

Intermediate Quantitative Analysis

Given by Juho Härkönen (Professor of Sociology) & Arnout van de Rijt (Professor of Sociology)

Teaching assistant: Mirko Wegemann, PhD Researcher

Contact: Monika.Rzemieniecka@eui.eu

Time & place:

(A couple of the sessions may take place in a hybrid form due to room restrictions).

Lecture: Monday 11:00-13:00

04 Oct	11:00 - 13:00	Refectory, Badia Fiesolana, Badia Fiesolana
11 Oct	11:00 - 13:00	Theatre, Badia Fiesolana, Badia Fiesolana (+ zoom room 2)
18 Oct	11:00 - 13:00	Refectory, Badia Fiesolana, Badia Fiesolana
25 Oct	11:00 - 13:00	Theatre, Badia Fiesolana, Badia Fiesolana (+ zoom room 2)
2 Nov	9:00 - 11:00	Refectory, Badia Fiesolana, Badia Fiesolana
08 Nov	11:00 - 13:00	Refectory, Badia Fiesolana, Badia Fiesolana
15 Nov	11:00 - 13:00	Refectory, Badia Fiesolana, Badia Fiesolana
22 Nov	11:00 - 13:00	Sala del Capitolo, Badia Fiesolana, Badia Fiesolana (+ zoom room 2)
29 Nov	11:00 - 13:00	Refectory, Badia Fiesolana, Badia Fiesolana
06 Dec	11:00 - 13:00	Refectory, Badia Fiesolana, Badia Fiesolana

Lab: Tuesday 11:00-13:00

05 Oct	11:00 - 13:00	Sala del Capitolo, Badia Fiesolana, Badia Fiesolana (+ zoom room 2)
12 Oct	11:00 - 13:00	Sala del Capitolo, Badia Fiesolana, Badia Fiesolana (+ zoom room 2)
19 Oct	11:00 - 13:00	Refectory, Badia Fiesolana, Badia Fiesolana
26 Oct	11:00 - 13:00	Sala del Capitolo, Badia Fiesolana, Badia Fiesolana (+ zoom room 2)
2 Nov	11:00 - 13:00	Refectory, Badia Fiesolana, Badia Fiesolana
09 Nov	11:00 - 13:00	Sala del Capitolo, Badia Fiesolana, Badia Fiesolana (+ zoom room 2)

16 Nov	11:00 - 13:00 Refectory, Badia Fiesolana, Badia Fiesolana
23 Nov	11:00 - 13:00 Refectory, Badia Fiesolana , Badia Fiesolana
30 Nov	11:00 - 13:00 Refectory, Badia Fiesolana
07 Dec	11:00 - 13:00 Refectory, Badia Fiesolana, Badia Fiesolana

Compulsory seminar:

Must be taken during the first two years of the programme. Can be substituted with "Introduction to Econometrics and Regression", offered by the Economics Department. Those with no or little previous experience with quantitative methods, or who want to brush up their basic skills can instead take Introduction to Quantitative Methods, which runs in parallel to this seminar.

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 first part of the seminar covers statistical inference and ordinary least squares (OLS) regression. The second part covers cases where the basic OLS assumptions need to be relaxed and some approaches to dealing with these situations. Topics cover statistical inference and hypothesis testing, linear regression and its assumptions, model specification, nested models (multilevel and panel regression), maximum likelihood estimation, and binary outcomes.

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 are deemed to 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. Many of the 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 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; knowing how to deal with and understanding ongoing debates about how to handle 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 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 required readings listed in the schedule below ahead of each meeting.

Seminar topics

I Inference and hypothesis testing

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

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

All should read:

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 reading:

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.

II Bivariate regression

- Regression anatomy “connecting the dots”: derivation and manual calculation of b_0 , b_1 , SEs, T and p values, ρ , df, TSS/ESS/RSS, F

Sources: Imai, Ch. 4; Agresti, Ch. 9; Gordon Ch 5; Wooldridge, Ch. 2.

Optional reading: Chapter 1 in Draper & Smith “Fitting a Straight Line by Least Squares”

III Multivariate regression

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

Sources: Agresti 14.2–14.3; Gordon, Ch. 6, 11; Angrist and Pischke Ch. 3.2-3.3; Wooldridge, Ch. 3-4.

Optional reading: Chapters 4 and 5 in Draper & Smith

IV Fixing regression problems

- Outliers and influential observations
- Nonlinearity
- Heteroscedasticity
- Multicollinearity

Sources: Agresti 14.5–14.6; Gordon, Ch. 9, 11; Wooldridge Ch. 5-6, 8.

V Model specification

- Confounders, mediators and colliders
- Interactions
- Measurement error

Sources: Gordon, Ch. 8, 10; Wooldridge Ch. 6-7, 9.

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>

Additional readings:

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

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

VI Nested data: Multilevel and panel regression analysis

- Nested data and non-independent observations
- ANOVA and multilevel data
- Random and fixed effects estimation
- Between-within model
- Random slopes
- Applications (sibling and twin studies, etc)

Sources: 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; Wooldridge Ch 10-12.

Optional:

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.

Wooldridge, J. M. 2012. *Econometric Analysis of Cross-Section and Panel Data*. Ch. 10-11.

VII Binary outcomes

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

Sources: Agresti, Ch.4.2, 5.2, 8

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>