Quantitative methods 2 - Linear and generalized linear models.

Instructor: Michal Kotnarowski, PhD (IFIS PAN and GSSR),
email: kotnarowski@ifispan.edu.pl

Academic year: 2025/2026, second semester

Planned course timetable:

Mondays 11.00-13.00

Office hours: Thursdays 14.00-16.00 (by prior appointment).

Teaching period: February 23, 2026 - June 8, 2026.

Format of the course: in-person classes, room 232.

General description.

The course will focus on the application of basic and intermediate regression techniques in social sciences. Various regression analyses are among the most commonly used analytical techniques in sociology, political science, and psychology. The critical skill of scholars in social sciences, regardless of substantive interests, should be an understanding of these techniques. The scholar, on the one hand, should be able to understand the work of other researchers applying these techniques, but on the other, should have the ability to use regression techniques correctly in their research.

The course assumes that participants have basic knowledge of descriptive and inferential statistics. During the course, participants will expand their statistical skills to an intermediate level. After completing the course, participants will be able to conduct regression analyses on their own at the level allowing for publication in academic journals. Moreover, the participants will gain the statistical foundations required to master more advanced analytical techniques, such as multi-level modelling, structural equation modelling, panel regression, time series analysis, event history analysis, or machine learning.

Goals of the course.

After completing the course, participants will be able to understand academic papers in which various regression models have been applied. Participants will learn how to interpret the published results of regression models. They will also gain the ability to critically evaluate the use of regression analyses in the work of other researchers. Finally, the course participants will be able to conduct their regression analyses correctly on their own, at least at an intermediate level.

Prerequisite knowledge.

Course participants should understand basic concepts of descriptive statistics such as mean, median, variance, standard deviation and conditional means. The practical parts of the course

will be conducted in R. Therefore, participants should have a basic knowledge of R as a statistical programming language and the interface of RStudio.

A detailed description.

The course will begin with an overview of the fundamental principles of statistics, including the basics of statistical inference and inferential reasoning, covering key concepts such as the Central Limit Theorem and the logic of constructing confidence intervals.

The next topic of the course will be the introduction of linear regression models, also known as ordinary least squares (OLS) models. In these models, the dependent (outcome) variable is a continuous variable defined on the interval scale. Participants will estimate these models, interpret their parameters, and assess the models' fit to the data. The regression models will then be extended by taking into account qualitative explanatory variables and introducing interactions between variables. The next meetings will concern the assumptions of the linear regression model, such as linearity, multicollinearity and heteroskedasticity. Participants will explore the meanings of these assumptions, the consequences of not meeting them, the methods of diagnosing whether the given assumption is met, and possible remedies for violations.

After the OLS framework is introduced, the course will present Analysis of Variance (ANOVA), treated as an extension of the OLS family of models.

In the second part of the semester, the course will cover regression models in which dependent variables are categorical. These are situations in which the dependent variable is either:

- a binary variable, when respondents select one out of two options (e.g., whether they voted in the last election);
- a nominal variable, when respondents select one out of three or more options (e.g., which party they voted for in the last election).

Generalized Linear Models (GLMs), which extend OLS by allowing non-normal outcome distributions and link functions, will be used to analyse this type of data. In particular, the course will include binary logistic regression and multinomial logit.

The course will focus on the practical application of the introduced statistical techniques. The emphasis will be placed on the presentation of regression analyses results both in tabular form as well as in the form of simple and complex statistical graphics. During the course, theoretical aspects of statistical models, which are crucial to their correct application, will be discussed.

Learning outcomes

The table below outlines the main learning outcomes.

Knows and understands		**Is able to**		**Is ready to**	
First stage generic descriptors (universal)		First stage generic descriptors (universal)		First stage generic descriptors (universal)	
P8U_W	the world's achievements in science and the resulting implications of this for practice	**P8U_U**	analyse and creatively synthesise scientific and creative achievements to identify and solve research problems as well as those related to innovative and creative activities; contribute new elements to these achievements;	**P8U_K**	conduct independent research which contributes to existing scientific and creative achievements; assume professional and public challenges taking into consideration: their ethical dimension, responsibility for their results and develop models

			independently plan one's own development as well as inspire the development of others; participate in the exchange of experiences and ideas, also in the international community		of good practice in such situations
Second stage generic descriptors typical for higher education		Second stage generic descriptors typical for higher education		Second stage generic descriptors typical for higher education	
P8S_WG	the world's achievements relating to advanced quantitative research methods at a level enabling the revision of existing paradigms; the main scientific developments of the quantitative research methods.	**P8S_UW**	creatively identify, formulate and innovatively solve complex problems using advanced quantitative methodology, especially: apply advanced quantitative research methods, draw conclusions on the basis of advanced quantitative research results perform critical analysis and evaluation of the results of scientific research, expert activities and other works applying advanced quantitative research methods; transfer the results of research applying advanced quantitative methods to the economic and social spheres	**P8S_KK**	critically evaluate the achievements of advanced quantitative research methods; critically evaluate one's contributions to the development of that field; recognize the value of knowledge in solving cognitive and practical problems
P8S_WK	the fundamental dilemmas of advanced quantitative research methods; the economic, legal, ethical and other essential conditions of conducting advanced quantitative research; basic principles of knowledge transfer to the economic and social sphere and commercialization of results of studies	**P8S_UK**	communicate on specialized topics to a degree that enables active participation in an international scientific environment; disseminate results of applications of advanced quantitative research methods, also to the general public.	**P8S_KO**	fulfill the social obligations of researchers; initiate activities on behalf of the public interest; think and act in an enterprising manner

	applying advanced quantitative research methods				
		P8S_UO	plan and implement one's own and a team's study implementing advanced quantitative research methods	**P8S_KR**	uphold and develop the ethos of the research communities, including: conducting research in an independent manner respecting the principle of the public ownership of academic research results, taking into account intellectual property rights
		P8S_UU	autonomously plan and act on behalf of personal development and inspire and organise the development of others.		

Course Requirements

1. Class Attendance: Regular attendance is required. Participation in class discussions and activities is essential to understanding the course content. If a student is unable to attend a class, they must inform the instructor in advance. Consistent attendance and active participation will contribute to the final grade. The course material is interconnected and follows a hierarchical structure, with later topics building on the knowledge from earlier sessions. Therefore**, being present in every class is crucial** to ensure continuous understanding and progress.
2. Assigned Readings: Before each class, students will be assigned specific readings crucial to the upcoming lecture and discussions. Students are expected to thoroughly read and understand the material before attending class. During class, students should be prepared to:
    - Discuss key concepts from the readings.
    - Summarize what they have learned.
    - Engage in a conversation about challenges, problems, or questions related to the assigned texts.

    The class will take the form of a detailed discussion of specific textbook sections assigned for that day. Without mastering these sections, participation in the class will not be meaningful. Therefore, **it is necessary to do the assigned readings thoroughly before each meeting**. Approximately **50 pages of reading** will be assigned for each class.
3. Practical Assignments in R: Students will be given practical assignments to complete in R. These assignments are designed to reinforce the statistical techniques and concepts covered in the lectures. Completed assignments must be **submitted the day before the next class**. Timely submission of these tasks is crucial, as they will be reviewed and discussed during the following class.

Grading Criteria

1. Class Attendance and Active Participation – 40% of the final grade. It includes regular attendance, engagement in class discussions, and demonstration that one has read the assigned chapters. What will be assessed is not so much whether the course participants understood the readings but the effort they put into understanding them. Active participation in discussions on the readings will be key to this part of the evaluation.
2. Timely Submission of Practical Assignments in R – 60% of the final grade. After some classes, students will be given practical assignments. Completing and submitting these assignments on time is essential. These tasks are designed to apply the statistical techniques learned in class.

The table below provides a grading scheme:

Grade	Percent	Mark
A	94-100	5
A-	90-93	5-
B+	87-89	4+
B	84-86	4
B-	80-83	4-
C+	77-79	3+
C	74-76	3
C-	70-73	3-
D+	65-69	2+
D	60-64	2
F	0-59	1

Policies on Attendance, Late Materials, and Make-Up

Attendance is mandatory and counts toward participants' classroom participation and attendance grade. Course participants should notify the instructor in advance if they cannot attend a given class session.

As a rule, this course meets in person. Under exceptional circumstances, such as documented participation in a short-term study/research visit or as a presenter at a scientific conference, and after discussing the case with the instructor in advance, attending up to three of the fifteen class sessions online may be possible.

All assignments must be turned in electronically via the Google Classroom portal by the specified deadlines. The instructor will accept late materials at his discretion and *only if* notified 24 hours before the deadline. Except for documented reasons (e.g., doctor's note), late assignments will incur a 10% penalty per day.

Course Information Platform

Google Classroom will be the primary platform for communication and information exchange between the instructor and students. All course-related materials will be posted there, including readings, presentations, links to external resources, and any additional materials.

Google Classroom will also be used to assign homework. Students are required to submit their assignments via Google Classroom; the instructor will not accept homework submissions sent via email. It is important to regularly check the platform for updates and new materials to stay informed and prepared for the course.

Software Requirement: R and RStudio

Throughout the course, students will use R and RStudio, powerful and free tools for statistical computing and data analysis. Students are required to install R and RStudio on their computers before the course begins to ensure they are ready for the practical assignments.

- R can be downloaded from here: https://cran.r-project.org/
- RStudio can be downloaded from here: https://posit.co/download/rstudio-desktop/

Please ensure both programs are installed and functioning properly before the first class. Participants who encounter difficulties during installation should contact the instructor for support before the course starts.

Detailed schedule of the course.

Date	Topic	Readings
February 23	1. Normal distribution and the Central Limit Theorem	QSS Chapter 6.
March 2	2. Confidence intervals and introduction to OLS regression	SDAM: Ch. 4
March 9	3. OLS regression - estimation, parameters, and goodness-of-fit measures	ARAGLM - Ch. 5-6, CAR - Ch. 4.1-4.4, 5.1-5.2
March 16	4. OLS regression - statistical inference	ARAGLM - Ch. 5-6, CAR - Ch. 4.1-4.4, 5.1-5.2
March 23	5. Regression with dummy variables and interaction terms	ARAGLM - Ch. 7; CAR - Ch. 4.5-4.9
March 30	6. Outliers and influential cases; nonlinearity	RD Ch. 4, 7; CAR - Ch. 8
April 13	7. Regression assumptions - collinearity and heteroskedasticity	RD Ch. 3, 8
April 27	8. Model of means - ANOVA	SDAM: Ch. 7; RCtS: Ch. 6
May 4	9. Repeated-measures ANOVA	SDAM: Ch. 10; RCtS: Ch. 8
May 11	10. Introduction to Generalized Linear Models - linear model vs. generalized linear model; linear predictor; link function	Long Ch. 3; ARAGLM Ch. 14.1
May 18	11. Maximum likelihood estimation; binary logistic regression vs. probit; interpretation of parameters; predicted probabilities	ARAGLM Ch. 15.1; Long Ch. 4
May 25	12. Binary logistic regression - goodness-of-fit measures; interaction terms; interpretation using statistical graphics	Long Ch. 4; Fox (2003); CAR - Ch. 6
June 1	13. Multinomial logit - interpretation of parameters; interaction terms; predicted probabilities; goodness-of-fit measures	Long Ch. 4; ARAGLM Ch. 14.2; Fox & Hong (2009)
June 8	Free week to catch up	

Handbooks

ARAGLM: Fox, John. 2016. *Applied Regression Analysis and Generalized Linear Models*. Third Edition. Los Angeles: SAGE.

CAR: Fox, John, and Sanford Weisberg. *2019 . An R Companion to Applied Regression.* Third edition. Los Angeles: SAGE.

Field: Field, Andy P., Jeremy Miles, and Zoë Field. 2012. Discovering Statistics Using R. London ; Thousand Oaks, Calif: Sage.

HiR: Kaufman, Robert L. 2013. *Heteroskedasticity in Regression: Detection and Correction*. Thousand Oaks, California: SAGE Publications.

Long: Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. 1st ed. Sage Publications, Inc.

QSS = Imai, Kosuke. 2017. Quantitative Social Science: An Introduction. Princeton; Oxford: Princeton University Press.

RD: Fox, John. 1991. *Regression Diagnostics*. Newbury Park, Calif: Sage Publications.

RCtS: Speekenbrink, Maarten. 2023. *An R Companion to Statistics: Data Analysis and Modelling*. https://mspeekenbrink.github.io/sdam-r-companion/index.html.

SDAM: Speekenbrink, Maarten. 2023. *Statistics: Data Analysis and Modelling*. https://mspeekenbrink.github.io/sdam-book/index.html.

Fox 2003: Fox, John. 2003. Effect Displays in R for Generalized Linear Models. Journal of Statistical Software.

Fox & Hong 2009: Fox, John, and Jangman Hong. 2009. Effect Displays in R for Multinomial and Proportional-Odds Logit Models: Extensions to the effects Package. Journal of Statistical Software.