Course syllabus «Longitudinal Data Analysis»

Approved by
Programme Academic Council
Protocol Nr. 05 from 27.08.2018

Author	Dr. Valentina Kuskova
Number of credits	4
Contact hours	48
Self-study hours	104
Course	1,2
Educational format	Without use of online course

I. Goals and Results of Mastering the Discipline; Prerequisites

This course is about quantitative methods, namely statistics, applied to social sciences. Specifically, we will focus on certain statistical competencies that help evaluate processes over time. I expect you to understand the basics of statistics you've learned previously in this course; everything else we will learn in this class. As you will see, we will use a lot of real-world datasets, and I am concerned more with your understanding on how statistic works as opposed to memorizing the formulas. This class will be unique in a sense that I will bring a lot of non-statistical material to help you understand the world of decision sciences.

As a result, students should:

Know:

- the theoretical foundation of longitudinal analysis
- modern applications of longitudinal analysis
- the variety of time-series models that are available to analyze real-life problems, starting with the simple OLS regression and ending with highly advanced models.

Be able to:

- Understand the meaning and use of longitudinal models.
- Present and/or interpret data in tables and charts.
- Understand and apply descriptive statistical measures to real-life situations.
- Understand and apply probability distributions to model different types of social processes.

Have:

- an ability to forecast future numbers based on historical data.
- an ability to use computer software to perform statistical analysis on data (specifically, STATA)
- an ability to resolve problems and recognize the most common decision errors and make tough decisions in a competent way.

Basic knowledge of introductory statistics are required for this course.

The basics of this discipline should be used all other program related courses

The course is strongly related and complementary to other compulsory courses provided in the first year (e.g. Applied Linear Models II, Contemporary Data Analysis) and sets a crucial pre-requisite for later courses and research projects as well as for the master thesis. The course gives students an important foundation to develop and conduct their own research as well as to evaluate research of others.

II. Content of the Course

Please note: due to their advanced nature, some topics will take more than one lecture to cover.

SESSION ONE: Introduction to the Framework of longitudinal data analysis

The Where, Why, and How of Longitudinal Data. Simple Linear Regression Model – A Review

SESSION TWO: Basics of Time Series I

Basics of Time Series Analysis. Static and Finite Distributed Lag models.

SESSION THREE: Basics of Time Series II

Trending, non-stationarity, serial correlation. Autoregressive (AR) proves and moving average (MA) process.

SESSION FOUR: ARIMA

Autoregressive integrated moving average model (ARIMA) with extensions. Box-Jenkins method for working with ARIMA.

SESSION FIVE: Advanced time-series models I

Cointegration. Equilibrium. Engle-Granger two-step procedure. Error correction models (ECM) and vector autoregression models (VAR). Reduced form VAR. Lag length selection and information criterion.

SESSION SIX: Advanced time-series models II

Structural vector autoregression models, including short-run (SVAR). Long-run restrictions. Structural equation models (SEM). The state-space approach to time series analysis. Predicted states, filtered states, smoothed states, forecasting.

SESSION SEVEN: Advanced time-series models III

Time-series with categorical predictors. Binary response. Random vs. fixed effects. Mixed model assumptions and estimation. Non-linear mixed effects. Observed marginal proportions, proportional and non-proportional odds.

SESSION EIGHT – Advanced time-series models IV

Panel and time series cross-sectional data (TSCS). Benefits of time-space data. Variable intercepts and slopes. Errors in the TSCS models. Heterogeneity and pooling. Fixed and random effects estimation.

III. Grading

Course grade will be completed as follows:

Course Element	**% Towards Final Grade**
Final Exam	**50%**
Final In-Class or Take-home exam (at the discretion of the instructor)	*50%*
Participation and responsibility grade	**50%**
Homework Assignments (5 x Varied points)	*20%*
In-Class Labs (9-10 x Varied points)	*20%*
Quizzes (Best 9 of 10, Varied points)	*10%*
Extra credit	As assigned
Total	100%

If the final grade is non-integer, it is rounded according to algebraic rules. If has a half (.5) at the end, we are rounding upward. Rounding of cumulative grades and other rounding issues are performed according to the HSE rules.

IV. Grading Tools

This class contains several assignments that test student knowledge and understanding throughout the course.

Quizzes

You cannot meaningfully participate in the seminar if you have missed my lecture and did not do any reading. Therefore, to encourage you to prepare for seminars, every seminar will have a quiz on the lecture material and all assigned readings for the week. This includes the very first seminar, which will focus on Lecture 1 material. You are allowed to miss any one quiz (skip a seminar, not prepare, etc.) – in other words, I will count the best 9 out of 10 quizzes that we will have. If you submit all ten, I will count best nine. All quizzes will be done online and submitted to me via SurveyMonkey (links will be given in class).

Important: I record IP addresses and only accept quizzes submitted from with the HSE IP address. Quizzes submitted from other locations are NOT counted towards your grade. In other words, to participate in a quiz, you have to be present in class.

In-class Labs

There will be a lab assignment in almost every seminar, depending on our progress. Since we will be learning SAS, and learning quickly, you will need to devote a substantial time to it. Seminar labs should help you with this task. At the end of the lab, you will submit your completed assignment for the day (or as much as you were able to complete) to me via LMS.

Homework assignments

There will be several homework assignments that will provide additional hands-on practice for the concepts we've learned in class and practiced during the seminar. Homeworks will be assigned as needed throughout the semester. All homework submissions must be done by the stated deadline via the LMS system.

Final exam

There will be comprehensive final exam in this course, which will consist of an analysis of a real-life dataset with problem formulation, appropriate tool selection, analysis and interpretation of results. More details will be provided during the course.

V. Resources

5.1 Main Literature

1. Taris, Toon W. A Primer in Longitudinal Data Analysis, SAGE Publications, 2000. ProQuest Ebook Central, URL https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=254689.
2. Derryberry, DeWayne R.. Basic Data Analysis for Time Series with R, John Wiley & Sons, Incorporated, 2014. ProQuest Ebook Central, URL https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=1719581.
3. Montgomery, Douglas C., et al. Introduction to Time Series Analysis and Forecasting, John Wiley & Sons, Incorporated, 2015. ProQuest Ebook Central, URL https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=1895570.
4. Tsay, Ruey S.. Analysis of Financial Time Series, John Wiley & Sons, Incorporated, 2010. ProQuest Ebook Central, URL https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=565117.
5. Turkman, Kamil, Manuel González Scotto, and Patrícia de Zea Bermudez. *Non-linear time series*. Springer International Pu, 2016. URL https://proxylibrary.hse.ru:2176/book/10.1007/978-3-319-07028-5. Springer Link.

5.2 Additional Literature

1. Beran, Jan. *Mathematical foundations of time series analysis: a concise introduction*. Springer, 2018. URL https://proxylibrary.hse.ru:2176/book/10.1007/978-3-319-74380-6. Springer Link.
2. Palma, Wilfredo. Time Series Analysis, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, URL https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=4517503.
3. Franses, Philip Hans, and Richard Paap. Periodic Time Series Models, Oxford University Press USA - OSO, 2004. ProQuest Ebook Central, URL https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=3052196.

5.3 Software

№ п/п	Name	Access conditions
1.	MicrosoftWindows 7 Professional RUS MicrosoftWindows 10	*From the university's internal network (contract)*

	MicrosoftWindows 8.1 Professional RUS	
2.	Microsoft Office Professional Plus 2010	*From the university's internal network (contract)*
3.	R, R studio	*Open access. URL: https://www.r-project.org/*
4.	STATA version 13 or above	*From the university's internal network (contract)*

5.3 Material and technical support

Classrooms for lectures on the discipline provide for the use and demonstration of thematic illustrations corresponding to the program of the discipline, consisting of:

- PC with Internet access (operating system, office software, antivirus software);
- multimedia projector with remote control.