

Introductory Statistics – Day 1

Introduction

*Data is the sword of the 21st century,
those who wield it well, the Samurai.*

– Jonathan Rosenberg

Brief Introduction

- Moodle
- Syllabus: (on Moodle)
- WeBWork: Online Homework System
- NearPod: Classroom Polling Software

Activity 1:

- Your instructor will give you a cup.
- Using your cup, take as many M&Ms as you like from the bowl at the front of the room.
- When you return to your seat please record the following on the Google Sheet linked from our class Moodle page.
 - 1 How many M&Ms did you take?
 - 2 What was the size of your cup?
 - 3 What is your gender?
 - 4 What is your class standing? (Freshman, Sophomore, Junior, Senior)
 - 5 Are you from Montana?
 - 6 Did you eat breakfast today?

We have our first data set!

- What do you notice? What do you curious about?
- Suppose that we are interested in investigating what factors affect how much food you consume (in this case, how many M&Ms you took).
- Pose several suggestions for what we should do with this data.

What counts as evidence?

It might be tempting to address your research questions with anecdotes.

Definition

Anecdotes are stories about individual data points.

Example:

Anecdote 1: The student at the end of the table took way more M&Ms. Maybe its because he skipped breakfast or maybe it was because he had a bigger cup.

Anecdote 2: The student in the corner didn't take much. Maybe its because she doesn't like chocolate, or maybe it was because she had a smaller cup.

The problem with anecdotes is that you can find one to support almost any theory or opinion. Statisticians rely on randomized controlled experiments to test theories, rather than anecdotal evidence.

Research question:

How does the size of the cup affect how much candy people tend to take?

In order to investigate our research question about the impact of cup size on candy quantity, there are a number of questions to consider.

Key Question 1

What **population** are we interested in studying?

- A. M&Ms
- B. cups
- C. students in our class
- D. students at Carroll
- E. adults in general
- F. people in general
- G. candy in general

Key Question 2

Why was randomization important?

First, discuss this with your neighbor.

Next, submit your best answer to Nearpod.

Key Question 3

How could we best describe the number of M&Ms chosen?

- A. List everyone's counts and cup size
- B. Create 2 graphs of the results for the two cup sizes,
- C. Provide the minimum and maximum for each cup size,
- D. Provide the mean for each cup size,
- E. Provide the median for each cup size,
- F. Other suggestions?

How does the size of the cup play a role?

Discuss the following with your neighbors:

- Do you think the size of the cup tends to play a role in M&M count?
- What graphical display might help us better understand the data?
- Does gender tend to have an impact too? How can we investigate?

Characteristics of a controlled experiment

Replication: Same treatments assigned to different experimental units.

Control: Grouping similar experimental units

- *Control Group:* Mimics “normal” as much as possible
- *Experimental Group:* Experimental treatment is given to this group.

Otherwise, the groups are treated exactly the same

Randomization: All treatments and groups (to the extent possible) must be assigned randomly.

Smooths biases and other confounding variables.

Every unit has equal probability of getting each treatment.

Ethical Considerations

External review is performed before the research begins.

© Carroll: by the Institutional Review Board (IRB).

Conclusion: *A cause/effect relationship can **only** be determined by a randomized controlled experiment*

Characteristics of an observational study

Researchers gather data without attempting to influence a variable.

Studies can be **prospective** (collecting data going forward) or **retrospective** (using existing data)

Random sampling

- *Simple random sampling*
- *Stratified random sampling* (take a random sample in each subsection of population)
- *Cluster sampling* (data is clumped into clusters and some clusters are randomly selected)

See section 1.4 in the OpenIntro text for more details.

Ethical considerations center more around privacy / confidentiality of data.

Conclusions may be about associations, but not cause/effect relations.