PS 211: Introduction to Experimental Design

Spring 2025 · Section C1

Exam 1 Review Session

Updates and reminders

- Exam 1 is on Thursday.
- You will not need a calculator for the exam.
- You can bring one 8.5"x11" sheet of notes (front and back).
- The exam will cover everything up to and including lecture 5: null and research hypotheses
- The exam will consist of 33 multiple choice questions. You will only need to answer 30 questions correctly to get 100%.
- Having your computer or phone out during the exam will result in a 0 for the exam.
- You will have the entire class period (75 minutes) to complete the exam.
- Please bring a pencil.

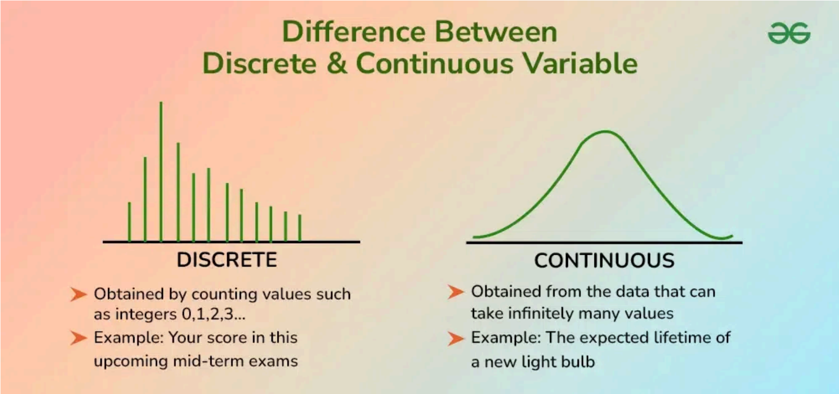
Subcategories of variables

Continuous (quantitative)

- Can assume any value within a range.

Discrete (quantitative or qualitative)

- Can only take on specific values.



Types of discrete variables

Nominal variables

- Used for labeling variables without any quantitative value.
- You can assign numbers to these variables, but the numbers do not have any mathematical meaning.
- Examples: hair color, gender, species.

Ordinal variables

- Used for labeling variables with a meaningful order or ranking.
- The intervals between the ranks may not be equal.
- No fractions or decimals.
- Examples: class rank, satisfaction ratings (e.g., 1-5 stars).

Types of continuous (or discrete) variables

Interval variables

- Used for labeling variables where the intervals between values are meaningful and equal.
- No true zero point (i.e., zero does not indicate the absence of the quantity being measured).

Ratio variables

- Interval variables with a true zero point (i.e., absence of the quantity being measured).

THE FOUR LEVELS OF MEASUREMENT:

	Nominal	Ordinal	Interval	Ratio
Categorizes and labels variables	✓	✓	✓	✓
Ranks categories in order		✓	✓	✓
Has known, equal intervals			✓	✓
Has a true or meaningful zero				✓

Can you think of examples of each type?

Types of variables: practice

Reliability and Validity

Reliability: consistency across repeated measures

Validity: measuring what it’s supposed to measure

Examples:

- A *reliable* scale gives the same weight each time.
- A *valid* scale gives the true weight.

Can a measure be...

Reliable but not valid?

Valid but not reliable?

Can a measure be...

Reliable but not valid?

- Yes! A broken scale that always reads 5 lbs too heavy is reliable (consistent) but not valid (not accurate).
- Yes! An IQ test may give you the same score each time (reliable) but not actually measure intelligence (valid).

Valid but not reliable?

Can a measure be...

Reliable but not valid?

- Yes! A broken scale that always reads 5 lbs too heavy is reliable (consistent) but not valid (not accurate).
- Yes! An IQ test may give you the same score each time (reliable) but not actually measure intelligence (valid).

Valid but not reliable?

- No! If a measure is not consistent, it cannot be valid because it cannot be accurate.
- A measure can't measure what it's supposed to measure if it gives different results each time.
- Example: A scale that gives different weights each time is neither reliable nor valid.

Reliability and Validity: Practice

A researcher wants to measure stress levels in college students. They develop a new questionnaire that asks about various stress-related symptoms. They administer the questionnaire to a group of students and find that the results are consistent when the same students take the test multiple times. However, when they compare the questionnaire results to physiological measures of stress (like cortisol levels), they find no correlation.

How would you describe the reliability and validity of this new questionnaire?

Answer: The questionnaire is reliable because it produces consistent results when the same students take it multiple times. However, it is not valid because it does not correlate with physiological measures of stress, indicating that it may not be accurately measuring stress levels.

Reliability and Validity: More Practice

You want to argue that SAT scores are not a reliable or valid measure of intelligence.

What evidence would you use to support your argument?

Answer: You could point to studies showing that SAT scores can vary significantly for the same individual when taken multiple times (indicating low reliability). Additionally, you could cite research showing that SAT scores do not strongly correlate with other measures of intelligence, such as IQ tests or academic performance (indicating low validity).

Operational definitions: Needed to test hypotheses

- **Operational definitions**: Specify the observations or procedures used to measure or manipulate a variable.
  - Example: Happiness could be operationally defined as "self-reported happiness on a 1-10 scale" or "number of smiles in a 5-minute video""

How could you operationalize musical ability?

Answer: Musical ability could be operationalized as "score on a standardized music listening test" or "number of musical instruments played proficiently."

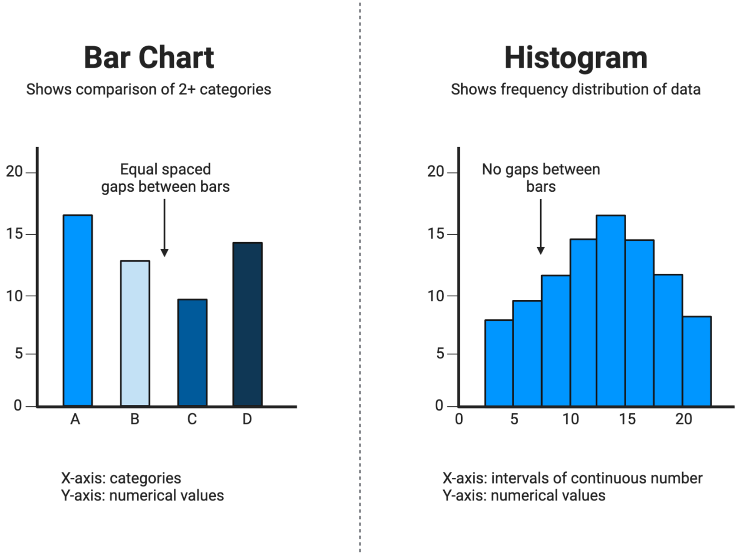
How could you operationalize stress?

Histograms

- Graphs are often more effective than tables for seeing patterns in data.
- A **histogram** is a graphical representation of a grouped frequency table.
- The x-axis shows the intervals (or "bins") of continuous values.
- The y-axis shows the frequency (or count) of observations in each interval.

Histograms (Continued)

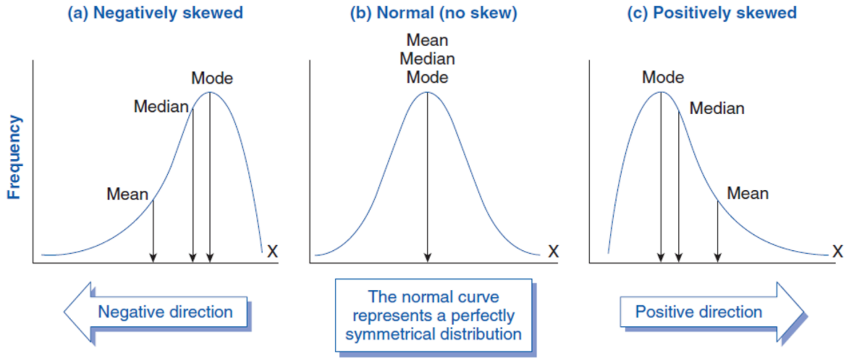
- A histogram *looks* like a bar graph, but **the y-axis always represents frequency or count**, not a separate variable.



Note: Histograms are for continuous variables only. Classically, the bars should touch to indicate this.

Skew

- When data are not symmetrically distributed, we say they are **skewed**.
- The ends of the distribution are called the "tails."
- **Positively skewed** (or right-skewed) distributions have a long tail on the right side.
- **Negatively skewed** (or left-skewed) distributions have a long tail on the left side.

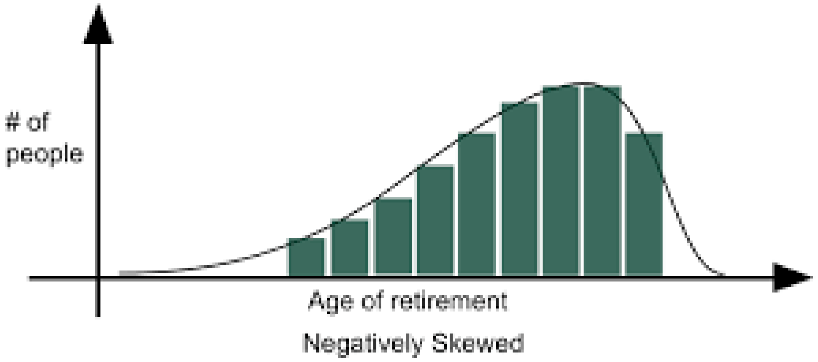


Positive skew and floor effects

- **Positive skew**: The tail of the distribution extends to the right.
- Sometimes a positive skew can indicate a **floor effect**. Not always!!
- A **floor effect** occurs when a large number of observations cluster at the lower end of the scale, with few observations at the higher end.

Negative skew and ceiling effects

- **Negative skew**: The tail of the distribution extends to the left.
- Sometimes a negative skew can indicate a **ceiling effect**. Not always!!
- A **ceiling effect** occurs when a large number of observations cluster at the higher end of the scale, with few observations at the lower end.



Practice: Understanding floor and ceiling effects

A teacher gives a very easy test to a class of students. Most students score between 90 and 100, with a few scoring slightly lower. What kind of skew would you expect in the distribution of test scores? Is there a floor or ceiling effect?

Answer: The distribution of test scores would likely be negatively skewed, with a ceiling effect present. Most students are clustered at the high end of the scale (90-100).

Why is this ceiling effect potentially problematic?

Answer: The ceiling effect is problematic because it limits the ability to differentiate between students' performance at the high end of the scale.

Practice: Distributions and floor/ceiling effects (continued)

Consider the following scenarios and determine whether the distribution of scores is likely to be positively skewed, negatively skewed, or normally distributed. Also, identify any potential floor or ceiling effects.

A researcher asks participants to rate their satisfaction with a new product on a scale from 1 to 10. Most participants give ratings between 8 and 10, with very few giving lower ratings.

Answer: The distribution is likely to be negatively skewed, with a ceiling effect present. Most participants are clustered at the high end of the scale (8-10).

A researcher asks parents how many hours their children spend playing video games each week. Most parents report that their children play between 0 and 2 hours, with a few reporting higher amounts.

Answer: The distribution is likely to be positively skewed, with a floor effect present. Most children are clustered at the low end of the scale (0-2 hours), but a few play significantly more.

Practice: Distributions and floor/ceiling effects (continued)

A researcher asks students how many hours they spend on social media, with a scale that ranges from 0 to 10 hours per week.

Answer: The distribution is likely to be negatively skewed, with a ceiling effect present. Most students are clustered at the high end of the scale (8-10 hours).

What about scores on the exam on Thursday?

Answer: We'll see! I hope that the distribution is either normally distributed or negatively skewed, with a ceiling effect present.

Practice: Distributions and floor/ceiling effects (continued)

Mean, median, and mode

Mean: The arithmetic average

Median: The middle score

Mode: The most common score

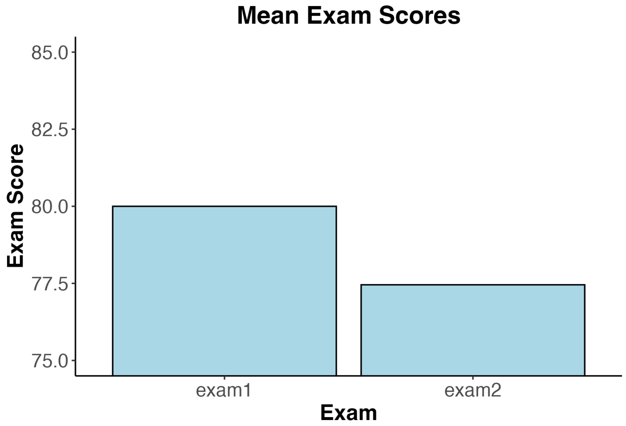
Comparing Mean, Median, and Mode

- In a normal distribution, the mean, median, and mode are equal!
- In a negatively skewed distribution, the mean is less than the median, which is less than the mode.
- In a positively skewed distribution, the mode is less than the median, which is less than the mean.

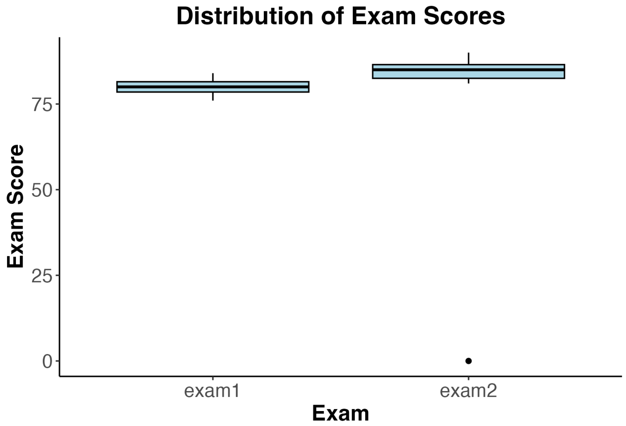
Outliers

- Outliers are extreme values that differ greatly from the rest of the data.
- Outliers can distort the mean, making it less representative of the dataset.
- The median is less affected by outliers and can be a better measure of central tendency in skewed data.

Visualizing outliers

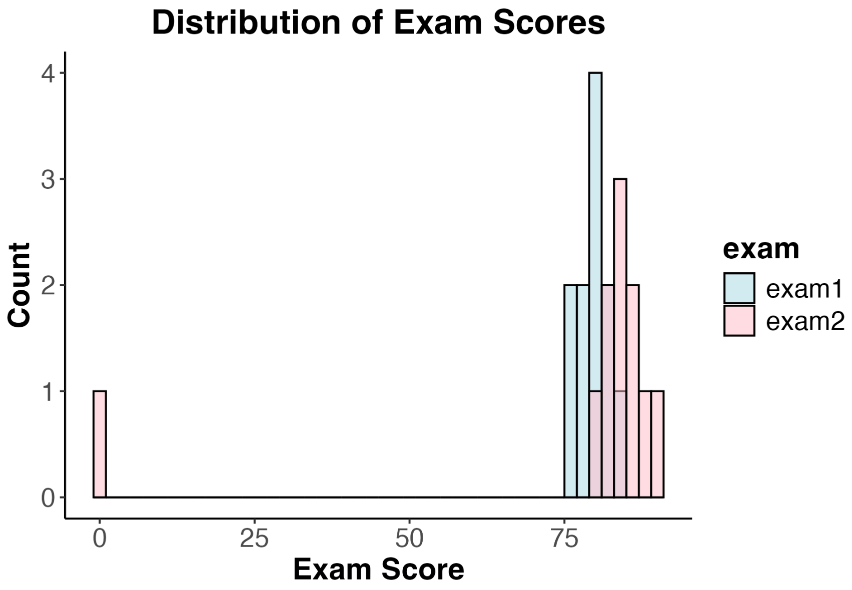


Which exam was harder?



Which exam was harder?

Visualizing outliers (continued)



Which exam was harder?

Variance

What if we want a measure of spread influenced by ALL scores?

- We can compute the distance each score is from the mean (the deviation), and take the average of that.
- But... if we add up all the deviations, they will always equal zero!

Example

- Scores: 70, 80, 90
- Mean: 80; Deviations: -10, 0, +10
- Sum of deviations: -10 + 0 + 10 = 0

To solve this, we square each deviation (to make them positive) before summing them!

Variance (continued)

Variance (continued)

Review: Variance

Why do we care about variance and standard deviations?

- Variance and standard deviations tell us how **spread out** scores are around the mean.
- They are used in many statistical tests (e.g., t-tests, ANOVA).
- They are also useful in the real world!

Review: Variance in the real world

Imagine you have the choice between two summer jobs: You can be a lifeguard or you can be a tour guide. Both jobs pay, on average, $15/hour. However, the lifeguard job has a standard deviation of $1/hour, while the tour guide job has a standard deviation of $10/hour. Which job would you choose? Why?

Imagine that you are a boating instructor and you need to order lifejackets for a group of 100 people. You know their average weight is 150 lbs. The options for lifejacket sizes range from XXS to XXL, which correspond to different weights. How would knowing the standard deviation of their weights help you decide how many of each lifejacket size to order?

Review: Variance in the real world

Now imagine that you need to estimate the standard deviation of the weights of the 100 people. You decide to randomly poll 10 of the 100 and ask their weights. How would you use the weights of the 10 people to estimate the standard deviation of the weights of all 100 people?

Answer: Calculate the standard deviation of the 10 weights (your sample) and use that as an estimate of the standard deviation of all 100 weights (your population).

Let's do it: Here are your sample weights (in lbs): $120, 130, 140, 150, 160, 170, 180, 190, 200, 210$

Step 1: Calculate the mean of the weights.

$$M = (120 + 130 + 140 + 150 + 160 + 170 + 180 + 190 + 200 + 210)/10 = 165$$

Step 2: Calculate the sum of squared deviations from the mean.

$$\sum (X_i - M)^2 = (120 - 165)^2 + (130 - 165)^2 + (140 - 165)^2 ... + (210 - 165)^2 = 8250$$

Review: Variance in the real world (continued)

Step 3: Calculate the *sample* variance.

$$Variance = \frac{\sum(X_i - M)^2}{n-1} = \frac{8250}{10-1} = 916.67$$

Step 4: Calculate the *sample* standard deviation.

$$SD = \sqrt{Variance} = \sqrt{916.67} \approx 30.28$$

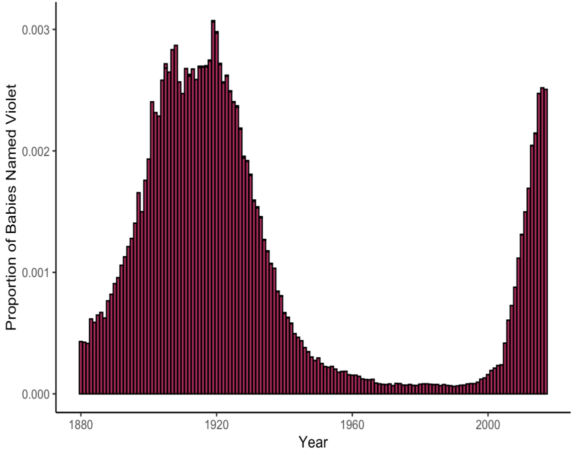
Now we have our *sample* standard deviation of 30.28 lbs, which we can use as an estimate of the *population* standard deviation of all 100 weights.

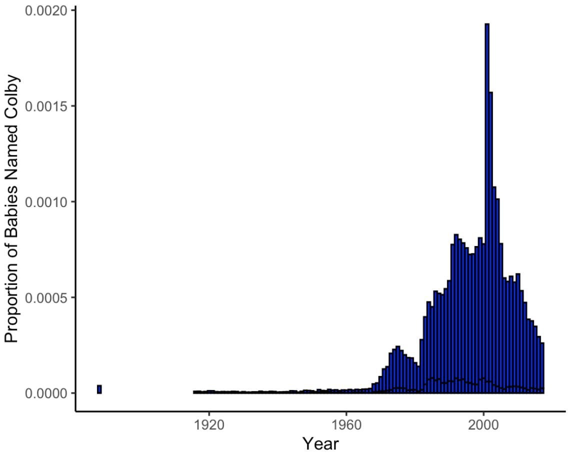
Variance is a very important concept in statistics!

It will come up again and again in this class and in future statistics classes.

Variability: Practice (Continued)

Which distribution likely has a greater standard deviation?





Answer: The "Violet" distribution likely has a greater standard deviation because its scores are more spread out from the mean.

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.

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.

That's all for today!

Good luck on Exam 1 on Thursday!