c	AIC	
d	χ^2	

23	**Based on the figure below, which plots OLS residuals against the predicted values of a regression model, which of the following statements concerning the model can be made?**  	Answer (2 points)
a	The assumption of homoscedasticity is likely fulfilled.	
b	The relationship between the dependent variable and the predictor(s) is strongly non-linear.	
c	The residuals follow a normal distribution.	
d	The model suffers from heteroscedasticity.	

24	**In a multiple regression analysis with three predictors (let's call them income, hsgpa, and sch2), assume that the VIFs are 1.187828, 1.187706, 1.000857 for income, hsgpa, and sch2, respectively. Based on these results, is multicollinearity a problem?**	Answer (2 points)
a	Yes, as for all variables the VIFs are clearly larger than zero.	
b	Yes, as for income, hsgpa and sch2 the VIFs are smaller than 10.	
c	No, as the VIFs are all close to 1.	
d	Yes, as all VIFs exceed the critical threshold of 1.	

25	**In the context of a logistic regression model, where $p_k(y=1) = \frac{1}{1+e^{-z_k}}$ and $z_k = b_0 + \sum_{j=1}^{J} b_j x_{jk}$ (and assuming all *b*s > 0), which of the following statements is correct? If all statements (a)-(c) are false, please tick (d).**	Answer (2 points)
a	If all *x*s decrease simultaneously, the probability of *y* = 1 increases.	
b	For $z_k = 0$, the probability of *y* = 1 is smaller than for $z_k < 0$.	
c	If all independent variables are zero, the probability of *y* = 1 is 0.5.	
d	All statements (a)-(c) are false.	

26	**In a logistic regression model predicting whether individual passengers on the Titanic survived the disaster based on the person's age and sex (1 = male, 0 = female), the following regression coefficients were estimated: $b_0 = 1.306$, $b_{age} = -0.006$, $b_{sex} = -2.466$. What are the odds ratios (OR) for age and sex?**	Answer (2 points)
a	$OR_{age} = .006, OR_{sex} = 2.466$	
b	$OR_{age} = .986, OR_{sex} = 0.003$	
c	$OR_{age} = .994, OR_{sex} = 0.085$	
d	$OR_{age} = -.006, OR_{sex} = -2.466$	

27	**Your company wants to increase the percentage of website visitors who buy a product (online conversion rate). To understand the variables that influence the online conversion rate, you collected data about whether a visitor has purchased a product (*Y*), and three other variables X1-X3 that you suspect might influence the probability of purchase. Based on the results of a logistic regression analysis shown below, what are the odds of purchase vs. no purchase when X1 = X2 = X3 = 1?** Coefficients ▼ Wald Test Estimate | Standard Error | z | Wald Statistic | df | p (Intercept) 0.873 | 0.123 | 7.106 | 50.494 | 1 | < .001 X1 1.984 | 0.139 | 14.233 | 202.592 | 1 | < .001 X2 −0.700 | 0.101 | −6.903 | 47.657 | 1 | < .001 X3 1.619 | 0.200 | 8.088 | 65.414 | 1 | < .001 *Note*. Y level 'Purchase' coded as class 1.	Answer (2 points)
a	43.64	
b	18.23	
c	3.78	
d	.98	

28	**What statement concerning factor analysis is correct? If all statements (a)-(c) are false, please tick (d).**	Answer (2 points)
a	The correlation of the variables bundled in a factor is supposed to be low.	
b	The identified factors are usually only weakly correlated.	
c	Factor analysis aims at identifying groups of objects that share variance.	
d	All statements (a)-(c) are false.	

29	**Below you can see the JASP output of a principle component analysis (PCA) of nine different cognitive ability test scores of seventh- and eighth-grade children from two different schools. Which of the following statements is correct?**	Answer (2 points)
a	The amount of variance explained does not differ between the rotated and the unrotated solution.	
b	The first and the second component of the retained solution are uncorrelated.	
c	Three components were retained and three items load highly on each of them before rotation.	
d	Three components were retained and three items load highly on each of them after varimax rotation.	

Principal Component Analysis

Chi-squared Test

	Value	df	p
Model	201.948	12	< .001

Component Loadings

	RC1	RC2	RC3	Uniqueness
x5	0.930			0.175
x4	0.903			0.186
x6	0.875			0.207
x3		0.814		0.369
x2		0.774		0.455
x1		0.655		0.413
x7			0.876	0.278
x8			0.822	0.308
x9			0.581	0.389

Note. Applied rotation method is promax.

Component Characteristics

	Unrotated solution			Rotated solution		
	Eigenvalue	Proportion var.	Cumulative	SumSq. Loadings	Proportion var.	Cumulative
Component 1	3.216	0.357	0.357	2.502	0.278	0.278
Component 2	1.639	0.182	0.539	1.899	0.211	0.489
Component 3	1.365	0.152	0.691	1.819	0.202	0.691

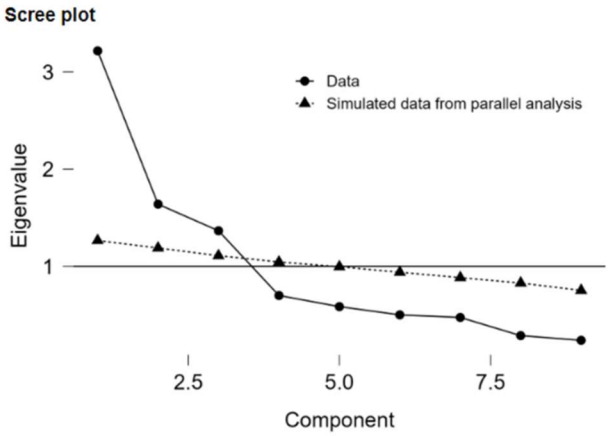
Component Correlations

	Component 1	Component 2	Component 3
Component 1	1.000	0.319	0.218
Component 2	0.319	1.000	0.274
Component 3	0.218	0.274	1.000

30	**In a factor analysis a variable X_1 has a communality of 0.6. Which of the following statements is correct? If all statements (a)-(c) are false, please tick (d).**	Answer (2 points)
a	All variables of the dataset together reflect 40% of the total information contained in X_1.	
b	60% of the information contained in the variable X_1 is lost.	
c	X_1 reflects 60% of the total variance explained in the factor analysis.	
d	All statements (a)-(c) are false.	

31	**A conjoint analysis was conducted to analyze respondents' preferences between different types of teas. The teas are characterized by the attributes price (low, medium, high), variety (black, green, red), kind (bags, granulated, leafy), and aroma (yes, no). Based on the following output of a conjoint analysis computed with the R-function conjoint(), what is the total predicted utility of a green tea of medium price, in bags, and without aroma?** Residuals: Min 1Q Median 3Q Max -5,1888 -2,3761 -0,7512 2,2128 7,5134 Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 3,55336 0,09068 39,184 < 0,0000000000000002 *** factor(x$price)1 0,24023 0,13245 1,814 0,070 . factor(x$price)2 -0,14311 0,11485 -1,246 0,213 factor(x$variety)1 0,61489 0,11485 5,354 0,0000001019348 *** factor(x$variety)2 0,03489 0,11485 0,304 0,761 factor(x$kind)1 0,13689 0,11485 1,192 0,234 factor(x$kind)2 -0,88977 0,13245 -6,718 0,0000000000276 *** factor(x$aroma)1 0,41078 0,08492 4,837 0,0000014751866 *** --- Signif. codes: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1 Residual standard error: 2,967 on 1292 degrees of freedom Multiple R-squared: 0,09003, Adjusted R-squared: 0,0851 F-statistic: 18,26 on 7 and 1292 DF, p-value: < 0,00000000000000022 [1] "Part worths (utilities) of levels (model parameters for whole sample):" levnms utls 1 intercept 3,5534 2 low 0,2402 3 medium -0,1431 4 high -0,0971 5 black 0,6149 6 green 0,0349 7 red -0,6498 8 bags 0,1369 9 granulated -0,8898 10 leafy 0,7529 11 yes 0,4108 12 no -0,4108 [1] "Average importance of factors (attributes):" [1] 24,76 32,22 27,15 15,88 [1] Sum of average importance: 100,01 [1] "Chart of average factors importance" >	Answer (2 points)
a	3.7513	
b	3.2173	
c	3.1713	
d	-0.3821	

32	Which of the measures of model performance for a cluster analysis trades off model complexity and model fit?	Answer (2 points)
a	Silhouette coefficient	
b	Within-cluster sum of squares	
c	Deviance	
d	Akaike Information Criterion	

33	**For the principle component analysis in Question 29, the figure below displays the scree plot with the factors' eigenvalues ordered from largest to smallest. Which method for determining the number of factors was used?**  	Answer (2 points)
a	Elbow criterion	
b	Scree test	
c	Parallel analysis	
d	Simple structure	

34	**Which of the following research practices is an example of p-hacking?**	Answer (2 points)
a	Conducting a study with low statistical power.	
b	Conducting a study without having conducted an a-priori power analysis.	
c	Excluding observations because they deviate from the pattern in the other observations and are responsible for a non-significant result.	
d	Planning to collect a very large number of observations to increase the chance that results will be significant.	

35	**Which of the following statements comparing parallel analysis and the scree test is correct?**	Answer (2 points)
a	Only the parallel analysis takes into account model complexity (rather than merely model fit) and therefore always suggests retaining fewer components than the scree test.	
b	The parallel analysis usually points to a less appropriate component solution than the scree test because it enforces a less clearly differentiated distribution of eigenvalues than the scree plot.	
c	Only the scree test takes into account the absolute size of the eigenvalues.	
d	Only the parallel analysis takes into account the distribution of eigenvalues that would be expected by chance.	

Empirical Research in Management and Economics
Answer Sheet for Multiple Choice Questions

Last name **First name**

Student ID (matriculation number)

Instructions:

- **ONLY ANSWERS ON THIS SHEET ARE GRADED!**
- There is only **one correct answer** per question.
- **Fill in the entire circle** that corresponds to your answer for each question on the exam.
- Use a **pen**, not a pencil.
- **To make a change, fill in the circle for the correct answer and mark it with an arrow next to it.**

1	A	B	C	D
2	A	B	C	D
3	A	B	C	D
4	A	B	C	D
5	A	B	C	D
6	A	B	C	D
7	A	B	C	D
8	A	B	C	D
9	A	B	C	D
10	A	B	C	D
11	A	B	C	D
12	A	B	C	D
13	A	B	C	D
14	A	B	C	D
15	A	B	C	D
16	A	B	C	D
17	A	B	C	D
18	A	B	C	D
19	A	B	C	D
20	A	B	C	D

21	A	B	C	D
22	A	B	C	D
23	A	B	C	D
24	A	B	C	D
25	A	B	C	D
26	A	B	C	D
27	A	B	C	D
28	A	B	C	D
29	A	B	C	D
30	A	B	C	D
31	A	B	C	D
32	A	B	C	D
33	A	B	C	D
34	A	B	C	D
35	A	B	C	D