We compare two means from *independent* samples this time. Remember we use the *t*-distribution for means.

Statistics
Class Notes
Inference about Two Means: Independent Samples (Section 11.3)

Let's test two drugs for the common cold against each other. We will determine the mean time for recovery (in days) for two groups of patients, one receiving drug 1 and the other receiving drug 2. We assume the mean (population) times are the same ($\mu_1 = \mu_2$) and seek (sample) evidence otherwise.

Do females and males have different reaction times? Would people perform better at Trivial Pursuit if they first spent five minutes thinking about being a professor versus thinking about soccer hooligans?

These are independent samples (with possibly different sample sizes and population standard deviations). Do you remember what independent means?

Recall: Definition: A sampling method is **independent** when an individual selected for one sample does *not* dictate which individual is to be in a second sample.

The procedure we use is called **Welch's test**, named after the English statistician Bernard Lewis Welch.

Sampling Distribution of the Difference of Two Means: Independent Samples with Population Standard Deviations Unknown (Welch's *t*):
Suppose that a simple random sample of size n_1 is taken from a population with unknown mean μ_1 and unknown standard deviation σ_1. In addition, a simple random sample of size n_2 is taken from a population with unknown mean μ_2 and unknown standard deviation σ_2.

If the two populations are normally distributed or the sample sizes are sufficiently large ($n_1 \geq 30$, $n_2 \geq 30$), then

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

approximately follows Student's t-distribution with the smaller of $n_1 - 1$ or $n_2 - 1$ degrees of freedom, where $\bar{x}_1$ is the sample mean and s_1 is the sample standard deviation from population 1, and $\bar{x}_2$ is the sample mean and s_2 is the sample standard deviation from population 2.

We will use this fact to test hypotheses much as we have done previously.

We must verify that the following is true before continuing with hypothesis testing:

- sample data (with no outliers) come from simple random sampling, or through a completely randomized experiment with two levels of treatment,
- the samples are independent,
- sample size is small relative to the population size ($n_i \leq 0.05N$), and
- the differences are normally distributed or the sample sizes are large ($n_i \geq 30$).

Small departures from normality will *not* cause trouble. However, outliers are a bigger problem. If outliers exist, do *not* use these procedures.

Summary of the *P*-value Approach:

Step 1: Determine the null and alternative hypotheses. Again, the hypotheses can be structured in one of three ways:

1. Equal versus *not* equal hypothesis **(two-tailed test)**

   H_0: $\mu_1 = \mu_2$

   H_1: $\mu_1 \neq \mu_2$

Our two samples come from two different populations.

2. Equal versus less than **(left-tailed test)**

   H_0: $\mu_1 = \mu_2$

   H_1: $\mu_1 < \mu_2$

3. Equal versus greater than **(right-tailed test)**

   H_0: $\mu_1 = \mu_2$

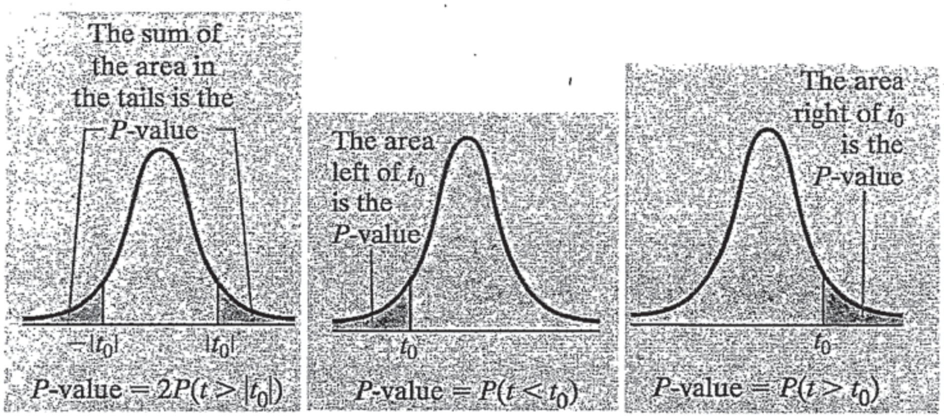
   H_1: $\mu_1 > \mu_2$

The value of $\mu_1 - \mu_2$ is assumed to be 0.

Step 2: Select a level of significance, α, depending on the seriousness of making a Type I error.

Step 3: We *could* compute the test statistic $t_0 = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ (using the *smaller* of $n_1 - 1$ or $n_2 - 1$ degrees of freedom) and use Table VII to approximate the *P*-value. However, we will often use the calculators or StatCrunch to perform the hypothesis testing where this calculation will be done for us.

Step 4: If the P-value $< \alpha$, reject the null hypothesis. For an understanding of the P-values, we will look at these pictures.

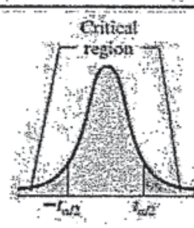
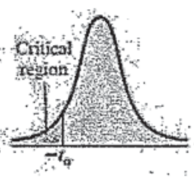


Step 5: State the conclusion.

Again, the value of $\mu_1 - \mu_2$ is assumed to be 0.

Alternatively, Steps 3 and 4 Using Classical Approach:

Step 3: We compute the test statistic $t_0 = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ which follows Student's t-distribution (using the *smaller* of $n_1 - 1$ or $n_2 - 1$ degrees of freedom). Use Table VII to determine the critical value according to the following pictures.

	Two-Tailed	Left-Tailed	Right-Tailed
Critical value	$-t_{\frac{\alpha}{2}}$ and $t_{\frac{\alpha}{2}}$	$-t_{\alpha}$	t_{α}
Critical region			

Some homework problems will require the P-value approach while others will ask for critical values (classical approach).

Step 4: Compare the critical value to the test statistic. If the test statistic is in the shaded region shown above for the appropriate test, we reject the null hypothesis.

Assessing normality and outliers:
We will draw normal probability plots and boxplots to check for normality and outliers as before. Do *not* use these procedures at all if you find outliers or non-normal populations (for sample sizes below 30). Other methods (*not* discussed in this course) are required.

expl 1: Dutch researchers conducted a study in which groups of students were asked to answer 42 Trivial Pursuit (TP) questions. The students in group 1 were first asked to spend five minutes thinking about what it would mean to be a professor. The students in group 2 spent five minutes thinking about soccer hooligans. The relevant data is below.

	Group	**Sample size**	**Mean score***	**Sample standard deviation**
μ_P	**Group 1 (Professor)**	$n_1 = 200$	$\bar{x}_1 = 23.4$	$s_1 = 4.1$
μ_S	**Group 2 (Soccer)**	$n_2 = 200$	$\bar{x}_2 = 17.9$	$s_2 = 3.9$

* The mean score is (presumably) the mean number of questions students got correct.

Follow the steps below to test whether the group that thought about being a professor did better on the TP questions at the 0.05 significance level. Assume the populations are normal and there are no outliers in the data.

a.) Write out the null and alternative hypotheses.

$H_0: \mu_P = \mu_S$ (right-tailed test) $\mu_P - \mu_S = 0$

$H_1: \mu_P > \mu_S$

b.) Compute the test statistic t_0 by hand. The calculator will verify this number.

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{23.4 - 17.9}{\sqrt{\frac{4.1^2}{200} + \frac{3.9^2}{200}}} \approx 13.746$$

✓

c.) Use the calculator to perform the test. Instructions follow. State the conclusion clearly.

$t \approx 13.746$

$p \approx 9.7 \times 10^{-36} \approx 0$

$df \approx 397$

Your test statistic should be large. Look for scientific notation in the *P*-value.

So, pvalue $< \alpha = 0.05 \rightarrow$ We reject null. We have evidence that $\mu_P > \mu_S$.

Instructions for TI Calculators:

1. If needed, enter raw data into columns **L1** and **L2** in **STAT** editor.

2. Press **STAT**, arrow over to **TESTS** and select **4:2-SampTTest.**

3. If the data are raw, highlight **Data**. If you have summary data (means, sample sizes, and standard deviations), highlight **Stats**. If you selected **Data**, tell it where the data is (**L1** and **L2**) and set frequencies to 1. If you selected **Stats**, tell it the mean, standard deviation, and sample size for sample 1 and then that of sample 2.

4. For either **Data** or **Stats**, we select the desired alternative hypothesis and tell it the data is **not pooled.**

5. Highlight **Calculate** and set that sucker off! Do your magic, oh-so-magic box! (**Draw** will draw a graph of the t-curve with the P-value shaded. It will also show the test statistic and P-value. Be sure you have no other functions listed in the y-editor that will graph.)

Why Does the Calculator Give a Different Number for Degrees of Freedom?

The truth is that the degrees of freedom using the *smaller* of $n_1 - 1$ or $n_2 - 1$ gives a conservative estimate. That means we would require more evidence against the null hypothesis (means that are further apart) to reject it. This reduces the probability of a type 1 error (rejecting the null hypothesis when it is true) to beneath the level of significance α.

A more accurate formula for degrees of freedom is used by the technology. It is here for reference but you do *not* need to use it.

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

You would always round this *down* to the nearest integer.

Why Pooled? Do Girls Just Wanna Have Fun?

If the samples come from populations that have the same variance, we use the pooled test. (Pooling refers to finding a weighted average of the two population variances.) However, since we do *not* know the population data, it is safer to assume it is *not*. The book has more to say on this but we will leave it there. When asked by technology, do *not* pool the variances.

Instructions for StatCrunch:

1. If you have it, enter raw data into first two columns and label them.

2. Select **Stat > T Stats > Two Sample** and then choose **With Data** or **With Summary**.

3. Put in all of the information. Notice the difference (under **Perform: Hypothesis test**) will be "Sample 1 minus Sample 2" so be sure to enter the data as you need it. Leave the box for **Pool variances** *unchecked* under **Calculation options**. Choose either **Hypothesis test** or **Confidence interval**, entering the relevant information. Click **Compute!**.

Confidence Intervals:

A $(1-\alpha)\cdot 100\%$ confidence interval for $\mu_1 - \mu_2$ is given by

Lower bound: $(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\cdot\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ and **Upper bound:** $(\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\cdot\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$

where $t_{\alpha/2}$ is the critical value using the *smaller* of $n_1 - 1$ or $n_2 - 1$ degrees of freedom (df).

(You can also use the formula from the last page for df.)

Instructions for TI Calculators:

Follow the steps given for hypothesis tests but select **0:2-SampTInt** from the **STAT > TESTS** menu. Set **Pooled** to **No** as before.

expl 2: Young children require lots of time and attention. Do parents of young children get less leisure time than adults who have no children? Researchers want to estimate the difference in mean daily leisure time between adults who have no kids (under the age of 18) and adults who do have kids (under 18). We obtain the following data through a random survey.

Group	Sample size	Mean daily leisure time	Sample standard deviation
Adults with no kids	$n_1 = 40$	$\bar{x}_1 = 5.62$ hrs	$s_1 = 2.43$ hrs
Adults with kids	$n_2 = 40$	$\bar{x}_2 = 4.10$ hrs	$s_2 = 1.82$ hrs

Questions are on the next page. Use a calculator to form the intervals.

The means are different. Can we quantify that difference for the populations? How confident are we that we got it right?

expl 2 (continued):

a.) Form a 90% interval for the mean difference in daily leisure time between adults with no kids and adults with kids.

We are 90% confident that the mean difference in leisure time for adults with no kids vs. adults with kids is between 0.72 and 2.32 hrs.

Be sure you enter the data in the correct order.

b.) Form and interpret a 95% interval for the mean difference in daily leisure time between adults with no kids and adults with kids.

→ 0.56 and 2.48 hrs.

We are 95% confident that the mean difference in daily leisure time between adults with no kids and adults with kids is between 0.56 and 2.48 hours.

If we want to be *more* confident that the interval contains the true population mean difference, the interval must become larger. ✓

c.) Notice that these intervals do *not* contain 0. What does that say about the difference between leisure time of adults with and without kids?

We are quite certain that people with no kids have more leisure time than those with kids.

(duh!)