Syllabus «Improving your statistical inferences»

Approved by the Academic Council
of the bachelor's programme "Political Science"
Protocol №25б от «26» June 2018

Author	Ilya M. Lokshin
Credits	5
Contact hours	0
Independent work hours	190
Year	4
Format of studying	blended learning

I. ABSTRACT AND LEARNING OBJECTIVES

This course aims to help you to draw better statistical inferences from empirical research. First, we will discuss how to correctly interpret p-values, effect sizes, confidence intervals, Bayes Factors, and likelihood ratios, and how these statistics answer different questions you might be interested in. Then, you will learn how to design experiments where the false positive rate is controlled, and how to decide upon the sample size for your study, for example in order to achieve high statistical power. Subsequently, you will learn how to interpret evidence in the scientific literature given widespread publication bias, for example by learning about p-curve analysis. Finally, we will talk about how to do philosophy of science, theory construction, and cumulative science, including how to perform replication studies, why and how to pre-register your experiment, and how to share your results following Open Science principles.

In practical, hands on assignments, you will learn how to simulate t-tests to learn which p-values you can expect, calculate likelihood ratio's and get an introduction the binomial Bayesian statistics, and learn about the positive predictive value which expresses the probability published research findings are true. We will experience the problems with optional stopping and learn how to prevent these problems by using sequential analyses. You will calculate effect sizes, see how confidence intervals work through simulations, and practice doing a-priori power analyses. Finally, you will learn how to examine whether the null hypothesis is true using equivalence testing and Bayesian statistics, and how to pre-register a study, and share your data on the Open Science Framework.

The course has the following main objectives:

- give students an introduction to the most widely used quantitative and qualitative data analysis methods;
- explain the quantitative and qualitative data analysis methods using real data and concentrating on complications that may occur during the analysis in real-life research;
- teach students how to organize their own research project using the knowledge obtained during the course;

The course will contribute to developing the following competences:

- ability to select appropriate methods of data analysis depending on the research question and types of empirical data;
- ability to select an appropriate data analysis software package to conduct quantitative and qualitative analysis of different types of data (statistical data, texts, pictures, media);
- ability to prepare empirical data for their further analysis in data analysis software packages;
- ability to formulate research hypotheses and construct models;
- ability to perform data analysis;
- ability to interpret the results of data analysis;
- ability to report research results to the audience.

Discipline studies are based on the following disciplines:
- Mathematics and statistics;
- Game Theory.

To master the discipline, students must possess the following knowledge and competencies:
• know the basic laws of classical algebra;
• know the simplest methods for solving mathematical problems;
• have the skills to work with qualitative and quantitative data analysis methods.

II. COURSE CONTENTS

Theme 1. Introduction + Frequentist Statistics.

Frequentism, Likelihoods, Bayesian statistics. What is a p-value. Type 1 and Type 2 errors.

Theme 2. Likelihoods & Bayesian Statistics.

Interview: Zoltan Dienes. Likelihoods. Binomial Bayesian Inference. Bayesian Thinking.

Theme 3. Multiple Comparisons, Statistical Power, Pre-Registration.
Type 1 error control. Type 2 error control. Interview Professor Dan Simons. Pre-registration.

Theme 4. Effect Sizes.

Effect Sizes. Cohen's d. Correlations.

Theme 5. Confidence Intervals, Sample Size Justification, P-Curve analysis.

Confidence Intervals. Sample Size Justification. P-Curve Analysis.

Theme 6. Philosophy of Science & Theory.

Philosophy of Science. The Null is Always False. Theory Construction.

Theme 7. Open Science.

Replications. Publication Bias. Open Science.

III. FORMS OF CONTROL

The grade for this course (G_{final}) is made up of the grade for the online-course ($G_{online-course}$) and final exam (G_{exam}).

The formula for calculating the final grade is the following:

$G_{final} = 0{,}6 \cdot G_{exam} + 0{,}4 \cdot G_{online\text{-}course}$

IV. EXAMPLES OF ASSESSMENT

Example of exam question:

1. When there is no true effect, and we simulate 100000 studies, what does the p-value distribution look like? P-values are normally distributed, centered on $p = 0.5$.

a. We will mostly observe very small p-values, and only very few high p-values.
b. We will mostly observe very high p-values, and only very few small p-values.
c. The p-value distribution is uniform.

2. If we have 90% power when we use an alpha level of 0.05, the power we will have when using an alpha level of 0.01 is always:

a. lower
b. the same
c. higher

V. RESOURCES

5.1 Basic resources

Discipline studies are carried out by students independently on the basis of an online course “Improving your statistical inferences” https://www.coursera.org/learn/statistical-inferences.

5.2 Software

№	Name	Access conditions
1.	Microsoft Windows 7 Professional RUS	*From the university’s internal network (contract)*
2.	Microsoft Office Professional Plus 2010	*From the university’s internal network (contract)*

5.3 Professional databases, information reference systems, Internet resources (electronic educational resources)

№	Name	Access conditions
	Internet resources (electronic educational resources)	

1.	Coursera	URL: https://www.coursera.org/

5.4 Technical Support for the Course

The classrooms for self-study in the discipline are equipped with personal computers with the ability to connect to the Internet and access to the electronic information and educational environment of the HSE.