Methods II, MACIS, Fall 2017:
Essential Mathematics, Basic Statistics, and Linear Regression

Dr. Dominik Schraff
dominik.schraff@eup.gess.ethz.ch

Dr. Liam McGrath
liam.mcgrath@ir.gess.ethz.ch

Tuesdays, 14:00-17:00
Room: IFW B 42

Dates

Weekly, starting 19.09.

Description

This course has several goals. The first part of the course serves as a refresher in essential math and statistics, ensuring that all students have the necessary background knowledge for more advanced topics. This part also offers an introduction to the statistical software R, which is used for later modules of the MACIS program. The second part of the course builds on the first and offers a thorough treatment of the classical linear regression model and extensions. Both parts also introduce students to simple simulation methods using R.

Assessment

There will be one home assigment with applied regression analysis in R, counting 60% towards the overall grade, as well as a mid-term written exam, counting 40%. The exam will be held on November 7.

Readings

The course mainly relies on two textbooks:

- Moore, Will H. & Siegel, David A. 2013. A Mathematics Course for Political & Social Research. Princeton: Princeton University Press.
- Wooldridge, Jeffrey M. 2009. Introductory Econometrics: A Modern Approach. Mason OH: Cengage Learning.

There are three further textbooks that complement the material to this course (i.e. are not essential). There are Peter Kennedy's *A Guide to Econometrics* and William Greene's *Econometric Analysis*. The Kennedy book focuses far more on the intuition of the material we are covering here, separating this from the specific technical details. Therefore it serves as a solid

companion piece. The Greene book is probably the most comprehensive and thorough textbook available for econometrics. It is more technically demanding than Wooldridge, but this enables a far clearer and more succinct presentation of the materials covered in this course. Consider it a good investment if you a) have a solid mathematical background b) plan to continue with statistics long into the future. Finally, there is Thomas M. Carsey's and Jeffrey J. Harden's *Monte Carlo Simulation and Resampling Methods for Social Science*. It has a nice approach to use R and simulation methods to understand linear regression and its assumptions. It also provides a (very) short introduction to R and uses example code throughout the book.

Software

All students will need to install R on their laptops and bring their laptops to class. We recommend installing R as part of the somewhat more user-friendly RStudio environment. *Laptops with R installed will be needed from session 1 onwards.* For a fairly detailed discussion of R, see *An Introduction to R*.

Miscellany

This course is designed to require very little background knowledge in math and statistics. The most important requirement is dedication in attempting to understand unfamiliar material. Maths and statistics are a language like any other, and thus cannot be learned by skim reading textbooks before a final examination. It will take time to understand and be comfortable with this language. In addition, the course material is cumulative, meaning lectures will frequently refer back to concepts covered earlier in the course. Therefore, continual attendance and outside reading are essential to succeed.

Given these particular demands, there will be two open sessions in weeks 7 and 14. These offer a chance to revisit material in the previous weeks that was not clear. Suggestions for topics to be covered should be emailed to the relevant lecturer 5 days before the date of the open session.

Sessions

Part 1: Essential Mathematics and Basic Statistics

1. Introduction
   (a) Course overview and learning goals
   (b) Notation
   (c) Basic arithmetic and algebra
   (d) R introduction (first part)

   Readings:

   - Moore & Siegel: Ch. 1 & 2
   - A (very) short introduction to R

2. Calculus

   (a) Differentiation and the integral
   (b) R introduction (second part)

   Readings:

   - Moore & Siegel: Ch. 5-8

3. Probability

   (a) Basic probability theory
   (b) Calculating probabilities
   (c) Distributions (discrete)

   Readings:

   - Moore & Siegel: Ch. 9 & 10

4. Random variables and descriptive statistics

   - Distributions (continuous)
   - Moments and summary statistics
   - Exploratory data analysis (EDA)

   Readings:

   - Moore & Siegel: Ch. 11
   - Wooldridge: Appendix B.3 & B.4

5. Estimators

   (a) Populations, parameters, and random sampling
   (b) Finite sample properties
   (c) Asymptotic sample properties

   Readings:

   - Wooldridge: Appendix C.1–C.3
   - Stock & Watson: Ch. 2 (recaps and nicely combines topics on probability, random variables, and estimators)

6. Hypothesis Testing

   (a) The normal and related distributions
   (b) Confidence intervals and hypothesis testing
   (c) P-values and pitfalls

   Readings:

   - Wooldridge: Appendix C.5–C.7, B.5

- Wasserstein, Ronald L. and Nicole A. Lazar (2016) *The ASA's Statement on p-Values: Context, Process, and Purpose* The American Statistician 70(2): 129-133.

7. Open Session

8. Written Exam (November 7, 14:00, IFW B 42).

Part 2: The Classical Linear Regression Model and Extensions

9. Bivariate Empirical Measures of Association

   (a) Scalar derivation of linear regression

   (b) Basic properties of regression

   (c) Goodness of fit

   Readings:

   - Wooldridge: 2.1–2.3, 2.6
   - King, Gary (1986) *How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science* American Journal of Political Science 30: 666–687. http://j.mp/jFQ4Zl

10. CLRM I

    (a) Matrix algebra refresher

    (b) Gauss-Markov assumptions: bias, efficiency, consistency, BLUE.

    (c) Hypothesis tests: normally distributed errors assumption, testing hypotheses about coefficients

    Readings:

    - Matrix Algebra: Wooldridge Appendix D.1–D.3 or Gill Ch. 3
    - Substance: Wooldridge: 3.3–3.5, 4.1–4.3, 5, Appendix E

11. CLRM IIa: Violations of Gauss-Markov Assumptions – Non-Spherical Errors

    (a) Relation to Gauss-Markov: which assumptions are violated, what does this imply.

    (b) Specific Forms: Heteroscedasticity, Autocorrelation, Spatial Correlation.

    (c) Diagnostics: graphical, statistical tests

    (d) Solutions: standard error adjustments, explicit modelling.

    Readings:

    - Wooldridge: 8.1–8.4
    - Beck, Nathaniel and Jonathan N. Katz (1995) *What to do (and not to do) with Time-Series Cross-Section Data.* American Political Science Review 89 (3): 634–647
    - King, Gary, and Margaret E Roberts (2015) *How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About It.* Political Analysis 23 (2): 159–179. http://j.mp/1BQDeQT

12. CLRM IIb: Violations of Gauss-Markov Assumptions – Endogeneity

    (a) Relation to Gauss-Markov: which assumptions are violated, what does this imply.
    (b) Specific Types: omitted variables, measurement error, simultaneity, selection.
    (c) Diagnostics: can we even diagnose with data?
    (d) Solutions: A brief tour.

    Readings:

    - Wooldridge: 3.3, 9.4, 9.5, 15.1
    - Clarke, Kevin A. (2005) *The Phantom Menace: Omitted Variable Bias in Econometric Research* Conflict Management and Peace Science 22: 341–352. https://www.rochester.edu/college/psc/clarke/CMPSOmit.pdf

13. CLRM III: Specification Choices

    (a) Interaction effects, functional forms.
    (b) Appropriate construction of uncertainty measures.
    (c) Multiple restriction tests.

    Readings:

    - Wooldridge: 6.2, 7.4, 4.5
    - Thomas Brambor, William Roberts Clark, and Matt Golder. (2006) *Understanding Interaction Models: Improving Empirical Analyses.* Political Analysis 14: 63–82. http://mattgolder.com/interactions

14. Open Session