Санкт-Петербургский филиал федерального государственного автономного образовательного учреждения высшего образования "Национальный исследовательский университет "Высшая школа экономики"

Факультет Санкт-Петербургская школа социальных и гуманитарных наук Национального исследовательского университета «Высшая школа экономики»

Департамент социологии

Рабочая программа дисциплины
Анализ данных в социологии (преподается на английском языке)

для образовательной программы «Социология»
направления подготовки 39.03.01 «Социология»
уровень бакалавр

Разработчики программы:
Тенишева К.А., ktenisheva@hse.ru
Широканова А.А., к.социол.н., ashirokanova@hse.ru

Согласована методистом ОСУП

«30» августа 2016 г.

Т.Г. Ефимова _____

Утверждена Академическим советом образовательной программы

«30» августа 2016 г., № протокола_____1_____

Академический руководитель образовательной программы

Д.А. Александров _____

Санкт-Петербург, 2016

National Research University Higher School of Economics

Saint - Petersburg School of Social Sciences and Humanities

Course Syllabus
Data Analysis in Sociology (offered in English)

Authors:
Anna Shirokanova, PhD, associate professor, ashirokanova@hse.ru
Ksenia Tenisheva, lecturer, ktenisheva@hse.ru

Saint Petersburg, 2016

Аннотация программы учебной дисциплины

Название дисциплины **Анализ данных в социологии (преподается на английском языке)**

Направление **39.03.01 «Социология»**
Код, название

Авторы программы **Тенишева К.А., ktenisheva@hse.ru, Широканова А.А. ashirokanova@hse.ru**

1. Краткое описание курса

Курс представляет собой анализ данных в социологии для продолжающих (intermediate-advanced). На 3 году обучения студентов курс строится вокруг двух главных тем: факторного анализа и статистического предсказания, включающего линейную регрессию и моделирование структурными уравнениями. Также обсуждаются такие ключевые для статистического анализа вопросы, как создание индексов и определение каузальности на основе полученных результатов. На 4 году обучения студентов курс посвящен многомерным методам категориальных данных и включает как специальные виды предсказательных моделей (логлинейный анализ, бинарная логистическая регрессия), так и методы снижения размерности пространства (анализ соответствий, многомерное шкалирование) и классификации (кластерный анализ).

Целью курса является обучение студентов осознанному использованию возможностей количественных исследований. Данный курс также является отправной точкой для студентов, нацеленных на более углубленное изучение методов статистики или планирующих применение количественных методов в собственных исследованиях.

2. План курса

1. Введение. Статистический вывод и основы регрессионного анализа
2. Линейная регрессия
3. Эксплораторный факторный анализ
4. Конфирматорный факторный анализ
5. Моделирование структурными уравнениями. Путевой анализ, модели структурных уравнений
6. Логлинейный анализ и логистическая регрессия
7. Многомерное шкалирование и анализ соответствий
8. Кластерный анализ

3. Требования к уровню знаний студентов

Студенты должны иметь базовое знание статистики в социальных науках: вероятность, тестирование статистических гипотез, линейная модель.

4. Преподаватели

Тенишева К.А., преподаватель департамента социологии
Широканова А.А., доцент департамента социологии

5. Тип экзамена

Письменная работа

Англоязычная версия

1. Outline

This course provides an intermediate-advanced statistical analysis for quantitative research in sociology. In the 3rd year of study, the course covers two main topics - factor analysis and statistical prediction, including linear regression and structural equation modeling. We also discuss key issues in statistical analysis, such as creating indices and identifying causality based on the results of the analysis. The 4th year of study focuses on multivariate analysis of categorical data. It includes special types of prediction models (loglinear analysis and logistic regression), techniques of dimension reduction (correspondence analysis, multidimensional scaling) and classification (cluster analysis).

The course is designed for senior students in sociology. The course covers the building blocks of quantitative data analysis with the goal of training students to be informed consumers of quantitative research. This course is also the starting point for students interested in pursuing advanced methods training or planning to use quantitative methods in their own research. This course is more applied and comprehensive than the basic statistics course that you might have taken earlier.

2. Syllabus

1. Introduction. Statistical inference. Basics of regression analysis
2. Linear regression model
3. Exploratory factor analysis
4. Confirmatory factor analysis
5. Structural equation modeling. Path analysis, SEM models
6. Loglinear analysis and binary logistic regression
7. Multidimensional scaling and correspondence analysis
8. Cluster analysis

3. Prerequisites

Students are expected to have taken some sort of basic/introductory statistics course for social science research or to have at least some experiences and knowledge about basic social statistics, such as probability, hypothesis testing, and linear regression, while they will revisit those topics in a considerable proportion during the course.

4. Authors

Ksenia Tenisheva, lecturer; Anna Shirokanova, associate professor

5. Examination type

Written assignment

1. Area of Application and Regulatory References

The program intends to lay the basic foundation of knowledge and determine the content and forms of educational activities and assessment.

The program is designed for the instructors of the Data Analysis in Sociology course, teaching assistants, and students of sociology (Bachelor Program 39.03.01 "Sociology").

The program of the Data Analysis course has been developed in accordance with:

- Educational standard of NRU HSE for Bachelor level education. URL: http://www.hse.ru/standards/standard
- Educational Program “Sociology” for the field of studies 39.03.01 “Sociology”.
- Curriculum of NRU HSE – Saint Petersburg for Bachelor level education, field of study 39.03.01 “Sociology”.

2. Course Goals

The goal of the course is to train students to be informed consumers of quantitative research. This course is also the starting point for students interested in pursuing more advanced methods training or those who are planning to use quantitative methods in their own research. This course is more applied and comprehensive than the basic statistics course that students should have taken earlier.

3. Students' Competencies to be developed by the Course

As a result of studying the discipline, the student is supposed to be able to:

- Conduct simple statistical analysis in the RStudio environment;
- Deal with basic types of sampling bias;
- Choose appropriate methods for certain types of variables and certain aims of the analysis;
- Give meaningful interpretation of statistical results: regression coefficients, tables, plots and diagrams (produced in R);
- Perform data transformations for normality correction and variables’ standardization;
- Graphically represent results of the analysis;
- Create analytical reports describing all the stages of the analysis and interpreting its results.

The competencies developed during the Course can be found in **Appendix 1**.

4. How the Course Fits in with the Curriculum

The present course is part of professional disciplines, the block of core disciplines. It is an essential element of studies for a comprehensive and in-depth understanding of the majority of professional disciplines and is crucial for developing research competency.

The “Data Analysis in Sociology” is an obligatory course within the Basic Syllabus for the area of studies 39.03.01 "Sociology" for the B.A. program at the Higher School of Economics.

The Course is based on the acquisition of the following Courses (at the bachelor level):

- Algebra and analysis
- Probability Theory and Mathematical Statistics
- Applied Software
- Sociological Theory
- Methodology and Methods of Sociological Research

The Course requires the following students' competencies and knowledge:

- Major social theories
- Basic statistical concepts
- Methods of quantitative social research

The main provisions of the Course should be used for further studies of the following Courses:

- Scientific Research Seminar

5. Course Schedule

This course is worth 6 credits.

In the 3rd year, a usual weekly procedure of this course will be a combination of a three-hour lecture given by the instructor and a one-hour group lab practice, during Modules 2 and 3.

In the 4th year, each method will be introduced in its basic version (formulas and assumptions) followed by another lecture on practice-based examples and a two-hour lab seminars. During the course, students develop three pieces of their own analysis which they present at the end of Module 1.

№	Topic	Total Aca-demic Hours	Classroom Activities			Self-Study
			Lec-tures	Semi-nars	Comput-er Labs	
Year 3						
1	Introduction	4	0	2	-	2
2	Linear regression: OLS. Diagnostics	14	0	2	2	10
3	Linear regression: Interaction effects	14	0	2	2	10
4	Exploratory factor analysis	18	4	2	2	10
5	Confirmatory factor analysis	20	4	2	4	10
6	Introduction in SEM	18	4	2	2	10
7	SEM: model specification	16	2	2	2	10
8	Path analysis	16	2	2	2	10
9	SEM with latent variables	16	2	2	2	10
10	Putting it all together	16	2	2	2	10
	Subtotal:	152	20	20	20	92
Year 4						
11	Overview of categorical data analysis	6	2	-	-	4
12	Loglinear analysis	14	2	4	-	8
13	Binary logistic regression	14	2	4	-	8
14	Multidimensional scaling	14	2	4	-	8
15	Correspondence analysis	14	2	4	-	8
16	Cluster analysis	14	4	2	-	8
	Subtotal:	76	14	18	-	44

6. Forms and Types of Testing

Type of control	Form of control	3rd year of study Modules				4th year of study Modules				Requirements
		1	2	3	4	1	2	3	4	
Current	Home-works		x	x		x				Problem sets will be given once in 1-2 weeks. Students may work together to discuss and find solutions, but they must write their own answers rather than team or group answers; identical answer sets will be considered as cheating.
Current	Activity		x	x		x				Students are expected to ask meaningful questions and participate in the discussions, as well as helping other students during practices.
Final exam	Exam (take-home written assignment)			x		x				At the end of the Module students are given written assignment to prove the skills they developed during the entire course. Students will be given two days (48 hours sharp) to complete and submit their answers after an exam is posted at the class website. Students will be supposed not to discuss or work with other students.

7. Grading criteria

The evaluation for the 3rd year is based on four criteria: activity, home works, mid-term exam, and final exam.

Activity in the class during lectures and seminars. Students are expected to ask meaningful questions and participate in discussions, as well as help other students during practices. Regular active participation in the classes is graded as perfect (10); no participation is graded as 0.

Home works. There are three basic features assessed: correct calculations and correct code (syntax); correct interpretations – students must describe trends properly, assess significance of the results, and predict values of dependent variable correctly; and produce correct graphics, with proper types of plots and formatting applied. Home works are graded from 0 ("extremely poor"="fail") to 10 ("perfect"="pass") each. Proficiency in the English language does not affect the grade.

Final exam is aimed at checking the skills students should have obtained during the course. Its structure is close to the structure of home assignments, though it covers all the topics studied. Criteria for the assessment of the exam are the same as for home works: correct calculations, correct interpretation and correct graphics.

The grades are calculated by the following scheme:

Cumulative score = 0,2*activity + 0,8*mean(home works)

Final grade = 0,8*cumulative score + 0,2*final exam

The evaluation for the 4th year is based on home works, in-class tests, and final grades earned in previous years. Each two weeks, one homework is due, which is 10% of the running grade. The homework

is a piece of data analysis using one of the core methods covered in the course (binary logistic regression, correspondence analysis, and cluster analysis). No works are accepted after the deadline. Every two weeks there will also be three short in-class tests on interpretation: one covering binary logistic regression (1st test, 4%), binary logistic regression and correspondence analysis (2nd test, 8%) and correspondence analysis and cluster analysis (3rd test, 8%). Half (50%) of the cumulative grade is transferred from previous classes on data analysis.

Final exam is aimed at checking the skills students should have obtained during the course. Its structure is close to the structure of home assignments, though it covers all the core methods studied. Criteria for the assessment of the exam are the same as for home works: correct calculations, correct interpretation and correct graphics (detailed instructions are provided during the course).

The grades are calculated by the following scheme:

Cumulative score = 0,2*final grade in sophomore year + 0,3*final grade in junior year + 0,1*homework 1 + 0,1*homework 2 + 0,1*homework 3 + 0,04*test 1 + 0,08*test 2 + 0,08*test 3

Final grade = 0,8*cumulative score + 0,2*final exam

8. The Course Contents

Topics 1-2. Introduction. Linear regression: OLS. Diagnostics

Covariance and correlation, basic concept and logics of linear regression, OLS estimator of linear regression, interpretation and statistic test of OLS estimators, fitted values and residuals, R-squared, addressing nonlinearity in linear regression framework, standardized coefficients, drawing plots, practice in R.

Core reading

John Fox and Sanford Weisberg. (2011). An R Companion to Applied Regression, Second Edition, Sage Publications.

Additional reading

Andrew Gelman and Jennifer Hill. (2007). Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge University Press.

Kurt Taylor Gaubatz (2014). A Survivor's Guide to R: An Introduction for the Uninitiated and the Unnerved, 1st edition, Sage.

Роберт И. Кабаков (2014). R в действии. Анализ и визуализация данных на языке R, ДМК Пресс.

Topic 3. Linear regression: Interaction effects

Main and multiplicative effects in regression models. Interaction effects, additive effects. Interpreting results. Choosing best model. Practice in R.

Core reading

John Fox and Sanford Weisberg. (2011). An R Companion to Applied Regression, Second Edition, Sage Publications.

Additional reading

Alan Agresti and Barbara Finlay. (2007). Statistical Methods for the Social Sciences, Fourth Edition, Pearson Prentice Hall.

Paul Teetor. (2011). R Cookbook (O'Reilly Cookbooks), 1st Edition, O'Reilly Media.

Роберт И. Кабаков (2014). R в действии. Анализ и визуализация данных на языке R – ДМК Пресс.

Topic 4. Exploratory factor analysis

Dimensionality reduction. Manifest and latent variables. Factors, graphical representation of factors. Exploratory factor analysis. Factor scores, factor space, types of rotation. Optimal number of factors. Interpretation of the results. Creating indices based on factor analysis. Practice in R.

Core reading

Andy Field, Jeremy Miles and Zoe Field (2012). Discovering Statistics Using R, Sage Publications.

Additional reading

C. E. Lance and R.J. Vandenberg (ed.) (2010). Statistical and methodological myths and urban legends: Doctrine, verity and fable in organizational and social sciences, Routledge.

Topic 5. Confirmatory factor analysis

Difference between exploratory and confirmatory factor analyses. Factor structure. Testing your (or somebody else's) scales. Types of latent variables. Constructing factor model in lavaan package. Calculation of degrees of freedom, minimal number of cases. Non-correlated and correlated latent factors. Interpreting results. Model diagnostics. Cronbach's alpha. Practice in R.

Core reading

Andy Field, Jeremy Miles and Zoe Field (2012). Discovering Statistics Using R, Sage Publications.

Additional reading

T.A. Brown (2015). Confirmatory factor analysis for applied research, Guilford Publications.

Topic 6. Introduction in SEM

Structural equation modeling as extension of confirmatory factor analysis. Exogenous and endogenous variables. Testing causal assumptions. Partial correlation, heterogeneous correlations (polychoric, tetrachoric and polyserial correlations). Practice in R.

Core reading

R. B. Kline (2015). Principles and practice of structural equation modeling, Guilford publications.

Additional reading

C. E. Lance and R.J. Vandenberg (ed.) (2010). Statistical and methodological myths and urban legends: Doctrine, verity and fable in organizational and social sciences, Routledge.

Topic 7. SEM: model specification and identification

Formulating theory-based causal hypotheses. Causal inference. Specification concepts. Mediation and moderation effects. Measurement error: correlated and uncorrelated. Practice in R.

Core reading

R. B. Kline (2015). Principles and practice of structural equation modeling, Guilford publications.

Additional reading

C. E. Lance and R.J. Vandenberg (ed.) (2010). Statistical and methodological myths and urban legends: Doctrine, verity and fable in organizational and social sciences, Routledge.

Topic 8. Path analysis

Concept of "path". Path analysis: only observed variables. Graphical representation. Identification of path model. Estimation of structural equation model. Model fit. Degrees of freedom, number of cases. Meaning of the indices. Corrected chi-square measures. Interpreting the results. Practice in R.

Core reading

R. B. Kline (2015). Principles and practice of structural equation modeling, Guilford publications.

Additional reading

C. E. Lance and R.J. Vandenberg (ed.) (2010). Statistical and methodological myths and urban legends: Doctrine, verity and fable in organizational and social sciences, Routledge.

Topic 9. SEM with latent variables

Introducing latent factors in the model. Identification of SEM. Estimation of structural equation model. Model fit. Meaning of the fit indices. Model modification. Interpreting the results. Practice in R.

Core reading

R. B. Kline (2015). Principles and practice of structural equation modeling, Guilford publications.

Additional reading

C. E. Lance and R.J. Vandenberg (ed.) (2010). Statistical and methodological myths and urban legends: Doctrine, verity and fable in organizational and social sciences, Routledge.

Topic 10. Putting it all together

Implementing all the methods to the real-life research. Combining factor analysis and regression analysis. Using SEM to test theoretical assumptions about causality. Advantages and disadvantages of the methods.

Additional reading

C. E. Lance and R.J. Vandenberg (ed.) (2010). Statistical and methodological myths and urban legends: Doctrine, verity and fable in organizational and social sciences, Routledge.

Topic 11. Overview of Categorical Data Analysis

Models for categorical outcome variables. Variety of goals of analysis with categorical data. Examples of empirical research for various methods, e.g. loglinear analysis, Poisson regression (count variable), binary logistic regression, ordinal regression, multinomial regression, correspondence analysis, conjoint analysis, multidimensional scaling, and cluster analysis. Typical goals of analysis and interpretation of results.

Core reading

Joseph F. Hair et al. (2014). Multivariate Data Analysis, Pearson New International Edition, Pearson Education Limited.

Additional reading

Alan Agresti (2013). Categorical Data Analysis, 2nd edition, John Wiley & Sons, Inc.

Andy Field, Jeremy Miles and Zoe Field (2012). Discovering Statistics Using R, Sage Publications.

Klaus Backhaus, Bernd Erichson and Rolf Weiber (2015). Fortgeschrittene Multivariate Analysemethoden. Eine anwendungsorientierte Einführung, 3. Afl., Springer.

Э. Понарин, А. Лисовский, Ю. Зеликова (2013). Модели для пуассоновских зависимых переменных: можно ли прогнозировать результативность футбольных матчей? // Социология: 4М. №36. С. 36-64.

Topic 12. Loglinear Analysis

Contingency tables. Pearson's chi-square test. The likelihood ratio. Loglinear models. Backward elimination for model building. Assumptions of loglinear analysis. Mosaic plots. Odds ratios. Reporting the results of loglinear analysis. Procedures in R.

Core reading

Andy Field, Jeremy Miles and Zoe Field (2012). Discovering Statistics Using R, Sage Publications.

Additional reading

D. Sharpe (2015). Your chi-square test is statistically significant: Now what?. Practical Assessment, Research & Evaluation, 20(8). URL: http://www.pareonline.net/getvn.asp?v=20&n=8

Г. Ястребов (2016). Социальная мобильность в советской и постсоветской России: новые количественные оценки по материалам представительных опросов 1994, 2002, 2006 и 2013 гг. Часть II // Мир России. Т. 25. № 2. С. 6-36.

Topic 13. Binary Logistic Regression

Logistic regression. Objectives of logistic regression. Logistic curve. Assumptions of logistic regression. Transforming a probability into odds and logit values. Maximum likelihood estimation. Goodness-of-fit measures for logistic regression. Interpretation of results (linear and dichotomous predictors). Stepwise model building. Diagnostics. Procedures in R.

Core reading

Joseph F. Hair et al. (2014). Multivariate Data Analysis, Pearson New International Edition, Pearson Education Limited.

Additional reading

Andy Field, Jeremy Miles and Zoe Field (2012). Discovering Statistics Using R, Sage Publications.

B.S. Everitt and T. Hothorn (2017). Logistic Regression and Generalised Linear Models: Blood Screening, Women’s Role in Society, and Colonic Polyps, in: HSAUR: A Handbook of Statistical Analyses Using R (1st Edition): https://cran.r-project.org/web/packages/HSAUR/vignettes/Ch_logistic_regression_glm.pdf

Topic 14. Multidimensional Scaling

Dimension reduction as an objective of data analysis. Idea of MDS. Perceptual map. MDS vs. factor and cluster analyses. Objectives of MDS. MDS algorithms. Decompositional and compositional approach. The number and selection of objects. Nonmetric vs. metric methods. Similarity data and preference data. Assumptions of MDS analysis. Selecting dimensionality. Ideal point. Goodness-of-fit measures. Interpreting MDS results. Procedures in R.

Core reading

Joseph F. Hair et al. (2014). Multivariate Data Analysis, Pearson New International Edition, Pearson Education Limited.

Additional reading

B.S. Everitt and T. Hothorn (2017) Multidimensional Scaling: British Water Voles and Voting in US Congress , in: HSAUR: A Handbook of Statistical Analyses Using R (1st Edition): https://cran.r-project.org/web/packages/HSAUR/vignettes/Ch_multidimensional_scaling.pdf

Klaus Backhaus, Bernd Erichson and Rolf Weiber (2015). Fortgeschrittene Multivariate Analysemethoden. Eine anwendungsorientierte Einführung, 3. Afl., Springer.

Todd D. Little (ed.) (2013). The Oxford Handbook of Quantitative Methods. Volume 2: Statistical Analysis, Oxford University Press.

Topic 15. Correspondence Analysis

Objectives of correspondence analysis. Assumptions of correspondence analysis. Perceptual mapping. Principal components analysis. The row and column problems. Correspondence analysis displays. Correspondence analysis biplots. Multiple correspondence analysis. MCA maps. Measures of fit for MCA. Interpreting correspondence analysis results. Canonical correspondence analysis. Procedures in R.

Core reading

Joseph F. Hair et al. (2014). Multivariate Data Analysis, Pearson New International Edition, Pearson Education Limited.

Additional reading

Klaus Backhaus, Bernd Erichson and Rolf Weiber (2015). Fortgeschrittene Multivariate Analysemethoden. Eine anwendungsorientierte Einführung, 3. Afl., Springer.

Jorg Blasius and Michael Greenacre (2006). Correspondence Analysis and Related Methods in Practices, in: Greenacre, M., & Blasius, J. (Eds.). (2006). Multiple correspondence analysis and related methods. CRC press. P. 3-40.

J.D. Jobson (2012). Applied multivariate data analysis: volume II: Categorical and Multivariate Methods. Springer Science & Business Media.

Todd D. Little (ed.) (2013). The Oxford Handbook of Quantitative Methods. Volume 2: Statistical Analysis, Oxford University Press.

М. Шафир (2009). Анализ соответствий: представление метода // Социология: 4М. №28. С. 29-44.

Topic 16. Cluster Analysis

Objectives of cluster analysis (taxonomy description, data simplification, and relationship identification). Necessity of conceptual framework. Similarity measures. Proximity matrix. Decision-process in cluster analysis. Dendrograms. Cluster profiles. Distance measures for various types of variables. Assumptions of cluster analysis. Measures of overall fit. Between- and within-cluster variation. Hierarchical and non-hierarchical clustering algorithms. Determining the number of clusters. Interpretation of clusters. Cross-classification from several solutions. Procedures in R.

Core reading

Joseph F. Hair et al. (2014). Multivariate Data Analysis, Pearson New International Edition, Pearson Education Limited.

Additional reading

B.S. Everitt and T. Hothorn (2017). Cluster Analysis: Classifying the Exoplanets, in: HSAUR: A Handbook of Statistical Analyses Using R (1st Edition): https://cran.r-project.org/web/packages/HSAUR/vignettes/Ch_cluster_analysis.pdf

Todd D. Little (ed.) (2013). The Oxford Handbook of Quantitative Methods. Volume 2: Statistical Analysis, Oxford University Press.

9. Educational Technologies

The course syllabus, reading texts, presentations, practical tasks and home assignments will be available in the Google group "HSE Data Analysis class" and sent by email or distributed via Dropbox groups.

9.1. Methodical recommendations for students

Students are recommended to read lots of scientific articles to learn how to apply statistical methods and, most importantly, describe the results. They are also expected to adopt the knowledge they got during these classes in their course papers and to discuss with the teaching staff any issues arising during this process. Start with basic examples and look for the available explanations in videos, step-by-step layouts, peers, etc. "Build your linear regression models first, then decide whether it is worth working harder to achieve better results" (Data Science Design Manual, 256).

9.2 Methodical recommendations for instructors

B. Rosenshine Principles of Instruction (2010). URL: http://www.ibe.unesco.org/sites/default/files/resources/edu-practices_21_eng.pdf

10. Diagnostics for Monitoring Students' Work and Progress

To assess their progress, students will be given home tasks each two weeks. Students are expected to replicate a research from a given scientific article using the methods discussed during the classes or to produce their own model following the lecture example. Each home task is based on the materials from preceding lectures and includes both theoretical questions and practical tasks that ought to be answered using R.

Example:

- *Replicate regression table given in the article. Interpret the results.*
- *What is the null hypothesis? What is the alternative hypothesis?*
- *Find the test statistic in the result reported by R. Show how the resulting test statistic could be computed.*
- *What is the p-value and what does the p-value mean?*
- *Should we reject the null hypothesis? Can we conclude that the estimated relationship between household size and satisfaction would hold true in the population?*

In addition to home tasks, in the 3rd year students are to pass midterm and final exam. Their structure is similar to home tasks, though it covers all topics studied:

- Module 2 – for the midterm exam;
- Modules 2 and 3 – for the final exam.

Example of exam questions (students are expected to answer at home):

- *Read the article carefully. Identify the measures used in the analysis, find them in the dataset.*
- *Were the variables transformed in any way? Prepare the data for analysis*
- *Replicate the regression table given in the article.*
- *Interpret the results. Do your results fit those presented in the article? What might be the reason of this difference?*

11. Information basis for the course

11.1. Core Textbook

Andy Field, Jeremy Miles and Zoe Field (2012). Discovering Statistics Using R, Sage Publications.

John Fox and Sanford Weisberg. (2011). An R Companion to Applied Regression, Second Edition, Sage Publications.

Joseph F. Hair et al. (2014). Multivariate Data Analysis, Pearson New International Edition, Pearson Education Limited.

R. B. Kline (2015). Principles and practice of structural equation modeling, Guilford publications.

11.2. Internet resources:

Institute for Digital Research and Education Data Analysis Examples : https://stats.idre.ucla.edu/other/dae/ UCLA

John Verzani, Simple R: Using R for Introductory Statistics. Available at http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf

LAVAAN project (CFA and SEM): http://lavaan.ugent.be/

LOGIT REGRESSION | R DATA ANALYSIS EXAMPLES (2013): https://stats.idre.ucla.edu/r/dae/logit-regression/

Multivariate Data Analysis (J. Hair) Companion Website: www.mvstats.com
StatSoft Electronic Statistics Textbook: http://www.statsoft.com/Textbook/
UCLA Stats Lab : http://www.ats.ucla.edu/stat/r

11.3. Supplementary Reading

Alan Agresti (2013). Categorical Data Analysis, 2nd edition, John Wiley & Sons, Inc.

Alan Agresti and Barbara Finlay. (2007). Statistical Methods for the Social Sciences, Fourth Edition, Pearson Prentice Hall.

Klaus Backhaus, Bernd Erichson and Rolf Weiber (2015). Fortgeschrittene Multivariate Analysemethoden. Eine anwendungsorientierte Einführung, 3. Afl., Springer.

Jorg Blasius and Michael Greenacre (2006). Correspondence Analysis and Related Methods in Practices, in: Greenacre, M., & Blasius, J. (Eds.). (2006). Multiple correspondence analysis and related methods. CRC press. P. 3-40.

T.A. Brown (2015). Confirmatory factor analysis for applied research, Guilford Publications.

B.S. Everitt and T. Hothorn (2017). Cluster Analysis: Classifying the Exoplanets, in: HSAUR: A Handbook of Statistical Analyses Using R (1st Edition): https://cran.r-project.org/web/packages/HSAUR/vignettes/Ch_cluster_analysis.pdf

B.S. Everitt and T. Hothorn (2017). Logistic Regression and Generalised Linear Models: Blood Screening, Women’s Role in Society, and Colonic Polyps, in: HSAUR: A Handbook of Statistical Analyses Using R (1st Edition): https://cran.r-project.org/web/packages/HSAUR/vignettes/Ch_logistic_regression_glm.pdf

B.S. Everitt and T. Hothorn (2017) Multidimensional Scaling: British Water Voles and Voting in US Congress , in: HSAUR: A Handbook of Statistical Analyses Using R (1st Edition): https://cran.r-project.org/web/packages/HSAUR/vignettes/Ch_multidimensional_scaling.pdf

Kurt Taylor Gaubatz (2014). A Survivor's Guide to R: An Introduction for the Uninitiated and the Unnerved, 1st edition, Sage.

Andrew Gelman and Jennifer Hill (2007). Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge University Press.

J.D. Jobson (2012). Applied multivariate data analysis: volume II: Categorical and Multivariate Methods. Springer Science & Business Media.

C. E. Lance and R.J. Vandenberg (ed.) (2010). Statistical and methodological myths and urban legends: Doctrine, verity and fable in organizational and social sciences, Routledge.

Todd D. Little (ed.) (2013). The Oxford Handbook of Quantitative Methods. Volume 2: Statistical Analysis, Oxford University Press.

D. Sharpe (2015). Your chi-square test is statistically significant: Now what?. Practical Assessment, Research & Evaluation, 20(8). URL: http://www.pareonline.net/getvn.asp?v=20&n=8

Роберт И. Кабаков (2014). R в действии. Анализ и визуализация данных на языке R, ДМК Пресс.

Э. Понарин, А. Лисовский, Ю. Зеликова (2013). Модели для пуассоновских зависимых переменных: можно ли прогнозировать результативность футбольных матчей? // Социология: 4М. №36. С. 36-64.

М. Шафир (2009). Анализ соответствий: представление метода // Социология: 4М. №28. С. 29-44.

Г. Ястребов (2016). Социальная мобильность в советской и постсоветской России: новые количественные оценки по материалам представительных опросов 1994, 2002, 2006 и 2013 гг. Часть II // Мир России. Т. 25. № 2. С. 6-36.

11.4. Program packages

1. R: R is a software that is widely used for data analyses among social scientists in these days, it has been getting popularity increasingly because of its several merits over other stats package tools. First, it is free and easy to get. You can download a recent version of R on the web (https://www.r-project.org/). Second, self-teaching materials, manuals or references for solving problems you may come across are abundant on the web, all free of charge. Third, it is a programming tool, so learning R, even at a very basic level, might provide an interesting experience of getting over any inner barriers in mind and finding a programming no scary thing with students who are afraid of programming or anything like that. Lastly, R will be extremely useful if you are considering to be a more serious researcher.

2. R-Studio: R-Studio is a software proving a more user-friendly environment when you run R. It is also available free of charge at https://www.rstudio.com.

12. Technical support

Each lecture is supported by Power Point presentations shown by OHP projector. Seminars and lab sessions are to be held in a fully-equipped computer class with personal computers available to every student in a group (in cases when there are more students than PCs they are welcome to bring their own computers).

13. Academic Integrity

Each student in this course is expected to abide by the Higher School of Economics’ Academic Honesty Policy. For this course, collaboration is allowed for pair and group work during seminar classes.

Appendix 1. Competences developed during the Course.

Уровни формирования компетенций:

РБ - ресурсная база, в основном теоретические и предметные основы (знания, умения)
СД - способы деятельности, составляющие практическое ядро данной компетенции
МЦ - мотивационно-ценностная составляющая, отражает степень осознания ценности компетенции человеком и готовность ее использовать

В результате освоения дисциплины студент осваивает следующие компетенции:

Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
Способен решать проблемы в профессиональной деятельности на основе анализа и синтеза	УК-3	СД, МЦ	Способен сформулировать гипотезы исследования, провести анализ, оценить степень подтвержденности гипотез и предложить научное объяснение	Работа на семинарах; выполнение домашних заданий	Домашние задания; экзамен
Способен вести исследовательскую деятельность, включая анализ проблем, постановку целей и задач, выделение объекта и предмета исследования, выбор способа и методов исследования, а также оценку его качества	УК-6 (СК-Б7)	СД	Способен определить цели и задачи исследования; в соответствие с ними сформулировать гипотезы и выбрать подходящий метод для их проверки. оценить результаты анализа	Работа на семинарах; выполнение домашних заданий	Активность на семинарах; домашние задания; экзамен
Способен осуществлять производственную или прикладную деятельность в международной среде	УК-10	СД	Способен понимать статьи на английском языке; понимать лекции на английском языке; обсуж-	Лекции на английском языке; участие в семинарах; выполнение домашних заданий (включая чтение ста-	Активность на семинарах; домашние задания; экзамен

			дать методы и результаты анализа на английском языке	тей)	
Способен использовать основные законы естественнонаучных дисциплин в профессиональной деятельности, применять методы математического анализа и моделирования, теоретического и экспериментального исследования при решении профессиональных задач	ПК-5	РБ, СД	Способен выбрать метод, необходимый для решения исследовательских задач; построить корректную модель, отвечающую техническим и содержательным требованиям; интерпретировать результаты	Работа на семинарах; выполнение домашних заданий	Домашние задания; экзамен
Способен самостоятельно формулировать цели, ставить конкретные задачи научных исследований в различных областях социологии и решать их с помощью современных исследовательских методов	ПК-6, ИК-Б5.1.1 НИД (С)	СД	самостоятельное построение моделей анализа данных, оценка качества модели и умение выбрать лучшую модель обработка реальных социологических данных с применением RStudio	Работа на семинарах; выполнение домашних заданий	Домашние задания; экзамен