36-617: Applied Linear Models

Introduction Brian Junker 132E Baker Hall brian@stat.cmu.edu

Classes Will Be Recorded

- I plan to record and share these classes with Zoom with you for later viewing/review.
 - They will be available at cmu.canvas.edu
 - I'm sure the first few lectures will be pretty choppy, since I am just getting used to the tech in this classroom!
- Watching recorded lectures does not replace being in class.
 - There is a "participation" component in your grade
 - If at any time you cannot be physically in class (travel problems, quarantine, etc.)
 I can give you a zoom link to join class remotely.

Outline

- Introduction
- Syllabus Stuff
- Pre-Class Survey
- Valid vs Useful
- R!
- Example
- HW00 ungraded review/intro to R
- Please take a look at Ch's 1 and 2 of Sheather for Wednesday!

Introduction – About Us

<u>Instructor</u>
 <u>Brian Junker</u>
 <u>brian@stat.cmu.edu</u>

<u>TA</u> <u>Lorenzo Tomaselli</u> <u>ltomasel@stat.cmu.edu</u>

- Office hours:
 - MW 11am-Noon
 - 132E Baker Hall
 - ...or by appt(in person or Zoom)
- Office hours:

 - □ ...or by appt

Introduction – About The Course

- <u>Technical material</u>: The machinery of linear regression and its generalizations
- Disposition: When is a model adequate?
- <u>Translation:</u> ABA⁻¹:
 - □ A: Translate from real world to quantitative question
 - B: Answer quantitative question using Statistics
 - A⁻¹: Translate back to real word
 - Lather, rinse, repeat...
- Communication:
 - IMRaD -> IDMRaD
 - Clear sentences, paragraphs, sections.

Introduction – Course Materials

<u>Technical material:</u>

- Sheather (2009). A Modern Approach to Regression with R. NY: Springer
 - Buy online, or get free pdf from link.springer.com (via CMU vpn)

Supplementary texts:

 Gelman, A. & Hill, J. (2007). Data Analysis Using Regression and Multilevel/Hierarchical Models. NY:Cambridge Univ Press.

Software:

- R, RStudio
- LaTeX (I like MikTeX; or just use overleaf.com online—free to CMU)
- □ All free software; all very flexible and general
- <u>Help:</u> I will post some links, but get used to googling!

Other course tools:

- Canvas! (canvas.cmu.edu)
 - I will make course materials available here; you will submit hw's and peer reviews here
- Piazza (available through Canvas)
 - Great asynchronous supplement to office hours; the TA and I will also monitor Piazza.

Introduction – Online Resources

- Online resources:
 - □ I hope to record lectures and make them available on Canvas.
 - Canvas (canvas.cmu.edu)
 - All course materials
 - Your grades
 - Monday quizzes
 - Submit and peer review project papers
 - **Gradescope** (within Canvas)
 - Submit weekly homework
 - Piazza (within Canvas)
 - Great for asking (and answering) questions outside of class
 - The TA and I will also monitor Piazza
- General schedule for the semester

General schedule for the semester

Week	Dates	Tentative Topics	Tentative Sources
Week 1	Aug 30, Sep 1	Intro, Appl Statistics, Regression Ba-	Ch* 1, 2
		sics	
Week 2	Sep 6 (no class ¹), Sep 8	Writing	Ch 1, handouts
Week 3	Sep 13, 15	Diagnostics & Transformations I	Ch 3
Week 4	Sep 20, 22	Multiple Regression	Ch 5
		Project 1 assigned	
Week 5	Sep 27, 29	Diagnostics II & Variable Selection	Ch 6
Week 6	Oct 4, 6	Variable Selection	Ch 7, handouts
Week 7	Oct 11, 13	Logistic Regression	Ch 8, handouts
		Project 1 due	
Week 8	Oct 18, 20	Generalized Linear Models (GLM's)	Handouts; stuff from G&H3
Week 9	Oct 25, 27	Causal Reasoning	Handouts; stuff from G&H
Week 10	Nov 1, 3	Generalized Least Squares	Ch 9, (Ch 4?)
		Project 2 assigned	
Week 11	Nov 8, 10	Multilevel and Mixed Effects Models	Ch 10; stuff from G&H
Week 12	Nov 15, 17	Residuals, Estimation and Model Se-	Handouts; stuff from G&H
		lection	
Week 13	Nov 22 (no class ² on 24 th)	Shrinkage, Examples	Lecture notes; handouts
		Project 2 due	
Week 14	Nov 29, Dec 1	Multilevel GLM's, & maybe	Lecture notes; handouts
		Bayesian Approaches	

*All Chapters from Sheather unless otherwise noted.

¹Labor Day (US Holiday).

²Thanksgiving (US Holiday).

³G&H: recommended text by Gelman & Hill (2009).

Syllabus Stuff – Work & Rules

- 20%: 10-ish HW's
 - Please feel free to work with each other on hw;
 <u>BUT you must list who you worked with.</u>
- 10%: Monday Quizzes (on weekly reading/materials)
- 50%: 2 Projects
 - You do these alone no collaboration
- 10% Peer review of projects
- 10% Participation (Do I remember your name? What you did in class? In office hours? On Piazza?)
- Credit where credit is due
 - Please list any <u>person</u> or any <u>source</u> you consulted in doing your work, in a list of references at the end of your writeup
- All hw will be submitted via Canvas (Gradescope)
 Generally we will not accept late hw or late take-homes...

Introduction – Level

- Hopefully you have seen calculus-based prob & stat, matrix algebra, and a little linear regression.
 We need to talk like statisticians!
- You have all different levels of experience with
 Applied regression and statistical modeling
 R
 - Writing scientific reports
- Fill in the gaps
 - Learn on your own (Google)! Help each other!
 - Ask Lorenzo and me!

Pre-Class Survey

Let's take a break and think about these two quotes...

- "[I]t makes sense to base inferences or conclusions only on valid models"
 S.J. Sheather (2007)
- "All models are wrong but some are useful"
 G.E.P. Box (1978)

R!

 Some people learning R for the first time; others have done extensive data analysis projects in R.

Some references on R:

- http://www.cookbook-r.com
- Quick-R: <u>https://www.statmethods.net/</u>
- Online course: <u>https://www.datacamp.com/courses/free-introduction-to-r</u>
- Lately I really like <u>https://kieranhealy.org/publications/dataviz/</u>
- If you have not used R before...
 - http://www.cs.cmu.edu/~10702/R2/Rintro.pdf provides a good start
 - Let's look at the first part of it now...

A first look at R...

- http://www.cs.cmu.edu/~10702/R2/Rintro.pdf is a good place to start...
 - If you are new to R, type into R all the commands and examples in rintro.pdf
 - If you have worked with R before, read through rintro.pdf and try to predict what would happen with each command.
 If you are not sure, type in that command/example.

Example

- Goldstein et al (1993) try to rank schools in London to distinguish the "best" from the "worst" of them.
- 01-london-schools.r

HW00

- Already online on canvas.cmu.edu
- Mainly review/crash course on data frames and graphing in R.
- Please go to Canvas and download:
 hw00.pdf
 - □and the accompanying files...

I will publish solutions later this week.

Summary

- Introduction
- Syllabus Stuff
- Pre-Class Survey
- Valid vs Useful
- R!
- Example
- HW00 ungraded review/intro to R
- Please take a look at Ch's 1 and 2 of Sheather for Wednesday!