Family name	
First name	
Study program	
Student ID number	

Important notes:

- Fill out this cover completely!
- Put your student card visibly on the table!

****

(leave blank)

Nr.	Korrektur	Nachkorrektur
1		
2		
3		
4		
5		
6		
Total		

Written Exam
(1 hours)

General remarks:

- You are allowed to use
  - a pocket calculator
  - a dictionary
  - a handwritten summary of at most 8 pages.
- All tasks (1 to 6) are multiple-choice questions. Only one answer is correct for each question.
  - A **correct answer** adds +1 point.
  - A **wrong answer** yields −1 point.
  - Giving **no answer** yields 0 points.

  Accumulation of points:
  In all tasks the points are accumulated over the whole task. In any case, you get a minimum of 0 points for each task.
- If not stated otherwise, all tests have to be done at the 5%-level.
- Do not stay too long at a part where you experience a lot of difficulties.
- Switch off your mobile phone!

Good Luck!

1. Descriptive analysis

a) A list of puls rates is: $70, 64, 80, 74, 92$. What is the median for this list?

- ☐ 72
- ☐ 74
- ☐ 77
- ☐ 80

b) If the mean of 10 blood pressure changes is negative, then also the standard deviation of these 10 values is negative.

- ☐ True
- ☐ False
- ☐ Cannot be told.

c) Which of the following data can be well visualized by a histogram?

1. The blood pressure of 50 patients.
   - ☐ True
   - ☐ False
2. The gender of 40 patients.
   - ☐ True
   - ☐ False
3. The weight of 5 patients.
   - ☐ True
   - ☐ False
4. The number of missing kidneys per patient over all 430 patients of a doctor.
   - ☐ True
   - ☐ False

d) The distribution of the blood-concentration of a certain doping drug in 120 randomly controlled athletes looks right skewed - which kind of data transformation can change the shape of a distribution?

- ☐ A square-root transformation
- ☐ Standardization
- ☐ An appropriate linear transformation
- ☐ A log transformation

e) Which of the following would indicate that a dataset is not bell-shaped?

- ☐ there are no outliers
- ☐ the mean is much larger than the standard deviation
- ☐ the mean is much smaller than the median
- ☐ the standard deviation is larger than 6

f) Which one of these statistics is least affected by outliers?

- ☐ Mean
- ☐ Interquartile range
- ☐ Standard deviation
- ☐ Median

2. Testing

A reaction test was performed by 12 young and 100 old men and supervised by a PhD-student. Since reaction times can be assumed to be normally distributed, the student uses an unpaired t-test on a significance level of $\alpha = 1\%$ to investigate if the mean reaction time of young men is different from the mean reaction time of old men. The resulting p-value is p=0.04.
Which of the following statements are correct based on this information?

a) There is statistical evidence on a significance level of 1% that there is no difference in the mean reaction time of young and old men.
☐ True ☐ False ☐ Cannot be told

b) There is statistical evidence on a significance level of 1% that there is a significant difference in the mean reaction time of young and old men.
☐ True ☐ False ☐ Cannot be told

c) The 99% confidence interval for the mean difference of reaction times does cover the zero.
☐ True ☐ False ☐ Cannot be told

d) If the test would have been conducted on a 5%-significance level then the test would have resulted a significant difference in the mean reaction time of young and old men.
☐ True ☐ False ☐ Cannot be told

e) If the sample sizes are increased then we have better chances to get a significant result.
☐ True ☐ False ☐ Cannot be told

f) It would have been also valid to use the unpaired Wilcoxon-Test.
☐ True ☐ False ☐ Cannot be told

g) It would have been better, if the student would have used a paired t-test
☐ True ☐ False ☐ Cannot be told

h) ANOVA would have yielded the same results as the t-test.
☐ True ☐ False ☐ Cannot be told

i) It is not possible to get the same results with a linear regression
☐ True ☐ False ☐ Cannot be told

3. Study design and the role of the different variables

A study was done to compare the lung capacity of coal miners to the lung capacity of farm workers. The researcher studied 200 workers of each type. Other factors that might affect lung capacity are smoking habits and exercise habits. The smoking habits of the two worker types are similar, but the coal miners generally exercise less than the farm workers.

a) Which of the following is the primary explanatory variable in this study?
- ☐ Exercise
- ☐ Lung capacity
- ☐ Smoking (Yes/No)
- ☐ Occupation

b) Which of the following is the response variable in this study?
- ☐ Exercise
- ☐ Lung capacity
- ☐ Smoking (Yes/No)
- ☐ Occupation

c) Which of the statistical methods are appropriate to compare the lung capacity of coal miners and farmers in the study?
- ☐ Regression
- ☐ Barplot
- ☐ Binomial Test
- ☐ Chi-Square Test

d) What is the study type of this study?
- ☐ Observational Study
- ☐ Non randomized experimental study
- ☐ Randomized experimental study

4. Correlation

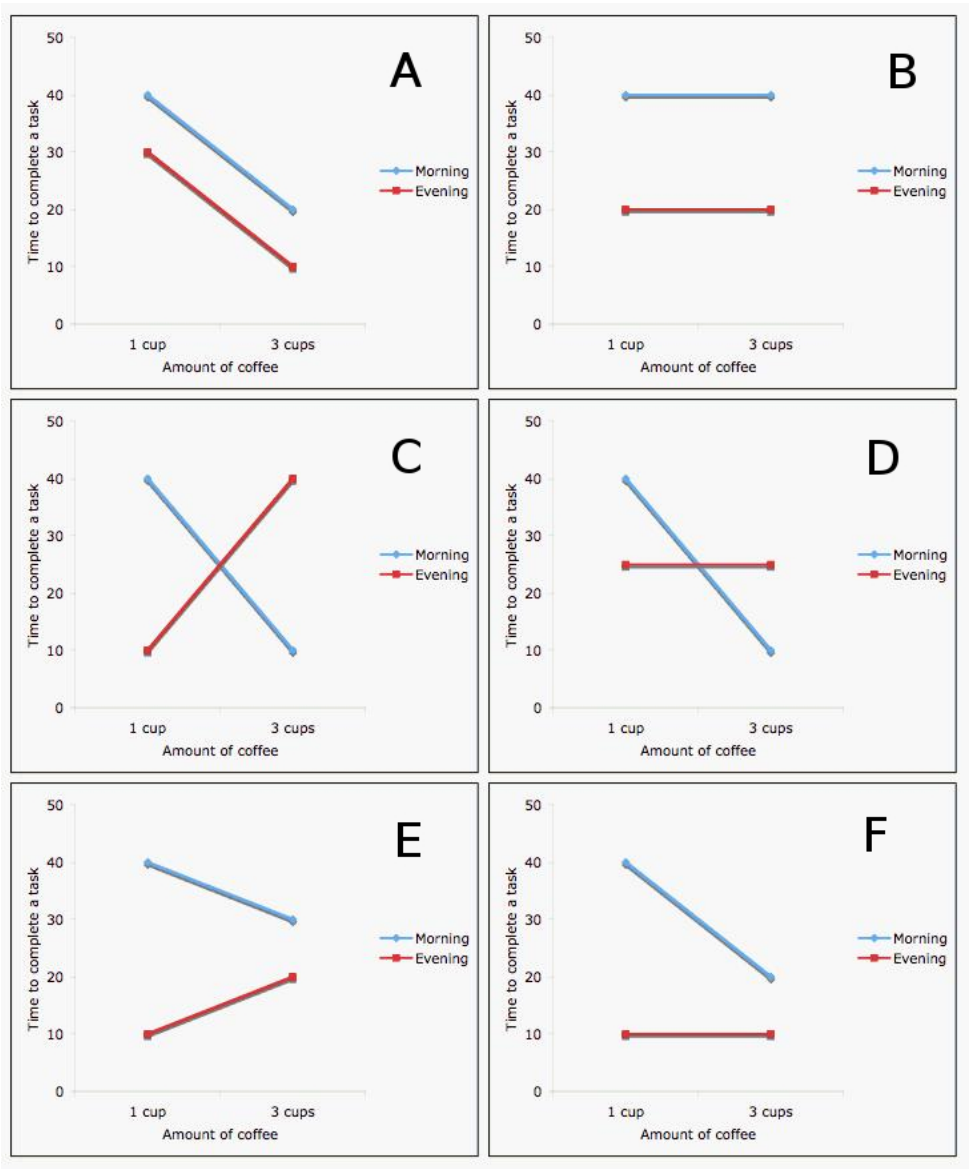
a) If the Pearson correlation between blood pressure and body weight of guinea pigs is zero, then we can conclude that body weight has no influence on the blood pressure in these animals.
☐ True ☐ False

b) If the Pearson correlation is an appropriate measure and yields a positive number then also the Spearman rank correlation would lead a positive number.
☐ True ☐ False

c) A scatter plot of the number of medical doctors and the number of people who suffer from diabetes for cities in Switzerland reveals a positive association. What is the most likely explanation for this positive association?
- ☐ The presence of medical doctors encourage people to have an un-healthy life style.
- ☐ Rich cities tend to have more medical doctors and more obese people.
- ☐ Larger cities tend to have both – more medical doctors and more sick people.
- ☐ Cities with many people suffering from diabetes attract a lot of medical doctors.

5. Statistical models and their interpretation

A number of possible models for the effect of `daytime` and `amount of coffee` on the `time to complete a task` are depicted in the following figure.



a) If we look only at the variable `amount of coffee`, in which model do we have an effect on the `time to complete a task`?

- ☐ A, B, C, D, E and F.
- ☐ A, C, D, E and F.
- ☐ A, B, C, E and F.
- ☐ only in C and D.

b) If we look only at the variable `daytime`, in which model do we have an effect on the `time to complete a task`?

- ☐ C, D, E and F.
- ☐ A, B, E and F.
- ☐ only in B.

c) In which model is an interaction present between `daytime` and `amount of coffee` in their effect on the `time to complete a task`?

- ☐ A and F.
- ☐ C, D, E and F.
- ☐ A, B, E and F.
- ☐ only in B.

d) Which variables are assumed as factor variables?

1. `amount of coffee`
   - ☐ True
   - ☐ False
   - ☐ Cannot be told.
2. `time to complete a task`
   - ☐ True
   - ☐ False
   - ☐ Cannot be told.
3. `daytime`
   - ☐ True
   - ☐ False
   - ☐ Cannot be told.

e) In which case could the Simpson effect lead to the false conclusion that the `amount of coffee` has no effect on `time to complete a task` when dropping the `daytime` variable from the model?

- ☐ A and F.
- ☐ C and D.
- ☐ C and E.
- ☐ E and F.

6. Linear regression

In a study it was analyzed if the ph in the soil has an influence on the height (measured in meter) of a specific species of trees. The study was based on 10 year old trees which were all planted on soils with ph values between 7 and 9.5. The following R output shows the result of a linear regression. You can assume that the model assumptions are fulfilled.

```
Residuals:
    Min      1Q  Median      3Q     Max
-3.7020 -0.5471  0.0874  0.6663  2.0033

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  28.7227     2.2395   12.82   <2e-16 ***
ph           -3.0034     0.2844  -10.56   <2e-16 ***
---

Residual standard error: 1.008 on 121 degrees of freedom
Multiple R-squared:  0.4797,        Adjusted R-squared:  0.4754
F-statistic: 111.5 on 1 and 121 DF,  p-value: < 2.2e-16
```

a) The soil ph has a statistically significant effect on the tree height

- ☐ True
- ☐ False
- ☐ Cannot be told.

b) There is a significant negative correlation between ph and height

- ☐ True
- ☐ False
- ☐ Cannot be told.

c) How many trees were included in the study?

- ☐ 28
- ☐ 100
- ☐ 121
- ☐ 122
- ☐ 123

d) A farmer claims that the height of a tree decreases in average by 0.35 meter when the ph increases by 0.1. Does the result of the regression contradict this statement?

- ☐ Yes
- ☐ No

e) Which mean height would you predict for trees grown on a soil with ph=8?

- ☐ 1.8 m
- ☐ 3 m
- ☐ 4.7 m

- ☐ 5.2 m

f) According to this study we would expect a height of 29m for a tree on a soil with ph=0

- ☐ True
- ☐ False
- ☐ Cannot be told.

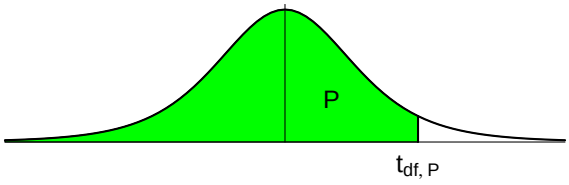
g) It is possible that the estimated coefficient of the variable `ph` becomes positive, if an additional explanatory variable is added to the model.

- ☐ True
- ☐ False
- ☐ Cannot be told.

h) If we want to account for the effect of the mean daily rain volume we should:

- ☐ work only with observations from trees which received the same mean rain volume.
- ☐ include the mean rain volume into the linear regression model
- ☐ fit a second model which uses only the mean rain volume as explanatory variable.
- ☐ use ANOVA instead of linear regression,
- ☐ use logistic regression instead of linear regression.

Perzentile der t-Verteilung



Bsp.: $t_{9;\ 0.975} = 2.262$

df	$t_{0.60}$	$t_{0.70}$	$t_{0.80}$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750
31	0.255	0.530	0.853	1.309	1.696	2.040	2.452	2.744
32	0.255	0.530	0.853	1.309	1.694	2.037	2.449	2.738
33	0.255	0.530	0.853	1.308	1.693	2.035	2.445	2.733
34	0.255	0.529	0.852	1.307	1.691	2.032	2.441	2.728
35	0.255	0.529	0.852	1.306	1.690	2.030	2.438	2.724
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
90	0.254	0.526	0.846	1.291	1.662	1.987	2.368	2.632
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576