# GEO Skills: Research Methods & Techniques

## Assignment 1. Application of Linear Regression Analysis

This assignment aims to answer a research question related to explaining costs of energy and water consumption in the household. You need to conduct a simple statistical analysis of a subsample of the WoON Survey 2015, in particular examining correlations between variables and applying a linear regression analysis. You can use lecture notes and the ‘classroom’ exercises in the first week. The results of the data analysis should be used to write a small paper (max 3 pages, including all tables and figures – only excluding do-file appendix). This paper ought to include basic components of an empirical paper, as given below.

While writing this paper, at least one academic article (either self-found or from the list of example articles in the course manual) should be closely read, understood and to serve as an example of how to report statistical findings (i.e. how to present the data clearly, what to include, what to report on, being concise while being transparent, etc.). Make sure to refer to this article in your report.

The assignment is centered on explaining *total costs of energy and water consumption* *in the household which is measured by* the variable *totener.* Costs of energy and water consumptionare assumed to be explained from characteristics of house and household, such as surface house *(gebruiksopp),* housingtenure *(huko),* number of people in household (*AantalPP) and residential location of household (g4\_5).*

**Answering a research question in a small report**

Please report your (only) relevant results in a small paper (max three pages) using (a selection of) the analysis above. The aim of this paper is answering a research question in a well-structured “paper” by examining data. The following work process can guide you in structuring your paper which should have about the format of a paper in quantitative research (see below).

* Introduce the topic, think up a research question and describe your method briefly.
* Formulate your expectations (hypotheses) regarding separate explanatory variables. The estimation results will provide necessary information to reject/accept these hypotheses.
* Report *only* results which contribute to telling a coherent argument and facilitate accountability and replication of your analysis.
* You should compose your own tables and figures from the stata-output to convey information optimally. Tables and/or figures should be well-formatted. Please pay attention to titles, labels, colors, content (when percentages or frequencies) etc. Ideally, your layout should satisfy a ready-to-submit state of an academic paper.
* Pay attention to significance levels, direction/magnitude of effects and possible multicollinearities; How can you interpret the estimation results? Do you have expected results?
* Try to present your results concisely and convincingly.

This assignment must be carried out in groups of three students as published on Canvas, with the help of lecture notes, do-file(s) and online resources.

## Structure of an academic paper in quantitative research

***Introduction***: relevance of the topic and introducing research question and how to answer (your strategy and method)?

***Theory*** (less relevant for this course): positioning in literature and theoretical arguments for the analysis from which a couple of hypotheses are driven.

***Data* *and Method***: description of data and (regression) method.

Introducing variables and presenting descriptive statistics.

**Bivariate analysis:** assessing correlations between dependent variable and separate independent variables (crosstabs)

***Regression analysis***: introducing, presenting and describing estimates and discussing results.

***Conclusions:*** a very sort summary of what has been done to answer the research question and highlighting major findings and their implications for theory. What have we learnt from the analysis? Some reflections on your data, methods, analysis and findings.

### Workflow of data analysis

**Relevant Stata commands:**

describe, summarize, tabulate, histogram, kdensity

Generate, replace

Scatter, lfit, correlation, Regression,

(ib#. ; Estimates store; Estimates table)

**Step 1 inspecting variables**

* Inspect your variables by descriptive statistics (*command: summarize*)
  + *Continue var: summarize, histogram, kdensity,*
  + *Factor var:* tabulate
* Make a scatter plot between costs of energy and water consumption and other continue variables and, plot linear prediction to see correlations.

**Step 2 Preparing variables**

Prepare your dependent variable (*total costs of energy and water consumption*; *totener*) and the independent variables (*gebruiksopp, AantalPP, huko, g4\_5*) for the regression analysis.

* Create new variables from originals if necessary; for example to exclude potential outliers, to make interpretation easier or to avoid skewed distribution.

**Step 3 Descriptive statistics (to report)**

* Calculate descriptive statistics for your variables.
* Compose a table for descriptive statistics, including ***only*** necessary information.

**Step 4 Regression analysis (to report)**

* Run linear regression model(s) using the explanatory variables prepared in step 2.
* Compose another table from your regression output(s), including ***only*** necessary information.

Please add your *do-file* to the paper as appendix.

|  |
| --- |
| Advanced (optional): using two additional explanatory variables  Choose two other explanatory variables that are potentially relevant to explain energy costs.  Run a new regression model including these two variables in addition to the earlier variables. Discuss the results. Are these estimates as expected? |

**Assessment Criteria**

Assessment of the quality of your work will be based upon:

* Meeting the submission criteria
* Efficient use of literature
* Structure and format of the paper
* Correct use of variables
* Transparency / replicability
* Presentation of results
* Interpretation of results
* Demonstration of understanding of statistical analysis
* Creativity to answer the research question
* Critical reflections

**Submission deadline**: **Tuesday 8th of September 2020 at 10:00**. Please post your paper (as a single document and do-file in appendix) through Canvas (box: Assignment 1).