Quantitative Research Methods

Koç University

Spring 2022

Syllabus

Instructor	**Class Schedule**
David Carlson	Monday 14.30 – 17.10
dcarlson@ku.edu.tr	**Office Hours**
Office: CASE 140	Monday 12.00 – 14.00 (or by appointment)
Teaching Assistant	**Online Access**
Oğuz Bulut Kök	https://ku.blackboard.com (for grades and readings)
okok20@ku.edu.tr	https://github.com/carlson9/QRM (for in-class material)

Introduction

This course is designed to primarily familiarize the student with advanced techniques in political methodology. As part of that goal, the statistical software R will be learned by the student and used to implement all of the discussed methods. While it is impossible to cover every method, part of the class is learning how to understand methods through appropriate mathematical notation that the student has never encountered. There is also therefore a mathematical modeling component. In this way, at the end of the semester the student is expected to be well-versed in political methodology.

Prerequisite: INTL 501 (or equivalent).

Required Book

This class requires extensive reading to be adequately prepared for class. While most readings are available online, we will be working with one text book. The book is written by two prominent political methodologists, and therefore uses political examples throughout. Readings, including those from the book, are listed on the schedule below. The details of the book are as follows:

Gill, Jeff, and Michelle Torres. 2020. *Generalized Linear Models: A Unified Approach.* 2nd Edition. Sage.

The book comes with an `R` package with all of the code and data used in the book. When doing the readings, work through the code and ensure you either understand everything, or come to class with questions. The earlier version of the text is freely available and can be used instead, but note that no reference to software is included in the earlier version. Details of the accompanying package are as follows:

`GLMpack` available on CRAN or at: https://github.com/smtorres/GLMpack.

We will also be using `Shiny` apps (which can be run through a browser) to gain a better intuition for the topics covered in class. These are taken from Metzger's *Using Shiny to Teach Econometric Models* (2020). The book is available pre-print at https://www.dropbox.com/s/p4mre993k3tn and all of the apps are available at https://metzgersk.github.io/shinyElement/. As with the Gill and Torres book, the code is available through the apps. Be sure to read through the code and output.

Requirements and Grading

Grades will not be rounded, these represent strict cut-offs. In the rare event of, for example, exactly a 90, the higher grade will be assigned. Pluses and minuses will be applied at the instructor's discretion and will only be used if there are clear separations within a given

grade. **Note that the Koç suggested grades are not followed in this course.**

A	90–100
B	80–90
C	70–80
D	60–70
F	<60

1) *Homework: 50%*

Both graded and ungraded homework assignments will be assigned throughout the semester. The ungraded assignments will not be checked, but it is essential to complete them for success in the course and to adequately learn the material. Four graded assignments will be checked. They will be posted on Blackboard at least one week before they are due. It is strongly encouraged that you start the homeworks as early as possible. Late work will not be accepted. The work must be completed before the class meets for the week of the due date.

2) *Thesis Development / Final Project: 50%*

A major component of the course is to develop methodological and presentation skills for the development of your thesis or a final project. Rather than work on unrelated research questions, the knowledge obtained in the course should be applied to your topic of interest. The methodological rigor and presentation style will be the key determinants of your grade for this section. In order to track your progress and keep you on-track, we will discuss your goals throughout the semester. If, at the end of the semester, you do not have a working thesis or project, you will be required to write a report detailing your progress and submit this for a final grade.

Course Schedule

Please note this schedule is subject to change.

Week 1: OLS review

Stats: Mathematical properties and assumptions of OLS
`R:` Loading a dataset, running a linear model, analyzing the output, linear algebra
Math: Matrix operations, simple derivatives

Reading:

1) App `leastSq`

2) Read through and run the code from https://www.scribbr.com/statistics/linear-regression-in-r/

3) Read and work through the math in https://www.cliffsnotes.com/study-guides/algebra/linear-algebra/matrix-algebra/operations-with-matrices

Week 2: Violations to the Gauss-Markov assumptions and variable transformations

Stats: Implications of violations, introduction to common solutions
R: Determining optimal variable transformations, transforming variables
Math: Exponentials, logarithms, and their derivatives
Reading:

1) Apps `whySurv`, `olsApp`, and `linRegEstms`

2) Ch. 14 of *Applied Statistics with R* (Dalpiaz 2020), available at https://daviddalpiaz.github.io/appliedsta markdown-7

3) Work through some of the math and read section 3.9 of *Calculus Volume 1* (Open-Stax), available at https://openstax.org/books/calculus-volume-1/pages/3-9-derivatives-of-exponential-and-logarithmic-functions

Week 3: Probability distributions, mathematical notation, other prerequisites

Graded homework 1 is due at the beginning of class.
Stats: Connecting probability distributions to OLS
R: Drawing from a distribution, calculating PDF/PMF and CDF, plotting densities
Math: Notation, understanding common distributions
Reading:

1) Gill and Torres Ch. 1

2) "Common Probability Distributions: The Data Scientist's Crib Sheet" (Owen 2018), available at https://medium.com/@srowen/common-probability-distributions-347e6b945ce4

3) Skim the Wikipedia page https://en.wikipedia.org/wiki/List_of_probability_distributions and click on some distributions

Week 4: The exponential family

Stats: Putting common distributions into exponential family form
R: More advanced plotting, introductory simulations, variable types and dealing with messy data
Math: Further derivatives, introduction to integration
Reading:

1) Gill and Torres Ch. 2

2) "Introduction to Integration" available at https://www.mathsisfun.com/calculus/integration-introduction.html (this is very elementary, but is actually fairly intuitive)

Week 5: Likelihood theory and the moments

Stats: MLE, the mean and variance of the exponential family
R: The structure of an `R` package
Math: The variance function, maximum likelihood estimation
Reading:

1) Gill and Torres Ch. 3

2) Apps `mleLM` and `mleLogit`

3) Chapter "Package structure" in *R packages* (Wickham 2015) available at http://r-pkgs.had.co.nz/packag

Week 6: Linear structure and the link function

Graded homework 2 is due at the beginning of class.
Stats: Generalization of OLS
R: Running models with a link function, reporting the results, prediction function
Math: Further distributions
Reading:

1) Gill and Torres Ch. 4

2) Review previous readings

Week 7: Estimation procedures

Stats: Interpreting and presenting the output of a GLM
`R:` Further GLMs, further plotting
Math: Likelihood ratio tests, introduction on the math behind estimation techniques
Reading:

1) Gill and Torres Ch. 5

2) Pick three articles (preferably from top journals) that present at least one GLM, and look at the methods write-up. Bring the articles to class.

Week 8: Residuals and model fit

Stats: Deviance functions, measuring and comparing goodness of fit
`R:` Analyzing deviances, determining outliers, sensitivity analyses
Math: The asymptotic properties of fit statistics
Reading:

1) Gill and Torres Ch. 6

2) Chapter "Outlier Treatment" from *r-statistics.co* (Prabhakaran 2017) available at http://r-statistics.co/Outlier-Treatment-With-R.html

Week 9: Extensions to generalized linear models

Graded homework 3 is due at the beginning of class.
Stats: Quasi-likelihoods, fractional regression models, Tobit models, zero-inflated models, hurdle models
`R:` Implementing extensions using packages
Math: Quasi-likelihoods, probabilistic modeling notation
Reading:

1) Gill and Torres Ch. 7

2) All of the remaining apps (Ch. 5 and Appendix C)

Week 10: Temporal models

Stats: Formalization of growth models, survival models, other temporal models
R: Implementation of temporal models, dealing with temporal data
Reading:

1) Chapter "Time Series Analysis" from *r-statistics.co* (Prabhakaran 2017) available at
   http://r-statistics.co/Time-Series-Analysis-With-R.html

2) Review the app `whySurv`

3) "Survival Analysis with R" (Rickert 2017) available at https://rviews.rstudio.com/2017/09/25/survival-
   analysis-with-r/. Be sure to work through the code as well.

Week 11: TSCS analyses

Stats: Random effects, fixed effects, standard error adjustments, GPs
R: `lme4`, implementing fixed effects with different contrasts, packages for error adjustment
Reading:

1) Run through the code and read (including the links) of the slides at https://dss.princeton.edu/training/P
   Note that the procedures are for learning purposes and not necessarily the best practice.

2) Bates, Douglas et al. 2015. "Fitting Linear Mixed-Effects Models Using `lme4`." *Journal
   of Statistical Software.* Available at https://cran.r-project.org/web/packages/lme4/vignettes/lmer.pdf

Week 12: Introduction to Bayesian statistics

Graded homework 4 is due at the beginning of class.
Stats: Conceptual differences between frequentist and Bayesian approaches, advantages of
the Bayesian approach
R: Introduction to Bayesian packages, implementing Bayesian models, checking for conver-
gence, presenting output
Math: Bayes' Theorem, building blocks of Bayesian models
Reading:

1) Chapter 1 from *An Introduction to Bayesian Thinking* (Clyde et al. 2020), available at
   https://statswithr.github.io/book/

Week 13: Conjugate priors and derivations

Stats: Overview of common models and their conjugate priors
R: Further implementation of Bayesian models
Math: Deriving posteriors with conjugate priors
Reading:

1) Chapter 2 from *An Introduction to Bayesian Thinking* (Clyde et al. 2020), available at https://statswithr.github.io/book/

Week 14: Implementing Stan in R

Stats: Overview of common priors and their properties
R: Coding a Stan block, implementing it in R, plotting output, convergence diagnostics, parallelizing code
Math: Properties of probability densities
Reading:

1) Chapter 7 from *An Introduction to Bayesian Thinking* (Clyde et al. 2020), available at https://statswithr.github.io/book/