Lesson 17

Inferences on the Difference Between Two Population Means

Learning Objectives

Upon completion of this lesson you should be able to do the following:

1. Use the difference between two sample means as a point estimator for the difference between two population means.

2. Generate the standard error of the point estimator for the difference between the means when assuming equal population variances and when not assuming equal population variances.

3. Construct a pooled variance estimate to estimate the variances of two populations under the assumption of equal population variances.

4. Construct interval estimates, called confidence intervals, to estimate the difference between two populations means, when assuming equal population variances and when not assuming equal population variances.

5. Test hypothses concerning the difference between two population means, when assuming equal population variances and when not assuming equal population variances.

Key Words

difference parameter, pooled variance estimator

Concepts

Earlier lessons dealt with estimating the mean of an individual population and testing hypotheses concerned with the mean of a population. This lesson deals with comparing the means of two populations. The population means are compared by considering their difference. The difference between two parameters is called a **difference parameter.** Estimating the difference between two means and testing a certain value for the difference between two means are the two major procedures discussed in this lesson.

Parameter is $\mu_1 - \mu_2$.

Point Estimator for the parameter is $\overline{X}_1 - \overline{X}_2$.

The point estimator for the difference between two population means, $\mu_1 - \mu_2$, is the difference between two sample means, $\overline{X}_1 - \overline{X}_2$. The point estimator constructed from the difference between the sample means is an excellent point estimator for the difference between the population means in the sense that it is unbiased and has minimum variance in the set of unbiased estimators. Recall that to say an estimator is unbiased means that the expected value or mean of the estimator is equal to the parameter being estimated. On average the difference between the sample means is equal to the difference between the population means. But, this point estimator has the same characteristic shared by all point estimators. A single observed value of the point estimator is not equal to the parameter being estimated. The standard error of a point estimator is a critical piece of information about the mistake made by the estimate. The standard error indicates the magnitude of the error associated with the point estimator based on a certain sample size.

Recall that the standard error for one sample mean is based on the ratio of the population standard deviation to the square-root of the sample size. The variance of the sample mean, which is the square of the standard error, is the ratio of the population variance to sample size. The standard error of the point estimator that is constructed from the difference between two sample means has two similar ratios involved. The ratios of the population variance divided by sample size are added together to form the variance of the difference between the sample means. The square-root of the sum of these two ratios is the standard error for the difference between the sample means.

Standard Error of $\overline{X}_1 - \overline{X}_2$ when the population variances, σ_1^2 and σ_2^2, are known is $\sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

Assume two random samples are drawn, one from each of two independent populations with means of μ_1 and μ_2, and variances of σ_1^2 and σ_2^2. The point estimator, $\overline{X}_1 - \overline{X}_2$, can have at least an approximately normal distribution either by large samples and CLT or by assuming the sampled populations are normally distributed. If a normally distributed sample statistic like the point estimator, $\overline{X}_1 - \overline{X}_2$, is standardized by substracting off the mean of the sampling distribution of the point estimator and dividing by the standard error of the point estimator then the resulting form has the standard normal distribution.

The form $\frac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sigma_{\overline{X}_1-\overline{X}_2}}$ has a standard normal distribution, or in other words the Z distribution, if variable $\overline{X}_1 - \overline{X}_2$ is normally distributed. In symbols it could be written that $Z = \frac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sigma_{\overline{X}_1-\overline{X}_2}} = \frac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \backsim N(0,1)$.

When the population variances, σ_1^2 and σ_2^2, are unknown how is the estimated standard error of the point estimator $\overline{X}_1 - \overline{X}_2$ constructed? There are two ways to estimate the standard error for the difference between the sample means. The population variances in the standard error equation can be estimated separately with the individual sample variances. Or, if the two population variances are assumed to be equal then the population variances can both be estimated with the same estimator, called a pooled variance estimator, which is a weighted average of the sample variances.

The estimated standard error of $\overline{X}_1 - \overline{X}_2$ is $\hat{\sigma}_{\overline{X}_1-\overline{X}_2} = S_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

when the population variances are not assumed equal and are estimated with the individual sample variances.

Notice that S_1^2 is used to estimate σ_1^2 and S_2^2 is used to estimate σ_2^2. When this form is used the assumption of equal population variances is not required. The population variances are estimated with different sample variances, so there is no need to assume that the population variances, σ_1^2 and σ_2^2 are equal.

The estimated standard error for the point estimator $\overline{X}_1 - \overline{X}_2$ can be constructed using a single estimator for both of the population variances, but it must be assumed that the population variances are equal. It is not reasonable to estimate two parameters with the same point estimator unless the parameters are believed to be equal. If it is assumed that the two population variances are equal then a pooled variance estimator is constructed to use as the estimator for both population variances.

The estimated standard error of $\overline{X}_1 - \overline{X}_2$is $\hat{\sigma}_{\overline{X}_1-\overline{X}_2} = S_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$

when the population variances are assumed equal and are both estimated with a single pooled variance estimator .

Both of the population variances, σ_1^2 and σ_2^2, were estimated with the pooled variance estimator, S_p^2, which is a weighted average of the two sample variances, weighted by their degrees of freedom.

The equation for the pooled variance estimator is $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$.

When the point estimator $\overline{X}_1 - \overline{X}_2$ is standardized to calculate a measure of relative standing for the observed point estimate value using the estimated standard error rather than the standard error based on the population variances what is the effect on the distribution of this standardized form? If the estimated standard error is used in the demoninator what is the set of resulting values and what is their probability structure?

If the pooled variance estimator is used to estimate both population variances then the standardized form of the normally distributed point estimator $\overline{X}_1 - \overline{X}_2$ has a t distribution with $n_1 + n_2 - 2$ degrees of freedom. In symbols it could be written that

$$t = \frac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{S_{\overline{X}_1-\overline{X}_2}} = \frac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}}} \sim t(n_1 + n_2 - 2).$$

If the individual sample variances are used to estimate the individual population variances then the standardized form of the normally distributed point estimator $\overline{X}_1 - \overline{X}_2$

has an approximate t distribution with approximate degrees of freedom calculated by the Satterthwaite degrees of freedom equation shown below. In symbols it could be written that

$$t^* = \frac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{S_{\overline{X}_1-\overline{X}_2}} = \frac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \backsim_{\text{approximately}} t(df), \text{ where } df \approx \frac{\left[\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right]^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}.$$

The distributions of these standardized forms are very important. If the null hypothesis is true then the distribution of the test statistic is known. This distribution is the basis for measuring the validity of the null hypothesis statement in a hypothesis test. The standardized forms of the point estimator provide measures of relative standing for the observed point estimate in the known distribution under a true null hypothesis. The standardized forms are the test statistics for the hypothesis tests on the difference between two population means. The distribution of the test statistics are based on the two distributions shown above.

$(1-\alpha)100\%$ Confidence Interval to Estimate the Difference Between Two Population Means when the population variances are assumed to be equal.

$$(\overline{X}_1 - \overline{X}_2) \pm t_{\frac{\alpha}{2},n_1+n_2-2} S_{\overline{X}_1-\overline{X}_2} \rightarrow (\overline{X}_1 - \overline{X}_2) \pm t_{\frac{\alpha}{2},n_1+n_2-2} \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

When testing $H_o : \mu_1 - \mu_2 = D_0$ with the assumption of equal population variances the test statistic has a t distribution with $n_1 + n_2 - 2$ degrees of freedom when the null hypothesis is true. The test statistic is:

$$t = \frac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{S_{\overline{X}_1-\overline{X}_2}} = \frac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}}} \backsim t(n_1+n_2-2) \text{ if } H_o \text{ is true.}$$

Although the above form is prevalently used to test the difference between the populations means the assumption of equal population varinaces is very strict. If this assumption is not valid then the above test statistic does not have the distribution indicated. Recent research indicates that it is probably better to use the processes indicated below for the confidence interval on the difference between two population means and the hypothesis test on the difference parameter since it is unlikely that the two population variances are equal. Furthermore, the researcher never knows for certain if the variances are equal or not, so the assumption of equal population variances is probably best avoided by using the processes shown next, which do not required the assumption of equal population variances.

$(1-\alpha)100\%$ Confidence Interval to Estimate the Difference Between Two Population Means when no assumption is made about the equality of the variances.

$$(\overline{X}_1 - \overline{X}_2) \pm t_{\frac{\alpha}{2},df} S_{\overline{X}_1 - \overline{X}_2} \rightarrow (\overline{X}_1 - \overline{X}_2) \pm t_{\frac{\alpha}{2},df} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

When testing $H_o : \mu_1 - \mu_2 = D_0$ the test statistic in this case has an approximate t distribution with degrees of freedom based on the equation shown when the null hypothesis is true. The test statistic is:

$$t^* = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_{\overline{X}_1 - \overline{X}_2}} = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \backsim_{\text{approximately}} t(df) \text{ if } H_o \text{ is true, } df \approx \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}.$$

Recall that if the degrees of freedom is adequately large, in practice $n > 30$, then both of the above test statistic forms can be reasonably well estimated with the Z-distribution. The hypothesis test on the difference between two population means is organized in the same manner as the other hypothesis tests that have been considered. A value is hypothesized for the difference parameter in the hypothesis statements, denoted as D_o. The sample data are analyzed to determine if significant evidence against the null hypothesis exists. The decision process follows the same criteria as earlier hypothesis tests.

Interpreting a Confidence Interval on a Difference Parameter

The confidence interval for the difference between two means can be interpreted in a similar manner as the earlier confidence intervals covered. The set of values listed in the confidence interval represents a set of reasonable and plausible values for the parameter. The equation provides an interval that does contain the value of the parameter (1-α) 100% of the time. Whether the interval generated from the data contains the actual parameter value depends on whether an extreme value has occurred for the difference between the sample means.

Whether the confidence interval to estimate the difference between the population means contains zero or not is of particular interest. If the interval contains zero then there is no significant evidence to indicate that the population means are unequal. A confidence interval on a difference parameter can be used to judge the relationship between the two parameters involved in the difference. The numeric result of a confidence interval on a difference parameter will be in one of the three following categories. What information each interval relates is indicated.

1. (positive number, positive number) would indicate that the first number in the difference is larger than the second.
2. (negative number, negative number) would indicate that the second number in the difference is larger than the first.
3. (negative number, positive number) would indicate no evidence against equality for the first and second value in the difference.

Summary Equations for Lesson 17

Parameter is $\mu_1 - \mu_2$.

Point Estimator for $\mu_1 - \mu_2$ is $\overline{X}_1 - \overline{X}_2$.

Standard Error of $\overline{X}_1 - \overline{X}_2$ when the population variances, σ_1^2 and σ_2^2, are known is $\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

Estimated standard error of $\overline{X}_1 - \overline{X}_2$ is $\hat{\sigma}_{\overline{X}_1 - \overline{X}_2} = S_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$ when the population variances, σ_1^2 and σ_2^2, are not assumed equal and are estimated with the individual sample variances.

Estimated standard error of $\overline{X}_1 - \overline{X}_2$ is $\hat{\sigma}_{\overline{X}_1 - \overline{X}_2} = S_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$ when the population variances, σ_1^2 and σ_2^2, are assumed equal and are both estimated with a single pooled variance estimator.

$(1-\alpha)100\%$ Confidence Interval to Estimate the Difference Between Two Population Means when the population variances are assumed to be equal.

$$(\overline{X}_1 - \overline{X}_2) \pm t_{\frac{\alpha}{2}, n_1+n_2-2} S_{\overline{X}_1-\overline{X}_2} \rightarrow (\overline{X}_1 - \overline{X}_2) \pm t_{\frac{\alpha}{2}, n_1+n_2-2} \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}.$$

$(1-\alpha)100\%$ Confidence Interval to Estimate the Difference Between Two Population Means when no assumption is made about the equality of the variances.

$$(\overline{X}_1 - \overline{X}_2) \pm t_{\frac{\alpha}{2}, df} S_{\overline{X}_1-\overline{X}_2} \rightarrow (\overline{X}_1 - \overline{X}_2) \pm t_{\frac{\alpha}{2}, df} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}.$$

Test Statistic to Test $H_o : \mu_1 - \mu_2 = D_0$ when the population variances are assumed equal

$$t = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_{\overline{X}_1-\overline{X}_2}} = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}} \backsim t(n_1 + n_2 - 2) \text{ if } H_o \text{ is true.}$$

Test Statistic To Test $H_o : \mu_1 - \mu_2 = D_0$ without the assumption of equal population variances

$$t^* = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_{\overline{X}_1-\overline{X}_2}} = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \backsim_{\text{approximately}} t(df) \text{ if } H_o \text{ is true,}$$

$$\text{where } df \approx \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$$

DETAILS FOR INFERENCES ON THE DIFFERENCE BETWEEN TWO MEANS

Not Assuming Equal Variances **BASED ON LARGE SAMPLES**

HYPOTHESIS TEST

1. **State the set of hypotheses.**

   $H_o: \mu_1 - \mu_2 = D_o$

   $H_a: \mu_1 - \mu_2 \neq D_o$ or $H_a: \mu_1 - \mu_2 < D_o$ or $H_a: \mu_1 - \mu_2 > D_o$

2. **Take two large independent random samples. Calculate the test statistic.**

   $$z = \frac{\bar{x}_1 - \bar{x}_2 - D_o}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{x}_1 - \bar{x}_2 - D_o}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

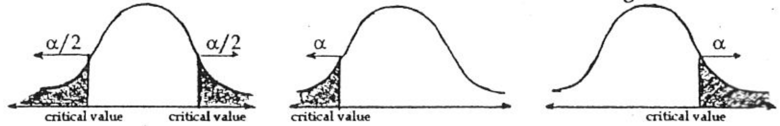
3. **Identify the distribution of the test statistic.**

   The variable $Z = \dfrac{\bar{X}_1 - \bar{X}_2 - D_o}{S_{\bar{x}_1 - \bar{x}_2}} \sim \mathrm{N}(0,1)$ if $\mathrm{H_o}$ is true.

4. **Generate the rejection region or the p-value to make the decision of whether to reject the null hypothesis or not.**

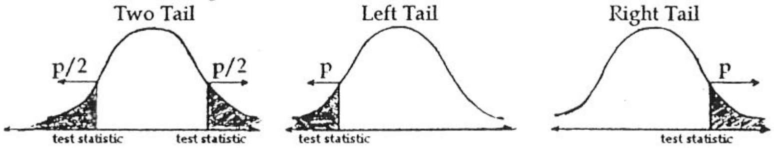
   Rejection Region

   Two Tail | Left Tail | Right Tail



   P-value or Observed Significance Level

   Two Tail | Left Tail | Right Tail



5. **Form a conclusion in words.** If the null hypothesis is rejected then the data do support the alternative hypothesis; if the null hypothesis is not rejected then the data do not support the alternative hypothesis.

(1-α) 100% CONFIDENCE INTERVAL

$$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2(\text{high df})} \cdot s_{\bar{x}_1 - \bar{x}_2} \Rightarrow \bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$ The high df t is estimated with z.

DETAILS FOR INFERENCES ON THE DIFFERENCE BETWEEN TWO MEANS

Assuming Equal Variances BASED ON SMALL SAMPLES

HYPOTHESIS TEST

1. **State the set of hypotheses.**

$$H_o : \mu_1 - \mu_2 = D_o$$

$$H_a : \mu_1 - \mu_2 \neq D_o \text{ or } H_a : \mu_1 - \mu_2 < D_o \text{ or } H_a : \mu_1 - \mu_2 > D_o$$

2. **Take two small independent random samples. Calculate the test statistic.**

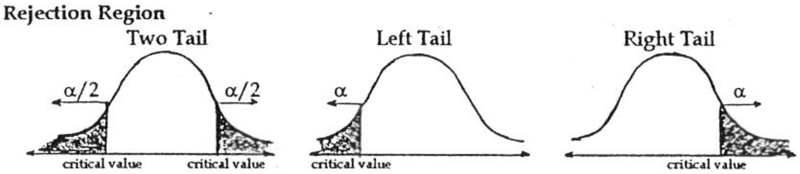
$$t = \frac{\bar{x}_1 - \bar{x}_2 - D_o}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{x}_1 - \bar{x}_2 - D_o}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}, \text{ where } s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$$

3. **Identify the distribution of the test statistic.**

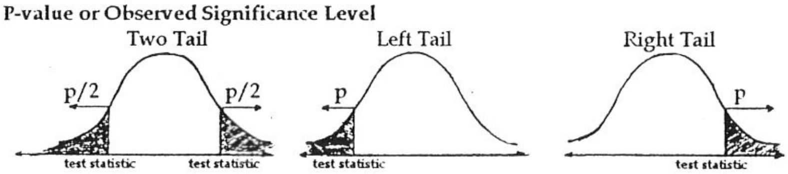
$$\text{The variable } t = \frac{\bar{X}_1 - \bar{X}_2 - D_o}{S_{\bar{x}_1 - \bar{x}_2}} \sim t(n_1 + n_2 - 2) \text{ if } H_o \text{ is true. Small n, assume } X_i \sim N.$$

4. **Generate the rejection region or the p-value to make the decision of whether to reject the null hypothesis or not.**

Rejection Region



P-value or Observed Significance Level



5. **Form a conclusion in words.** If the null hypothesis is rejected then the data do support the alternative hypothesis; if the null hypothesis is not rejected then the data do not support the alternative hypothesis.

(1-α) 100% CONFIDENCE INTERVAL

$$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2 (n_1 + n_2 - 2)} \cdot s_{\bar{x}_1 - \bar{x}_2} \Rightarrow \bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2 (n_1 + n_2 - 2)} \cdot \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}.$$ If $n_1 + n_2 - 2 > 30$ then est t with z.

Example

Confidence Interval on the Difference Between Two Means Assuming Equal Variances

An automotive manufacturer is interested in the difference between the mean operating cost per thousand miles for cars with rotary engines and cars with standard engines. Two small independent samples produced the following summary statistics on cost:

Rotary	Standard
$n_1 = 8$	$n_2 = 12$
$\overline{x}_1 = \$56.9$	$\overline{x}_2 = \$52.73$
$s_1 = \$4.85$	$s_2 = \$6.35$

Construct a 90% confidence interval to estimate $\mu_1 - \mu_2$.

$$\overline{x}_1 - \overline{x}_2 \pm t_{\frac{\alpha}{2},(n_1+n_2-2)} s_{\overline{x}_1-\overline{x}_2}$$

$$\overline{x}_1 - \overline{x}_2 \pm t_{\frac{.10}{2},(8+12-2)} \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}, \text{ where } s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}.$$

$$\overline{x}_1 - \overline{x}_2 \pm t_{\frac{.10}{2},(18)} \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\overline{x}_1 - \overline{x}_2 \pm t_{.05,(18)} \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$56.96 - 52.73 \pm 1.734\sqrt{\frac{(8-1)4.85^2+(12-1)6.35^2}{8+12-2}\left(\frac{1}{8} + \frac{1}{12}\right)}$$

$$4.23 \pm 1.734\sqrt{33.789\left(\frac{1}{8} + \frac{1}{12}\right)}$$

$$4.23 \pm 1.734(2.6532)$$

$$4.23 \pm 4.6006$$

$$(-0.3706, 8.8306)$$

The difference in average operating cost per thousand miles between cars with rotary engines and those with standard engines is between –\$0.37 and \$8.83. This interval is associated with a 90% confidence level. Notice that the value of zero is contained in the interval.

The interval represents a set of plausible and reasonable values for the parameter, the difference between the population means. Any value contained in the confidence interval would not be rejected if it were tested as the value in a null hypothesis in a two tail test where the signficance level is the same as the confidence interval. The next example tests whether cars with rotary engines and cars with standard engines have the same operating costs. The confidence interval in this example could also be used to answer that question. Since zero is contained in the confidence interval, it is possible that $\mu_1 = \mu_2$. If a possible value for the difference between the means is zero, then it is plausible to think that the means might be equal . If a hypothesis test on is performed with $\alpha = 0.10$, the same error rate as in the prior confidence interval the null hypothesis of equal population means will not be rejected. That is examined in the next example.

Example

Hypothesis Test on the Difference Between Two Means Assuming Equal Variances

An automotive manufacturer is interested in the difference between the mean operating cost of cars with rotary engines and cars with standard engines. View the summary statistics for this example in the prior example. Do the data provide evidence that there is a difference in the mean operating cost of the two types of cars? The parameter in this problem is the difference between two population means and the parameter value of interest is 0 since we are checking to see if the two populations means are equal. Checking equality of means is the same as checking if $\mu_1 - \mu_2 = 0$. Recall that the confidence interval constructed from these data to estimate the difference between the means does contain the value zero. The set of hypotheses that are needed for this problem are:

$$H_o : \mu_1 - \mu_2 = 0$$

$$H_a : \mu_1 - \mu_2 \neq 0$$

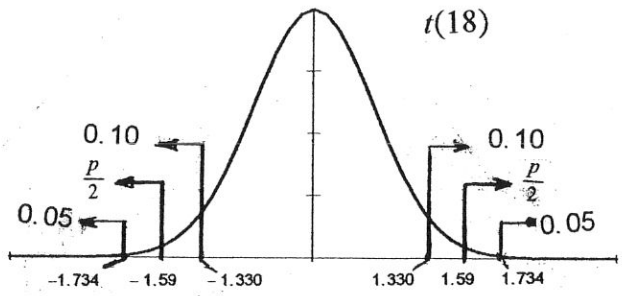
The test statistic is

$$t = \frac{\bar{x}_1 - \bar{x}_2 - D_o}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{x}_1 - \bar{x}_2 - D_o}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2 - D_o}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\bar{x}_1 - \bar{x}_2 - D_o}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} =$$

$$= \frac{56.96 - 52.73 - 0}{\sqrt{\frac{(8-1)4.85^2 + (12-1)6.35^2}{8+12-2}\left(\frac{1}{8} + \frac{1}{12}\right)}} = \frac{4.32}{\sqrt{33.789\left(\frac{1}{8} + \frac{1}{12}\right)}} = \frac{4.23}{2.6532} = 1.5943.$$

What is the distribution of this test statistic if the null hypothesis is true? The student's *t*-distribution with $(n_1-1)+(n_2-1)=n_1+n_2-2=18$ degrees of freedom is the set of all possible values that occur for the test statistic if the null hypothesis is true. If the test statistic value is unlikely in this distribution, then that is evidence that the null hypothesis is false. If the test statistic value is likely in this distribution, then that is lack of evidence that the null hypothesis is false. The *p*-value or OSL is the probability of observing a test statistic at least as extreme as the one observed in the data. The *p*-value or OSL is the tail areas associated with the test statistic and is calculated in the following way:

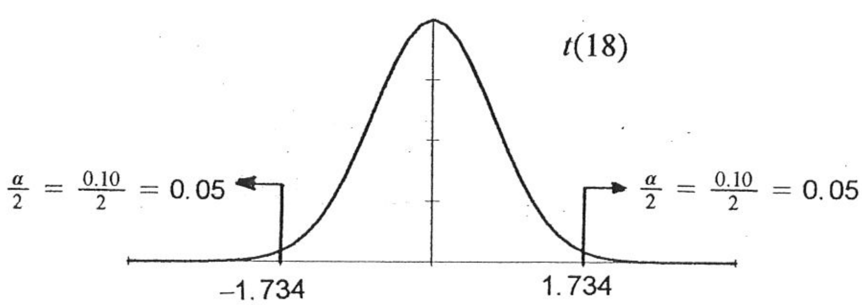
$P = P(t < -1.5943) + P(t > 1.5943)$.



Two-tail p-value, area left of -1.5943 and area right of 1.5943.

From the graph it can be seen that $0.05 < \frac{p}{2} < 0.10 \rightarrow 0.10 < p < 0.20$. If the null hypothesis is rejected the observed error rate that must be tolerated is between 10 to 20%. Remember, the p-value is the error rate that must be tolerated is the null hypothesis is rejected. When the p-value is so large there is too much chance of being wrong when the null is rejected. Do not reject the null hypothesis. Conclude that these two samples do not provide significant evidence that there is a difference between the mean operating cost of cars with rotary engines and cars with standard engines. Notice that is the same conclusion that was reached with the confidence interval. The interval contained zero so the conclusion was that the means may not differ. The null hypothesis of equal means was not rejected so it is concluded that the evidence indicates there is not sufficient evidence to say that the means are different.

Another way is to make the rejection decision is to identify the rejection region and recognize whether or not the test statistic value is in the rejection region. The rejection region for this test is illustrated below. Note that the test statistic value does not fall in the unlikely test statistic values identified by the rejection region. This method will result in the same conclusion as the p-value method to not reject the null hypothesis. We can use either method to reach our rejection conclusion.
To set up the rejection region consider the two-sided alternative hypothesis and = 0.10. The rejection region is below –1.734 and above +1.734 since these values cutoff 0.05 in each tail for a total of 0.10, the -value.



Rejection Region, two-tail test, $\alpha = 0.10$.

The test statistic value, 1.594, is not in the rejection region, so the null hypothesis is not rejected. Conclude that these two samples do not provide evidence to claim that the mean operating costs of cars with rotary engines and cars with regular engines are different.

Make sure you understand how the decision in this hypothesis test with $\alpha = 0.10$ to not reject the null hypothesis of equal means relates to the 90% confidence interval generated in the prior example. Since the value of zero was contained in the confidence interval to estimate the difference between the means the hypothesized difference of zero was not rejected in the hypothesis test. The confidence interval provides a set of values that would not be rejected in a two-tail hypothesis test with the same significance level.

In this hypothesis test, we were assuming that the variances of the two populations were the same. It is only valid to use a pooled variance estimator to estimate the variances of both populations if the variances of the two populations are equal.

Exercise with Answers

Inferences on the Difference Between Two Population Means Not Assuming Equal Variances

Assume that you are the manager of a furniture store in Tulsa. Your store currently carries two lines of custom-built lawn furniture; you want to limit that to one. In an attempt to choose the line of furniture on which your customers spend more on average, the mean order price for the two lines is being compared. One hundred orders from each furniture line were sampled and the price of each order was recorded. The two samples resulted in the following sample statistics.

$\overline{x}_1 = \$780 \qquad \overline{x}_2 = \$810$

$s_1^2 = \$1400 \qquad s_2^2 = \$2200$

1. If there is no difference between the two population means then the set of all possible values for the point estimator $\overline{X}_1 - \overline{X}_2$. is a known set which is called the sampling distribution of the point estimator. State the sampling distribution for the difference between the sample means if there is no difference between the two population means of the order prices for the two lines of furniture. Estimate the standard error of the difference between the sample means from the sample data. A blank graph is provided for you to graph the sampling distribution.

2. Do the data indicate that the mean order prices for the two lines of custom-built lawn furniture are different? State the set of hypotheses that would be used to address this question.

3. Give the value of the test statistic, which would occur from the data, to test whether or not the mean order prices are the same. State the equation; show the work; state the value.

4. Calculate the p-value that would result from these data. State the probability statement that describes the p-value. Draw a graph, label the value of the test statistic on the graph and shade the area that represents the p-value.

5. Based on the p-value of the test what is the decision about the validity of the null hypothesis? Based on the p-value of the test what is the conclusion about the mean price of the orders for the two lines of furniture? Answer the questions with complete sentences.

6. Which line of custom-built lawn furniture do you think your store should continue to carry? Why? Why couldn't this decision be reached without the above hypothesis test since one sample mean is larger than the other sample mean? Answer the questions with complete sentences.

7. Construct a 95% confidence interval to estimate the difference between the mean order prices for the two lines of custom-built furniture. Give the basic equation, show all work, state the interval estimate. Write a one-sentence description of the interval estimate. Notice that the interval does not contain the value 0. How does that information agree with the conclusion of the above hypothesis test?

8. What does it mean for the confidence interval to be associated with 95% confidence?

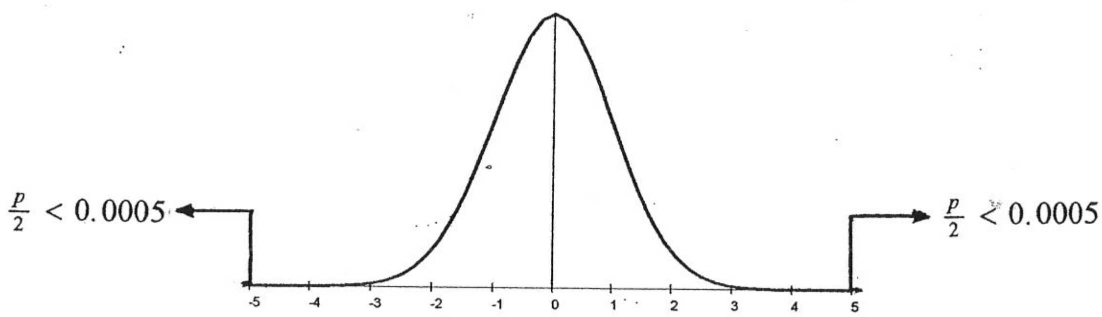
Answers

1. $\overline{X}_1 - \overline{X}_2 \backsim N[0, (\frac{1400}{100} + \frac{2200}{100})]$ and $s_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{1400}{100} + \frac{2200}{100}} = 6.$

2. $Ho: \mu_1 - \mu_2 = 0$ vs $Ha: \mu_1 - \mu_2 \neq 0$

3. $z = \frac{\overline{x}_1 - \overline{x}_2 - D_o}{s_{\overline{x}_1 - \overline{x}_2}} = \frac{\overline{x}_1 - \overline{x}_2 - D_o}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{780 - 810 - 0}{\sqrt{\frac{1400}{100} + \frac{2200}{100}}} = \frac{-30}{6} = -5$

4. $p - value = P(Z < -5) + P(Z > 5) < 0.0005 + 0.0005 = .001.$



p-value<0.001 for the two-tail test

5. The p-value of the hypothesis test is less than 0.001. This means there is less than a.1% chance of observing the sample data that were drawn if the population means are really equal. The p-value of less than.001 indicates that there is less than a.1% chance of error if the decision is to reject Ho and conclude that the two lines of custom-built lawn furniture have different mean order prices. Assuming that an error rate of less than.1% is tolerable reject Ho and conclude that the two lines of custom-built lawn furniture have different mean order prices.

6. Assuming the store wants to carry the line of furniture that has the higher mean order price, the choice should be line 2. This line would be chosen over line 1 since the observed average of line 2 is greater than the observed average of line 1. This statement cannot be made prior to the hypothesis test since the conclusion of the hypothesis test confirms that the two sample means are adequately different to assume there is a difference at all between the population means. The sample means being unequal does NOT indicate that the population means are unequal. The population means must be shown to be unequal through a hypothesis test.

7. $\overline{x}_1 - \overline{x}_2 \pm z_{\frac{\alpha}{2}} s_{\overline{x}_1 - \overline{x}_2} \Rightarrow \overline{x}_1 - \overline{x}_2 \pm z_{\frac{0.05}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \Rightarrow 780 - 810 \pm 1.96\sqrt{\frac{1400}{100} + \frac{2200}{100}} \Rightarrow$

$$-30 \pm 1.96(6) \Rightarrow -30 \pm 11.76 \Rightarrow (-41.76, -18.24)$$

The set of plausible and reasonable values for the difference between the mean order prices of the two lines of furniture is between the values of -41.76 and -18.24 at the 95% confidence level. The value 0 is not contained in the 95% confidence interval to estimate the difference between the population means. This indicates that the means are different, since 0 is not a reasonable value for the difference. This agrees with the conclusion of the hypothesis test in number 5.

8. The equation used to calculate the confidence interval has a 95% chance of providing an interval that does contain the difference between the means. The interval is centered on the observed difference between the values of the two sample means. There is a 95% chance of observing a difference between the sample mean values that would center the interval adequately to contain the value of the difference between the population means.

Exercise with Answers

Inferences on the Difference Between Two Population Means Assuming Equal Population Variances

Assume that you are the manager of a financial institute. With the stock markets on the decline your institution is interested in increasing the amount of investment in bonds and wants to compare the average returns on two types of bonds. A sample of each type of bond, measured in percent quarterly return and recorded in decimal fractions, resulted in the following sample statistics. Solutions next page.

$n_1 = 8 \qquad n_2 = 10$

$\overline{x}_1 = 0.043 \quad \overline{x}_2 = 0.021$

$s_1^2 = 0.006 \quad s_2^2 = 0.008$

1. State the point estimate for the difference between the mean quarterly returns for these two types of bonds.

2. State the value of the pooled variance estimate based on the two samples of data.

3. State the estimated standard error of the point estimate for the difference between the mean quarterly returns for these two types of bonds.

4. Do the data indicate that the mean quarterly returns for the two types of bonds are different? State the set of hypotheses that would be used to address this question.

5. Give the value of the test statistic, which would occur from the data, to test whether or not the mean quarterly returns are the same. State the equation; show the work; state the value.

6. Approximate the p-value that would result from these data. State the probability statement that describes the p-value. Draw a graph, label the value of the test statistic on the graph and shade the area that represents the p-value.
7. Based on the p-value of the test what is the decision about the validity of the null hypothesis? Based on the p-value of the test what is the conclusion about the difference in mean returns for the two types of bonds? Answer the questions with complete sentences.
8. Should your financial institution have a preference of one bond type over the other? Why or why not? Why does this decision result when one sample mean is larger than the other sample mean? Answer the questions with complete sentences.
9. Construct a 95% confidence interval to estimate the difference between the mean returns for the two types of bonds. Give the basic equation, show all work, state the interval estimate. Write a one-sentence description of the interval estimate. Notice that the interval does contain the value 0. What does that indicate? Does that information agree with the conclusion of the above hypothesis test?

Answers for the Exercise

1. $\overline{x}_1 - \overline{x}_2 = 0.043 - 0.021 = 0.022.$

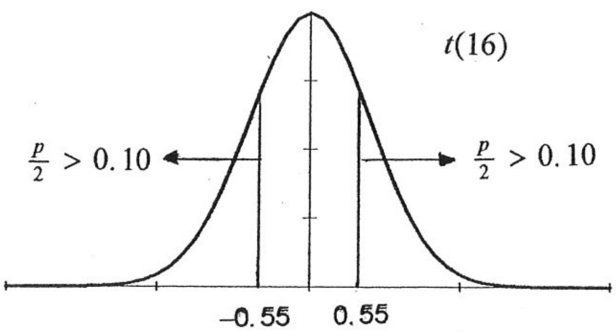
2. $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = \frac{7(0.006)+9(0.008)}{16} = \frac{0.114}{16} = 0.007125.$

3. $s_{\overline{x}_1-\overline{x}_2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} = \sqrt{\frac{0.007125}{8} + \frac{0.007125}{10}} = 0.040039$

4. $Ho : \mu_1 - \mu_2 = 0$ vs. $Ha : \mu_1 - \mu_2 \neq 0$

5. $t = \frac{\hat{x}_1-\hat{x}_2-D_o}{s_{\hat{x}_1-\hat{x}_2}} = \frac{\hat{x}_1-\hat{x}_2-D_o}{\sqrt{\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}}} = \frac{\hat{x}_1-\hat{x}_2-D_o}{\sqrt{s_p^2(\frac{1}{n_1}+\frac{1}{n_2})}} = \frac{0.022}{\sqrt{0.007125(\frac{1}{8}+\frac{1}{10})}} = \frac{0.022}{0.040039} = 0.54946 \approx 0.55.$

6. $p-value = P[t(16) < -.55] + P[t(16) > .55] > 0.10 + 0.10 = 0.2 \rightarrow p-value > 0.2.$



p-value>0.2 for the two-tail test

7. The p-value of the hypothesis test is more than 20%, the null hypothesis should not be rejected. The error rate to be tolerated would be more than 20% if the null hypothesis were rejected. There is more than a 20% chance of observing sample means that are at least as far apart as what we observed in our samples if the population means are really equal. There is more than a 20% chance of observing data at least as extreme as our data even if the null hypothesis is true.

8. These two samples do not provide evidence that a difference exist between the mean returns for the two types of bonds. It would seem that one would want to choose the type of bond that has the higher sample mean, but the hypothesis test indicates that these sample means do not indicate a significant difference between the population means. The sample sizes are too small for the difference between the sample means to indicate a significant difference between the population means.

9. $\overline{x}_1 - \overline{x}_2 \pm t_{\frac{\alpha}{2},(n_1+n_2-2)} s_{\overline{x}_1-\overline{x}_2}$

$\overline{x}_1 - \overline{x}_2 \pm t_{\frac{.05}{2},(8+10-2)} \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$, recall $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$. from above.

$\overline{x}_1 - \overline{x}_2 \pm t_{.025,(16)} \sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}$

$0.043 - 0.021 \pm 2.12\sqrt{0.007125(\frac{1}{8} + \frac{1}{10})}$

$0.022 \pm 2.12(0.040039)$

$0.022 \pm 0.084883 \rightarrow (-0.063, 0.107)$.

The interval, (-.063,.107), represents a set of plausible and reasonable values for the difference between the mean quarterly returns for the two types of bonds at the 95% confidence level. The value 0 is contained in the 95% confidence interval to estimate the difference between the population means. This indicates that there is inadequate evidence to claim that the means are different. This agrees with the conclusion of the hypothesis test.

Exercise

Inferences on the Difference Between Two Population Means Work Two Ways: Assume Equal Variances and Do Not Assume Equal Variances

Assume that you are the owner of a construction company in Enid. Your company has to lay many miles of underground cable at your construction sites each year. There are two possible trenching methods and your company is interested in comparing the mean cost per foot of trench for each method. In an attempt to choose the most cost efficient method four hundred feet of each type of method were sampled and the average cost per foot for each method was recorded. The two samples resulted in the following sample statistics.

$\overline{x}_1 = 0.62 \quad \overline{x}_2 = 0.58$

$s_1^2 = 0.40 \quad s_2^2 = 0.60$

1. State the sampling distribution for the difference between the sample means if there is no difference between mean cost of the two trenching methods. Estimate the standard error of the difference between the sample means from the sample data. Graph this sampling distribution.
2. Do the data indicate that the mean trenching price per foot for the two methods are different? State the set of hypotheses that would be used to address this question.
3. Give the value of the test statistic, which would occur from the data, to test whether or not the mean trenching costs are the same. State the equation; show the work; state the value.
4. Calculate the p-value that would result from these data. State the probability statement that describes the p-value. Draw a graph, label the value of the test statistic on the graph and shade the area that represents the p-value.
5. Based on the p-value of the test what is the decision about the validity of the null hypothesis? Based on the p-value of the test what is the conclusion about the difference in the mean cost per foot for the two trenching methods? Answer the questions with complete sentences.
6. Which trenching method, one or two, do you think your company should use? Or, is the difference between the cost of the two methods so small it doesn't matter which method you use? Why couldn't this decision be reached without the above hypothesis test since one sample mean is larger than the other sample mean? Answer the questions with complete sentences.
7. Construct a 95% confidence interval to estimate the difference between the mean cost per foot of the two trenching methods. Give the basic equation, show all work, state the interval estimate. Write a one-sentence description of the interval estimate. Notice that the interval does contain the value 0. What does that indicate? Does that information agree with the conclusion of the above hypothesis test?
8. What is meant by the comment that, "the confidence interval is associated with 95% confidence?" Answer in complete sentences.

Exercise

Inferences on the Difference Between Two Population Means Work Two Ways: Assume Equal Variances and Do Not Assume Equal Variances

Huge gypsum and limestone caverns and deposits exist in the extreme southwest corner of Oklahoma. This natural resource is utilized in a variety of ways including the recreational activity of spelunking, manufacturer of wallboard, and cut stone for building exteriors much like hard white sandstone. Assume that after you graduate from State you acquire a stone quarry between Eldorado and Olustee at Creta, which was once a railstop on a western cattle trail. Creta stone is a very valuable construction material, but is quite expensive to cut from the ground. Two methods for quarrying Creta stone exist, one a method where the saw is cooled with water and the other a method where the saw is cooled with air. The water-cooled method is expensive but is also more productive, yielding a higher quality of stone. You want to compare the average costs of the two methods to cut $1000 worth of stone. For each method 40 observations were randomly selected. Each observation was the cost to cut $1,000 worth of stone. The two samples resulted in the following means and standard deviations.

Water-Cooled	Air-Cooled
$\overline{x}_1 = \$285$	$\overline{x}_2 = \$264$
$s_1^2 = \$400$	$s_2^2 = \$600$

1. State the sampling distribution for the difference between the sample means if there is no difference between mean cost of the two cutting methods. Estimate the standard error of the difference between the sample means from the sample data. Graph this sampling distribution.
2. Do the data indicate that the mean cost of cutting $1,000 worth of stone is more for the water-cooled method than for the air-cooled method? State the set of hypotheses that would be used to address this question.
3. Give the value of the test statistic, which would occur from the data, to test whether the two cutting methods have the same average costs per $1,000 worth of stone. State the equation; show the work; state the value.
4. Calculate the p-value that would result from these data. State the probability statement that describes the p-value. Draw a graph, label the value of the test statistic on the graph and shade the area that represents the p-value. (Use the most extreme value in the t-table on the bottom row.)
5. Based on the p-value of the test what is the decision about the validity of the null hypothesis? Based on the p-value of the test what is the conclusion about whether the mean cost of the water-cooled cutting method is more than the mean cost of the air-cooled cutting method? Answer the questions with complete sentences.
6. Construct a 95% confidence interval to estimate the difference between the mean cost of the two cutting methods. Give the basic equation, show all work, state the interval estimate. Write a one-sentence description of the interval estimate. Notice that the interval does not contain the value 0. What does that indicate? Does that information agree with the conclusion of the hypothesis test?
7. The $21 difference in the two observed average costs per $1,000 worth of stone has been shown to be statistically significantly different from zero. Would you perhaps still want to use the water-cooled cutting method if you had a bigger market for the higher quality stone produced by that method? Comment on these issues in two complete sentences.

Exercise

Inferences on the Difference Between Two Population Means Work Two Ways: Assume Equal Variances and Do Not Assume Equal Variances

Two independent random samples were drawn from normally distributed populations. The resulting data are summarized below.

$n_1 = 11 \quad n_2 = 14$

$\overline{x}_1 = 543 \quad \overline{x}_2 = 355$

$s_1^2 = 66 \quad s_2^2 = 58$

1. State the point estimate for the difference between the means of the two populations.
2. State the value of the pooled variance estimate based on the two samples of data.
3. State the estimated standard error of the point estimate for the difference between the means of the populations.
4. Do the data indicate that the mean of population one is more than 175 units greater than the mean of population two?
5. Give the value of the test statistic, which would occur from the data, to test that the mean of population one is only 175 units greater than the mean of population two. State the equation; show the work; state the value.
6. Approximate the p-value that would result from these data. State the probability statement that describes the p-value. Draw a graph, label the value of the test statistic on the graph and shade the area that represents the p-value.
7. Based on the p-value of the test what is the decision about the validity of the null hypothesis? Based on the p-value of the test what is the conclusion about whether the mean of population one is more than 175 units greater than the mean of population two? Answer the questions with complete sentences.

Exercise

Inferences on the Difference Between Two Population Means Work Two Ways: Assume Equal Variances and Do Not Assume Equal Variances

As the production manager of a ballet company you are concerned about the increase in the average price of toe-shoes for ballerinas. The shoes are expensive but do not last very long in terms of numbers of shows in which they can be worn. Of the two types of shoes that the ballet company has used within the last six months statistics have been recorded concerning in how many shows each pair was used. The summary data are shown below.

$n_1 = 18 \quad n_2 = 24$

$\overline{x}_1 = 8.6 \quad \overline{x}_2 = 11.4$

$s_1^2 = 4.6 \quad s_2^2 = 5.8$

1. State the point estimate for the difference between the mean number of shows in which the two types of shoes can be used.
2. State the value of the pooled variance estimate based on the two samples of data.
3. State the estimated standard error of the point estimate for the difference between the mean number of shows in which each type of shoes were used.
4. Do the data indicate that the mean number of shows in which each type of shoes was used are different? State the set of hypotheses that would be used to address this question.
5. Give the value of the test statistic, which would occur from the data, to test whether or not the mean returns are the same. State the equation; show the work; state the value.
6. Approximate the p-value that would result from these data. State the probability statement that describes the p-value. Draw a graph, label the value of the test statistic on the graph and shade the area that represents the p-value.
7. Based on the p-value of the test what is the decision about the validity of the null hypothesis? Based on the p-value of the test what is the conclusion about the difference in the mean number of shows in which the two types of shoes could be used? Answer the questions with complete sentences.
8. Which type of shoes should your ballet company purchase for the ballerinas if the qualities of the two types are similar? Why? Why couldn't this decision be reached without the above hypothesis test since one sample mean is larger than the other sample mean? Answer the questions with complete sentences.
9. Construct a 95% confidence interval to estimate the difference between the mean number of shows in which the two types of shoes could be used. Give the basic equation, show all work, state the interval estimate. Write a one-sentence description of the interval estimate.