BOSTON UNIVERSITY

bu.edu

PS 211: Introduction to Experimental Design

Fall 2025 · Section C1

Lecture 11: Independent-Sample *t* Tests

Updates & Reminders

- The mid-semester survey has been posted. Please take a few minutes to fill it out!
- If 38 people fill it out by next class, everyone will get 2 bonus percentage points on Exam 3.
- I'm already finding it super useful, so thank you to those who have filled it out!
  - One note: Answer keys for the homeworks *ARE* posted (on Blackboard and Slack) and are not taken down.
- Homework 3 will be posted by the end of the week and is due on Friday, Oct. 31.

Updates & Reminders

- Exam 2 grades have been posted.
- This exam was hard!
- We changed the denominator from 30 to 28 to account for the fact that several questions were overly confusing and challenging.
- We will go over the challenging questions today, in a few moments.

Exam 2: Why was it so hard?

Review: Hard questions from Exam 2

We are going over these outside of the slide deck.

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.

Before Exam 2!

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

Last class!

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

A brief aside: Where did *t* tests come from?

- *t* tests were developed by William Sealy Gosset in 1908 while working at the Guinness Brewery in Dublin, Ireland. Gosset wanted to monitor the quality of raw materials (e.g., barley) used in brewing, but he had only small samples to work with.
- To address this problem, he developed the *t* distribution and the *t* test, which allowed him to make inferences about population means from small samples.
- He published his work under the pseudonym "Student" to avoid conflicts with Guinness's policy against employees publishing research.

More contributions from Guinness Brewery!

- Stella Cunliffe, a statistician at Guinness in the 1950s, developed quality control methods that improved the consistency of beer production.
- At the time, workers had to either accept or reject barrels of beer by moving them up or down a hill.
- She used t testing to determine that *accepting* barrels was easier than *rejecting* them, leading to bias in quality control. She redesigned the process to reduce this bias.
- She became the first woman president of the Royal Statistical Society in 1975.

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.

   Before Exam 2!

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

   Last class!

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

   Today!

Between- vs Within-Subjects Designs

Within-Subjects / Within-Groups

- 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 / Between-Groups

- 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?

Within-Subjects Designs: Pros & Cons

Between-Subjects Designs: Pros & Cons

Advantages

- No carryover effects
- No order effects
- Sometimes the only ethical design (e.g., drug trials where participants cannot receive two different treatments)

Disadvantages

- More participants needed
- Individual differences can introduce variability

Which *t* Test Should I Use?

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

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.

- We know the mean of the comparison distribution is 0 under H_0. Now we need to compute the standard error (SE) of the mean differences.
- This requires estimating the population standard deviation (σ) of the difference scores.
- Our best estimate of the population standard deviation will be the *sample* standard deviation (s) of the difference scores.
- First, we compute the difference scores (With Music – Without Music).
- Then, we need to compute their mean. *We will also need this later!*
- Finally, we compute the sample standard deviation (s) of the difference scores.

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

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

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 and SE = 0.318
- Remember, this is the mean and SE of the *difference scores* we would expect if we drew many samples of 5 students from a population where music has no effect on quiz performance.

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

- Now, we need to determine where the middle 95% of this comparison distribution lies (since we are using an alpha level of .05 for a two-tailed test).
- Remember, we are using a *t* distribution to account for our additional uncertainty due to estimating the population standard deviation from our sample.
- We can find out critical values from a t-table or computer program for a two-tailed test with df = (n – 1) = 4.: → $t_{crit} = \pm 2.776$

Practice: Paired-Samples t Test (Solution)

Thinking critically: Paired-Samples *t* Test

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

Why can we conclude that we have a significant effect without computing a *p*-value?

A. We can't. This is a mistake.

B. Because our calculated *t* value falls in the critical region, we know that the *p*-value is less than .05.

C. Because our calculated *t* value is negative, we know that the *p*-value is less than .05.

D. Because our calculated *t* value is greater than 1, we know that our results are significant.

Thinking critically: Paired-Samples *t* Test

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

Given that our calculated mean differences is -1, and our *t* value is -3.14, which of the following are the most likely values for the 95% confidence interval?

A. -1, 1

B. -3.14, -1

C. -3.14, 3.14

D. -1.9, -0.1

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.

Can you think of another study design where you would use an independent-samples *t* test?

Paired- vs. Independent-Samples *t* Tests

Paired-Samples *t* Test

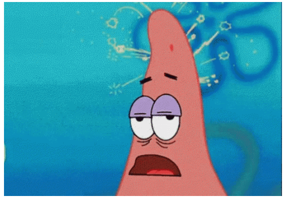
1. Compute difference scores for each participant.
2. Compute mean of these difference scores.
3. Determine probability of observing this mean difference under null hypothesis

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.

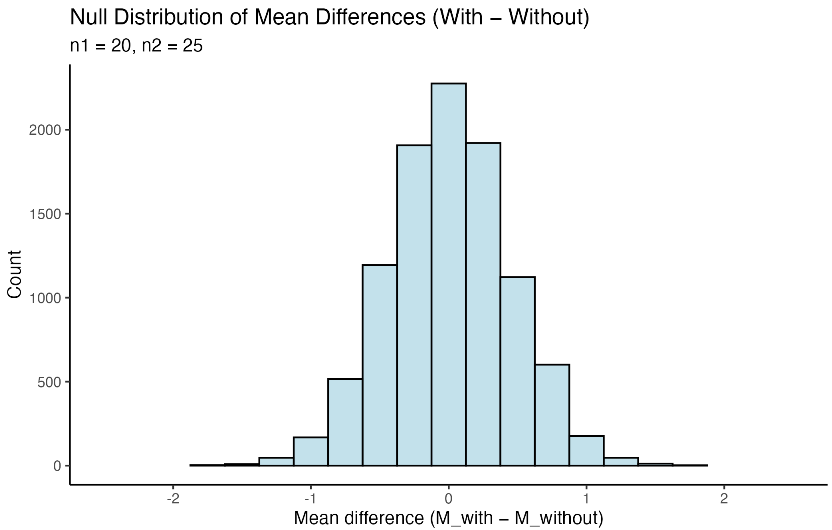
Here, our null distribution is a distribution of differences between independent group means.



Paired- vs. Independent-Samples *t* Tests

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.

What are the characteristics of the distribution of differences between independent group means?

Remember, **normally we do not have to build this distribution by simulating many samples.** Instead, we can compute its characteristics directly.

- Under H_0, the mean of this distribution is 0.
- The standard error (SE) of this distribution is computed using both groups' sample standard deviations and sample sizes.
- We use both groups' data to estimate the *pooled variance*, which is our best estimate of the population variance.
- From the pooled variance, we can compute the standard error of the difference between independent means.

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.

How do we compute the standard error for an independent-samples *t* test?

Five steps to compute the standard error:

1. Compute each sample's variance (s^2).
2. Compute the pooled variance (s^2pooled).
3. Convert the pooled variance from the squared standard deviation (SD^2) to the squared standard error (SE^2).
4. Add the two squared standard errors together to get the variance of the difference between means (SE^2difference).
5. Take the square root of the variance of the difference to get the standard error of the difference between means (SEdifference).

How do we compute the standard error for an independent-samples *t* test?

Step 1: calculate the variance for each sample:

$$s_1^2 = \frac{\sum_{i=1}^{n_1}(X_{1i} - M_1)^2}{n_1 - 1}$$

$$s_2^2 = \frac{\sum_{i=1}^{n_2}(X_{2i} - M_2)^2}{n_2 - 1}$$

Why do we divide by (n – 1) instead of n?

A. To correct for bias in estimating the population variance from a sample.

B. To account for the fact that we are working with two samples.

C. To account for different numbers of participants in each group.

D. There is no particular reason; it's just a convention.

How do we compute the standard error for an independent-samples *t* test?

Step 2: Compute the pooled variance:

- We have to weight our estimates because the sample sizes are different.
- To do this, we compute the degrees of freedom (df) for each sample:
  - $df_1 = n_1 - 1$
  - $df_2 = n_2 - 1$
- Then, we compute the pooled variance as a weighted average of the two sample variances:

$$s^2_{pooled} = \frac{df_1}{df_{total}} s^2_1 + \frac{df_2}{df_{total}} s^2_2$$

How do we compute the standard error for an independent-samples *t* test?

Step 2: Compute the pooled variance:

$$s^2_{pooled} = \frac{df_1}{df_{total}} s^2_1 + \frac{df_2}{df_{total}} s^2_2$$

If our first sample has 10 participants and our second sample has 15 participants, which variance will have more weight in the pooled variance calculation?

A. They will have equal weight.

B. The variance from the first sample ($n_1 = 10$).

C. The variance from the second sample ($n_2 = 15$).

D. We cannot tell without knowing the variances.

How do we compute the standard error for an independent-samples *t* test?

Step 3: Convert the pooled variance from the squared standard deviation (SD^2) to the squared standard error (SE^2).

- We now have our best estimate of the population variance (s^2pooled).
- Remember, variance = s^2.
- Remember, $SE = \frac{s}{\sqrt{n}}$.
- That means that $SE^2 = \frac{s^2}{n}$.
- So, for each sample, we convert the pooled variance to the squared standard error by dividing by the respective sample size.

How do we compute the standard error for an independent-samples *t* test?

Step 4: Add the squared standard errors (SE^2) together.

- We ultimately want the standard error of the *difference* between means.
- The squared standard error of the difference between means is the sum of the two squared standard errors.
  - This is a mathematical property of variances that is beyond the scope of this course.

$$SE^2_{difference} = SE^2_1 + SE^2_2$$

How do we compute the standard error for an independent-samples *t* test?

Step 5: Take the square root of the variance of the difference to get the standard error of the difference between means (SEdifference).

- Now that we have the squared standard error of the difference between means, we can take the square root to get the standard error.

$$SE_{difference} = \sqrt{SE_1^2 + SE_2^2}$$

How do we compute the standard error for an independent-samples *t* test?

Putting it all together:

1. Compute each sample's variance (s^2).
2. Compute the pooled variance (s^2pooled).
3. Convert the pooled variance from the squared standard deviation (SD^2) to the squared standard error (SE^2).
4. Add the two squared standard errors together to get the variance of the difference between means (SE^2difference).
5. Take the square root of the variance of the difference to get the standard error of the difference between means (SEdifference).

$$SE_{difference} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

Are you kidding me?

$$SE_{difference} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Here is the good news:

- In practice, you will almost never have to compute this by hand.
- It's important to understand the steps so that you understand *where* this formula comes from.
- But in practice, you will use statistical software (e.g., R) to compute independent-samples *t* tests for you.

Let's test our intuitions

If we increase the sample sizes (n_1 and n_2) while keeping the sample variances (s_1^2 and s_2^2) constant, what happens to the standard error of the difference between means (SEdifference)?

A. It increases.

B. It decreases.

C. It stays the same.

D. We cannot tell without knowing the sample sizes.

Answer: B. It decreases.

Why is this the case?

Answer: As sample sizes increase, our estimates of the population parameters become more precise, leading to a smaller standard error.

What are the characteristics of the distribution of differences between independent group means?

- Under H_0, the mean of this distribution is 0.
- We now know how to compute the standard error (SE) of this distribution using both groups′ sample standard deviations and sample sizes.

Step 4: Determine Critical Values (Cutoffs)

- We now need to determine where the middle 95% of this comparison distribution lies (since we are using an alpha level of .05 for a two-tailed test).
- We are using a *t* distribution to account for our additional uncertainty due to estimating the population standard deviations from our samples.
- We can find out critical values from a t-table or computer program.
- Here our degrees of freedom (df) is computed as:
  - $df = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$
- This is because we are estimating the population standard deviation twice (once for each sample).

Step 5: Compute the Test Statistic

- We now need to compute the test statistic (*t*), by subtracting the hypothesized population mean difference (0 under H_0) from the observed mean difference between our two samples, and dividing by the standard error of the difference between means.

$$t = \frac{M_D - \mu_D}{SE}$$

- Here:
  - M_D = observed mean difference between the two samples
  - μ_D = hypothesized population mean difference (0 under H_0)
  - SE = standard error of the difference between means
- This simplifies to:

Step 6: Make a Decision

- We use exactly the same decision rule as before:
  - If the calculated t value falls in the critical region (beyond the critical values), we reject H_0.
  - If the calculated t value does not fall in the critical region, we fail to reject H_0.

Confidence Intervals for Independent-Samples *t* Tests

- We use exactly the same calculation as before:

$$CI = (M_1 - M_2) \pm (t_{crit} \times SE_{difference})$$

Effect Sizes for Independent-Samples *t* Tests

- We use exactly the same calculation as before:

$$d = \frac{(M_1 - M_2)}{s}$$

Where s is the pooled standard deviation:

$$s = \sqrt{s^2_{pooled}}$$

Worksheet time!

BOSTON UNIVERSITY

- The shared worksheet walks you through the steps of an independent-samples *t* test.
- Work through it with a partner or small group.
- We will go over the solution together on Tuesday.

That's all for today!