BOSTON UNIVERSITY

bu.edu

PS 211: Introduction to Experimental Design

Fall 2025 · Section C1

Lecture 10: Paired-Sample *t* Tests and Independent-Sample *t* Tests

Updates & Reminders

- We'll post Exam 2 grades tomorrow afternoon and go over questions in class on Thursday.
- Lots more to say about it on Thursday!
- Homework 3 will be posted by the end of the week and will be due Friday, Oct. 31.
- A mid-semester survey will be sent out today or tomorrow. Anyone who completes it by Monday will get an extra percentage point on Exam 3.

Review: Distributions of Scores

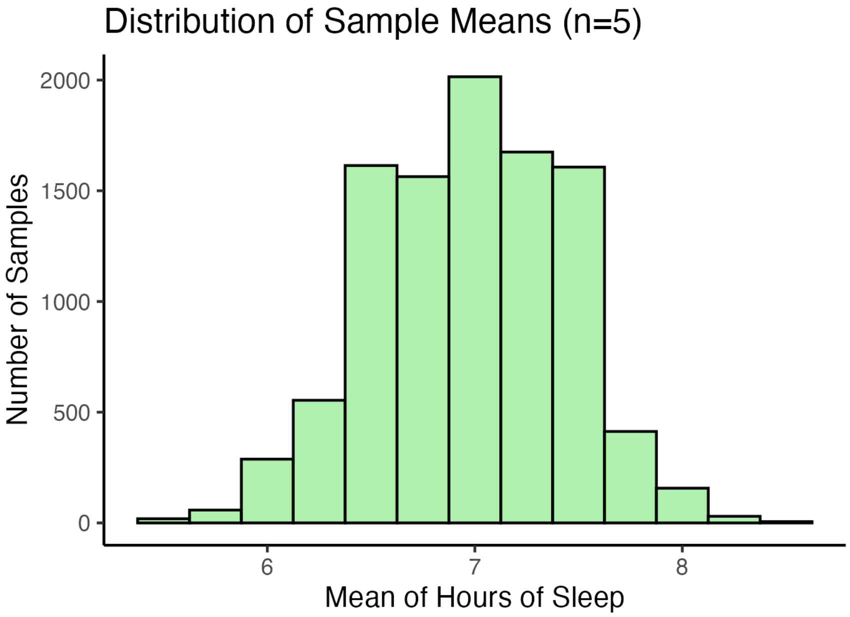
- Data in the world can be represented as **distributions of scores**.
- Examples: heights, test scores, reaction times, anxiety levels, hours of sleep per night
- We can summarize these distributions with statistics that describe their **central tendency** (e.g., mean) and **variability** (e.g., standard deviation).

Review: Distributions of Means

- In research, we often take **samples** from populations to make **inferences** about population parameters.
- For example, I measured the average number of hours of sleep per night for 5 randomly selected PS 211 students.
  - I found a sample mean of 6.4 hours, which means I can use as an estimate of the population mean.
- However, sample means vary from sample to sample due to **sampling error**. How *uncertain* should I be about my estimate?

Review: Distributions of Means (continued)

- To understand variability in our estimates of the sample mean, we can create a **distribution of means** by repeatedly sampling from the population and calculating the mean for each sample.



Review: Distributions of Means (continued)

- We normally don’t have access to the whole population to create a distribution of means.
- Usually we just have ONE sample mean.
- However, we can use our knowledge of sampling distributions to quantify our uncertainty about the sample mean using the **standard error** (SE).
- The **standard error (SE)** quantifies how much sample means vary from sample to sample.

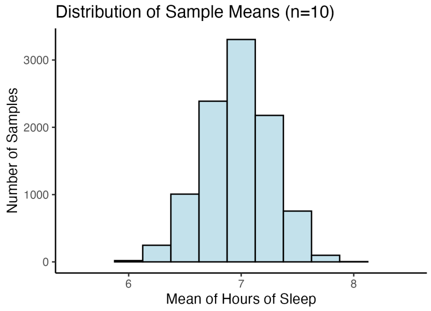
Review: Standard Error

Review: Why is quantifying our uncertainty about our sample mean important?

- In research, we often want to know if our sample mean is **significantly different** from a known or hypothesized population mean.
- To determine this, we can use different statistical tests, such as *z* tests or *t* tests.
- Fundamentally, when we run these tests, we are asking: **Is the difference between my sample mean and the population mean larger than what I would expect due to sampling error alone?**

Review: *z* Tests

- Sometimes, we *know* a population mean (μ) and standard deviation (σ).
- From this, we can compute the distribution of means we would expect if we took many samples of a given size (n) from that population, and we can directly compute the standard error.



- Then, we can compare our sample mean to this distribution using a *z* test, which tells us the probability of observing a sample mean as extreme as ours if that mean came from the population.

Review: *t* Tests

- Most of the time, we *do not know* the population standard deviation (σ).
- In this case, we estimate the standard error using the sample standard deviation (s) instead.
- However, this introduces additional uncertainty, especially with small sample sizes.
- To account for this, we use the ***t* distribution**, which is wider and has heavier tails than the *z* (normal) distribution.
- Then, we can compare our sample mean to this distribution using a *t* test, which tells us the probability of observing a sample mean as extreme as ours if that mean came from the population.

Review: *t* Distributions

Review: *t* Tests

- Normally, we don't **create** a *t* distribution from scratch.
- Instead, we use **predefined *t* distributions** based on degrees of freedom (df = n – 1).
- The shape of the *t* distribution changes with df:
  - Smaller df (smaller n) → wider distribution with heavier tails.
  - Larger df (larger n) → approaches normal distribution.
- Once we know our sample size, we can select the appropriate *t* distribution to compare our sample mean against.
- Like with a *z* test, we can determine critical values and make decisions about our hypotheses.
- Here, we our similarly asking, **Is the difference between my sample mean and the population mean larger than what I would expect due to sampling error alone?** *What is the probability of observing this sample mean if it came from the population?*

Types of *t* Tests

There are 3 types of *t* tests, used in different research scenarios:

1. **Single-sample *t* test** – Compare a sample mean to a population mean when the population SD is unknown.

   This is what we went over already!

2. **Paired-sample *t* test** – Compare two samples when every participant is in both samples (within-subjects design).

3. **Independent-samples *t* test** – Compare two samples when participants are in only one group (between-subjects design).

Between- vs Within-Subjects Designs

Within-Subjects

- Each participant experiences **all** levels of the independent variable.
- Comparisons are made *over time or conditions* for the same people.

What is an advantage of a within-subjects design?

Between-Subjects

- Each participant experiences **only one** level of the independent variable.
- Comparisons are made *between different people*.

What is an advantage of a between-subjects design?

Which *t* Test Should I Use?

- **Within-groups design** → use a *paired-samples t test*

- **Between-groups design** → use an *independent-samples t test*

💡 For the paired *t*, we must first create **difference scores** for each participant (e.g., after – before; Condition 1 - Condition 2).

What is a Paired-Samples *t* Test?

- Compares two means for a **within-groups** design in which each participant experiences both levels of an independent variable.
- Can also be used for **before-and-after** comparisons.

Imagine you want to measure "Stroop Interference" by testing participants' reaction times on a color-naming task where the words are either congruent or incongruent with the ink color. How would you design this study to test for differences in reaction times using a paired-samples *t* test?

Answer:

- Have each participant complete both the congruent and incongruent conditions of the Stroop task.
- Measure their reaction times for each condition.

Distributions of Mean *Differences*

Another distribution!

To start a paired-samples *t*, we must create a **distribution of mean differences** by subtracting the two scores for each participant and then sampling means of those differences repeatedly.

- Step 1: Randomly sample n participants.
- Step 2: Compute each participant's difference score (Condition 2 – Condition 1).
- Step 3: Compute the mean of these differences across all participants in a sample.

Example: Distribution of Mean Differences

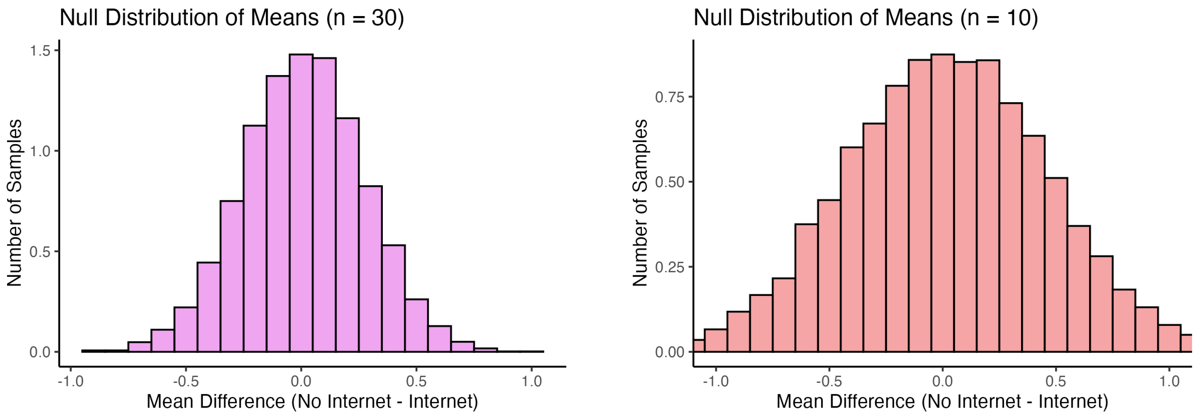
A researcher wants to determine whether students sleep more when internet access in their dorms is restricted at night. She has 30 students report their average nightly sleep duration for one week *with* internet access and one week *without* internet access.

To conduct a paired-samples *t* test, she needs to determine the probability that the observed mean difference in sleep duration occurred by chance, assuming no true difference in sleep duration across conditions.

To do this, she needs to create a **distribution of mean differences** under the null hypothesis:

- Assume no true difference in sleep duration across conditions.
- Randomly sample 30 (or n) students from this (simulated) population and compute the mean difference in sleep duration for each sample.
- Repeat this process many times to build a distribution of mean differences.

Example: Distribution of Mean Differences



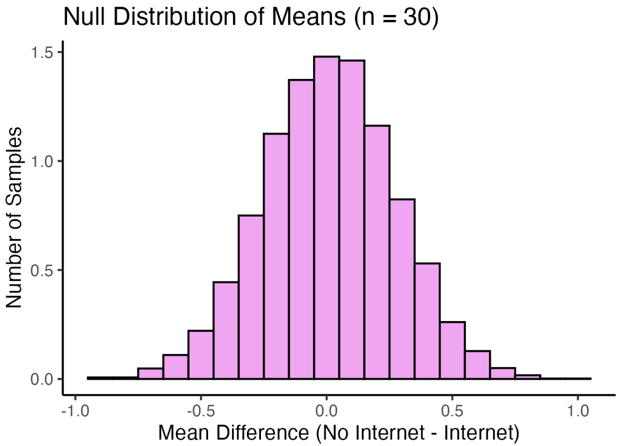
Will the shape of these null distributions always be normal?

Answer: Yes, if the sample size is sufficiently large. Due to the Central Limit Theorem, the distribution of the mean differences will approach normality as the sample size increases.

Conducting a Paired-Sample *t* Test

Under the **null hypothesis**, we assume there is no average difference in sleep duration across the population:

$$\mu_1 - \mu_2 = 0$$



- We can use this distribution of mean differences to determine the **probability of observing a mean difference as extreme as the one in our sample**, assuming the null hypothesis is true.
- If the researcher observes a mean difference in sleep duration of 0.5 hours in her sample of 30 students, she can calculate the *t* statistic to see how many standard errors this observed mean difference is away from the null hypothesis mean difference of 0.

Steps of a Paired-Sample *t* Test

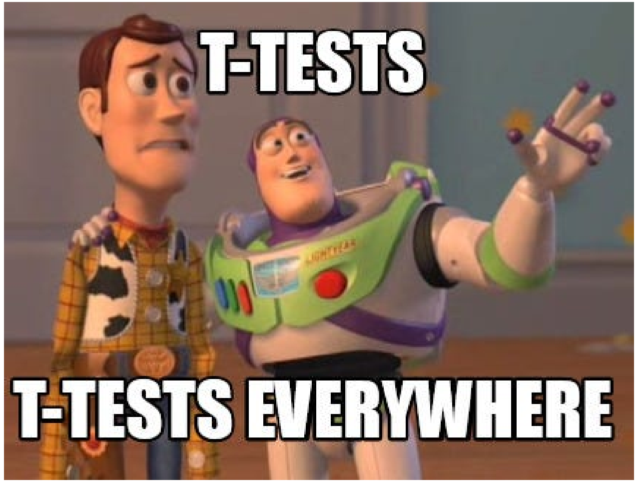
1. Identify populations, distribution, & assumptions.
2. State null and research hypotheses.
3. Determine characteristics of comparison distribution.
4. Determine critical values (cutoffs).
5. Calculate test statistic.
6. Make a decision.

Example: Beliefs About Violence & Mental Illness

A researcher wants to determine whether students sleep more when internet access in their dorms is restricted at night. She has 5 students report their average nightly sleep duration for one week *with* internet access and one week *without* internet access.

- Hours of sleep **with** internet access:
  - 6.5, 7.0, 6.0, 5.5, 6.0
- Hours of sleep **without** internet access:
  - 7.0, 7.5, 5.5, 6.0, 7.0

We'll test if mean hours of sleep differ across conditions using a **paired-samples *t* test.**



Step 1: Populations & Assumptions

- Group 1: Sleep reports with internet access.
- Group 2: Sleep reports without internet access.
- Distribution: Differences between paired scores.
- Assumptions:
  - DV is numeric.
  - Random sample of students.
  - Population distribution (including SD in average numbers of sleep) is unknown.

Step 2: Hypotheses

- **Null (H_0):** $\mu_1 = \mu_2$ (no mean difference)
  - or equivalently: $\mu_1 - \mu_2 = 0$
  - In words: Students sleep the same amount with or without internet access.
- **Research (H_1):** $\mu_1 \neq \mu_2$ (there is a difference)
  - or equivalently: $\mu_1 - \mu_2 \neq 0$
  - In words: Students sleep different amounts with vs. without internet access.

Step 3: Determine Characteristics of the Comparison Distribution

Under the null, the mean of the comparison distribution = 0. Because we don't know the population SD, we must estimate the standard deviation from the sample data, just like in a single-sample *t* test.

Participant	With Internet	Without Internet	Difference (D)	D – Mean D (0.4)	Squared Deviation
1	6.5	7.0	0.5	0.1	0.01
2	7.0	7.5	0.5	0.1	0.01
3	6.0	5.5	-0.5	-0.9	0.81
4	5.5	6.0	0.5	0.1	0.01
5	6.0	7.0	1.0	0.6	0.36

Step 3: Determine Characteristics of the Comparison Distribution

- Now, we can compute the standard error (SE) of the mean differences:

$$SE = \frac{s}{\sqrt{n}} = \frac{0.55}{\sqrt{5}} = 0.246$$

- Thus, our comparison distribution has:
  - Mean = 0
  - SE = 0.246

Step 4: Determine the Critical Values (or Cutoffs)

- df = N – 1 = 4
- The critical values are determined by the t-distribution with 4 degrees of freedom.
- We will use an alpha level of .05 for a two-tailed test.
- We can find out critical values from a t-table or computer program: → $t_{crit} = \pm 2.776$

Steps 5: Calculate the Test Statistic

- We calculate our t value using the formula:

$$t = \frac{M_{diff} - \mu_D}{SE}$$

where:

- M_{diff} = sample mean difference
- μ_D = hypothesized population mean difference (0 under H_0)
- SE = standard error of the mean differences

$$t = \frac{M_{diff} - 0}{SE} = \frac{0.4}{0.246} = 1.63$$

Steps 6: Make a Decision

- Compare calculated t to critical values:
  - Calculated $t = 1.63$
  - Critical values = $\pm$ 2.776

$1.63 < 2.776$ → **Fail to Reject H_0**

Confidence Intervals for Paired-Sample *t* Tests

- A **confidence interval (CI)** provides a range of values within which we are fairly certain the true population parameter lies.
- For a paired-samples *t* test, we can construct a CI around the sample mean difference to estimate the range of plausible values for the true mean difference in the population.
- A 95% CI means that if we were to take many samples and construct a CI for each sample, approximately 95% of those intervals would contain the true population mean difference.

Computing a 95% Confidence Interval for the Mean Difference

- To compute a 95% confidence interval for the mean difference, we can use the following formula:

$$CI = M_{diff} \pm t_{crit} \times SE$$

Why can we use this formula?

- This formula is similar to the one used for single-sample *t* tests, but here we are using the mean difference and the standard error of the mean differences.

- Plugging in our values:

$$CI = 0.4 \pm 2.776 \times 0.246$$

- This gives us:

Interpreting the Confidence Interval

- The sample mean (0.4) is centered within CI -0.283, 1.083.
- If we repeated this study many times, ≈ 95% of CIs would contain the true population mean difference.
- Because the CI includes 0, we cannot rule out the possibility that there is no true difference in sleep duration across conditions.

Beyond Hypothesis Testing: Effect Size

- The effect size indicates the magnitude of the difference between conditions in standard deviation units.
- For paired-samples *t* tests, we can use **Cohen's d** to quantify effect size:

$$d = \frac{M_{diff} - 0}{s}$$

where:

  - M_{diff} = sample mean difference
  - s = sample standard deviation of the differences

- Using our example:

$$d = \frac{0.4}{0.55} = 0.73$$

Beyond Hypothesis Testing: Effect Size

$$d = \frac{0.4}{0.55} = 0.73$$

- Interpretation:
- $d = 0.2$ → small effect.
- $d = 0.5$ → medium effect.
- $d = 0.8$ → large effect.

If this study were repeated with a larger sample size, we might find a significant (and meaningful) effect!

Practice: Paired-Samples t Test

Practice: Paired-Samples *t* Test (Solution)

1. Step 1: Populations & Assumptions

- Group 1: Quiz scores with music.
- Group 2: Quiz scores without music.
- Distribution: Differences between paired scores.
- Assumptions:
    - DV is numeric.
    - Random sample of students.
    - Population distribution (including SD in quiz scores) is unknown.

2. Step 2: Hypotheses

- **Null (H_0):** $\mu_1 = \mu_2$ (no mean difference)
    - or equivalently: $\mu_1 - \mu_2 = 0$
    - In words: Music does not affect quiz performance.
- **Research (H_1):** $\mu_1 \neq \mu_2$ (there is a difference)
    - or equivalently: $\mu_1 - \mu_2 \neq 0$
    - In words: Music affects quiz performance.

Practice: Paired-Samples *t* Test (Solution)

3. Step 3: Determine Characteristics of the Comparison Distribution

Participant	With Music	Without Music	Difference (D)	D – Mean D (-1)	Squared Deviation
1	7	8	-1	0	0
2	6	6	0	1	1
3	5	7	-2	-1	1
4	8	9	-1	0	0
5	7	8	-1	0	0
				Sum of Squares (SS)	**2.00**
				Sample Variance (s^2 = SS / 4)	**0.50**
				Sample SD ($\sqrt{s^2}$)	**0.71**

Practice: Paired-Samples *t* Test (Solution)

3. Step 3: Determine Characteristics of the Comparison Distribution (continued)

- Now, we can compute the standard error (SE) of the mean differences:

$$SE = \frac{s}{\sqrt{n}} = \frac{0.71}{\sqrt{5}} = 0.318$$

- Thus, our comparison distribution has:
  - Mean = 0
  - SE = 0.318

4. Step 4: Determine the Critical Values (or Cutoffs)

- df = N – 1 = 4
- The critical values are determined by the t-distribution with 4 degrees of freedom.
- We will use an alpha level of .05 for a two-tailed test.
- We can find out critical values from a t-table or computer program (and were told it here): → $t_{crit} = \pm 2.776$

Practice: Paired-Samples *t* Test (Solution)

5. Step 5: Calculate the Test Statistic

- We calculate our *t* value using the formula:

$$t = \frac{M_{diff} - \mu_D}{SE}$$

where:

- M_{diff} = sample mean difference
- μ_D = hypothesized population mean difference (0 under H_0)
- SE = standard error of the mean differences

$$t = \frac{M_{diff} - 0}{SE} = \frac{-1}{0.318} = -3.14$$

6. Step 6: Make a Decision

- Compare calculated *t* to critical values:
  - Calculated *t* = -3.14
  - Critical values = ± 2.776
  - -3.14 < -2.776 → **Reject H_0**

The results suggest that background music has a significant effect on quiz performance, with students performing *worse* when music is played during the quiz.

What if the researcher has used a between-subjects design?

The researcher realizes his study is flawed because students took the quiz twice, which may have influenced their performance. How could the researcher redesign the study to use a between-subjects design?

Answer:

- The researcher could randomly assign students to one of two groups: one group with music and another group without music.
- The researcher would then compare the quiz scores of the two independent groups using an independent-samples *t* test.

Independent-Samples *t* Tests

- Compares two means from **independent groups.**
- Each group has different people that each experience **only one** level of the independent variable.
- Example: Comparing quiz scores of students who took the quiz with music vs. students who took the quiz in silence
- The scores are independent because no student is in both groups. There are no paired scores.

Paired- vs. Independent-Samples *t* Tests

Paired-Samples *t* Test

1. Compute difference scores for each participant.

2. Compute mean of these difference scores.

3. Determine probability of observing this mean difference under null hypothesis

Here, our null distribution is a distribution of mean differences.

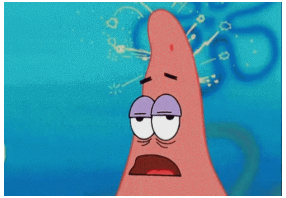
Independent-Samples *t* Test

1. Compute mean scores for each group.

2. Compute difference between these group means.

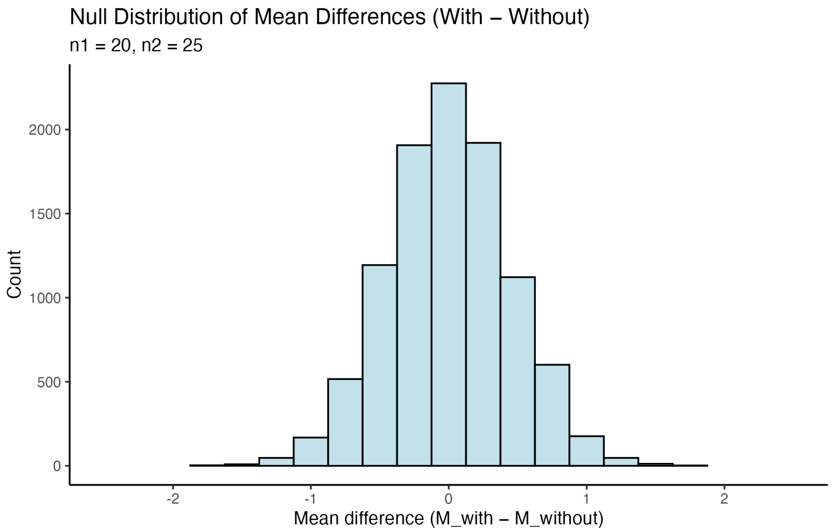
3. Determine probability of observing this mean difference under null hypothesis.

Here, our null distribution is a distribution of differences between independent group means.



How do we create the distribution of differences between independent group means?

1. Randomly sample n_1 participants for Group 1 and n_2 participants for Group 2 from the population.
2. Compute the mean for each group.
3. Compute the difference between the two group means ($M_1 - M_2$).
4. Repeat many times to build a distribution of mean differences between independent groups.
5. Use this distribution to determine the probability of observing a mean difference as extreme as the one in our sample, assuming the null hypothesis is true.



Note that here we have **two** ns (n_1 and n_2) because the two groups can have different sample sizes.

Steps of an Independent-Samples *t* Test

1. Identify populations, distribution, & assumptions.
2. State null and research hypotheses.
3. Determine characteristics of comparison distribution.
4. Determine critical values (cutoffs).
5. Calculate test statistic.
6. Make a decision.

Steps 3-5 are similar to those for paired-samples t tests, but the formulas differ slightly because we are dealing with two independent groups.

3. Determine characteristics of comparison distribution.
4. Determine critical values (cutoffs).
5. Calculate test statistic.

That's all for today!

We will continue with independent-samples *t* Tests next class!