Regression Analysis for Educational Research 7.5 Credits
Study Guide and Preliminary Schedule
Fall 2018
Regression Analysis for Educational Research

Welcome to the course!

This study guide is meant to serve as a complement to the course syllabus and as a support for participants to plan and manage your study during this QRM course of *Regression Analysis*. It contains general information about the course contents, structure, schedule, and assignments and examination. During the course, participants will be given more detailed information concerning the on-going course activities, schedule and assignments on our online learning platform, GUL ([http://www.gul.gu.se](http://www.gul.gu.se)), and we therefore strongly encourage you to be active on the course platform and follow the updates so you do not miss out on anything!

**General information about the course**
The course comprises 7.5 higher education credits, corresponding to five weeks of full-time studies. The course runs with 50 % study pace during the second half of the autumn semester (November-December)

The course follows a blended learning model that starts with three days on campus teaching and training on the 12-14th, November, 2018 in Gothenburg, and thereafter 4 weeks of online self-regulated learning. For the campus part, we start at 13.00 on the 12th of November and end at 15.00 on the 14th of November (see schedule). During the 3 campus days, we will have lectures briefly introducing the contents of the course, hands-on practice in performing statistical analyses with the software SPSS. After these important introductory days, our lectures, seminars, and communication will be carried out on the course webpage on the learning platform GUL used by Gothenburg University ([https://www.gul.gu.se](https://www.gul.gu.se)).

The participants will be given a username and a password to log into the course website on GUL as a course member. Once you log in to the course website, you will get full access to all the resources of the course whenever and wherever you are. On the course website, you can also exchange ideas, ask questions, and discuss different issues with both your fellow students and the course leaders. Assignments and examination task are also posted on the course website. A time slot will be open for submission of each assignment and examination task. After deadline, it will no longer be possible to submit the assignment/examination task online. A participant will then have to contact the course leaders for manual submission. If you have questions concerning any aspect of the course, please do not hesitate to contact us course leaders!

**Course leaders/teachers:**
Kajsa Yang Hansen e-post: kajsa.yang-hansen@ped.gu.se
Magnus Wikström e-post: magnus.wikstrom@umu.se
**Overall structure and schedule**

Our ambition with this course is that the participants will learn to handle different types of regression models to answer research questions that involve causal claims. In order to be able to make the right model choice, and to properly apply these regression models, basic principles and assumptions of the models need to be understood. Participants are therefore required to have sufficient knowledge in basic statistics, such as, measurement level of a variable, the concepts of variance, covariance, standard error, hypothesis testing, confidence interval, p-value, and normal distribution, which makes up the content of the QRM Basic statistics course.

The course starts Monday, the 12th of November, 2018, with two and half mandatory on-campus days in Gothenburg, where lectures will be given covering the core contents of the course, work with educational data in SPSS and practical assignments. After these days, the course continues as an online course on GUL. There, the students will be able to listen to short pre-recorded lectures elaborating on course content not covered during the on-campus lectures. Each lecture will be accompanied by practical hands-on analysis tasks. At the course website on GUL, students will also be able to submit course assignments, and engage in discussions with other participants, as well as to take part in seminars and examinations.

The course content will primarily be distributed as follows:

I. Introducing correlation measures when modeling relationship between two variables, controlling for the effect of a third variable. This leads us to the method of predicting a continuous outcome variable by one or several predictors, i.e., independent variable or explanatory variables, and to see the effects of these independent variables on the dependent variable in linear regression analysis. We will learn the model estimation method (the least square method), model fit evaluation, as well as interpretation of model parameters. The different assumptions of a multiple linear regression model, such as linearity between dependent and independent variables, normality, independency, homoscedasticity of residuals will be discussed. The multicollinearity problem will be introduced and its consequences and possible solutions to the problem will also be discussed.

II. General linear model (GLM) will be introduced when one wants to compare mean differences among several dummy coded predictors of a continuous outcome variable. A typical use of this GLM is to examine the effects of interventions on the outcome variable in experimental designs. Belonging to an intervention group (Yes=1) or not (No=0) will be brought into the General linear model as predictors to study the mean differences in the outcome variable for individuals in different intervention groups. Such intervention effects can be adjusted for one or more predictors (i.e., covariates) as well as factors (factorial design, including the main factorial effects and interaction effects).

III. Analytical methods with categorical dependent variables will be introduced. We start with examining the association between two or more categorical variables in contingency tables and chi square tests in a loglinear model, for example, we predict a
categorical outcome with only categorical predictors. We extend this method by allowing both categorical and continuous predictors in the model in logistic regression analysis. Issues such as model assumptions, model fit evaluation, relative contribution of each independent variable to the dependent variable, model specification and parsimony will be touched upon.

IV. Examination (see below)
<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Content</th>
<th>Teacher</th>
<th>Room</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Week 46</strong></td>
<td></td>
<td><strong>Introduction of the course;</strong> 1. Correlation; 2. Linear regression.</td>
<td>KYH</td>
<td>MW</td>
</tr>
<tr>
<td>Mon 12/11</td>
<td>13-17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tue 13/11</td>
<td>9-17</td>
<td><strong>General Linear Model (GLM):</strong> 1. Using GLM to compare several means; 2. Using GLM to compare means, adjusting for other predictors; 3. GLM that has 2 or more categorical independent predictors (factorial designs).</td>
<td>KYH</td>
<td>MW</td>
</tr>
<tr>
<td>Wed 14/11</td>
<td>9-15</td>
<td><strong>Methods handling categorical outcomes:</strong> 1. Chi-square and loglinear analysis; 2. Logistic regression.</td>
<td>KYH</td>
<td>MW</td>
</tr>
<tr>
<td></td>
<td>15.00</td>
<td><strong>Ajourn</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Week 47-48</strong></td>
<td></td>
<td><strong>Students continue working with/repeating SPSS tasks from Campus days, focusing on chapter 8-9 correlations and linear regression. Study video clips/lecture material that are available on GUL for this week. Read literature and work with assignment</strong></td>
<td>WM</td>
<td>GUL</td>
</tr>
<tr>
<td>19-30/11</td>
<td></td>
<td></td>
<td>KYH</td>
<td></td>
</tr>
<tr>
<td>30/11</td>
<td></td>
<td><strong>Submit practical assignment I (see below)</strong></td>
<td>GUL</td>
<td></td>
</tr>
<tr>
<td><strong>Week 49-50</strong></td>
<td></td>
<td><strong>Study lecture material available on GUL, elaborating on course content on GLM (chapter 12-14). Students work independently with reading literature and work with assignment</strong></td>
<td>WM</td>
<td>GUL</td>
</tr>
<tr>
<td>03/12-14/12</td>
<td></td>
<td></td>
<td>KYH</td>
<td></td>
</tr>
<tr>
<td>14/12</td>
<td></td>
<td><strong>Submit practical assignment II (see below)</strong></td>
<td>GUL</td>
<td></td>
</tr>
<tr>
<td><strong>Week 51-52</strong></td>
<td></td>
<td><strong>Study lecture material available on GUL, elaborating on course content Categorical data analysis, loglinear, logistic regression analysis (chapter 19-20). Students work independently with reading literature and work with assignment</strong></td>
<td>WM</td>
<td>GUL</td>
</tr>
<tr>
<td>18-29/12</td>
<td></td>
<td></td>
<td>KYH</td>
<td></td>
</tr>
<tr>
<td><strong>Week 1-2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Throughout the course, there will be supervision available in GUL at given time points. The above is a preliminary schedule that may be subjected to change. Students are advised to actively participate in campus days and be active on the course’s GUL course website where changes in tasks and schedule, if any, will be informed. To be able to follow the course and perform the analyses, the participants will need access to the SPSS software. It is also recommended that the participants start reading the course literature and get familiar with the data management procedures in the SPSS program.

**Exercises and dataset:**

For teaching and hands-on practices, data set from PIRLS (the Progress in International Reading Literacy Study) will be used. PIRLS large-scale comparative study focuses on reading proficiency among students in Grade 4 around world. More specifically, we will use questionnaire data and reading results data from all Swedish students who participated in PIRLS 2016. The dataset is available on the GUL course site (under Regression Analyses for Educational Research → Document).

**Examination**

The course examination consists of three individual assignments and one final report. Two of the assignments consist of practical analyses in SPSS using statistics reviewed in lectures and used in hands-on practice. The third assignment is a critical reading of an article that applies basic statistical methods. The final report is to be presented as a short scientific article where students are supposed to independently perform analyses on a set of data, present and discuss their findings using the APA format. More detailed instructions for the assignments will be available on the GUL course website.

**Assignment I** (last day for submission 30/11).

Using the PIRLS 2016 dataset to specify a multiple regression model to predict reading achievement of students. You can choose your own predictors, i.e., explanatory variables or independent variables, according to your theoretical framework (theoretical model). If you have only dichotomous predictors, you may consider using a General linear model. Your task will be to:

1. Set up the multiple regression model;
2. Test model assumptions;
3. Interpret the model results;
4. Report your model results in APA standard;
5. Discuss the advantages and disadvantages of the method that you applied.

Assignment II (last day for submission: 14/12)

Using the Evaluation through Follow-up (UGU) dataset to predict the probability to enter higher education for individuals. You can choose your own predictors, i.e., explanatory variables or independent variables, according to your theoretical framework (theoretical model). If you have only dichotomous predictors, you may consider using a General linear model. Your task will be to:

1. Set up the multiple regression model;
2. Test model assumptions;
3. Interpret the model results;
4. Report your model results in APA standard;
5. Discuss the advantages and disadvantages of the method that you applied.

Assignment III (last day for submission 04/1, 2019, when we will also discuss in a web seminar on 07/1, 2019):

In this assignment, your task is to perform a critical read of an article that applies regression analysis methods to analyze data (articles provided by the course leader). After having read the article, write an analysis focusing on the following aspects (maximum 3 pages):

- What are the research questions and the hypotheses?
- Is the method chosen appropriate for testing the hypotheses? Why?
- Is the method chosen appropriate given the characteristics of data? Why?
- Are the parameter estimates of the model interpreted properly?
- Are the causal inferences drawn in the study valid?

The analysis of the article should reflect upon the course literature. Your assignment will be reviewed and discussed by a peer student.

Final examination task (21/01, 2019)

For your final examination task, you may choose from one of the below two alternatives:

Alternative I: data analysis with a given dataset

The International Association for the Evaluation of Educational Achievement (IEA) is an independent, international cooperative of national research institutions and governmental research agencies. It conducts large-scale comparative studies of educational achievement and other aspects of education. The IEA’s Progress in International Reading Literacy Study
(PIRLS) collects data to provide information on reading literacy achievement of fourth-grade students, as well as questionnaire data to investigate the experiences young children have at home and school in learning to read.

Your task is to apply one the regression methods that have learned in the current course and examine a research question that you yourself find it more interesting. If necessary, you need to compute new variables, or recode variables, so that the variables involved are at the right measurement level.

Your report (max 6 pages) should include the following:

1. Very brief about the research field and theory(ies);
2. Research question(s) and hypotheses;
3. Data, variables and method description;
4. Descriptive statistics for your variables;
5. Short description of the method;
6. Presentation and interpretation of major results;
7. Conclusion and discussion.

Alternative II: data analysis with your own data

The participants can supply their own data materials to complete the Alternative I (consult course leaders first for the suitability of your own data)

Written tasks should be written with Times New Roman, 12pt. All individual assignments and tasks are to be uploaded to the course page on GUL. To pass the course, all mandatory assignments must have been completed, submitted, and given a pass grade. Students who are absent from mandatory group seminars will be given an individual task/corresponding.

As noted above, all further information will be available in GUL, where we teachers also be will be available on a regular basis throughout the course. The course literature is stated in the literature list; below we also have some alternative/reference literature for the interested student.

We hope that the course guide will provide you some brief information of the course. With that, we hope that it can help you to navigate through the course and successfully gain knowledge and confidence in understanding regression analysis and employing these methods to investigate research questions of interest.

Once again, welcome to the course! We hope that you will have an enjoyable and rewarding experience together with peers and teachers within the QRM research school over the next few weeks!

----------