MAT1372, Classwork25, Fall2025

7.3 Difference of Two Means

1. Using the t-Distribution for a Difference in Means

   The t-distribution can be used for inference when working with the standardized difference of two means if *Independence, extended*. The data are independent within and between the two groups, e.g. the data come from independent random samples or from a randomized experiment.

   Normality. We check the outliers rules of thumb for each group separately.

2. The standard error may be computed as

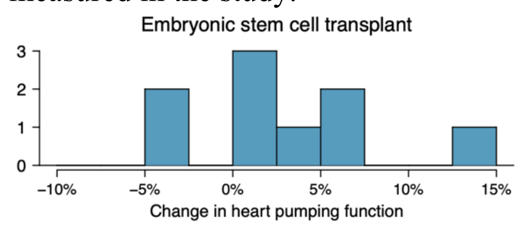
$$SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

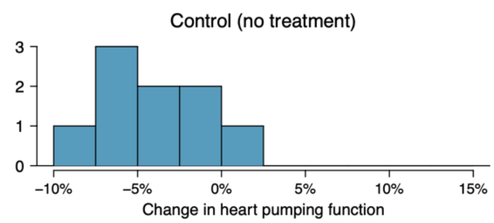
   The degree of freedom is the smaller of $n_1 - 1$ and $n_2 - 1$ if software isn't available.

3. Confidence Interval for a Difference of Means.

   To test if treatment using embryonic stem cells (ESCs) helps improve heart function following a heart attack, each of these sheep (test subjects) was randomly assigned to the ESC or control group, and the change in the hearts' pumping capacity was measured in the study.

	n	$\bar{x}$	s
ESCs	9	3.50	5.17
control	9	-4.33	2.76





(a) What is the point estimate of the difference in the heart pumping variable? $\bar{x}_{ESC} - \bar{x}_{Ctrl}$=3.5-(-4.33)=7.83

(b) Can the t-distribution be used to make inference using this point estimate?

Independence. the sheep were randomized into the groups, independence within and between groups is satisfied.

Normality. The histograms of the two data sets do not reveal any clear outliers in either group.

(c) With both conditions met, we can use the t-distribution to model the difference of sample means. Calculate a 95% confidence interval for the effect of ESCs on the change in heart pumping capacity of sheep after they've suffered a heart attack.

We will use the sample difference that we computed earlier calculations: $\bar{x}_{ESC} - \bar{x}_{Ctrl}$=7.83

and calculate SE and df: $SE = \sqrt{\frac{s_{ESC}^2}{n_{ESC}} + \frac{s_{Ctrl}^2}{n_{Ctrl}}} = \sqrt{\frac{5.17^2}{9} + \frac{2.76^2}{9}} = 1.95$, df=min($n_{ESC} - 1, n_{ESC} - 1$)= 9-1=8

Using df = 8, we can identify the critical value of $t_8^\star$ = 2.31 for a 95% confidence interval.

Finally, we can enter the values into the confidence interval formula:

$$\text{point estimate} \pm t_8^\star \times \text{SE} \rightarrow 7.83 \pm 2.31 \times 1.95 \rightarrow (3.32, 12.34)$$

We are 95% confident that ESCs improve the heart's pumping function in sheep by 3.32% to 12.34%.

4. The outline to build **Confidence Interval for a Difference of Means:**

   Prepare. Retrieve critical contextual information, and if appropriate, set up hypotheses.

   Check. Ensure the required conditions are reasonably satisfied.

   Calculate. Find the SE, and then construct a confidence interval.

   Conclude. Interpret the results in the context of the application.

5. Hypothesis Testing For a Difference of Means.

To test that if newborns from mothers who smoke have a different average birth weight than newborns from mothers who don't smoke, We use the North Carolina sample to try to answer this question. The smoking group includes 50 cases and the nonsmoking group contains 100 cases.

(a) Set up appropriate hypotheses to evaluate if there is a relationship between a mother smoking and average birth weight μ_n and μ_s where μ_n represents non-smoking mothers and μ_s represents mothers who smoked

H_0: $\mu_n - \mu_s = 0$ (There is no difference in average birth weight for newborns from mothers who did and did not smoke.)

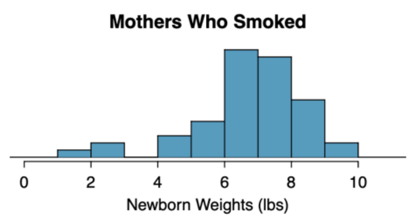
H_A: $\mu_n - \mu_s \neq 0$ (There is some difference in average newborn weights from mothers who did and did not smoke)

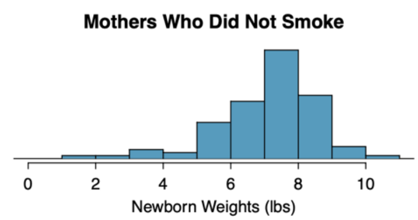
(b) Check the two conditions necessary to model the difference in sample means using the t-distribution.

Independence: a simple random sample both within and between samples.

Normality: the data in the histograms have no particularly extreme outliers.

	smoker	nonsmoker
mean	6.78	7.18
st. dev.	1.43	1.60
samp. size	50	100





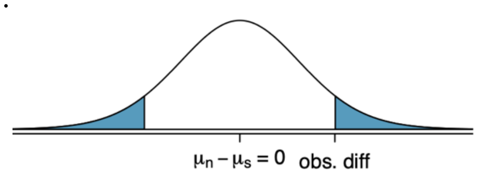
(c) What is the point estimate of the population difference, $\mu_n - \mu_s$? $\mu_n - \mu_s$=0.4

(d) Compute the standard error of the point estimate. $SE = \sqrt{\frac{s_n^2}{n_n} + \frac{s_s^2}{n_s}} = \sqrt{\frac{1.6^2}{100} + \frac{1.43^2}{50}} = 0.26$

(e) Find the degree of freedom. df=min($n_n - 1, n_s - 1$) = min (99,49) = 49.

(f) Complete the hypothesis test by using a significance level of $\alpha = 0.05$.

test statistic for this test: $T = \frac{\text{point estimate} - \text{null value}}{\text{SE}} = \frac{0.4-0}{0.26} = 1.54$



The p-value is represented by the two shaded tails in the following plot with df=49 and the single tail area is between 0.05 to 0.10 and the p-value is between 0.1 to 0.2.

The p-value is larger than the significance value, 0.05, so we do not reject the null hypothesis. There is insufficient evidence to say there is a difference in average birth weight of newborns from North Carolina mothers who did smoke during pregnancy and newborns from North Carolina mothers who did not smoke during pregnancy.

6. The outline to perform **Hypothesis Testing For a Difference of Means:**

Prepare. Retrieve critical contextual information, and if appropriate, set up hypotheses.

Check. Ensure the required conditions are reasonably satisfied.

Calculate. Find the SE, and conduct a hypothesis test, find a test statistic and p-value.

Conclude. Interpret the results in the context of the application.