

Course syllabus
«Methodology and Research methods in Sociology: Quantitative Research methods»

Approved by
Programme Academic Council
Protocol Nr. 01 from 25.06.2018.

Author	Anna Almakaeva
Number of credits	4
Class work (hours)	52
Individual work	100
Year	1
Discipline's format	Without using online courses

I. GOAL, RESULTS OF STUDY AND PRE-REQUISITES

The course aims to provide students with understanding of key concepts and methods of modern statistical data analysis. It gives an overview and practice of the skills necessary for conducting independent research with quantitative survey data, using R software. The course also puts these skills into the broader academic context by reviewing how statistics are used in published scientific journal articles.

As a result, students should:

Know:

- key concepts of statistics;
- main methods and techniques of statistical data analysis;
- main procedures of data transformation and data analysis using R;
- academic standards of reporting results and professional scientific ethics.

Be able:

- to choose correct statistical methods and procedures according to the research questions;
- to interpret and present the results of data analysis in oral and written form;
- to analyze secondary data at the level required for an MA thesis;

Have skills in:

- using R Studio for statistical data analysis;
- interpreting results and writing research papers;
- evaluation of the quality of published research papers.

This discipline requires following knowledge:

- Basic Statistics and Introduction into „R“
- Contemporary Sociological Theory

Main ideas of the discipline might be applicable in following courses:

- Research Workshop
- Research Seminar
- Introduction to SEM
- Multilevel regression analysis

- Bayesian Statistics
- Applied Research on Inequalities: Race, Gender, Migration and Ethnicity

II. THE CONTENT OF THE COURSE

Topic 1. Statistical inference in social sciences

Descriptive and inferential statistics. Random and non-random sampling, the case of “Literary Digest”. Normal distribution and z scores. Logic of statistical inference and Central limit theorem. Confidence level and confidence intervals.

Topic 2. T-tests, One-way ANOVA

Steps of hypothesis testing, Two-tailed/one tailed assumptions, Type I/Type II errors. One sample t-test. Independent samples test, paired samples test. T-distribution for small samples. ANOVA (Analysis of variance), F-distribution, post hoc tests. Equality of variance tests and tests for normal distribution.

Topic 3. Non-parametric analogs of T-tests and one-way ANOVA

Differences between parametric and non-parametric methods. Signed-Rank Test, Mann-Whitney U-test, Wilcoxon signed rank test, Kruskal-Wallis test, post hoc tests for non-parametric methods.

Topic 4. Pearson's chi-square for contingency tables

Properties of Pearson's chi-square, independence of two variables, expected and observed frequencies. Chi-square distribution, Yates correction for 2x2 tables. Limitations of chi-square test.

Topic 5. Correlations

Association between two variables, types of dependence. Pearson's r for metric scales, Pearson's R for metric scales and test of significance. Spearman's rank correlation coefficient, Kendall's rank correlation coefficient. Partial correlation.

Topic 6. Linear regression

Bivariate regression, quality of bivariate model. Multiple regression and quality of multiple regression model. Interpretation of regression coefficients, significance of regression coefficients. Standardization and interpretation of standardized coefficients.

Topic 7. Regression diagnostics: heteroscedasticity, multicollinearity, non-linearity, etc.

Testing regression assumptions: linearity, multicollinearity, heteroscedasticity, normal distribution of errors, autocorrelation, outliers.

Topic 8. Categorical independent variables in linear regression and interaction effects

Regression with dummy variables, interpretation of coefficients. Regression with categorical and nominal independent variables. Interaction effects, interpretation of interaction effects. Limitations of traditional regression tables for interpretation of interactions effects.

Topic 9. Binary logistic regression

Key concept of binary logistic regression: probability, odds, odds ratio, logit. Interpretation of coefficients. Quality of a model.

III. GRADING

Home assignments – 10%

Written paper - 30%

Quiz – 10%

Written exam – 50%

Final grade= $HA*0.1+Q*0.1+WP*0.3+WE*0.5$

If the final grade is non-integer, it is rounded according to algebraic rules. If has a half (.5) at the end, we are rounding upward. For submission home assignment after deadline – 50% downgrade.

IV. Grading Tools

Final essay rules:

7 pages long, font 14, Times New Roman, typed, 1.5-spaced, standard margins, not including abstract and appendices of tables. Using any type of regression is necessary.

Structure of the essay:

This paper, like all academic research papers, must contain the following sections:

1. Abstract
2. Introduction
3. Literature review
4. Methods
5. Findings
6. Discussion/conclusion
7. References.
8. Appendix

ABSTRACT: A one-paragraph summary of the research question and (only) main findings. (on a separate page, not counted in the page count).

INTRODUCTION: Contains the research question, establishes sociological relevance (ie the “why”, may also include social or policy relevance) of the topic and novelty.

LITERATURE REVIEW: Examines research question in terms of the theory that generated it, and reviews existing sociological research on the research question. The literature review generally includes a mention of how the current research replicates previous research, contradicts previous research, or somehow modifies or extends previous research. At the end of this section, you must clearly list the hypothesis or hypotheses to be tested. It should be obvious how the hypotheses fit with the themes written in the literature review.

DATA & METHODS: This section briefly describes the dataset. This section should explain how the research question is operationalized into testable hypotheses. This section should clearly state the concepts to be tested in the hypotheses, as well as clearly label the

independent variables, the dependent variable, and any intervening or control variables also to be included. This section should describe which variables are measuring which concepts. Lists of descriptive statistics on all variables will be included in Tables in the Appendix.

FINDINGS: This is the section in which you present your findings and explain which tests were conducted in R. This is where written sentences report and discuss the most interesting findings. This section should address whether the analysis of data confirms your hypotheses.

DISCUSSION/CONCLUSION: This section should BRIEFLY summarize the findings. It should also explain how your findings represent a contribution to the literature reviewed at the beginning of the paper. Discuss whether your findings support or contradict some previous research. This section should briefly discuss limitations of the findings, especially in terms of your methods. Evaluate whether the concepts could have been measured differently or different tests run. This section may include suggestions for future research or implications for policy or both, but future research or policy implications are not required.

REFERENCES: All references cited in the text must be listed in the bibliography. This paper must contain at least 5 peer-reviewed journal articles but may also contain additional sources.

APPENDIX: This section contains additional tables and materials.

GRADING CRITERIA OF FINAL PAPER

1) Content

A) Logical Reasoning (Does the paper demonstrate understanding of key ideas? Can the reader easily see the answers to these questions in the paper?)

What concepts are being measured?

How are concepts being measured?

Why are these measures chosen?

What hypotheses are being tested?

Where do these hypotheses come from? (i.e connection to literature)

How will these hypotheses be tested?

Does testing these hypotheses address the research question as stated?

Do variables chosen measure the hypotheses as stated?

Are the hypotheses confirmed or rejected?

B) Skill at interpreting statistics (Does the paper demonstrate understanding of key ideas, such as, including but not limited to the questions below?)

What does r^2 mean?

Are results statistically significant?

What do linear regression results mean?

What do logistic regression results mean?

2) Form

Does the paper follow the paper sections and content described in the assignment?

Does the paper have nice visualization and correct citation?

V. Sources

5.1. Main literature:

1. Agresti, A. (2013). Categorical Data Analysis. 3. Wiley. URL <https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=1168529&query=data+analysis> Proquest
2. Lynch, S. M. (2013). Using Statistics in Social Research: A Concise Approach. Springer Science & Business Media. URL <https://www.springer.com/gp/book/9781461485728> Springer Link
3. Miah, A. Q. (2016). Applied statistics for social and management sciences. Springer. URL <https://link.springer.com/book/10.1007/978-981-10-0401-8> Springer Link
4. Schumacker, R., & Tomek, S. (2013). Understanding statistics using R. Springer Science & Business Media. URL <https://www.springer.com/gp/book/9781461462262> Springer Link.
5. Stowell, S. (2014). Using R for statistics. Springer Science & Business Media. URL <https://link.springer.com/book/10.1007/978-1-4842-0139-8> Springer Link

Additional literature:

1. Othmar W. Winkler (2009). Interpreting Economic and Social Data: A Foundation of Descriptive Statistics. Springer Science & Business Media, URL <https://www.springer.com/gp/book/9783540687207> Springer Link
2. May, T. (2011). Social research. McGraw-Hill Education (UK), URL <https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=729519&query=social+research> Proquest
3. Wickham, H. (2016). ggplot2: elegant graphics for data analysis. Springer Science & Business Media, URL <https://link.springer.com/book/10.1007/978-3-319-24277-4> Springer Link
4. Wiley, M., & Wiley, J. F. (2016). Advanced R. Apress. URL <https://link.springer.com/book/10.1007/978-1-4842-2077-1> Springer Link
5. Zuur, A., Ieno, E. N., & Meesters, E. (2009). A Beginner's Guide to R. Springer Science & Business Media. URL <https://link.springer.com/book/10.1007/978-0-387-93837-0> Springer Link

5.3. Software:

№ n/n	Title	Access
1.	Microsoft Windows 7 Professional RUS Microsoft Windows 10 Microsoft Windows 8.1 Professional RUS	<i>From internal network of HSE (contract)</i>
2.	MicrosoftOfficeProfessionalPlus 2010	<i>From internal network of HSE (contract)</i>
3.	R, RStudio. Packages for R	<i>Open-source license</i>

5.4. Professional data bases, information systems, internet resources (online educational resources):

№ п/п	Title	Access
<i>Professional data bases</i>		
1.	World Values Survey	Open access. URL: http://www.worldvaluessurvey.org/wvs.jsp
<i>Internet resources (online educational resources)</i>		
2.	Datacamp	URL: https://www.datacamp.com/

5.5. Material and technical base

Classrooms should be equipped with:

- PC with Internet access (operational system, office software, anti-virus software, R, RStudio);
- Multimedia projector with remote control.