Exam exercise 2: Seasonal wind speed

It is highly recommended that you answer the exam using Rmarkdown (you can simply use the exam Rmarkdown file as a starting point).

The data set for this exam problem contains measurements of the quarterly average wind speed in the years 2001-2019. The measurements were made at the Danish weather station Sjælsmark in the years 2001-2019.

We first read in the data:

```
library(mosaic)
wind <- read.delim("https://asta.math.aau.dk/datasets?file=speed_quarterly.txt",sep="")
head(wind)
```

```
##   quarter     speed
## 1  winter 3.571196
## 2  winter 4.448287
## 3  winter 3.590417
## 4  winter 3.306593
## 5  winter 4.047176
## 6  winter 3.395046
```

We saw in Exam exercise 1 that wind speed measurements are Weibull distributed. However, because of the central limit theorem, when we take the average of hourly measurements from three months, these values will be close to having a normal distribution. Hence, we will assume in this workshop that the quarterly averages are normally distributed.

Throughout the exercise, we use a significance level of $\alpha = 0.05$.

Part I: Comparison of wind speed for spring and autumn

In part I we compare the wind speed in spring and autumn. The questions for Part I should be answered using pen and paper. You may use R as a calculator. You can use the output from `favstats` in your computations:

```
favstats(~speed|quarter,data=wind)
```

```
##   quarter      min       Q1   median       Q3      max     mean        sd
## 1  autumn 2.397207 3.010508 3.097482 3.230105 3.627427 3.094134 0.2532494
## 2  spring 2.877559 3.117040 3.380616 3.572803 3.794384 3.346272 0.2772962
## 3  summer 2.246513 2.699706 2.848324 3.059420 3.438134 2.878752 0.2751879
## 4  winter 2.940046 3.382014 3.805370 4.069144 4.448287 3.721420 0.4084774
##    n missing
## 1 19       0
## 2 19       0
## 3 19       0
## 4 19       0
```

1. Make an F-test to check whether the variance is the same in spring and autumn.

2. Make a t-test to compare the mean wind speed in spring and autumn. What is the null hypothesis and the corresponding alternative hypothesis? Explain how the test statistic and p-value are computed? What is the conclusion of the analysis?

3. Make a 95% confidence interval for the difference between mean wind speed in spring and autumn. What is the interpretation of the confidence interval?

Part II: Comparing all four quarters

In this part we compare all four quarters of the year.

1. Make a box plot of the variable `speed` for each quarter.

```
# gf_boxplot(...~...,data=...)
```

2. Test the null-hypothesis that the mean wind speed is the same for all four quarters by editing the code below. Explain the parameter estimates in the output.

```
# model <- lm(... ~ ..., data = ...)
# summary(model)
```

To find out which quarters differ, we can make pairwise comparisons. We could use the pairwise t-tests as in Part I, Exercise 3. However, there is a problem with this approach:

3. How many pairwise comparisons of quarterly mean values can we make? What is the probability of rejecting a single true null-hypothesis? If we test several true null-hypotheses, explain why the probability of falsely rejecting at least one of them is larger than the significance level.

This phenomenon is known as multiple testing. We can adjust the pairwise t-tests for multiple testing by performing a so-called Tukey's test. It replaces the t-distribution by a different distribution (the studentized range distribution) that takes multiple testing into account.

4. Tukey's test can be performed using the code below. Which quarters differ significantly?

```
#model<-aov(speed~quarter,data=wind)
#TukeyHSD(model, conf.level=.95)
```

5. Compare the confidence interval for the difference between spring and autumn from Tukey's test to the one you computed in Part I, Exercise 3. What is different?