STAT 311 WINTER 2015

- Instructor: Professor Elizabeth Thompson
- TA's: Nilanjana Laha: Sections AA,AB

    Yicheng Li: Sections AC,AD

    Grainne Ne Shi McKnight: Sections AE, AF

- The same team as Fall 2014
- The web page:

    https://canvas.uw.edu/courses/950686/

See the Canvas web page for

-- lecture notes/schedule, office hours, study center

-- homework info, lab info (do you have a laptop?)

-- book info, exams info, grading info, and more

- Add codes: sign up for add code at end of class. – Please give your UW NetId – NOT student number.

The text book and Aplia

- The text book is
    Utts and Heckard: Mind on Statistics, 5th edition
- We will cover Chapters 1-16: 1 or 2 per week.
- The course moves fast: and we have only 28 lecture sessions as compared to 31 in Fall.
- Two options:
(1) Bookstore – hardcopy text+ Aplia web access (more expensive; and make sure the copy is good)
(2) With Aplia access electronic copy: (less expensive and quite convenient– if hate reading online can download and print chapters.)
- Homeworks will be through Aplia online system – see Canvas page for details on how to register.
- Please use your UW name on Aplia

Acknowledgements

- I did teach this class in Fall but –

- Professors Eli Gurarie and Martina Morris have regularly taught STAT 311 in recent quarters, and developed many lab and lecture materials.

- Professor Martina Morris has used the Aplia system for STAT 311:

  - The Aplia assignments evolve, but remain very close to those of Martina from Spring 2014.

  - Many figures etc. are taken from Martina's slides.

- Other figures are taken from the slides and other instructor materials provided by the Aplia provider:

  Cengage Learning: www.cengage.com

3

Two schedules: follow both!

1. Canvas schedule of Lecture and quiz sections:

- Module 1 on Canvas contains much info, including
  - PDF schedule of lecture and lab topics
  - Will be updated as often as possible.
  - Also, note Midterm date Friday Feb 6
- Tied to the Canvas course modules which are organized by week (Mon to Sun as per Aplia)
  - Always look ahead to the next week– especially for labs
- Full Lecture note slides will be uploaded AFTER each lecture-
  - An outline version will be available before class.
  - Remind me if the final version does not appear.

Two schedules – Aplia Schedule

2. Aplia homework and lab assignments schedule:
- Two types of Aplia homework assignments:
    - Practice: will show answers, can do any number of tries.
    - Graded: generally basis for homework grade,
      but may be "excluded" from course score
      (e.g. Math practice assignment due Friday 11:00 p.m.)
- Graded Aplia problems:
    - No "grade-it-now" multiple try problems in scored graded hwks.
    - "Grade at deadline": 1 try only --but can be edited up to the deadline – what ever is there at deadline is graded.
    - Most (maybe all) homeworks are same for all students.
- Labs will also be submitted via Aplia. (all 1 try only)
- Until Midterm: Homeworks due Monday night, and labs Tuesday night -- look ahead every week!
    - Homeworks/labs available at latest 7 days ahead. (ASK if not)
- Due time: 11:00 p.m. Pacific time – if Aplia apparently tells you otherwise CHECK your Aplia time zone, and/or ASK.

View forward for the week

- Today, Monday: Overview and an example.
  - Check out the canvas site; register for Aplia;
  - get add codes and register for class if you can.
- Tomorrow, Tuesday sections:
  - Go to Quiz section – even if not yet registered
  - Get help with Aplia and R-studio– take your laptop!!
- Wednesday – Graphical summaries of data (U/H 2.1-4)
- Thursday-- Aplia and Rstudio practice and more practice.
- Friday: "Graded" (but "excluded")) Math prep assignment due
- Friday -- Numerical data summaries (U/H 2.5-7) (Hwk 1)
- Monday 11:00 p.m.: first actually graded homework is due
- Tuesday -- quiz section -- more R towards lab 1
- Tuesday   11:00 p.m.:   Lab 1 is due.

Getting help: it is a heavy class

- Your instructor and TAs are there for you
  - – but you also have to do your part.
- Check the web sites, come to class, go to quiz sections
  - Use the class discussion board
- Office hours
  - TA Office hours will be posted asap
  - Mine are posted, but may need to change some weeks. Tomorrow, Tues, 1-2pm in C-317 PDLFD
- Statistics Tutoring & Study Center (STSC)– FREE HELP
  - Basement of Communications (B-023, CMU)
  - Your TA's and I will do (some) office hours there
  - Get help from your peers– study together.
  - Get help from tutors – 10am-5pm, 7-9pm, Mon-Thurs (?)

What is Statistics?

- Quantitative facts, numerical descriptions (data)
- Set of tools for the collection and analysis of data
- --- to assess evidence in the face of variation and uncertainty. (NOT true/false Math proofs.)
- --- so we can make decisions or draw conclusions.

- People (even here in JHN 102) vary in
  - height, gender, blood type….. Easy to measure
  - Income/job, math/music/athletic capabilities…OK
  - Opinions, beliefs, approach to life, …. Hard

Statistics uses terms like "error", "deviation",….

VARIATION is not ERROR

Where are statistical data?

- News reports; Crime statistics, Traffic statistics
- Weather reports; Record highs and lows; precipitation.
- School records, grades, course evaluations
- Consumer reports, Election polls
- Environmental standards, air pollution,

  forest diversity, salmon catch, endangered species
- Medical and dental records, diagnostic procedures.
- Stock market, business plans, marketing surveys
- --- and many more
- Note the news EVENT is not a statistical study, – although the event record may become part of one.

What is "doing Statistics"?

The course will cover:

- Describing data (Weeks 1-3):

  types of data: discrete counts; continuous variables
  graphical and numerical summaries

- Principles of study designs for collecting data (Weeks 4-5)

- Modeling data with probabilities (Weeks 6-7):
  Random variables. Sampling distributions

- Statistical Inference   (Weeks 8-10):
  Confidence intervals and Hypothesis testing ,
  for discrete count data (how many ….?)
  and continuous data variables
         (height, income, blood pressure….?)

Some key words & concepts

- Populations and samples
  - Observational studies and experiments
  - Cases: the unit of measurement
  - Variables: what we measure on the cases
  - Distributions: the patterns of variation for the measured variables in the population/sample.
- Parameters and Statistics
  - Description: summary statistics about samples
  - Estimation: from the statistic to the parameter
  - Inference: from the sample to the population
- Explanatory and response variables
  - Causation and association

Workers losing days to 'flu

- What proportion of workers in Seattle will lose work-days to 'flu next winter?
- Question: What proportion of workers in Seattle lost work-days to 'flu last winter?
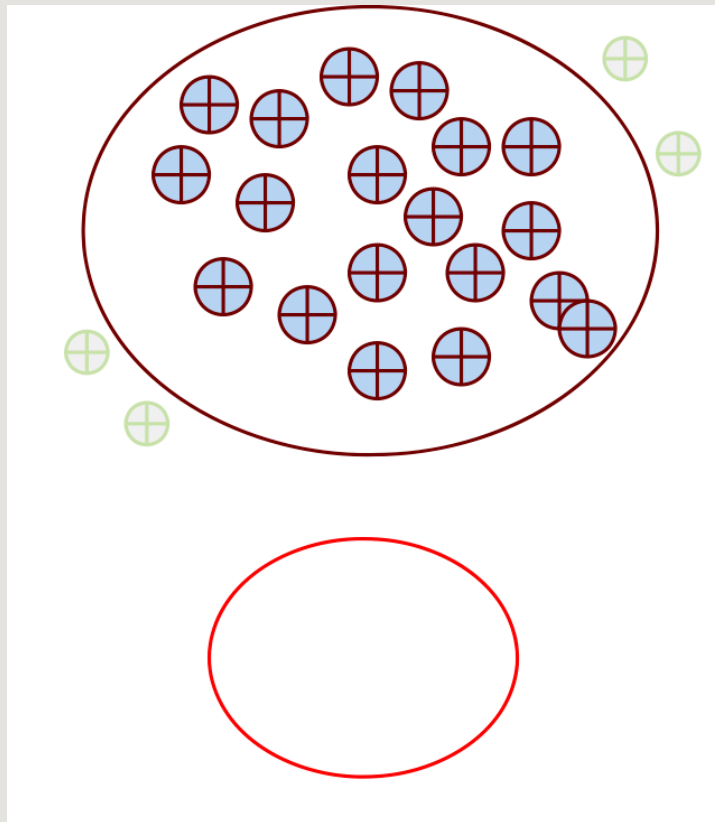
- Define the population (size N)
  ?? Seattle
  ?? Workers (full/part time?)
- Decide how to sample
  ?? representative sample
- Take a sample (size n)
  These workers are the cases or units of observation
- Measure a variable on each case: ask the question: Measurement (data) is YES/NO.
- Form a summary measurement: proportion in sample who lost days to flu last winter.

Parameters and Statistics

- Population parameter: the unknown proportion of Seattle workers who lost days to flu last winter.
- Sample statistic: the observed proportion in the sample who lost days to flu last winter.

- Estimation: assuming the sample is representative (!!), we can estimate the parameter by the statistic.

- AFTER next winter we can repeat the study.
- Statistical hypothesis:
  the population proportion is the same in the two years.
  - Test by comparing
    the sample proportions for last/next year.
- How large a sample (n)??
  This is in U/H Chapter 1, but we defer it for now.

Do flu-shots reduce workdays lost?

- An Observational Study:
- Ask the (sample) workers who lost workdays whether they had a flu shot. Useful information? Why/Why NOT?
- Ask all the (sample) workers whether they had a flu shot.
  Why is this a better design? Case-control comparison.
- Statistical question: Is the proportion of workers who lost work-days to flu the same among those who did/did not get flu shots?
- This is a question statistics can answer.
- Suppose we find a difference (in expected direction?):
  - Is it statistically significant?
  - If yes, can we conclude flu shots cause the reduction?
- Generally NO: there could be confounding factors:
  age, types of employment, economic/health-care status.

Do flu-shots reduce workdays lost?

- A Controlled Experiment:
- For next year's study: Suppose we could randomly assign workers to get a flu shot or not?
- Finding (for next year) a statistically significant flu-shot effect:
  - Could we conclude flu shots cause the reduction?
  - Effect of belief in flu shot protection, if have/not shot?
  - Can we make the study "double blind"?-- Placebo?
- If we randomize and control the experiment, then maybe can infer cause, but ….
- Also, is the flu-shot effect large enough to be of practical importance? -- there may be personal costs/ risks to flu shots.

Two kinds of Study designs

- Controlled Experiments
    - Investigator controls which subjects get which treatment
    - Intentional manipulation of subjects.
    - Apply a treatment, and measure data outcomes.
- Observational studies
    - Investigator does not control assignment of factors of interest to subjects.
    - Subjects are a sample from a population.
    - Data are measurements of variables of interest on subjects.

Explanatory and Response Variables

- Many statistical questions about

    the relationship between two variables.

    Did you get flu-shot? Did you lose days to flu?

- We (often) identify one as the explanatory variable and the other as the response variable**.**

    (In many cases, no good reason to choose which.)

- General idea**:** the value of the explanatory variable for an individual is thought to partially explain the value of the response variable for that case.

    - Having flu-shot reduces chance of losing days to flu?

- But association does not show causation

    - in an observational study there are confounding factors

(age, types of employment, economic/health-care status,…)

    - in a designed experiment, maybe if well-designed, but..