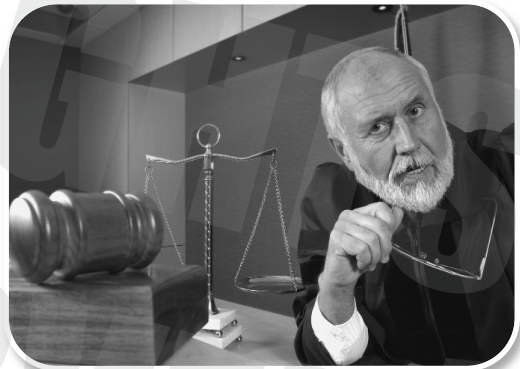


# Confidence Intervals—One Population Mean and Proportion Testing

**Courtroom humor** – *actual exchanges based on a book by Charles M. Sevilla*

- Attorney:** Can you describe the individual?  
**Witness:** He was about medium height and had a beard.  
**Attorney:** Was this a male or female?  
**Witness:** Unless the circus was in town, I'm going with male.



## Confidence Intervals

By way of review, let's look at a table to aid you in keeping the terminology straight. This is the same as Table 6.8 from the previous chapter.

The first row is associated with the population. The second row is a single sample ( $n = 1$ ). The third row is the sampling distribution of sample means ( $SD$  of  $SM$  when  $n > 1$ ). The second column consists of means. A mean is a mean is a mean. The third column is all variances. A variance is a variance is a variance. The fourth column is the standard deviation and the fifth is the size of the population, sample and  $SD$  of  $SM$ . All calculations are done the same way for a mean, for a variance, and for a standard deviation.

**Table 7.1** Review Table for Symbols.

	Mean	Variance	Standard Deviation	Size of the ...	Use of Z-Process
Population	$\mu$	$\sigma^2$	$\sigma$	$N$	
Sample	$\bar{X}$	$S^2$	$S$	$n$	$Z = \frac{X - \mu}{\sigma}$ When $n = 1$
SD of SM	$\bar{\bar{X}}$	$\sigma_{\bar{X}}^2$	$\sigma_{\bar{X}}$	$K$	$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$ When $n > 1^*$

\*Note: Use of Z will be modified when I discuss the  $t$ -distribution in this Chapter.

Because different concepts are stressed with each row of symbols, they are named differently to keep the conceptual confusion to a minimum.

Okay, I know you are thinking, "Getting confused is what I have been several times in this course." It will get better the more exposure you have to the language and concepts. The best method of leaning is to work the assigned homework problems.

I trudge on.

The last column in Table 7.1 shows two different Z-processes. One applies to a single sample and one applies any time  $n > 1$ .

$$Z = \frac{X - \mu}{\sigma}, \text{ when } n = 1.$$

Or  $Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$ , where  $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$  which allows me to re-write the formula

$$\text{as } Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}, \text{ when } n > 1.$$

I have two Z-processes. Use of either requires the distribution to be normal. Two things can make the distribution normal. First is the central limit theorem (CLT), which ensures the distribution of sample means is normal, if  $n \geq 30$ . Second, for sample sizes less than 30, you must be told the distribution is normal. If the distribution is not normal, a different statistical procedure is required, which I will cover a bit later.

How am I going to use all of this stuff, you ask? Remember, I have previously demonstrated that anytime I take a sample I have a very high probability of sampling error. Sampling error is defined as the difference between  $\bar{X}$  and  $\mu$ . My problem is what I do with the sampling error. Can I assert  $\bar{X}$  represents  $\mu$  or does  $\bar{X}$  not represent  $\mu$ ?

Sampling error is NOT standard error, so don't let the two get jumbled up in your mind.

Sampling error is  $\bar{X} - \mu$ .

Standard error is  $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ .

When I take a sample, I know I have sampling error. I want to assert that  $\bar{X}$  is representative of  $\mu$ . Two statistical methods are available to help me evaluate my assertion. One method is confidence intervals and the other method is hypothesis testing. In the next chapter, I will address hypothesis testing. In this chapter, I will address the use and interpretation of confidence intervals.

There are two types of estimators. One is a point estimate and the other is an interval estimate.

Okay, you have probably forgotten what a point estimate is, right?

The point estimate uses the *mean of the sample* as an *estimate of the mean of the population at a given point* ( $\bar{X}$  is a point estimate of  $\mu$ ). Point estimates are estimates of the population parameter (mean in this case). One difficulty with a point estimate is the sample mean changes each time a new sample is selected. Remember, permutation or combination arrangements demonstrate the number of samples or subsets I can develop for any population data set.

To use the sample mean and expect it to exactly equal the true, unknown population mean is not a practical or realistic expectation. If I can place around the point estimate ( $\bar{X}$ ) an interval that would be based on some statistically-consistent procedure, I would have a better chance of the sample mean actually representing the true, unknown population mean. In essence, this is what confidence intervals accomplish.

The interval estimate begins with the point estimate and adds an interval based on an acceptable level of error. The relationship is as follows.

$$\bar{X} \pm Z_{\alpha} \sigma_{\bar{X}} \text{ which basically comes from solving } Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \text{ for } \bar{X}$$

I begin with the point estimate of  $\bar{X}$ . Around the point estimate, an interval is calculated using  $\pm Z_{\alpha} \sigma_{\bar{X}}$ . The interval is  $Z$  at a chosen alpha error level times the standard error.

Also remember the standard error can be the following:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}, \text{ so the interval can be stated } \pm Z_{\alpha} \frac{\sigma}{\sqrt{n}}.$$

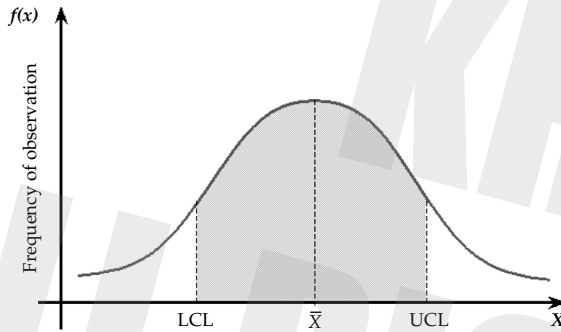
When everything is put together, the final formula looks like the following:

$$\bar{X} \pm Z_{\alpha} \sigma_{\bar{X}} \text{ or the mathematically equivalent formula, } \bar{X} \pm Z_{\alpha} \frac{\sigma}{\sqrt{n}}.$$

Okay, I have thrown a bunch of formulas at you, but the last two are the important ones to recall. Now let's talk about how using an interval estimate improves my ability to make a conclusion about the true, unknown population mean.

My process needs to be consistent. Confidence intervals provide me with that opportunity. Once I have determined the interval, I have two values, one below the mean (LCL) and one above the mean (UCL).

Graphed it looks like the following (Figure 7.1):



**Figure 7.1** Area under the curve between the LCL and UCL.

Since I never know the true mean of the population, confidence intervals allow me to place around the point estimate (the sample mean) an interval. I can make certain statements about the true, unknown population mean (as long as the sample is selected randomly). Error can still occur, but the error is limited. There are three common error values (1%, 5%, and 10%). Each of these is referred to as an alpha error. Each alpha error has a corresponding confidence coefficient (CC, which are 99%, 95%, and 90%). These are also referred to as confidence limits (CL).

More on this shortly, so be patient. Just soak up the wonderful knowledge I am imparting to you.

Okay, knock off the snickering.

The **alpha error** by definition is the *rejection of a true, correct outcome*. In other words, the result of my study is true, but I erroneously say the results are false. An alpha error is the most serious error in many testing processes. Everything is okay, but I incorrectly conclude it is not okay. If

**Table 7.2** Confidence coefficients and alpha.

Confidence Coefficients	Alpha Error	Total Area
0.99	0.01	1.00
0.95	0.05	1.00
0.90	0.10	1.00

I am in the manufacturing business, this might result in me shutting down my production line for an invalid reason. A shutdown when none is necessary could be a costly error.

When you use any particular confidence coefficient, there will be a corresponding alpha error level. Again, the alpha error is the *probability of rejecting a true statement*. If you have a CC of 95%, you will have a complementary alpha of 5%. Together they total 100% of the area under the curve. Table 7.2 summarizes the most common values.

Could I have other confidence coefficients?

Yes.

I could have anything from 1% to 100%, but the three most common are the 90%, 95%, and 99% confidence levels (confidence coefficients). Attorneys use a confidence interval of 51% to win a case, although they don't state it in that manner, nor do they probably consciously understand it that way.

Let's see if I can sort out what I am really trying to say.

A confidence interval is always a two-tailed rejection region test. The alpha value is in the tails of the curve. If alpha is 0.10, one-half will be in the right tail (5%) and one-half will be in the left tail (5%). My confidence statement will be associated with the chosen alpha value or the corresponding confidence level.

I know this is not clear yet, so keep reading.

The interval I develop will always begin with  $\bar{X}$ . I will choose an alpha value, which will be written  $Z_\alpha$ . For instance, if alpha was 0.05, you would see  $Z_{0.05}$ . This means I have chosen the 95% confidence interval or confidence level.

To repeat the formula, I have the following:

$$\bar{X} \pm Z_\alpha \sigma_{\bar{X}}$$

The true mean of the population is unknown. By using confidence intervals, I am able to state how confident I am the true, unknown population mean is between the LCL and UCL. For example, if I use an alpha error of 0.05, I can make the following statement: I am 95% confident the true, unknown population mean is between the LCL and the UCL. This means I will be correct 95% of the time and wrong 5% of the time (the alpha error).

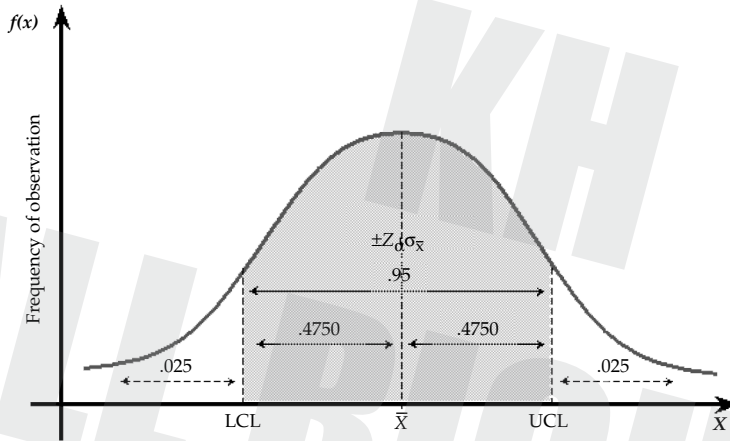
**Another Example:** In the formula,  $\bar{X} \pm Z_\alpha \sigma_{\bar{X}}$ ,  $Z$  at any alpha level can be determined by looking up a value in a table or using software such as Excel. Knowing how to look up  $Z$  in the normal table is crucial. For example, if I were using an alpha level of 0.05, the area between the UCL and LCL would be 95%. Because a confidence interval is a two-tailed rejection region, this means 0.0250 of the area would lie in the upper tail (to the right of the UCL) and 0.0250 of the area would lie in the lower tail (to the left of the LCL).

Well, you have done it. You have me confused, you comment.

Okay, let's keep trudging on and see if I can clear this up for you. Maybe looking at it graphically will help you to visualize what I am trying so desperately to explain verbally.

Remember, a normal distribution is symmetrical, thus 50% lies to the right of the mean and 50% lies to the left of the mean. The entire distribution must total 100%. In my example,  $0.9500 + 0.0250 + 0.0250 = 100\%$ .

See the below graph.



**Figure 7.2** Area under the curve for alpha 0.05.

Okay, you say. I think I partially understand it, but go over the interpretation once more.

Good question, since many students struggle with two things—looking up the Z-value and interpreting the results.

Once I have established the UCL and the LCL, I can then interpret the outcome.

There are only two possible interpretations.

1. I am 95% confident that the true, unknown population mean lies between the UCL and the LCL.

OR

2. If I construct many confidence intervals, 95% of them will contain the true, unknown population mean; however, 5% will not.

Don't get creative with your own interpretations. Mine are demanding enough. In either case, there is a 5% chance of error. Said differently, I will be correct in my assertion 95% of the time and wrong 5% of the time.

Confidence intervals have nothing to do with probability. The probability that the true, unknown population mean lies between the UCL and LCL is either 0 or 1. The true mean is either between the UCL and the LCL, or it is not there. You cannot interpret the result as that there is a 95% probability that the interval will contain the true mean. Probability and confidence levels are not the same.

## Putting It Together (I hope)

Let's suppose that I am a manufacturer of a paper product that must measure 11.0 inches in length. This value is my specified value, so it is the target mean length for all of my sheets of

paper ( $\mu_H$ ). Let us also suppose that I have determined that the standard deviation of my manufacturing process is 0.02 inches. I want to take samples on a periodic basis to determine if my production process is yielding paper that is 11.0 inches long within the limits of a tolerable error (one-half the interval width by definition). There may be times when the cutting blade may become dull, which may cause the paper to not be properly cut.

Suppose I determine I want to be 95% confident that the paper I am producing falls within tolerable error. Acceptable tolerable error is random error. I can calculate a 95% confidence limit. I take a sample of 100 sheets of paper ( $n = 100$ ). For this sample, the sample mean is determined to be 10.998 inches. While the sample reflects my production process is less than the 11.0 inches desired, do I have enough information to assert the blade has become too dull? Am I producing paper outside an acceptable length (outside tolerable error)?

Said another way, can I be 95% confident my production process is in control and the difference between the 10.998 and 11.0 is due to random sampling error? Random sampling error is okay in explaining the difference.

To test this I calculate a 95% confidence interval. The calculation is as follows:

$$\begin{aligned}\bar{X} \pm Z_{\alpha} \sigma_{\bar{X}} \\ 10.9980 \pm Z_{0.05} \left[ \frac{0.02}{\sqrt{100}} \right] \\ 10.9980 \pm (1.96) \left[ \frac{0.02}{10} \right] \\ 10.9980 \pm (1.96)[0.002] \\ 10.9980 \pm 0.0039 \\ 10.9941 \text{ through } 11.0019\end{aligned}$$

**Interpretation:** Since the interval contains the specified 11.0 inches, I can assert that I am 95% certain that the production process is in control and the blade is not too dull to cut the paper to the proper length.

Whoa, you say. I know where you got the 10.998, and I know how you got the standard error by using the standard deviation divided by the sample size, but where on earth did you get the 1.96?

Oh, how quickly you forget, I respond.

Go back and look at the Figure 7.2. A confidence interval is a *two-tailed rejection test*. One-half of alpha goes in each tail.

Okay, have you turned back to find that Figure 7.2 yet? It will give you a visual of the Z-calculation you must make. To find Z at alpha 0.05, you must look at the Z-table (normal table), which was given to you at the ends of Chapter 5 and Chapter 6. It is also in the back of the textbook. I will wait until you find that table, remove it from your textbook, and look at it as I tell you the following.

The alpha level is 0.05. This is a two-tailed rejection region test, so I will divide 0.05 by 2, which equals 0.0250 for each tail (since a normal distribution is symmetrical). Subtracting the 0.0250 from

0.5000 (50%), I have 0.4750 in the two equal center sections. ( $0.4750 + 0.4750 = 0.9500$ ). This is the 95% confidence interval.

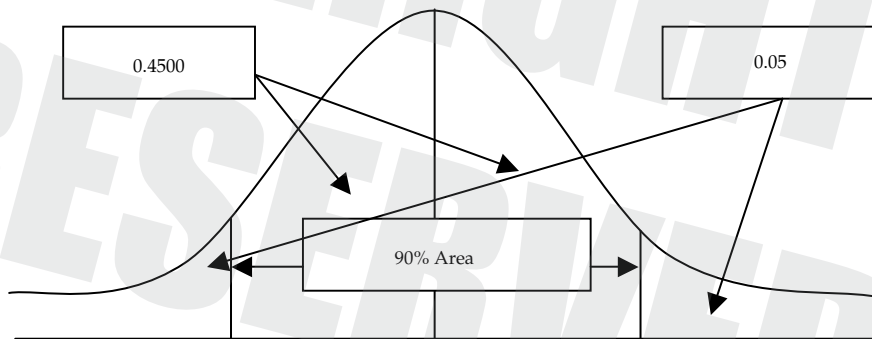
Using the normal table, I will look up the 0.4750 in the *body of the table*. When I do, I get a Z-value of 1.96. Take a minute and see if you get the same Z-value. If you do not, you need to see the Professor or another student in your class who understands this process. You must have an absolute grasp of looking up Z-values. If you do not, you will not do well on examinations or understand the majority of the remainder of these next few chapters.

**To repeat:** Did you find the 0.4750? Good. Now trace that value back to the Z-value on the left side of the normal table and to the value above in the normal table. You should have now solved the mystery of the 1.96. **THIS IS A CRITICAL ISSUE AND YOU MUST UNDERSTAND HOW TO LOOK UP Z-VALUES.**

Let's look further at the application of the Z-process.

## Z-Values

What I want you to do next is fill in the table below (Table 7.3). I want you to use the normal distribution table to determine the Z-values corresponding to the four listed confidence coefficients. I will give you some help on one of the values and let you determine the other two. (HINT: I just did the 95% area in the example problem just above.) The 90% CI would look like the following (Figure 7.3):



**Figure 7.3** Area under the curve.

If I want to determine the Z-value associated with the 90% confidence limit, I would first divide 0.90 by 2. This gives me the value of 0.4500. This value of 0.4500 represents the area under the curve from the mean to the LCL or the UCL. I next look in the body of the table (not the edges of the table) to find the value closest to the 0.4500. I find the value 0.4505. I next move to the outer edges of the normal table to find the Z-value. The answer is 1.65. Actually you will find you are midway between 1.64 and 1.65. I usually take the conservative approach and round up when the value is between two values. Rounding is okay since interpolation will not yield significant differences. Of course, using Excel will yield more precise values.

Take a couple of minutes and fill out the Z-values for the following confidence coefficients (CC or CL). I have already done a couple of them for you.

**Table 7.3** Exercise in looking up the Z-value.

Confidence Coefficient or Confidence Limits	Alpha Error	Z-Value
0.99	0.01	
0.98	0.02	
0.95	0.05	
0.90	0.10	

In general, to find the values, you will divide alpha by two and place it in each tail. Then you subtract the alpha value you just calculated from 0.5000. The result will be the value you look up in the body of the normal table. Usually use the closest value or round up. The same result can be obtained by dividing the CC or CL by two and looking up that value in the body of the normal table. Notice either approach yields the same result.

From the top of Table 7.3, the four answers are 2.58, 2.33, 1.96, and 1.65. The 2.58 is rounded up as is the 1.65. For the 2.33, the closest value in the body of the normal table is 0.4901 (you need 0.4900). I repeat, you must learn how to look up Z-values in the normal table. Before moving on, please make sure you know how to look up these values in the normal table.

I gave you an example just above about a paper manufacturer who wanted to develop a 95% confidence interval for his manufacturing process. Let's suppose that the manufacturer wanted to test his process control at the 99% level rather than the 95% level.

What would change? Think about it before reading the answer just below.

The Z-value would change from 1.96 to 2.58. Since the 2.58 is larger than the 1.96, it is logical to expect the interval to expand.

It does.

The new confidence interval is now the following.

10.9928 to 11.0032.

Make the substitution of 2.58 for the 1.96 and see if you come up with these same numbers. By comparing two intervals, I can see the 99% interval is wider.

95% Interval: 10.9941 to 11.0019 (width 0.0078)

99% Interval: 10.9928 to 11.0032 (width 0.0104)

**Interpretation:** My conclusion does not change, however. I have no reason to expect the process is not in control at the 95% level, and I have no reason to expect the process is out of control at the 99% level.

If I use the 90% level, the interval narrows from both the 99% and the 95% level, but my conclusion does not change. My statement of confidence changes (90%, 95% or 99% confident) in all three instances; however, the conclusion remains the same. There is no problem with the production process regardless of using the 99%, the 95%, or the 90% confidence levels. One word

of warning: The alpha level is a value that should be determined *before* the study begins. Do not play with it to achieve a pre-conceived result.

For example, let's say I am the quality control inspector for a soft drink bottler. I want to ensure that I fill my 12 oz cans to the net weight level stated on the can. I do not want to put too much in the can or too little in the can. I can never know the true, unknown population mean of the 12 oz cans, because every time I fill a new can, the true mean changes.

You wonder, why isn't the true mean 12 oz?

I respond, some cans are under filled slightly and some are overfilled slightly. The only way I can determine the true mean is to dump out the liquid in every can and measure it. As you can see, this would be impractical, since I would have to destroy my product. This would lead to zero production, which would lead to a very unhappy sales force and stockholders. The point is you will never know the true, unknown population mean. If you know the true mean, there is no need for a sample. There is a true, unknown population mean ( $\mu_T$ ) and a hypothesized (unproven, but asserted) population mean ( $\mu_H$ ). The hypothesized mean is a representative of the true, unknown population mean.

## Relationship Between Standard Error, Sample Size, and Interval Width

*Remember, as the sample size increases, the standard error decreases.* This relationship is demonstrated by looking at the formula for the standard error.

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

As "n" increases,  $\sigma_{\bar{X}}$  decreases. This is an important inverse relationship.

The interval width is represented by the following formula:

$$\pm Z_{\alpha} \sigma_{\bar{X}}$$

*Two things will affect the interval width.* First, if you increase the sample size, the standard error will decrease. If the standard error decreases, the interval will become narrower assuming you hold Z constant. Second, if you select an alpha level of 0.01 ( $Z = 2.58$ ) versus an alpha level of 0.10 ( $Z = 1.65$ ), you will have a narrower interval at the 0.10 level than you will at the 0.01 level, assuming you hold the standard error constant. In summary, the interval width is controlled by the size of the sample (the larger the sample, the smaller the standard error) and by your selection of an alpha level. The lower the confidence interval, the narrower the interval will be. The higher the confidence interval, the broader the interval will be.

Okay, Professor, you have covered sample sizes which are equal to or greater than 30, but I can imagine a time when smaller sample sizes must be taken. What do you do in these instances when  $n < 30$ ?

*I have been working with the Z-test. This is known as a test statistic. There are other test statistics, however. There are several circumstances that must be considered as you determine the proper test statistic you will use in developing confidence intervals.*

- First, you should always ask yourself “Is the population normally distributed?”
- Second, you should ask yourself “How large is the sample size?”
- Third, you should ask yourself “Is the standard deviation of the population known?”

Your answer to these three questions will tell you which test statistic to use. I have three choices of test statistics at this point: Z-test, *t*-test, and nonparametric tests. I won't challenge you with the nonparametric tests at this point, and you have been exposed to the Z-test. This only leaves one other: the *t*-test. Let's look at the *t*-test.

Wait, a minute, you ask. Tell me again, what is a test statistic?

Oh, I respond. Let me amplify it a bit more.

The term **test statistic** simply refers to the type of test you are going to conduct in solving any particular problem. The one to which you have been introduced is the Z-process. I am going to bring into focus the *t*-test, shortly. Others include nonparametric tests, the *F*-test, Chi Square, ANOVA, and others, but don't worry about those just yet. There is plenty of time left in the textbook pages to muddle your thinking in those areas.

Let's briefly go back and look at the three questions I posed just above.

If the population is known to be normal and the sample size is 30 or greater and the standard deviation of the population is known, you *always use the Z-test*.

You would also use the Z-test if the population is known to be normal and the sample size is less than 30, but the *standard deviation of the population is known*.

I can give you an even broader picture of when you are to use the Z-test. *If the population is either normal or non-normal and **the standard deviation of the population is known**, use the Z-test statistic regardless of the size of the sample. The sample size may be equal to or greater than 30 or it may be less than 30.*

Good gosh, you say. This is getting confusing. Can't you simplify these statements in some way, so I can understand when to use what?

You are correct, I say. I am beginning to confuse myself, too.

Let's put the rules in a table and see if I can simplify the selection process. The three questions I ask form the headings of the columns.

This should simplify things for you.

Yeah, right, you comment. Maybe it makes is simpler for you, but I am still confused.

Okay, look down the right side of Table 7.4. Notice you *use the Z-test almost always, unless you are using the t-test or a nonparametric test*. Okay, so relax. As I have already said, you don't

**Table 7.4** Conditions for using various test statistics ( $Z$ ,  $t$  or nonparametric).

First Question	Second Question	Third Question	
Is the Population...?	Is the Sample Size ...?	Is the Standard Deviation of the Population...?	Use
Normal–If Yes	Greater than or Equal to 30	Known	Z-test
Normal–If Yes	Greater than or Equal to 30	Unknown	Z-test
Normal–If Yes	Less Than 30	Known	Z-test
<b>Normal–If Yes</b>	<b>Less Than 30</b>	<b>Unknown</b>	<b>t-test</b>
Normal–If No	Greater than or Equal to 30	Known	Z (CLT)
Normal–If No	Greater than or Equal to 30	Unknown	Z (CLT)
Normal–If No	Less Than 30	Known	<b>Nonparametric</b>
Normal–If No	Less Than 30	Unknown	<b>Nonparametric</b>

have to be concerned about the nonparametric tests at this point—that’s later. The only one you have to learn is when to use the  $t$ -test, since *all other conditions* require you to use the  $Z$ -test. I could list each leg of Table 7.4 (normal distribution,  $n \geq 30$ , and  $\sigma$  is known =  $Z$ -test, for example), but listing all of the legs is not really necessary. In reality all you have to do is remember to use the  $Z$ -test in all circumstances, *unless* the distribution is normal,  $n < 30$ ,  $\sigma$  is unknown. Under those three conditions and those three alone, you use the  $t$ -test. I will repeat them below.

There are three conditions which must *all, repeat all*, be met to use a  $t$ -test.

- First, the population must be normal.
- Second, the sample size must be less than 30.
- Third, the standard deviation of the population must be unknown.

When all three of these conditions are met, you use a  $t$ -test rather than a  $Z$ -test. Use  $Z$  unless all three conditions are met.

So now you wonder silently, what is a  $t$ -test?

Clairvoyance taking over again I say: Okay, let’s look at the meaning and use of a  $t$ -test.

You continually are amazed at my ability to read your mind. There is a secret to this skill, but I will not disclose it until much, much later.

## t-test

Quite naturally, the  $t$ -test has three conditions that must be assumed or stated in order for you to successfully use  $t$ -test values. It seems as though there are always assumptions, aren't there?

Those three conditions are repeated here a third time for emphasis: (Don't you just hate redundancy?)

- *The distribution is normal* through any statement in the problem that ascribes normality to the distribution.
- *The sample size is less than 30*. The  $t$ -test is used for small size samples.
- *The standard deviation of the population is unknown*. Notice this is the population standard deviation and not the sample standard deviation.

So what is the  $t$ -test, you ask again?

Essentially the  $t$ -test is the same thing as a  $Z$ -test, but for small samples ( $n < 30$ ).

The  $t$ -distribution is driven by *degrees of freedom*. For now, the degrees of freedom are  $n - 1$ . In an earlier chapter, I introduced you to degrees of freedom.

Let's review the idea of degrees of freedom. Let's say that I have four numbers that equal 100.

$$X_1 + X_2 + X_3 + X_4 = 100.$$

I can set three of the values at any number that I wish. The first could be 20, the second could be 30, and the third could be 40. The total of these three is 90. However, once these three are selected, the fourth number is fixed and allows me no choice—it is 10.

By definition, the degrees of freedom are  $n - 1$  or three degrees of freedom. I have three free choices out of four total choices. The fourth is not a free choice.

### Solving Problems with the $t$ -test

Okay, now that I know about the  $t$ -distribution or  $t$ -test, what do I do next?

The formula for the  $t$ -test is very similar to the formula for the  $Z$ -test.

$$\bar{X} \pm Z_{\alpha} \sigma_{\bar{X}}, \text{ where } \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \text{ for the } Z\text{-test approach.}$$

The formula revised for the  $t$ -test approach is as follows:

$$\bar{X} \pm t_{\alpha} S_{\bar{X}}, \text{ where } S_{\bar{X}} = \frac{S}{\sqrt{n}}$$

Okay, you cheerily comment. These look much like the same formula. The only difference seems to be the  $t$ -value and the fact you use sample values and not population values for the standard deviation. I think I can do all of this formula, but do I look up the  $t$ -value in the normal table?

You are close to a breakthrough, dear student, very close. However, there is something called the  $t$ -distribution or  $t$ -table. It is different from the  $Z$ -table (normal table). I have reproduced one at the end of this chapter and the end of the textbook. You will need to go to it now or you will not understand anything about the next few paragraphs. It is a tear-out sheet, so remove it and place it side by side to the text material, so you can follow my comments.

Mark your place here and retrieve the chart. I'll wait.

Got it? Good!

Fortunately, the statisticians have provided a  $t$ -table for you rather than making you work with a complicated formula or use the very helpful Excel. (The command for the Excel function is TINV.) There is an explanation at the end of the table itself on how to look up values using the  $t$ -table. Read it. There are three keys to reading the table: know the alpha level desired; know the degrees of freedom; know if you have a one or two-tailed rejection region. For confidence intervals, the test is always two-tailed.

For example, let's suppose that I have set a 5% alpha value with a sample size of 20 and you want to find the appropriate  $t$ -value. You will look in the  $t$ -value table (Table 7.7) where the alpha of 0.05 intersects with the degrees of freedom of 19 ( $n - 1$ ).

Degrees of freedom =  $(20 - 1 = 19)$ .

The  $t$ -value is 2.093.

Find it. I will wait.

If you do not see this value, then get help. Make sure you understand how to get this value. **IT IS CRITICAL THAT YOU ARE ABLE TO LOOK UP  $t$  and  $Z$ -VALUES IN TABLES or USE THE EXCEL FUNCTION.** (I may have said this a time or two before.)

## Substitutions

Most of the time, you will not have the population values. This will often require that you use sample values as estimates of the population values. I will present a brief review of the formulas of interest and the possible substitutions.

If you do not know  $\sigma_{\bar{x}}$ , then you may substitute  $S_{\bar{x}}$ . If you do not know  $\sigma$ , then you may substitute  $S$ . The following substitution is possible.

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

This, like  $Z$ , means that I could rewrite the  $t$ -formulas several ways.

$$\bar{X} \pm t_{\alpha} S_{\bar{x}} \text{ or the equivalent formula } \bar{X} \pm t_{\alpha} \frac{S}{\sqrt{n}}$$

To recap the *Z-formulas*, I could write them as follows.

$$\bar{X} \pm Z_{\alpha} \sigma_{\bar{X}} \text{ or the equivalent formula } \bar{X} \pm Z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

Previously, I asked you to calculate for *Z-values* for each of the three conventional confidence coefficients (99%, 95%, and 90%). I am now asking you to complete the table below (Table 7.5) for the *t-values* using the same 99%, 95%, and 90% confidence coefficients (CC or CL). The sample sizes are 20, 15, and 10. These are small-sized samples. Remember, the *t-values* are sensitive to degrees of freedom. You will look up the correct values under the degrees of freedom ( $n - 1$ ), which will be 19, 14, and 9 in this instance.

Complete the following table (Table 7.5).

**Table 7.5** Exercise in looking up *t-values*.

Confidence Coefficient	Alpha Error	Z-Value	<i>t</i> -value for $n = 20$	<i>t</i> -value for $n = 15$	<i>t</i> -value for $n = 10$
0.99	0.01	2.58			
0.95	0.05	1.96			
0.90	0.10	1.65			

Take the time to look up these values. The solution in Table 7.6 should be used only after you make an attempt at looking up the values. Fill in the nine values in Table 7.5 before continuing. I used the *t-table* at the end of the chapter, but have rounded the values in Table 7.6 to three decimal places, even though the *t-table* at the back of the chapter carries the decimals to five places (d.f. stands for degrees of freedom).

**Table 7.6** Completed results of exercise in Table 7.5.

Confidence Coefficient	Alpha Error	z-Value	<i>t</i> -value for $n = 20$ d.f. = 19	<i>t</i> -value for $n = 15$ d.f. = 14	<i>t</i> -value for $n = 10$ d.f. = 9
0.99	0.01	2.58	2.861	2.977	3.250
0.95	0.05	1.96	2.093	2.145	2.262
0.90	0.10	1.65	1.729	1.761	1.833

Make sure you understand how these values were developed. Use the two-tailed test, since confidence intervals are two-tailed.

Now that you have completed the table, back away from the detail and look at any patterns that exist. Can you make an observation about any obvious pattern?

I am looking, you comment.

See anything yet?

Nope. I may need some help.

If you will notice, the smaller the sample size, the larger the  $t$ -value. With a sample size of 10, the value of  $t = 3.250$  at an alpha of 0.01. With a sample size of 20, the value of  $t = 2.861$ . The reverse is also true. The larger the sample size, the smaller the  $t$ -value.

What do you think happens at  $n \geq 30$ ?

Actually, the  $t$ -value merges with the  $Z$ -value because of the central limit theorem.

If you stop and think about the logic of this, you would agree. Taking a small size sample gives you more opportunity to make an error, so you compensate for the larger chance of error by expanding the interval width.

Think about it. Using the following formula, do the following mental exercise:

$$\bar{X} \pm t_{\alpha} S_{\bar{X}}$$

If you hold the mean and the standard error constant, the interval width hinges on the  $t$ -value you insert. The  $t$ -value depends on the sample size. The smaller the sample size ( $n = 10$  versus  $n = 20$ ), the larger the  $t$ -value. For example, assuming a 95% confidence interval, the  $t$ -value for  $n = 10$  is 2.262, but for  $n = 20$  the  $t$ -value is 2.093. Holding the mean and standard error constant, you can see that the interval width grows larger as the sample size decreases.

$$\bar{X} \pm (2.262)(S_{\bar{X}}) \quad n = 10 \text{ (smaller sample = larger interval)}$$

$$\bar{X} \pm (2.093)(S_{\bar{X}}) \quad n = 20 \text{ (larger sample = smaller interval)}$$

## Rules to Remember

**Table 7.7** Rules for  $Z$  and  $t$  Solutions.

Rule for $Z$ -Solution	Rule for $t$ -solution	Exception to $t$ -Rule
If $n \geq 30$ always use a $Z$ -Solution	If 1) Normal Distribution 2) $n < 30$ 3) $\sigma$ is unknown Always use a $t$ -solution	If 1) Normal Distribution 2) $n < 30$ 3) $\sigma$ is known Always use a $z$ -solution

## One Key Thing

My interpretation does not change regardless of the  $t$ -value or sample size. I am still 95% confident that the true, unknown population mean is between the UCL and LCL. I do, however, hedge my bet by making the interval larger as the sample size is smaller. As the sample size decreases, I have more propensity for error, so to compensate for that error, I broaden my interval. The opposite is true: As my sample size increases, I am still 95% confident, but my interval itself decreases.

Wow, those concepts take a lot of thinking, you comment.

Yep. Go back and think about them again, if you did not pick the concept up the first time.

This part of the course demands a lot of critical reasoning and deductive logic.

## Controlling or Affecting the Interval Width

You may do two things to control the interval width. Remember, the interval width is composed of two pieces— $Z$  or  $t$  and the standard error.

The interval width is as follows:

$\pm Z_{\alpha} \sigma_{\bar{X}}$  if you know population values or  $\pm t_{\alpha} S_{\bar{X}}$  if you know the sample values.

What two things might I do?

- Increase the sample size (as  $n$  increases, the standard error decreases, thus narrowing the interval width).
- Select a smaller or larger alpha value [as alpha decreases from 10% (1.65) to 1% (2.58), the interval increases, thus broadening the interval width].

## Sample Size

As a quality-control engineer or marketing research manager, proper sample size might be very important to you. You do not want to take too large a sample, especially if destructive sampling must be used and the cost of your product is quite expensive. Examples of destructive sampling are testing ammunition or wine tasting.

The formula is

$$n = \frac{(Z)^2 (\sigma)^2}{(e)^2}, \text{ where "e" is the sampling error } (\bar{X} - \mu).$$

Remember, the tolerable error allowable in most manufacturing operations is one-half of the interval range. In manufacturing operations, the interval might be three centimeters as you drill a hole into which you are going to fit a round part. The tolerance could not be three centimeters around the circumference, because this would make the interval six centimeters rather than the three centimeters as specified. The tolerable error is one-half of the interval or 1.5 centimeters, which leaves the interval at three centimeters. In this instance, however,  $e = 3$  centimeters.

Often in real business settings, these three values ( $Z$ , standard deviation, and tolerable error) are not easily determined. The researcher may have to estimate them.

For example, the  $Z$ -value most often used is the one associated with the 95% level. This value would then be 1.96.

The standard deviation is often unknown, but can be estimated by using the range divided by six. The range is the difference between the high and the low, not the difference in the LCL and the UCL. The six is developed by using plus and minus three standard deviations on either side of the mean. This would, of course, assume a normal distribution and would encompass 99.7% of the area under the curve.

You might have to make an educated determination of the error you would find acceptable. Acceptable error is generally what a researcher can tolerate.

Once you have determined the value of " $n$ ", you would usually round the value up to the next integer, because it is always better to make the sample slightly too large rather than too small.

For example, let's assume I want a 95% confidence interval ( $Z = 1.96$ ) and I know the standard deviation is 3 cm with sampling error of 0.5 cm. What sample size must I take to ensure these values are maintained?

$$n = \frac{(1.96)^2(3)^2}{(0.5)^2} = \frac{(3.8416)(9)}{0.25} = \frac{34.5744}{0.25} = 138.2976 = 139$$

My sample size would be 139 to hold a 5% alpha with sampling error of 0.5 cm and a standard deviation of 3 cm.

## Four Properties of a Good Estimator

You will probably never be faced with these issues, but I toss them in to fill space. Don't get me wrong; these are important, but often not something you will ever see again unless your job requires this knowledge. Read about them if you want to. Some Professor somewhere may test you on these, but the only question I might ask is a multiple choice or true/false question on them, such as "The four properties of good estimators are unbiased, efficient, consistent, and sufficient". True or False?

Anytime you have an estimate, it will be composed of two parts—the **process** and the **result**. The *estimator is the process*. The *estimate is the results*. Estimators must be good in order to have acceptable and statistically-accurate results. There are four properties of a good estimator (process).

**In summary they are as follows:**

The estimator must be **unbiased**. This tells me the mean of the sampling distribution of sample means must equal the population mean.

The estimator must be **efficient**. Given the first statement is true, then the estimator must have the smallest possible variance or standard error.

The estimator must be **consistent**. As “ $n$ ” increases, the value of the statistic must approach the parameter. If you are using the mean as the parameter, then the mean of the sample must approach the mean of the population.

The estimator must be **sufficient**. No other estimator can provide more information about the parameter.

Okay, with that over, let’s move on with one final subject—proportions. I have mentioned proportions one other time but have never developed the ideas associated with them. Now I have to take a paragraph or two to develop the idea of proportions.

## Proportions (Same Material as in Chapter 6)

**Proportions** are associated with business decisions that have binary results. As such, you will be faced with developing percentages or ratios. These are referred to as proportions.

Just as there is a sample mean which estimates the population mean, there is a sample proportion which estimates the population proportion. The concept of proportions is exactly the same as the concept associated with means.

As you might imagine, the formulas are going to be different but not impossible to understand and solve. The confidence coefficients for proportions are the same as for  $Z$  and  $t$ -values, so those tables are appropriate to use. The confidence interval formula is as follows:

$$CI = p \pm Z_{\alpha} S_p \text{ where } S_p = \sqrt{\frac{p(1-p)}{n}}$$

This formula is the sample proportion— $p$  is an estimate of  $\pi$ , the population proportion.

Let’s look at an example. Suppose you have an auditor who wants to determine the frequency of errors in a company’s invoicing. Assume that a sample of 100 invoices is selected at random. The auditor want to maintain a 95% confidence interval of the proportion of errors contained in the invoicing process. The sample yields invoices with 10 errors. What is the maximum and minimum invoice error this auditor might expect using a 5% alpha error?

**Solution:**


$$S_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.10)(0.90)}{100}} = \sqrt{\frac{0.09}{100}} = \sqrt{0.0009} = 0.03$$

The confidence Interval is as follows:

$$0.10 \pm (1.96)(0.03) = 0.10 \pm 0.0588 = 0.0412 \leq p \leq 0.1588$$

**Interpretation:** My conclusion is I am 95% confident the true, unknown population proportion of invoices with errors will be between 4.12% and 15.88%. Management must now decide if this error range is acceptable. Proportions are important, so don't just gloss over them.

**You know  
You are  
a Texan  
If:**



- You owe more money on your tractor than on your car.
- Hot water comes out of both taps.

**ALL RIGHTS  
RESERVED**

## Tear-Out Sheet: *t*-Distribution Table

**Table 7.7** *T*-Table.

	Second Row is for One-Tailed HT			Top Row is for CI & Two-Tailed HT	
Two-Tailed Test Alpha/C.L.	0.20/0.80	0.10/0.90	0.05/0.95	0.02/0.98	0.01/0.99
One-Tailed Test Alpha/C.L.	0.10/0.90	0.05/0.95	0.025/0.975	0.01/0.99	0.005/0.995
Degrees of Freedom					
1	3.077684	6.313752	12.70620	31.82052	63.65674
2	1.885618	2.919986	4.30265	6.96456	9.92484
3	1.637744	2.353363	3.18245	4.54070	5.84091
4	1.533206	2.131847	2.77645	3.74695	4.60409
5	1.475884	2.015048	2.57058	3.36493	4.03214
6	1.439756	1.943180	2.44691	3.14267	3.70743
7	1.414924	1.894579	2.36462	2.99795	3.49948
8	1.396815	1.859548	2.30600	2.89646	3.35539
9	1.383029	1.833113	2.26216	2.82144	3.24984
10	1.372184	1.812461	2.22814	2.76377	3.16927
11	1.363430	1.795885	2.20099	2.71808	3.10581
12	1.356217	1.782288	2.17881	2.68100	3.05454
13	1.350171	1.770933	2.16037	2.65031	3.01228
14	1.345030	1.761310	2.14479	2.62449	2.97684
15	1.340606	1.753050	2.13145	2.60248	2.94671
16	1.336757	1.745884	2.11991	2.58349	2.92078
17	1.333379	1.739607	2.10982	2.56693	2.89823
18	1.330391	1.734064	2.10092	2.55238	2.87844
19	1.327728	1.729133	2.09302	2.53948	2.86093
20	1.325341	1.724718	2.08596	2.52798	2.84534

(Continued)

Table 7.7 *T*-Table (Continued).

Two-Tailed Test Alpha/C.L.	0.20/0.80	0.10/0.90	0.05/0.95	0.02/0.98	0.01/0.99
One-Tailed Test Alpha/C.L.	0.10/0.90	0.05/0.95	0.025/0.975	0.01/0.99	0.005/0.995
Degrees of Freedom					
21	1.323188	1.720743	2.07961	2.51765	2.83136
22	1.321237	1.717144	2.07387	2.50832	2.81876
23	1.319460	1.713872	2.06866	2.49987	2.80734
24	1.317836	1.710882	<b>2.06390</b>	2.49216	2.79694
25	1.316345	1.708141	2.05954	2.48511	2.78744
26	1.314972	1.705618	2.05553	2.47863	2.77871
27	1.313703	1.703288	2.05183	2.47266	2.77068
28	1.312527	1.701131	2.04841	2.46714	2.76326
29	1.311434	1.699127	2.04523	2.46202	2.75639
30	1.310415	1.697261	2.04227	2.45726	2.75000
Infinity	1.281552	1.644854	1.95996	2.32635	2.57583

The *t*-table is driven by degrees of freedom ( $n - 1$ ). Knowing the sample size, you will subtract one from  $n$ . This will give you degrees of freedom (free choices). You will then decide if you are using a two-tailed test or a one-tailed test. For now, I am using two-tailed rejection regions, since I am addressing only confidence intervals. Confidence intervals are always two-tailed tests. You will use the references in the top row and trace those to the intersecting cell for the proper degrees of freedom. For example, assuming  $n = 25$ , the degrees of freedom would be 24. At the alpha level of 0.05, I would trace the top row alpha level of 0.05 to the intersecting point of *d.f.* = 24. The value I find is 2.06390 as bolded in Table 7.7.

You can also use the Excel function to calculate each value for *t*. Open Excel. Double click on *f(x)* in the tool bar. Find the function *TINV* which returns the inverse of the *t*-distribution. You have two questions to answer—the probability desired and the degrees of freedom. In the above example, you would type in 0.05 as the probability and 24 as the degrees of freedom. The result will be 2.06390 from Excel which is the same as the table value.

Note one interesting fact (interesting to some of you, but to others a bore). See the last row? If reads “infinity.” At infinity (wherever that is), the values in the *Z*-table (normal table) and the values in the *t*-table become the same. Notice for alpha 0.05, the value from the *Z*-table is 1.96 and from the *t*-table it is 1.95996 or rounded 1.96. Isn't that amazing?

## In-Class Problems Chapter 7

Name: \_\_\_\_\_

Class Day and Time/Section: \_\_\_\_\_ Date: \_\_\_\_\_

Tear out these pages after recording your answers and turn them in as homework. Detailed answers may be found in the solutions and comments answer section.

The following problems should be solved to aid you in understanding confidence intervals.

1. As the marketing director for TXU Electric, you are concerned about the change in average household consumption of electricity each month. You take a sample of 169 houses and find the mean usage to be 1,834 KWH per month. Previously you have determined that the *standard deviation of all consumption* ( $\sigma$ ) has been 260 KWH. You decide to run some statistical tests using an alpha level of 0.10. You have reason to believe the household consumption is normally distributed.
  - a. Calculate the standard error. Remember, the standard error is not the same as the sampling error.
  - b. Calculate the 90% confidence interval.
  - c. What can we say about the size of the sampling error with 90% confidence?
  - d. Explain the meaning of the 90% confidence interval.
2. You are the proprietor of a New York city boutique. You want to know the average age of your customers. You take a *random sample of 25* customers, which yields an average age of 32 with a standard deviation of 8. You have good reason to believe the distribution is normally distributed.
  - a. Determine a 95% confidence interval for the age of your customers.
  - b. Explain the meaning of the interval.
3. In order to determine the life expectancy of the picture tubes of a particular brand of portable television, *a sample of six tubes* was selected randomly from a normal

Picture Tube	Life of the Picture Tube ( $X$ ) (in thousands of hours)	$\bar{X}$	$X - \bar{X}$	$(X - \bar{X})^2$
1	8.2			
2	7.5			
3	9.5			
4	6.5			
5	8.5			
6	7.8			
<b>Total</b>	<b>48.0</b>			

distribution. You want to determine a 90% confidence interval, but you do not have a population, sample variance, or standard deviation, which requires you to calculate a sample value. The results of the sample are shown in the following table.

- a. What is the first step in calculating a sample standard deviation?
  - b. Calculate a 90% confidence interval for the life expectancy of all picture tubes for this brand of portable television.
  - c. Explain the meaning of the interval.
4. A *random sample of 64 children* with working mothers showed the children were absent from school an average of 5.3 days per term with a standard deviation of 1.8 days.
- a. Provide a 96% confidence interval for the average number of days absent per term for all of the children.
  - b. Explain the meaning of the confidence interval.
5. The monthly starting salaries of students who receive an MBA degree have a standard deviation of \$70. *What size sample* should be selected so that there is 0.90 confidence of estimating the mean monthly income within a *sampling error* of \$15 or less? Here again, remember, there is a difference between sampling error and standard error. Here I am interested in the sampling error and not the standard error.
6. The editor of a New York magazine wants to determine the proportion of magazines that have some sort of nonconforming attribute such as excessive rub off, improper page setup, missing pages, duplicate pages, typographical errors, etc. The editor commissions a 200-magazine study and finds 35 contain some type of nonconformance. Using a 90% confidence interval, estimate the true, unknown population proportion.

RESERVED

## Homework Problems Chapter 7

Name: \_\_\_\_\_

Class Day and Time/Section: \_\_\_\_\_ Date: \_\_\_\_\_

Tear out these pages after recording your answers and turn them in as homework. Detailed answers may be found in the solutions and comments answer section.

### Problems:

1. Explain the difference between point estimates and an interval estimate.

### Large-Sized Samples for Means:

2. One hundred 16-ounce jars of Darla Sifuentes' Homemade Hot Sauce average 15.2 ounces. The population standard deviation is known to be 0.96 ounces. At the 95 percent level of confidence, can you assert the jars are filled with a mean of 16 ounces?
3. To estimate the mean expenditure of customers at a local Sonic, students in a marketing research class sample 200 customers and find a mean expenditure of \$5.67 with a standard deviation of \$1.10. What is the 95 percent confidence interval for the mean expenditure of all the customers? Interpret the results.

4. A sample of 121 calls to the 900 number you operate has a mean duration of 16.6 minutes and a standard deviation of 3.63 minutes. You offer a discount, which will be discontinued unless the mean call duration exceeds 18 minutes. At the 90 percent level of confidence, what is your decision?
  
5. A 95 percent confidence interval is constructed such that it yields a lower confidence limit of 62 and an upper confidence limit of 69. Can you conclude from the above that there is a 95% probability that the true, unknown population parameter is between 62 and 69? Explain.
  
6. Jose has a thriving business in Acapulco selling authentic plastic Inca relics to tourists. He selects  $n = 60$  days to estimate his daily profit. However, Jose does not know whether the population of daily profits is normally distributed and is uncertain of how to proceed. Does he proceed with the sample? Why or Why not?
  
7. Georgia Pacific (GP), a major U.S. paper company, decides to harvest a timber stand if it can get an average of at least 700 board feet per tree (bf). A sample of 1,000 trees yields a mean of 695 bf, with a standard deviation of 22.1 bf.
  - a. Calculate the 90 percent confidence interval.
  
  - b. Interpret your answer.
  
  - c. Should GP harvest the stand?

8. Your product requires that a certain component used in your manufacturing process must average 15.2 grams. If you purchase 100 components and find  $\bar{X} = 14.8$  grams with  $s = 3.2$  grams, what would a confidence interval tell you about the advisability of buying more from this supplier? Your product is very delicate and you feel you can tolerate only a 1 percent alpha error.
  
9. A researcher found that a sample of 100 with  $\bar{X} = 50.3$  and  $s = 10.1$  generated a confidence interval of 48.3204 to 52.2796. What level of confidence can be attributed to this interval?
  
10. The Pizza Parlor is considering increasing the price of its large sausage pizza if the average price of its competitors exceeds Pizza Parlor's price of \$12.95. Thirty-seven other pizza places report a mean price of \$12.50 with a standard deviation of \$1.01. Pizza Parlor wants to be 90 percent confident in the decision to raise prices. Can they raise prices?

**Student's t-distribution:**

11. What three conditions must be met before the t-distribution can be used?
  
  
  
  
  
  
  
  
  
  
12. Fat Harry's, a popular student hangout, sells 16-ounce glasses of beer. Ten students purchase a total of 22 glasses and, using their own measuring cup, estimate the mean contents. The sample mean is 15.2 ounces with  $s = 0.86$ . At the 95 percent level of confidence, are the students getting their money's worth? Assume a normal distribution. Interpret the interval.

13. Kendall-Hunt Publishing samples 23 shipments to estimate the mean postal cost. The sample mean cost is \$23.56 with  $s = \$4.65$ . Assume a normal distribution.

a. The senior editor hopes to keep the mean cost under \$23.00. Calculate and interpret the 99 percent interval. Will the editor be happy?

b. Compare the results in “part a” to a 99 percent interval if  $s = \$2.05$ . Explain why there is a difference.

c. Leaving the standard deviation at \$4.65, compare the results in “part a” to a 95 percent interval. Explain the difference.

d. What happens to your calculation if previous records revealed that  $\sigma$  was known? If there is a difference, explain it.

14. During the last Super-bowl Sunday, Adrian Sifuentes and his buddies ordered 27 pizzas from *Pizzas To Go* at different times. The average delivery time proved to be 23.7 minutes, with a standard deviation of 10.7 minutes. Assume a normal distribution. Feeling this was far too long a delay in their culinary world, Adrian and his friends decided to buy the 28<sup>th</sup> pizza elsewhere if it appeared that the delivery time for *Pizzas To Go* exceeded 30 minutes. Set alpha at 1 percent. Will they order elsewhere?

### Determining the Sample Size:

15. Holiday Inn wants to develop a 99 percent interval for the mean number of rooms occupied each night at its locations around the nation. How many nights must be included in the sample if an error of 50 rooms can be tolerated and a pilot sample reveals  $s = 165$  rooms?

### Proportions:

16. The Washington Post reported that 68% of all high-school students had computers at their home. If a sample of 1,020 students is taken, which reveals that 673 students have computers at home, does a 99% confidence interval support The Washington Post?

17. A local travel agency with national affiliations commissioned a survey of travelers to England. What is the 96% confidence interval for those who hold American passports if 1,098 out of 3,769 in the survey hold a passport?

18. In order to curtail insider trading, the SEC requested information on the proportion of bank holding companies whose officers directly hold more than 50% of the bank's outstanding stock. Of the 200 banks selected at random, 79 reported that insiders held a majority of their stocks. What is the 90% confidence interval for the proportion of all bank holding companies whose officers hold at least 50% of the outstanding stock?

**KH**

**ALL RIGHTS  
RESERVED**

## Extra Problems Chapter 7

Name: \_\_\_\_\_

Class Day and Time/Section: \_\_\_\_\_ Date: \_\_\_\_\_

Tear out these pages after recording your answers and turn them in as homework. Detailed answers may be found in the solutions and comments answer section.

1. Papa's Pizza delivers pizza throughout its local market area at no charge to the customers. Customers, however, often tip the driver. The owner is interested in estimating the mean tip income per delivery. A random sample of 12 deliveries has been selected and the tips were as follows. Assume a normal distribution.

Tips	If Needed	If Needed	If Needed
\$2.25			
\$2.50			
\$2.25			
\$2.00			
\$2.00			
\$1.50			
\$0.00			
\$2.00			
\$1.50			
\$2.00			
\$3.00			
\$1.50			

- a. Using three decimal point accuracy, what is the best point estimate of the true, unknown mean of the population?
- b. Determine the sample standard deviation. Carry your answers to 6 decimal points as you make the calculations. Round final answer for  $S$  to four decimal points for use in part c. Remember the standard deviation is the square root of the mean of the squared deviations around the mean.

- c. Develop a 90% Confidence Interval.
  - d. Interpret the interval calculated in part c.
2. Tom Thumb, a grocery chain of several hundred stores, operates pharmacies inside the grocery store. An internal audit of 300 randomly-selected prescriptions from 10 stores in the North Central Region of Texas revealed a sample mean of \$14.23 with a sample standard deviation of \$3.00.
- a) Determine a 90% Confidence Interval.
  - b) Interpret the results.
  - c) One retail outlet recently reported monthly revenue of \$7,392 from 528 prescriptions. Should this outlet be audited? Why or Why Not?
3. BMW's Mini Cooper has a base sticker price of \$20,200. Often an automobile will sell above sticker when the customer purchases add-on extra options. Mini Cooper conducted a survey of 179 recent purchases and found the average add-on's totaled \$5,000. Suppose that the accessories purchased by all Mini Cooper customers have a standard deviation of \$1,500.
- a) Calculate a 95% Confidence Interval.
  - b) Interpret the results.
  - c) Determine the margin of error in estimating the average cost of accessories for Mini Coopers.
  - d) Using the same alpha level of 0.05, what sample size would be required to reduce the margin of error by 50%?
4. Grandy and Morgan CPA, Inc. recently conducted an audit for a discount chain. One part of the audit involved developing an estimate for the mean dollar error in total charges that occur during the checkout process. In a sample of 20 bills, the sample mean was found to be \$0.77 with a standard deviation of \$1.19. Assume a normal distribution.

- a) Develop the standard error.
  - b) Calculate a 90% confidence interval.
  - c) Interpret the results.
  - d) Based on the sample data and the interval, by what amount is the customer being over- or under-charged?
5. Billbank Software operates a call service center in Oklahoma City where questions about the software may be answered. A team of employees was asked to study the average length of time the service representatives spend with customers. A random sample of 25 calls was selected. The sample mean was 7.088 minutes with a standard deviation of 4.64 minutes.
- a) What is the best point estimate of the population mean?
  - b) Why is it not best to solely depend on a point estimate?
  - c) Determine a 95% Confidence Interval.
  - d) Interpret the results.
6. National Recycling, Inc. has a contract with the City of Southlake to pick up recycled products. Part payment of the contract depends in part on the amount of recycled product that is processed. The city has requested both a 99% and a 90% confidence interval prior to contract signing. A sample of 100 customers is selected. The sample mean is 40.78 pounds with a standard deviation of 12.6 pounds.
- a) What is the standard error?
  - b) Determine a 99% Confidence Interval.

- c) Determine a 90% Confidence Interval.
  - d) What is the margin of error for the 99% Confidence Interval?
  - e) What is the margin of error for the 90% Confidence Interval?
  - f) Does lowering the confidence interval lower or raise the margin of error? Comment on this issue.
7. Tarrant Appraisal District for Tarrant County, Texas wants to know the mean amount of dollars that home owners pay annually for property taxes for single family homes. To determine this, a random sample of 200 tax bills is selected. From the sample, the sample mean is \$5,230 with a sample standard deviation of \$500.
- a) Develop the standard error.
  - b) Develop a 90% confidence interval.
  - c) Interpret the result.
8. Dr. Pepper® randomly sampled 100 cans of Diet Dr. Pepper® to determine the mean fill level of their 12 ounce cans. The sample yielded a mean of 12.09 ounces. There is a known population standard deviation of 0.2 ounces.
- a) What value serves as a point estimate of the unknown population mean?
  - b) Develop the standard error.
  - c) Develop a 95% Confidence Interval.
  - d) Interpret the results.
  - e) What is the margin of error?

- f) Suppose you increase the sample size to 400 cans, what happens to the standard error?
- g) Using the sample size 400, comment on the margin of error and the revised 95% Confidence Interval.
9. A & A Convenience Stores is owned and operated by Art and Annie Wilcox. As proprietors, they are interested in knowing the average gallons used in filling up customers' cars each time they fill up. They are comfortable with a 90% confidence interval. The tolerable error (margin of error) is 0.5 gallons. The standard deviation is felt to be 4.85 gallons from a previous pilot sample of 20 customers. Calculate the required sample size to hold the error to 0.5 gallons at an alpha level of 10%.
10. Quick Lube, Inc. operates an automobile lubrication center. In hope of increasing repeat business, the manager has a system that tracks the following: type of service, date of service and the mileage of the car when service was rendered. About every three month a reminder card is sent to the customer that offers a discount. A random sample of 100 customers has found that 62 out of the 100 will return within a month of the mail out.
- Determine the point estimate of the true, population proportion.
  - Calculate a 95% Confidence Interval for this proportion.
  - Interpret the results.

**KH**

**ALL RIGHTS  
RESERVED**