

Welcome

Paul E. Johnson¹ ²

¹Department of Political Science

²Center for Research Methods and Data Analysis, University of Kansas

2015

PSYC650/790 POLS 706

What is this Course?

- Regression: “standard” first graduate course in stat “modeling”
 - What’s regression: prediction of a dependent variable from a formula you create!
- Prerequisites: statistics and college algebra
- Calculus not “required”, but will help in understanding
- Determination, computer savvy required

Online material in 2 Locations

- KU Blackboard
- My Dreamhost account:
 - Course Homepage: <http://pj.freefaculty.org/stat>
 - Lectures, etc <http://pj.freefaculty.org/guides>

Partially Flipped

- On Wednesdays, I'll generally ask you to watch a video online, then we'll discuss that for 30 minutes or so, after which a lab session will occur.
- Will schedule 1 or 2 other lab sessions during the week to address same/similar material.

Exercises, Homeworks and Tests

- We will have graded homework assignments (probably 5 of them). You must hand them in *on paper*.
- Expect some glitches, we constantly try to keep this interesting by swapping in new data sets
- I like this idea for exams.
 - 1 Create a “take home” exam based on student-specific data sets
 - 2 Ask students (on their own) to prepare tables and graphs
 - 3 During the “in class” portion of the test, ask for
 - 1 25% of the tables and graphs
 - 2 explanations and analysis of those tables

Homework vs Tests

- Worried about the tests? Please go look at the old ones. They are available in the Blackboard website. Look under the Course Documents.
- We are trying to make the homeworks “representative” of the tests—not more difficult, but not easier. This is difficult. You should look at the old tests to see what kinds of questions I tend to ask.

Where does Regression fit into the long term plan of study?

- Most of the things you would publish are taught in courses that are 1 or 2 steps downstream from here.



Undergrad “baby stats”



You Are Here: Regression is the analytical foundation of everything that follows



Courses on regression diagnostics & applications, generalized linear models,



repeated measures(aka longitudinal), hierarchical models, time series, structural equation models

My Slogans

- Scholars work at one level “below” their level of understanding.
- You’ll understand this class when you take the next stats class.
- Work on small example problems that isolate particular elements.
- Study math and computers wherever possible.
- Write down what you learn, you are likely to forget it.

I'm just warning you...

- I expect you to work hard, all the time.

I'm just warning you...

- I expect you to work hard, all the time.
- Graduate school is supposed to be a full time job. I spent 60hrs per week on school as a student. You should too.

I'm just warning you...

- I expect you to work hard, all the time.
- Graduate school is supposed to be a full time job. I spent 60hrs per week on school as a student. You should too.
- I'm hypocritical. I am usually behind schedule, but expect you to be on time, on schedule.

Software

- For statistics, I use R.
- “rockchalk” package for R I prepared for this class

```
install.packages("rockchalk", dep = TRUE,  
  repos="http://rweb.quant.ku.edu/cran")
```

- I don't insist you use R, but you probably should try it.
- I can give you some help if you use SAS. If you use Stata, I can sit with you and read the manual.
- Everybody needs a good programmer's file editor that can interact with R (Emacs, RStudio, or Notepad++ are leading contenders).

Why Prefer R?

- Free
- Open Source
- Extensible
- Plenty of Good Documentation
 - Documents are built into R (Run “help.start()”)

Running R

- This is not an “R Programming” course. We are only teaching you enough to “get by” in the regression course.
- We have some elementary guides available.
 - Summer Stat Institute R lectures
<http://crmda.ku.edu/summer-institute>
 - Paul Johnson’s R Web resource
<http://pj.freefaculty.org/R>
 - Basic R Steps
<http://pj.freefaculty.org/guides/Rcourse/First-R>.
Slides prepared for interns.
 - KUANT Guide R writeups (more advanced)
<http://crmda.ku.edu/kuant-guides>. Look for guides 20, 21, 25.

Replication is Important.

Don't “type commands into a program” or “Use Pull Down Menus”

DO Write out a “script” of commands that can be re-run, exactly.

- This is true not just with R, but also SAS, SPSS, Stata, etc.
 - You need to know exactly what you did
- Neatness counts!
 - Leave your code in a runnable state, from top to bottom

Creature Comforts Depend on Type of Computer

- I push everybody to learn to use Emacs with R (see Vincent Goulet's website for special versions of Emacs for Windows and Macintosh)
- Many novices drawn to RStudio (<http://www.rstudio.org>)
- Trouble setting up your Windows computer?
<http://crmda.ku.edu/windows-admin-tips>

Document Preparation

- I use \LaTeX and LyX (www.lyx.org)
- You too can learn to use \LaTeX
 - KU Grad School a \LaTeX template for theses/dissertations (written by a fine professor you know)
<http://pj.freefaculty.org/guides/Computing-HOWTO/KU-thesis>
 - latex tip sheet <http://pj.freefaculty.org/latex>
- Word is frustrating, but you can learn to use it. Learn to use the equation editor and shortcut keys. And Style Sheets.
- rockchalk has R functions to make regression tables, in 2014 (for the first time) in a format that Libre Office and MS Word can import.

This Class is Large

- Assignments have to be turned in on time.
- I have to be more scripted, more “mapped out”

I upload everything to

`http://pj.freefaculty.org/guides`

- Its a mirror of my working folder.
 - stat
 - Computing-HOWTO
- The online content includes everything, all the files I use to make the lectures.
- Material is “Sweaved”: The document runs the statistical analysis runs, weaves the results

Different types of files

- Browse a folder:

<http://pj.freefaculty.org/guides/stat/Regression/Overview>.

```
Regression-Overview-Lecture.lyx (I edit that)
Regression-Overview-Lecture.pdf (Results you see)
Regression-Overview-Lecture.Rnw (The Noweb file, an
    intermediate stage document)
Regression-Overview-Lecture.R (R code extracted from lecture
    )
plots:      folder of output files from R (not just plots,
    also statistical output)
```

Other Material

- [http://pj.freefaculty.org/R: Rtips, Working Examples](http://pj.freefaculty.org/R:Rtips,WorkingExamples)

Limited Use of Blackboard

- Use Blackboard for Announcements, Assignments, Readings.
- Why? Uploading and revising material in Blackboard is too tedious, frustrating, awful, and difficult.

FAQ

- Why don't you number out your lectures 1, 2, 3, etc, in a single folder
- I tried that, but got sick of re-numbering everything all the time
- The guide folder is topical, I do have a “lecture map” on the homepage (<http://pj.freefaculty.org/stat>).

Data Analysis Proceeds in Stages

- 1 Read in the Data
- 2 Summarize it
- 3 Estimate with it
- 4 Prepare results

Step 1. Read in the data

- Create a directory structure where the project will live
- Copy the data file into that directory
 - Through whatever means, copy this file
`http://pj.freefaculty.org/guides/stat/DataSets/AcademicSalary/academicsalary.txt`.
- Create an R file and write some commands that bring in the data.

```
dat <- read.table("academicsalary.txt", header  
  = TRUE, sep = "\t")
```


Get descriptives

```
## Could use R's summary function, but  
## rockchalk summarize is my favorite  
library(rockchalk)  
summarize(dat)
```

summarize output

```
$ numerics
      DEPART      PUB      SALARY      SEX      TIME
0%      1.0000      1.0000      30832.84      0.0000      1.0000
25%      1.0000     10.0000      51166.33      0.0000      4.0000
50%      2.0000     16.0000      63076.18      0.5000      7.0000
75%      3.0000     20.0000      76661.51      1.0000     11.7500
100%     3.0000     39.0000     108452.61      1.0000     25.0000
mean     1.9067     15.4867      64115.17      0.5000      8.0867
sd        0.8382      7.5064      17110.15      0.5017      5.2379
var        0.7026     56.3455  292757150.43      0.2517     27.4354
NA 's      0.0000      0.0000          0.00      0.0000      0.0000
N       150.0000  150.0000      150.00  150.0000  150.0000

$ factors
NULL
```

Then we look for relationships

- Regression analysis is THE bedrock on which many advanced research methods rest.
- I have written an Overview of regression in a separate set of slides, so lets go look at them