

Welcome to STAT 100B!

Instructor: Lauren Cappiello, M.S.

September 27, 2019

Welcome!

A little about me:

- Why statistics?

Course Website: lgpcappiello.github.io/teaching/stat100b/f19

- Bookmark this page!
- Website includes course calendar, assignments, links, and syllabus. There is also a link to download our open access textbook.
- Course site is linked to on iLearn.
- Grades will be posted on iLearn.

Syllabus

The syllabus is your first and only required reading.
(Reading the textbook is also highly recommended!)

OpenIntro Statistics by David Diez, Mine Cetinkaya-Rundel, and Christopher Barr. Free download at leanpub.com/openintro-statistics

(Optional) *Introduction to Probability & Statistics* by Mendenhall, Beaver, and Beaver, Brooks / Cole Cengage Learning. Any of the 12th, 13th, or 14th editions are fine.

Discussions

Discussions will consist of

- Review of material
- Student questions
- Weekly quiz

There is no discussion this week.

Labs

Labs will consist of

- Review
- R-based lab activities

There will be approximately one lab due each week. Each lab session will give you an opportunity to work on the lab with the help of your TA.

There is no lab meeting this week. Your first lab meeting will be Monday/Tuesday (Lab A) of Week 1.

Some FAQs

- Yes, I post my slides online!
- I will also post supplemental R code.
- No, I don't post anything I write on the board.
- Homeworks will be posted approximately one week before the due date.

Are there any other questions?

My Classroom

I have only one formal classroom policy: you can do whatever you want *as long as it doesn't disrupt anyone else's learning.*

A Few Requests

- In-class computer use is fine, but if you are going to be doing anything other than taking notes, please sit in the back so that nobody behind you can get distracted by your screen.
- Use professional language when emailing myself and your TAs. This is a good habit to practice whenever you send an email!
- If you have any problems - with me, the class, other students, your TA - please let me know!

Student Survey

There is a link to a brief survey on the course website. The survey is **due October 1 at midnight**.

There are no wrong answers! Everyone who fills out the survey will receive full credit. Points will go toward your class participation grade.

Why Study Statistics?

Just a few reasons (of many!)...

- ① You can't do scientific research without statistics. Having a solid understanding of statistics will be a huge benefit if you want to apply to Masters or PhD programs.
- ② Even if you don't want to get a PhD, many jobs require either some research or the ability to read technical reports.
- ③ A basic understanding of statistics will make you a more critical consumer of media. In a world of biased or "fake" news, this is a really important skill!

This term, labs will be completed in R.

- Open source statistical programming.
- Less of a "black box" than Minitab.
- Many companies and academic researchers use it.

Programming in R

Let's open up R and take a look at what we'll be working with.

The Normal Distribution

We start our review with the **normal distribution**. This is one of the most common distributions you will see in practice.



The Normal Distribution Model

- The symmetric, unimodal, bell-shaped curve of the normal distribution can vary based on:
 - Mean
 - Standard deviation
- These adjustable details are called **model parameters**.

Normal Distribution Notation

For a normal distribution with mean μ and standard deviation σ , we write

$$N(\mu, \sigma)$$

For a variable X with a normal distribution, we may write

$$X \sim N(\mu, \sigma).$$

where " \sim " denotes "is distributed".

Normal Distribution Notation

For a normal distribution with mean 19 and standard deviation 4, we write

$$N(\mu = 19, \sigma = 4)$$

- The mean and standard deviation describe a normal distribution fully and exactly.
- This is what we mean by a distribution's **parameters**.

Standard Normal Distribution

The **standard normal distribution** is a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$.

$$N(\mu = 0, \sigma = 1)$$

Standardizing with Z-Scores

We often want to put data onto a standardized scale, which can make comparisons more reasonable.

Standardizing with Z-Scores

We compute the Z-score for an observation x that follows a distribution with mean μ and standard deviation σ using

$$z = \frac{x - \mu}{\sigma}$$

Example

Let X represent a random variable from $N(\mu = 3, \sigma = 2)$

$$X \sim N(\mu = 3, \sigma = 2)$$

and suppose we observe $x = 5.19$.

- 1 Find the Z-score of x .
- 2 Use the Z-score to determine how many standard deviations above or below the mean x falls.

Example: Brushtail Possums



Head lengths of brushtail possums follow a normal distribution with mean 92.6 mm and standard deviation 3.6 mm.

Compute the Z-score for a possum with head length 85.8 mm.

Finding Tail Areas

There are several techniques for finding tail areas:

- 1 Integrate.
- 2 Use a graphing calculator.
- 3 Use a probability table.
- 4 Use a statistical software.

Finding Tail Areas

- For quizzes and exams, you will be provided with information from R .
- I will do the work in R , but you may need to use a Z -score to pick the correct tail probability from a list.

For example

Z-score	Lower Tail Area
1	0.8413
1.5	0.9332

Example: Normal Probability

Cumulative SAT scores are well-approximated by a normal model, $N(\mu = 1100, \sigma = 200)$.

Shannon is a randomly selected SAT taker, and nothing is known about her SAT aptitude. What is the probability Shannon scores at least 1190 on her SATs?

Finding Areas to the Right

- Software programs usually return the area to the left (left tail) when given a Z -score.
- To get the area to the right
 - 1 Find the area to the left.
 - 2 Subtract this area from one.

Example

Edward earned a 1030 on his SAT. What is his percentile? Recall that cumulative SAT scores are well-approximated by a normal model, $N(\mu = 1100, \sigma = 200)$.