Chapter 11
Comparing Two Populations or Treatments

Notation - Comparing Two Means

<table>
<thead>
<tr>
<th>Population or Treatment</th>
<th>Mean Value</th>
<th>Variance</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population or Treatment 1</td>
<td>$\mu_1$</td>
<td>$\sigma_1^2$</td>
<td>$\sigma_1$</td>
</tr>
<tr>
<td>Population or Treatment 2</td>
<td>$\mu_2$</td>
<td>$\sigma_2^2$</td>
<td>$\sigma_2$</td>
</tr>
</tbody>
</table>

Notation - Comparing Two Means

<table>
<thead>
<tr>
<th>Sample</th>
<th>Size</th>
<th>Mean</th>
<th>Variance</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample$_1$</td>
<td>$n_1$</td>
<td>$\bar{x}_1$</td>
<td>$s_1^2$</td>
<td>$s_1$</td>
</tr>
<tr>
<td>Sample$_2$</td>
<td>$n_2$</td>
<td>$\bar{x}_2$</td>
<td>$s_2^2$</td>
<td>$s_2$</td>
</tr>
</tbody>
</table>
11.1: Sampling Distribution for Comparing Two Means

If the random samples on which \( \bar{x}_1 \) and \( \bar{x}_2 \) are based are selected independently of one another, then

1. \( \mu_{\bar{x}_1 - \bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2 \)

2. \( \sigma_{\bar{x}_1 - \bar{x}_2}^2 = \sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \)

\( \sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \)

This implies that the sampling distribution of \( \bar{x}_1 - \bar{x}_2 \) is also normal or approximately normal.

If \( n_1 \) and \( n_2 \) are both large or the population distributions are (at least approximately) normal then \( \bar{x}_1 \) and \( \bar{x}_2 \) each have (at least approximately) a normal distribution.

The distribution of \( \bar{x}_1 - \bar{x}_2 \) is described (at least approximately) by the standard normal (z) distribution.
Hypothesis Tests Comparing Two Means

Large size sample techniques allow us to test the null hypothesis $H_0: \mu_1 - \mu_2 = \text{hypothesized value}$ against one of the usual alternate hypotheses using the statistic

$$z = \frac{\bar{x}_1 - \bar{x}_2 - \text{hypothesized value}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Hypothesis Test Example Comparing Two Means

We would like to compare the mean fill of 16 ounce cans of beer from two adjacent filling machines. Past experience has shown that the population standard deviations of fills for the two machines are known to be $\sigma_1 = 0.043$ and $\sigma_2 = 0.052$ respectively.

A sample of 35 cans from machine 1 gave a mean of 16.031 and a sample of 31 cans from machine 2 gave a mean of 16.009. State, perform and interpret an appropriate hypothesis test using the 0.05 level of significance.

Since $n_1$ and $n_2$ are both large (> 30) we do not have to make any assumptions about the nature of the distributions of the fills.

This example is a bit of a stretch, since knowing the population standard deviations (without knowing the population means) is very unusual.

Accept this example for what it is, just a sample of the calculation. Generally this statistic is used when dealing with “what if” type of scenarios and we will move on to another technique that is somewhat more commonly used when $\sigma_1$ and $\sigma_2$ are not known.
Hypothesis Test Example
Comparing Two Means

$\mu_1 =$ mean fill from machine 1
$\mu_2 =$ mean fill from machine 2
$H_0: \mu_1 - \mu_2 = 0$
$H_a: \mu_1 - \mu_2 \neq 0$

Significance level: $\alpha = 0.05$

Test statistic:
$$z = \frac{\bar{x}_1 - \bar{x}_2 - \text{hypothesized value}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Calculation:
$$z = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(16.031 - 16.009)}{\sqrt{\frac{0.043^2}{3} + \frac{0.052^2}{3}}} = 1.86$$

$P$-value:

$P$-value = 2P(z > 1.86) = 2P(z < -1.86) = 2(0.0314) = 0.0628$

The $p$-value of the test is 0.0628. There is insufficient evidence to support a claim that the two machines produce bottles with different mean fills at a significance level of 0.05.

Equivalently, we might say.

With a $p$-value of 0.063 we have been unable to show the difference in mean fills is statistically significant at the 0.05 significance level.
Two-Sample t Test for Comparing Two Population Means

If we don’t know the population standard deviations, we compute a “t” instead of a ‘z’.

But remember with a “t” we need df.

Can’t just average the df from each group.

df should be truncated to an integer.

Two-Sample t Tests for Difference of Two Means

Alternate hypothesis and finding the P-value:

1. $H_a: \mu_1 - \mu_2 >$ hypothesized value
   
P-value = Area under the $z$ curve to the right of the calculated $z$

2. $H_a: \mu_1 - \mu_2 <$ hypothesized value
   
P-value = Area under the $z$ curve to the left of the calculated $z$
Two-Sample t Tests for Difference of Two Means
3. $H_0: \mu_1 - \mu_2 = \text{hypothesized value}$
   i. $2 \cdot (\text{area to the right of } z) \text{ if } z \text{ is positive}$
   ii. $2 \cdot (\text{area to the left of } z) \text{ if } z \text{ is negative}$

Hypothesis Test Example
In an attempt to determine if two competing brands of cold medicine contain, on the average, the same amount of acetaminophen, twelve different tablets from each of the two competing brands were randomly selected and tested for the amount of acetaminophen each contains. The results (in milligrams) follow. Use a significance level of 0.01.

<table>
<thead>
<tr>
<th>Brand A</th>
<th>Brand B</th>
</tr>
</thead>
<tbody>
<tr>
<td>517, 495, 503, 491</td>
<td>493, 508, 513, 521</td>
</tr>
<tr>
<td>503, 493, 505, 495</td>
<td>541, 533, 500, 515</td>
</tr>
<tr>
<td>498, 481, 499, 494</td>
<td>536, 498, 515, 515</td>
</tr>
</tbody>
</table>

State and perform an appropriate hypothesis test.

Assumptions: The samples were selected independently and randomly. Since the samples are not large, we need to be able to assume that the populations (of amounts of acetaminophen) are both normally distributed.
Hypothesis Test Example

Assumptions (continued):

As we can see from the normality plots and the boxplots, the assumption that the underlying distributions are normally distributed appears to be quite reasonable.

Hypothesis Test Example

\[ \mu_1 = \text{the mean amount of acetaminophen in cold tablet brand A} \]
\[ \mu_2 = \text{the mean amount of acetaminophen in cold tablet brand B} \]

\[ H_0: \mu_1 = \mu_2 \ (\mu_1 - \mu_2 = 0) \]
\[ H_a: \mu_1 \neq \mu_2 \ (\mu_1 - \mu_2 \neq 0) \]

Significance level: \( \alpha = 0.01 \)

Test statistic:

\[ t = \frac{\bar{x}_1 - \bar{x}_2 - \text{hypothesized value}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \]

Calculation:

\[ n_1 = 12, \bar{x}_1 = 497.83, s_1 = 8.830 \]
\[ n_2 = 12, \bar{x}_2 = 515.67, s_2 = 15.144 \]
\[ t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{497.83 - 515.67 - 0}{\sqrt{\frac{8.830^2}{12} + \frac{15.144^2}{12}}} = -3.52 \]
Hypothesis Test Example

**Calculation:**

\[ V_1 = \frac{s_1^2}{n_1} = \frac{8.8300^2}{12} = 6.4974 \]
\[ V_2 = \frac{s_2^2}{n_2} = \frac{15.144^2}{12} = 19.112 \]
\[ \text{df} = \frac{(V_1 + V_2)^2}{V_1 + V_2} = \frac{(6.4974 + 19.112)^2}{6.4974 + 19.112} = 17.7 \]
\[ \text{df} = \frac{11}{11} + \frac{11}{11} \]

We truncate the degrees of freedom to give df = 17.

---

**P-value:** From the table of tail areas for t curve (Table 4) we look up a t value of 3.5 with df = 17 to get 0.001. Since this is a two-tailed alternate hypothesis, P-value = 2(0.001) = 0.002.

**Conclusion:** Since P-value = 0.002 < 0.01 = \( \alpha \), \( H_0 \) is rejected. The data provides strong evidence that the mean amount of acetaminophen is not the same for both brands. Specifically, there is strong evidence that the average amount per tablet for brand A is less than that for brand B.

---

**Confidence Intervals Comparing Two Means**

The general formula for a confidence interval for \( \mu_1 - \mu_2 \) when

1. The two samples are independently chosen random samples, and

2. The sample sizes are both large (generally \( n_1 \geq 30 \) and \( n_2 \geq 30 \)) OR the population distributions are approximately normal is

\[ (\bar{X}_1 - \bar{X}_2) \pm (t \text{ critical value}) \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \]
Confidence Intervals Comparing Two Means

The t critical value is based on

\[
df = \frac{(V_1 + V_2)^2}{\frac{V_1^2}{n_1 - 1} + \frac{V_2^2}{n_2 - 1}}
\]

where \( V_1 = \frac{s_1^2}{n_1} \) and \( V_2 = \frac{s_2^2}{n_2} \)

df should be truncated to an integer. The t critical values are in the table of t critical values (table 3).

Confidence Intervals Example Comparing Two Means

Two kinds of thread are being compared for strength. Fifty pieces of each type of thread are tested under similar conditions. The sample data is given in the following table.

Construct a 98% confidence interval for the difference of the population means.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Sample mean</th>
<th>Sample Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thread A</td>
<td>50</td>
<td>78.3</td>
</tr>
<tr>
<td>Thread B</td>
<td>50</td>
<td>87.2</td>
</tr>
</tbody>
</table>

\[
V_1 = \frac{s_1^2}{n_1} = \frac{5.62^2}{50} = 0.632 \\
V_2 = \frac{s_2^2}{n_2} = \frac{6.31^2}{50} = 0.796 \\
df = \frac{(0.632 + 0.796)^2}{\frac{0.632^2}{49} + \frac{0.796^2}{49}} = 96.7 \\
\]

Truncating, we have df = 96. Conservative df is smallest n-1 or 50-1= 49, if sig with this, will be sig with actual...
Confidence Intervals Example
Comparing Two Means

Looking on the table of t critical values (table 3) under 98% confidence level for df = 96, (we take the closest value for df, specifically df = 120) and have the t critical value = 2.36.

\[
(78.3 - 87.2) \pm 2.36 \sqrt{\frac{5.62^2}{50} + \frac{6.31^2}{50}}
\]

\[-8.9 \pm 2.82\]

The 98% confidence interval estimate for the difference of the means of the tensile strengths is:

\((-11.72, -6.08)\)

Can we determine significant differences???

Confidence Intervals Example
Comparing Two Means

A student recorded the mileage he obtained while commuting to school in his car. He kept track of the mileage for twelve different tanks of fuel, involving gasoline of two different octane ratings. Compute the 95% confidence interval for the difference of mean mileages. His data follow:

<table>
<thead>
<tr>
<th>87 Octane</th>
<th>90 Octane</th>
</tr>
</thead>
<tbody>
<tr>
<td>26.4, 27.6, 29.7</td>
<td>30.5, 30.9, 29.2</td>
</tr>
<tr>
<td>28.9, 29.3, 28.8</td>
<td>31.7, 32.8, 29.3</td>
</tr>
</tbody>
</table>

Comments about assumptions

By looking at the following normality plots, we see that the assumption of normality for each of the two populations of mileages appears reasonable.

Given the small sample sizes, the assumption of normality is very important, so one would be a bit careful utilizing this result.
Confidence Intervals Example

Let the 87 octane fuel be the first group and the 90 octane fuel the second group, so we have
\( n_1 = n_2 = 6 \) and

\[ \bar{x}_1 = 28.45, \ s_1 = 1.228, \ \bar{x}_2 = 30.73, \ s_2 = 1.392 \]

\[ V_1 = \frac{s_1^2}{n_1} = \frac{1.228^2}{6} = 0.2512 \]
\[ V_2 = \frac{s_2^2}{n_2} = \frac{1.392^2}{6} = 0.3231 \]

\[ df = \frac{(0.2512 + 0.3231)^2}{\frac{0.2512^2}{5} + \frac{0.3231^2}{5}} = 9.8 \]

Truncating, we have \( df = 9 \). Conservative \( df = 5 \)

Looking on the table under 95% with 9 degrees of freedom, the critical value of \( t \) is 2.26.

\[ \bar{x}_1 - \bar{x}_2 \pm (t \text{ critical value}) \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \]

\[ = 28.45 - 30.73 \pm 2.26 \sqrt{\frac{1.228^2}{6} + \frac{1.392^2}{6}} \]

\[ -2.28 \pm 1.71 \]

The 95% confidence interval for the true difference of the mean mileages is \((-3.99, -0.57)\).

Confidence Intervals Example

Comments about assumptions

Comments: We had to assume that the samples were independent and random and that the underlying populations were normally distributed since the sample sizes were small.

If we randomized the order of the tankfuls of the two different types of gasoline we can reasonably assume that the samples were random and independent. By using all of the observations from one car we are simply controlling the effects of other variables such as year, model, weight, etc.
11.2: Comparing Two Population or Treatment Means – Paired t Test

Null hypothesis: $H_0: \mu_d = \text{hypothesized value}$

Test statistic:

$$t = \frac{\bar{X}_d - \text{hypothesized value}}{\frac{s_d}{\sqrt{n}}}$$

Where $n$ is the number of sample differences and $\bar{X}_d$ and $s_d$ are the sample mean and standard deviation of the sample differences. This test is based on $df = n-1$.

Comparing Two Population or Treatment Means – Paired t Test

Assumptions:

1. The samples are paired.
2. The $n$ sample differences can be viewed as a random sample from a population of differences.
3. The number of sample differences is large (generally at least 30) OR the population distribution of differences is approximately normal.

Comparing Two Population or Treatment Means – Paired t Test

Alternate hypothesis and finding the P-value:

1. $H_a: \mu_d > \text{hypothesized value}$
   
   P-value = Area under the appropriate t curve to the right of the calculated $t$

2. $H_a: \mu_d < \text{hypothesized value}$
   
   P-value = Area under the appropriate t curve to the left of the calculated $t$

3. $H_a: \mu_d \neq \text{hypothesized value}$
   
   i. $2 \times$ (area to the right of $t$) if $t$ is positive
   
   ii. $2 \times$ (area to the left of $t$) if $t$ is negative
Paired t Test Example

A weight reduction center advertises that participants in its program lose an average of at least 5 pounds during the first week of the participation. Because of numerous complaints, the state’s consumer protection agency doubts this claim. To test the claim at the 0.05 level of significance, 12 participants were randomly selected. Their initial weights and their weights after 1 week in the program appear on the next slide. Set up and perform an appropriate hypothesis test.

Paired Sample Example continued

<table>
<thead>
<tr>
<th>Member</th>
<th>Initial Weight</th>
<th>One Week Weight</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>155</td>
<td>165</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>153</td>
<td>151</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>174</td>
<td>170</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>125</td>
<td>123</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>149</td>
<td>144</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>152</td>
<td>149</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>135</td>
<td>131</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>143</td>
<td>147</td>
<td>-4</td>
</tr>
<tr>
<td>9</td>
<td>139</td>
<td>138</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>198</td>
<td>192</td>
<td>6</td>
</tr>
<tr>
<td>11</td>
<td>215</td>
<td>211</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>153</td>
<td>152</td>
<td>1</td>
</tr>
</tbody>
</table>
Paired Sample Example continued

μ_d = mean of the individual weight changes (initial weight–weight after one week)

This is equivalent to the difference of means:

\[ \mu_d = \mu_1 - \mu_2 = \mu_{\text{initial weight}} - \mu_{\text{1 week weight}} \]

H_0: μ_d = 5
H_1: μ_d < 5

Significance level: \( \alpha = 0.05 \)

Test statistic:

\[ t = \frac{\bar{x}_d - \text{hypothesized value}}{s_d / \sqrt{n}} = \frac{\bar{x}_d - 5}{2.674 / \sqrt{12}} \]

Paired Sample Example continued

Assumptions: According to the statement of the example, we can assume that the sampling is random. The sample size (12) is small, so from the boxplot we see that there is one outlier but never the less, the distribution is reasonably symmetric and the normal plot confirms that it is reasonable to assume that the population of differences (weight losses) is normally distributed.

Paired Sample Example continued

Calculations: According to the statement of the example, we can assume that the sampling is random. The sample size (12) is small, so

\[ n = 12, \bar{x}_d = 2.333, s_d = 2.674 \]

\[ t = \frac{\bar{x}_d - 5}{s_d / \sqrt{n}} = \frac{2.333 - 5}{2.674 / \sqrt{12}} = -3.45 \]

P-value: This is an lower tail test, so looking up the t value of 3.5 under df = 11 in the table of tail areas for t curves (table 4) we find that the P-value = 0.002.
Paired Sample Example continued

Conclusions: Since $P$-value $= 0.002 < 0.05 = \alpha$, we reject $H_0$. We draw the following conclusion. There is strong evidence that the mean weight loss for those who took the program for one week is less than 5 pounds.

Paired Sample Example continued

Minitab returns the following for a paired t test:

This is substantially the same result.

Paired T-Test and CI: Initial, One week

Paired T for Initial - One week

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>12</td>
<td>160.92</td>
<td>8.14</td>
</tr>
<tr>
<td>One week</td>
<td>12</td>
<td>158.58</td>
<td>7.93</td>
</tr>
<tr>
<td>Difference</td>
<td>12</td>
<td>2.333</td>
<td>0.772</td>
</tr>
</tbody>
</table>

95% upper bound for mean difference: 3.720
T-Test of mean difference $= 5$ vs $< 5$: T-Value $= -3.45$ P-Value $= 0.003$

Confidence Intervals Paired Means

The general formula for a confidence interval for $\mu_d$ when

1. The two samples are paired random samples, and
2. The number of paired samples is large (generally $n \geq 30$) OR the population distribution is approximately normal

$$ \bar{X}_d \pm (t \text{ critical value}) \left( \frac{s_d}{\sqrt{n}} \right) $$
Paired $t$ CI Example

Because of numerous complaints, the state’s consumer protection agency is looking into the claim that the alcohol content of a certain wine company was higher than advertised. They decided to compute a 95% CI to test the claimed percent alcohol by randomly testing 6 bottles of wine from the company.

Actual: 14.2 14.5 14.0 14.9 13.6 12.6
Label: 14.0 14.0 13.5 15.0 13.0 12.5

Is there enough evidence to say the labels overstate the actual percent of alcohol? The data on percent alcohol is known to be normal.

Paired $t$ CI Example

Actual: 14.2 14.5 14.0 14.9 13.6 12.6
Label: 14.0 14.0 13.5 15.0 13.0 12.5
Difference: 0.2 0.5 0.5 -0.1 0.6 0.1

$x_d = 0.3$, $s_d = 0.2757$  
Crit $t$ for $\alpha = 0.05$ & df = 5 is 2.02 (one sided, so look under .90)

$0.3 \pm (2.02)(0.2757/\sqrt{6}) \implies 0.3 \pm 0.2274$

$(0.0726$ to $0.5273) \implies$ since 0 not in range, sig. differ

11.3: Large-Sample Inferences

Difference of Two Population (Treatment) Proportions

Some notation:

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Population Proportion of Successes</th>
<th>Sample Proportion of Successes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population or treatment 1</td>
<td>$n_1$</td>
<td>$\pi_1$</td>
</tr>
<tr>
<td>Population or treatment 2</td>
<td>$n_2$</td>
<td>$\pi_2$</td>
</tr>
</tbody>
</table>
Properties: Sampling Distribution of \( p_1 - p_2 \)

If two random samples are selected independently of one another, the following properties hold:

1. \( \pi_{p_1-p_2} = \pi_1 - \pi_2 \)
2. \( \sigma_{\hat{p}_1-\hat{p}_2}^2 = \sigma_{\hat{p}_1}^2 + \sigma_{\hat{p}_2}^2 = \frac{\pi(1-\pi)}{n_1} + \frac{\pi(1-\pi)}{n_2} \) and
   \[ \sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{\pi(1-\pi)}{n_1} + \frac{\pi(1-\pi)}{n_2}} \]
3. If both \( n_1 \) and \( n_2 \) are large \([n_1 \pi \geq 10, n_1(1- \pi) \geq 10, n_2 \pi \geq 10, n_2(1- \pi) \geq 10]\), then \( \hat{p}_1 \) and \( \hat{p}_2 \) each have a sampling distribution that is approximately normal.

Large-Sample z Tests for \( \pi_1 - \pi_2 = 0 \)

The book likes to do a combined estimate of the common population proportion, which is

\[ \hat{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2} = \frac{\text{total number of successes in two samples}}{\text{total sample size}} \]

I seldom use this. It is a little shorter computationally & IF \( n_1 \approx n_2 \) & \( \hat{p}_1 \approx \hat{p}_2 \), you get basically same result. However, if these are substantially different: way off

Large-Sample z Tests for \( \pi_1 - \pi_2 = 0 \)

Null hypothesis: \( H_0: \pi_1 - \pi_2 = 0 \)

Test statistic:

\[ z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} \]

Assumptions:

1. The samples are independently chosen random samples OR treatments are assigned at random to individuals or objects (or vice versa).
2. Both sample sizes are large:
   \( n_1 \hat{p}_1 \geq 10, n_1(1- \hat{p}_1) \geq 10, n_2 \hat{p}_2 \geq 10, n_2(1- \hat{p}_2) \geq 10 \)
Large-Sample z Tests for $\pi_1 - \pi_2 = 0$

Alternate hypothesis and finding the P-value:

1. $H_a: \pi_1 - \pi_2 > 0$
   P-value = Area under the z curve to the right of the calculated z

2. $H_a: \pi_1 - \pi_2 < 0$
   P-value = Area under the z curve to the left of the calculated z

3. $H_a: \pi_1 - \pi_2 \neq 0$
   i. $2 \times$ (area to the right of z) if z is positive
   ii. $2 \times$ (area to the left of z) if z is negative

Example - Student Retention

A group of college students were asked what they thought the “issue of the day”. Without a pause the class said “student retention”. The class then went out and obtained a random sample (questionable) and asked the question, “Do you plan on returning next year?”

The responses along with the gender of the person responding are summarized in the following table.

<table>
<thead>
<tr>
<th>Response</th>
<th>Yes</th>
<th>No</th>
<th>Maybe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>211</td>
<td>45</td>
<td>19</td>
</tr>
<tr>
<td>Gender</td>
<td>Female</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>141</td>
<td>32</td>
<td>9</td>
</tr>
</tbody>
</table>

Test to see if the proportion of students planning on returning is the same for both genders at the 0.05 level of significance.

Example - Student Retention

$\pi_1$ = true proportion of males who plan on returning
$\pi_2$ = true proportion of females who plan on returning
$n_1$ = number of males surveyed
$n_2$ = number of females surveyed
$p_1 = x_1/n_1$ = sample proportion of males who plan on returning
$p_2 = x_2/n_2$ = sample proportion of females who plan on returning

Null hypothesis: $H_0: \pi_1 - \pi_2 = 0$

Alternate hypothesis: $H_a: \pi_1 - \pi_2 \neq 0$
Assumptions: The two samples are independently chosen random samples. Furthermore, the sample sizes are large enough since
\[ n_1 p_1 = 211 \geq 10, \quad n_1(1- p_1) = 64 \geq 10 \]
\[ n_2 p_2 = 141 \geq 10, \quad n_2(1- p_2) = 41 \geq 10 \]

Example - Student Retention

Significance level: \( \alpha = 0.05 \)

Test statistic:
\[
z = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \text{ or } \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}\]

Calculations:
\[
p_1 = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{211 + 141}{275 + 182} = 0.7702
\]
\[
z = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} = \frac{0.76727 - 0.77473}{\sqrt{0.77024(0.77024) + 0.77024(1 - 0.77024)}}
\]
\[
= \frac{-0.0074525}{0.040198} = -0.19
\]

Example - Student Retention

Calculations:
\[
p_1 - p_2 = 0.774 - 0.767 = 0.007
\]
\[
\frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} = \frac{0.0074525}{0.040198} = -0.19
\]

Example - Student Retention

Calculations:
\[
p = \frac{p_1 + p_2}{2} = 0.7702
\]
\[
z = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} = \frac{0.76727 - 0.77473}{\sqrt{0.77024(0.77024) + 0.77024(1 - 0.77024)}}
\]
\[
= \frac{-0.0074525}{0.040198} = -0.19
\]

p value = 2(0.42645) = 0.8529
Example - Student Retention

P-value:
The P-value for this test is 2 times the area under the z curve to the left of the computed z = -0.19.
P-value = 2(0.4247) = 0.8494

Conclusion:
Since P-value = 0.849 > 0.05 = \( \alpha \), the hypothesis \( H_0 \) is not rejected at significance level 0.05.
There is no evidence that the return rate is different for males and females.

Example

A consumer agency spokesman stated that he thought that the proportion of households having a washing machine was higher for suburban households than for urban households. To test to see if that statement was correct at the 0.05 level of significance, a reporter randomly selected a number of households in both suburban and urban environments and obtained the following data.

<table>
<thead>
<tr>
<th>Number surveyed</th>
<th>Number having washing machines</th>
<th>Proportion having washing machines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suburban</td>
<td>300</td>
<td>243</td>
</tr>
<tr>
<td>Urban</td>
<td>250</td>
<td>181</td>
</tr>
</tbody>
</table>

\( \pi_1 \) = proportion of suburban households having washing machines
\( \pi_2 \) = proportion of urban households having washing machines
\( \pi_1 - \pi_2 \) is the difference between the proportions of suburban households and urban households that have washing machines.

\( H_0: \pi_1 - \pi_2 = 0 \)
\( H_a: \pi_1 - \pi_2 > 0 \)
Example

Significance level: $\alpha = 0.05$

Test statistic:

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

Assumptions: The two samples are independently chosen random samples. Furthermore, the sample sizes are large enough since

$$n_1 p_1 = 243 \geq 10, \quad n_1(1-p_1) = 57 \geq 10$$
$$n_2 p_2 = 181 \geq 10, \quad n_2(1-p_2) = 69 \geq 10$$

Calculations:

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{243 + 181}{300 + 250} = \frac{424}{550} = 0.7709$$

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} = \frac{0.810 - 0.742}{\sqrt{0.7709(1-0.7709)\left(\frac{1}{300} + \frac{1}{250}\right)}}$$

$$= 2.390$$

P-value:

The P-value for this test is the area under the z curve to the right of the computed $z = 2.39$.

The P-value = 1 - 0.9916 = 0.0084

Conclusion:

Since P-value = 0.0084 < 0.05 = $\alpha$, the hypothesis $H_0$ is rejected at significance level 0.05. There is sufficient evidence at the 0.05 level of significance that the proportion of suburban households that have washers is more that the proportion of urban households that have washers.
Large-Sample Confidence Interval for $\pi_1 - \pi_2$

When

1. The samples are independently selected random samples OR treatments that were assigned at random to individuals or objects (or vice versa), and
2. Both sample sizes are large:
   
   $n_1 p_1 \geq 10, n_1(1-p_1) \geq 10, n_2 p_2 \geq 10, n_2(1-p_2) \geq 10$

A large-sample confidence interval for $\pi_1 - \pi_2$ is

$$(p_1 - p_2) \pm (z \text{ critical value}) \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Example

A student assignment called for the students to survey both male and female students (independently and randomly chosen) to see if the proportions that approve of the College’s new drug and alcohol policy. A student went and randomly selected 200 male students and 100 female students and obtained the data summarized below.

<table>
<thead>
<tr>
<th></th>
<th>Number surveyed</th>
<th>Number that approve</th>
<th>Proportion that approve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>100</td>
<td>43</td>
<td>0.430</td>
</tr>
<tr>
<td>Male</td>
<td>200</td>
<td>61</td>
<td>0.305</td>
</tr>
</tbody>
</table>

Use this data to obtain a 90% confidence interval estimate for the difference of the proportions of female and male students that approve of the new policy.

$$(0.430 - 0.305) \pm 1.645 \sqrt{\frac{0.430(1-0.430)}{100} + \frac{0.305(1-0.305)}{200}}$$

$$(0.125) \pm 0.097 \quad \text{or} \quad (0.028, 0.222)$$

Based on the observed sample, we believe that the proportion of females that approve of the policy exceeds the proportion of males that approve of the policy by somewhere between 0.028 and 0.222.
11.5: Communicating & Interpreting Results

Like when working with the one sample hypothesis test, you should include:
• all hypotheses
• test procedure used
• value of test statistic
• p-value, and a conclusion in context.

Typically for CI’s, the key with interpretation is whether the interval includes zero.

Look For’s

• Are ONLY 2 groups being compared? If more groups are compared it’s a more complicated test.
• Were the samples independent or dependent?
• What Hypotheses were being tested
  • Were they one sided or two sided?
• Were all assumptions met?
• What p-value was associated with the test?
  • Reject or Fail to reject?
• Were conclusions in context & consistent with test results?
  If null was rejected, was there also a practical/clinical significance?

Cautions & Limitations

• Remember the results of a hypothesis test can never show support for the null.
  • So for a 2 sample test, this means we can never show the means are the same, only that they are different.
• If you have population data, don’t do an inference test or a CI.
• Don’t confuse statistical significance with practical/clinical significance.
• Make sure the data were collected appropriately for the test used (independent vs dependent).