

## List of tasks

You have three tasks to complete. For tasks 2 and 3 you should also include the SPSS outputs of your analysis (correlation in task 2 and regression in task 3). To provide these outputs either directly export the tables or figures from SPSS or use the printscreen command and paste the results into your file. In case of any technical troubles do not hesitate to contact me. Everything needed to solve the task can be found in the chapters of the textbook by Field that I included as readings to our session. Deadline is **January 10, 2022**. Ideally deliver one text file (docx, pdf or anything similar) with all your answers and solutions.

### Task 1

For each of the 10 enlisted variables decide whether they are nominal, ordinal or scale and shortly explain your decision.

Variables:

- Names of countries
- Percentage growth of GDP of countries
- Military ranks (captain, general etc.)
- Number of members of households
- Blood type of hospital patients
- Age of people measured in years
- Size of countries in square miles
- Educational levels of people (coded as 0 for people without any education, 1 for people with elementary education, 2 for high school graduates and 3 for college graduates)
- Color of hair
- Unemployment rate of countries

### Task 2

The file PresidentCZ.sav contains the results of presidential election in Czech Republic for all districts in the country. The elections use a two round system and if none of the candidates obtains majority of votes in the first round, two strongest candidates continue to the second round where the one with more votes wins. Your task is to calculate two correlations. First, calculate the correlation between the support of Zeman (in the first round) with support of Zeman (in the second round). Second, calculate the correlation between Schwarzenberg (in the first round) with Sobotka. For both calculations choose the proper correlation coefficient and explain why you have chosen this coefficient. Also answer the following questions:

What is the correlation between the support of Zeman (first round) and Zeman (second round)?

What is the correlation between the support of Schwarzenberg (first round) and Sobotka?

Are these correlations significant?

Briefly describe your findings.

### Task 3

The file GreenParty.sav contains a dataset of electoral results of Green party in all regions of a country. The dataset includes one dependent variable (GreenParty) and three independent variables (Graduates, Pensioners and Campaign). Your task is to provide a regression analysis to find out the effects of independent variables on your dependent variable. The characteristics of the variables are as follows:

Region – only the ID of regions (not used for the regression)

GreenParty – percentage share of votes obtained by Green Party in the region (dependent variable)

Graduates – percentage share of people in the region with university degree (independent variable)

Pensioners – percentage share of pensioners in the population of the region (independent variable)

Campaign – a binary variable which has value 1 if the Green Party has campaigned in the region before the election and value 0 if the party has not campaigned in the region before the election (independent variable)

In this task you have to calculate a linear regression model and answer these questions:

Based on  $R^2$  what is the fit of your regression model?

What is the effect of Graduates on the dependent variable GreenParty?

What is the effect of Pensioners on the dependent variable GreenParty?

What is the effect of Campaign on the dependent variable GreenParty?

Which of these effects are statistically significant?

Which independent variable is the best predictor of the dependent variable and based on which indicator can we know that?

In one paragraph explain briefly the main findings based on your regression model.