

Final Exam SA.01

Saturday, 19. 1. 2019

Duration: 90 Minutes

Surname, Forename: _____

IDS Nr.: _____ (On the back of your HSLU-Card) L _____

Stick number: _____

Signature: _____

Problem	1	2	3	4	Total
max. Points	5	10	10	10	35
achieved Points					

Important Information

- Allowed aids:
 - a) Printed *R reference card* (annotations allowed).
 - b) Ten (10) pages DIN-A4, with arbitrary content. This corresponds to five sheets DIN-A4 paper written on both sides.
 - c) Calculator
 - d) Formulary
- Other aids are not allowed! Switch off all phones.
- Place your HSLU-Card in front of you on your table.
- **All of your answers should come with explanations! Solutions without understandable justifications obtain no credit.**

Good luck!
Peter Büchel

Saturday, 19. 1. 2019

Problem 1: (5 Points)

A lie detector test is done routinely to employees working in sensitive positions. Let $+$ denote the event of a positive test outcome, meaning the lie detector claims the employee is lying. Let W denote the event, that the employee is actually telling the truth, and L the event, that the employee is actually lying.

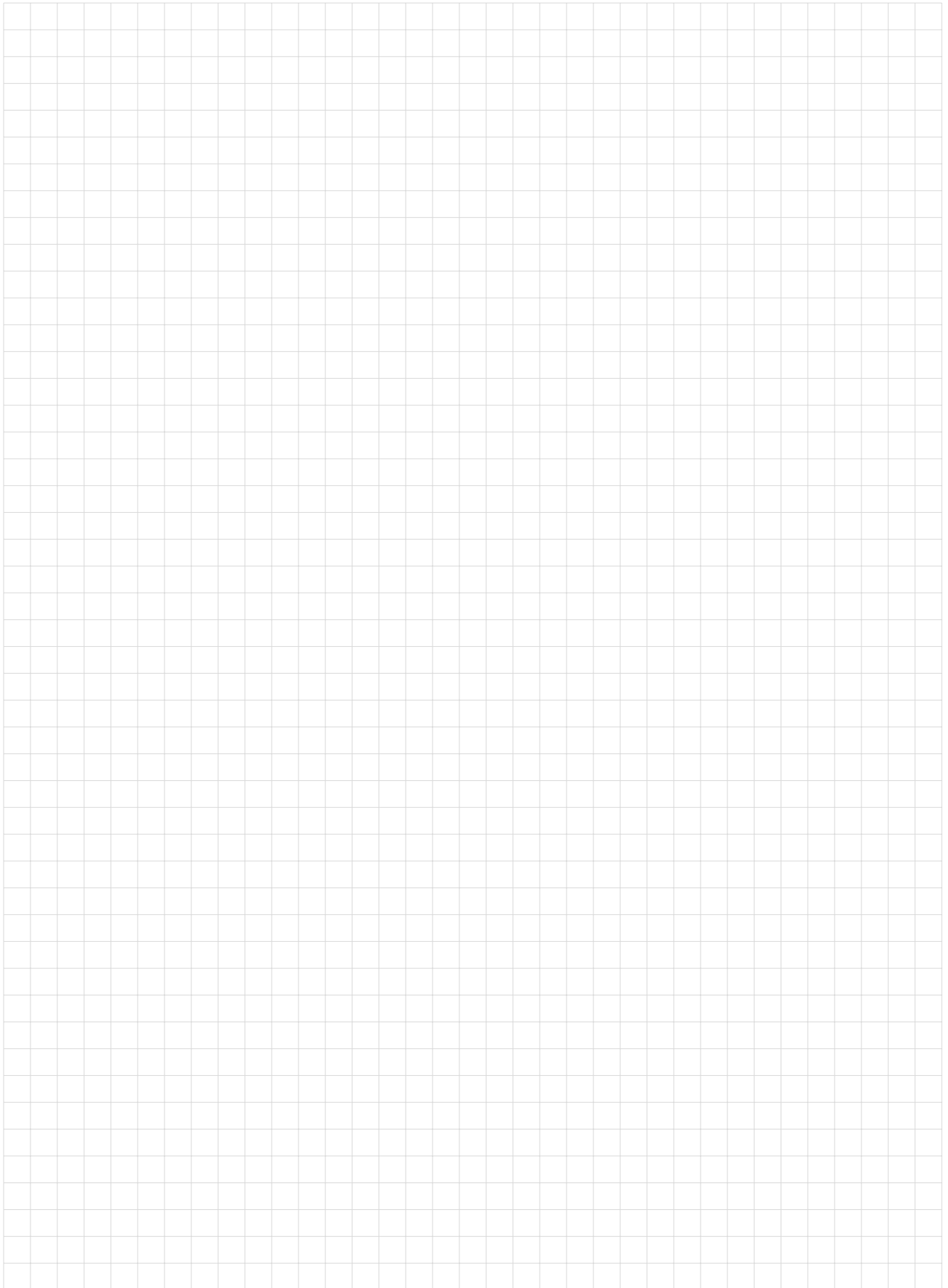
Trials with the lie detector show that

$$P(+|L) = 0.88 \quad \text{and} \quad P(-|W) = 0.86$$

Further it is assumed that

$$P(W) = 0.99$$

- (a) Interpret the meaning of $P(W)$ and $P(+|L)$. [1]
- (b) The lie detector claims an employee is lying. What is the probability that this employee actually lied? [2]
- (c) Interpret the result of (b) in 2-3 sentences. How convincing is the use of the lie detector? [2]



Problem 2:(10 Points)

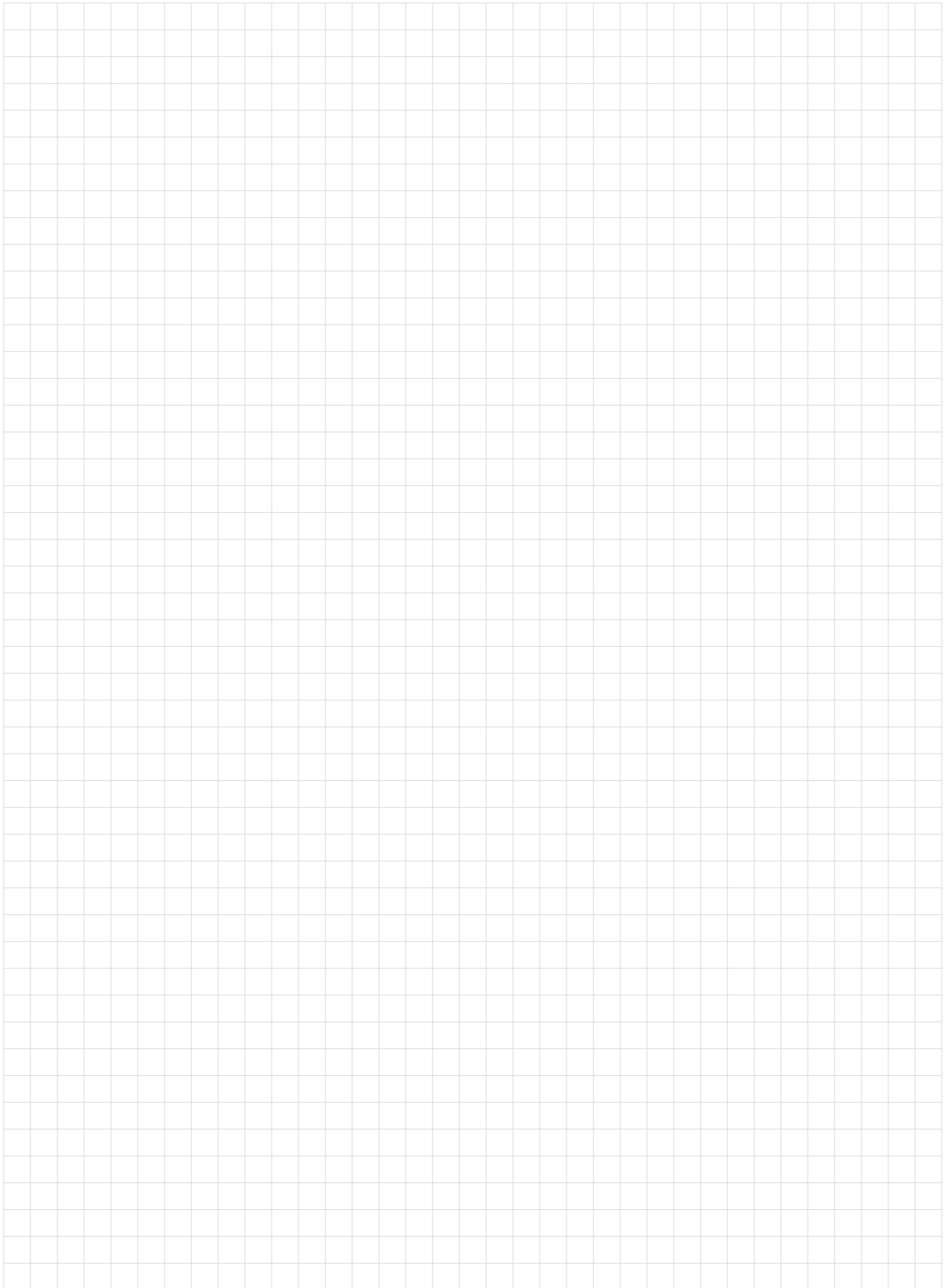
On the 14th of April 1912, the *Titanic* hit an iceberg and sank. Of about 2245 passengers approximately 1500 died. The data sets **survived.dat** and **died.dat** contain the age of the (known) passengers who survived or died respectively.

We want to investigate whether there is an age difference between those passengers who survived and those who died.

Hint: The commands for loading the data sets are included in the **R**-file on the stick.

- (a) We want to investigate the age difference with an hypothesis test. [2]
 - i) Do you choose a test for paired or unpaired samples? Justify your answer!
 - ii) Do you choose a one-sided or two-sided test? Justify your answer!
- (b) Calculate the average and standard deviation and give an interpretation of these values. [2]
- (c) Initially we perform a t -test. [4]
 - i) What are the requirements to justify a t -test?
 - ii) Describe the null hypothesis and the alternative hypothesis.
 - iii) Perform the test and use the p -value for the test decision on a significance level of 5 %.
 - iv) Use the confidence interval for the test decision.
- (d) We perform a Wilcoxon-test. [2]
 - a) Give an argument to use the Wilcoxon-test instead of a t -test.
 - b) What is your test decision in this case?
 - c) Explain the difference of the p -values from the t - and Wilcoxon-test?





Problem 3:(10 Points)

The data set **potato.csv** relates to a study which investigated the quality of potatoes with respect to texture, flavor and moisture depending on several factors (A. Mackey and J. Stockman (1958). „Cooking Quality of Oregon-Grown Russet Potatoes“, American Potato Journal, Vol.35, pp395-407).

In this problem, we just consider the factors **Size** (size of the potato: Large, Medium) and **CookingMethod** (Boil, Steam, Mash, Bake.350 (Fahrenheit), Bake.450).

Hint: The command for loading the data set is included in the **R**-file on the stick.

- (a) Generate two boxplots with response variable **MoistnessScore** (the smaller this value, the drier the potato) depending on the predictors **Size** and **CookingMethod** respectively. [2]
Interpret these plots using the notions of medium value.
- (b) Generate an interaction plot of **Size**, **CookingMethod** and **MoistnessScore**. [3]
Interpret this plot.
- (c) Perform an analysis of variance for the response variable **MoistnessScore** depending on the predictors **Size** and **CookingMethod**. [3]
Take interaction into account. Describe null hypotheses, alternative hypotheses and give an interpretation of your test decision on a significance level of 5 %.
- (d) Repeat (c) for **FlavorScore** (the higher this value, the better the taste of the potato). Interpret this result. [2]

