Quantitative Methods in the Social Sciences 2 (QMSS2 and SO7033), Spring 2024
Version 2024 02 22
Max Thaning, Siddartha Aradhya, Sunnee Billingsley, and Gunnar Andersson

Course topics and design

The course covers advanced applications of linear regression (e.g., marginal effects, decomposition methods, measurement error), extensions of linear regression (multilevel and panel data models), and models for discrete outcomes (binary, ordered, and multinomial logit). The course emphasizes the craftsmanship of conducting and interpreting these methods. The course requires prior knowledge of linear and logistic regression from QMSS1 (Quantitative Methods in the Social Sciences I, SO7032).

QMSS2 is an advanced level regression course. This means that a basic and intermediate understanding of regression (such as that offered by QMSS1) is required. It is also necessary to know the Stata program package: we will not help with basic level stata skills. For those who need a Stata refresher nonetheless, we recommend the IDRE facilities at UCLA (https://stats.idre.ucla.edu/stata/).

The course design centers on active student participation. Lectures will only give an introduction (1-hour max). The bulk of theoretical learning will take place by reading and discussing texts and conducting analyses in labs or assignments.

Labs are in a DIY (do-it-yourself) form. The lab will be conducted independently by the students following instructions (doing it in a group is a good idea). The teaching focuses on giving feedback on and discussing the lab. Expect the lab to take at least two hours of independent work to conduct. This preparation is necessary to benefit from the feedback seminar.

The literature and article seminars are mandatory. Each student will be assigned a paper to present (max 10 min) and prepare five questions to discuss during the seminar (while still reading all papers). Any student who cannot attend will instead write a report (max 2 pages), see further instructions under assignments.

The examination is in the form of group and individual assignments, which will be of a formative kind (i.e., focused on furthering skills, not only a tick of what is understood vs. not understood). You will find the grading criteria below.

The literature consists of both methodological and research article texts, all in article form. The textbook Regression analysis and causal inference by Best and Wolf (2014b) is available online (for free via SU computer) http://methods.sagepub.com/book/regression-analysis-and-causal-inference and print (for purchase).

Note that the use of Chat GPT or other LLM services is not allowed. If teachers suspect any use of these services to answer assignments, we have the option to proceed with an oral examination.

Course readings

[See schedule overview for the course outline]

Best and Wolf (2014b) is of handbook type; it contains independent chapters on specific topics written by leading scholars. We will also use some chapters in an older handbook; Bryman and Hardy (2004/2009). For reference purposes, Morgan (2013) is also a very comprehensive handbook (we provide some suggestions for alternative readings from this book below).

Refresher (not part of curriculum, but indicative of course requirements)
Stolzenberg (2004, p. 165-180) or Wolf and Best (2014).

Course introduction
Mustillo, Lizardo and McVeigh (2018): whole article

Literature seminars

Literature seminar 1: Marginal effects, decomposition, measurement error
Stolzenberg (2004): Assumptions and computational requirements, and Assumptions (178-180), Nonlinear and non-additive models (188-203)
Jaccard and Dodge (2004): whole chapter (237-257)
Jann (2008): whole article, except Estimation of sampling variances (458-460). 464-468 can be read very briefly.
Solon (1989): mainly part I
Baron and Kenny (1986)

For those unacquainted with regression diagnostics: read Fox (1991) and the chapters on Non-Normally Distributed Errors, Non-constant Error Variance and Nonlinearity. An alternative on non-linear effects is Lohmann (2014)

Literature seminar 2 – Inference New order 2021
Stolzenberg (2004): Regression inference about population parameters (180-188)
Wolf and Best (2014): Statistical Inferences of Regression Results
Little (2004): whole article
Aho, Derryberry and Peterson (2014): whole article
Bernardi, Chakhaia and Leopold (2017): whole article
Wasserstein and Lazar (2016): whole article

Literature seminar 3 – LPM and logit with issues New order 2021
Best and Wolf (2014a): whole article
Mood (2010): whole article
Karlson, Holm and Breen (2012): whole article
Long and Mustillo (2018): whole article

Literature seminar 4 – Ordered logit, multinomial logit, and choice models New order 2021 Long (2014): whole chapter (skip the "Adjacent Category Logit Model", "Stereotype Logit Model")
Hoffman and Duncan (1988): whole article

Literature seminar 5: Fixed effects methods, multilevel methods New order 2021
Brüderl and Ludwig (2014): whole chapter
Allison (2009): Chapters 1 and 2
Hox and Wijngaards-de Meij (2014): whole chapter except Statistical Tests and Multilevel Analysis of Dichotomous Data

An alternative reading on fixed effects is Firebaugh, Warner and Massoglia (2013) or Petersen (2004)

Literature seminar 6 – Causality
Goldthorpe (2001): whole article
Gangl (2010): whole article

Article seminars

Article seminar 1 – Non-linear models New order 2021
Breen et al. (2009) : whole article
Chan and Goldthorpe (2007) : whole article

Article seminar 2 – Linear models New order 2021
England et al. (1988)
Daymont and Andrisani (1984)
Sampson, Raudenbush and Earls (1997)

Grading criteria

The course grade is based on a total assessment of all assignments. Each assignment examines one or more of the expected learning outcomes, and each expected learning outcome is examined through one or more assignments. Which expected learning outcomes are examined by which assignments are given in Table 1 below. Expected learning outcomes are graded for each task in three steps: "Very Good (VG)", "Good with some shortcomings (G)" and "Not sufficient".

The final grade is an average of the results for each graded learning outcomes (each "X" in Table 1). However, *a non-sufficient result on an expected learning outcome is not offset by a strong result on another learning outcome*. The weights are VG = 2, G =1 and the averaging is first done within assignments, and then across. The grade limits are:

A: $2.0 < \text{average} \leq 1.8$
B: $1.8 < \text{average} \leq 1.6$
C: $1.6 < \text{average} \leq 1.4$
D: $1.4 < \text{average} \leq 1.2$
E: $1.2 < \text{average} \leq 1.0$

Expected learning outcomes [Translated from Swedish Original]
Knowledge and understanding: the student will

1. have good knowledge of advanced applications of linear models (such as interactions, transformations, margin effects, hypothesis test, decomposition methods, measurement errors) focusing on how the models are interpreted, as well as the statistical basics of these approaches
2. have good knowledge of linear model extensions (multilevel regression and panel data analyzes), focusing on how the models are interpreted, as well as the statistical basics of these methods
3. have good understanding of models for discrete outcomes (linear probability models, binary logit, ordered logit and multinomial logit), focusing on how the models are interpreted, as well as the statistical basics of these methods
4. understand the basic problems of causal inference in social sciences.
5. have knowledge of different research traditions within quantitative social sciences

Skills and competence:

6. the student will independently be able to perform, present and interpret results for advanced linear model applications
7. independently be able to perform, present and interpret the results for linear model extensions
8. independently be able to perform, present and interpret results for discrete outcomes models.
9. practically as well as theoretically analyze model specification and functional form, and address potential problems and sources of error, be able to compare and evaluate different regression models.

Values and approaches: the student is able to

10. assess and critically evaluate the results of published quantitative analyses in social science research based on the methods discussed in the course
11. understand which methods are appropriate as well as unfit to use, given data, dependent variables, and research questions.

Table 1. Expected learning outcomes by assignment. New order 2021

	ass 1	ass 2	ass 3	ass 4	ass 5	ass 6
	Margins, Oaxaca, ME	Inference	Logit, LPM	Ologit, Mlogit	FE, multilevel	Causality
1. Knowledge - linear model applications	X	X				
2. Knowledge - linear model extensions (multilevel, panel)					X	
3. Knowledge - discrete outcome models			X	X		
4. Understand causal inference						X
5. Knowledge research traditions		X	X			X
6. Perform, present and interpret - linear model applications	X	X				
7. Perform, present and interpret - linear model extensions (multilevel, panel)					X	
8. Perform, present and interpret - discrete outcome models			X	X		
9. Model specification and evaluation	X	X	X			
10. Assess evaluate published studies					X	
11. Understand feasibility (RQ, data, depvar)		X	X		X	

References

Aho, Ken, DeWayne Derryberry, and Teri Peterson. 2014. "Model selection for ecologists: the worldviews of AIC and BIC." *Ecology* 95:631-636. http://www.jstor.org/stable/43495189

Allison, Paul D. 2009. *Fixed effects regression models*: SAGE publications. http://methods.sagepub.com/book/fixed-effects-regression-models

Baron, Reuben M, and David A Kenny. 1986. "The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations." *Journal of personality and social psychology* 51:1173. https://search.proquest.com/docview/614307043

Bernardi, Fabrizio, Lela Chakhaia, and Liliya Leopold. 2017. "'Sing Me a Song with Social Significance': The (Mis)Use of Statistical Significance Testing in European Sociological Research." *European Sociological Review* 33:1-15. http://dx.doi.org/10.1093/esr/jcw047

Best, Henning, and Christof Wolf. 2014a. "Logistic Regression." Pp. 153-172 in *The SAGE Handbook of Regression Analysis and Causal Inference*, edited by Henning Best and Christof Wolf. London: SAGE Publications Ltd. http://methods.sagepub.com/book/regression-analysis-and-causal-inference/n8.xml

— (Eds.). 2014b. *The Sage handbook of regression analysis and causal inference*: SAGE. http://methods.sagepub.com/book/regression-analysis-and-causal-inference

Breen, Richard, Ruud Luijkx, Walter Müller, and Reinhard Pollak. 2009. "Nonpersistent Inequality in Educational Attainment: Evidence from Eight European Countries." *American Journal of Sociology* 114:1475-1521. http://www.journals.uchicago.edu/doi/abs/10.1086/595951

Brüderl, Josef, and Volker Ludwig. 2014. "Fixed-effects panel regression." Pp. 327-357 in *The SAGE handbook of regression analysis and causal inference*, edited by Henning Best and Christof Wolf: Sage. https://methods.sagepub.com/book/regression-analysis- and-causal-inference/n15.xml

Bryman, Alan, and Melissa A Hardy. 2004/2009. *Handbook of data analysis*. London: Sage. http://methods.sagepub.com/book/handbook-of-data-analysis

Chan, Tak Wing, and John H. Goldthorpe. 2007. "Social Status and Newspaper Readership." *American Journal of Sociology* 112:1095-1134. http://www.journals.uchicago.edu/doi/abs/10.1086/508792

Daymont, Thomas N., and Paul J. Andrisani. 1984. "Job Preferences, College Major, and The Gender Gap in Earnings." *Journal of Human Resources* 19:408-428. https://www.jstor.org/stable/145880

England, Paula, George Farkas, Barbara Stanek Kilbourne, and Thomas Dou. 1988. "Explaining Occupational Sex Segregation and Wages: Findings from a Model with Fixed Effects." *American Sociological Review* 53:544-558. http://www.jstor.org/stable/2095848

Firebaugh, Glenn, Cody Warner, and Michael Massoglia. 2013. "Fixed Effects, Random Effects, and Hybrid Models for Causal Analysis." Pp. 113-132 in *Handbook of Causal Analysis for Social Research*, edited by Stephen L. Morgan. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-6094-3_7

Fox, John. 1991. *Regression Diagnostics*. Thousand Oaks, California: SAGE Publications, Inc. http://methods.sagepub.com/book/regression-diagnostics

Gangl, Markus. 2010. "Causal inference in sociological research." *Annual review of sociology* 36:21-47

Goldthorpe, John H. 2001. "Causation, Statistics, and Sociology." *European Sociological Review* 17:1-20. http://dx.doi.org/10.1093/esr/17.1.1

Hoffman, Saul D., and Greg J. Duncan. 1988. "Multinomial and Conditional Logit Discrete-Choice Models in Demography." *Demography* 25:415-427. http://www.jstor.org/stable/2061541

Hox, Joop, and Leoniek Wijngaards-de Meij. 2014. "The Multilevel Regression Model." in *The SAGE Handbook of Regression Analysis and Causal Inference*, edited by Henning Best and Christof Wolf. London: SAGE. http://methods.sagepub.com/book/regression-analysis-and-causal-inference/n7.xml

Jaccard, James, and Tonya Dodge. 2004. "Analyzing contingent effects in regression models." in *Handbook of data analysis*, edited by Mellisa A Hardy and Alan Bryman: SAGE. http://methods.sagepub.com/book/handbook-of-data-analysis/n10.xml

Jann, Ben. 2008. "The Blinder-Oaxaca decomposition for linear regression models." *The Stata Journal* 8:453-479. http://www.stata-journal.com/article.html?article=st0151

Karlson, Kristian Bernt, Anders Holm, and Richard Breen. 2012. "Comparing Regression Coefficients Between Same-sample Nested Models Using Logit and Probit: A New Method." *Sociological Methodology* 42:286-313. http://smx.sagepub.com/content/42/1/286.abstract

Little, Jani S. 2004. "Goodness-of-Fit Measures." in *The SAGE Encyclopedia of Social Science Research Methods*, edited by Michael Lewis-Beck, Alan Bryman, and Tim Futing Liao. Thousand Oaks, California. http://methods.sagepub.com/reference/the-sage-encyclopedia-of-social-science-research-methods/n377.xml

Lohmann, Henning. 2014. "Non-linear and non-additive effects in linear regression." Pp. 111 in *The SAGE Handbook of Regression Analysis and Causal Inference*, edited by Henning Best and Christof Wolf. http://methods.sagepub.com/book/regression-analysis-and-causal-inference/n6.xml

Long, J Scott. 2014. "Regression models for nominal and ordinal outcomes." Pp. 173-204 in *The Sage handbook of regression analysis and causal inference* edited by Henning Best and Christof Wolf: Sage. http://methods.sagepub.com/book/regression-analysis-and-causal-inference/n9.xml

Long, J. Scott, and Sarah A. Mustillo. 2018. "Using Predictions and Marginal Effects to Compare Groups in Regression Models for Binary Outcomes." *Sociological Methods & Research* 0:0049124118799374. https://journals.sagepub.com/doi/abs/10.1177/0049124118799374

Mood, Carina. 2010. "Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It." *European Sociological Review* 26:67-82. http://esr.oxfordjournals.org/cgi/content/short/26/1/67

Morgan, Stephen L. 2013. *Handbook of causal analysis for social research*: Springer. https://doi.org/10.1007/978-94-007-6094-3

Mustillo, Sarah A., Omar A. Lizardo, and Rory M. McVeigh. 2018. "Editors' Comment: A Few Guidelines for Quantitative Submissions." *American Sociological Review* 83:1281-1283. https://journals.sagepub.com/doi/abs/10.1177/0003122418806282

Petersen, Trond. 2004. "Analyzing panel data: Fixed-and random-effects models." Pp. 331-345 in *Handbook of data analysis*, edited by Mellisa A Hardy and Alan Bryman: SAGE. http://methods.sagepub.com/book/handbook-of-data-analysis/n14.xml

Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls. 1997. "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy." *Science* 277:918-924. http://www.sciencemag.org/cgi/content/abstract/277/5328/918

Solon, Gary. 1989. "Biases in the Estimation of Intergenerational Earnings Correlations." *The Review of Economics and Statistics* 71:172-174. http://www.jstor.org/stable/1928066

Stolzenberg, Ross M. 2004. "Multiple regression analysis." Pp. 165-207 in *Handbook of data analysis*, edited by Mellisa A Hardy and Alan Bryman: SAGE. http://methods.sagepub.com/book/handbook-of-data-analysis/n8.xml

Wasserstein, Ronald L., and Nicole A. Lazar. 2016. "The ASA's Statement on p-Values: Context, Process, and Purpose." *The American Statistician* 70:129-133. https://doi.org/10.1080/00031305.2016.1154108

Wolf, Christof, and Henning Best. 2014. "Linear regression." in *The SAGE Handbook of Regression Analysis and Causal Inference*, edited by Christof Wolf and Henning Best. London: SAGE Publications Ltd. http://methods.sagepub.com/book/regression-analysis-and-causal-inference

NB! All links provided requires a connection via the Stockholm university domain.