Welcome to QuACK!

Week 4

7/23/2025

Nina, Sophie, and Victoria

Structure of the bootcamp

Session 1 (6/11)	Session 2 (7/2)	Session 3 (7/16)	Session 4 (7/23)
Intro to R and RStudio	Data cleaning and wrangling	Data visualization	Intro to stats with R

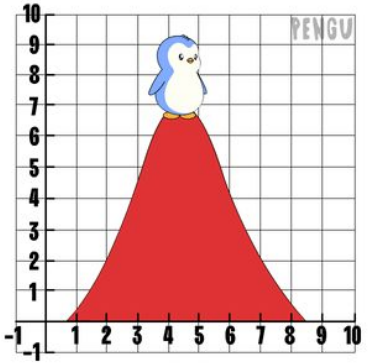
We will work with some of the same data throughout the bootcamp to get a feel for the process of exploring data and getting it ready for analysis

Sessions build on each other and we will pick up where we left off in the last session – you will get the most out of the bootcamp if you attend all four weeks!

UC Berkeley

Today's Agenda: Intro to Stats

- Warm-up activity
- Analyzing your data
  - Goals
  - Methods
  - Steps
- Group activity
- Demo: data analysis



UC Berkeley

Week 4 Warm-up – 10 minutes

1. Download this week's materials from this link: https://tinyurl.com/quack-summer
2. Find the number of respondents in the clean dataset and save it to a variable "n"
3. Find the mean, median, and range of ExperienceOther
4. Make a boxplot comparing ExperienceR for respondents born in CA vs. those not born in CA (hint: you will need to group_by() and summarize() for this!)

UC Berkeley

What is the point of analyzing your data?

→ to know to what degree you should believe that any patterns you found in your data are representative of the underlying population

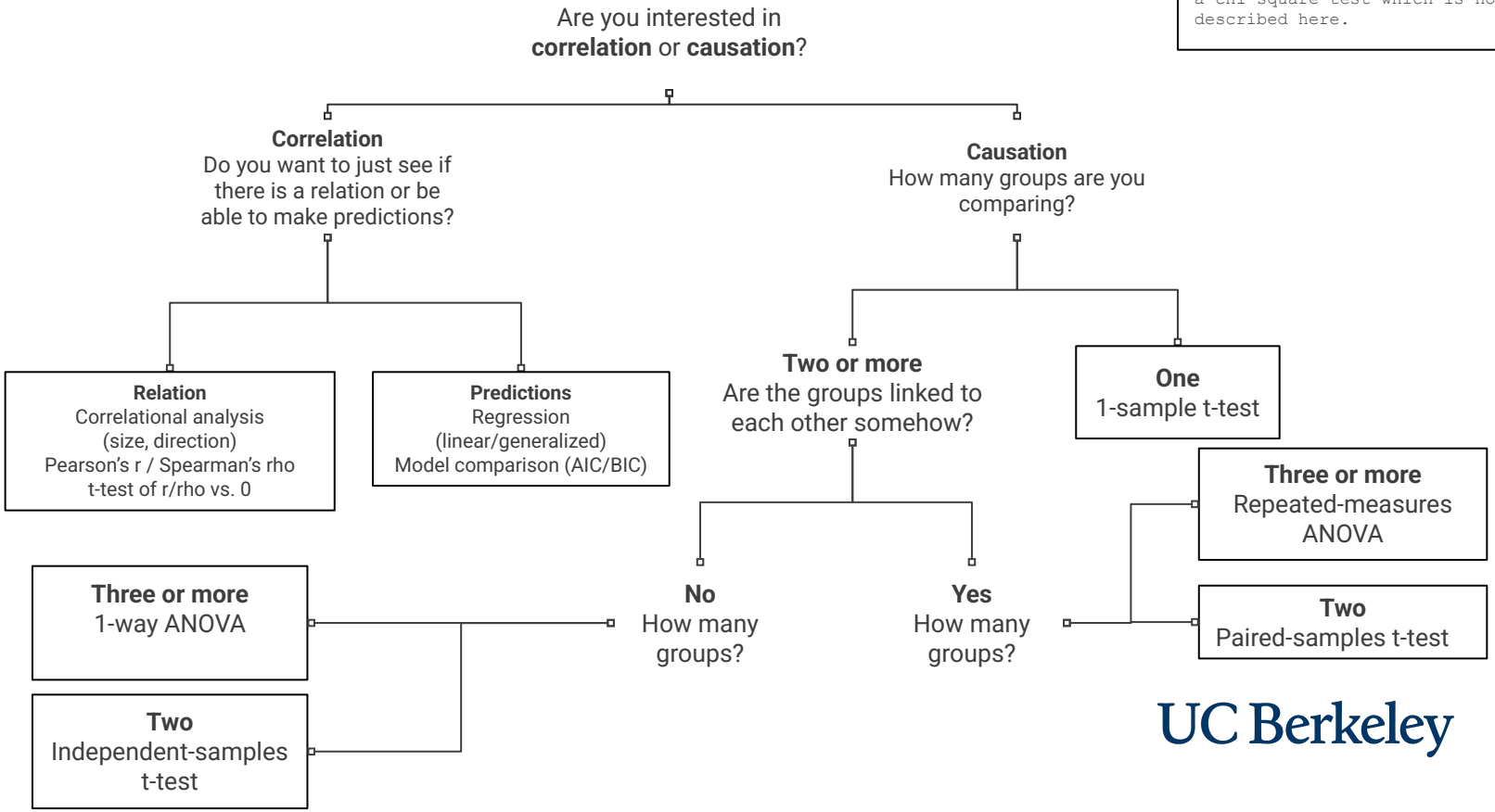
UC Berkeley

How do you decide how to analyze your data?

- Tricky because there is no one-size-fits-all answer
- For common types of datasets (e.g., comparing means of a numeric variable across two groups of equal size), there are well-established tests
- For others (e.g., time-series data or analyses on especially small datasets), you may need some more niche methods
- We are going to focus on building up intuitions and R skills for basic statistical tests

Overview of basic statistical tests

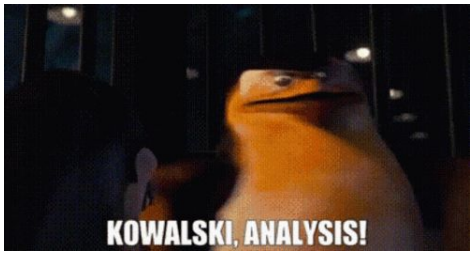
Note: the specific test you choose will also depend on the type of variables you have (continuous, categorical, count, etc.). For example, if you have categorical independent and dependent variables, you may need a chi-square test which is not described here.



UC Berkeley

Basic steps to analyzing your data
(some of these steps happen long before you even look at your data!)

- Identify your independent and dependent variables (and groups/levels)
- Decide what your null and alternative hypotheses are (H_0 and H_a)
- Pick your test, α, and do a power analysis
- Calculate your test statistics (e.g., t for t-tests or F for ANOVAs)
- Decide whether to retain or reject H_0
- Make sure you can clearly explain your conclusion in words!



Group activity!
Discuss with 2-3 people around you

In our dataset, what statistical test would you use to:

1. Determine whether LCD lab members are taller than Gopnik lab members?
2. Determine whether knowledge of R and general coding knowledge are related?
3. Determine whether year in college predicts general coding knowledge?
4. Determine whether taller people are more likely to think there was a correct answer to the last question on the survey?

And as usual, resources you can look back on even when QuACK is over 🥹

https://www.statsflowchart.co.uk/

quiz that will tell you which statistical test to use

https://statsandr.com/blog/what-statistical-test-should-i-do/

decision tree with explanations of different statistical tests

UC Berkeley

Time to get started in R!

- Go to https://tinyurl.com/quack-summer
- Download and unzip the course materials
- You will see many files:
  - Two R Markdown (.rmd) files
  - A .csv file
  - Two PDF files – slides and a practice doc
  - Some other folders – ignore these for now
- Open the R Markdown file marked “starter_code” – this is where we will code together for the first part of the session
- The .csv file contains the data we will be working with
- The practice PDF contains practice exercises for you to do during the second part of the session

UC Berkeley