

PS 211: Introduction to Experimental Design

Spring 2025 · Section C1

Lecture 5: Overview of Sampling, Probability, and Hypothesis Testing

Updates and reminders

- **Homework 1** was due yesterday. Hopefully you have submitted. If not, please do so ASAP.
- We will post homework answers and grades on Friday morning, at which point we will no longer accept late submissions.
- We will not be returning Homework 1. If you are not sure where you got points off, please review the posted answers, and come to office hours if you have questions.
- The next homework will be posted after Exam 1, and will be due on **Friday, October 10**.
- We are going to go over some more information about how to use **R** and **R Studio** today.

Updates and reminders (Continued)

- Exam 1 is scheduled for **Thursday, September 25** during our regular class time.
- Exam 1 will cover Lectures 1–5.
- I will post a review sheet by Friday with a list of all the topics the exam will cover.
- You can print this review sheet and take hand-written notes on it (both sides) to use during the exam.
- Alternatively, you can write your own notes on a blank sheet of paper (both sides).
- You will *not* be tested on R code, but you should be able to interpret plots and output from R. For example, we may show you a histogram or frequency table created in R and ask you questions about them.
- The class before the exam (Tuesday, September 23) will be a review session.

Before we begin discussing sampling and probability...

- Resolving key points of confusion about R:
 - Opening and closing R Studio without losing work.
 - Files: .R, .Rmd, .Rproj, .html, .pdf. *What are they and when to use each?*
 - Saving your work: saving code vs. workspace variables.

Review: Central Tendency

- **Mean (μ , **M**):** Average score. Sensitive to outliers.
- **Median:** Middle score. Robust to outliers.
- **Mode:** Most frequent score.

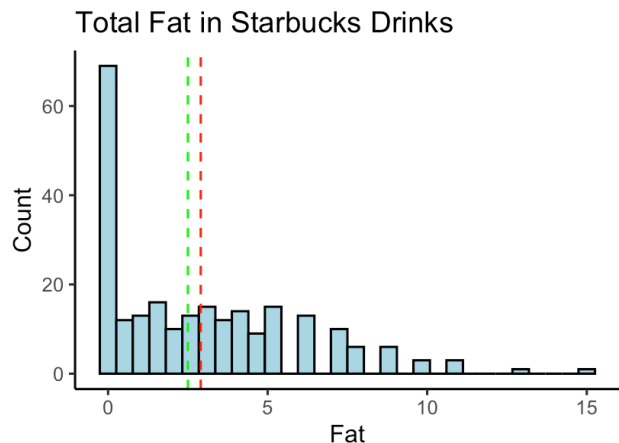
Which one of these measures is useful for nominal data?

Answer: Only the mode is useful for nominal data.

Nominal data represent categories without a meaningful order or numerical value.

The mean and median require numerical values and are not meaningful for nominal data.

Review: Central Tendency



The histogram shows the fat content (in grams) of Starbucks drinks.

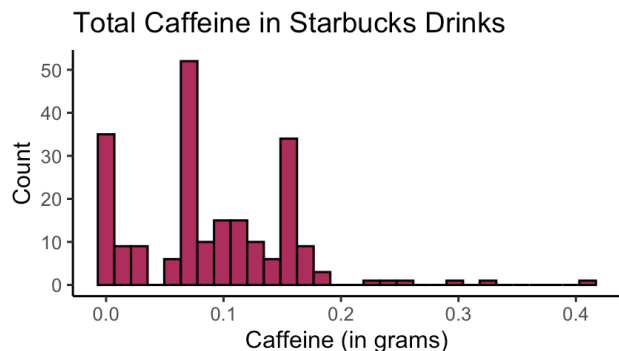
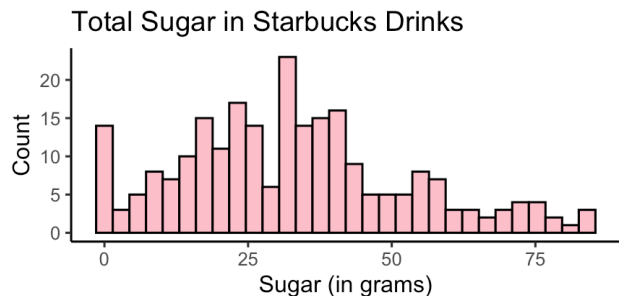
Which line represents the mean and which line represents the median? How do you know?

Answer: The red line represents the mean, and the green line represents the median. The mean is higher than the median because the distribution is positively skewed (right-skewed) due to a few drinks with very high fat content.

What is the mode?

Answer: The mode is 0 grams of fat, as indicated by the tallest bar in the histogram.

Review: Variability



Which is more variable: sugar content of Starbucks drinks or their caffeine content? How do you know? Which measure of variability would you use to quantify this?

Answer: The sugar content is much more variable than the caffeine content. We could use any measure of variability (range, variance, standard deviation) to quantify this.

Note: Look at the x-axis scales. The sugar histogram ranges from 0 to 100 grams, while the caffeine histogram ranges from 0 to .4 g!

Review: Population Variance: Step by Step

1. Compute the population mean (μ).
2. For each score X_i , compute its deviation: $(X_i - \mu)$.
3. Square each deviation.
4. Sum all squared deviations.
5. Divide by the total number of scores (N).

Formula:

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$$

Review: Population Standard Deviation

- The standard deviation is the square root of the variance.

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}}$$

- It tells us the **typical amount that scores deviate from the mean**.

Sample Variance: Step by Step

1. Compute the sample mean (M).
2. For each score X_i , compute $(X_i - M)$.
3. Square each deviation.
4. Sum all squared deviations.
5. Divide by $N - 1$.

Formula:

$$s^2 = \frac{\sum_{i=1}^N (X_i - M)^2}{N - 1}$$

Why do we use the sample mean instead of the population mean here?

Answer: We use M (the sample mean) instead of μ (the population mean) because we typically do not know the population mean when working with a sample. The sample mean is our best estimate of the population mean based on the data we have.

Here, we use our samples to estimate *both* the population mean and the population variance. That means our estimate of the mean is biased towards the samples.

Why divide by $N-1$?

- Using M instead of μ makes our estimate **too close** to the data.
- This tends to **underestimate the true population variance**.
- Dividing by $N - 1$ (instead of N) corrects this bias.
- This adjustment is called **Bessel's correction**.



Bessel's correction: dividing by $N - 1$ (instead of N) corrects bias in the estimate of the population variance.

Sample Standard Deviation

- The sample SD is the square root of the sample variance:

$$s = \sqrt{\frac{\sum_{i=1}^N (X_i - M)^2}{N - 1}}$$

- This provides an **unbiased estimate** of the population's spread.

Why divide by $N-1$?

Khan Academy Video Explanation:

<https://www.youtube.com/watch?v=KkaU2ur3Ymw&t=520s>

New stuff! Sampling and probability

- Now that we understand variance and SD, we can see **why sampling matters**.
- We will also discuss **probability**, which is the foundation of inferential statistics.

Why do we sample?

- We usually want to know about a **population** (e.g., all college students).
- But we can't measure everyone, so we take a **sample** (e.g., 20 BU students who participate in a study).
- We use the sample to estimate population parameters (e.g., mean, variance).
- We also use samples to test general hypotheses about broader populations. Inferential statistics lets us do this!

Two main types of samples in Psychological Science

- **Random Sample:** Every member of the population has an equal chance of being selected.
 - Advantage: More representative.
 - Disadvantage: Expensive, often impossible.
- **Convenience Sample:** Uses participants who are readily available.
 - Common in psychology (e.g., Psych 101 students).
 - Easier and cheaper, but may introduce bias.

Examples: Samples in Psychological Studies

Participants. Ninety-two participants, evenly distributed between the ages of 10 and 25 years, were recruited from New York University and the surrounding community for an in-person behavioral study. Participants were recruited through advertisements on social media, flyers around New York University, and science fairs and events throughout New York City.

Is this a random or convenience sample? How do you know?

Answer: This is a convenience sample. The participants volunteered to take part in the study, and they were recruited through flyers and online postings. This means they were not randomly selected from the general population.

A researcher wants to study whether giving BU students free coffee improves their mood. They randomly select 50 students from the entire BU student directory to participate in the study.

Is this a random or convenience sample? How do you know?

Answer: This is a random sample. The researcher randomly selected participants from the entire BU student directory, giving every student an equal chance of being chosen.

Random Sampling: One way to do it

- Use a computer program to generate random numbers.
- Use those numbers to select participants from a list.

You use a computer program to randomly select 10 students from your 100-person Psych 101 class to participate in a study. Your computer spits out the following numbers: 1, 2, 3, 4, 44, 56, 57, 58, 83, 99. They don't look very random to you. Should you re-run the program to get a different set of numbers?

Answer: No! Randomness can produce streaks or clusters of numbers. The set of numbers you got is just as random as any other set.

Convenience Sampling

- In practice, true random samples are rare.
- Psychologists often rely on **convenience samples**.
- Example: Volunteer samples (people who sign up for studies).
 - Benefits: Motivated and willing participants.
 - Downsides: Self-selection bias. The people who volunteer may differ from those who don't.

Generalizability & External Validity

Modern Sampling Techniques

- Online crowdsourcing solicits input from a large group of people.
- There are many online platforms for this:
 - e.g., Amazon Mechanical Turk (MTurk), Prolific, Qualtrics Panels.
- Participants sign up to complete surveys or tasks for pay.
- These platforms provide access to large, diverse samples quickly and cheaply.

Crowdsourcing

What are some potential advantages and disadvantages of using crowdsourcing platforms for research?

Advantages:

- Access to a large and diverse participant pool.
- Cost-effective and time-efficient.
- Ability to reach participants from various geographic locations.

Disadvantages:

- Potential for low-quality data (e.g., inattentive respondents).
- Limited control over the research environment.
- Limited types of data can be collected (e.g., no brain imaging)

Random Assignment

- **Random sampling:** Selecting participants for a study.
- **Random assignment:** Assigning participants to conditions within a study.
- Ensures every participant has an equal chance of being in any condition (i.e., level of the independent variable).

Why should researchers care about random assignment if they don't seem to care about random sampling?

Answer: Random assignment helps to reduce the risk of confounding variables by ensuring that the groups are comparable at the start of an experiment. This is important for understanding causality even if the sample itself is not representative of the broader population.

Example: Random Assignment

- Imagine you want to run a between-subjects version of the Stroop task.
- You decide to run this experiment on your PS 211 classmates.
- You divide the class into two groups:
 - Group 1: Congruent condition. (e.g., the word "red" printed in red ink).
 - Group 2: Incongruent condition. (e.g., the word "red" printed in blue ink).
- You assign the first 10 students who arrive to Group 1, and the next 10 to Group 2.

Is this a good method of random assignment? Why or why not?

Review: Descriptive vs. Inferential Statistics

- **Descriptive stats:** Summarize data from a sample. "Describe" the sample.
 - Examples: "The average amount of sugar in the five drinks I tried from Starbucks is 30 grams."
- **Inferential stats:** Allow us to draw conclusions about populations from samples.
 - Examples: "Based on my sample, I estimate that the average amount of sugar in all Starbucks drinks is between 20 and 50 grams."
- But ... how do we ensure that these inferences are valid? **Probability!**



Probability

- **Probability** = the likelihood that a specific outcome will occur, out of all possible outcomes.
- Ranges from 0 (impossible) to 1 (certain).
- Example: Probability of flipping heads on a fair coin = 0.5.
- Probability helps us understand and quantify uncertainty.
- **Experimental probability** = relative frequency of an outcome in repeated trials.
- **Theoretical probability** = based on known possible outcomes (e.g., coin flips, dice rolls).

Probability Terms

- **Trial:** Each repetition of an event or each occasion a procedure is carried out (e.g., a dice roll).
- **Outcome:** Result of the trial.
- **Success:** The outcome of interest (e.g., heads).

If you want to determine the probability of a coin landing on heads, what is a trial, an outcome, and a success?

Answer:

- Trial: Each flip of the coin.
- Outcome: The result of the coin flip (either heads or tails).

Independent Trials

- Independent = one trial's outcome does not affect another.
- Example: slot machines, coin flips.
- Yes, rare streaks can occur!

Probability: Our intuitions are bad!

- We are generally bad at intuitively grasping probability.
- We see patterns where none exist (e.g., gambler's fallacy).
- We overestimate the likelihood of vivid events (e.g., shark attacks).
- We underestimate the likelihood of common events (e.g., car accidents).
- We notice evidence that confirms our beliefs, and ignore contrary evidence (confirmation bias).

Probability intuitions: The birthday paradox

- What's the probability that two people in this class of 45 have the same birthday (out of 365 days)?

What do you think the answer is closest to: 1%, 10%, 30%, 50%, 70%, or 90%?

Answer: The probability is surprisingly high! It's actually 94.1%.

Probability that at least two people share a birthday in a group of 45 is calculated using the complementary probability that no one shares a birthday, which is then subtracted from 1.

- Probability that no one shares a birthday:

$$P(\text{no shared birthday}) = \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \dots \times \frac{321}{365}$$

Probability is important for ... hypothesis testing!

- Hypothesis-testing: Using sample data to evaluate two competing hypotheses about a population.

Example: Are students who take PS 211 happier than other BU students?

Experiment:

- Administer survey measuring happiness (scale 1-10) to a random sample of 10 PS 211 students and a random sample of other BU students.
- Results: PS 211 students: 8, 7, 9, 6, 8, 7, 9, 10, 8, 7, $M = 8.0$
- Other BU students: 5, 6, 4, 5, 6, 5, 4, 5, 6, 5, $M = 5.0$

Are the 10 students in PS 211 on average happier than the 10 other BU students?

Probability is important for ... hypothesis testing!

Can we conclude that PS 211 students are happier than other BU students? Why or why not?

Answer: We cannot conclude that PS 211 students are happier than other BU students based solely on this sample. The observed difference in means could be due to random chance, especially given the small sample sizes. We need to conduct a hypothesis test to determine if a real difference likely exists in the population.

Hypothesis Testing

- In psychological research, we define two hypotheses:
 - **Null Hypothesis (H_0):** There is no difference.
 - The null hypothesis is a statement that postulates that there is no difference between populations or that the difference is in a direction opposite of that anticipated by the researcher.
 - Any observed difference is due to random chance or sampling error.
 - **Research Hypothesis (H_1):** There is a difference.
 - There is a difference between populations or sometimes, more specifically, that there is a difference in a certain direction, positive or negative; also called an alternative hypothesis.
 - An observed difference reflects a true effect in the population.
- Goal: Use data to test which is more plausible.

Hypothesis Testing: Examples

In the PS 211 happiness example, what are the null and research hypotheses?

Answer:

- Null Hypothesis (H_0): There is no difference in happiness levels between PS 211 students and other BU students.
- Research Hypothesis (H_1): PS 211 students are happier than other BU students.

Hypothesis Testing: Examples (Continued)

In the Stroop task example, what are the null and research hypotheses?

Answer:

- Null Hypothesis (H_0): There is no difference in reaction times between the congruent and incongruent conditions.
- Research Hypothesis (H_1): Reaction times are slower in the incongruent condition compared to the congruent condition.

Hypothesis Directions

- Both of these examples had **directional** research hypotheses (PS 211 students are happier; Stroop incongruent is slower).
- Sometimes, we have **non-directional** research hypotheses (e.g., there is a difference, but we don't know the direction).
- The type of hypothesis affects the type of *significance test* we use. We will cover *significance testing* in more detail later.

One-tailed vs. Two-tailed Tests

- **One-tailed test:** Tests for an effect in one direction (e.g., PS 211 students are happier).
- **Two-tailed test:** Tests for an effect in either direction (e.g., there is a difference in happiness, but we don't know which group is happier).
- One-tailed tests have more statistical power, but can miss effects in the opposite direction.

Rejecting the Null Hypothesis

The goal of statistics is to determine whether we can reject the null hypothesis.

- We either reject or fail to reject H_0 based on the data.
- If the data are very unlikely under H_0 , we reject H_0 in favor of H_1 .
- If the data are not unlikely under H_0 , we fail to reject H_0
- We never "accept" H_0 , because we can't prove it true.
- We use probability to assess how likely our data are under H_0 .

What's wrong with this statement of experiment results: "PS 211 students are equally as happy as other BU students."?

Answer: This statement suggests that we have proven the null hypothesis (H_0) to be true, which is not possible. We can only fail to reject H_0 .

How should we fix it?

Answer: A better way to state the results would be: "We did not find evidence that PS 211 students are happier than other BU students."

Two ways to be wrong: Type I and Type II Errors

- **Type I error:** Rejecting H_0 when it is true (false positive).
 - "You said PS 211 students are happier, but they aren't."
- **Type II error:** Failing to reject H_0 when it is false (false negative).
 - "You said PS 211 students aren't happier, but they are."
- Both errors have consequences.

Type I and Type II Errors: Example

You have a cough and decide to take a covid test. What is your null and alternative hypothesis?

Answer:

- Null Hypothesis (H_0): You do not have COVID-19. (i.e., you are not 'different' from your usual healthy state)
- Alternative Hypothesis (H_1): You have COVID-19. (i.e., you are 'different' from your usual healthy state)

If the test comes back positive, but you actually don't have COVID-19, what type of error is this?

Answer: This is a Type I error (false positive). You rejected the null hypothesis (you do not have COVID-19) when it was actually true.

That's all for today!

See you Thursday – we will combine probability with hypothesis testing and inference.