D2 Exam AS 2020 – grading key

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)

```
# simple summary (only 0.5)
# grouped summary (1)
```

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

```
# suitable representation with output, ref, vacuum (1)
# general description (0.5)
# possible interactions (0.5)
```

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)

```
# Linear model with interactions (1)
# P-value F-test = 2.667 e-11 (0.5)
# Reject H0 that all treatments have the same theoretical mean (0.5)
```

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)

```
# Independent observations (1)
# Normality (1)
# Homoskedasticity (1)
```

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

```
# anova(lm1) (0.5)
# the interaction can be removed (0.5)
```

Problem 2

The **yield_data** data frame contains data on the **yield** of a chemical reaction conducted at different temperatures **temp**. For logistical reasons, the experiment had to be carried out over several days (denoted by **day**). At each day, the sequences 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. Is the design balanced? Does it have a special name? (1)

```
# balanced (0.5)
# Randomized complete block = RCBD (0.5)
```

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

```
# Plot with yield, temp, and day (1)
# General description of temp effect (0.5)
# Comments on day effect (0.5)
```

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

```
# main goal: temp effect on yield (day as random effect) (0.5)
# Linear mixed effects model with temp and RE for day (1.5, either all or not
hing)
```

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)

```
# R code provided, functioning (1)
# p-value and answer reported for each comparison (4 x 0.25)
```

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)

```
# Intercept: mean yield for ref temperature (after removing day-effect) (0.5)
# high: difference in average yield from intercept (0.5)
```

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)

```
# SDs: same order of magnitude (0.5)

# day III, based on ranef (0.5)
```