BOSTON UNIVERSITY

PS 211: Introduction to Experimental Design

Fall 2025 · Section C1

Exam 2 Review Session

Updates and Reminders

- Homework 2 is due Friday.
  - We are going to try to return grades quickly so you can use feedback to prepare for Exam 2.
- Please submit as a PDF.

Updates and Reminders (Exam 2)

Exam 2 is next Thursday.

- The exam will focus on lectures 6 - 9, but may also include some cumulative content from earlier in the course.
- The exam will consist of 31 multiple choice questions. You will only need to answer 30 questions correctly to get 100%.
- You will not need a calculator.
- You can bring one 8.5"x11" sheet of handwritten notes (front and back).
- If you need to use a z table or t table, we will provide the table.
- Please bring a pencil or dark pen.
- Having your computer or phone out during the exam will result in a 0 for the exam.

Example: Counseling Session Contracts

- We are studying counseling sessions attended by students at a university. We want to know if signing a contract to attend counseling improves attendance.
- We ask students to sign contracts to attend a set number of counseling sessions (10)
  - Sample: Students at this counseling center who sign the contract to attend at least 10 sessions
  - Population: All students who attended counseling sessions at this university and did not sign the contract

- We sample 5 students who sign a contract. They attended 6, 6, 12, 7, and 8 counseling sessions
- The university average of students who did not sign contracts is 4.6 counseling sessions attended.

Did students who sign the contract attend a different number of sessions than those who did not?

Example: Counseling Session Contracts

Step 1: Identify the populations, distributions, and assumptions **Populations:** Contract sample, non-contract population

Distributions: Distributions of *means*. We want to know whether a sample mean is different from a population mean. Our population distribution is a *t* distribution because we do not know the population standard deviation, so we will have to estimate it from the sample. This means we will use a *t* test.

Assumptions met for a *t* test?:

- We don't know anything about the population distribution or how the sample was collected.
- However, we will assume that the sample was collected randomly and that the population is approximately normal.
- If the population was not normal, we would need a larger sample size to use a *t* test. (so that the Central Limit Theorem applies)

Example: Counseling Session Contracts

Step 2: State the hypotheses

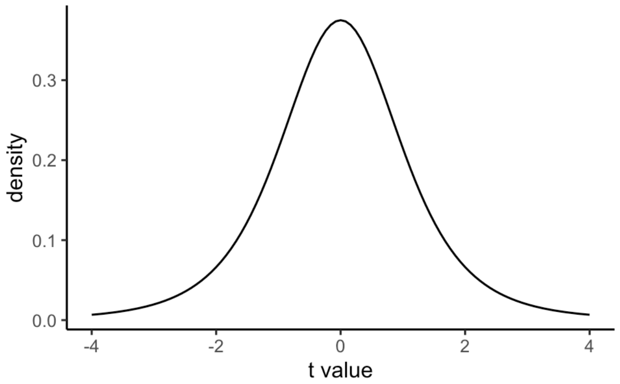
- H_0: $\mu_1 = \mu_2$ (Students who sign the contract attend the same number of sessions as those who do not)
- H_1: $\mu_1 \neq \mu_2$ (Students who sign the contract attend a different number of sessions than those who do not)

Step 3: Determine characteristics of the **comparison distribution**

- Here, we how "extreme" our sample mean is, *assuming the null hypothesis is true*.
- We have already decided that the distribution is a *t* distribution because we know the population mean but not the population standard deviation.
- So we need to compute the *t* statistic for our sample mean.

Example: Counseling Session Contracts

Example: Counseling Session Contracts



Now we have our *t* distribution for df = 4.

Step 4: Determine critical values or cutoffs

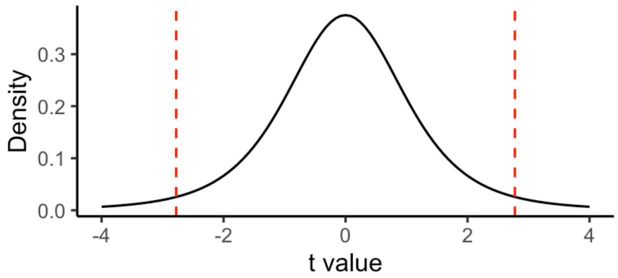
- We need to know how extreme our *t* statistic needs to be to reject the null hypothesis.
- Remember, *t* distributions vary based on degrees of freedom (df = n - 1). Here, df = 5 - 1 = 4.
- We will use α = 0.05 and a two-tailed test (because our alternative hypothesis is that the means are different, not specifically higher or lower).

Example: Counseling Session Contracts

Example: Counseling Session Contracts

Step 4: Determine critical values or cutoffs

- How do we convert α = 0.05 into a critical value for *t*?
- We can use a *t* table or an online calculator to find the critical value for df = 4 and α = 0.05 (two-tailed).



df:	One-Tailed Tests: 0.10	One-Tailed Tests: 0.05	One-Tailed Tests: 0.01	Two-Tailed Tests: 0.10	Two-Tailed Tests: 0.05	Two-Tailed Tests: 0.01
1	3.078	6.314	31.821	6.314	12.706	63.657
2	1.886	2.920	6.965	2.920	4.303	9.925
3	1.638	2.353	4.541	2.353	3.182	5.841
4	1.533	2.132	3.747	2.132	2.776	4.604
5	1.476	2.015	3.365	2.015	2.571	4.032

- The critical value is approximately ±2.776.

Example: Counseling Session Contracts

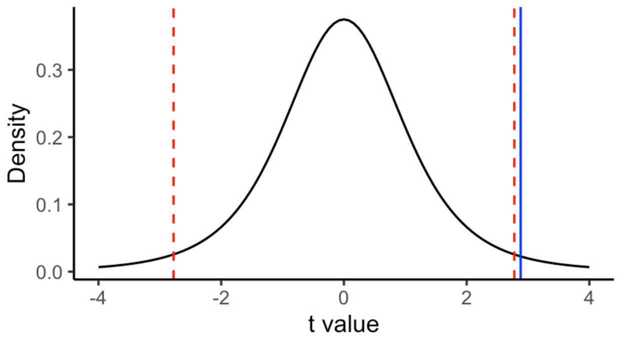
Step 5: Calculate the test statistic

- We can use the formula for the *t* statistic:

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

- Plugging in our values:

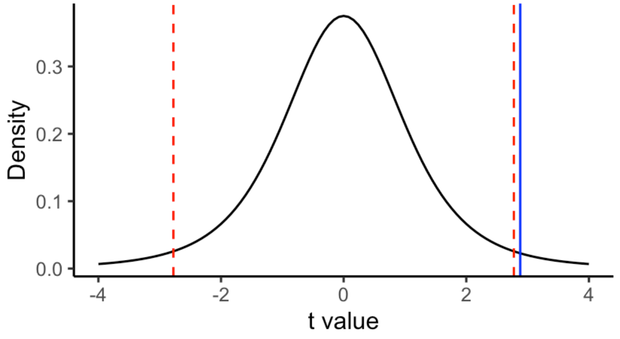
$$t = \frac{7.8 - 4.6}{1.11} = 2.88$$



Example: Counseling Session Contracts

Step 6: Make a decision

- Our calculated *t* statistic is 2.88.
- Our critical values are ±2.776.
- Since 2.88 is greater than 2.776, we reject the null hypothesis.



This means that if there is no difference in attendance between students who sign the contract and those who do not, we would expect to get a t statistic this extreme or more extreme less than 5% of the time. So we reject the null hypothesis, and instead we conclude that there is a significant difference in attendance.

Review: Degrees of freedom

- To use a t table and run a t test, we need to determine our **degrees of freedom**.

Degrees of freedom = number of scores that are free to vary when we estimate a population parameter from a sample.

- Degrees of freedom reflect the amount of **independent information** available.
- More degrees of freedom means more independent information, which means a more accurate estimate of the population parameter.

Degrees of freedom for a single-sample t test

df = n-1

- Here, we are estimating the population *standard deviation* from our sample data.
- We *know* the mean.
- Our degrees of freedom reflect the number of scores that could vary (amount of independent information we have) when a given parameter is known.
- Because our mean is known, that means all the scores in our dataset could vary, except one. Once we know the values of the first n-1 scores, the last score MUST take on a specific value.

Example:

- If we have 5 scores that sum to 40 (mean = 8), and we know the first 4 scores are 6, 8, 10, and 4, what must the last score be?
- The last score must be 12, because 6 + 8 + 10 + 4 + 12 = 40.
- So, with 5 scores, we have 4 degrees of freedom (5 - 1 = 4).

Review: Z Scores and Z Tests

- A z score tells us how many standard deviations a score is from the mean.
- Formula: $z = \frac{X - \mu}{\sigma}$
  - X = score
  - μ = population mean
  - σ = population standard deviation

- A z test compares a sample mean to a population mean, when we know the population standard deviation.
- Formula: $z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$
  - $\bar{X}$ = sample mean
  - μ = population mean
  - σ = population standard deviation
  - n = sample size

Remember, when we are dealing with distributions of means, we use the standard error ($\frac{\sigma}{\sqrt{n}}$*).*

Review: Raw Scores, Z Scores, and Percentiles

- A raw score is the original score (e.g., 85 on a test).
- A *z* score tells us how many standard deviations a score is from the mean.
- A percentile tells us the percentage of scores that fall below a given score.
- We can convert between raw scores, z scores, and percentiles using the mean and standard deviation of the distribution.

Moving between raw scores and z scores

- To convert a raw score to a z score:

$$z = \frac{X - \mu}{\sigma}$$

- To convert a z score to a raw score, we do some algebra to solve for X!

$$z = \frac{X - \mu}{\sigma}$$

$$z * \sigma = X - \mu$$

$$z * \sigma + \mu = X$$

Moving between raw scores and z scores

- Let's practice. The average grade on Homework 1 was 90 with a standard deviation of 5. Your grade was 85. What is your z score?

- To convert a raw score to a z score:

$$z = \frac{X - \mu}{\sigma}$$

$$z = \frac{85 - 90}{5} = -1$$

You scored 1 standard deviation below the mean.

Moving between raw scores and z scores

- Let's practice. The average grade on Homework 1 was 90 with a standard deviation of 5. Your friend scored .5 standard deviations above the mean ($z = .5$). What was their raw score?

- To convert a *z* score to a raw score:

$$z * \sigma + \mu = X$$

$$X = 0.5 * 5 + 90 = 92.5$$

Your friend scored 92.5 on Homework 1.

Moving between raw scores and z scores for means

- To convert a sample mean to a z score:

$$z = \frac{M - \mu}{\sigma_M}$$

Where σ_M is the standard error: $\sigma_M = \frac{\sigma}{\sqrt{n}}$

- To convert a z score to sample mean, we do some algebra to solve for M!

$$z = \frac{X - \mu}{\sigma_M}$$

$$z * \sigma_M = M - \mu$$

$$z * \sigma_M + \mu = M$$

Moving between raw scores and z scores for means: Practice

- Let's practice. The average grade on Homework 1 was 90 with a standard deviation of 5. A sample of 25 students had a mean score of 85. What is the z score for this sample mean?

- First, we need to find the standard error:

$$\sigma_M = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{25}} = 1$$

- Now we can find the z score:

$$z = \frac{M - \mu}{\sigma_M} = \frac{85 - 90}{1} = -5$$

This sample of students scored 5 standard errors below the mean.

Moving between raw scores and z scores for means: Practice

- The average grade on Homework 1 was 90 with a standard deviation of 5. Nine students used chatGPT to help with their homework. Their mean score was 2 standard errors above the class mean. What was their mean score?

- First, we need to find the standard error:

$$\sigma_M = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{9}} \approx 1.67$$

- Now we can find their mean score:

$$M = z * \sigma_M + \mu = 2 * 1.67 + 90 \approx 93.34$$

Moving between raw scores, z scores, and percentiles for means

- To convert between z scores and percentiles, we can use a *z* table or computer program.
- This will enable us to find the percentage of scores below a given z score, or the z score that corresponds to a given percentile.
- We can also use this to find critical values for hypothesis tests or confidence intervals.
- We can then convert between z scores and raw scores using the formulas from the prior slides.

Moving between raw scores, z scores, and percentiles for means: Practice

- Is there a significant difference between the mean score of the 9 students who used chatGPT ($M = 93.34$) and the class mean ($\mu = 90$, $\sigma = 5$)? Use $\alpha = .05$.
- We already calculated the z score for this sample mean:

$$z = \frac{M - \mu}{\sigma_M} = \frac{93.34 - 90}{1.67} \approx 2$$

- Now we need to find the percentage of scores that fall below this z score. We can use a z table or computer program to find this.
- Using a z table, we find that the percentage of scores below a z score of 2 is approximately 97.72%.

With $\alpha = .05$, we would reject the null hypothesis if the percentage of scores below our z score is less than 2.5% or greater than 97.5%. Since 97.72% is greater than 97.5%, we reject the null hypothesis.

Finding confidence intervals for means

What is the 95% confidence interval for the mean score of the 9 students who used chatGPT (M = 93.34)?

- First, we can imagine that this **sample mean** lies in the center of a **sampling distribution of means**.
- Because we know the population standard deviation, we can compute the standard error of the mean directly and use a z distribution.
- To find the 95% confidence interval, we need to find the z scores that correspond to the middle 95% of the distribution.
- Second, we need to determine what percentiles correspond to the middle 95%. This means we need to leave 2.5% in each tail.
- Our lower percentile is 2.5% and our upper percentile is 97.5%.

Finding confidence intervals for means (continued)

- Using a z table, we find that the z score that corresponds to 2.5% is approximately -1.96 and the z score that corresponds to 97.5% is approximately 1.96.
- Now we can convert these z scores to raw scores using the formula: $M = z * \sigma_M + \mu$
- For the lower bound:

$$M = -1.96 * 1.67 + 93.34 \approx 90.06$$

- For the upper bound:

$$M = 1.96 * 1.67 + 93.34 \approx 96.62$$

Effect sizes: Building an intuition

- The average grade on Homework 1 was 90 with a standard deviation of 5. One hundred students used chatGPT to help with their homework. Their mean score was 2 standard errors above the class mean. What was their mean score?

- First, we need to find the standard error:

$$\sigma_M = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{100}} \approx 0.5$$

- Now we can find their mean score:

$$M = z * \sigma_M + \mu = 2 * 0.5 + 90 \approx 91$$

Effect sizes: Building an intuition

- Is there a significant difference between the mean score of the 100 students who used chatGPT (M = 91) and the class mean ($\mu = 90$, $\sigma = 5$)? Use $\alpha = .05$.

- We already calculated the z score for this sample mean:

$$z = \frac{M - \mu}{\sigma_M} = \frac{91 - 90}{0.5} = 2$$

- Now we need to find the percentage of scores that fall below this z score. We can use a z table or computer program to find this.

- Using a z table, we find that the percentage of scores below a z score of 2 is approximately 97.72%.

With $\alpha = .05$, we would reject the null hypothesis if the percentage of scores below our z score is less than 2.5% or greater than 97.5%. Since 97.72% is greater than 97.5%, we reject the null hypothesis and conclude there is a *significant* difference.

Effect sizes: Building an intuition

- We know there is a *significant* difference between the mean score of the 100 students who used chatGPT ($M = 91$) and the class mean ($\mu = 90$, $\sigma = 5$).
- But is this a **meaningful** difference?
- To answer this, we can consider the *size* of the difference between the two means, while *ignoring* sample size.
- Remember, with a large enough sample size, even a tiny difference can be statistically significant.
- The sample size affects the standard error, which affects the *z* score, which affects the *p* value.
- But the sample size does not affect the **actual** difference between the two means ($M - \mu$).
- So, is a difference of 1 point (91 - 90) meaningful in this context?

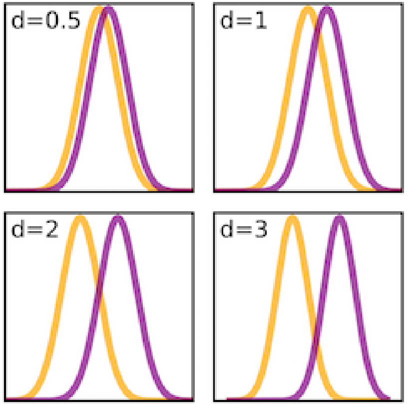
Effect sizes: Computing Cohen’s d

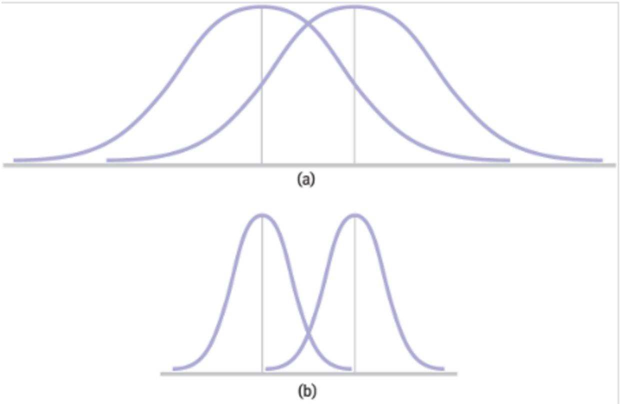
- Sometimes, we can use our intuition to determine if a difference is meaningful.
- But other times, we may want a more objective measure of the size of the difference.
- We can use an effect size measure called Cohen’s d to quantify the size of the difference between two means.
- Cohen’s d is calculated as the difference between two means divided by the standard deviation.

$$d = \frac{M - \mu}{\sigma} = \frac{91 - 90}{5} = 0.2$$

Cohen's d of 0.2 is considered a small effect size. This suggests that while the difference between the two means is statistically significant, it may not be practically meaningful.

Effect size tells us how much two populations do not overlap





Nolan/Heinzen, *Statistics for the Behavioral Sciences*, 5e, © 2020 Worth Publishers

Overlap can be decreased in two ways:

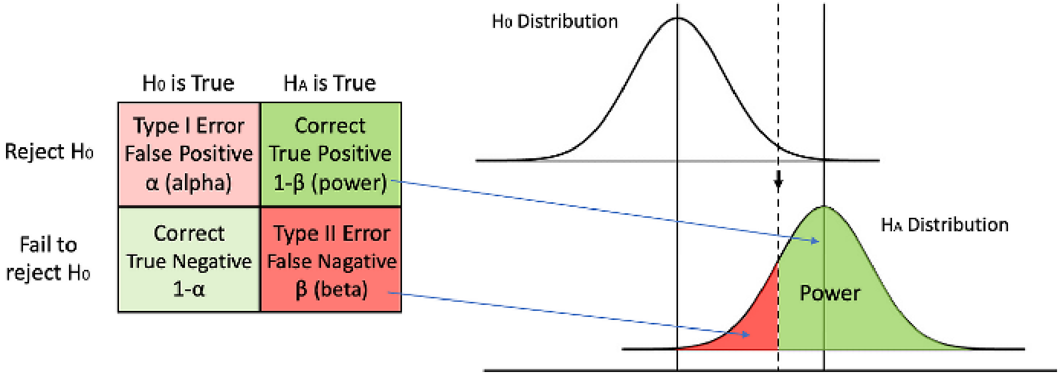
1. When two population means are far apart, the overlap of the distributions is less and the effect size is bigger.

2. When variability within each distribution is smaller, overlap decreases and effect size increases.

Statistical Power

- **Statistical Power** = probability of correctly rejecting H_0 when it's false (avoiding a Type II error).
- In other words, power is the likelihood we will reject the null hypothesis *when we should.*
- Ranges from probability of 0.00 to probability of 1.00
- Probability of 0.80 (80%) is the conventional goal.
  - Many studies in psychology are underpowered!

Statistical Power (Continued)



- When testing hypotheses, there are two ways we can be correct and two ways we can be wrong:

1. Correctly rejecting the null hypothesis (true positive)
2. Correctly failing to reject the null hypothesis (true negative)
3. Incorrectly rejecting the null hypothesis (Type I error)

Five factors that influence power

Power increases when:

1. Alpha **increases**

- This is usually not a good idea: This is like changing the rules of a basketball game by shortening the basket height, or widening the goalposts in football or soccer

- Increasing the alpha level from 0.05 to 0.1 increases the probability of type I error from 5% to 10%!

2. Turn a **two-tailed test** into a **one-tailed test**

- This is only appropriate if you have a strong theoretical reason to predict the direction of the effect.

Five factors that influence power (continued)

Power increases when:

3. Sample size (n) **increases**

- This is a good idea: More data gives us a clearer picture of the population.
- Larger samples give us more precise estimates of population parameters, reducing standard error and increasing the likelihood of finding a statistically significant effect

4. Difference in means **increases**

- This is usually not under our control, but we can try to design studies that maximize effect size.
- For example, we can use extreme groups (e.g., comparing very high vs. very low anxiety individuals) to increase the difference between means.

Five factors that influence power (continued)

Power increases when:

5. Standard deviation **decreases**

- This is also a good idea: Populations with less variability make it easier to detect differences between groups.
- Often not under our control, but we can try to use reliable measures and reduce measurement error to decrease variability within groups.
- For example, if we are measuring anxiety, we can use a well-validated questionnaire rather than a single-item measure to reduce measurement error and variability within groups.

When and how do we use power?

We use power calculators in two ways:

1. Calculate power after conducting study from several pieces of information (*post hoc*).
2. Conduct power analyses before conducting study to determine sample size necessary to achieve given level of power given estimate of effect size (*a priori*).

A priori power calculations are especially useful because they help us determine the sample size needed to achieve 80% power with an alpha level of 0.05

We can use online calculators or packages for R to conduct power analyses.

> Computing power is largely beyond the scope of this course, but it is very important you understand power at a conceptual level.

That's all for today!

See you next Thursday for Exam 2!