Федеральное государственное автономное образовательное учреждение высшего образования
"Национальный исследовательский университет "Высшая школа экономики"

Факультет социологии

Программа дисциплины

Многоуровневый анализ данных / Multi-level regression analysis

для направления подготовки магистра 39.04.01

программы «Сравнительные социальные исследования»

Автор программы: В.В.Костенко, к.соц.н. vvkostenko@hse.ru

Одобрена на заседании совета магистерской программы «Сравнительные социальные исследования» «___»_____ 2018 г

Руководитель магистерской программы:

Рекомендована секцией УМС: Профессиональной коллегией по направлению «Социология» «___»_____ 2018 г.

Москва, 2018

1. Course Description

a. Title of a Course

Multi-level regression analysis

b. Pre-requisites

The basic competence in statistics (such as sampling principles, scaling, linear and binary logistic regression models) ar e prerequisites for the course.

c. Course Type

Compulsory

d. Abstract

Analysts have to deal with hierarchical data structures increasingly more ofte n. In particular, one encounters them in the context of cross-country comparisons. Classic regression methods applied to such data result in biased e stimates. There are several ways to deal with this problem. One popular method is the multile vel regression. This course covers the basic tenets of this method with applications to international survey research data.

The course assumes the student's knowledge of linear and binary logistic regression modeling. The workload of the course includes participation and preparation for classro om activities, use of open datasets for analyzing individual and country effects in cross-country perspective, and an individual project in essay form that could be developed into a journal article.

2. Learning Objectives

- To study principles of multilevel modeling, its advantages and limitations.
- To be able to read and interpret results of multilevel modeling usin g formulae and regression tables.
- To learn analyzing sociological and political issues using multilevel research design.
- To employ real co mparative survey data for studying sociological and political issues in multilevel paradigm.

3. Learning Outcomes

- To understand the basic principles of multilevel modeling.
- To be able to acce ss the results of multilevel modeling and interpret them statistically and sociologically.
- To model individual cases within groups choosing the best model.
- To apply multileve l modeling techniques in practical research.

4. Course Plan

Topic	Total	Directed Study		Self-directed Study
		Lectures	Tutorials	
1. Introduction. The idea of hierarchical modeling. Pre-requisites for multilevel modeling. Alternative s to multilevel modeling.	8	2	2	4
2. A basic (empty) multilevel model. Intra -class correlation coefficient. Individual - level predictors. Group- level predictors. Fixed intercept. Fixed slopes.	16	2	2	12
3. Varying intercepts. Varying slopes. Cross-level interaction in multilevel models.	16	2	2	12
4. Multilevel binary logi stic regression.	16	2	2	12
5. Mid-term presentation of the individual project prop osals.	16	2	2	12
6. Testing and model specification, model comparisons.	16	2	2	12
7. Discrete dependent variables in multilevel models.	16	2	2	12
8. Missing data issues in multilevel structures.	16	2	2	12
9. Imperfect hierarchies, cross-classified data structures.	16	2	2	12

10. Class discussion of issues in individual projects prior to submission.	16	2	2	12
Total study hours	152	20	20	112

Topic 1. Introduction. The idea of hierarchical modeling. Pre-requisites for m ultilevel modeling. Alternatives to multilevel modeling.

This class is designed to expla in the limitations of multilevel research, and th e opportunities it gives. Basing on Chapters 2 and 3 from Snijders and Bosker "Multilevel Analysis, 2nd edition" book, we work on issues of statistical treatment of clustered data, including multi-level propositions, micro-macro relations, the concept of levels of data and hierarc y as an organizational principle of certain types of data. We overview the most comm on situations when multilevel modeling is used in social and political science, and figure out the principle of grouping, nestedness, and cross-classification. We discuss what happens if w e do not apply multilevel modeling to hierarc hically organized data (taking it as pooled), what to do if the number of second level units is too low, if the sizes of groups vary strongly etc.

Topic 2. A basic (empty) multilevel model. Intra -class correlation coefficient. Individual - level predictors. Group- level predictors. Fixed intercept. Fixed slopes.

At this class we start with the basic regression formula, and unfold the mathe matical backstage of multilevel modeling by co mplicating it stepwise. We introduce the idea of dividing explained variance between levels and discuss intra-class correlation coefficient as a measure of explained variance at the second level. We also train to distinguish predictors between levels. Basing on Chapter 4 from Snijders and B osker, we estimate a regression model with fix ed effects only and learn how to interpret its coefficients from the output, and match those coeffic ients to the regression formula. Using *stargazer* and *sjPlot* packages we start working on better visualization of the tables, paying special attention to non-default solutions and modificatio ns of the tables received.

Topic 3. Varying intercepts. Varying slopes. Cross-level interaction in multile vel models. We continue complicating the regression formula by randomizing intercepts first, and slopes second, and accessing the interpretation of the coefficients provided by R. We deploy *lme4* package in R to get calculate the regressions. Cross-level interaction effects are discussed and

visualized with *sjPlot* packag e. We discuss the topic of marginal effects and work extensively on interpretation of the results received. This class is based on Chapter 5 from Sn ijders and Bosker. The students are supposed to practice the skills acquired in their first home assignment that they submit by class 4.

Topic 4. Multilevel binary logistic regression.

We return to the basic logistic regression formula to repeat the concept of odd s, log odds, logit, etc., and make it more complex by adding multi-level perspective to it. We estimate a model step-by-step in the same mann er we used to practice with the basic multilevel model to see the difference. Some time is dedi cated to discussing better ways to report logit models (choosing between odds and estimates in the tables, for example). Chapter 17 can be use d for reference. Students are supposed to practice this type of modeling in their second home assignment that has to be submitted by class 7.

Topic 5. Mid-term project proposal presentation.

See section 8 of this syllabus for details of the presentation.

Topic 6. Testing and model specification, model comparisons.

We work extensively on methods of quality assessment, as in multilevel mod els this process requires more sophistication due to numerous assumptions and complicated nature of modeling. Chapters 6 and 10 give a sense of the major issues here. We also extensively discuss the process of modeling, and recommend to follow the principles suggested by Snijders and Bosker starting from individual level, then adding second-level predictor, followed by interaction effects, and continue with randomization of slopes. Alternatives are also discussed.

Topic 7. Discrete dependent variables in multilevel models.

Being a more advanced topic, discrete variable multilevel analysis unites multinomial and ordered regressions. We discuss complications of such modeling, and apply packages *MCMCglmm* and *ordinal* to th ose issues. As the former implies Bayesian approach, and is somewhat more complicated, we extensively interpret the results received in class. For ordinal models, we discuss the conce pt of thresholds, equidistance, and the assumptio n of latent concept behind the variable. Chapters 17.3 and 17.4 can be used for reference.

Topic 8. Missing data issues in multilevel structures.

In this session we study differences between MCAR, MAR, and MNAR missingness patterns, and apply Little's test. We work on multiple imputation techniques using *Am elia* package, and its extensions for hierarchically structured data. Following Chapter 9 from Snijders and Bosker, we learn why list-wise deletion leads to biased results, and how to cope with the issue of missingness in multilevel designs. We also work with *mice* package, deploying md.pattern command to get more sense of missingness patterns of the datasets. Using flexibility of *mice,* we proceed with imputing multiple datasets and extracting estimates from the resulting complete samples.

Topic 9. Imperfect hierarchie s, cross-classified data structures.

Here we address the issue of imperfect hierarchies, when, for example, one pe rson belongs to two groups that are non-nested, but rather residing on the same level (e.g. mig rants who can be grouped to their country of origin as well as to their country of residence). Em ploying Chapter 13 from Snijders and Bosker, learn how to estimate, visualize, and interpret m odels with multiple group membership.

Topic 10. Class discussion of issues in individual projects prior to submission .

This is a wrap-up session in an Q&A style. Students have an opportunity to a sk their questions regarding all stages of modeling and reporting their results prior to final pape r submission. Students are welcome to show their results so that their classmates can help w ith interpretation. Students are also encouraged to show some of their tables to see how other pe ople perceive them. It helps figuring out the best way of presenting the results of complicat ed models. Graphs are also discussed.

5. Reading list

a. Required

Mandatory:

Snijders T., Bosker R. Multile vel Analysis. An introduction to Basic and Advanced Multilevel Modeling (2^{nd} edition). Sage, 2012.

b. Optional:

Gelman A., Hill J. Data Analysis Using Regression andMultilevel/Hierarchic al Models. Chapters 1, 11-14.

6. Grading System

- 2 homework assignm ents (25%, 12.5% per each assignment).
- Mid-term presentation of the individual project proposal (25%).
- Individual research project essay in English (final project): 3000-350 0 words (50% of total grade).
- Late assignments will be graded down.
- Plagiarism will result in failure. Papers submitted for other classes c annot be reused.

7. Rounding the final g rade

If the final grade is non-integer, it is rounded according to algebraic ru les. If has a half (.5) at the end, we are rounding upward.

8. Guidelines for Knowledge Assessment

In-Class Participation and Attendance:

Participation is require d and expected. During the lectures active participation in discussion is needed, so the students are supposed to read the textbook . Seminars will include working with data.

Homework

Homework will be given twice for long intermissions between the classes. The first homework will be on a multilevel model with a normally distributed dependent variable. The second will be on a multilevel binomial logistic regression. Students will have to repeat modeling that h ad been studied during the class on a different dataset, modifying the given script and give statistical and sociological interpretations of the results received. Students are supposed to submit their script and their descriptive text as homework.

Mid-term presentation: individual project proposals

Students are supposed to show their preliminary work on the final project on multilevel modeling. Only brief overview of theory or literature review is required at this stage. The

focus of presentation is research design and operationalization of concepts, especially the limitations. The prese nter is supposed to prepare a PDF or a PowerPoint presentation of 6 to 7 slides with the following content:

Slide 1: The name of the student and the project title

Slide 2: Research puzz le (what is of interest/understudied/puzzling)

Slide 3: Brief overview of the theoretical framework and/or contemporary literature on the proposed topic

Slide 4: Chosen dataset, potential advantages and limitations when using that dataset

Slide 5: Proposed second-level variables, sources of those variables

Slide 6: Operationalization of the proposed concepts. Variables that the student plans to use (including their coding)

Slide 7 (optional): D escriptive statistics on the variables of interest (distributions, correlations, etc.)

Individual research project essay in English (final project)

The final work for the course is an essay of about 3000-3500 words in English related to sociology or political science in cross-national comparative perspe ctive conducted in multi-level statistical paradigm. This text is intended to be a draft for an article that can be published in a peer-reviewed journal after some revisions. The essay is supposed to include an abstract, an introduction, a theoretical section and\o r literature review, hypotheses derived from the theory, some methodological discussion, a model built on one of the cross-country datasets, and a results section. Discussion section should follow to wrap up and emb ed the empirical results into the existing d iscourse. The most important aspects t o be graded are the creativity of the r esearch idea, the operationalization, proper modeling, and clear understanding of the li mits of research.

Please do not hesita te contact me via email at veronika.koste nko@gmail.com or vvkostenko@hse.ru with “Multilevel modeling” as the topic on any issues considering this course.

9. Timeline for control

- Homework assignment 1 has to be submitted before Class 4 (Multilevel binary logistic regression)

- Homework assignment 2 has to be submitted before Class 7 (Discrete dependent variables in multil evel models)
- Mid-term presenta tion has to be submitted 1 week after the lecture on Multilevel binary logistic regr ession
- Final essay has to be submitted 1 week prior to the date of final exam published by the study office

10. Methods of instructio n

The course consists of lectures\seminars and labs. Labs will follow lectures a nd will familiarize the student with multilevel re gression analysis in cross-cultural research.

11. Special Equipment and Software Support

R is preferable softwa re used throughout the course. Seminars are based on R scripts and real life open-source datasets (World Values Survey, European Social Survey, European Values Study, Arab B arometer and others).