Regression Course Check List

Paul E. Johnson[1] [2]

[1]Department of Political Science

[2]Dept of Psychology

2020

A few students have asked

- What preparation is required? What courses do you assume I've taken?
- How horribly difficult is this?
- What if I forgot everything from college?

We'll address these in separate bits, involving

1. Math
2. Statistics
3. Computing

You did take a stats class, right?

- Presumably, you've had a class that involved
  - data collection and organization (aka "research design")
  - elementary statistics.
- KU offers many preparatory courses, such as COMS 356, PSYC 210 & 500, Math 356, POLS 306.
- Consider sitting in on one of those.
- Or do some background reading to catch up (next slide).

Got baby stat book?

You can get these for cheap (old editions are adequate).

1. Jack Levin and James Alan Fox. (1994) *Elementary Statistics in Social Research* (New York: Harper Collins). Mark Joslyn brought this to my attention. Looks clear and concise.

2. Frederick J Gravetter and Larry B. Wallnau, (2009) *Statistics for the Behavioral Sciences, 8th Ed* (Belmont, CA: Wadsworth Cengage). My daughter used this in PSYC200, it looks OK to me.

3. Howell, D.C. (2007) *Fundamental Statistics for the Behavioral Sciences* (6ed). Suggested by Wei Wu as background readings.

4. Schumacker, Randall E. (2014) *Learning Statistics with R* (Sage). I really like this one!

What about math background?

- College algebra is assumed.
- If you took calculus, great! We come across calculus notation from time to time
- If you never took calculus, some parts you have to "take on faith." There is a calculus overview in my *Stuff Worth Knowing* (in the blackboard)
- If you took matrix algebra, hooray, you can teach the class for me :) Seriously, I include introductory matrix notation in the lecture material.

Any particular math?

Glad you asked!

- Almost everybody recognises the importance of these mathematical symbols
    - $+$
    - $\times$
    - $\sum_{i=1}^{N}$
- We use Greek letters *A Lot!*
    - β, α, and σ^2.
- Subscripted variables x_i and y_i.

Vital: plot lines and curves

- On a piece of graph paper, plot $y = 3 + 0.4x$.
- Remember the appearance of $y = 3 + 0.4x + 2x^2$? What if $+2$ changes to -2?
- Remember the appearance of $y = log(x)$? Remember how to calculate $log(x1 + x2)$.
- If you have trouble with that, there's a review of that in my manuscript *Stuff Worth Knowing*. Look for a chapter with a title like "Plotting Curves" or such.

What to Review in Stats?

- Forgetting details is not fatal, because
  - We will re-teach you the details. Or assign readings.
- But you do need some "Statistical Maturity", the frame of mind.
  - tolerate jargon and symbols!

Basics #1. Data Types

- Unordered Categorical (aka "nominal" or "factor" in R). Mathematical transformations not meaningful (squaring or logging) not meaningful.
- Ordered Categorical (aka "ordinal" or "orded factor" in R)
- Numeric (aka interval level). Makes math transformations meaningful.

Basics #2. Data is in Columns

- A "data set" is a set of columns with the same number of observations in each one.
- Numbers or labels (representing factor levels) allowed
- Procedures may transform these into different numbers

id	age	weight	height	rel
001	21	100	5.5	cath
002	14	120	5.1	prot
003	23	199	6.1	cath
004	44	201	5.9	musl

Basics #3. Summarizing Observations

- I watched my kids take 9th grade math
  - They calculated means, medians, modes.
  - They created bar plots, scatter plots.
- mean, median, and mode. You better know what they mean.
- If you have never calculated the ___ of a small set of observations, now is a GOOD time.
 - mean
  - variance

Basics #4. Hypothesis testing jargon

- We hope this terminology is at least faintly recognizable to you

null hypothesis	alternative hypothesis	p value
confidence interval	type I error	type II error
parameter	parameter estimate	critical value
t-test	standard error	

- We will teach these things, but it is tough for people who remember nothing from undergrad stats.

Basics #4. Hopefully this is not completely foreign to you
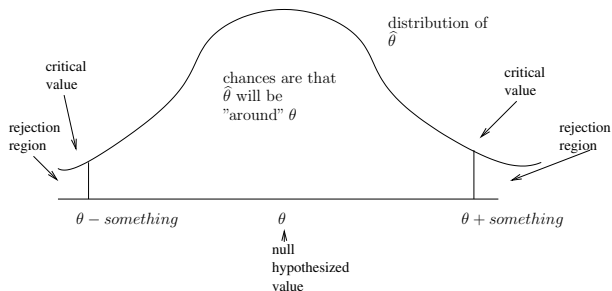
We will review this, but if you have never seen anything like it,
look up t-test in a baby stat book.

Computing

- I'd like you to know more about computing, not just "your computer."
- Necessary to tolerate transition from one operating system to another
  - We run Linux on the high performance compute cluster
- "Friends don't let friends use Excel" ("Microsoft has lots of top researchers so it's hard for me to understand how Excel can remain so crappy." (Gelman, 2013) http://andrewgelman.com/2013/04/17/excel-bashing)

Use R for Stats?

- R is
    - free
    - open source
    - available for every Operating System
    - extensible
    - used by almost all of the leading statisticians and methods developers on Earth

Install R and an Editor

- Before the semester starts, you need to have a working copy of R in your computer
  - In the CRMDA, one of my last projects was creating clear instructions on installing R in both Windows and Maintosh.
  - Those materials are available:
    - Windows_R_setup (See PDF slides and an OGV video file that can play in your browser)
    - Mac_R_setup(See PDF slides and MOV and MP4 video file (same presentation, just compressed differently) that can play in your browser). We also toyed with a tool to convert slides into an "article". You see that there as well.

Getting started with R

Step 1. Get an orientation.
Slideshows on the bare beginning R user experience and some very important fundamental points that are absolutely vital to becoming an effective researcher. Please check this folder:

```
http://pj.freefaculty.org/guides/Rcourse/First-R
```

In there, you will find a sequence of separate presentations, First-R-01, First-R-02, First-R-03. There may be more by the time you get there, maybe. In the top folder, you should see a file First-R-roadmap.txt where I've tried to explain what I'm up to.

Getting started with R

Step 2. Overview my web offerings
My R Webpage, `http://pj.freefaculty.org/R`,
offers tons of information

1. Rtips
   as well as a large collection of R programs that you should try
2. Working Examples:
   `http://pj.freefaculty.org/R/WorkingExamples`