EUI DEPARTMENT OF POLITICAL AND SOCIAL SCIENCES

Intermediate Quantitative Methods

Time & place: **Seminar** | Mondays | 11.00-13.00 | Sala del Capitolo
                       LAB | Tuesdays | 11.00-13.00 | Sala del Capitolo

Given by: **Filip Kostelka**, Professor of Political Science
                       Arnout van de Rijt, Professor of Sociology

Teaching assistant: **Julie Fournier**, PhD Researcher

Administrative assistant: Pia Dittmar (Pia.Dittmar@eui.eu)

Compulsory seminar:

This seminar must be taken during the first or second year of the program. It can be substituted with "Introduction to Quantitative Methods", which runs in parallel to this seminar. It can also be substituted with "Introduction to Econometrics and Regression", offered by the Economics Department.

Course Description and Objectives

The objective of this seminar is to refresh and extend knowledge of basic quantitative methods typically used in the political and social sciences. The seminar covers statistical inference and ordinary least squares (OLS) regression, as well as cases where the basic OLS assumptions need to be relaxed and some approaches for dealing with these cases. Topics include statistical inference and hypothesis testing, linear regression and its assumptions, model specification, nested models (multilevel and panel regression), maximum likelihood estimation, and binary outcomes. Many of these topics are also covered in the "Introduction to Quantitative Methods" course, but this course goes into greater depth, seeking a more grounded understanding.

The seminar consists of weekly lectures and lab exercises. The seminar schedule is flexible so that the themes covered during the seminar do not completely follow a weekly schedule but some flexibility is left to extend the coverage of topics that require more attention.

Many of the assignments include exercises using statistical software and they can be done either with R or Stata (or some other software); the focus of the labs will be substantive and not coding. This means that the seminar participants have to be sufficiently familiar with statistical analysis software because instruction on coding will be limited. A number of exercises using software include data simulation and application of the covered methods on simulated data. The basics of data simulation will be covered in the first lab.

The seminar is suitable for researchers who have previously taken introductory courses in quantitative methods and know the basics of linear regression, and who are sufficiently comfortable with a statistical software package like R or Stata. Being comfortable with high-school level mathematics is important so it is a good idea to take Mathematics for Political and Social Scientists in September to brush up. For those with no or little previous experience with quantitative methods, or who want to brush up their basic skills,

we recommend they instead take Introduction to Quantitative Methods, which runs in parallel to this seminar. Those who have extensive prior background in quantitative methods can consider Introduction to Econometrics and Regression in the Economics Department. The Intermediate Quantitative Methods seminar is also open to students who previously took Introduction to Quantitative Methods.

Learning Outcomes

Having a thorough grasp of the logic of inference and hypothesis testing, being familiar with well-known problems with inference in scientific practice: multiple comparisons, replication, publication bias, having an intuition for the basic mathematics behind OLS; understanding the Gauss-Markov assumptions on which OLS is built, when their violations is problematic and what can be done about it in everyday research practice; understanding issues with handling binary outcomes; understanding how fixed and random effects models may be used for inference; understanding the basics of maximum likelihood estimation.

Readings

Readings are assigned for each week and the course participants should become prepared to class. There is no single course textbook. The topics for each class are covered in various textbooks and we point to alternative sources that you can use; typically it is enough to use just one of them. The library has five copies of the Agresti book.

The main source books include:

Agresti, Alan (2018). Statistical Methods for the Social Sciences, Fifth edition. Pearson.

Draper, N. R., & Smith, H. 1998. Applied Regression Analysis. Wiley.

Gordon, Rachel (2015). Regression Analysis for the Social Sciences, 2nd Edition. Routledge.

Imai, Kosuke (2017). Quantitative Social Science: An Introduction. Princeton.

Wooldridge, Jeffrey M (2015). Introductory econometrics: A modern approach, 4-6th Edition. Nelson Education.

For those who want to learn more about simulation: Carsey, Thomas M., and Jeffrey J. Harden. *Monte Carlo simulation and resampling methods for social science*. Sage Publications, 2013.

Requirements for credit

This course requires regular attendance (80% excluding valid documented absences), active participation, and timely handing in of assignments. Researchers are expected to do the compulsory readings listed in the schedule below ahead of each meeting. Finally, researchers have to pass a **final exam** (passing grade 60%) that takes place on **December 19, 2022**

Seminar topics

I   Inference and hypothesis testing (Week 1)

- The logic of statistical inference
- Hypothesis testing and statistical significance

Compulsory readings:

Ioannadis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine* 2(8): e124. 10.1371/journal.pmed.0020124

Colquhoun, David (2014). An investigation of the false discovery rate and the misinterpretation of p-values. *Royal Society Open Science*, 1(3): 140216. **https://doi.org/10.1098/rsos.140216**

Optional readings:

Troeger, V. E. (2019) 'To P or Not to P? The Usefulness of P-Values in Quantitative Political Science Research'. *Swiss Political Science Review* 25 (3): 281–87.

Bernardi, F., Chakhaia, L. & Leopold, L. (2017) 'Sing me a song with social significance': The (mis)use of statistical significance testing in European sociological research. *European Sociological Review*, 33(1): 1-15.

Agresti, Ch. 2, (3), 4–5.3, pp. 173-4 (ch. 6); Imai, Ch. 3.1-3.3, 6, 7.1

II   Bivariate regression (Weeks 2-3)

- Regression anatomy "connecting the dots": derivation and manual calculation of b0, b1, SEs, T and p values, conf intervals, rho, df, R^2, TSS/ESS/RSS, F

Compulsory readings:

Agresti, 9.1-9.4

Optional readings:

Draper & Smith, Ch 1; Imai, Ch 4; Gordon Ch 5; Wooldridge, Ch 2.

III   Multivariate regression (Weeks 4-5)

- Algebraic intuitions for regression anatomy in 3D and beyond
- Regression assumptions and diagnostics
- Variable transformations
- Interactions

Compulsory readings:

Agresti, 9.5, 14.2-14.3, 14.5-14.6

Optional readings:

Draper & Smith, Ch 4, 5; Gordon, Ch 6, 9, 11; Angrist and Pischke Ch 3.2-3.3; Wooldridge, Ch 3-5,8.

IV   Model specification (Weeks 6-7)

- Confounders, mediators and colliders
- Measurement error
- Sensitivity analysis

Compulsory readings:

Brambor, Thomas, William Roberts Clark, and Matt Golder. "Understanding interaction models: Improving empirical analyses." *Political analysis* 14.1 (2006): 63-82.

Cunningham, Scott (2021): *Causal inference: the mixtape*. Yale University Press. Pp. 96-118. [e-book in the library].

Optional readings:

Morgan, Stephen L. & Winship, Christopher. 2007. Counterfactuals and causal inference. Cambridge University Press. Chs. 1.6, 3.1. [e-book in the library]

Hainmueller, Jens, Jonathan Mummolo, and Yiqing Xu. (2019) "How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice." *Political Analysis* (2019) vol. 27:163–192

Elwert, Felix & Winship, Christopher (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology* 40:31-53. **https://doi.org/10.1146/annurev-soc-071913-043455**

Gordon, Ch 8, 10; Wooldridge Ch 6, 9.

VI   Nested data: Multilevel and panel regression analysis (Weeks 7-8)

- Nested data and non-independent observations
- Multilevel data
- Random and fixed effects estimation
- Between-within model

Compulsory readings:

Bell, A., Fairbrother, M., & Jones, K. (2019). Fixed and random effects models: making an informed choice. *Quality & quantity*, 53(2), 1051-1074.

Optional readings:

Snijders, T.A.B., & Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage. Ch 1-5, 7, 10, 11, 15

Halaby, C. N. (2004). Panel models in sociological research: Theory into practice. *Annual review of sociology*, 507-544.

Allison, P. D. (2009). *Fixed effects regression models.* SAGE

Wooldridge Ch 10-12.

VII   Binary outcomes (Week 9)

- Linear probability and logistic regression models
- Probit
- Calculating maximum likelihood by hand

Compulsory readings:

Mood, C. 2010. Logistic regression: Why we cannot do what we think we can do, and what to do about it. *European Sociological Review* 26(1): 67-82. **https://doi.org/10.1093/esr/jcp006**

Optional readings:

Kuha, J. & Mills, C. 2020. On group comparisons with logistic regression models. Sociological Methods and Research 49(2) **https://doi.org/10.1177%2F0049124117747306**

Background Readings:

Agresti, Ch.4.2, 5.2, 8

VIII   TSCS Data (Week 10)

Compulsory reading:

Wilson, S. E., & Butler, D. M. (2007). A Lot More to Do: The Sensitivity of Time-Series Cross-Section Analyses to Simple Alternative Specifications. Political Analysis, 15(2), 101–123. **https://doi.org/10.1093/pan/mpl012**

Optional readings:

Beck, Nathaniel, and Jonathan N. Katz. 1995. "What to Do (and Not to Do) with Time-Series Cross-Section Data." The American Political Science Review 89 (3): 634–47.

Petersen, Trond. 2004. Analyzing panel data: Fixed and random effects models. Pp. 332-345 in hardy, M & Bryman, A. (eds) Handbook of Data Analysis. Sage.

Background readings:

Woolridge, chapters 13 and 14., Wooldridge, J. M. (2012). Econometric analysis of cross section and panel data. MIT Press, chap. 10-11.