EXAM COVER SHEET

Module:	**D2, Design and Analysis of Experiments**
Date of exam:	11 June 2019, 1.00 – 3.00pm
Duration:	120 min
Type of exam:	Open book, open internet, no communication with other human beings (i.e. printed material allowed, laptop allowed, wlan allowed, not allowed to communicate with other people)
Module coordinator:	Christoph Kopp (BFH)

Name of the student (readable!): ……
Please write your name also on <u>every</u> page (at the top of the sheet)!

School: ❒ ZHAW ❒ BFH ❒ FHNW ❒ HES-SO

Venue of exam: ……

Exam

Version A

Briefing

- Next to each problem, the number of points is indicated in parentheses, e. g. (3).
- The level of significance is 5% unless otherwise mentioned. Accordingly, confidence bounds and intervals have a confidence level of 95% by default.
- Include a short reasoning (e. g. *"I used a marginal F test and obtained a p value of . . . "*) for your results. This permits giving points for your approach even in the final answer should not be not correct.
- Give numeric results (such as p values) to at least three digits.
- All answers must be copied on this exam, file submissions are not accepted.

Best of luck!

For correction, please do not write into.

Prob. 1	Prob. 2

Problem 1 (16)

The federer.tobacco data frame from the agridat library contains data on the total height (in cm, the sum of 20 tobacco plants per treatment and block) which had been exposed to seven different doses of radiation (dose, in roentgen). The seedlings were transplanted in a randomized complete block design with eight blocks. It is the main aim to model the effect of the radiation on the height, while accounting for potential block effects.

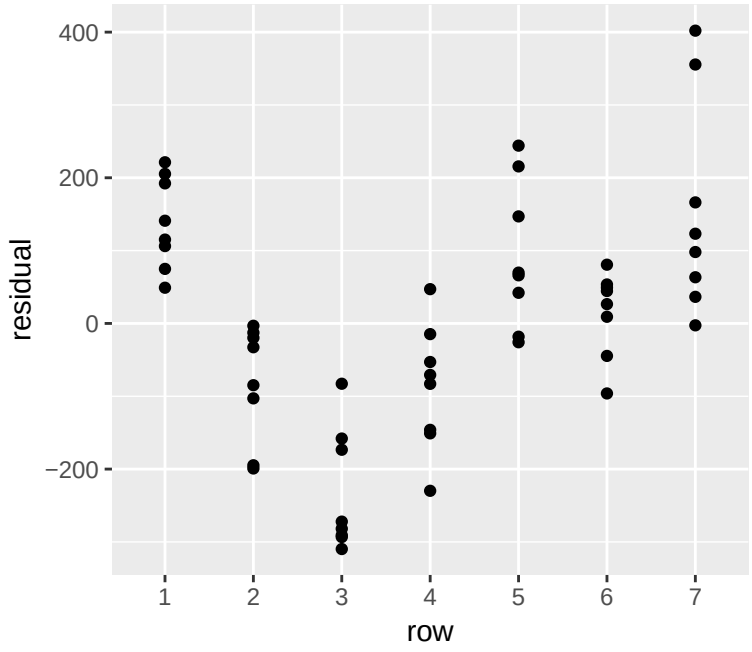
Do turn dose, block and row into factors now, e. g. as follows:

```
> federer.tobacco$dose <- factor(federer.tobacco$dose)
```

a.) Give the R code to fit a suitable parametric model to this data set to answer the research question. (1)

b.) Is the dose effect significant according to the model from a.)? Name the tools used, give the R code to perform the tests, report the p value and use it to answer the question.(2)

c.) The layout of the experiment was in eight rows and seven columns. Shortly after the plants were transplanted to the field it became apparent that an environmental gradient existed. The following residual plot for the model in a.) was produced:



Use this residual plot to explain that there is a gradient and discuss what it looks like.(2)

d.) To take care of the mentioned gradient, one can add the fixed effect of the row to the model. Give the R code to fit this model. (1)

e.) According to the model in d.), is the radiation effect significant? And the row effect? Use the same tools as in b.), give your R code, p values and your decision. (1)

f.) According to the model in d.), by how many cm does the total plant height (of the 20 plants) decrease on average for a "typical" block if the dose was 5'000 roentgen, compared to 0 roentgen? Explain your approach. (2)

g.) Is the answer to f.) different for different rows or not? Why or why not? (1)

h.) Assess the model assumptions: explain what you assess, with which method, give your R code, discuss the results (give p values where you calculate them) and report your conclusions regarding the assumptions. (2)

i.) You should have found a problem with the assumptions. What could you do about it? Give a precise answer for full points, not only a rough idea. (2)

j.) We want to compare each of the dose levels only to its adjacent levels (0 to 250, 250 to 500, ...). Show how to test which of these differences are significant, and give the significant results only, with p values. (2)

Problem 2 (14)

The data set gregory.cotton from the agridat package contains data from an experiment to study the effects of the nitrogen level (levels: N0: None, N1: 600 rotls/feddan [a local unit in Sudan]), the sowing date (levels: D1 up to D4, see help for exact dates), and two other factors (water and spacing) on the yield (in a local unit) of cotton plants. Data were recorded for two years (Y1, Y2). For this problem, we focus only on the effect of the sowing date, the nitrogen and the year on the yield.

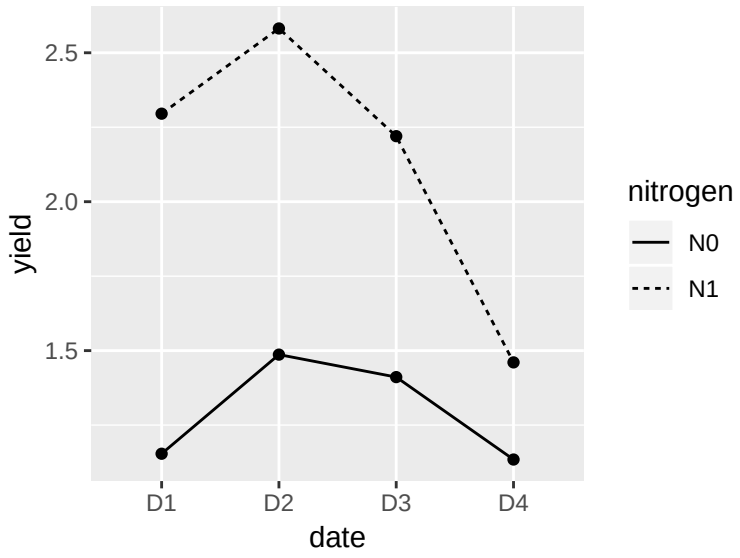
a.) According to the sample data from year Y1: which of the four dates should we choose to get the highest average yield if we use the nitrogen fertilizer, and how high is the average yield for this sowing date in year Y1 for the fertilized plots? (1)

b.) We begin with a model that has no year effect. Give the R code to model the effect of the sowing date, the nitrogen fertilizer and their interaction on the yield. (1)

c.) Does the effect of the nitrogen fertilizer on the yield depend on the date according to the model from b.)? Give your code, and answer the question, argue with p values. (2)

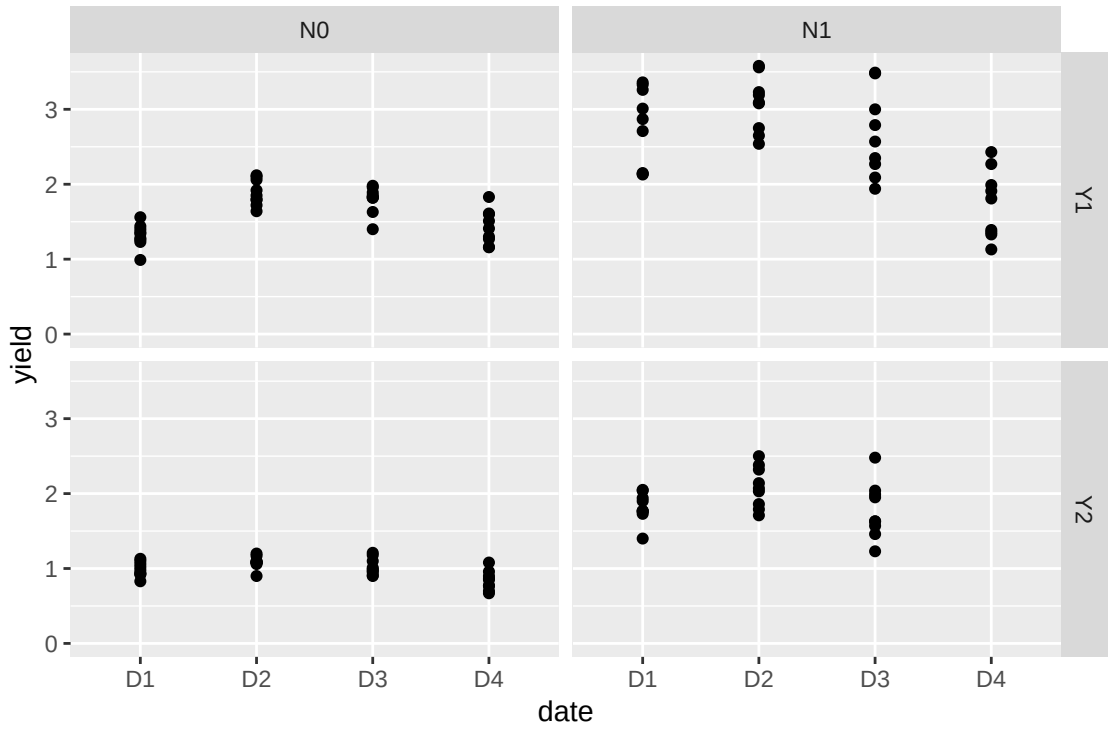
d.) How many interaction terms are involved in the interaction effect of this model? (1)

e.) According to the interaction plot on the following page, it seems that only a part of the interaction terms may be needed from a statistical point of view. Which one looks the most needed, and why? (2)



f.) Use significance tests for the interaction terms to check your answer to problem e.). Give your code and argue with p values to show that your answer was correct. (2)

g.) We now start to investigate the effect of the year. Add the missing yield values for date D4 in year Y2 with nitrogen level N1 to the plot below. (1)



h.) Add the year to the model. Keep in mind that the year effect could also depend on the levels of the nitrogen and the date. Give your R code to fit a suitable model. (1)

i.) Refer to your model from h.). Which (if any) of the interaction effects with the year is significant? Give your R code and argue with p values. (1)

j.) According to your model in h.) (without removing any nonsignificant effects), how high is the average yield for a plot with N1 nitrogen level, at date D4 in year Y2? Give your code and your answer. (2)