



Course and Examination Fact Sheet: Autumn Semester 2020

7,354: Data Analytics and Causal Inference

ECTS credits: 6

Overview examination/s

(binding regulations see below)

Decentral - Group examination paper (all given the same grades) (50%)

Examination time: term time

Central - Written examination (50%, 90 mins.)

Examination time: inter-term break

Attached courses

Timetable -- Language -- Lecturer

[7.354.1.00 Data Analytics and Causal Inference](#) -- Englisch -- [Epper Thomas Frank](#)

[7.354.2.01 Data Analytics and Causal Inference: Exercises, Group 1](#) -- Englisch -- [Bonev Petyo](#)

[7.354.2.02 Data Analytics and Causal Inference: Exercises, Group 2](#) -- Englisch -- [Bonev Petyo](#)

Course information

Course prerequisites

Students with no prior knowledge or experience in the methodology of empirical research in the social sciences may be required to familiarize themselves with some basic empirical methods. This can be done by studying relevant textbooks, such as Imai, Kosuke (2018) *Quantitative Social Science: An Introduction*. Princeton University Press. Note that the book covers topics beyond those discussed in this course. Refer to the course outline for the content of the course.

In order to hone their skills in programming skills for data analysis, students are encouraged to attend the MIA mornings workshops 7,364 'Programming in R'; 7,366 'Data Handling and Manipulation' and 7,368 'Effective Data Visualization' taught by André Walter.

We also recommend Gary King's excellent math refresher (<https://bookdown.org/kuriwaki/prefresher/>).

Learning objectives

The course aims to impart a *basic understanding* and *intuition* of quantitative research methods. The goal is to provide students of International Affairs with the foundation necessary to analyze data in their own research and to become critical consumers of statistical claims made in the news media, in policy reports, and in academic research. The following learning objectives should be attained after successful completion of both the lecture and the associated group meetings:

- Students have a strong working knowledge and the skills to identify and apply adequate methods in tackling problems in International Affairs.
- Students identify appropriate evidence and evaluate critically existing social science scholarship to an advanced level.
- Students systematically appraise theoretical and applied knowledge and understanding of quantitative research techniques.
- Students assess diverse theoretical and methodological approaches and perspectives as they frame and evaluate evidence in a systematic and critical manner.
- Students demonstrate proficiency in assessing the adequacy of statistical and causal inference methods to answer relevant research and policy questions in social sciences

Course content



Substantive questions in empirical scientific and policy research are often causal. Does voter outreach increase turnout? Are job training programs effective? Can a universal health insurance program improve people's health? This course introduces students to both statistical theory and practice of causal inference, with a focus on applications in International Politics, Economics, Law, and Business Strategy.

The course puts special emphasis on an understanding of and the intuition of the relevant concepts, rather than the formal definitions and technicalities. It has an applied focus: Students learn what methods can help them to answer particular research questions in their field, they learn how to make sense out of data, and to interpret and to critically assess existing data analyses. Examples adopting the different analysis methods and causal identification strategies are discussed to illustrate how the methods are used in practice.

In this course we use the open-source statistical software R (<http://www.r-project.org>). To help make using R easier, we'll be using RStudio (<http://www.rstudio.com/>)-a user interface that simplifies many common operations.

Course structure

The course takes place during the first half of the semester. It includes two key components: lectures and exercise groups. For the effective completion of the course, it is necessary to attend both the lecture sessions and exercise group meetings.

The lecture intends to provide a brief introduction to the following topics in causal inference. The kick-off session introducing the course structure and requirements is followed by three block sessions:

Session I:

- i. Data analytics I: Basics
 - Working with data
 - Probability and statistics
 - Variable relationships

Session II:

- i. Data analytics I: Testing
 - Hypothesis testing
- ii. Causal inference I: Basics
 - Correlation vs. causation
 - Experimental reasoning and identification
 - Understanding decision-making processes: Process tracing

Session III:

- i. Causal inference II: Methods
 - Counterfactuals, controls and matching
 - Difference-in-differences
 - Regression discontinuity
 - Instrumental variables

The lecture is complemented by **group exercise sessions**. There will be four weekly two-hour sessions. The exercise group meetings center on basic descriptive statistics and linear regression. The methods are applied to specific problems in class. Exercise group meetings are discussion-based. Participation should reflect solid preparation-done the required readings for the week's topic, thought about the material and done the best to understand it.

Group meeting I: Elements of R and Descriptive statistics



- A very brief introduction to R
- The necessity of descriptive statistics
- Frequency tables, empirical distribution functions, histograms

Group meeting II: Descriptive statistics for the analysis of two or more variables

- Scatter plots
- Merged histograms, 2D density visualization, covariance matrices

Group meeting III: Simple linear regression

- The regression function and its causal interpretation; basic assumptions
- Running regression in R and interpreting the output
- Confidence intervals; hypothesis testing

Group meeting III: Multivariate linear regression; model diagnostics

- Is my model correct? Choice of functional form and variables; discussion of endogeneity.

Course literature

The following course books on causal inference are recommended:

- Stephen L. Morgan and Christopher Winship (2014) "Counterfactuals and Causal Inference: Methods and Principles for Social Research", Cambridge University Press, 2nd edition, ISBN: 1107694167.
- Rebecca Morten and Kenneth C. Williams (2012) "Experimental Political Science and the Study of Causality - From Nature to the Lab", Cambridge University Press, ISBN: 9780511762888.

These books lack a comprehensive discussion of two submodules of the course. These are (i) regression discontinuity and (ii) difference-in-difference methods. This gap will be filled with additional handouts and resources. Specifically, the following provides an excellent (albeit more technical) coverage of these two topics:

- Joshua D. Angrist and Jörn-Steffen Pischke (2009) "Mostly Harmless Econometrics- An Empiricist's Companion", Princeton University Press, ISBN 9781400829828" (in particular, Chapter 4/5)

The course covers the intuition of these methods, but not the technicalities.

Additional course information

Thomas Epper is an assistant professor at the School of Economics and Political Science. Please find more information here: <https://www.thomasepper.com>

Petyo Bonev is an assistant professor of econometrics at the School of Economics and Political Science. Please find more information here: <https://sites.google.com/site/petyobbonev>

COVID-19 note: This course is scheduled to be taught face-to-face in the fall semester. In case that this course is disrupted by a second wave of the coronavirus pandemic or a local lockdown, teaching will take place according to university guidelines (e.g., online via Zoom). There are no changes necessary for the centrally organized exam.

Examination information

Examination sub part/s



1. Examination sub part (1/2)

Examination time and form

Decentral - Group examination paper (all given the same grades) (50%)

Examination time: term time

Remark

Problem sets to be solved in groups

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

2. Examination sub part (2/2)

Examination time and form

Central - Written examination (50%, 90 mins.)

Examination time: inter-term break

Remark

Sit-in exam (90 min.)

Examination-aid rule

Extended Closed Book

The use of aids is limited; any additional aids permitted are exhaustively listed under "Supplementary aids". Basically, the following is applicable:

- At such examinations, all the pocket calculators of the Texas Instruments TI-30 series and mono- or bilingual dictionaries (no subject-specific dictionaries) without hand-written notes are admissible. Any other pocket calculator models and any electronic dictionaries are inadmissible.
- In addition, any type of communication, as well as any electronic devices that can be programmed and are capable of communication such as notebooks, tablets, mobile telephones and others, are inadmissible.
- Students are themselves responsible for the procurement of examination aids.

Supplementary aids

-



Examination languages

Question language: English

Answer language: English

Examination content

Students are expected to make the following contributions:

- Problem sets (50%). There will be four problem sets throughout the course. The problem sets provide an opportunity for students to conduct data analysis and learn important statistical concepts. Note: Programming necessitates that you reach your own understanding of the problem and discover a path to its solution.
- Central exam (50%). The exam assesses the understanding of concepts and the ability to correctly interpret and criticize research results. Students should also be able to interpret R code and output.

Examination relevant literature

Materials covered on the exam are provided through Canvas. All basic readings will be made available prior to the lectures. All exam relevant parts will be made available online prior to the last lecture.

Please note

Please note that only this fact sheet and the examination schedule published at the time of bidding are binding and takes precedence over other information, such as information on StudyNet (Canvas), on lecturers' websites and information in lectures etc.

Any references and links to third-party content within the fact sheet are only of a supplementary, informative nature and lie outside the area of responsibility of the University of St.Gallen.

Documents and materials are only relevant for central examinations if they are available by the end of the lecture period (CW51) at the latest. In the case of centrally organised mid-term examinations, the documents and materials up to CW 42 are relevant for testing.

Binding nature of the fact sheets:

- Course information as well as examination date (organised centrally/decentrally) and form of examination: from bidding start in CW 34 (Thursday, 20 August 2020);
- Examination information (regulations on aids, examination contents, examination literature) for decentralised examinations: in CW 42 (Monday, 12 October 2020);
- Examination information (regulations on aids, examination contents, examination literature) for centrally organised mid-term examinations: in CW 42 (Monday, 12 October 2020);
- Examination information (regulations on aids, examination contents, examination literature) for centrally organised examinations: two weeks before the end of the registration period in CW 44 (Thursday, 29 October 2020).