**Course Syllabus**

|  |  |
| --- | --- |
| Title of the course | Advanced Statistical Modeling (offered in English) |
| Title of the Academic Programme  | Modern Social Analysis |
| Type of the course  | Elective |
| Prerequisites | 1. Methodology and Methods of Sociological Research
2. Quantitative Analysis of Sociological Data
 |
| ECTS workload | 5 |
| Total indicative study hours | Directed Study | Self-directed study  | Total |
| 32 | 158 | 190 |
| Course Overview | The course gives an introduction to a variety of extensions of linear regression analysis which are widely used in the modern social sciences for identification of causal effects, as well as implementations of the respective methods in R, a popular, free programming language for statistical computing. The topics covered by the course are (a) the basics of counterfactual inference and the potential outcomes framework (including graphical identification methods), (b) matching methods, (c) instrumental variables estimation, (d) regression discontinuity design, and (e) difference-in-differences estimation. In addition, practical issues of estimation, visualization and presentation of the statistical models studied during the course are discussed. Students are assumed to have basic knowledge of statistics and be familiar with several conventional statistical methods, most importantly linear regression, and with the R programming environment.  |
| Intended Learning Outcomes (ILO) | After completing this course, students are expected to be able:* to read and understand most academic social sciences articles that use quantitative approach
* to design a quantitative social study
* to choose statistical methods appropriate to their data and substantive research problem
* to use R programming language for complex statistical computations
 |
| Teaching and Learning Methods | Tutorials, labs, project work |
| Content and Structure of the Course |
| **№** | **Topic / Course Chapter** | **Total** | **Directed Study** | **Self-directed Study** |
| **Lectures** | **Tutorials** |
| 1. | Introduction to the potential outcomes framework | 40 | 0 | 8 | 32 |
| 2. | Matching | 38 | 0 | 6 | 32 |
| 3. | Instrumental variables estimation | 38 | 0 | 6 | 32 |
| 4. | Regression discontinuity design | 38 | 0 | 6 | 32 |
| 5. | Difference-in-differences estimation | 36 | 0 | 6 | 30 |
| **Total study hours** | 190 | 0 | 32 | 158 |
| Indicative Assessment Methods and Strategy  | Your final grade is composed of four components: (a) home assignments (unweighted average), (b) class work, (c) midterm presentation, and (d) final project. The first three components (taken with different weights) define your accumulated grade (which comprises 65% of the overall grade) which is then combined with the score for the final paper (35% of the overall grade).Formally, your cumulative grade for the courseis calculated in the following way:**Gcumulative = 0.5 \* Ghome assignment + 0.1 \* Gclass work + 0.4 \* Gmid-term presentation (1)**where Ghome assignment is the mean [unweighted] grade for home assignments prepared throughout the course; Gclass assignment is the [single, *post hoc,* on a 1-10 scale] assessment of your overall participation in class activities; and Gmid-term presentation is the mean [unweighted] grade for final paper progress reports. The final (overall) grade for the course (the one you will have in your course records) is calculated in the following way:***G*final *=* 0,65 \* Gcumulative + 0.35 \* Gfinal paper (2)**where Gcumulative is the accumulated grade for the course calculated using the formula above and Gfinal paper is the grade for the final research paper.Both the cumulative grade and the final grade are rounded according to the rules of algebraAll late submissions (both home assignments, progress reports, and the final paper) will be graded down (the maximum grade for a one-day delay is 7 points; this quantity decreases by one point per extra day of delay; all works submitted with a delay of three and more days are evaluated on a 0-4 scale). If you plagiarize, you will fail. However, it is fine to combine your final research project with other class papers (e.g., term paper). |
| Readings / Indicative Learning Resources | MandatoryPearl, Judea, et al. Causal Inference in Statistics : A Primer, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, <https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=4383483> OptionalVan-Nam HuynhVladik KreinovichSongsak Sriboonchitta (2016). *Causal Inference in Econometrics*. [https://link.springer.com/book/10.1007%2F978-3-319-27284-9#about](https://link.springer.com/book/10.1007/978-3-319-27284-9#about) Tabachnick, B. G., and Fidell, L. S. (2014). Using Multivariate Statistics. 6th edition. Harlow, Essex: Pearson.  URL:<http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1418064>  |
| Indicative Self- Study Strategies | **Type** | **+/–** | **Hours** |
| Reading for seminars / tutorials (lecture materials, mandatory and optional resources) | + | 38 |
| Assignments for seminars / tutorials / labs | + | 40 |
| E-learning / distance learning (MOOC / LMS) | **–** | **–** |
| Fieldwork | **–** | **–** |
| Project work | + | 40 |
| Other (please specify) | **–** | **–** |
| Preparation for the exam | + | 40 |
| Academic Support for the Course | Academic support for the course is provided via LMS, where students can find: guidelines and recommendations for doing the course; guidelines and recommendations for self-study; samples of assessment materials |
| Facilities, Equipment and Software | It is better to have your own personal laptop with R (and maybe Rstudio) software installed (though it is also possible to use university computers for completing home assignments and performing in-class labs). You can read detailed R installation instructions [here](https://cran.r-project.org/doc/manuals/r-release/R-admin.html) or simply download the latest version [here](http://cran.us.r-project.org/). |
| Course Instructor | Boris Sokolov, associate professor, bssokolov@hse.ru  |