Methods II, MACIS, Fall 2015:
Essential Mathematics, Basic Statistics, and Linear Regression

Jørgen Bølstad
joergen.boelstad@eup.gess.ethz.ch

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

Tuesdays, 10:15-13:00
Room: ML F 34

Dates

15.09; 22.09; 29.09; 06.10; 13.10; 20.10; 27.10; ~~03.11~~; 10.11; 17.11; 24.11; 01.12; 08.12; 15.12.

Description

This course serves several purposes. 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.

Assessment

There will be one home assigment with applied regression analysis in R, as well as a written exam at the end of the term. Each will count 50% towards the overall grade. The exact dates will be determined during the first week of the term.

Textbooks

Gill, J. (2006). *Essential Mathematics for Political and Social Research.* Cambridge University Press, Cambridge.

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach.* South-Western, Cengage Learning, Mason, OH, 5th edition.

There are also two further textbooks that complement the material to this course (i.e. are not essential). These 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.

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 2 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 13. 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

   Basics: Notation, arithmetic, functions, polynomial functions, logarithms.

   Readings:

   - Gill: Ch. 1

2. Calculus & Introduction to R

   (a) Scalar Calculus: Limits, changes and derivatives, derivative rules, areas, slices, integrals, partial derivatives, Taylor series expansion.

   (b) Introduction to R: Objects and types, libraries, performing operations, using functions, loading data, creating random data, subsetting and ordering matrices, combining data, summarising, plotting, using scripts, loops. (Some topics will most likely be saved for later sessions).

   Readings:

   - Gill: Ch. 5 [ex. 5.7], 6.1–6.4.2
   - A (very) short introduction to R

3. Probability Theory

   (a) Probability Theory: Counting rules and permutations, sets and operations on sets, the probability function, calculations with probabilities, conditional probability and Bayes law, independence, odds.

   (b) Introduction to R continued.

   Readings:

   - Gill: Ch. 7

4. Random Variables

   Levels of measurement, distribution functions, probability mass functions, probability density functions, measures of central tendency, measures of dispersion, correlation and covariance, expected value.

   Readings:

   - Gill: Ch. 8 [ex. 8.9]

5. Estimators and Assessment Criteria

   Populations, parameters, and random sampling, estimators and estimates, unbiasedness, the sampling variance of estimators, efficiency, MSE, consistency, asymptotic normality, the central limit theorem.

   Readings:

   - Wooldridge: Appendix C.1–C.3

6. Hypothesis Testing

   The normal and related distributions, confidence intervals, hypothesis testing, problems with p-values and current practices, frequentism versus Bayesianism.

   Readings:

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

7. Open Session

– Reading Week –

Part 2: The Classical Linear Regression Model and Extensions

8. 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

9. 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

10. 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

11. 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

12. 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

13. Open Session