Unit 3

Inferential Statistics

Unit Objectives

Upon completion of this unit, you should be able to:

1. estimate population means by constructing confidence intervals;

2. investigate how sample size affects the precision of the estimate of the population mean;

3. introduce hypothesis testing procedures – an inferential procedure that uses sample data to draw conclusions about a population characteristic;

4. identify which type of hypothesis test should be performed given the number of populations and certain assumptions about the sample size and standard deviation; and

5. compute and interpret *p*-values.

Instructor's Notes

Confidence Intervals for One Population Mean

In Unit 2, we operated under the assumption that we knew the actual values of the population mean and the population standard deviation. Generally, this is not the case. In fact, more often than not, we attempt to use sample data to **estimate** characteristics of a population. For example, we could use the sample mean to estimate the population mean.

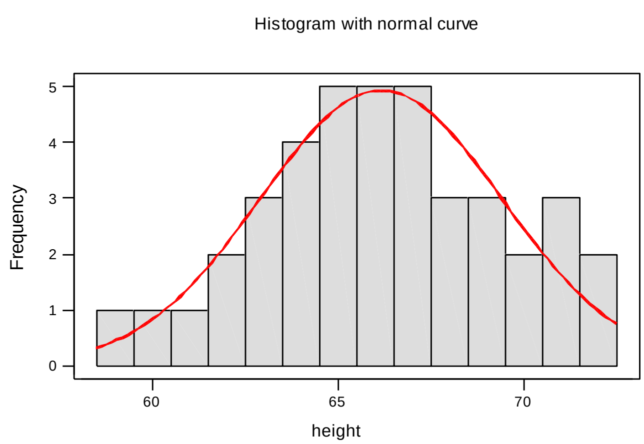
If we are interested in knowing the heights of all women in America over the age of 18, it would be too time consuming and costly to take a census. Instead, we are more likely to gather data from a random sample of women that is representative of the population. After collecting the data we could compute a sample mean and use that as an estimate of the population mean. Because of sampling error, we cannot expect that the population mean will actually equal our sample mean.

The sample mean is called a **point estimate** of the population mean. Unfortunately, point estimates are not the most useful estimates because they are expressed using only one number and do not allow for any room for error. How confident are we in using this sample mean to estimate the population mean? Can we develop a means of reporting information about the accuracy of our estimate? In order to address this concern, we will give a **confidence interval estimate** for the population mean. The basic idea is this: we will construct an interval of numbers using the sample mean and then state how confident we are that the population mean will lie inside the interval.

Example: Suppose that the population standard deviation for women's height is known to be 3.25. Consider the following sample of 40 women's heights.

64	72	65	62	67	62	71	64	65	67
68	66	67	66	64	71	65	70	67	66
63	69	63	72	69	71	60	65	67	66
69	65	59	61	66	63	64	68	68	70

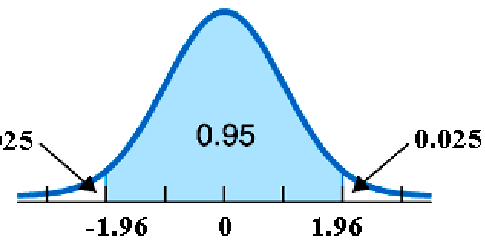
The sample mean is 66.175 inches with a sample standard deviation of 3.249. As previously mentioned, we can use the sample mean of 66.175 inches as a point estimate of the mean height for all American women. However, if we were to randomly select one woman from the population of all women and then measure her height it is very unlikely that she would be precisely 66.175 inches tall. In other words, the probability of her being 66.175 inches tall is extremely small. Recall from the previous unit that it would be more useful to try to compute the probability that her height falls within a range of specific values.



In order to apply some of the ideas of the previous unit, we must first verify that our sample is at least approximately normal. We can check this informally by looking at the histogram of the sample data. Below is a histogram for our sample data, along with a superimposed normal curve. We can see that the distribution of our sample is roughly bell-shaped. In other words, our sample of women's heights is approximately normally distributed. We may now feel justified in applying some of the techniques on normal distributions.

How confident are we that this sample mean of 66.175 represents the population mean? To answer this question we will build an interval **centered** around the sample mean of 66.175 and then express our level of confidence in the possibility of the population mean falling within that interval. Suppose we want the possibility of the population mean falling within the specified interval to be 95%. In other words, we want to say with 95% confidence that the interval contains the population mean.

To do this, we will build an interval around 66.175 inches that contains 95% of the area under the normal curve. Below is a graph of the standard normal distribution and the *z*-scores that bound 95% of the area (centered around the mean 0) under the normal curve.



Recall that a *z*-score represents the number of standard deviations away from the mean. Therefore, if we find two numbers that are each 1.96 standard deviations away from the mean of 66.175, then the area bound by those numbers under the normal curve will equal 95%. In other words, we need to take a *z*-score and convert it back to the same scale of our problem.

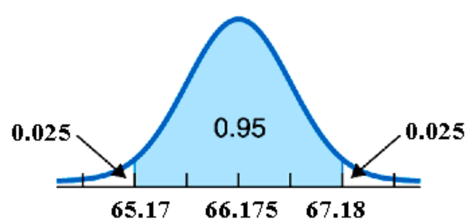
Recall, that to find the *z*-score that corresponds to *x* we subtract the mean and then divide by standard deviation. Therefore, working backwards to find *x* requires that we do the "opposite" operations in reverse order. In particular, to find *x* should involve multiplying *z* by the standard deviation and then adding the mean.

A word of caution: since we are dealing with a sample and we are using $\bar{x}$ to estimate μ, then we need to consider the sampling distribution of $\bar{x}$. It follows from the Central Limit Theorem that women's mean height $\bar{x}$ has a normal distribution with a mean of μ and a standard deviation of $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{3.25}{\sqrt{40}} = 0.514$.

So for our example we get the following:

for $z = -1.96$: $\quad x_1 = (-1.96)\cdot(0.514) + 66.175 = 65.17$

for $z = 1.96$: $\quad x_2 = (1.96)\cdot(0.514) + 66.175 = 67.18$



We have now constructed an interval for which we can estimate the population mean of women's heights. In particular, we are 95% confident that the population mean for women's height should fall between 65.17 inches and 67.18 inches.

The length of our confidence interval will increase if we decide to increase the level of confidence, α. Likewise, we can construct a confidence interval for any confidence level we want. The confidence level will determine the number of standard deviations we are willing to go away from the mean, i.e., the *z*-scores will change.

It is important to notice that at the beginning of this example we assumed that the value of the population standard deviation was known. Generally, this is not the case. When the population standard deviation is not known, it would be most desirable to substitute the sample standard deviation in its place. In order to do so, we will have to use a distribution other than the *z*-distribution.

When population standard deviation is unknown we will use the **student's *t*-distribution**. The *t*-distribution is a symmetric, bell-shaped curve. The *t*-curves differ from normal curves in that the distribution is more spread out. We will compute *t*-scores much like we do *z*-scores except we will use the sample standard deviation *s* instead of σ. Unlike the normal distribution, we do not identify a **t-**curve by its mean and standard deviation. Instead we identify a given **t-**distribution by its **degrees of freedom**, which is 1 less than the size of the sample (d.f. = $n-1$). Be sure to read the section in the book that details how we can use the associated **t-**tables in the back of the book.

Constructing confidence intervals when the population standard deviation is unknown will require that we identify the number of degrees of freedom and compute *t*-scores corresponding to the level of confidence (instead of *z*-scores). Other than that, the formula for confidence limits is constructed analogously to the situation when σ is known.

As sample size increases so does the number of degrees of freedom. Consequently, when the sample size is large the **t-**distribution will be closely approximated by the *z*-distribution. In other words, for large samples ($n \geq 30$) the *t*-scores and *z*-scores will be equal for a given confidence level.

Section 8.3 is fairly straightforward and deals with margin of error. In particular, you will learn how to answer questions regarding how to determine sample size when estimating population mean within a desired margin of error for a certain degree of confidence. It is important to note that, in general, increasing sample size will improve upon the precision of the estimate.

Written Assignment

Reminder: these written assignments are for your benefit and are **NOT** to be turned in for a grade.

Do problems 8.3, 8.13, 8.20, 8.42, 8.36, 8.52, 8.55, 8.57, 8.81, 8.94, 8.101

Hypothesis Tests for One Population Mean

What is the average body temperature for a healthy adult? Most people would argue 98.6°F. However, in 1992, researchers from the University of Maryland published a study arguing that perhaps we should reject the common belief that the average body temperature of a healthy adult is 98.6°F (Journal of the American Medical Association, 1992, Vol. 268, pp.1578-1580). In fact, in their survey of 106 healthy adults the mean body temperature was $\bar{x} = 98.20°$ F with a standard deviation of s = 0.62°F.

If we construct a 95% confidence interval for the mean body temperature we find

$$\bar{x} \pm z_{.025} \cdot \frac{\sigma}{\sqrt{n}} = 98.2 \pm 1.96 \cdot \left(\frac{0.62}{\sqrt{106}}\right) = 98.2 \pm 0.12$$

In other words, we can be 95% confident that the mean body temperature for a healthy adult should fall between 98.08°F and 98.32°F. This interval does not contain the generally accepted value of 98.6°F. This leads us to the natural question: should we abandon the common belief that the mean body temperature of a healthy adult is 98.6°F?

To answer this type of question we need to perform a **hypothesis test**. This is a very common technique employed throughout statistics.

Informally, a **hypothesis** is a statement about the population distribution, e.g., mean body temperature μ = 98.6°F. A **test** is a statistical procedure employed to decide if the given hypothesis is true or false.

There are two types of hypotheses:

1. the **null hypothesis**, H_o, which represents the current belief.
2. the **alternative (or research) hypothesis**, H_a, which represents the new belief.

When performing any hypothesis test, we are required to make a decision. We either **reject** the null hypothesis or we **fail to reject** the null hypothesis. Note: if we fail to reject the null hypothesis we do not assume that this means that the null hypothesis will be **accepted**. It merely implies that there is insufficient evidence to warrant the rejection of the null hypothesis.

Example: Consider a jury trial. In the United States the defendant is assumed to be innocent until he/she is proven to be guilty. Since the current belief is that the defendant is innocent, this becomes our null hypothesis. The alternative hypothesis is generally taken to be the "opposite" of the null hypothesis. Therefore, the null and alternative hypotheses may be defined as

H_o: the defendant is innocent vs. H_a: the defendant is guilty

The jury will decide on the innocence or guilt of the defendant. Consider the following two scenarios.

Choice 1: There is not enough evidence to rule out the innocence of the defendant, therefore the defendant goes free. Making this choice could lead to the possible error that a guilty person is set free.

Choice 2: There is enough evidence to suggest that the defendant is guilty, therefore the defendant is sent to prison. Making this choice could lead to the possible error that an innocent person is sent to prison while the guilty person remains free and unpunished.

Upon further reflection, we can see that the possible error that can result from the jury making the second choice may be considered a more serious error. In the first instance, a guilty person is set free, but in the second, an innocent person is wrongly imprisoned while the guilty party remains unpunished.

More generally, every time we perform an hypothesis test and make a decision we run the risk of making an error like the ones described above. There are two types of errors for a hypothesis test:

Type I: rejecting the null hypothesis when, in fact, it is true.
Type II: failing to reject the null hypothesis when, in fact, it should be rejected.

For the previous example, finding an innocent person guilty would be considered a type I error. Allowing a guilty person to go free would be classified as a type II error.

Typically, type I errors are considered to be more serious. As a result, we try to control type I errors by attempting to keep the probability of a type I error to less than 5% while at the same time keeping type II errors as small as possible. The probability of a type I error will be denoted by α and will be referred to as the **significance level** of a hypothesis test.

There are five major steps for an hypothesis test:

1. Define the null and alternative hypotheses.
2. Note the value of α. If no α value is given, then a *p*-value will need to be calculated.
3. Select and compute the appropriate **test statistic**. The test statistic measures the standardized discrepancy between the claim (under the null hypothesis) and the evidence.
4. Draw the **rejection region**. The rejection region contains the values of the test statistic for which the null hypothesis should be rejected. This region is found by determining the appropriate critical value(s) based on the significance level α.
5. Make a decision. Reject the null hypothesis if the test statistic falls in the rejection region; otherwise, fail to reject the null hypothesis.

We will be performing many types of hypotheses tests throughout this course. A specific test statistic is associated with each type of hypothesis test. Furthermore, the alternative hypothesis will be classified as either a two-sided alternative or a one-sided (left or right) alternative. In order to draw the rejection region, it will be necessary to know both the general shape of the graph of the distribution associated with the corresponding test statistic *and* whether H_a is one or two-sided.

Example: Return to the mean body temperature scenario. University of Maryland researchers sampled 106 healthy adults and found the mean body temperature was $\bar{x} = 98.20°$ F with a standard deviation of s = 0.62°F. Does the sample provide enough evidence to warrant the rejection of the common belief that mean body temperature is 98.6°F? Test at the α = 0.05 significance level.

This question may be answered using a hypothesis test.

Step 1: Define the null and alternative hypotheses.

H_o: μ = 98.6 (mean body temperature is 98.6)
H_a: $\mu \neq$ 98.6 (mean body temperature is **not** 98.6)

Step 2: α = 0.05

Step 3: Since the claim is made about the population mean, the sample statistic most relevant to this test is $\bar{x} = 98.20°$ F. The Central Limit Theorem applies because the sample size $n \geq 30$. Therefore, $\bar{x}$ has a distribution which is approximately normal with $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. When computing the test statistic we may use s = 0.62 to approximate σ since the sample size $n \geq 30$. The test statistic is

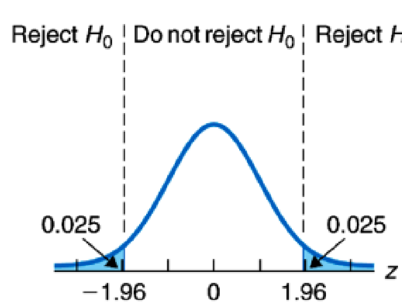
$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{98.20 - 98.6}{\left(\frac{0.62}{\sqrt{106}}\right)} = -6.64$$

Step 4: Draw the rejection region.

To find the critical z-values first notice that H_a: $\mu \neq 98.6$. In other words, getting a mean significantly less or greater than 98.6 provides strong evidence against H_o. Therefore, the test is two-tailed. Divide the significance level, 0.05, equally between the two tails:

$$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025 \text{ in each tail } \Rightarrow \text{ critical values } z = -1.96 \text{ and } z = 1.96$$

The rejection region is given below.



Step 5: Make a decision.

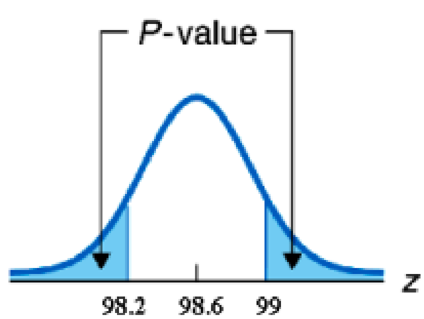
The test statistic $z = -6.64$ lies in the left tail of the rejection region. Therefore, we reject the null hypothesis and conclude that there is sufficient evidence to suggest that mean body temperature differs from 98.6°F. In other words, there is evidence to suggest we should abandon the common belief that the mean body temperature is 98.6°F.

The danger inherent in the above argument is that the test does not provide proof that the original assumption was wrong. We have merely shown that there is evidence to support the belief that mean body temperature of 98.6°F may be incorrect. We must acknowledge that the sample mean of 98.2°F may have happened by **chance**.

The *p*-value approach to hypothesis testing is somewhat more intuitive. Considering *p*-values requires that we think of our problem in terms of probability. The question is, "Should we abandon the common belief that the mean body temperature of a healthy adult is 98.6°F?" In other words, "Is the sample mean body temperature of a healthy adult significantly different from our assumed population mean of 98.6°F?" Basically, we are interested in determining the probability of getting a sample mean of 98.2°F or something more extreme, in a random sample of 106 healthy adults, assuming the population mean body temperature is 98.6°F. If the probability of getting a sample mean of 98.2°F or more extreme is very small, then perhaps our original assumption of 98.6°F as an average body temperature is **incorrect**. If the probability of getting a sample mean of 98.2°F or less is large then we may conclude that it is not significantly different from the suspected mean of 98.6°F.

Formally, a *p*-value is the probability of observing a value of the test statistic as extreme or more extreme than what has been observed.

In the context of our problem on mean body temperature, the *p*-value equals the probability of getting 98.2°F or something more extreme. When considering what is more extreme we first need to realize that we are dealing with a two-sided hypothesis test. Therefore for this problem we need to find $\boldsymbol{P}(\bar{x} \le 98.2 \text{ or } \bar{x} \ge 99) = 2 \cdot \boldsymbol{P}(\bar{x} \le 98.2)$.



In order to determine such a probability we need to standardize the problem by converting to a *z*-score.

$$p\text{-value} = 2 \cdot \boldsymbol{P}(\bar{x} \le 98.2) = 2 \cdot \boldsymbol{P}\left(z \le \frac{98.2 - 98.6}{0.62/\sqrt{106}}\right) = 2 \cdot \boldsymbol{P}(z \le -6.64)$$

Using the ***z***-table, we see that the associated probability is 0.0000 to four decimal places (i.e., $p < 0.0005$) for $z < -3.90$. Therefore, our *p*-value = 2(0.0000) = 0.0000. Using a TI-83 calculator yields a more precise *p*-value of $3 \times 10^{-11} = 0.0000000000\ 3$. This probability is exceptionally small. Therefore we may conclude that the probability of getting 98.2°F or something more extreme by **chance** is very unlikely. Hence, we conclude that our original assumption that the mean body temperature is 98.6°F must be incorrect.

The method of computing *p*-values for one-sided tests is similar.

Thus far our discussion has only dealt with problems where the population standard deviation, σ, has been known. In most real life situations the population standard deviation will be unknown. In spite of this fact, we would still like to perform hypothesis tests. Likewise, how will we deal with small samples? As with confidence intervals, if the population standard deviation is unknown **and/or** if the sample size is 30 or less we will use the *t*-distribution. In other words, our test statistic will be a *t*-score and our rejection region will be constructed using a *t*-curve and the appropriate number of degrees of freedom. The *t*-score will make use of the **sample** standard deviation instead of the population standard deviation.

Written Assignment

Do problems 9.5, 9.11, 9.13, 9.29, 9.35, 9.37, 9.59, 9.67, 9.69, 9.95, 9.103, 9.105, 9.125, 9.127

Inferences for Two Population Means

Many practical applications of statistics involve comparisons of two different population means. For example, a researcher might be interested in determining if men perform better than women on math tests. Similarly, a researcher might be interested in determining if a particular weight-loss program is effective. These two examples compare two different means; however, there is a fundamental difference between these examples. In the first, the two samples are **independent** – men and women. We should think of independence as meaning that there is no relation between specific values of the two populations. Given that fact, it hardly seems possible to consider the second example as the comparison between two independent populations. (The only way that could be possible, is if the "before" weights were recorded from one group of subjects while the "after" weights were recorded from an entirely different group of subjects. This would not be a useful method for determining the effectiveness of a weight-loss program.) Thus for the second example, the samples are **dependent** – the weight measurements are recorded from the same subjects before and after the implementation of the weight-loss program.

When making such comparisons we are essentially interested in the difference between the two population means, $\mu_1 - \mu_2$. If the two population means are the same then this difference should be zero, since $\mu_1 = \mu_2$ implies that $\mu_1 - \mu_2 = 0$. Without going deeply into the statistical theory, it turns out that when we make such comparisons it will require that we identify the sampling distribution associated with $\bar{x}_1 - \bar{x}_2$, including the new mean and standard deviation that will be used when computing the test statistic. This is explained in detail in the text. We will be required to choose a test statistic (*z*-statistic or *t*-statistic) and its associated formula depending on various assumptions about sample size and population standard deviation (assumed to be equal or not equal).

The most important thing to remember is that the procedure for performing hypothesis tests is essentially the same. The only things that change from one test to another are the formulas for the test statistic and the distribution used when constructing rejection regions. Likewise, the methods of computing *p*-values for one or two-sided tests remain the same, provided we select the appropriate distribution curve when computing area.

When testing the difference of two means for **independent** samples we may define the null and (options for) the alternative hypothesis as follows:

$$H_o:\ \mu_1 = \mu_2 \text{ (or } \mu_1 - \mu_2 = 0\text{)} \quad \text{vs.} \quad H_a: \begin{cases} \mu_1 \neq \mu_2 \\ \quad \text{or} \\ \mu_1 < \mu_2 \\ \quad \text{or} \\ \mu_1 > \mu_2 \end{cases}$$

When dealing with two or more populations, calculating critical value(s) and test statistics becomes very tedious and time consuming. Such extensive calculations are prone to error. It is very important not to lose sight of the fact that we are more concerned with the interpretations of our results than we are with the actual computations involved. Whenever possible it is preferable to use computers or calculators to perform hypothesis tests and compute confidence intervals for the difference of two means.

Example: In the summer of 1988, Yellowstone National Park suffered the devastating effects of several major forest fires that destroyed large tracts of old timber near many famous streams. Many fishermen have been concerned about the long-term effects of the fires on the streams. Some biologists have argued that the growth of new meadows would produce more of the insects on which the fish feed. In other words, there is a possibility that improved fishing could be a long-term effect of the 1988 fires.

Rangers have collected data detailing the daily number of fish caught by fisherman. The Park's publication, Yellowstone Today, has indicated that the biologists' claim have been correct and that fisherman have been satisfied by the increase in their daily catch.

Suppose that the instructor went to a Yellowstone Ranger station and gathered some actual data from fishing reports. In the years prior to the fires, a random sample of 167 fishing reports was selected and revealed that the average catch per day was 5.2 trout with a sample standard deviation of 1.9. After the fires, a random sample 125 fishing reports was selected and revealed that the average catch per day was 6.8 trout with a sample standard deviation of 2.3.

Looking solely at means it would appear that the biologists' claim is true. However, it would be nice to verify this belief using a hypothesis test.

First we should organize and label the data:

Before	After
$n_1 = 167$	$n_2 = 125$
$x_1 = 5.2$	$x_2 = 6.8$
$s_1 = 1.9$	$s_2 = 2.3$

The sample standard deviations are not incredibly close, so we will assume that the two populations have unequal standard deviations. This means we will need to use a **non-pooled** test. Since we are hoping to establish that fishing has improved since the fire we will chose the left-tailed alternative hypothesis that $\mu_1 < \mu_2$.

Using a TI-83 calculator to perform the extensive computations yields the following: *t*-statistic = –6.33, degrees of freedom = 236.86, and *p*-value = 0.0000000006. Given the exceptionally small *p*-value, we may confidently reject the null hypothesis. There is sufficient evidence to suggest that the daily catch of fish has increased since the 1988 forest fires.

Recall that when the degrees of freedom get large, the *t*-distribution closely approximates a *z*-distribution. In other words, critical value(s) for the rejection region will be the same for both the *t*-distribution with very large degrees of freedom and the *z*-distribution. This could simplify your work when determining the rejection region.

Suppose that instead of using a hypothesis test, we chose to find a 95% confidence interval to address our question. My handy calculator provides the following:

95% CI: (-2.098, -1.102)

In other words, we are 95% confident that $\mu_1 - \mu_2$ lies between –2.098 and –1.102. Since both the upper and lower limits of the confidence interval are **negative** we are 95% confident that $\mu_1 - \mu_2 < 0$. This means that we are 95% confident that $\mu_1 < \mu_2$, i.e., we are 95% confident that fishing has improved.

Example: My friend, Frieda, is a nutritionist. She has been working towards developing an effective weight-loss program. Frieda knows nothing about statistics so she asked me to help her determine if her latest program is truly effective. Below are the weights of 10 subjects before and after implementing the experimental weight-loss program.

Before	156	183	172	147	197	149	162	176	153	128
After	147	189	167	140	204	138	156	179	153	126

Quickly looking at this information reveals that 6 people lost weight, 3 people actually gained weight, and one person maintained their original weight.

This example is different from our previous example because the samples are **not** independent. This means that we will need to alter our procedure for comparing the two means. Basically we are interested in determining if, on average, the subjects lost weight. Given that objective, we really only need to consider the difference in weight from before and after. Therefore, we are interested in testing the difference of **paired** samples.

To do so we will need to define a new variable, ***d***, to be the difference in samples. For this example, ***d*** = before − after. The differences are given below:

Difference	9	-6	5	7	-7	11	6	-3	0	2

Our problem has been reduced to answering the question: "is the mean difference, $\bar{\boldsymbol{d}}$, greater than zero?" This is because if subjects, on average, lose weight then the "before" weight should be greater than the "after" weight, and hence, $\bar{\boldsymbol{d}}$ = before − after > 0. If we think about this for just a moment we can see that our hypothesis test will be reduced to a *one*-sample *t*-test on the sample statistic $\bar{\boldsymbol{d}}$.

When testing the difference of two means for **paired** (or dependent) samples we may define the null and (options for) the alternative hypothesis as follows:

$$\text{H}_\text{o}\text{:}\ \ \bar{\boldsymbol{d}} = 0 \ \text{(or } \mu_1 = \mu_2\text{)} \quad \text{vs.} \quad \text{H}_\text{a}\text{:} \begin{cases} \bar{\boldsymbol{d}} \neq 0 & (\mu_1 \neq \mu_2) \\ \quad \text{or} & \\ \bar{\boldsymbol{d}} < 0 & (\mu_1 < \mu_2) \\ \quad \text{or} & \\ \bar{\boldsymbol{d}} > 0 & (\mu_1 > \mu_2) \end{cases}$$

Specifically for our problem, our hypotheses are defined as H_o: $\bar{\boldsymbol{d}} = 0$ and H_a: $\bar{\boldsymbol{d}} > 0$.
If we use Minitab to perform the one-sample *t*-test we get the output below.

Minitab -- Paired T-Test and Confidence Interval

```
(1) Paired T for before - after

                         N            Mean          StDev         SE Mean
(2) before               10           162.30        20.08         6.35
(3) after                10           159.90        24.63         7.79
(4) Difference           10           2.40           6.26         1.98
(5) 95% CI for mean difference: (-2.08, 6.88)
(6) T-Test of mean difference = 0 (vs > 0): T-Value=1.21  P-Value=0.128
```

Line 1 identifies how we have defined our difference. Lines 2 – 4 provide descriptive statistics for each of the samples as well as the difference between them.

Line 6 defines both the null and alternative hypotheses while providing both the value of the test statistic and the associated *p*-value. Since the *p*-value is large, i.e., greater than 0.10, we will reject the null hypothesis. Therefore, we conclude there is **insufficient** evidence to suggest that the weight-loss program is truly effective.

Line 5 supplies us with a 95% confidence interval for the mean difference. We could use this confidence interval to arrive at the same conclusion provided by the hypothesis test. The interval contains zero, since the lower confidence limit is negative while the upper confidence limit is positive. Thus, we are 95% confident that the mean difference in weights is between –2.08 and 6.88. Since zero lies inside the interval then we must acknowledge that the mean difference **could** equal zero. In other words, there is not enough evidence to suggest that the weight-loss program is truly effective.

Written Assignment

Do problems 10.3, 10.12, 10.33, 10.45, 10.65, 10.73, 10.95, 10.97

(Note: For problem 10.65: $\bar{d} = -7.26$ and $s_d = 7.16$)

You are now ready to take the Unit 3 test. Please send in the test, with the attached assignment cover, as your assignment for Unit 3.

Unit 3 Test

Use only hand calculators on this exam; Minitab or any other computer software is **not** allowed. The textbook (but no other notes) may be used during the exam. Be sure to submit all your work in order to receive partial credit. For example, when performing hypothesis tests it is necessary that you define the hypotheses, sketch the rejection region, and show computations for the test statistic.

1. Suppose a wildlife service wishes to estimate, with 99% confidence, the mean number of days of hunting per hunter for all licensed hunters in the state during a given season, with a bound on the error of estimation equal to 2 hunting days. How many hunters must be included in the survey? Assume that the data collected in earlier surveys have shown $\sigma \approx 10$.

2. We wish to estimate the mean serum indirect bilirubin level of 4-day-old infants. The mean for a sample of 160 infants was found to be 5.98 mg/100 cc. Assume that bilirubin levels in 4-day-old infants are approximately normally distributed with a standard deviation of 3.5 mg/100 cc. Construct a 95% confidence interval for the true mean bilirubin level in 4-day-old infants.

3. A researcher studied the effects of pancuronium-induced muscle relaxation on circulating plasma volume on newborn infants weighing more than 1700 grams who required respiratory assistance within the first 24 hours of birth and met other clinical criteria. The following table is a result of measurements of plasma volume (ml) made during mechanical ventilation.

subject group	sample size	sample mean	sample st. dev.
paralyzed	5	48.0	8.1
nonparalyzed	7	56.7	8.1

   a. You may assume equal standard deviations. Construct a 98% confidence interval for the difference of the two population means.
   b. What are your conclusions? Do you think there is a difference between the two population means?

4. Consider the following scenario:

   Dr. Jeffrey M. Barrett of Lakeland, Florida, reported data on eight cases of umbilical cord prolapse. The maternal ages were 25, 28, 17, 26, 27, 22, 25, and 30. He was interested in determining if he could conclude that the mean age of the population of his sample was greater than 20 years. Let $\boldsymbol{\alpha = .01}$.

   The null and alternative hypotheses for this problem are defined as:

$$\mathbf{H_o : \mu = 20} \quad \textbf{\textit{and}} \quad \mathbf{H_a : \mu > 20}$$

a. Explain the meaning of a Type I error in the context of this problem.
b. Explain the meaning of a Type II error in the context of this problem.
c. Suppose Dr. Barrett expanded his study to include a total of 50 subjects and then calculated a test statistic of $\boldsymbol{z = 2.33}$ for his hypothesis test. What is the associated p-value? What is the conclusion for this test?

5. A local weight-loss company suggests that the average client loses 15 pounds during the first month. A consumer advocate group feels that the actual number of pounds lost is much less than this. To test the claim, they select 30 of the clients at random and obtain the following data:

Pounds Lost During the 1st Month:

16	20	14	15	14	18	19	20	9	13
20	11	12	18	13	17	11	21	14	15
15	13	17	12	9	20	16	18	21	11

A hypothesis test is conducted.
a. State the null and alternative hypotheses.
b. The test statistic is $t = 0.6107$ with p-value = 0.7269. Based on this information, what is the conclusion of the hypothesis test? Justify your answer.

6. The manufacturer of a particular all-season tire claims that the tires last for 22,000 miles. After purchasing the tires you discover that yours did not last the full 22,000 miles. Suppose that a sample of 100 tires made by that manufacturer lasted on average 21,819 miles with a sample standard deviation of 1,295 miles. Is there sufficient evidence to refute the manufacturer's claim that the tires last 22,000 miles? Let $\alpha = 0.05$. Assume that the population standard deviation is $\boldsymbol{\sigma = 1300}$.
a. Define the null and alternative hypotheses.
b. Find the appropriate rejection region.
c. Compute the test statistic.
d. What is your conclusion? Explain.

7. The following data presents tests scores on a Behavioral Vignettes Test for self-help skill teaching for the primary parents of a random sample of families with mentally retarded children before and after a training program.

Before	7	6	10	16	8	13	8	14	
After		11	14	16	17	9	15	9	17

Based on this information, can you conclude that the training program is effective? Use a significance level of 0.05. State your assumptions.