Examination Statistics
Prof. Dr. Falkenberg, Prof. Dr. Schrader
Course of Study: Computer Sciences
3.3.2022
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	10	15	10	15	80	
Obt. scores							

Authorized examinations aids:
Scientific calculator, one file folder

Further instructions:

1. Please write your handwritten solutions directly into the distributed exam.
2. Do not use your own paper. Blank paper is attached to this question paper. Extra sheets will be provided on demand.
3. Please save your solutions in R to a computer file in the folder Ergebniss on drive Z. Mark the task to which the respective solutions belong.
4. You are allowed to submit totally ONE (1) computer file. The file with the last time stamp will be corrected, other files NOT!!!
5. The computer file should be a .pdf-document.
6. The name of the computer file must be your complete name and matriculation number.
7. The exam and additional files can be found in the folder Vorlagen on drive Z.
8. Please notice, not only the solution but the derivation of the solution has to be given.
9. Switch off all electronic devices und remove them from the table.

Good Luck!
Dr. Falkenberg, Dr. Schrader

Descriptive Statistics

1. (a) Import the file melanoma.csv as a tibble called melanoma.
   (b) Determine type and scale of all variables. You find a description of the dataset in the file melanoma_description.pdf
   (c) Change the values of the variables sex, status, ulcer to strings describing their values and add a new variable live.status describing whether the patient is alive or dead.
   (d) Create a contingency table for the variables sex and live.status.
   (e) Evaluate the relative risks to survive at least 3 years for the variable sex and interpret the values.
   (f) Create a summary describing the distribution of the variable age containing min, max, mean, the three quartiles depending on the variable sex.
   (g) Create side by side boxplots for the age of persons depending on their sex and interpret the diagram.
   (h) The csv file add.data.melanoma.csv contains data from another study. Import the dataset as a tibble called add.data.melanoma.
   (i) Is this dataset tidy?

Probability

2. In a computer science course the projects P1, P2, P3, P4 are offered. Each of the 60 students must sign up for one of the projects offered. All projects are equally popular among the students. Determine the probability that

   (a) exactly 15 students register for every of the four projects

   (b) more than 15 students sign up for project P1.

   Assume that the number of places in the projects is unlimited:
   Hint: The R functions factorial(n) and choose(n,m) evaluate n! and $\binom{n}{m}$.

3. Every year, an introductory computer course is held at the beginning of the winter semester. From many years of experience, we know that about 11% of the registered course participants do not show up for the course. Since each participant needs his own computer during the course, no more participants can take part in the course than there are free computers. In total, there are ten rooms with 22 seats each and a total of 240 first-year students. Using an approximation by the central limit theorem, calculate

   (a) the probability that all students who are present for the course will find a seat if all first-year students have registered for the course.

   (b) the minimum number of computers needed so that there is at least a 99% probability that all students who show up will have a computer?

   (c) how many registrations may be accepted at most, if with probability 0.99 all students who show up for the course will find a place in the course with 220 places.

Inferential Statistics

4. To estimate the prevalence of a disease in a population, a random sample of size n=200 is drawn from the population and the individuals are screened for for the disease. In the process, m=12 individuals are found to have the disease.

   (a) Show that $X = \frac{m}{n}$ is an unbiased estimator of prevalence.

   (b) Determine the variance of the estimator X.

   (c) Determine an upper 95% confidence bound for prevalence from the sample data.

   (d) What is the minimum sample size needed for the upper 95% confidence bound is 0.01 greater than the estimate. Assume that the prevalence is ≤ 0.1 and use a normal approximation of the confidence bound.

5. A company produces chocolade bars with a standard weight of 100 gr. As a measure of quality controls he weighs 15 bars and obtains the following results:

   ```
   98.32,97.26,99.85,99.52,95.73,95.56,100.49,98.19,95.16,
   98.26,96.46,100.23,99.76,98.58,97.43
   ```

   (a) What is an appropriate hypothesis regarding the expected weight μ for a two-sided-test?

   (b) If weights can be assumed to be nomally distributed, which test should used to test these hypothesis?

   (c) Conduct the test that was suggested to be used in b) at a 5% level. What is your test decision. Specify the p-value.

   (d) Based on the sample, the producer changes the settings in production. To check whether the correction has led to an improvement, he again takes 15 chocolate bars and weighs them.

   ```
   100.14,100.05,96.51,98.70,98.22,101.06,103.55,100.16,
   100.60,102.85,103.15,100.66,102.52,102.09,100.84
   ```

   What is an appropriate hypothesis for comparing the expected weights of the two samples?

   (e) Provide an appropriate statistical test to test the hypothesis and perform at the 5% level. Assume that the variances of the populations of the two samples are equal. What is your test decision? Specify the p-value.

   (f) In question e) the population variances of the two samples are assumed to be equal. Verify that the variances are equal using an appropriate test at the 10% level.