

PS 211: Introduction to Experimental Design

Fall 2025 · Section C1

Exam 3 Review Session

Updates & Reminders

- Homework 3 answers and grades were posted.
 - Feel free to come to office hours to go over your homework.
- Exam 3 is on **Thursday, Nov. 6th** during our regular class time.
 - A review sheet has been posted.
 - You will *not* need a calculator.
- Office hours this week:
 - Today: 12:30 PM – 2:30 PM (Juneau)
 - Wednesday: 9:00 – 10:00 AM (Kate)

Review: Effect Size

- Effect sizes quantify the **magnitude** of an effect.
- For z and t tests, we use **Cohen's d** :

$$d = \frac{M - \mu}{SD}$$

Where:

- M = sample mean
- μ = population mean
- SD = standard deviation of the sample
 - For a z test, this is the population standard deviation (σ).
 - For an single-sample or paired-samples t test, this is the sample standard deviation (s).

Review: Effect Size (Continued)

- Effect sizes are often interpreted using Cohen's guidelines:
 - Small effect: $d = 0.2$
 - Medium effect: $d = 0.5$
 - Large effect: $d = 0.8$
- Effect sizes are useful because they are **standardized** and can be compared across studies.
- They tell us about the **practical significance** of our results.
 - If an effect is statistically significant but has a small effect size, it may not be *practically* important.

Review: Effect Size (Continued)

- Effect sizes are not influenced by sample size, unlike test statistics, p values, and confidence intervals.
- Effect sizes are not influenced by sample size because they quantify the magnitude of the effect relative to the variability in the data, rather than the precision of the estimate.
- This means that even with a small sample size, you can have a large effect size if the difference between groups is substantial relative to the variability.
- Conversely, with a large sample size, you can have a small effect size if the difference between groups is minimal relative to the variability.

Review: Effect Size (Continued)

Practice question: What would it mean if you found a large effect size but a non-significant p value?

- A. It could indicate that the sample size was too small to detect the effect statistically, even though the effect itself would be important in practice, if it were real.
- B. It could indicate that the effect is not practically significant, even though it is statistically significant.
- C. It could indicate that there was a calculation error, since effect sizes and p values always align.
- D. It could indicate that the null hypothesis is true.

Answer: A

- A large effect size with a non-significant p value often suggests that the sample size was too small to detect the effect statistically, even though the effect itself may be practically important.

Review: Statistical Power

- **Statistical power** is the probability of correctly rejecting the null hypothesis (i.e., detecting a true effect).
- Power is influenced by several factors:
 - Effect size: Larger effects are easier to detect, increasing power.
 - Sample size: Larger samples provide more information, increasing power.
 - Significance level (α): A higher α increases power but also increases the risk of Type I errors.
 - Variability in the data: Less variability increases power.
- A common target for power is 0.80, meaning there is an 80% chance of detecting a true effect if it exists.

Review: Statistical Power (Continued)

Practice question: Why do within-subjects designs generally have higher statistical power than between-subjects designs?

- A. Because they have larger sample sizes.
- B. Because they control for individual differences, reducing variability.
- C. Because they use a higher significance level (α).
- D. Because they always have larger effect sizes.

Answer: B

- Within-subjects designs control for individual differences by using the same participants in all conditions, which reduces variability and increases statistical power.

Review: Statistical Power (Continued)

Practice question: Why does running a one-tailed test instead of a two-tailed test increase statistical power?

- A. Because it increases the effect size.
- B. Because it allocates all of the alpha level to one tail of the distribution, making it easier to reach significance in that direction.
- C. Because it increases the sample size.
- D. Because it reduces variability in the data.

Answer: B

- Running a one-tailed test allocates all of the alpha level to one tail of the distribution, making it easier to reach significance in that direction.

Review: APA Style Reporting

Practice question: Which of the following is the correct way to report the results of a single-sample t test in APA style? Imagine the first part of the sentence read, "A single-sample t test revealed that the sample mean ($M = 15.2$, $SD = 4.5$) was significantly higher than the population mean ($\mu = 12$), "

- A. $t(29) = 2.45$, $p < .05$, $d = 0.45$
- B. $t = 2.45$, $df = 29$, $p = .03$, $d = 0.45$
- C. $t(29) = 2.45$, $p = .03$, $d = 0.45$
- D. $t(29) = 2.45$, $p = .03$, Cohen's $d = 0.45$

Answer: C

- The correct APA style reporting includes the test statistic, degrees of freedom in parentheses, the p value, and the effect size. Optionally, you can also include the confidence interval.

Review: APA Style Reporting

Practice question: What's wrong with the following APA style reporting of a paired-samples t test? "A paired-samples t test indicated that the mean score on the post-test was significantly higher than the mean score on the pre-test, $t(24) = 3.12$, $p = .004$, $d = 0.62$, 95% CI [1.2, 4.5]."

- A. The means and standard deviations should be reported.
- B. The degrees of freedom should not be included.
- C. The p value should only be reported to two decimal places.
- D. There is nothing wrong with this reporting.

Answer: A. The means and standard deviations for both the pre-test and post-test should be reported to provide context for the results.

Review: t Tests

- The t test is used to compare means when the population standard deviation is unknown.
- Types of t tests:
 - Single-sample t test: Compares a sample mean to a known population mean.
 - Paired-samples t test: Compares means from the same group at different times or under different conditions.
 - Independent-samples t test: Compares means from two different groups.
- Key assumptions of t tests:
 - The data are numeric.
 - The data are collected from a random sample.
 - The observations are independent (for independent-samples t tests).

Review: t Test Formulas

$$t = \frac{M - \mu}{SE}$$

Where:

- M is the sample mean
- μ is the population mean
 - In a single-sample t test, this is a known value.
 - In a paired-samples t test, this is the mean of the difference scores under the null hypothesis (which is 0).
 - In an independent-samples t test, this is the difference between the two sample means under the null hypothesis (which is 0).

Review: t Test Formulas

Single-Sample t Test

$$t = \frac{M - \mu}{SE}$$

Paired-Samples t Test

$$t = \frac{M_D}{SE_{Difference}}$$

Independent-Samples t Test

$$t = \frac{M_1 - M_2}{SE_{Difference}}$$

- Can think of all these equations as representing a ratio of "signal" (the difference between the sample mean and population mean) to "noise" (the variability in the data, represented by the standard error).

Review: t Test Formulas

Single-Sample t Test

$$t = \frac{M - \mu}{SE}$$

Paired-Samples t Test

$$t = \frac{M_D}{SE_{Difference}}$$

Independent-Samples t Test

$$t = \frac{M_1 - M_2}{SE_{Difference}}$$

Practice question: What happened to the population mean (μ) in the paired-samples and independent-samples t test formulas?

Answer: The population mean (μ) is replaced by the mean of the difference scores (for paired-samples) or the difference between the two sample means (for independent-samples) under the null hypothesis, which is assumed to be 0.

Review: t Test Formulas

$$t = \frac{M - \mu}{SE}$$

Practice question: Which of the following changes would increase the value of the t statistic?

- A. Increase the difference between the sample mean and population mean of the null hypothesis.
- B. Increase the standard error.
- C. Decrease the sample size.
- D. Increase the variability in the data.

Answer: A

- Increasing the difference between the sample mean and population mean of the null hypothesis increases the numerator of the t statistic, leading to a larger t value.

Review: t Tests

Practice question: If you run a t test and do not observe a significant effect, in which cases would you expect your confidence interval to include 0?

- A. Paired-samples t test, independent-samples t test, and single-sample t test.
- B. Paired-samples t test only.
- C. Independent-samples t test only.
- D. Paired -samples t test and independent-samples t test only.

Answer: D

- The confidence interval for the mean difference in paired-samples t tests and the difference between means in independent-samples t tests is constructed around the observed difference. If the observed difference is not statistically significant, the confidence interval is likely to include 0, indicating no difference. In a single-sample t test, the confidence interval is constructed around the sample mean, so it may not include 0 even if the result is not significant.

Review: Paired-Samples t Tests

Practice question: In which of the following cases would you use a paired-samples t test?

- A. When comparing heights of children and adults.
- B. When comparing test scores of the same students at three different time points.
- C. When comparing blood pressure before and after administering a medication to the same group of patients.
- D. When comparing how number of hours spent studying relates to the number of hours spent sleeping.

Answer: C

- A paired-samples t test is appropriate when comparing measurements taken from the same group of participants under two different conditions or at two different times, such as before and after administering a medication.

Review: Paired-Samples t Tests

Practice question: What type of distribution is used to determine the critical values for a paired-samples t test?

- A. Distribution of sample means
- B. Distribution of mean differences.
- C. Distribution of individual scores.
- D. Distribution of differences between independent observations.

Answer: B

- The distribution of mean differences is used to determine the critical values for a paired-samples t test, as it focuses on the differences between paired observations.

Review: Paired-Samples t Tests: The Comparison Distribution

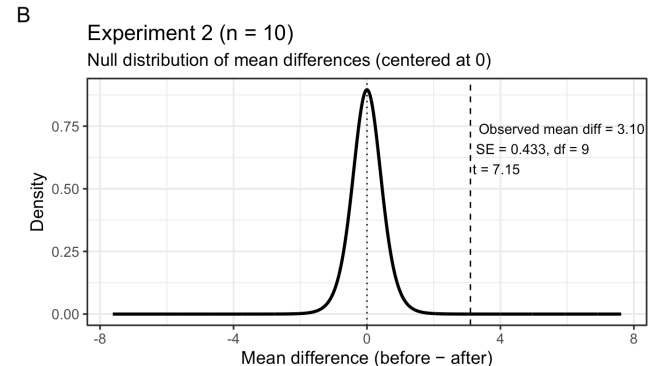
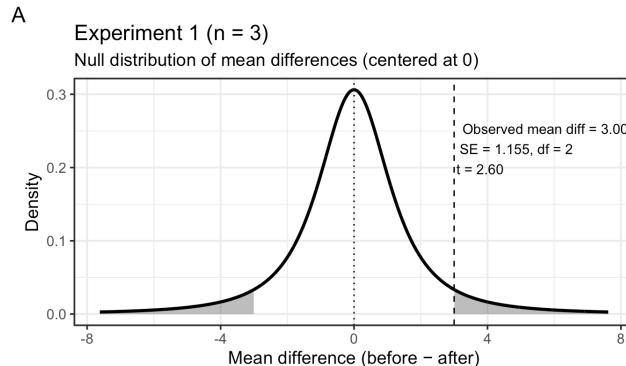
- When we run a paired-samples t test, we are comparing our distribution of sample mean differences to a comparison distribution.
- The comparison distribution is the distribution of mean differences that we would expect to see if the null hypothesis were true.
 - Because this distribution reflects the **null hypothesis**, **its mean will *always* be 0**.
 - However, the spread of this distribution (i.e., its standard error) will depend on the variability in our data and our sample size.
 - Why? Because even if the null hypothesis were true, **we would still see some variability in the mean differences due to random sampling error**.
 - The amount of variability we would expect to see in the mean differences is captured by the standard error of the difference scores, which takes into account both the variability in the difference scores and the sample size.

Review: Paired-Samples t Tests: The Comparison Distribution

- Imagine that we want to determine whether students run faster after drinking a cup of coffee.
- Now imagine that the null hypothesis is true: **drinking coffee has no effect on run times.**
- If we were to conduct our study with three students, we might still observe some differences in their run times before and after drinking coffee, just due to random chance.
- If student run times were highly variable, we would expect to see even larger differences in their run times before and after drinking coffee, again just due to random chance.
- However, if we increased our sample size to 300 students, we would expect the differences in run times to become more consistent and less influenced by random chance.
- This is why the standard error of our comparison distribution depends on both our sample size and the variability of the data.

Review: Paired-Samples t Tests: The Comparison Distribution

- Our t statistic tells us where (how many standard errors away from mean) in this comparison distribution our observed mean difference falls.
- A larger absolute value of t indicates that our observed mean difference is further away from the mean of the comparison distribution.
- If our t statistic is large enough (i.e., falls in the critical region), we can reject the null hypothesis and conclude that there is a significant difference between the two conditions.



Review: Paired-Samples t Tests: The Comparison Distribution

Practice question: Why are we more likely to find a significant effect with a larger sample size?

- A. Larger sample sizes will both increase our t statistic AND change the shape of the t distribution, decreasing the critical value needed to reach significance.
- B. Larger sample sizes will increase our t statistic, but will not change the shape of the t distribution.
- C. Larger sample sizes will change the shape of the t distribution, decreasing the critical value needed to reach significance, but will not affect our t statistic.
- D. Larger sample sizes will increase our effect size.

Answer: A

- Larger sample sizes decrease the standard error, which increases the t statistic. Additionally, larger sample sizes lead to a t distribution that more closely resembles the normal distribution, which has lower critical values for significance.

Review: Independent-Samples t Tests: The Comparison Distribution

- Independent-samples t tests compare means from two different groups.
- The comparison distribution for an independent-samples t test is the distribution of differences between sample means that we would expect to see if the null hypothesis were true.
- Because this distribution reflects the **null hypothesis**, **its mean will *always* be 0**.
- However, the spread of this distribution (i.e., its standard error) will depend on the variability in our data and our sample sizes in both groups.
- The standard error of the difference between means takes into account both the variability in each group and the sample sizes of each group.

Review: 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^2_{pooled}).
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 ($SE^2_{\text{difference}}$).
5. Take the square root of the variance of the difference to get the standard error of the difference between means ($SE_{\text{difference}}$).

Review: Independent-Samples t Tests

Practice question: Which of the following is NOT a computation involved in conducting an independent-samples t test?

- A. Compute each sample's variance (s^2).
- B. Compute the pooled variance (s^2_{pooled}).
- C. Compute the mean of the two sample means.
- D. Convert the pooled variance from the squared standard deviation (SD^2) to the squared standard error (SE^2).

Answer: C

- There is no reason to compute the mean of the two sample means when conducting an independent-samples t test. The focus is on the difference between the two means, not their average.

Review: Independent-Samples t Tests

Practice question: Why do we compute the pooled variance in an independent-samples t test?

- A. We have two different samples, so we need to combine their variances to get a better estimate of the population variance.
- B. This enables us to calculate the difference between the sample means.
- C. We need the pooled variance so we know which sample has more variability.
- D. There is no reason to compute the mean of the two sample means when conducting an independent-samples t test.

Answer: A

- In an independent-samples t test, we compute the pooled variance to combine the variances from both samples, providing a more accurate estimate of the population variance when the two groups are assumed to have equal variances.

Review: Independent-Samples t Tests

Practice question: In which of these scenarios would you use an independent-samples t test?

- A. You want to know if there is a difference in SAT scores between students at BU, BC, Northeastern, and Harvard.
- B. You want to know if the number of shark attacks each month relates to the average monthly temperature.
- C. You want to know if student responses on an ordinal scale measuring satisfaction with library services (e.g., strongly agree to strongly disagree) differs between undergraduate and graduate students.
- D. None of the above.

Answer: D

- A involves more than two independent groups.
- B involves two *continuous* variables, not two groups.
- C involves an ordinal variable, which violates a key assumption of the t test: that the data are numeric.

Review: Independent-Samples t Tests

Practice question: If you decide to run a one-tailed independent-samples t test instead of a two-tailed test, which of the following is TRUE?

- A. You will need a larger t statistic to reach significance.
- B. You will need a smaller t statistic to reach significance.
- C. The shape of the t distribution will change.
- D. None of the above.

Answer: B

- In a one-tailed test, all of the alpha level is allocated to one tail of the distribution, making it easier to reach significance in that direction, which means you will need a smaller t statistic to reach significance. For example, if you are using an alpha level of .05, the critical value for a two-tailed test would be split between both tails (i.e., .025 in each tail), whereas in a one-tailed test, the entire .05 would be in one tail. This means

That's all for today!

Good luck on Exam 3!