EXAM COVER SHEET – Version B, Problem 2

Module:	**D2, Design and Analysis of Experiments**
Date of exam:	19.01.2021, 1.00 – 3.00pm
Duration:	2x 45 min
Type of exam:	Open book: Distributed printed course material allowed, personal notes allowed, laptop allowed, access to Internet allowed, pocket calculator allowed. Any form of oral or electronic communication with other students or persons from outside is forbidden. Furthermore, Videos and Screencasts are not allowed.
Module coordinator:	Lorenzo Tanadini (BFH)

Name of the student: *Clara Iglhaut*

School: *ZHAW*

Venue of exam: *online examination upload/download on Moodle*

Declaration of Independent Work

By taking part to this exam, I hereby affirm that the examination is my own work and that I have not used any sources other than those explicitly allowed for the exam. Furthermore, I have not assisted any other students with their online examination.

Exam Briefing

- Write your name and affiliation on the first page
- Next to each problem, the number of points is indicated in parentheses, e.g. (3). Partial credit can be accredited for partially correct answers.
- The level of significance is 5%. Give numeric results (such as p values) to at least three digits.
- Always include a short reasoning (e.g. I applied a marginal F-test and obtained a p value of … , and therefore I conclude ….“)
- **Report all your answers on this document. Convert it as a PDF file before submission.**

Best of luck!

Problem 2

The **yield_data** data frame contains data on the **yield** of a chemical reaction conducted at different temperatures **temp**. For logistic reasons, the experiment had to be carried out over several days (denoted by **day**). At each day, the sequence of applied temperatures was randomized.

While day-to-day variation in yield is to be expected (due to e.g. different personnel, calibration of equipment, …), the main goal of the experiment is to model the effect of temperature on yield.

Set your working directory appropriately and import the data set using:

```
mydata2 <- readRDS("yield_data.rds")
```

1. Give the **R** code to produce suitable descriptive statistics to describe the data set. (1)

   str(mydata2)

   with(mydata2, tapply(yield, temp, mean))
   with(mydata2, tapply(yield, temp, sd))

   For all days: The highest average yield is produced with the temperature very high, the highest variability in yields has the temperature high.

2. Give the **R** code to produce suitable graphical representations of the data set. What do you observe? (2)

   library(ggplot2)

   ggplot(mydata2, aes(temp, yield)) + geom_point() + facet_grid(~day)

   We can observe a positive linear relationship between the temperature and the yield (in day 1 the temperature labeled high seem to be the only exception from this), higher temperatures seem to bring a higher yield. We can also see some differences from day to day but no obvious block effects of the days can be observed.

3. What is the main goal of the analysis? Give the **R** code to fit a suitable parametric model to this data set. (2)

   We want to asses the fixed effect of the temperature (treatment) on the yield, while accounting for the random effects that are introduced by the days (block). We fit a fixed effect model.

```
library(nlme)
mydata2.lme = lme(yield~temp, random=~1|day, data=mydata2)
```

4. Compare the estimated marginal mean of all treatments to the estimated marginal mean yield for very low temperature. Which ones differ significantly? Give your **R** code, report the p-values and use them to answer the question. (2)

```
library(emmeans)

mydata2.emm = emmeans(mydata2.lme, 'temp')
pairs(mydata2.emm)
```

   The difference of estimated marginal means between very low with middle temperature (p-value = 0.0034) and very low between high (p-value = 0.0003) and between very low and very high (p-value = <.0001) are all significant. Only the difference between very low and low is not significant (p-value = 0.2485)

5. Interpret your regression model: extract the estimated fixed effects and explain in your own words how to interpret the value provided for the intercept and the effect of high temperature. (1)

```
fixef(mydata2.lme)
```

```
fixef(mydata2.lme)['(Intercept)'] + fixef(mydata2.lme)['temphigh']
```

The reference level was set to temperature very low, this means that the intercept gives the average yield for a typical day at very low temperatures with 5.105756 units.

The difference of average yield on a typical day between temperature very low and temperature high is 9.302813 units, this means we get the average yield of 14.40857 units on a typical day for temperature high.

6. Interpret your regression model: how does the estimated standard deviation for the random effects compare with the residuals standard deviation? Based on your regression model, which day was most favorable in terms of yields? (1)

```
summary(mydata2.lme)
ranef(mydata2.lme)
```

The estimated standard deviation for the random effects introduced by the blocks is given with 2.391598 while the estimated standard deviation of the residuals is given with 2.096523. This shows that we should not ignore the block effects otherwise the variability introduced by the random effects will also fall into the variability of the residuals and detecting significant treatment effects will become harder.

The random effects show that day 3 was the most favorable in terms of yields, since it has the highest intercept with 2.3924711.

END of problem 2