Graduate School for Social Research

MethodsNET
Methods Excellence Network

Quantitative Methods Course: Generalized Linear Models, GLM

March 20-24, 2023

Graduate School for Social Research, Warsaw, Poland

Call for Applications: International Travel Awards for PhD Students at Higher Education Institutions outside of Poland

From Monday, March 20, to Friday, March 24 2023, the Graduate School for Social Research, GSSR (gssr.edu.pl), of the Polish Academy of Sciences' Institutes of Philosophy and Sociology (IFiS), Political Studies (IPS) and Psychology (IP), organizes the methodological course ***General Linear Models, GLM.*** This course is offered within the **Methods Excellence Network,** MethodsNET (www.methodsnet.org).

We invite PhD students in the social sciences at Higher Education Institutions (HEI) outside of Poland to apply for funding for travel and accommodation to participate in the **GLM** course at GSSR. Travel scholarships to GSSR are made possible by the PROM program of the Polish National Agency for Academic Exchange, NAWA (nawa.gov.pl).

Basics at a glance

Target audience:	PhD students in the Social Sciences
Instructor:	dr. Michał Kotnarowski, IFiS and GSSR
Dates:	Monday, March 20 – Friday, March 24, 2023
Location:	Room 268, Staszic Palace, Nowy Swiat 72, 00-330, Warsaw, Poland
Mode of participation:	In person (i.e. offline)
Language of instruction:	English
Scholarships:	Up to **12 scholarships** covering travel, accommodation and meal expenses for PhD students from foreign HEIs
Course attendance fees:	No Fees
Application deadline:	**Thursday, February 23, 2023**
Notification of acceptance:	March 7, 2023

Eligibility requirements

We invite PhD students who conduct social science dissertation research at HEIs outside of Poland to apply for participation in the GSSR GLM course Applicants must have the following documented prerequisite knowledge:

1. **Understanding the logic of inferential statistics**
   Participants should be familiar with hypothesis testing and concepts such as confidence intervals and significance level.
2. **Familiarity with rudiments of the linear regression model estimated using the Ordinary Least Squares method**
   Participants should be familiar with the logic of linear regression analysis, assumptions of the linear regression model and regression with dummy variables.
3. **Have at least a basic understanding of the R language**
   Participants should be able to import into R a dataset written in the SPSS format and run a linear regression model. You should be able to conduct a set of basic data manipulations in R, such as: selecting observations, selecting variables and computing

Familiarity with the use and interpretation of interaction terms in OLS regression is an advantage.

Application process

Organizers can provide up to 12 Scholarships to cover transportation costs to and from Poland to your country of residence (the scholarships covers insurance and travel expenses up to 5000 PLN for travel from outside Europe, and up to 3000 PLN for travel within Europe), accommodation and meals for the duration of the GSSR GLM course (successful applicants will be paid per diems for duration of the course up to 4840 PLN). Participation in the course is free of charge for any accepted student.

To apply, you need to submit the following materials electronically:

1. Academic CV

2. Cover letter (600-800 words) that discusses:

   a. Your experience with using quantitative methods for the social sciences in your research. Please address your familiarity with (i) hypothesis testing and main concepts in inferential statistics, (ii) OLS regression and its main assumptions.
   b. Your level of familiarity with R.
   c. How the GSSR GLM course contributes to your research goals, in particular, advancing your PhD research and/or work on a specific empirical paper.

3. The completed Application form: https://forms.gle/V3st9DMsE61FiAEn7

4. Copy of valid ID (e.g. Passport, National ID Card): a scan or high-resolution picture (JPEG, PNG, or PDF).

All materials that form the application are due **Thursday, February 23.** Please send them electronically to John Fells, john.fells@gssr.edu.pl

A Selection Committee composed of GSSR faculty and administration will review the applications. Applicants will be notified of the decision via email by March 7, 2023.

Selected participants will be asked to fill in paperwork once at GSSR, to receive the reimbursement for travel expenses and accommodation.

About the GLM Course

Instructor: **Michał Kotnarowski**

Michał Kotnarowski is an Assistant Professor at the Institute of Philosophy and Sociology, the Polish Academy of Sciences, a lecturer at GSSR, and the head of the European Social Survey (ESS) in Poland. Michał has extensive teaching experience in quantitative methods for the social sciences, including as instructor of the GLM course at the 2018-2020 Winter Schools organized by the European Consortium for Political Research, ECPR. His work on voting behavior, comparative politics and political methodology has been published in, among others, *Party Politics, Electoral Studies, Communist and Post-Communist Studies*, *Acta Politica*, and the *International Journal of Sociology*.

Course Dates & Times

Monday March, 20 – Friday, March 24 2023, 11:00 – 15:30 (including a break for lunch) *15 hours over five days*

Course Description

Researchers working in broadly defined social sciences often have to deal with analyses in which the dependent variable is not a continuous variable defined on the interval scale. These are situations in which the dependent variable is either:

1. a binary variable, when respondents select one out of two options (e.g., whether they voted in the last election)
2. a nominal variable, when respondents select one out of three or more options (e.g., which party they voted for in the last election)
3. an ordinal variable (e.g., when a respondent chooses an answer on the Likert scale) or
4. a variable counting the number of occurrences of a phenomenon (e.g., how many times a respondent participated in protest actions).

For this type of dependent variable, the appropriate technique are General Linear Models (GLMs), which are estimated in a different way from linear regression models.

Interpretation of GLMs is much more complex than for OLS models. Although GLMs are often used in social sciences, their use and correct interpretation still give researchers difficulties.

This course is an introduction to GLMs, and participants will gain practical skills related to their use. Instructor will also introduce theoretical aspects of GLMs so participants can understand and interpret them properly. The course will be framed in such a way to be understandable to those without rudimentary background in matrix algebra or calculus.

By the end of the course, participants will be able to:

- run GLMs on their own datasets
- interpret GLM regression coefficients and odds ratios
- assess the goodness of fit of the models
- estimate the uncertainty of predicted effects using simulations
- present the results using statistical graphics techniques.

Course Structure

Day 1

The regression model with a binary dependent variable. We start by discovering why it is inappropriate to use OLS models in such cases; in particular, which OLS model assumptions are not met, and what might be the negative consequences of using OLS models for this type of data. Next, I will show you how to generalize a linear model so that it can be applied to models with a limited dependent variable. You will learn the linear predictor and the link function. We close with a presentation of Maximum Likelihood Estimation as a technique for estimating logistic regression model parameters.

Days 2 & 3

You will develop practical skills related to the interpretation of the binary logistic regression model, i.e. the interpretation of regression coefficients and odds ratios. I will present measures of goodness of fit of the models, and various versions of pseudo-R-squared measures. I demonstrate the extension of additive logistic regression models by introducing interactions between independent variables. You will learn how to correctly interpret a logistics binary regression model that incorporates interaction terms, and how to report the results using predicted probabilities, in particular through statistical graphics.

Day 4

I will introduce models with a nominal dependent variable, i.e. multinomial logistic regression. We will learn how to estimate and correctly interpret such models.

Day 5

We will learn about techniques appropriate for an ordinal dependent variable, i.e. ordinal logistic regression, and regression models for counts. Poisson regression and negative binomial models.

I will illustrate the application of each method using analyses based on real-world data, presenting GLMs with their constraints and limitations.

Course requirements

The course assigns readings, and you will be given assignments. Students who want ECTS credits (2 ECTS) must complete practical exercises related to the techniques introduced on a given day. You can use your own data for these (strongly recommended) or data provided by the instructor.

The lab sessions and assignments use the open-source statistical software R, enabling efficient implementation and advanced interpretation of GLMs. R's graphical capabilities allow effective and relatively simple presentation of GLM results.

Day	Topic & Readings	Details
Day 1	Introduction to General Linear Models	**90-minute Workshop with elements of lecture** Linear model vs. general linear model, linear predictor, link function, Maximum Likelihood Estimation. **90-minute Lab session** Running first binary regression models.
	Long (1997), Chapter 3 Fox (2008), Chapters 14.1 and 15.1	
Day 2	Binary Logistic Regression	**90-minute Workshop with elements of lecture; 90-minute Lab session** Interpretation of parameters of binary logistic regression models, goodness of fit measures, interaction terms within a binary logistic regression model, predicted probabilities, measures of uncertainty of predicted effects.
	Long (1997), Chapter 4 Fox (2011), Chapters 5.1 – 5.3 Brambor et al. (2006)	
Day 3	Binary Logistic Regression - continuation	**90-minute Workshop with elements of lecture; 90-minute Lab session** Developing skills in the interpretation of binary logistic regression models. Presentation of logistic regression models using tools of statistical graphics.
	Fox (2003)	
Day 4	Models for Nominal Outcomes	**90-minute Workshop with elements of lecture; 90-minute Lab session** Multinomial logistic regression model. Interpretation of parameters of the model, goodness of fit measures, interaction terms, predicted probabilities, measures of uncertainty of predicted effects.
	Long (1997), Chapter 6 Fox (2008), Chapter 14.2 Fox & Hong (2009)	
Day 5	Models for Ordinal Outcomes and Count Data	**90-minute Workshop with elements of lecture; 90-minute Lab session** Ordinal logistic regression model. Poisson regression and negative binomial model. Interpretation of parameters of the models, goodness of fit measures, interaction terms, predicted probabilities/predicted values, measures of uncertainty of predicted effects.
	Long (1997), Chapters 5 and 8 Fox (2008), Chapter 15.2 Fox (2011), Chapter 5.5	

Software Requirements

Download the newest version of R

Download the newest version of R Studio

Hardware Requirements

Please bring your own laptop.

Literature

Agresti, A. (2007) *An introduction to categorical data analysis* (2nd ed.) Hoboken, NJ: Wiley-Interscience

Agresti, A. (2013) *Categorical data analysis* (3rd ed.) Hoboken, NJ: Wiley

Brambor, T., Clark, W. R., & Golder, M. (2006) Understanding Interaction Models: Improving Empirical Analyses *Political Analysis*, *14*(1), 63–82

Cameron, A. C., & Trivedi, P. K. (2013) *Regression analysis of count data* (2nd ed.) Cambridge ; New York, NY: Cambridge University Press

Fox, J. (2003) Effect Displays in R for Generalised Linear Models *Journal of Statistical Software*, *8*(15)

Fox, J. (2008) *Applied Regression Analysis and Generalized Linear Models* (2nd ed.) Sage Publications, Inc.

Fox, J., & Hong, J. (2009) Effect Displays in R for Multinomial and Proportional-Odds Logit Models: Extensions to the effects Package *Journal of Statistical Software*, *32*(1), 1–24

Fox, J., & Weisberg, H. S. (2011) *An R Companion to Applied Regression* (2nd ed.) Sage Publications, Inc.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013) *Applied logistic regression* (3rd ed.) Hoboken, New Jersey: Wiley

King, G., & Zeng, L. (2001) Logistic Regression in Rare Events Data *Political Analysis*, *9*(2), 137–163

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

Funding and organizational support for the GSSR GLM Course

International PhD student funding comes from the Polish National Agency for Academic Exchange, NAWA (nawa.gov.pl), "International Scholarship Exchange of PhD Candidates and Academic Staff," PROM, project (PPI/PRO/2019/1/00043/U/00001).

Organizational support for the course comes from GSSR administration, the Methods Excellence Network, MethodsNET (methodsnet.org/), IFiS PAN (ifispan.edu.pl), and CONSIRT of The Ohio State University and the Polish Academy of Sciences (consirt.osu.edu).