University of St.Gallen

Course and Examination Fact Sheet: Spring Semester 2021

8,386: Data Analytics: Applications and Visualization

ECTS credits: 4

Overview examination/s

(binding regulations see below)
Decentral - Group examination paper with presentation (all given the same grades) (100%)
Examination time: term time

Attached courses

Timetable -- Language -- Lecturer
8,386,1.00 Data Analytics: Applications and Visualization -- Englisch -- Bonev Petyo

Course information

Course prerequisites

Students attending the course should be familiar with basic probability concepts such as: 1) what is a probability 2) what is a random variable? 3) what is a distribution of random variable 4) what is expectation and variance?

In the first lecture, a very brief overview of R will be given, so no prior knowledge required. However, students are expected to have R or R studio on their computers.

Attendance of the HS2020 course "Data Analytics and Causal Inference" (7,354) is not a must, and some of the concepts there will be repeated but (1) in a more detailed and (2) basic way and (3) with a strong implementation focus.

Learning objectives

The objectives of the course are twofold. The first objective is to give students comprehensive tools for the analysis of statistical data. This includes 1) intuitions of the most important research designs, 2) the ability to critically analyse empirical assumptions that underly causal analysis, as well as (most importantly) programming skills.

The second objective is to teach participants how to 1) visualize complex causal relationships with R (i.e. create credible and convincing graphs and 2) present research/project results to third parties.

One implicit objective of the course is to teach students how to write and present a data-oriented (empirical) Master thesis.

The main way to meet these targets will be to require attendants to actively participate in the course. In particular, they will write own projects with real-world datasets that will be presented at the end of the course.

Course content

The first part of the content is dedicated to visualizing relationships between variables. 2 dimensional densities, 2-dimensional histograms, contingency tables and other methods will be introduced. In addition, basic concepts of presenting results will be taught. How do I recognize that the graphic is trying to manipulate me? How do I present a graph that is objective? How do I create a graph that does not confuse the reader? What are the limits of a graph in terms of information content?

The second part is dedicated to the linear regression analysis and to other research methods that are related. We will use a real-world dataset throughout the course and study the research designs via step-by-step implementation using this dataset. The major focus is two-fold: (1) how do I interpred the results of these methods (and under which assumptions)? and (2) how do I perform model diagnostics: (i) are the assumptions of my research design plausible? (ii) how do I perform sensitivy analysis/robustness checks? (iii) how do I select the best model among many?

The third part consists of three extentions: instrumental variable method, difference-in-differences approach and LASSO. The last is a Machine Learning/Artificial intelligence method. The fourth part consists of project presentations. Students will present in groups a real-world-dataset project that they have worked on during the semester. This will be simultaneously also a preparation for writing the Master thesis.

Course structure

I. Visualization

1. Brief introduction into R

2. Introduction into visualization: principles, strategies, mistakes.

3. Introduction into descriprive statistics with R: 2-D plots, 2-D densities, 2-D histogramms

II. The linear regression in R

1. Causal interpretation of the linear regression model

2. Assumptions and identification

3. Model diagnostics:

3.1 Checking the assumptions

3.2 Robustness checks and sensitivity analysis

3.3. Model selection

III. Extensions (if time allows it):

1. Instrumental variables in R

2. Artificial intelligence: LASSO

3. Difference-in-differences

Course literature

The must literature for the course consists of a lecture script.

Additional sources (nonobligatory):

* Stock and Watson (Introduction to econometrics), the first 4 chapters.

* Mostly harmless econometrics (2008) by Angrist and Pischke, pdf file free of charge in the internet

Additional course information

In the case of the President's Board having to implement new directives due to the SARS-CoV-2 pandemic in SpS2021, the course information listed above will be changed as follows:

- The course is conducted online via the platform zoom.
- The recordings of the course are available for 30 days;
- The lecturer informs via Canvas and or email on the changed implementation modalities of the course;

The examination information listed below would be changed as follows:

- The presentations are conducted online and are being recorded;

University of St.Gallen

Examination information

Examination sub part/s

1. Examination sub part (1/1)

Examination time and form
Decentral - Group examination paper with presentation (all given the same grades) (100%)
Examination time: term time

Remark
--

Examination-aid rule
Term papers

Term papers must be written without anyone else's help and in accordance with the known quotation standards, and they must contain a declaration of authorship which is a published template in StudentWeb.

The documentation of sources (quotations, bibliography) has to be done throughout and consistently in accordance with the chosen citation standard such as APA or MLA.

For papers in law, the legal standard is recommended (by way of example, cf. FORSTMOSER, P., OGOREK R. et SCHINDLER B., Juristisches Arbeiten: Eine Anleitung für Studierende, newest edition respectively, or according to the recommendations of the Law School).

The indications of the sources of information taken over verbatim or in paraphrase (quotations) must be integrated into texts in accordance with the precepts of the applicable quotation standard, while informative and bibliographical notes must be added as footnotes (recommendations and standards can be found, for example, in METZGER, C., Lern- und Arbeitsstrategien, newest edition respectively.

For any work written at the HSG, the indication of the page numbers is mandatory independent of the chosen citation standard. Where there are no page numbers in sources, precise references must be provided in a different way: titles of chapters or sections, section numbers, acts, scenes, verses, etc.

Supplementary aids
--

Examination languages
Question language: English
Answer language: English

Examination content

The grade will consist of two components: a grade on the term project and a grade on the group presentation of the term project. Project papers will be evaluations of real-world causal questions using real-world datasets. The projects will be written in groups of up to 5 and minimum of 2 students. The projects have to be submitted before the first presentation takes place. Topics will be assigned at the begin of the semester.

Presentations will be short talks that aim to explain the object and results of analysis of the term projects. Dates will be given at the begin of the semester. The duration will depend on the total number of presentations but will be up to 20 min.

Examination relevant literature
The same as the course literature.

University of St.Gallen