Assignment $\mathcal{N}^{\underline{o}}$ 1

released: 25.02.2020 at 23:30 **due:** 19.03.2020 at 23:59

Task 1 **3 points**

A multiple linear regression model has been estimated to study the relationship between Y = violent crime rate (per 100,000 people), X_1 = poverty rate (percentage with income below the poverty line) and X_2 = percentage living in urban area. Data are collected in 51 cities in the U.S.

The table below reports the output.

	Est.	s.e.	t-value	p-value
Intercept	-498.683	140.988		0.009
X_1		6.677	4.885	0.001
X_2	9.112		6.900	0.001
R^2	0.5708			
SST				
SSReg				
SSR	1841257.15			

(1) Fill in the missing information.
(Please report the formula and computation).

(2) Interpret the coefficient of determination R^2.

(3) Compute and interpret the overall F-test.

Task 2 **5 points**

The table below shows the scores of the first test (maximum score 10 points) in a beginning German course. Students in the course are grouped as follows:

- Group A:
  Never studied foreign language before, but have good English skills
- Group B:
  Never studied foreign language before, have poor English skills
- Group C:
  Studied other foreign language

Group A	Group B	Group C
4	1	9
6	5	10
8		5

(1) Use an adequate method to compare the mean scores of the groups. Specify the assumptions, hypotheses, test statistics and p-values.

(2) Suppose that the first observation in the second group was actually 9, not 1. Then, the standard deviations are the same, but the sample means are 6, 7 and 8, rather than 6, 3 and 8. Do you think that the F-test statistic would be larger, the same or smaller? Explain your reasoning without doing any computation.

(3) Suppose you have the same means as these data, but the sample standard deviations were 1.0, 1.8 and 1.6, instead of the actual 2.0, 2.8 and 2.6. Do you think that the F-test statistic would be larger, the same or smaller? Explain your reasoning without doing any computation.

(4) Suppose you have the same means and standard deviations as these data, but the sample size were 30, 20 and 30, instead of 3, 2 and 3. Do you think that the F-test statistic would be larger, the same or smaller? Explain your reasoning without doing any computation.

(5) In (1), (2), (3) and (4), would the p-values of the F-tests be larger, the same or smaller? Why?

Task 3 **4 points**

The compressive strength of concrete is being studied, and four different mixing techniques are being investigated. The following data have been collected. For each mixing technique, 4 compressive strength measurements (in pounds per square inch) have been recorded.

	Compressive strength			
Mixing	1	2	3	4
1	3129	3000	2865	2890
2	3200	3300	2975	3150
3	2800	2900	2985	3050
4	2600	2700	2600	2765

Does the technique affect the compressive strength?
To answer this question draw comparative box plots and perform an analysis of variance using R.

Task 4 **8 points**

Consider the data set `munich.csv`. The data set contains information on the rent prices of apartments in Munich. The variables in the data set are:

- *rent*: net rent per month (in Euro)
- *area*: living area in square meters
- *yearc*: year of construction
- *location*: quality of location according to an expert assessment (0 = average location, 1 = good location, 2 = top location)

(1) Read the data into R and tell R that location is a categorical variable.

(2) Discuss whether a linear regression model would be adequate to investigate the relationship between *rent*, *area*, *yearc* and *location*.

(3) Estimate the model

$$Y = \beta_0+\beta_1 X_{area}+\beta_2 X_{loc:good}+\beta_3 X_{loc:top}+\beta_4 X_{loc:good}X_{area}+\beta_5 X_{loc:top}X_{area}+\beta_6 X_{yearc} .$$

(4) Run and comment on the model diagnostics.
Hint: use the Rcode
`mod <- lm(rent ~ location+area+area:location+yearc, data=munich)`

(5) Interpret the model results.

(6) Compare the fit of the model in (3) with the model including only the variable *area*.

Please submit your solutions using moodle.
The solutions should also include the R script.
Only one person has to submit the solutions for each group.
Please do not forget to specify the name of all the members of the group.