Linear Regression Models
P8111

Lecture 26

Jeff Goldsmith
April 28, 2016

THE DEPARTMENT OF BIOSTATISTICS

Columbia University MAILMAN SCHOOL OF PUBLIC HEALTH

Today's Lecture

- Last class! 
  - How to write statistics 
  - Tips on the final project

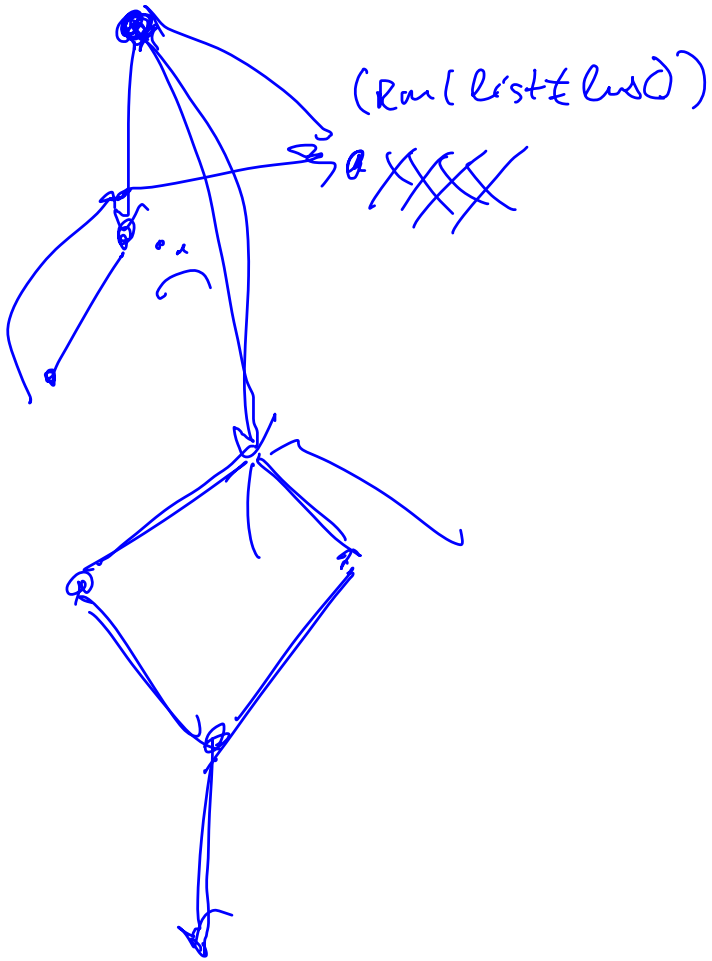
Writing a statistical report

General tips:

- Know your audience
  - Are they statistically knowledgeable?
  - How many details do they want / need?
- Say exactly what you did
  - Don't leave any thing important out
  - Not the same as a step-by-step list of what you typed into R – that's what the RMD is for

Analysis vs Report structure

What analysis looks like



What a report looks like

Report structure

- Introduction
- Methods
  - EDA
  - Formal analysis
- Results
- Discussion

Report structure: Introduction

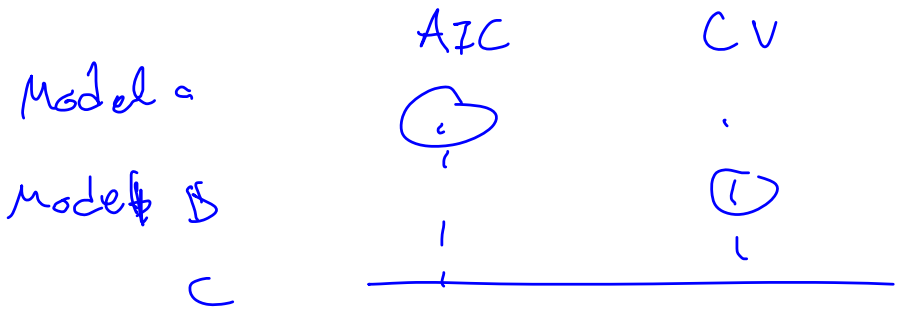
- What is the context for this problem?
- What kind of data were gathered?
- What do you hope to learn?

Report structure: Methods

- Exploratory analyses
  - Preprocessing and cleaning (creation of variables; identifying missing values; coding)
  - Exploratory plots
  - Exploratory analyses
- Formal analyses
  - Model components
  - Model strategy
  - Decision process (what tests / comparisons; thresholds for significance)

Report structure: Results

- What did you find in exploratory analyses (any missing values? data distributions? notable features?)
- What happened in your modeling?
- What is your final model, and what are the important quantities?



Report structure: Discussion

- What do your results say about the question you hoped to answer?
- What were the limitations of your data or your analysis?
- Are there broader implications to note?

Final project info

- No TA office hours – email me
- I do the grading

Final project

Final project tips

- Think before you do anything
- Think while you're doing stuff
- Think after you've done things

Final project tips

- There is no right answer – I'm interested in your thought process and justification
- Use the above structure to write your report
- Follow the instructions (e.g. stick to the page limit)

Course's big ideas

- Regression is neat
- Statistics is neat

$(X^T X)^{-1} X^T y$