Class 16: The T-Distribution for One and Two Samples (Text: Sections 7.1 and 7.2)

When can we safely use a T-Test? Robustness of T-Procedures

An inference procedure, such as a *t*-testis **robust** if it is not sensitive to violations for the assumptions. Assumptions we need to make for the *T*-test:

Normal

- The t-distribution applies if original distribution normal
- Since $\bar{x}, s$ are not resistant to outliers, this procedure not robust to outliers
- Provided there are no outliers, procedure is robust with respects to skewness unless very extreme. (That is, unless there is a very noticeable tail in one direction.)

Size of sample: By CLT larger samples mean $\bar{x}$ is more normally distributed

- If $n < 15$, use t-procedure if original population is close to normal, no outliers
- If $n \geq 15$, use t unless outliers, strong skewness, or strongly not normal
- If $n \geq 40$, use t even if skewed
- If $n \geq 120$, can use Z because Z and t distributions are approximately the same for large df.

We have a SRS: It is important that the sampling be well done.

One versus Two Samples

When testing a hypothesis, we ask whether the sample mean is significantly different from some particular value—for example, if coached students had significantly higher SAT scores than the average.

- But how do we know that the coached students are like the rest of the population? Indeed, the fact that they signed up for coaching suggests that they may be different.
- In practice, we often study a *control* group and compare two samples

Today we see how to see if there is a significant difference between two samples.

MATCHED PAIRS (Section 7.1)
Suppose we had SAT results for a group of students before and after coaching. Then we have *matched pairs*: we have two samples (the pre-coaching scores and post-coaching scores) and they are matched in pairs coming from the same person. We test to see if the gain in scores is significant.

Ex: Forty students who are coached have SAT scores where the mean gain is 9.3 and the standard deviation of the gain is 8.1. Is the gain significant?
Consider all students are coached. Let μ be the mean gain in score.
Step 1: Null and alternate hypothesis are
$H_0: \mu = 0$
$H_a: \mu > 0$
This is a one-sided test because the alternate hypothesis is >
Step 2: Calculate the test statistic:

$$t = \frac{9.3 - 0}{8.1/\sqrt{40}} = 7.26 \quad \text{with df} = 39.$$

Step 3: Computing P-values
With the table, the t-value is off the chart, so the P-value is very small
With a calculator, we can find the P-value more precisely:

$$P(T > 7.26) = \text{tcdf}(7.26, 100, 39) = 4.7 \cdot 10^{-9} \text{ tiny}$$

Step 4:
Since the P-value is so small, we reject the null hypothesis and conclude that there is evidence that coaching has an effect.

Ex: Find a 95% confidence interval for the mean increase in SAT score with coaching
Since $n = 40$ and we want $t = 2.021$, we have

$$\left(9.3 - 2.021\,\frac{8.1}{\sqrt{40}}, 9.3 + 2.021\,\frac{8.1}{\sqrt{40}}\right) = (6.71, 11.89)$$

So the mean increase in scores is scores is likely to be between 6.7 and 11.9 points.

Notice that even if the increase in scores is *statistically significant*, it may not be of *practical significance*.

Assumptions for t-test for matched pairs
- Random variable in original population is normally distributed, so difference is normal

Matched pairs are useful because:
- Comparative studies enable us to avoid the avoid effects of confounding (for example, choosing to get coaching with an increase in SAT scores—the students who chose coaching may have been more likely to see an increase in scores anyway)
- Can be used even when randomization is not possible

But students who choose to go for coaching may not represent the rest of the population, so we may not be able to generalize the results to other students. In other words, coaching may not help other students in the same way. Better to compare increases in coached students' scores with increases of other students' scores; this is *difference in differences*. (Difference in differences is not in Math 263.)

COMPARING TWO MEANS : UNMATCHED SAMPLES (Section 7.2)
Suppose we want to compare the means of two groups, but
- No paring of individuals (not matched pairs)
- Samples can be different sizes

We will assume that
- We want to compare means of two groups
- Each group is a sample from a different population
- Responses in each group are independent of those from other group

Notation:

	Variable	Mean	Standard deviation	Size
Population 1	x_1	μ_1	σ_1	
Population 2	x_2	μ_2	σ_2	
Sample 1		$\bar{x}_1$	s_1	n_1
Sample 2		$\bar{x}_2$	s_2	n_2

Goal: To estimate $\mu_1 - \mu_2$

Estimator is $\bar{x}_1 - \bar{x}_2$; we want to know the distribution of $\bar{x}_1 - \bar{x}_2$
- If original distributions of x_1 and x_2 are both normal, then distribution of $\bar{x}_1 - \bar{x}_2$ is normal
- Mean is given by
$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2$$
(This true even without the normality.)
- $\text{Var}\,(\bar{x}_1 - \bar{x}_2) = \text{Var}\,(\bar{x}_1) + \text{Var}\,(\bar{x}_2)$ if $\bar{x}_1$ and $\bar{x}_2$ are independent. Then
$$\text{Var}\,(\bar{x}_1 - \bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$
- So standard deviation is given by
$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Notice this formula for the SE depends on the two samples being independent.

Two-Sample *Z*-statistic for Hypothesis Test
If we know the standard deviations of both populations, then

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Two-Sample *T*-statistic for Hypothesis Test
If we use the sample standard deviations instead of the population standard deviations, then we have a statistic that does not have exactly a T-distribution, but can be approximated by one:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_1 - \mu_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim T(k)$$

where the degree of freedom is $k = \min\,(n_1 - 1, n_2 - 1)$ if we calculate by hand.

Degrees of Freedom by Computer: (Optional) If we calculate the df by computer, the formula used is that below, which gives values close to those obtained by hand—but may not be integers:

$$\text{df} = k = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

Assumption for Two Sample Test:

- The two samples must be independent.

Hypothesis Test for Two Sample Means

Follows the usual steps; calculate the standard deviation using the new formula.

Ex: Are the following two schools comparable in SAT scores?

School 1: A random sample of 43 students has mean SAT of 502, and standard deviation of 60
School 2: A random sample of 35 students has mean SAT of 480, and standard deviation of 50

Show all the steps in the reasoning and include an interpretation of the P-value.

First, look at the data and notice that the difference in sample means, $502 - 480 = 22$, is not large in comparison to either of the standard deviations, 60 and 50. Thus we suspect that the difference in sample means could probably be explained by sampling variation and is probably not significant. To be sure, we do the steps in the hypothesis test.
Is the test one or two sided? Two sided because the question asks about "comparable", not "smaller" or "larger".
The two populations are the students in each school. The variable is quantitative; each student's score.
Step 1: Null and alternate hypothesis are
$H_0: \mu_1 = \mu_2$
$H_a: \mu_1 \neq \mu_2$
This is a two-sided test because the alternate hypothesis is ≠
Step 2: Calculate the test statistic:

$$t = \frac{502 - 480 - (\mu_1 - \mu_2)}{\sqrt{\frac{60^2}{43} + \frac{50^2}{35}}} = \frac{502 - 480 - 0}{\sqrt{\frac{60^2}{43} + \frac{50^2}{35}}} = 1.767 \quad \text{with df} = 34.$$

Step 3: Computing P-values
With the table,

$$P(|T| > 1.767) = 2 \cdot P(T > 1.767) \approx 2(0.05) = 0.10 = 10\%$$

With a calculator, we can find the P-value more precisely:

$$P(|T| > 1.767) = 2 \cdot \text{tcdf}\,(1.767, 20, 34) = 2(0.043) = 8.6\%$$

Step 4: Because 8.6% is above 5%, **we** do not reject H_0 and we conclude there is no significant difference between the two schools' SATs.

Interpretation of the P-value: If the schools have the same mean, there's 8.6% chance of getting sample means that differ by as much as we saw. This is a reasonable large probability, so we do not have evidence to reject H_0.

<u>Confidence Intervals for Difference in Two Population Means</u>

If we know the population standard deviations, the confidence interval for the difference in population means is given by

$$\left(\bar{x}_1 - \bar{x}_2 - z \cdot \sigma_{\bar{x}_1 - \bar{x}_2},\ \bar{x}_1 - \bar{x}_2 + z \cdot \sigma_{\bar{x}_1 - \bar{x}_2} \right)$$

If we do not know the population standard deviations, the confidence interval for the difference in population means is given by

$$\left(\bar{x}_1 - \bar{x}_2 - t \cdot SE_{\bar{x}_1 - \bar{x}_2},\ \bar{x}_1 - \bar{x}_2 + t \cdot SE_{\bar{x}_1 - \bar{x}_2} \right)$$

Where

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad \text{and} \quad SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

<u>Ex</u>: Find a 95% confidence interval for difference in population means for the two schools in the previous example.

For 95% confidence with df $= 30$, we use $t = 2.042$ f and we have that the

$$SE = \sqrt{\frac{60^2}{43} + \frac{50^2}{35}} = 12.46$$

Thus the confidence interval is

$$\left(22 - 2.042\ (12.46), 22 + 2.042(12.46)\right) = (-3.4, 47.4).$$

<u>Ex</u>: Interpret the 95% confidence interval in the previous example.

The true difference in means of the two schools is likely (95% chance) to be in this interval. More precisely, the procedure used to generate this interval has a 95% chance of producing and interval which contains the true difference.

Relationship Between Confidence Intervals and Hypothesis Tests

Ex: What does the confidence interval tell us about the difference in population means of the two schools?

The fact that the interval, $(-3.4, 47.4)$, includes 0 is evidence that the difference in mean SATs of the two schools could be 0. Thus we do not have evidence that the mean SATs of the two schools are significantly different.

If 0 is in a confidence interval, the corresponding *two* sided hypothesis test is not significant.

Pooled estimate for $SE_{\bar{x}}$ (Optional)

If we have two samples from populations that we believe have the same standard deviation (even though we may not know what that value is), we can estimate the shared standard deviation and get a statistics which exactly has the T-distribution. This is called a **pooled t-test**.

If both populations have same σ, then s_1^2 and s_2^2 both estimate σ^2 . We take a weighted average:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad \text{with } df = n_1 + n_2 - 2$$

We can use this standard error in confidence intervals and hypothesis tests.