Tools and Sources

Intended for POLI492 students

Charles Breton
Department of Political Science
University of British Columbia
cbreton@alumni.ubc.ca

Introduction

In this document, you will find sources and tools that can all be considered to be optional for those of you who want to write a thesis using quantitative methods. Do not feel compelled to use any of them, that's not the point. They are just here if you need them. Some of the readings are truly facultative, others less so. A good example of the former is everything that has to do with R and ggplot2. You can write a good thesis without using R or without representing your tables as graphs. Making beautiful graphs will not make a bad thesis better either. It's icing on the cake. Other readings however can be considered to be more important. I strongly encourage you to read articles or books that are on the specific method you want to use. For instance, if your dependant variable is categorical, the Long and Freese (2006) book is a must.

Note that I will try to update this document as we progress through the year.

Basics

We will cover the basics of regression analysis in class but if you are curious and want to read/learn/know more, these books can be useful. Most of them are about the stuff that's going on behind the scene when you run a regression in Stata. Do not buy any of them (especially not Wooldridge, it's $248 on Amazon). They are all available at the library.

Wooldridge, J. M. (2012). *Introductory Econometrics: A Modern Approach 5th edition.* SouthWestern.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press, Princeton, NJ.

Freedman, D. A. (2009). *Statistical Models: Theory and Practice.* Cambridge University Press, Cambridge.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press, Cambridge.

In addition, the Sage series on quantitative applications for the social sciences has more than 100 little monographs on specific topics. The required readings for the quantitative lecture are from this series. Most of them will be available at the library.

Data Analysis Software

We will give you access to the latest version of Stata. As for R, it is a free software that can be downloaded on CRAN. In both cases, there's a large number of forums dedicated to questions and problems. This means that whatever your problem is, there's a strong chance that you are not the first to face it and someone might have already answered the question. Just google your problem and you are bound to find a solution.

Stata

Later in the term (November 14), a whole lecture will be dedicated to an introduction to Stata. In the meantime, you can have a look at the UCLA modules and the Princeton's tutorial. They are really good places to start if you are new to Stata and you can also refer back to them when you start playing with your own data.

R

The Cookbook for R is a good place to start. You can also look at:

Fox, J. and Weisberg, S. (2011). *An R Companion to Applied Regression.* Sage, Thousand Oaks, CA

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press, Cambridge

You may find it easier to use R through R Studio which is an interface that makes interacting with R a bit more user-friendly.

How do I do this?

The following references are supplementary readings if you don't find what you are looking for in the different Stata tutorials. There are many more, usually published by Stata Press. Just look them up on the library website.

Binary or Categorical Variables

Long, J. S. and Freese, J. (2006). *Regression Models for Categorical Dependent Variables Using Stata.* Stata Press

Time-Series

Becketti, S. (2013). *Introduction to Time Series Using Stata.* Stata Press

Workflow

If you are new to any of the statistical softwares, this will help you get off the right foot. It will also probably save you time by decreasing the amount of "trial and error" you will need to go through to make it work (although a good amount of trial and error is always healthy!).

Stata: Long, J. S. (2009). *The Workflow of Data Analysis Using Stata.* Stata Press

R: Hadley Wickham's tidy data

Interpretation

Depending on your estimation technique, your statistical models may produce coefficients that are hard to interpret (coefficients from an ordered probit for example). The Clarify package in Stata helps you transform these coefficients into quantities of interest that better convey the information. The following source explain how it works:

King, G., Tomz, M., and Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science*, 44(2):347–361

The Zelig package does the same in R (although it also does much more).

Interactions

Braumoeller, B. F. (2004). Hypothesis testing and multiplicative interaction terms. *International Organization*, 58(4):807–820

Brambor, T., Clark, W. R., and Golder, M. (2006). Understanding interaction models: Improving emprirical analyses. *Political Analysis*, 14:63–82

Data Visualization

You can use Stata to generate tables and graphs. The UCLA Stata modules on this topic can be a great help. However, the ggplot2 package in R does wonders. The learning curve might be steep but you will be rewarded for your effort. Many questions about ggplot2 can be answered by looking at the documentation on the website or on R Cookbook for Graphs or just by Googling them (or Binging them or whathever).

Other sources:

Kastellec, J. P. and Leoni, E. L. (2007). Using graphs instead of tables in political science. *Perpectives on Politics*, 5(4):755–771

Wickham, H. (2009). *ggplot2: Elegant Graphic for Data Analysis*. Springer, New York

Other softwares and tools for your thesis

You are about to start a project that will probably be the longest and most detailed paper you've ever written. This type of work may demand that you organize yourself better. These are tools available for free that may be useful for that.

Dropbox: If you still don't have a dropbox account (or another similar service), do it now. The only disadvatange to using such a service is that you lose an excuse. You will not be able to say that you lost all your work because your hard drive failed. External hard drives that you use for backups are also a good investment.

Workflowy: Workflowy is nothing more than a list maker. I find it useful to keep track of multiple projects or to organize the structure of a longer paper.

Thomas Basboll: This is a blog with ideas and tips on how to organize your writing process.

Skim: Useful to take notes on PDF. Tree saver.

The inspiration for this document comes from a similar but more detailed document produced by Kieran Healy a sociology professor at Duke. Choosing Your Workflow Applications might be a bit too technical for what you actually need but have a look, you might find other useful tools not listed here.

Finally, this document was created using LaTeX a free typesetting system that lets you describe the structure or format of the document you are producing and "transforms" it into a PDF (well...it's a bit more complicated than that). Here again, the learning curve might be steep, less so if you have dabbled with HTML before. In other words, it's not going to save you time right off the bat but it might in the future.

LaTeX usually works in tandem with a bibliographical tool to manage your references. Bibdesk does that and comes with LaTeX when you download it. Even if you don't write your paper using LaTeX, I strongly encourage you to familiarize yourself with a reference manager software such as Endnote or Refworks. These tools will save you tons of time. The library offers workshops on how to use them.