

EXAM COVER SHEET – Version C, Problem 1

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: *Memeti Nurdzane*

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 1

You are involved in a project with the goal to reduce food waste due to improper storing conditions. To avoid food spoilage, your team investigated the results of a controlled experiment in which four storing settings were applied to similar and independent batches of food.

The data set **output_data** contains the data collected in this experiment:

- **output** is the quantity of food that is still edible after storage
- **refrigeration** specifies whether the food was stored in a refrigerated chamber
- **vacuum** indicates whether the food was stored in a vacuum chamber

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

```
mydata1 <- readRDS("output_data.rds")
```

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

```
str(mydata1)
head(mydata1)
with(mydata1, tapply(output, list(refrigeration, vacuum), mean))
with(mydata1, tapply(output, list(refrigeration, vacuum), sd))
with(mydata1, tapply(output, list(refrigeration, vacuum), length))
with(mydata1, tapply(output, list(refrigeration, vacuum), summary))
```

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

```
library(ggplot2)
ggplot(mydata1, aes(x=refrigeration, y=output, col=vacuum)) +geom_point()
```

In this case, I observe two block from refrigeration with each 7 treatments. I can see that “no” in Refrigeration have deeper outputs, where Yes refrigerator have a higher outputs.

3. Give the **R** code to fit a suitable parametric model to this data set. Perform an overall F-test, report its p-value and state your conclusion. (2)

```
library(nlme)
mydata1.lm=lm(output~refrigeration*vacuum, mydata1)
summary(mydata1.lm)
anova(mydata1.lm)
```

I applied a overall F-test and obtained a p value of 2.942e-12 for effect “refrigeration” and $p = 0.0008651$ for “vacuum” effect, and therefore I conclude they don’t show a significant effect, because smaller than 5%. You got the same result with `Anova(mydata1.lm, type=2)`.

4. Assess the model assumptions for your final model: explain what you assess, with which method, give your **R** code, discuss the results and state your conclusions. (3)

Testing the Normality we use normal QQ plots and the Shapiro-Wilk test with. $p\text{-value} = 0.6576$. The QQ plot shows some problems with a perhaps slightly heavy down tail and 2 outliers. You can see a wave-form of the dots. Robust or nonparametric methods could be used to control the effect of the outlier. We have two options now. We can abandon normal distribution based methods for e. g. nonparametric, robust or permutation methods, or we can try to find a transformation g such that an ANOVA of the sample $g(Y)$ produces residuals which looks as if they came from a normal distribution.

another thing which has to be discussed: Equal variance : Barlett / Levene test Testing with Barlett Test we have a $p\text{-value} = 0.5492$, with Levene Test a 0.5305, boxplot. If $p > 0.05$ samples have equal variance = homoskedacity

```
library(car)
Anova(mydata1.lm, type=2)
qqPlot(resid(mydata1.lm))
shapiro.test(resid(mydata1.lm))
```

```
plot(mydata1.lm, which = 1)
plot(fitted(mydata1.lm), resid(mydata1.lm), las = 1, xlab = "Fitted values", ylab =
"Residuals"); abline(h=0)
bartlett.test(output~interaction(refrigeration,vacuum), data=mydata1)
leveneTest(output~interaction(refrigeration,vacuum), data=mydata1)
```

5. Can you simplify the model from 3.? Give your **R** code and state your conclusions (1)

We still conclude that the interaction is needed and cannot simplify the model.

END of problem 1