Chapter 10: Hypothesis Testing With Two Samples

10.3: Comparing Two Independent Population Proportions

Example 1. When the author recently purchased a new car, he was surprised when the dealer told him that he would be given two license plates, but only the rear plate would be installed. The salesman said that state law requires license plates on the front and rear of the car, but police don't bother to ticket drivers with only a rear license plate. Are the license plate laws enforced as laws or are they treated as guidelines? Is adherence to the laws different in different states?

To help answer these questions, the author collected the sample data provided in the table below. The cars included in the table are non-commercial passenger cars. Connecticut and New York are contiguous states, both having laws that require front and rear license plates.

	Connecticut	New York
Cars with rear license plate only	239	9
Cars with front and rear license plates	1810	541
Total	**2049**	**550**

The notation for comparing two proportions is as follows:

	Population Proportion	Sample Size	Number with Trait	Sample Proportion
Population 1	p_1	n_1	X_1	$\hat{p}_1 = \dfrac{X_1}{n_1}$
Population 2	p_2	n_2	X_2	$\hat{p}_2 = \dfrac{X_2}{n_2}$

The notation applied to Example 1:

	Population Proportion	Sample Size	Number with Trait	Sample Proportion
Connecticut	p_1	2049	239	$\hat{p}_1 = 0.117$
New York	p_2	550	9	$\hat{p}_2 = 0.016$

Steps for the Hypothesis Test for Comparing Two Population Proportions

Step 1: Determine the Null and Alternative Hypotheses

One of the three choices:

1. $H_0 : p_1 = p_2$ versus $H_a : p_1 \neq p_2$ (two-tailed)
2. $H_0 : p_1 \geq p_2$ versus $H_a : p_1 < p_2$ (left-tailed)
3. $H_0 : p_1 \leq p_2$ versus $H_a : p_1 > p_2$ (right-tailed)

Step 2: Verify That The Conditions Are Met And State the Level of Significance

Conditions for the Difference in Two Proportions

1. The samples are representative of the population.
2. Independent samples are available from the two populations.
3. The number of success and failures is at least 5 in each sample.
   $n_1\hat{p}_1 \geq 5, \qquad n_1(1-\hat{p}_1) \geq 5, \qquad n_2\hat{p}_2 \geq 5, \qquad n_2(1-\hat{p}_2) \geq 5$

Thus, the distribution for the difference in proportions is:

$$\hat{p}_1 - \hat{p}_2 \sim \text{ normal } \left[0, \sqrt{p_C(1-p_C)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\right]$$

where $p_C = \dfrac{x_1 + x_2}{n_1 + n_2}$.

Step 3: Summarize the Data into an Appropriate Test Statistic

The Test Statistic

We assume the null hypothesis is true, that is $p_1 = p_2$. First, we must estimate the common population proportion p using all of the data:

$$p_C = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$

The test statistic is as follows:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{p_C(1-p_C)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{p_C(1-p_C)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}},$$

Step 4: Find the p-Value OR determine the critical value

If the null hypothesis is true, this z-statistic ia approximately the standard normal curve, so the standard normal curve is used to find the p-value.

- For $H_a : p_1 > p_2$ (a one-tailed test), the p-value is the area above z, even if z is negative.
- For $H_a : p_1 < p_2$ (a one-tailed test), the p-value is the area below z, even if z is positive.
- For $H_a : p_1 \neq p_2$ (a two-tailed test), the p-value is $2 \times$ area above $|z|$.

The Rejection Region Approach to Hypothesis Testing for a Proportion

We compare the test statistic to the rejection region. If the test statistic falls in the rejection region, the null hypothesis is rejected.

	Rejection Region Rule	
Alternative Hypothesis	$\alpha = 0.05$	$\alpha = 0.01$
$H_a : p_1 < p_2$	Reject H_0 if $z \leq -1.645$	Reject H_0 if $z \leq -2.33$
$H_a : p_1 > p_2$	Reject H_0 if $z \geq 1.645$	Reject H_0 if $z \geq 2.33$
$H_a : p_1 \neq p_2$	Reject H_0 if $\lvert z\rvert \geq 1.96$	Reject H_0 if $\lvert z\rvert \geq 2.575$

A boundary of a rejection region is called a **critical value**. The critical value is the value is the table above, while the rejection region is the area more extreme than this value.

Step 5: Make a decision based on either the p-value or the rejection region

- If p-value $< \alpha$, Reject H_0
- If the test statistic is in the shaded region (rejection region), reject H_0

Step 6: State your conclusion in terms of the problem

Interpret the conclusion in context of the situation. We should also consider the manner in which the data were collected.

1. Set up hypothesis:

Two-Tailed	Left-Tailed	Right-Tailed
$H_0 : p_1 = p_2$	$H_0 : p_1 = p_2$	$H_0 : p_1 = p_2$
$H_1 : p_1 \neq p_2$	$H_1 : p_1 < p_2$	$H_1 : p_1 > p_2$

2. Verify the conditions and select a level of significance, α.
   (a) The samples are representative of the population.
   (b) Independent samples are available from the two populations.
   (c) The number with the trait of interest and the number without the trait of interest is at least 5 in each sample.
   $n_1\hat{p}_1 \geq 5, \qquad n_1(1-\hat{p}_1) \geq 5, \qquad n_2\hat{p}_2 \geq 5, \qquad n_2(1-\hat{p}_2) \geq 5$

3. Compute the **test statistic**

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{p_C(1-p_C)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

   where $p_C = \dfrac{x_1 + x_2}{n_1 + n_2}$.

4. Find the p-Value. Use the z-table.

4. OR determine the **critical value**, z^*.

5. If the p-value $< \alpha$, reject the null hypothesis.

5. OR **Decision Rule:** If the test statistic is MORE EXTREME than the critical value, reject the null hypothesis.

6. Conclusion - State your decision and your conclusion in terms of the problem.

Example 2. In clinical trials of Nasonex, 3774 adult and adolescent allergy patients (patients 12 years and older) were randomly divided into two groups. The patients in group 1 (experimental group) received 200 μg of Nasonex, while the patients in group 2 (control group) received a placebo. Of the 2103 patients in the experimental group, 547 reported headaches as a side effect. Of the 1671 patients in the control group, 368 reported headaches as a side effect. Is there significant evidence to conclude that the proportion of Nasonex users that experienced headaches as a side effect is greater than the proportion in the control group at the $\alpha = 0.05$ level of significance?

Step 1

Step 2. First we must verify the requirements to perform the hypothesis tests.

(a) Subjects were assigned to the treatment randomly.

(b) The samples are independent, and subjects were assigned to the treatment randomly.

(c) We have $x_1 = 547$, $n_1 = 2103$, $x_2 = 368$, and $n_2 = 1671$.

Step 3

Step 4

Step 4

Step 5

Step 5

Step 6

Example 3. The Harris Poll conducted a survey in which they asked, "How many tattoos do you currently have on your body?" Of the 1205 males surveyed, 181 responded that they had at least one tattoo. Of the 1097 females surveyed, 143 responded that they had at least one tattoo. Test the claim that the proportion of males with at least one tattoo differs significantly from the proportion of females with at least one tattoo at the $\alpha = 0.05$ level of significance.

Step 1

Step 2. First we must verify the requirements to perform the hypothesis tests.

(a)

(b)

(c)

Step 3

Step 4

Step 4.

Step 5

Step 5

Step 6

Construct and Interpret Confidence Intervals for the Differnce between Two Population Proportions

To construct a confidence interval for the difference between two population proportions, the following requirements must be satisfied:

1. The samples are representative of the population.
2. Independent samples are available from the two populations.
3. The number with the trait of interest and the number without the trait of interest is at least 5 in each sample.
   $n_1\hat{p}_1 \geq 5, \qquad n_1(1-\hat{p}_1) \geq 5, \qquad n_2\hat{p}_2 \geq 5, \qquad n_2(1-\hat{p}_2) \geq 5$

Provided that these requirements are met, a $(1-\alpha)\cdot 100\%$ confidence interval for $p_1 - p_2$ is given by

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Notice that we do not pool the sample proportions. This is because we are not making any assumptions regarding their equality, as we did in hypothesis testing.

The notation applied to Example 1:

	Population Proportion	Sample Size	Number with Trait	Sample Proportion
Connecticut	p_1	2049	239	$\hat{p}_1 = 0.117$
New York	p_2	550	9	$\hat{p}_2 = 0.016$

Example 4. Construct a 90% confidence interval for the data from Example 1.

Example 5. The Gallup organization surveyed 1100 adult Americans on May 6-9, 2002, and conducted an independent survey of 1100 adult Americans on May 10-13, 2007. In both surveys they asked the following: "Right now, do you think the state of moral values in the country as a whole is getting better or getting worse?" On May 10-13, 2007, 902 of the 1100 surveyed responded that the state of moral values is getting worse; on May 6-9, 2002, 737 of the 1100 surveyed responded that the state of moral values is getting worse. Construct and interpret a 90% confidence interval for the difference between the two population proportions.

First we must verify the requirements to perform the hypothesis tests.

1. Subjects were assigned to the treatment randomly.
2. The samples are independent.
3. We have $x_1 = 902$, $n_1 = 1100$, $x_2 = 737$, and $n_2 = 1100$.

Use the formula to find the lower and upper bounds

We are 90% confident that the difference between the proportion of adult Americans who believed that the state of moral values in the country as a whole was getting worse from 2002 to 2007 is between 0.12 and 0.18.

To put this statement into everyday language, we might say that we are 90% confident that the proportion of adult Americans who believe that the state of moral values in the country as a whole was getting worse increased between 12% and 18% from 2002 to 2007. Because this interval does not contain 0, we can say that a higher proportion of the country believed that the state of moral values was getting worse in the United States in 2007 than in 2002.

Example 6. The Pew Research Group conducted a poll in which they asked, "Are you in favor of, or opposed to, executing persons as a general policy when the crime was committed while under the age of 18?" Of the 580 Catholics surveyed, 180 indicated they favor capital punishment; of the 600 seculars (those who do not associate with a religion) surveyed, 238 favored capital punishment. Construct a 99% confidence interval. Interpret the interval.

Example 7. A study was done to determine whether there is a relationship between snoring and the risk of heart disease (Norton and Dunn, 1985). Among 1105 snorers in the study, 85 had heart disease, while only 24 of 1379 nonsnorers had heart disease. Determine a 95% confidence interval that estimates $p_1 - p_2$ = difference in proportions of nonsnorers and snorers who have heart disease. Interpret the interval.

10.4: Two Population Means: Matched or Paired Samples

Distinguish Between Independent and Dependent Sampling

Consider two scenarios:

Scenario 1: You might theorize that men and women who marry tend to have similar IQ. To test this theory, you randomly select 30 married individuals and measure the IQ of each spouse. Your goal is to determine whether the difference in their IQs is significantly different from zero.

Scenario 2: Do individuals who make fast-food purchases with a credit card tend to spend more than those who pay with cash? To answer this question, a marketing manager randomly selects 30 credit-card receipts and 30 cash receipts to determine if the credit-card receipts have a significantly higher dollar amount, on average.

Is there a difference in the approach taken to select the individuals in each study? Yes! In scenario 1, once a husband (or wife) is selected to be in the study, we automatically *match* the individual up with his or her spouse. In scenario 2, the receipts selected from the credit-card group have nothing at all to do with the receipts selected from the cash group.

In this regard the husbands selected *depend* on the wives selected (or vice versa), while the credit-card receipts selected are *independent* of the cash receipts.

Two samples are **independent** if the individuals selected for one sample do not dictate which individuals are to be in a second sample.

Two samples are **dependent** if the individuals selected to be in one sample are used to determine the individuals in the second sample. Dependent samples are often referred to as **matched-pairs** samples.

So, the sampling methods in scenario 1 is dependent (or a matched-pairs), while the sampling method in scenario 2 is independent.

The term **paired data** is used to describe data that are collected in natural pairs.

Example 8. For each of the following experiments, determine whether the sampling method is independent of dependent.

(a) Researcher Steven J. Sperber, MD, and his associates wanted to determine the effectiveness of a new medication in the treatment of discomfort associated with the common cold. They randomly divided 430 subjects into two groups: Group 1 received the new medication and group 2 received a placebo. The goal of the study was to determine whether the mean of the symptom assessment scores of the individuals receiving the new medication (group 1) was less than that of the placebo group (group 2).

(b) In an experiment conducted in a biology class, Professor Andy Neill measured the time required for 12 students to catch a falling meter stick using their dominant hand and nondominant hand. The goal of the study was to determine whether the reaction time in an individual's dominant hand is different from the reaction time in the nondominant hand.

(c) A sociologist wishes to compare the annual salaries of married couples. She obtains a random sample of 50 married couples in which both spouses work and determines each spouse's annual salary.

(d) A researcher wishes to determine the effects of alcohol on people's reaction times to a stimulus. She randomly divides 100 people aged 21 or older into two groups. Group 1 is asked to drink 3 ounces of alcohol, while group 2 drinks a placebo. Both drinks taste the same, so the individuals in the study do not know which group they belong to. Thirty minutes after consuming the drink, the subjects in each group perform a series of tests meant to measure reaction time.

The Six Steps for a Paired t-Test

Step 1: Determine the Null and Alternative Hypotheses

1. $H_0 : \mu_d = 0$ versus $H_a : \mu_d \neq 0$ (two-tailed)
2. $H_0 : \mu_d = 0$ versus $H_a : \mu_d < 0$ (one-tailed)
3. $H_0 : \mu_d = 0$ versus $H_a : \mu_d > 0$ (one-tailed)

Step 2: Verify That The Conditions Are Met And State the Level of Significance

Conditions Required for a Paired t-Test To Be Valid

1. The sample data are dependent (matched pairs)
2. The matched pairs are a representative sample.
3. One of the following:
   (a) The population must be normally distributed, OR
   (b) The sample size needs to be *large* enough, $n \geq 30$, OR

Step 3: Summarize the Data into an Appropriate Test Statistic

The Test Statistic

$$t = \frac{\bar{x}_d - \mu_d}{s_d/\sqrt{n}}$$

This particular t-statistic has approximately a t-distribution with $df = n - 1$.

Step 4: Find the p-Value OR determine the critical value

Using the t-distribution with $df = n - 1$, the p-value is the area in the tail(s) beyond the test statistic t, as follows:

- For $H_a : \mu_d < 0$ (a one-tailed test), the p-value is the area below t, even if t is positive.
- For $H_a : \mu_d > 0$ (a one-tailed test), the p-value is the area above t, even if t is negative.
- For $H_a : \mu_d \neq 0$ (a two-tailed test), the p-value is $2 \times$ area above $|t|$.

The Rejection Region Approach for t-Tests

We compare the test statistic to the rejection region. If the test statistic falls in the rejection region, the null hypothesis is rejected.

A boundary of a rejection region is called a **critical value**, denoted t^*. and is found using $df = n - 1$ from the t-table.

Step 5: Make a decision based on either the p-value or the rejection region

- If p-value $< \alpha$, Reject H_0
- If the test statistic is in the shaded region (rejection region), reject H_0

Step 6: State your conclusion in terms of the problem

Interpret the conclusion in context of the situation. We should also consider the manner in which the data were collected.

Procedures for Inferences with Dependent Samples

1. Verify that the sample data consist of dependent samples (or matched pairs), and verify that the requirements are satisfied.
2. Find the difference, d, for each pair of sample values. (*Caution:* Be sure to subtract in a consistent manner, such as "before - after.")
3. Find the value of $\bar{x}_d$ (mean of the differences) and s_d (standard deviation of the differences).
4. For hypothesis tests and confidence intervals, use the same t-test procedures used for a single population mean (Chapter 9).

1. Set up hypothesis:

Two-Tailed	Left-Tailed	Right-Tailed
$H_0 : \mu_d = 0$	$H_0 : \mu_d = 0$	$H_0 : \mu_d = 0$
$H_1 : \mu_d \neq 0$	$H_1 : \mu_d < 0$	$H_1 : \mu_d > 0$

2. Verify the following two requirements are satisfied:
   (a) The sample data are dependent (matched pairs)
   (b) The matched pairs are a representative sample.
   (c) One of the following:
      i. The population must be normally distributed, OR
      ii. The sample size needs to be *large* enough, $n \geq 30$, OR

3. Compute the **test statistic**

$$t = \frac{\bar{x}_d - \mu_d}{s_d/\sqrt{n}}$$

   with $df = n - 1$

4. Find the p-Value. Use the t-table to find a p-value range.

4. OR determine the **critical value**, t^*.

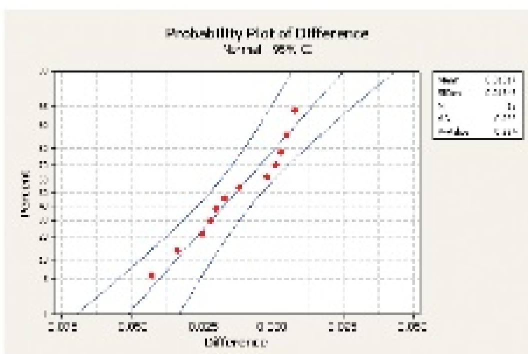
5. If the p-value $< \alpha$, reject the null hypothesis.

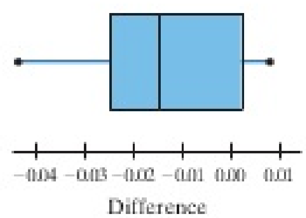
5. OR **Decision Rule:** If the test statistic is MORE EXTREME than the critical value, reject the null hypothesis.

6. Conclusion - State your decision and your conclusion in terms of the problem.

Example 9. Professor Andy Neill measured the time (in seconds) required to catch a falling meter stick for 12 randomly selected students' dominant hand and nondominant hand. Professor Neill wants to know if the reaction time in an individual's dominant hand is less than the reaction time in his or her nondominant hand. A coin flip is used to determine whether reaction time is measured using the dominant or nondominant hand first. Conduct the test at the $\alpha = 0.05$ level of significance.

Student	Dominant Hand, X_i	Nondominant Hand, Y_i	Difference, d_i
1	0.177	0.179	0.177 - 0.179 = -0.002
2	0.210	0.202	0.210 - 0.202 = 0.008
3	0.186	0.208	-0.022
4	0.189	0.184	0.005
5	0.198	0.215	-0.017
6	0.194	0.193	0.001
7	0.160	0.194	-0.034
8	0.163	0.160	0.003
9	0.166	0.209	-0.043
10	0.152	0.164	-0.012
11	0.190	0.210	-0.020
12	0.172	0.197	-0.025





Step 1

Step 2

Step 3

Step 4

Step 4

Step 5

Step 5

Step 6

Example 10. Assume the data below are a random sample of matched pairs data that come from a population that is normally distributed.

Observation	1	2	3	4	5	6	7	8
X_i	19.4	18.3	22.1	20.7	19.2	11.8	20.1	18.6
Y_i	19.8	16.8	21.1	22.0	21.5	18.7	15.0	23.9

(a) Determine $d_i = X_i - Y_i$ for each pair of data.

(b) Compute $\bar{x_d}$ and s_d.

(c) Test if $\mu_d \neq 0$ at the $\alpha = 0.01$ level of significance.

In most cases, paired data are collected because the researchers want to know about the differences and not about the original observations. In particular, it is often of interest to know whether the mean difference in the population is different from 0.

Confidence Interval for the Difference Between Two Dependent Means

Conditions Required for Using the t Confidence Interval for the Mean of Paired Differences

1. The sample data are dependent (matched pairs)
2. The matched pairs are a representative sample.
3. One of the following:
   (a) The population must be normally distributed, OR
   (b) The sample size needs to be *large* enough, $n \geq 30$, OR

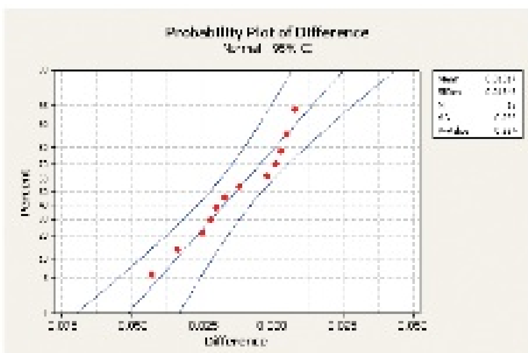
Confidence Interval Formula

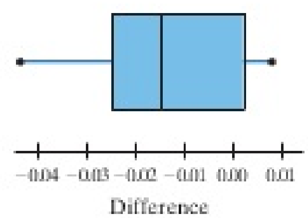
$$\bar{x}_d \pm t^* \frac{s_d}{\sqrt{n}}$$

where $df = n - 1$.

Example 11. Professor Andy Neill measured the time (in seconds) required to catch a falling meter stick for 12 randomly selected students' dominant hand and nondominant hand. A coin flip is used to determine whether reaction time is measured using the dominant or non-dominant hand first. Construct a 95% confidence interval estimate of the mean difference, μ_d.

Student	Dominant Hand, X_i	Nondominant Hand, Y_i	Difference, d_i
1	0.177	0.179	0.177 - 0.179 = -0.002
2	0.210	0.202	0.210 - 0.202 = 0.008
3	0.186	0.208	-0.022
4	0.189	0.184	0.005
5	0.198	0.215	-0.017
6	0.194	0.193	0.001
7	0.160	0.194	-0.034
8	0.163	0.160	0.003
9	0.166	0.209	-0.043
10	0.152	0.164	-0.012
11	0.190	0.210	-0.020
12	0.172	0.197	-0.025





Example 12.

Observation	1	2	3	4	5	6	7	8
X_i	19.4	18.3	22.1	20.7	19.2	11.8	20.1	18.6
Y_i	19.8	16.8	21.1	22.0	21.5	18.7	15.0	23.9

Compute a 99% confidence interval about the population mean difference $\mu_d = \mu_X - \mu_Y$.

10.1: Two Population Means: Independent Samples

It is often of interest to determine whether the means of populations represented by two independent samples of a quantitative variable differ.

The procedure for testing the null hypothesis is called the **two-sample t-test** or **t-test for the difference in two means**. The null hypothesis is that the two means are equal:

$$H_0 : \mu_1 - \mu_2 = 0 \qquad (\text{ or } \mu_1 = \mu_2)$$

The Six Steps for a Two-Sample t-Test

Step 1: Determining the Null and Alternative Hypotheses

One of the three choices:

1. $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$ (two-tailed)
2. $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 < \mu_2$ (one-tailed)
3. $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 > \mu_2$ (one-tailed)

Step 2: Verify That The Conditions Are Met And State the Level of Significance

Conditions Required for a Two-Sample t-Test To Be Valid

1. The two samples are independent.
2. Both samples are representative of the population.
3. One of the following:
   (a) The population must be normally distributed, OR
   (b) The sample size needs to be *large* enough, $n \geq 30$, OR

Step 3: Summarizing the Data into an Appropriate Test Statistic

The Test Statistic

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

where $\mu_1 - \mu_2$ is often assumed to be 0, and the $df = min(n_1 - 1, n_2 - 1)$.

Step 4: Find the p-Value OR determine the critical value

Using the t-distribution with $df = min(n_1 - 1, n_2 - 1)$, the p-value is the area in the tail(s) beyond the test statistic t, as follows:

- For $H_a : \mu_1 < \mu_2$ (a one-tailed test), the p-value is the area below t, even if t is positive.
- For $H_a : \mu_1 > \mu_2$ (a one-tailed test), the p-value is the area above t, even if t is negative.
- For $H_a : \mu_1 \neq \mu_2$ (a two-tailed test), the p-value is $2 \times$ area above $|t|$.

The appropriate degrees of freedom are found by a complicated formula called *Welch's approximation*.

If computer software is used, it will provide the numerical value of degrees of freedom found from the approximation.

If software is not available, a conservative approach is to use the smaller of $n_1 - 1$ and $n_2 - 1$ as the degrees of freedom: $df = min(n_1 - 1, n_2 - 1)$.

Step 5: Make a decision based on either the p-value or the rejection region

- If p-value $< \alpha$, Reject H_0
- If the test statistic is in the shaded region (rejection region), reject H_0

Step 6: State your conclusion in terms of the problem

Interpret the conclusion in context of the situation. We should also consider the manner in which the data were collected.

Testing Hypotheses Regarding Two Population Means, $\mu_1 - \mu_2$

1. Set up hypothesis:

Two-Tailed	Left-Tailed	Right-Tailed
$H_0 : \mu_1 = \mu_2$	$H_0 : \mu_1 = \mu_2$	$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$	$H_1 : \mu_1 < \mu_2$	$H_1 : \mu_1 > \mu_2$

2. Verify the following two requirements are satisfied:
   (a) The two samples are independent.
   (b) Both samples are representative of the population.
   (c) One of the following:
      i. The population must be normally distributed, OR
      ii. The sample size needs to be *large* enough, $n \geq 30$, OR

3. Compute the **test statistic**

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

   with $df = min(n_1 - 1, n_2 - 1)$.

4. Find the p-Value. Use the t-table to find a p-value range.

4. OR determine the **critical value**, t^*.

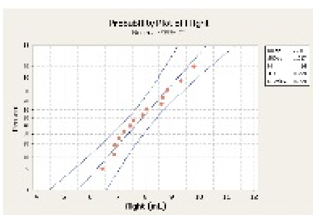
5. If the p-value $< \alpha$, reject the null hypothesis.

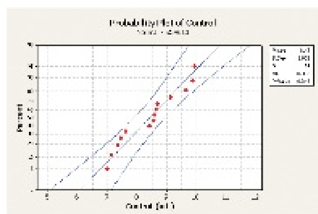
5. OR **Decision Rule:** If the test statistic is MORE EXTREME than the critical value, reject the null hypothesis.

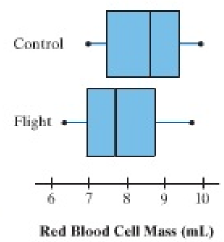
6. Conclusion - State your decision and your conclusion in terms of the problem.

Example 13. In the Spacelab Life Sciences 2 payload, 14 male rats were randomly selected and sent to space. Upon their return, the red blood cell mass (in milliliters) of the rats was determined. A control group of 14 male rats was held under the same conditions (except for space flight) as the space rats and their red blood cell mass was also determined when the space rats returned. The project, led by Dr. Paul X. Callahan, resulted in the data listed below. Does the evidence suggest that the flight animals have a different red blood cell mass from the control animals at the $\alpha = 0.05$ level of signficance.

Flight				Control			
8.59	8.64	7.43	7.21	8.65	6.99	8.40	9.66
6.87	7.89	9.79	6.85	7.62	7.44	8.55	8.70
7.00	8.80	9.30	8.03	7.33	8.58	9.88	9.94
6.39	7.54			7.14	9.14		







Step 1

Step 2

Step 3

Step 4

Step 4

Step 5

Step 5

Step 6

Example 14. Assume the two populations are normally distributed and test whether $\mu_1 \neq \mu_2$ at the $\alpha = 0.1$ level of significance for the given sample data.

	Group 1	Group 2
n	25	18
$\bar{x}$	50.2	42.0
s	6.4	9.9

Step 1

Step 2

Step 3

Step 4

Step 4

Step 5

Step 5

Step 6

	Population Mean	Sample Size	Sample Mean	Sample Standard Deviation
Population 1	μ_1	n_1	$\bar{x}_1$	s_1
Population 2	μ_2	n_2	$\bar{x}_2$	s_2

Confidence Interval for the Difference Between Two Independent Means

An approximate **confidence interval for** $\mu_1 - \mu_2$ is

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

The difficulty is that for this confidence interval procedure, it's not exactly mathematically correct to use a t-distribution to determine the multiplier. It can be used to find an approximate multiplier, but the apprixmation involves a complicated formula for the degrees of freedom. That formula (called **Welch's approximation**) is

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

Statistical software will determine this quantity and will also automate the process of finding the multiplier t^*. If software is not availabe, a conservative "by hand" approach is to use the lesser of $n_1 - 1$ and $n_2 - 1$ for the degrees of freedom:

$$\text{df} = min(n_1 - 1, n_2 - 1)$$

Conditions for Which a t Confidence Interval for the Difference Between Two Means is Valid

1. The two samples are independent.
2. Both samples are representative of the population.
3. One of the following:
   (a) The population must be normally distributed, OR
   (b) The sample size needs to be *large* enough, $n \geq 30$, OR

Example 15. Are male professors and female professors rated differently by students? Below are student course evaluation scores for courses taught by female professors and male professors. Construct a 90% confidence interval for the difference in evaluation scores for the professors.

Female	4.3	4.3	4.4	4.0	3.4	4.7	2.9	4.0	4.3	3.4	3.4	3.3			
Male	4.5	3.7	4.2	3.9	3.1	4.0	3.8	3.4	4.5	3.8	4.3	4.4	4.1	4.2	4.0

Example 16. For a sample of 35 men, the mean head circumference is 57.5 cm with a standard deviation equal to 2.4 cm. For a sample of 36 women, the mean head circumference is 55.3 cm with a standard deviation equal to 1.8 cm. Find an approximate 95% confidence interval for the difference between population mean head circumferences for men versus women.

Mixed Problems

Example 17. The following data represent the muzzle velocity (in feet per second) of rounds fired from a 155-mm gun. For each round, two measurements of the velocity were recorded using two different measuring devices, with the following data obtained:

Observation	1	2	3	4	5	6	7	8	9	10	11	12
A	793.8	793.1	792.4	794.0	791.4	792.4	791.7	792.3	789.6	794.4	790.9	793.5
B	793.2	793.3	792.6	793.8	791.6	791.6	791.6	792.4	788.5	794.7	791.3	793.5

1. Is there a difference in the measurement of the muzzle velocity between device A and device B at the $\alpha = 0.01$ level of significance? Note: Q-Q- Plots indicate the data are normally distributed.

Example 18. A STAT 200 instructor wants to know how traditional students and adult learners differ in terms of their final exam scores. She randomly sampled 239 traditional students and found their mean exam score to be 41.48 with a standard deviation of 6.03. A different random sample of 138 adult learners resulted in a mean exam score of 40.79 with a standard deviation of 6.79. Test the instructor's claim using $\alpha = 0.05$.

Example 19. The headline for an article in the *Sacramento Bee* read, "Women appear to be better investors than men in study" (Jack Sirard, April 24, 2005, p. D1). The conclusions was based on a telephone poll of 500 men and 500 women. One quote in the article was "men are much more likely to stick with a losing investment than women (47 percent to 35 percent)." Test this claim using $\alpha = 0.05$.