This homework assignment is to be presented on exercise session on **December 10, 2019**. Students should tick in TUWEL problems they have solved and are prepared to present their detailed solutions on blackboard. The solved problems should be ticked by **23:30h on December 9, 2019**.

---

(1) **ANOVA by hand**
Let three quadrupel of data be given as $x = (2, 3, 3, 4)^t$, $y = (3, 4, 4, 5)^t$ and $z = (4, 5, 5, 6)^t$. For a significance level of $\alpha = 1\%$, test the null hypothesis that all data derive from the same distribution using an ANOVA (by hand) as follows

(a) Give the degrees of freedom
(b) Calculate the f-Statistic
(c) Give the rejection region using the following table which shows the 99%-Quantile of the $\mathscr{F}(df_1, df_2)$-distribution

		df_1 2	3	4
	9	8.02	6.99	6.42
df_2	10	7.56	6.55	5.99
	11	7.21	6.22	5.67

(d) Formulate the outcome of the test
(e) If you were very strict with the assumptions of the ANOVA, would you be allowed to apply it at first?

(2) **Processors - Part 1**
The heating of five different types of processors is to be compared. For each type of processor the temperature (in celsius) was measured multiple times (each time a new processor of the same type was used). The technical setup was the same throughout. The data is stored in the file `temperatures.Rdata`

(a) Plot the data using a stripchart
(b) For each group add the mean and the standard error of the mean
(c) From what you see in the plot, would you say that the shift between the groups happened easily if the data was sampled from the same distribution?

(3) **Processors - Part 2**
Use the ANOVA to test the null hypothesis that all data were sampled from the same distribution, on the 5% significance level

(a) Calculate the f-Statistic and the P-value (without using `anova()`)
(b) Double-check your statistics using `anova()`
(c) Formulate a result
(d) Would you say that the assumptions of the ANOVA are fulfilled?

(4) **Processors - Part 3**

Perform pairwise t-Tests ($\alpha = 5\%$) and correct the P-values according to Bonferroni

(a) How many pairs are to be tested?

(b) Give all P-values. Which null-hypotheses are to be rejected prior Bonferroni correction?

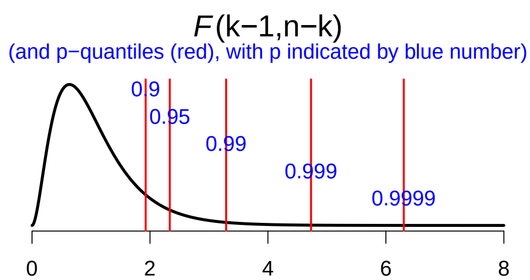
(c) Which null-hypotheses are to be rejected after Bonferroni correction?

(5) **ANOVA by table**

There where k groups compared with an ANOVA. Given the following incomplete output, and a plot of the associated $\mathscr{F}(k-1, n-k)$-distribution, answer the questions below.

```
Analysis of Variance Table

Response: x
          Df  Sum Sq  Mean Sq  F value  Pr(>F)
gr         6    72       ?       ?        ?
Residuals  ?   240       6
```



(a) What are the degrees of freedom?

(b) What are the numerator and and denominator of the f-statistic?

(c) What is the value of the f-statistic?

(d) Using the plot, on which significance level can we reject the null hypothesis?

(e) Using the plot, in which interval do we find the P-value?

(f) How many groups are compared?

(g) How many post-hoc pairwise t-tests are to be performed if all pairs were considered?

(6) **Which statement is correct?**

In the situation of an ANOVA with k groups of k observation each, assume that the null hypothesis $H_0 : \mu_1 = \cdots = \mu_k$ was rejected on the 1%-level. Comment on the following statements.

(a) The rejection boundary is derived from the $\mathscr{F}(k-1, (k-1)k)$-distribution

(b) The test statistic f exceeded the 1%-Quantile of the associated $\mathscr{F}$-distribution

(c) The test statistic f exceeded the 99%-Quantile of the associated $\mathscr{F}$-distribution

(d) The null hypothesis is also rejected on the 5% level

(e) With probability less than 5% we obtain a value of the associated F-statistic that is greater than the observed f, if the null hypothesis is true.

(f) If the null hypothesis was not true, we would have neither made the α- nor the β-error

(g) We can not conclude that at least two expectations μ_i and μ_j are unequal

(7) **Distribution of the P-value in simulations**

Simulate the distribution of the P-value in the (two-sided) two-sample t-test:

Let $X_1, \ldots, X_{20}, Y_1, \ldots, Y_{20}$ be independent random variables with $X_i \sim N(0,1)$ and $Y_i \sim N(d,1)$ for all $i = 1, 2, \ldots, 20$. For each $d \in \{0, 0.25, 0.5\}$, derive P-values in 10000 simulations ($H_0 : d = 0$) and plot them in a histogram of unit area. Comment on your three histograms.