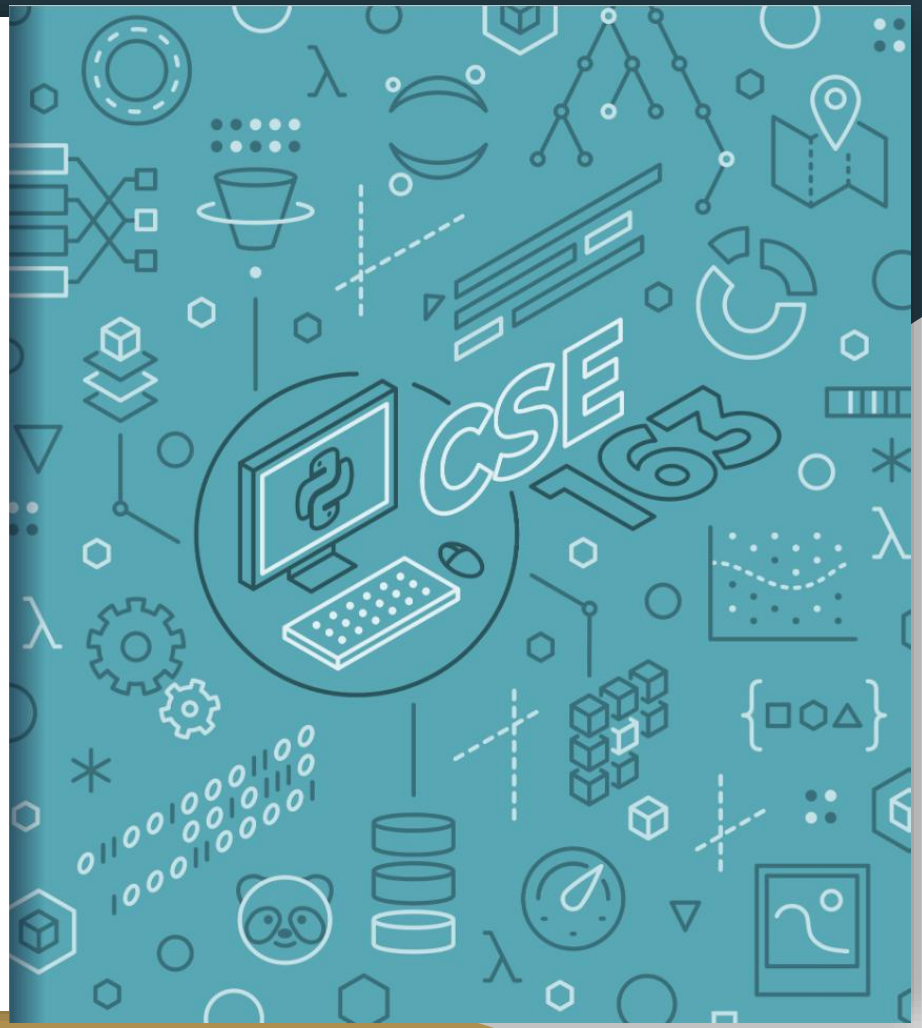
CSE 163
Statistical Learning

Adrian Salguero
Spring 2026

💭 Icebreaker (discuss with neighbors): ‘
Halfway through the quarter! How are you feeling?
Add to our Slido!

slido.com
#cse163



Announcements

- **Take Home Assessment 4: Education** released tonight, due Thursday May 7th at 11:59pm!
- **Peer Reviews for THA 3** released tonight, due Wednesday May 6th 29th at 11:59pm!
- **Project Part 1** due tonight at 11:59pm on Gradescope! (No late submissions!)
- **Lesson 14 Canvas Quiz** due tonight at 11:59pm!
- **Checkpoint 3** due Monday, May 4th at 11:59pm!

Mid Quarter Feedback

- Things going well
  - *THAs* – reinforce course content and are reasonable, creative components are good for exploring data in a less restrictive space
  - *Quiz sections* – formatted well, good additional practice
  - *Class exercises, resubmission cycles, pre-class readings*

- Areas for improvement
  - *Quiz sections* – more practice problems, more direct help, be less of a repeat of lecture
  - *Reading assignments* – seem irrelevant to course topics or assignments
  - *Lecture* – not enough time spent going over all practice problems
  - Too many assignments scattered across different platforms

- If you have additional feedback (or just want to share this feedback again for credit towards a weekly token), please fill out this feedback form.

What is Statistics?

"The study of models to gather, understand, and draw conclusions from real-world data."

- Hunter Schafer, 2020 (and probably other people too)

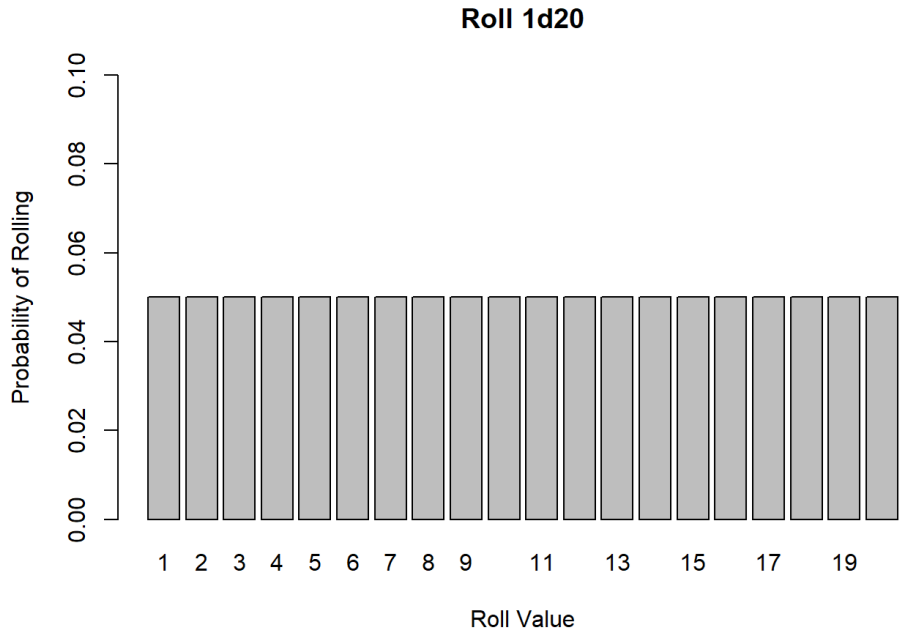
- Applications of statistics:
  - *Predicting disease outbreaks*
  - *Product testing*
  - *Sports analytics*
  - *Machine learning*
  - *…and many more*

Statistics 101

- Summary Statistics
  - Number of values in our data set
  - ***Mean*** - the “average” value
  - ***Median*** - the value in the middle of our data set
  - ***Standard deviation*** - how spread out our values are
  - ***Min***, ***max***, ***mode***, ***range***, etc.

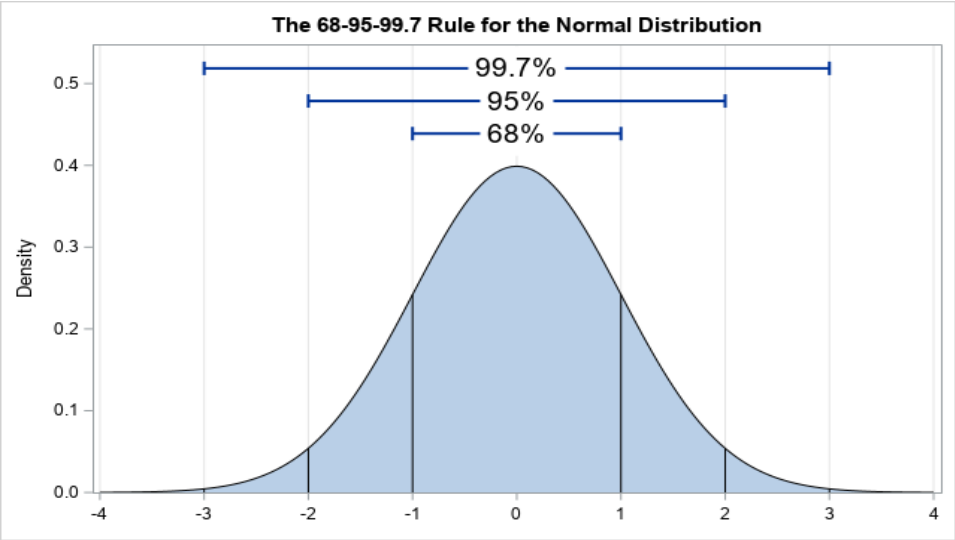
- Distributions
  - How are values in our data set distributed?
  - Uniform distribution
  - Normal distributions
  - Other types of distributions (e.g. bimodal)

Uniform Distribution - everything equally likely



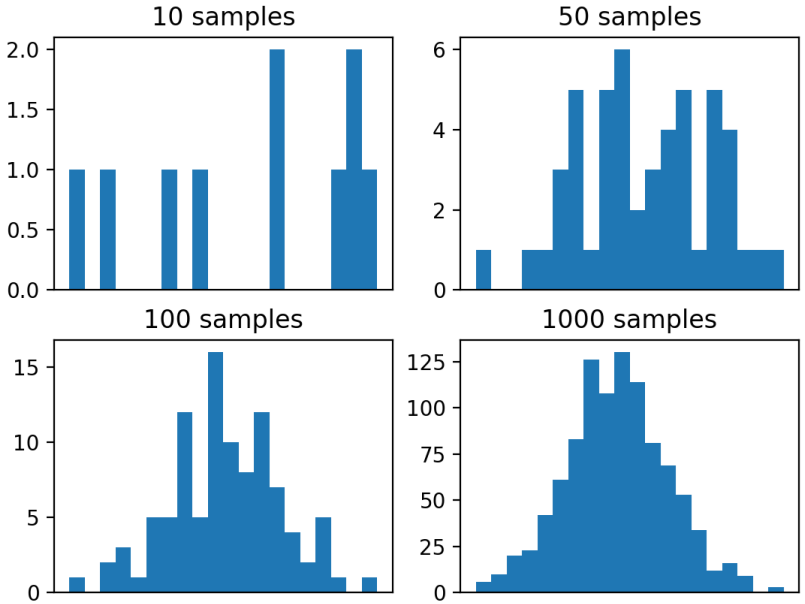
- Coin flip - Heads or Tails
- Dice roll - all sides
- ***Perfect*** random number generator

Normal Distribution - bell curved, mean centered



- Height
- IQ
- Grades

Central Limit Theorem

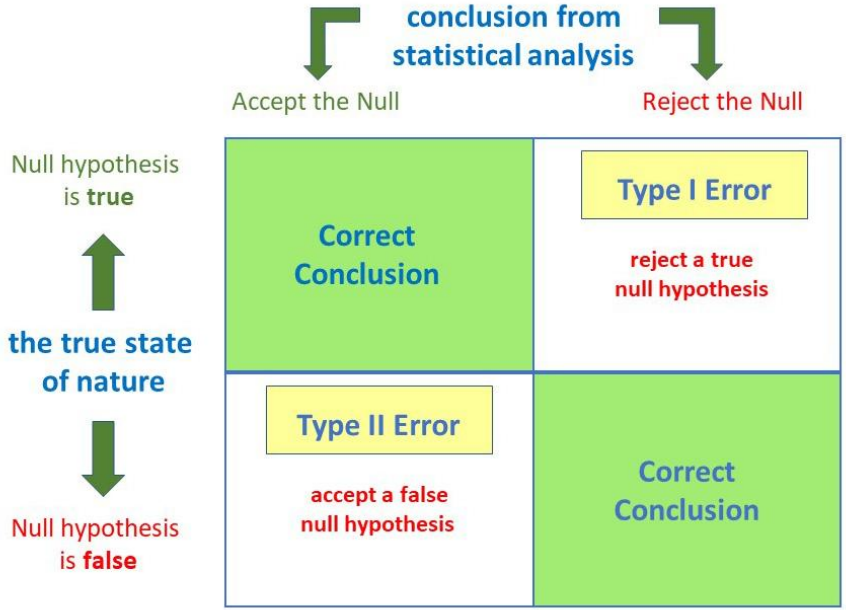


Distribution of sample means approaches a normal distribution as number of samples increases!

Hypothesis Tests

- ***Null Hypothesis*** is the hypothesis that supports the pre-existing expectations of probability (e.g. the status quo)
- ***Alternative Hypothesis*** is the hypothesis that something is different, one that if true, would allow us to reject the null hypothesis.
- Our ***p-value*** is the chance that this situation would happen in a world bound by the null hypothesis
  - If our p-value is lower than a given significance level (usually 0.05), it is significant enough to reject the null hypothesis
  - Be mindful of "***p-hacking***" - misusing data analysis techniques or data in the hopes of getting significant results
  - A non-significant result is still valuable and should be reported!

Type 1 and Type 2 Errors



Accept = "fail to reject"

Type 1 and Type 2 Errors

You visited Hall Health Center to check if you have the flu. Your doctor says you don't have it, while you actually have the flu. ***What type of error did the doctor make?***

Null hypothesis

You don't have the flu