

POLSCI-702: Advanced Techniques of Political Science Research **4:00-6:40PM, Mondays 1/25/2016-5/10/2016, BOL 293**

Instructor: Dave Armstrong

Office: Bolton Hall # 666

Hours: Mondays 2-3PM or by appointment

E-mail: armstrod@uwm.edu

Course website: <http://quantoid.net/teachuwm/uwm702>

This course builds explicitly on the foundation that was laid in POLSCI 701. In that class, you learned the nuts and bolts behind statistical inference. In this course, we extend those tools to cover the linear model. The linear model is the workhorse of political science research. Nearly all of the techniques you see in published Political Science journals are either direct extensions or close relatives of the linear model. It is a very powerful tool for understanding relationships among a wide variety of types of variables. The course will teach you the ins and outs of estimation, interpretation, diagnostics and presentation of the linear model. If time permits, we will move on to models for binary dependent variables, but I suspect there will be little time for that.

This course will also push you in the direction of becoming producers of research. Your work in this course will be, at least in part, more like “real” data analysis. The idea is that by the end of this course, you will be able to effectively both evaluate quantitative work and present your own analyses.

As you are all graduate students, I expect that you will take this class seriously regardless of your inherent interest in the subject matter. I expect that you will attend class regularly, do the readings and ask questions when something is confusing. Pay particular attention to this last instruction - **ASK QUESTIONS IF YOU ARE CONFUSED**. In a similar fashion to last term, the things we learn in later weeks depend fundamentally on previous material. If you do not understand something it will cause problems later on. Please don't hesitate to ask either in class or outside of class if you are confused. If you miss class, you are responsible for learning the material you missed in a manner that proves least distracting for the other participants in the course. Also, the late work policy is that make-up exams are not given to graduate students and late papers are not accepted (rare exceptions may be allowed on a case-by-case basis).

As with 701, students will use **R** as the main piece of statistical software. You will also need a calculator for the course. This does not have to be a fancy calculator, it would be helpful if it had the square function (usually something like x^2) and the square-root function (usually either $\sqrt{\quad}$ or \sqrt{x}) is essential.

Your final grade in the course will depend on the following:

Quizzes	15%
Homework	30%
Short Data Papers	25%
Final Exam	30%

Quizzes

Given their wild popularity last semester, the weekly quizzes are back. We'll work on the same system - you can drop your three lowest quiz grades and there will be one every week covering the readings due that week.

Homework

You will get weekly assignments of various lengths and types some of the questions can/should be answered with pencil and paper while others will utilize the computer. You should consider your colleagues a resource and I encourage you to discuss the problem sets with your them. That said, each person must turn in their own, original answers to the homework problems. Further, you'll notice that only 30% of the grade comes from the homework, thus it makes sense for you to do your own work so you can pass the other 70% of the course.

Short Data Papers

One of the problems I have noticed in past semesters is that people get done with the course and are still lost about how to *do* empirical research on their own. The short data papers are meant to remedy this problem. I want you to come up with a question (presumably at least tangentially related to Political Science) and find some data to answer it. You don't need to do a literature review of any sort, just pose a question and find some data to answer it and answer it. There are five of these due throughout the semester. The first one will be a bit more guided than the others. We will go over an example of one of these in class. The due dates for the five papers are:

Exam

The final exam will test you on material from the entire course. This will be entirely a pencil and paper affair; you will not need to use the computer at all. To put your mind at ease now, you will not need to memorize formulae. For each exam, you will be allowed to bring with you one (1) standard sheet of letter-sized paper (8.5" × 11") with whatever types of information you want written on either side. You will obviously need to know which formulae apply to which problems, but that is not a task that requires memorization so much as it is a task that requires understanding of which procedures are appropriate for which types of problems.

Textbooks

The required text for the course is:

Fox, John. 2016. "Applied Regression Analysis and Generalized Linear Models, 3rd edi-

tion” Thousand Oaks: Sage Publications, Inc.

Fox, John and Sanford Weisberg. 2011. “An R Companion to Applied Regression, 2nd edition” Thousand Oaks: Sage Publications, Inc.

noindent I may provide some additional reading materials that will elucidate points covered in the lecture, but I will make those available electronically as they are needed.

Miscellaneous

You may obtain information on UWM policies concerning academic issues and course conduct here: <http://www.uwm.edu/Dept/SecU/SyllabusLinks.pdf>

Outline

The outline below offers a chronological list of topics covered, but each topic does not necessarily represent a complete lecture’s worth of material. Rather, some topics will require more than one class period to complete, while others can be covered in less than a single class period.

1. Loose Ends/Intro (1/25)
 - (a) Go over 701 Final
 - (b) Introduction to Statistical Modeling
2. Intro to Regression (2/1)
 - (a) What is Regression?
 - (b) Non-parametric Regression.
 - (c) Why parametric regression is “better”.
 - (d) Examining Data

Read: Fox (2016) Chapters 1-3

Homework: `Week2_Homework.pdf` (due 11/6)
3. Least Squares Regression (2/8-15)
 - (a) Simple Linear Regression in Scalar Form
 - (b) Multiple Regression
 - (c) Standardized Coefficients **Read:** Fox (2016) Chapter 5, Fox and Weisberg (2011) 149-157

Homework: `Week4_Homework.pdf` (due 11/13)

4. Least Squares Regression II (2/22)
 - * First Short Data Paper Due
 - (a) Statistical Properties of OLS Estimators
 - (b) Empirical versus Structural Relationships
 - (c) Regression in Matrix Form
 - Read:** Fox (2016) Chapter 6, Fox and Weisberg (2011) 190-197 (available on D2L)
 - Homework:** `Week5_Homework.pdf` (due 11/17)
5. Presentation and Categorical Independent Variables (2/29)
 - (a) Regression with Dummy Regressors
 - (b) Dealing with Categorical Regressors
 - (c) General advice on model presentation
 - Read:** Fox (2016) 120-131, Armstrong (2013) on D2L, Fox and Wesiberg (2011) 157-164 and 171-175
 - Homework:** `Week67_Homework.pdf` (due 3/21 - includes problems from both 2/29 and 3/7 lectures)
6. Testing and Presenting Conditional Hypotheses (3/7)
 - * Second Short Data Paper Due
 - (a) Specifying and interpreting conditional hypotheses
 - Read:** Fox (2016) 131-140, Fox and Weisberg (2011) 164-166 and 175-177 Brambor, Clark and Golder (2006) and Berry, Golder and Milton (2012) (available electronically on D2L)
7. Assessing the Linearity Assumption I (3/21)
 - (a) Diagnostic plots and methods for Non-linearity including transformations.
 - Read:** Fox (2016) Chapter 4, 277-290 and 291-296, Fox and Wesiberg (2011) 125-140, 285-293, 302-314.
 - Homework:** `Week8_Homework.pdf` (D4.1, D4.2 and D4.3, due 12/4)
8. Assessing the Linearity Assumption II (3/28)
 - (a) Non-parametric regression
 - Read:** Fox (2016) 451-462 and 476-508, Fox and Wesiberg (2011) 177-181

9. Assessing and Homoskedasticity and Normality (4/4) * Third Short Data Paper Due
(a) Diagnostic plots and methods for non-constant error variance and non-normality of the errors
Read: Fox (2016) 267-277, Fox and Weisberg (2011) 184-186, 314-317
Homework: `Week9_Homework.pdf` (due 4/11)
10. Outliers and Influential Observations (4/11) * Fourth Short Data Paper Due
(a) How to diagnose and treat influential observations.
Read: Fox (2016) Chapter 11, Fox and Weisberg (2011) 294-302
Homework: `Week10_Homework.pdf` (D11.1 and D11.3, due 4/25)
11. Logit and Probit I (4/18)
(a) Understanding the basics of models for binary dependent variable.
Read: Fox (2016) 335-355, Fox and Weisberg (2011) 229-240
Homework: `Week11_Homework.pdf` (D14.1, due 5/9)
12. Logit and Probit II (4/25)
(a) Presentation and interpretation of logit and probit models
Read: Fox “Effect Displays in R for Generalized Linear Models” (Available on D2L)
13. Clean Up, Final Thoughts 5/2
14. Class Cancelled 5/9.

Final Exam - 5/16, 4:00 PM, BOL 657

Below are the four short data paper assignments and their due dates.

1 Data Paper 1

This is the first of the short data assignments and it will be the most “guided” one. The others will ask you to go out on your own and answer questions you have. Here, I provide the question and tell you where to get the data. The question we’re going to answer is the following:

Q Are people’s evaluations of the way the president is handling the economy are related to political partisanship.

That is, do democrats feel that the president is doing a better job than independents and republicans. Perhaps this is not the most novel question, but novelty is hardly the goal - the goal is to come up with a question and find data to answer it. I want you to go to the ICPSR data archive <http://icpsr.umich.edu> and find the ABC News/Washington Post Monthly Poll, February 2010 (study #30202). You may have to create a ‘MyData’ account. You should choose the link to “download selected files” and click on the link for the Stata file. You can read the .dta file into R using the `read.dta()` function. There will also be a codebook there as well that will have more detail about the question wording and response coding. Make sure that responses coded DK (Don’t Know), No Opinion, System Missing are recoded in R as `NA`. Also, I don’t want to see R output here. I want to see the R code attached at the end, but I want to see you present the information in a format that is informative, but that omits any superfluous/redundant information. Imagine someone was paying you to answer this question form them as a consultant. Present to me what you would present in that situation.

2 Data Paper 2

Pose a question about the relationship between a continuous dependent variable and one continuous and one binary independent variable. Find data to operationalize the different pieces and estimate the linear regression model. Present and interpret your findings. What did you find as the answer to your question?

3 Data Paper 3

1. Pose a conditional hypothesis where one independent variable is continuous and the other is binary.
2. Pose a conditional hypothesis where both independent variables are continuous

In both cases, collect data that will allow you to evaluate the hypotheses (the two models can have the same dependent variable if you like) and evaluate the hypothesis

using linear regression. This will definitely require you to do some graphical as well as numerical presentation. Make sure that you present all relevant information and provide conclusions about whether the hypotheses were ultimately borne out by the data or not.

4 Data Paper 4

Add another continuous variable and a categorical (> 2 categories) to the data you collected for Data Paper 3. Estimate a linear regression where all variables (three continuous and one binary) enter the regression equation additively (i.e., do not include interactions). Assess the model for problems of non-linearity, non-normality of the residuals, heteroskedasticity and outliers. If you find problems, fix them in the best way you know how. Tell me what you did along the way and present/interpret the final model (i.e., the one with everything fixed).