|  |
| --- |
| **Page 1** |

**Semester assignment**

**MET2010 H2021**

*Perform an analysis of a simple regression model and a multiple regression model. You*

*shall be based on the ESS8 dataset located on Blackboard.*

This semester assignment must be passed in order to sit for the exam. The task can

delivered individually or in groups of 2. **Deadline for submission: Friday 29 October at 14:00.**

The analysis will be carried out in SPSS. The results presented must be interpreted by the student.

The assignment should be about 8-10 pages.

**The multivariate analysis must contain at least:**

- A continuous and / or a binary response variable

- A binary explanation variable

- A variable that is operationalized as a set of dummy variables

- An interaction led

- A higher order member

These parts do not have to be presented in the same table, but can be included as separate

analyzes. SPSS tables can be pasted directly into the document, but it is important that the answer

must be a document with text. Emphasis will be placed on good interpretations and understanding

the results.

**The assignment should also include:**

- A presentation of descriptive statistics (number of data points, standard errors and

average for all variables you analyze)

- The tables must contain coefficients, standard errors, p-value, valid for the whole model and

determination coefficient R 2 .

- A critical discussion of generalization, ie whether the results are valid for

the population (for example, the country or countries' population or working population

population). This is where the probability tests (eg t-test) come into play, and the p-values, which are

an expression of these tests. The thesis must contain a critical assessment of the results. Can

there be reasons why the connection is not credible?

**The dataset**

We will use the 8th round of the European Social Survey (ESS8), which was conducted in 2016.

The dataset is located on Blackboard in SPSS format. Unlike the dataset you have used in

The SPSS exercises, which only had 1760 Norwegian respondents, have this data set over 44,000

respondents in a number of European countries. You choose whether you want to use the whole

the dataset, or limit yourself to a single or a selection of countries. Some variables are specific

for each country, for example «Highest level of education». The data is based on a

|  |
| --- |
| **Page 2** |

questionnaire completed during an interview. Interested parties can find the questionnaire

("Questionnaires" and "showcards") and other details about ESS8 at:

[http://www.europeansocialsurvey.org/data/download.html?r=8](https://translate.google.com/translate?hl=en&prev=_t&sl=auto&tl=en&u=http://www.europeansocialsurvey.org/data/download.html%3Fr%3D8)

Description of all variables is for example here:

[www.europeansocialsurvey.org/docs/round8/survey/ESS8\_appendix\_a7\_e01\_0.pdf](https://translate.google.com/translate?hl=en&prev=_t&sl=auto&tl=en&u=http://www.europeansocialsurvey.org/docs/round8/survey/ESS8_appendix_a7_e01_0.pdf)

**Choice of variables**

Everyone should find a variable. If a continuous variable is selected, it must have

at least 5 values ​​and must be able to be considered as a continuous variable. It's more ideal

it also has more than 5 values.

Here are some examples of variables to choose from:

ccnthum: Climate change caused by natural processes, human activity, or both

gincdif: Government should reduce differences in income levels

imbgeco: Immigration bad or good for economy

stfeco: How satisfied with present state of economy in country

stflife: How satisfied with life as a whole

uemp12m: Any period of unemployment and work seeking lasted 12 months or more

icpdwrk: Respondent in paid work

Then one has to find some -variables that one assumes may have significance for this -en.

For example, it may be natural to think that the variable lrscale (Placement on left right scale)

has an impact on several of the variables above. There are also several possible background variables about

the individuals in question. At least one of these variables must have at least 5 values ​​and be able to

is considered a continuous variable.

Tips for setting up the task

1. Make a short introduction where you say a little about why you choose your dependent variable, and what

you want to find out in your analyzes.

2. Present your variable with descriptive statistics as well as frequency table and histogram.

Tables and graphs can be taken directly from SPSS.

3. Construct a bivariate analysis and present your variable. This should only have a descriptive

statistics. Interpret the results with respect to the specific requirements.

4. Construct a multivariate linear analysis, without interaction links. Present your variables, say

something about why you have chosen each of them and interpret your results.

|  |
| --- |
| **Page 3** |

5. Expand the multivariate analysis with an interaction link. Try to find an interaction

which is important, and assess the extent to which the interaction link improves

the regression model.

6. Expand the multivariate analysis with a higher order term. Try to find a variable there

there is reason to believe that a higher order member can help, and consider the extent to which it

improves the regression model.

7. Use the multivariate model to create an example with prediction. The points above

represents the minimum requirements for the task.

You are free to include other analyzes. Remember, however, that the task should not be on more

than 10 pages, and that there is therefore no room for much irrelevant material. The assignment should include

a general assessment of the results.

In the assessment, emphasis will be placed on whether the student has understood the basics within

regression analysis. Collaborate willingly across groups, but the assignments that are delivered,

should not look too much like each other. If the answer is very poor, the student (s) can

refused to sit for the exam