Examination Statistics
Prof. Dr. Falkenberg, Prof. Dr. Schrader
Course of Study: Computer Sciences
23.02.2023
Duration: 90 Minutes

Name : _____
Matriculation Number : _____

I hereby confirm with my signature

- **that I have composed this report on my own without external aid and with only those resources explicitly allowed in the exam formulation,**
- **that I feel in good health and that I am able to participate in the exam. I am aware of the fact that having received the exam formulation I have attempted the exam and my exam will be counted as an attempt and evaluated by the examiner.**

Signature student _____

Problems	1	2	3	4	5	Sum	Mark
Max. scores	30	15	15	15	15	90	
Obt. scores							

Authorized examinations aids:
Scientific calculator, one file folder

Further instructions:

1. Please use RStudio on the Windows system.
2. Submit all you want to be assessed (derivations, answers, interpretations, commands, diagrams, etc.).
3. You are allowed to submit totally **ONE (1)** computer file. The file with the last time stamp will be corrected, other files NOT!!!
4. The computer file should be a **.pdf-document**.
5. The name of the computer file must be your complete name and matriculation number.
6. Write your name and your matriculation number at the beginning of the computer file.
7. The exam and additional files can be found in the folder Vorlagen on drive Z.
8. Your computer file must be saved in the folder Ergebnisse on drive Z.
9. Please notice, not only the solution but the derivation of the solution has to be given.
10. Do not use your own paper. Blank paper is attached to this question paper. Extra sheets will be provided on demand.
11. Switch off all electronic devices und remove them from the table.

Good Luck!
Dr. Falkenberg, Dr. Schrader

Descriptive Statistics

1. The two files stud-data.csv and exam-data.csv contain the data of students of two study programs as well as the results of the exams from the 3rd semester of the winter semester 21/22.

mat.nr	matriculation number
gender	sex
semester	current semester
course	course of study
exam	matriculation number/exams
attempt	number of attempt
score	achieved score

   (a) Import the files stud.data, exam.data as tibbles.

   (b) Determine the scale and type of all variables.

   (c) Add a variable grade that indicates the grade of the exam. The grade is derived from the score as follows:

   - Grade 5 if score < 50
   - Grade 4, if $50 \leq$ score < 65
   - Grade 3, if $65 \leq$ score < 80
   - Grade 2, if $80 \leq$ score < 90
   - Grade 1, if $90 \leq$ score

   (d) Split the variable exam into 2 columns containing the matriculation number and the exam subject.

   (e) Determine the total number of tests in each exam and the number of students participating

   (f) For each subject, determine the absolute frequencies of the grades and store the result in a tibble with the variables grade, Computer Networks, Data Bases, Formal Languages, Mathematics, OOP and Software Engineering.

   (g) For each subject, determine the minimum, maximum, the three quartiles, the mean of the variable score, the number of participants and the dropout rates.

   (h) Create side by side boxplots of the score for each subject and interpret the results

   (i) Determine the contingency table of the variables attempt and grade and determine the indifference table and chi-square value..

Probability

2. A biased coin (head with probability 1/3) is tossed. If the coins shows tail a fair die is rolled 5 times and if the coin shows head a biased die (6 with probability 0.4 and the other sides of the die each have the same probability) is rolled 5 times. The number of sixes are counted.

   (a) Determine the density of the random X which counts the number of sixes.

   (b) Evaluate the expected value and the variance of the random variable X.

   (c) What is the probability that the coin had shown a head if 3 sixes has been in the 5 rolls?

3. The weight of bags of grain can be assumed to be a random variable with expected value 50 kg and standard deviation 2 kg. The price for one kilogram grain, which a farmer achieves, is 0.53 Euro per kg. 300 bags of grain fit into a truck. Let X be the price a farmer can obtain for a fully loaded truck.
   You can assume that the weights of bags are independent and identically distributed random variables.

   (a) Determine an approximate distribution of the random variable X.

   (b) Find the probability that X is bigger than 8000 Euro.

   (c) What are the lower and upper bounds of the interval containing the middle 80% of X?

   (d) What is the minimum number of bags a farmer must sell to earn at least 20000 euros with a probability of 95%?

Inferential Statistics

4. In a representative sample 100 people were asked if they prefer candidate A or not if the election were held next Sunday.
   The following values were obtained (1=yes, 0=no)

```
1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0,
1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0,
0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0,
0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0,
0, 0, 1, 0, 1, 0, 0, 0
```

   (a) Determine the relative frequency of voters prefering A.

   (b) Show that the relative frequency is an unbiased point estimator for the proportion of voters prefering A in the whole population.

   (c) Determine a normal approximation of two-sided 95% confidence intervals for the unknown proportion of voters of A.

   (d) Find an approximate value of the minimal sample size so that the length of the approximate confidence interval for the proportion of voters prefering A is less than 0.1.

   (e) Determine the confidence level so that the width of the approximate confidence interval of the unknown proportion of voters preferring A is equal to 0.1.

5. According to the manufacturer the car model XYZ should consume no more than 5l per 100 km in the average. A car magazine doubts this claim and tests the consumption of 20 randomly selected vehicles. The average consumption per 100 km of the 20 cars is:

   ```
   5.51, 5.42, 5.39, 5.07, 5.03, 4.85, 5.31, 4.96, 5.14, 5.24,
   4.95, 4.84, 5.16, 5.23, 5.19, 5.47, 5.06, 5.23, 5.04, 5.36
   ```

   (a) What are the null hypothesis and alternative for an appropriate statistical test?

   (b) Perform an appropriate statistical test to verify the null hypothesis at the 5% level. What is the test decision and what is the p-value?

   (c) In making the above conclusion, which type of error are you risking, type I or type II?

   (d) The manufacturer promotes a new fuel supplement, which leads to a reduction of the average consumption per 100 km of car model XYZ. The car magazin doubts this claim, too and believes that the fuel supplement has no influence on the average consumption. To test this, the car magazin conducts an experiment in which they measure the consumption of the same 20 cars with the fuel supplement.

   ```
   5.11, 5.20, 4.91, 4.94, 5.22, 4.59, 5.45, 4.80, 5.13, 4.79,
   4.69, 4.14, 4.64, 5.20, 5.16, 5.49, 4.94, 5.75, 5.36, 4.69
   ```

   Perform an appropriate statistical test to verify the conjecture at the 5% level and determine the pvalue.

   Hint: Assume that the average consumption follows a normal distribution.