BTRY 6020: Principles of Statistics II

Giles Hooker

Spring 2016
MW 08:40 - 9:55
Caldwell 100

Instructor

- Professor: Giles Hooker, BSCB
- Office: 1186 Comstock Hall
- Email: giles.hooker@cornell.edu
- Phone: 255-1638
- Office Hours: Wednesday 10:00 - 12:00
- Webpage:

  www.bscb.cornell.edu/∼hooker/

- Class notes and discussion boards on

  blackboard.cornell.edu

TA and Labs

David Sinclair (dgs242)

Office Hours in Comstock 1181

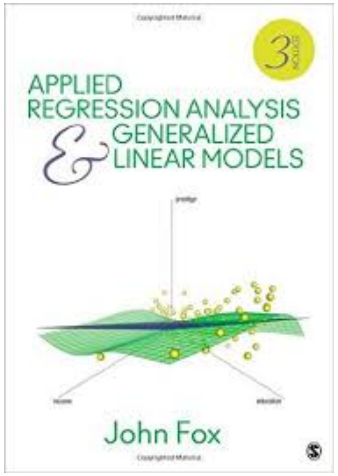
- Thursdays 13:00 - 14:00
- Thursdays 16:00 - 17:00

Labs will be held (mostly) the week after homework is handed out in Man B30B

- Monday, 14:55 - 16:10
- Tuesday 13:25 - 14:40

Texts

Fox, 2016, "Applied Regression Analysis and Generalized Linear Models"



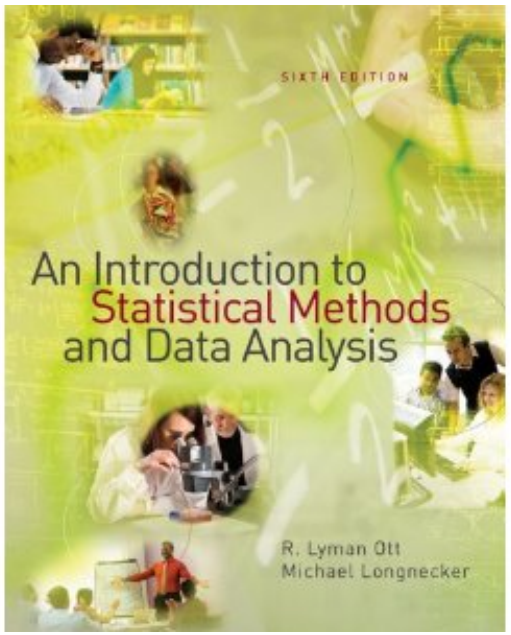
Other freely available references:

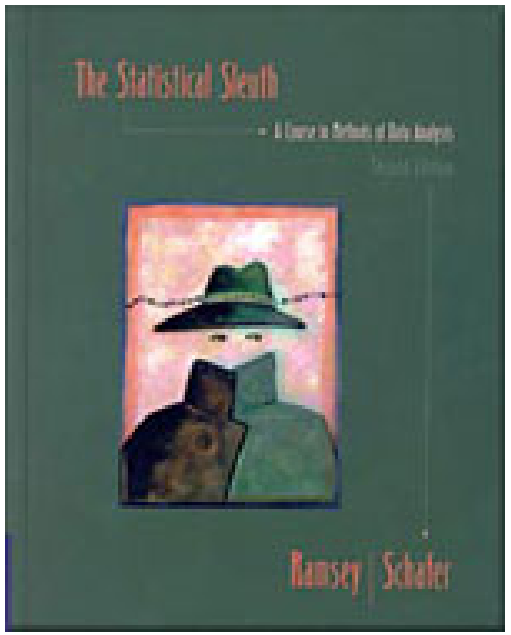
- Diez ,Barr and Cetinkaya-Rundel, 2012, "OpenIntro Statistics" https://www.openintro.org/stat/textbook.php?stat_book=os
- Agresti,2007, "An Introduction to Categorical Data Analysis" http://newcatalog.library.cornell.edu/catalog/6223313
- Bates (???) " lme4: Mixed Effects Modeling with R" Available on blackboard

Old Texts

On short loan in Mann library

- Ott and Longnecker (2001)
  "Statistical Methods and Data Analysis", 6th Edition
  ISBN: 0-534-25122-6
- Ramsey and Schafer (2005)
  "The Statistical Sleuth" 2nd Edition
  ISBN: 0-534-38670-9





Software

- BTRY 6020 will be taught using R.
- R is available in Mann library computer labs and may be downloaded from

  www.r-project.org

  Cost: free!
- R is different from menu-driven statistics packages (JMP, STATA, Minitab...).
- Students in BTRY 6010 in 2014 have already used R; a crash course will be given in labs in week 1.
- R scripts to reproduce most analyses in lectures will be supplied.
- Homework will be expected to be submitted as in 6010.

R cont...

- The easiest way to use `R` is with RStudio

  www.rstudio.com

- In addition to running `R`, the `knitr` package lets you turn `R` scripts (with additional commentary and output results) into .pdf or .doc or .html documents.
- In general, this is a really good way to document your analysis and make sure that everything you do is reproducible.
- To run `knitr` you also need LaTeXavailable from

  www.miktex.org

- If you were not in 6010, please install `R`, RSTudio and LaTeXbefore the first lab.
- Course Info section of blackboard has some documentation for getting started with R as well as a list of commonly used R functions.

Homework and Grading

- Grades will be based on (approximately) bi-weekly homework assignments, two prelims and a final exam.
- Homework assignments will be posted on the course website on Mondays and will be due after class on Friday the following week.
- Homework must be uploaded to blackboard as a .pdf file. (from 6010: if you prefer to knit into .doc and then save as .pdf, this is fine and can save LaTeXerrors).
- *Students may discuss homework problems with one another, but must write up their solutions on their own. Do not share your homework file with other students.*
- Late Work/Regrading – see class syllabus.
- POINTS: Prelims 25% each. Final 25%. Homework 25%.

Exam Dates

- Evening Prelims March 3 and April 14.
- EXAM: TBD

Weather cancelation policy: if Cornell closes or TCAT suspends services.

Curving and letter grades:

- Individual items will not be curved.
- Letter grades will be assigned based on distribution of scores among students.
- Formula not pre-set; aim is for steps of about 5%, median B+/A-; credence given to gaps between students.

Communication

- All course announcements and materials will be posted on

  blackboard.cornell.edu

- Slides are pedagogical tools, not substitutes for notes.
- Discussion boards are also available for
  - general questions
  - each homework assignment

  We will check them regularly. Please use them!
- Questions can be posted anonymously; we will also post answers to questions that are e-mailed to us or asked in office hours if we think they will be useful to others.
- Communication goes two ways. Please provide feedback.

Syllabus

- Review of simple linear regression (Chapters 5.1, 12.1)
- Multiple linear regression (Chapters 5 - 12)
- Generalized linear models (Chapters 14 and 15)
- Basic design concepts
- ANOVA for standard designs (Chapter 8.2, 8.3)
- Random and mixed effects models (Chapter 23)
- Repeated measures and Cross-Over Designs (Chapter 23)
- Additional material (notes provided in class)

Please remind me to indicate which chapters material is coming from!

Assumed Math

- BTRY 6020 not intended to be mathematically intensive.
- Some mathematical manipulation is unavoidable.
- In particular, you will need to work with
  - Algebraic manipulation:

$$(x + y)z = zx + zy$$
$$ax + b = c \Rightarrow x = (c - b)/a$$

  - Summations:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

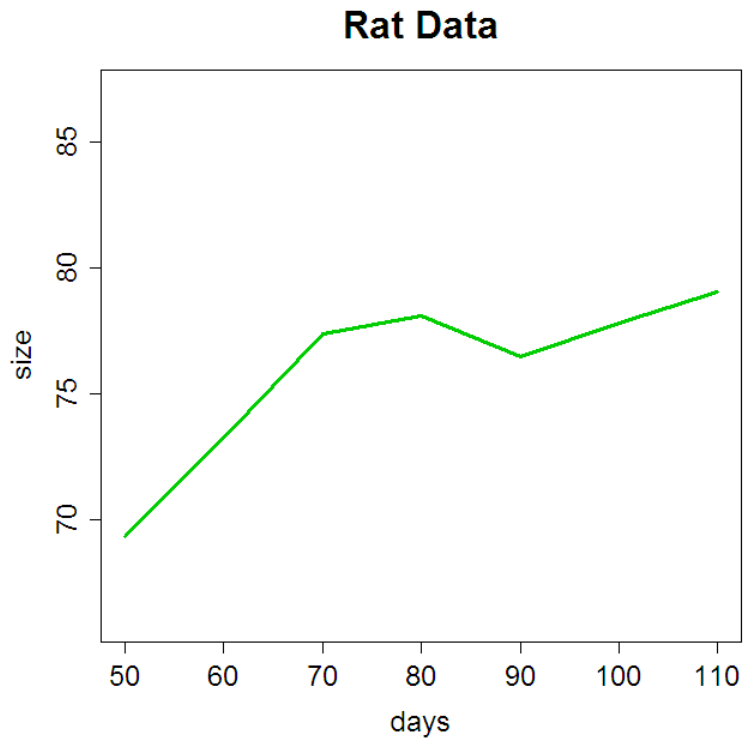
  - Powers and square roots.
  - Logs and exponents:

$$\log(e^x) = x, \ e^{x+y} = e^x e^y, \ \log(xy) = \log(x) + \log(y)$$

    (note that log in this class and in R means natural logarithm in all cases).
- Some matrix algebra will be introduced, but you will not need to use it.

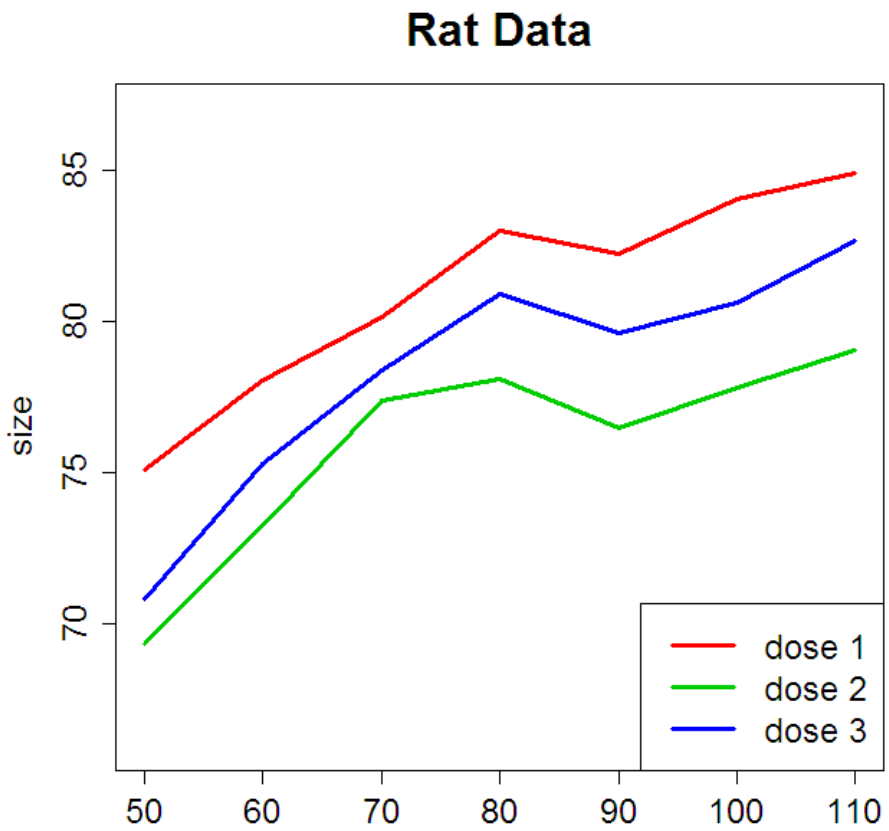
Rats on Hormones

- Interest in rats response to growth hormone
- Width of head measured at 7 time points



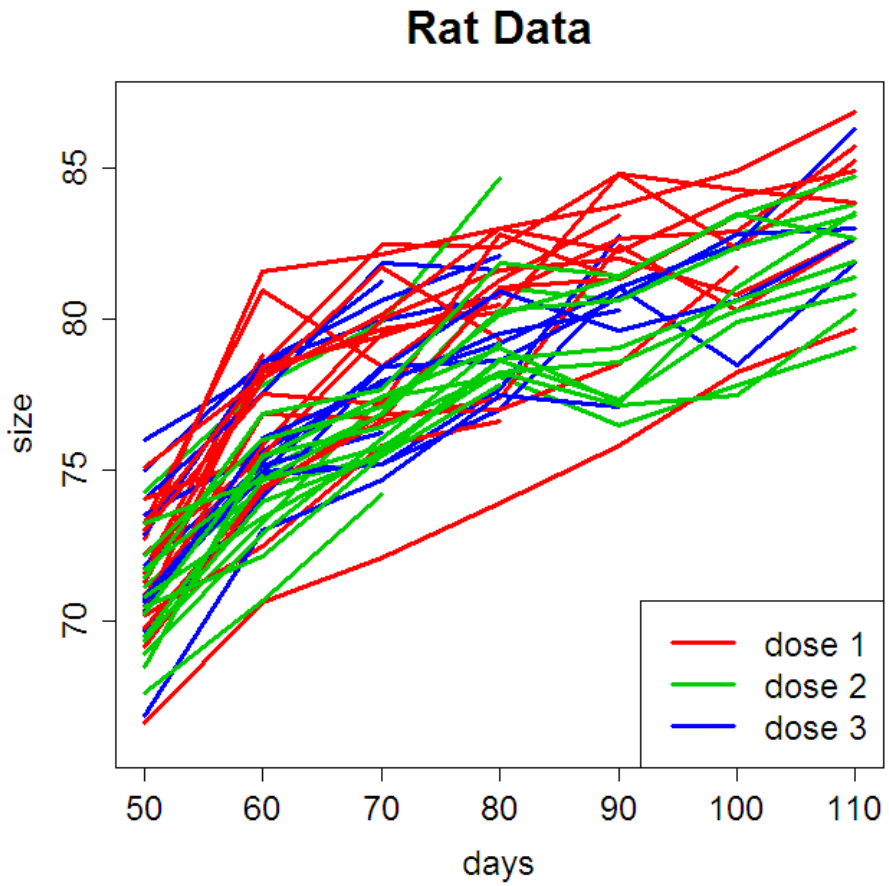
Rats on Hormones

- Need some baselines in order to obtain a more complete picture.
- 3 different doses tried, on 3 different rats.
- But how do we know the difference is because of dose, rather than something about these particular rats?



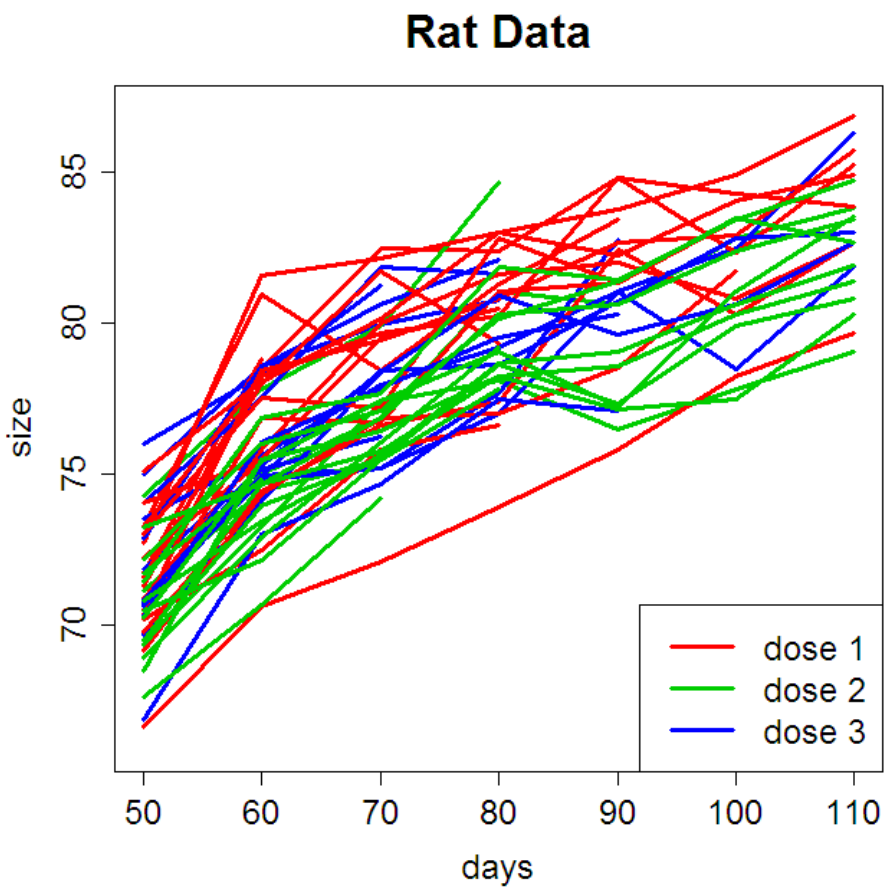
Rats on Hormones

- 64 rats, divided into 3 dose levels.
- Interested in *average* slope of rats within a dose level.



Rats on Hormones

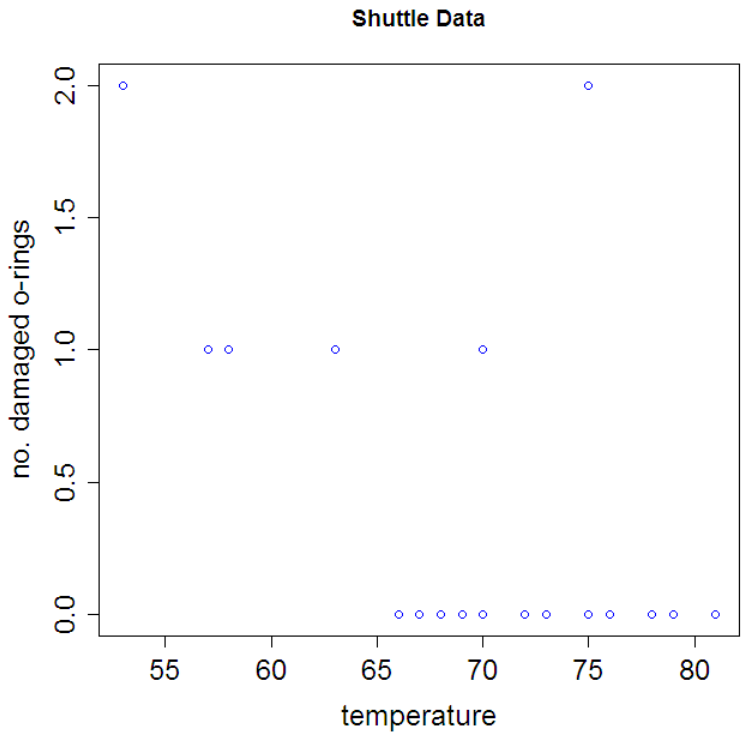
Statistics: "What would happen if you repeated the experiment?"



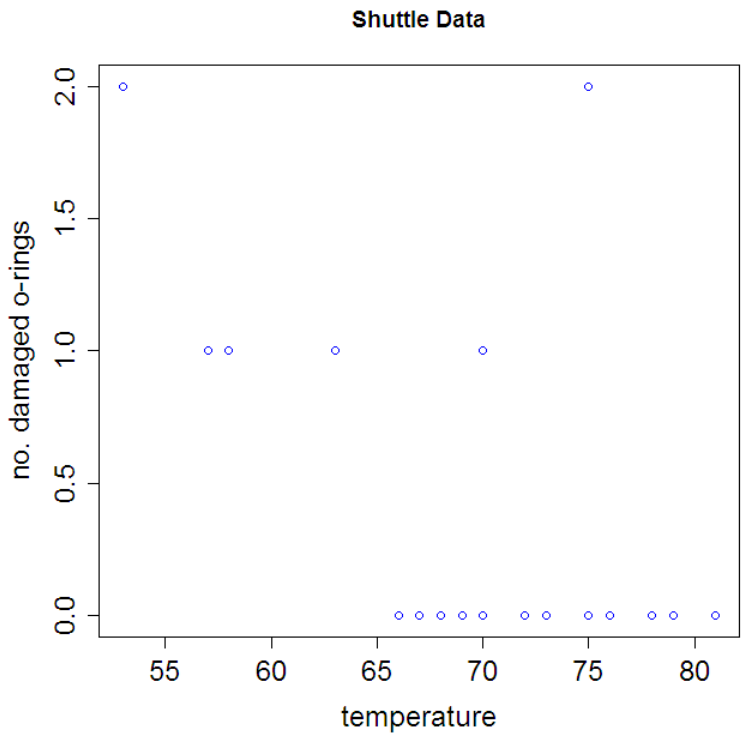
- Linear regression: the slope would still be the same, just deviations are different.
- This experiment = new rats $\Rightarrow$ new slopes!
- But the *average* slope for rats in a level should be the same.
- Need to account for variation due to new rats as well as new measurements.

Space Shuttle O-rings

- In 1986, Space Shuttle Challenger blew up on take-off.
- Cause of explosion was fuel leaking through an O-ring seal.
- Concern had been expressed about the effect of low temperatures (about 34F) on O-rings, but was ignored.



Space Shuttle Data



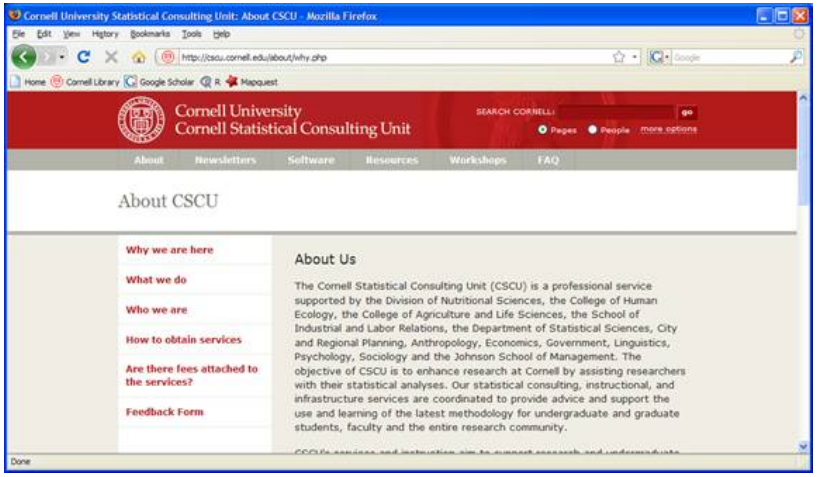
- How do we determine if a relationship exists between number of damaged O-rings and take-off temperature?
- Need to account for type of data: *counts*, cannot be negative, must be whole numbers.
- How do we assess confidence in prediction at 34 degrees?

Your Own Data

If you have your own data

1. Great!
2. BTRY 6020 is not a consulting service.
3. Free statistical consulting is available from the Cornell Statistical Consulting Unit:

cscu.cornell.edu



A First Assignment

- **This Week** A short survey (link on the blackboard website); it is intended to generate data for in-class examples. Please log on and fill it out this week.

  `https://cornell.qualtrics.com/SE/?SID=SV_e2Gz3K6rKBIgoi9`

- **Homework 1 Question 1**: make up a data set relevant to you. Each homework you will be asked how an analysis could be applied to it.
  - Don't make up numbers, just what variables are measured and how many experimental units are there?
  - Measured variables should include include one of each
    - Continuous variables (could take any value)
    - Binary variables (yes/no, true/false, 1/0)
    - Counts
  - Also include some variables you have controlled (which treatment applied, selected experimental units of a particular type etc)

Course Goals and Assessment Philosophy

Three main subject areas:

- Statistical models
- Experimental design
- Accounting for sources of random variation

Assessment, you should be able to:

- Carry out analysis in R, report correct quantities, calculate confidence intervals etc.
- Give real-world meaning to the results of analysis. Discuss/check model assumptions and the implications if they are wrong.
- Decide on what model/analysis to use to answer a real-world question. Eg:
  Hypothesis: the gender-gap in pay expectations of students in engineering is smaller than for students in ILR.

Next Time

- Covariance and Correlation
- Simple Linear Regression - Models and Mechanics.