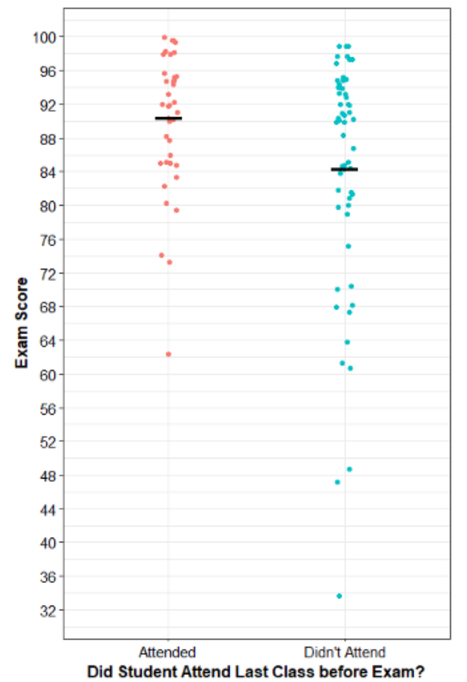
Chapter 11 Comparing Two Means

Hypothesis test for Means of Independent Samples

- In the previous chapters, we focused on inference for a mean or a proportion for a singular population.
- More often though, we might want to make an inference involving the comparison of two populations/conditions, in which we take a sample from each one.

Example: Let's return to an earlier example, in which we compared the exam scores of students who attended the previous class period to those of students who didn't attend the last class.

	Attended	Didn't Attend
Mean Exam Score	$\bar{x}_1 = 90.3$	$\bar{x}_2 = 84.3$
Sample Standard Deviations	$s_1 = 8.4$	$s_2 = 14.2$
Sample Size	$n_1 = 41$	$n_2 = 57$



Guiding Question: We know that the mean exam score of the students who attended is higher than the mean of those who didn't in this sample

$\bar{x}_1 - \bar{x}_2 =$ ________

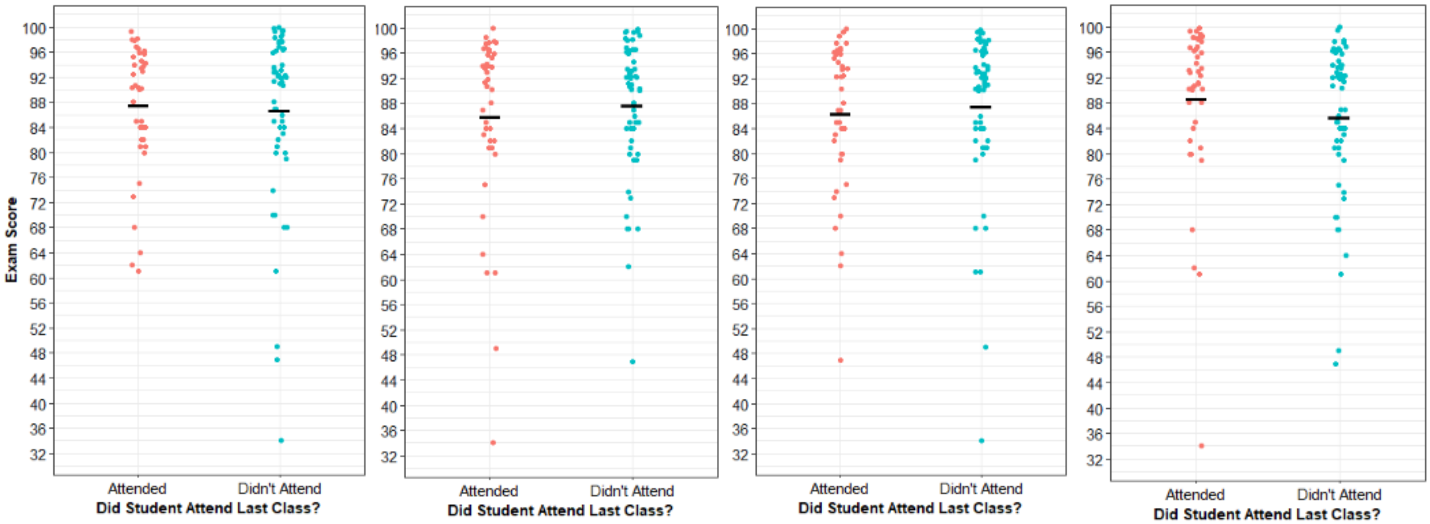
…but does this imply that there is an underlying difference in these two populations of students ($\mu_1 > \mu_2$), or is this the amount of sample mean difference we would expect to see by random chance sampling variability?

- **The Null Hypothesis:** ________________________________
  - Non-directionally: $\mu_1 = \mu_2$
  - Directionally: *Mirror the alternative*
- **The Alternative Hypothesis:** ______________________________________
  - Non-directionally: $\mu_1 \neq \mu_2$
  - Directionally: $\mu_1 > \mu_2$ or $\mu_1 < \mu_2$

Practice: Identify the null and alternative hypotheses for the exam score example both symbolically, and in words.

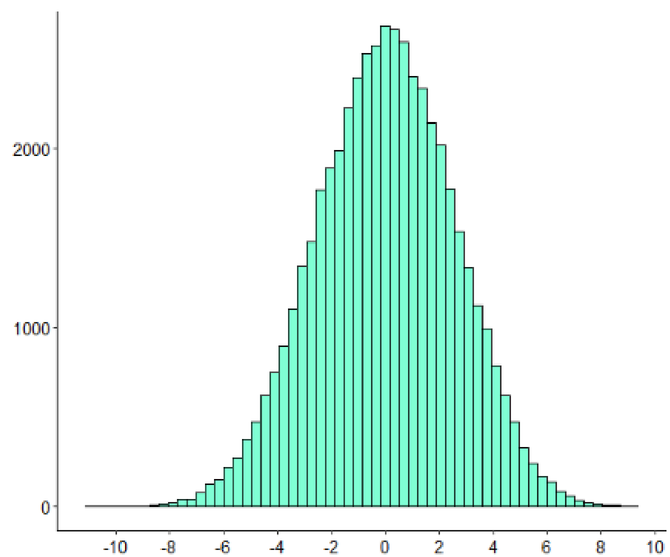
- Simulate the distribution of $\bar{x}_1 - \bar{x}_2$
  - **IF the null is true and there is no underlying difference between these two populations of students on average…**we can "shuffle" the group designation randomly and simulate sample mean differences we should expect under the null hypothesis.

87.4 86.5 | 85.9 87.6 | 86.2 87.4 | 88.5 85.7



$\bar{x}_1 - \bar{x}_2 =$ _____ $\bar{x}_1 - \bar{x}_2 =$ _____ $\bar{x}_1 - \bar{x}_2 =$ _____ $\bar{x}_1 - \bar{x}_2 =$ _____

Simulate 50,000 Times…plot each $\bar{x}_1 - \bar{x}_2$



Distribution of sample mean differences under null hypothesis

Our **real** sample mean difference was 6.0. How often do we see a sample mean difference at least that large under the null hypothesis?

- **Calculating the Standard Error for $\bar{x}_1$ - $\bar{x}_2$**
  - Even if the null hypothesis were true, we expect some difference in our sample means by sampling variability alone.
  - We measure the expected error in our sample mean difference with a ____________________!
  - If we can assume that each population has approximately the same variance, we can use a "pooled" method to calculate this value.
  - The standard error of $\bar{x}_1$ - $\bar{x}_2$ would be calculated as follows:

$$\text{(pooled) } SE_{(\overline{x_1} - \overline{x_2})} = S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

  - S_p represents the __________________________________. This would be the standard deviation of our numeric variable if we ignore the grouping variable.
  - In the exam score example, S_p = 12.1

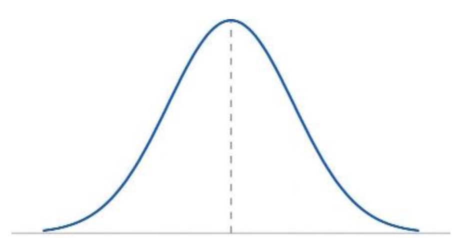
Practice: Calculate the standard error for $\bar{x}_1$ - $\bar{x}_2$ for the exam score comparison data.

In cases where it is not reasonable to assume that each population has approximately the same variance, you can use an "unpooled" method, however, we will not calculate this by hand.

- **Test statistic and p-value**
  - Once we have our standard error, we can calculate the test statistic to see how many standard errors our sample means are away from each other.
  - If sample sizes are not very large, we will need to use an **"independent samples t-test"** to account for standard deviation estimates.
    - To use a "two-sample z-test," we need a sufficiently large sample for each group.
    - A rule of thumb we will use is 1) the combined sample size is > 120 and each individual sample is > 30.
  - Either way, the test statistic is calculated with the same formula.

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sim SE(\overline{x_1} - \overline{x_2})} \qquad z = \frac{\overline{x_1} - \overline{x_2}}{SE(\overline{x_1} - \overline{x_2})}$$

Practice: Calculate and mark the test statistic and shade the p-value range on this t distribution.



- Next, we can use a t-table to approximate the p-value for our test result.
- Note that our degrees of freedom will now be n – 2, where n is the combined sample size of both groups.

Practice: Approximate our p-value from the exam score comparison (complete this as a one-sided test). Decide whether to reject the null hypothesis using $\alpha = 0.05$

- **Making Conclusions**
  - Be very specific what claims we can draw from a hypothesis test.

Practice: Which statement best describes what we have found specifically from this hypothesis test result?

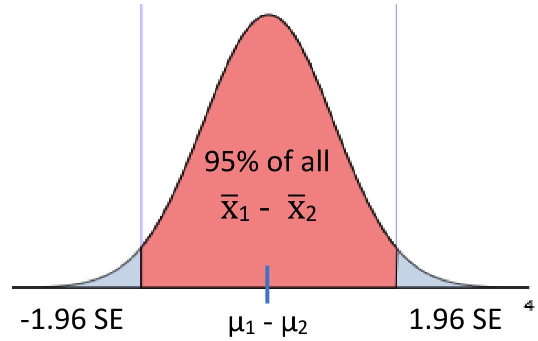
A. Attending the last class before the exam directly leads to higher exam scores on average

B. Students who attend the last class before the exam score *much* higher than students who do not.

C. All students who attend the last class before the exam score higher than all students who do not attend the last class.

D. Students who attend the last class before the exam score higher on average than those who don't. This may be directly related to attendance, or explained by one or more associated variables.

Confidence Interval for $\mu_1 - \mu_2$



- P-values help us determine how confident we are in *any* departure from the null. However, they alone cannot tell us how large that difference is or whether we should care.
    - Very low p-values do **not** necessarily mean a large ________________ from the null
    - Complementing a p-value with a measure of magnitude (like a confidence interval) can add much more context about the inference.
- We can also estimate the parameter $\mu_1 - \mu_2$ using a confidence interval.
- Our point estimate for this parameter is: _____________
- We again need to extend out a certain number of standard errors from our point estimate.

z-interval: $\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} * SE_{(\bar{x}1 - \bar{x}2)}$

90%: $z_{0.05} = 1.645$
95%: $z_{0.025} = 1.960$
98%: $z_{0.01} = 2.326$
99%: $z_{0.005} = 2.576$

t-interval: $\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2} * SE_{(\bar{x}1 - \bar{x}2)}$

t depends on confidence *and* degrees of freedom

Practice: Calculate a 95% t-interval for the true difference in average exam scores between these two populations of students.

Assumptions for z- and t-test comparison of two means

- ✓ The distribution of $\bar{x}_1 - \bar{x}_2$ is normally distributed
    - Each population is already approximately normally distributed OR the Central Limit Theorem would apply (e.g., sample size > 30 and no long tail/large skew).
- ✓ z-test/z-interval specific
    - σ_1 and σ_2 known, or reasonably approximated with large sample sizes (n_1 and n_2 > 30 and df > 120)
- ✓ Pooled t-test/t-interval specific
    - Variances of each group are reasonably close (for this class, we will just state if this is reasonable!)

Practice: Does drinking caffeinated coffee make students more productive? To test this, a researcher gathers 60 college students who don't normally drink coffee. 30 of them are assigned to a decaf coffee in the morning, while 30 are assigned to drink caffeinated coffee in the morning. None of the students knew which group they were assigned to. The coffee brand and amount were prescribed to all participants. At the end of the day, students provide a self-report on their productivity from 1 to 10. Results of the productivity scores are in the table below.

	Caffeine	**No Caffeine**
Mean Productivity Score	$\bar{x}_1 = 6.21$	$\bar{x}_2 = 5.94$
Sample Standard Deviations	$s_1 = 1.98$	$s_2 = 1.90$
Pooled Standard Deviation	$s_p = 1.94$	

Complete a hypothesis test to determine if there is evidence for a difference in average productivity between students who take caffeinated coffee vs. those who don't. Use $\alpha = 0.05$ as your benchmark.

1) **State your Hypotheses symbolically**

2) **Calculate the SE of your sample mean difference**

3) **Calculate the test statistic**

4) **Use a table to find the approximate the p-value**

5) **Make a conclusion!**

If there had been evidence for a difference, would it be appropriate to attribute the difference to the caffeine? Why or why not? *Hint: What type of study design is this?*

Practice: A weight-loss app collects data from 3,682 regular users. These users record their caloric intake for the day, along with specific items they consumed. A dietician decides to record whether the participant included “coffee” as one of their items for the day. This dietician then compares the caloric intake of coffee drinkers to non-coffee drinkers.

	Coffee	**No Coffee**
Mean Caloric Intake	$\bar{x}_1 = 1{,}765$	$\bar{x}_2 = 1{,}691$
Standard Deviation	$s_1 = 289$	$s_2 = 313$
Sample Size	$n_1 = 2{,}075$	$n_2 = 1{,}607$
Pooled Standard Deviation	$s_p = 299$	

This dietician would like to use this sample to generalize to the broader population of people who might be trying to lose weight. For this example, let’s assume this is a representative sample of that population.

Create a 99% confidence interval for the true mean difference in caloric intake between coffee drinkers and non-coffee drinkers who are trying to lose weight.

If we had done a (non-directional) hypothesis test, the p-value would be quite small (<0.001). The dietician is studying whether coffee might be an appetite enhancer (causes people to consume more calories). Does this data provide evidence for that? Why or why not? *Hint: What type of study design is this?*