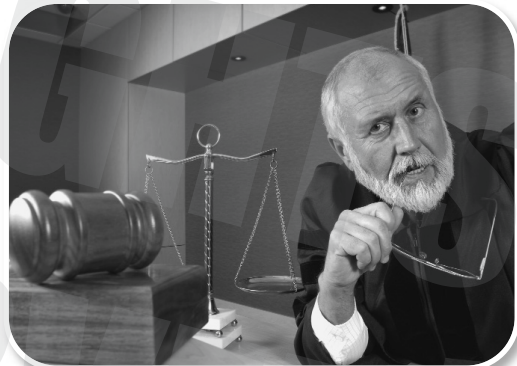


Hypothesis Testing and Confidence Intervals for Two Populations

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

Attorney: Doctor, how many of your autopsies have you performed on dead people?

Witness: All of them. The live ones put up too much of a fight.



Two Populations

What exactly do I mean by **two-population hypothesis testing**? In many business settings, the business manager will often find himself (not a gender statement) faced with asking questions about a product his company produces or a product his company wants to produce.

The question might be as simple as comparing the mean life of two products—his product and his competitor's product. The goal is to determine if his product is better or worse than the competitor's product.

Are North Slope Ski Boots more durable than Head Ski Boots?

Plant A is producing 200 metric tons and Plant B is producing 185 metric tons even though the two have similar equipment and processes. The question might be:

Do the workers in Plant A produce more than the workers in Plant B?

Your plant uses method A to produce products. A time and motion engineer claims if you switch to method B, the production output will increase significantly. Before spending the money on the switch over, you need to test the claim. The question might be:

Does production Method A yield higher output than Method B?

These are real questions often asked by the owner or manager of many businesses. To stay or become competitive, you must always be cognizant of how your product or service compares to your competitor's product or service. The continual demand for your product and the proper pricing of your product will depend on how your product matches your competitor's product. If they are better and less expensive, you will lose market share. Often, consumer advocacy groups will test the claims made about the quality of products. This testing will serve to provide information to the general public so they can make informed decisions before they purchase a product.

In order to understand two-population testing, let's organize it into two general approaches— independent sampling and paired sampling. In both cases I am interested in the *differences in the two parameters* not the individual values generated from each of the samples. For example, assume there are two companies that produce tires. One tire has a mean life of 36,000 miles and the other comparable tire has a mean life of 33,000 miles. I have no interest in the 36,000 or the 33,000, but have a great deal of interest in the 3,000 difference.

The design of my experiment might compare two samples independently drawn at random from two separate and distinct populations (independent sampling). Alternatively, the design of my experiment might match two separate pairs of data (dependent testing, or paired sampling).

Independent sampling collects separate samples from each population. The researcher would then evaluate the difference in the means of the samples. The sample sizes do not have to be equal. For example, if $n_1 = 25$ for one population and $n_2 = 50$ for a second population, the results will not be compromised because the sample sizes are unequal.

Paired Sampling, on the other hand, takes the same number of observations from each population. Each group is matched as closely as possible. These populations vary in only one relevant detail—the test characteristic being measured. Matched pairs are most often found in medical research, where for example two groups will be matched as closely as possible as to age, gender, weight, level of activity, cholesterol, and any other factor which might affect blood pressure. These two groups are often referred to as twins. One group would be given Medication A and the other Medication B. One of the two medications is a placebo and the other is the real medication. Neither the nurse nor the patient knows which group is receiving the actual medicine. Over a period of time, the blood pressure for each group is measured often and the result will determine if the group taking the real medication responds better than the group taking the placebo. Paired sampling can also work for on the job training programs with output before training being matched to output after training.

I can use hypothesis testing or confidence intervals with both independent sampling or paired sampling. Either hypothesis testing or confidence intervals testing will reveal if there is

a statistically-significant difference between the two populations. However, there is one main difference in the two approaches. If I am interested in knowing *how much the difference is*, I must (repeat *must*) use confidence intervals.

The hypothesis testing approach will answer the question—*yes*, there is a difference or *no*, there is no difference. Confidence intervals, on the other hand, will tell you if there *is a statistically-significant difference and how much that difference is*.

The formulas used for confidence intervals and the formulas for hypothesis testing are essentially the same formulas.

Wait a minute, you say, while wiping the sweat from your brow. I do not understand how you can claim the formulas are the same.

Easily, I reply, since they are. All you do is solve the same equation for a different value.

You respond, I am not from Missouri, I still need you to “*splain yourself*” as Ricky Ricardo often said to Lucy (shows my age).

Let’s look at the formulas associated with large sample sizes for confidence intervals as compared to the formulas for hypothesis testing.

Confidence Intervals—Large Samples

$$\bar{X}_1 - \bar{X}_2 \pm Z_{\alpha} S_{\bar{X}_1 - \bar{X}_2}$$

(Definition of terms later)

Hypothesis Testing—Large Samples

$$Z_{\text{test}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_{\bar{X}_1 - \bar{X}_2}}$$

(Definition of terms later)

As you look at the above comparisons, notice that the hypothesis testing formula and the confidence interval formula are essentially the same. The hypothesis test formula solves the confidence interval formula for Z . All formulas in this chapter have the same mirrored image. There are two basic categories for two-population testing. Under independent testing there are two sub-categories, with two other ones under small samples.

Independent Testing:

Large size samples

Small samples

Equal variances

Unequal variances

Dependent Testing:

Matched pairs

Given this similar relationship, the test statistic you select depends on the following question “*Am I interested in knowing if there is a difference or must I know the amount of the difference?*” The answer to this question will determine which solution—hypothesis testing or confidence intervals—you will use. Either the confidence interval or the hypothesis testing approach tells you if there is a difference. Confidence intervals is the only one, however, that tells you the amount of the difference. (I think I have said this before, right?)

A summary comparison of the respective formulas follows (Table 9.1), as well as some of the key factors associated with each of the methods. You may want to copy this table for future reference.

Table 9.1 Comparison of CI and HT mirrored image formulas.

Confidence Interval	Hypothesis Testing
<p>Independent Samples Large–CI:</p> <ul style="list-style-type: none"> * Large samples. * Both n's must be ≥ 30. * Solution will use Z. 	<p>Independent Samples Large–HT:</p> <ul style="list-style-type: none"> * Large samples. * Both n's must be ≥ 30. * Solution will use Z. * $\mu_1 - \mu_2 = 0$ (always)
$\bar{X}_1 - \bar{X}_2 \pm Z_\alpha S_{\bar{X}_1 - \bar{X}_2}$	$Z_{\text{test}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_{\bar{X}_1 - \bar{X}_2}}$
$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$	$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$
<p>Independent Samples Small–CI:</p> <ul style="list-style-type: none"> * Small samples—equal variances. * $\sigma_1^2 = \sigma_2^2$. * S_1^2 will not equal S_2^2. * Either n is less than 30. * Variances may be pooled. * The pooled variance is a weighted average and will fall between the two sample variances. * Solution will use t. * Degrees of freedom are $n_1 + n_2 - 2$. 	<p>Independent Samples Small–HT:</p> <ul style="list-style-type: none"> * Small samples—equal variances. * $\sigma_1^2 = \sigma_2^2$. * S_1^2 will not equal S_2^2. * Either n is less than 30. * Variances may be pooled. * The pooled variance is a weighted average and will fall between the two sample variances. * Solution will use t. * Degrees of freedom are $n_1 + n_2 - 2$.
$\bar{X}_1 - \bar{X}_2 \pm t_\alpha \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$	$t_{\text{TEST}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$
<p>Pooled Variance Calculation.</p> $S_p^2 = \frac{S_1^2 (n_1 - 1) + S_2^2 (n_2 - 1)}{n_1 + n_2 - 2}$	<p>Pooled Variance Calculation.</p> $S_p^2 = \frac{S_1^2 (n_1 - 1) + S_2^2 (n_2 - 1)}{n_1 + n_2 - 2}$

<p>Independent Samples Small-CI:</p> <ul style="list-style-type: none"> *Small samples—unequal variances. *$\sigma_1^2 \neq \sigma_2^2$. *$S_1^2$ will not equal S_2^2. *Either n is less than 30. *Variances cannot be pooled. *Solution will use t-prime. *Degrees of freedom must be calculated. *Always round degrees of freedom down. 	<p>Independent Samples Small-HT:</p> <ul style="list-style-type: none"> *Small samples—unequal variances. *$\sigma_1^2 \neq \sigma_2^2$. *$S_1^2$ will not equal S_2^2. *Either n is less than 30. *Variances cannot be pooled. *Solution will use t-prime. *Degrees of freedom must be calculated. *Always round degrees of freedom down.
$\bar{X}_1 - \bar{X}_2 \pm t'_\alpha \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$	$t_{\text{TEST}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$
<p>Degrees of freedom are calculated.</p> $\text{d.f.} = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$	<p>Degrees of freedom are calculated.</p> $\text{d.f.} = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$
<p>Dependent Matched Pairs-CI:</p>	<p>Dependent Matched Pairs-HT:</p>
<p>All matched pairs must have two distinct data sets which are matched as closely as possible for all characteristics, with the exception of the one characteristic which is the characteristic you wish to measure.</p>	
$\bar{d} \pm t_\alpha \frac{S_d}{\sqrt{n}}$	$t_{\text{test}} = \frac{\bar{d} - (\mu_1 - \mu_2)}{\frac{S_d}{\sqrt{n}}}$
$\bar{d} = \frac{\sum d_i}{n}$	$\bar{d} = \frac{\sum d_i}{n}$
$S_d = \sqrt{\frac{\sum d_i^2 - n(\bar{d}^2)}{n - 1}}$	$S_d = \sqrt{\frac{\sum d_i^2 - n(\bar{d}^2)}{n - 1}}$

Let's look at several examples. I will work through several examples of confidence interval problems. Let's begin with independent sampling with large size samples.

Independent Sampling—Large Size Samples—CI

Let's assume you have two answering devices you are considering for purchase. You are the purchasing agent of a Fortune 500 company, so your decision could have a major impact. These devices are useful in answering and queuing calls from your customers. You authorize an outside bureau to test the two devices. In the trial tests, you determine the following (Table 9.2).

Table 9.2 Answering device comparison.

Device A	Device B
$n_1 = 75$ calls	$n_2 = 70$ calls
$S_1 = 4.8$ seconds	$S_2 = 3.8$ seconds
$\bar{X}_1 = 25.2$ seconds	$\bar{X}_2 = 21.3$ seconds

First, I observe that the sample sizes both exceed 30. This is important. If the sample sizes equal or exceed 30, the central theorem may be assumed, which in turn assures the distributions are normally distributed. Both sample sizes must be equal to or exceed 30 for me to apply the large size sample approach. If one of the samples is less than 30, then I must apply a different technique (t -distribution and small size samples).

Second, I observe that I do not know the standard deviation or the standard error of the population. I do, however, know the standard deviation of the sample. This again is important because it determines the technique I will use.

The question is simply this: "Which device do I select?" Or, said differently, "Is there a statistically-significant difference between the mean answering times of the two devices or can the difference be explained by sampling error?" I want to select the device that responds more promptly in servicing my customers. Price considerations are about the same, so this does not factor into my decision. From the initial observations, Device B looks as though it is faster (21.3 versus 25.2 seconds). I know enough about the comparison of products to know that I must test the results and *cannot rely on the sample means directly*. I want to look at the differences in the sample means rather than the value of each sample mean alone.

When I say I want to look at the differences, I am asserting that the difference in the two means is equal and that any error that occurs is due to sampling error alone.

This can be written as follows:

$$\mu_1 = \mu_2$$

or

$$\mu_1 - \mu_2 = 0 \text{ (This is an important concept. Remember this for later.)}$$

I want to perform my test using an alpha level of 0.10. Since this is a confidence interval, there is no need for a formal statement of the null and alternate hypotheses.

From Table 9.2 the confidence interval formula is as follows:

$$\bar{X}_1 - \bar{X}_2 \pm Z_\alpha S_{\bar{X}_1 - \bar{X}_2}$$

Go to Table 9.2 and make sure you can find this formula. It is a mirror image of the hypothesis formula.

$\bar{X}_1 - \bar{X}_2$ is the difference between the two means.

Z_α is the table look up Z-value at the chosen alpha level.

$S_{\bar{X}_1 - \bar{X}_2}$ is the standard error of the difference between the two means.

First, calculate the standard error of the difference.

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Insert the values using the information given in Table 9.2.

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{(4.8)^2}{75} + \frac{(3.8)^2}{70}}$$

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{(23.04)}{75} + \frac{(14.44)}{70}}$$

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{0.3072 + 0.2063}$$

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{0.5135}$$

$$S_{\bar{X}_1 - \bar{X}_2} = 0.7166$$

Next, I return to the confidence interval formula for large size samples. Inserting the proper values I will have the following calculations.

$$\bar{X}_1 - \bar{X}_2 \pm Z_\alpha S_{\bar{X}_1 - \bar{X}_2}$$

$$(25.2 - 21.3) \pm (1.65)(0.7166)$$

$$3.9 \pm 1.18$$

$$2.72 \text{ to } 5.08 \text{ seconds.}$$

Incredible, you say, but I have a few questions.

Okay, shoot.

Well, first I understand where you got the 0.7166.

Good start.

I do not, however, know where you got the 1.65.

Okay, I say. This is a normal table look up of the Z-value at an alpha of 0.10.

I found it as follows: Since I am dealing with confidence intervals, I will divide the alpha by 2 ($0.10 \div 2 = 0.05$). One-half of alpha will be in the left tail and one-half will be in the right tail. I am assuming a normal distribution. Since both sample sizes are greater than 30, the central limit theorem applies. I know the normal distribution is symmetrical, so I can subtract 0.05 from 0.5000 (one-half of the area to the left of the mean and one-half to the right of the mean) to get a value of 0.4500. This is the area under the curve, so I will then go to the normal distribution table and find the area closest to 0.4500. I find this value in the body of the table and not on the edges. The edges are Z and the middle of the table is area under the curve. The closest value is 0.4505, which then cross references to a Z-value of 1.65.

Are there any other questions?

Yes, where did you get the 25.2 and the 21.3?

Oh, you are going to feel funny when I answer that.

Why, you ask?

Well, what was the mean of the sample of Device A?

You look and sheepishly say, oh, 25.2.

Okay, Okay, you say, enough already. I can see the mean of Device B is 21.3, too.

Great!

Now the rest of the solution is interpretation.

I will recap the calculations to this point.

The confidence interval is 2.72 seconds to 5.08 seconds. Both are positive values.

How do I interpret it, you ask?

Notice that the interval does not contain zero.

So what, you say.

For an interval to contain or not contain zero is significant.

If the interval *does* contain zero, I would conclude there is *no statistically-significant difference* in the two devices. Both devices would be considered to be statistical equivalents. Any difference would be attributed to sampling error. Sampling error that is random is okay.

Remember I am really making the following statement: $\mu_1 - \mu_2 = 0$ (This was established just above in the chapter, and I suggested you remember it for later. Now is later).

However, since the interval *does not* contain zero, I will conclude that I am 90% confident there is a *statistically-significant difference* between the two devices. At the alpha 0.10 level, the difference is too great to be explained by sampling error alone.

Having said that, I must now conclude which device is better. To answer that question I look at the means of the two samples. Device A had a mean response time of 25.2 seconds. Device B had a mean response time of 21.2 seconds. Device B is faster than Device A. *I can conclude at the 90% level of confidence that Device B will respond from 2.72 seconds to 5.08 seconds faster than Device A.* If the price consideration is not a real factor and speed is the determining factor, then Device B should be chosen over Device A.

One word about the signs of 2.72 and 5.08. Both signs are positive in this instance. If the signs were negative by subtracting Device B from Device A, I could actually conclude Device A responds 2.72 to 5.08 seconds *slower*, which is another way of saying the same thing. The point is this: when both signs are the same, the positive or negative sign really does not matter but is only an aid in interpreting the results. If you interpret the result from the viewpoint of Device B, you may say it is faster by 2.72 to 5.08 seconds. If you interpret the results from the viewpoint of Device A, you may say Device A is slower than Device B by 2.72 to 5.08 seconds. If you interpret the results as faster or slower, it makes no difference as long as your point of reference is understood.

Of course it goes without saying, but I am going to say it anyway: Any conclusion will depend on the question posed in the problem. Sometimes slower is better and sometimes faster is better. Sometimes higher is better and sometimes lower is better. The conclusion depends on the problem content and must be accessed individually at the time you make your final statement as to the meaning of the interval. Making a clear statement in your concluding remarks is extremely important in proper interpretation.

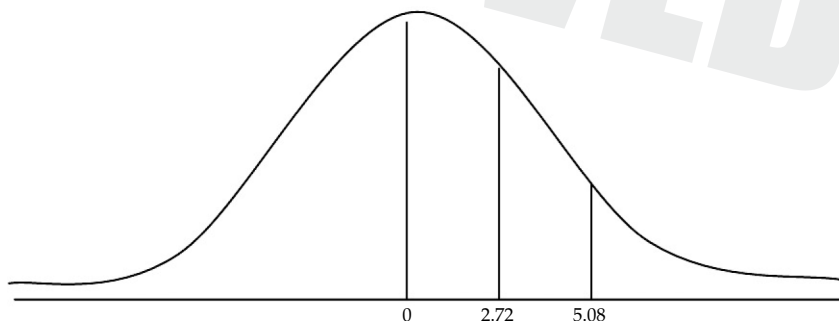


Figure 9.1 Answering Devices CI Solution.

Figure 9.1 shows zero is not in the interval, which indicates there is a significant difference between the two answering devices. Zero must be in the interval for there to be no significant difference. If there is no significant difference, any variation can be explained by random sampling error. This is not the case in this problem. Here, there is a statistically-significant difference and random error has not occurred.

Independent Sampling—Small Size Samples—CI

Independent sampling can have both large and small size samples. I have just shown you the calculations for large size samples. Large samples use a Z-value. For small samples I will switch to t -values. There are, however, two approaches to small size sample calculations based on the *variances of the two populations*.

To use the small size sample and t -value approach, *at least one* of the samples must have a sample size less than 30. Both can be less than 30 but *one must be less than 30*. One sample size can be 40 and the second sample size can be 28. When this occurs, I will use the t -approach.

Small samples can have two approaches.

Approach One: The *populations have equal variances*.

$$\sigma_1^2 = \sigma_2^2$$

Approach Two: The *populations have unequal variances*.

$$\sigma_1^2 \neq \sigma_2^2$$

Note that the parameter in both approaches is the *population variance not the sample variance*. The sample variances ($S_1^2 \neq S_2^2$) will *never equal* each other.

Approach One: Equal Variances (Small Sample)—CI

If the variances of the two populations can be assumed to be equal but unknown, you can pool the variances of the samples. Since I am using a t -solution, I need the degrees of freedom. Pooling the variances makes life much simpler, because I determine the degrees of freedom by $(n_1 + n_2) - 2$. Degrees of freedom are a necessary evil when I look up an appropriate t -value. Let's look at an example.

Let's assume a certain concession stand dispenses soft drinks at an event. A sample of 15 cups yields a mean fill level of 15.3 ounces with a *variance* of 3.5 ounces. This appears to be under the desired level of 16 ounces, so you make an adjustment to the dispensing equipment. After adjusting the fill level, a second sample of 10 cups is taken with a mean fill level of 17.1 ounces and a *variance* of 3.9 ounces. It is reasonable to assume the population variances are equal, since I am using one machine to fill the cups. I want to construct a 95% interval for *the difference* in fill levels.

Side Note

There is a statistical method you can use to determine if the variances are equal. It is the F -test. The F -test places the larger variance over the smaller variance. The formula and calculations are as follows:

$$F = \frac{S_1^2}{S_2^2} = \frac{3.9}{3.5} = 1.114 \text{ (This is my calculated value.)}$$

The null and alternate hypotheses I am testing are as follows:

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_a: \sigma_1^2 \neq \sigma_2^2$$

I have the calculated value (1.114), but now need a critical value of F . Since I place the larger variance over the smaller variance, the F -value will be greater than 1. This forces a two-tailed test (see the null hypothesis) to become a one-tailed, right rejection region test. The procedure requires alpha to be divided by 2, so at the 0.05 alpha level, the alpha in the right tail would be 0.0250.

I now must find a critical F -value. The best approach is to use Excel. Go to Excel. Click on $f(x)$ in the toolbar. Scroll down and find "FINV." Open it. You have three questions to answer. First, the probability is 0.0250 (one-half of the alpha). Second, you are asked to provide two different degrees of freedom. One will be the numerator (degrees_freedom1) and the second will be the denominator (degrees_freedom2). Here the numerator has 9 degrees of freedom ($n-1$ or $10-1$). The denominator has 14 degrees of freedom ($n-1$ or $15-1$). After inserting these values, the resultant will be 3.2093.

Figure 9.2 shows the results of my example:

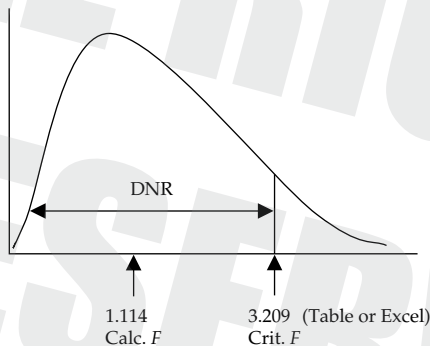


Figure 9.2 F -Distribution (Right Skewed Distribution).

Alternately, you can type into your browser " F distribution" to search the Internet for an F -table. I discovered one at www.statsoft.com/textbook/sttable.html.

Table 9.3 is an F -table and is partially reproduced by using the Excel function for alpha 0.0250. Other tables can be reproduced for other alpha levels—0.10, 0.050, or 0.010 for example.

To read Table 9.3, I use the degrees of freedom for the numerator of 9 ($10 - 1$) and the denominator of 14 ($15 - 1$). The critical F -value is 3.209. Since 1.114 (calculated F -value) is less than 3.209 (critical F -value), I would DNR (accept) the null hypothesis. The null hypothesis was stated above and the variances of the populations are equal, $H_0: \sigma_1^2 = \sigma_2^2$. I DNR the null and conclude the variances of the populations are equal. This allows me to pool my variances as intuitively suggested. To test the variance of the population, I use the variance of the sample as

representative of the population. Once again, this is parametric testing. I use a sample to test the population value.

I am now ready to determine my confidence interval using the following formula. I suggest you look at Table 9.1 and find the formula for small samples with equal variances.

Table 9.3 F-table for alpha 0.0250.

		Numerator Degrees of Freedom for Alpha 0.0250									
d.f. 1		1	2	3	4	5	6	7	8	9	10
D e n o m i n a t o r s D e g r e e s o f F r e e d o m	1	647.78	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28	968.63
	2	38.506	39.000	39.165	39.248	39.298	39.331	39.355	39.373	39.387	39.398
	3	17.443	16.044	15.439	15.101	14.885	14.735	14.624	14.540	14.473	14.419
	4	12.218	10.649	9.979	9.605	9.365	9.197	9.074	8.980	8.905	8.844
	5	10.007	8.434	7.764	7.388	7.146	6.978	6.853	6.757	6.681	6.619
	6	8.813	7.260	6.600	6.227	5.988	5.820	5.700	5.600	5.523	5.461
	7	8.073	6.542	5.890	5.523	5.285	5.119	4.995	4.899	4.823	4.761
	8	7.571	6.060	5.416	5.053	4.817	4.652	4.529	4.433	4.357	4.295
	9	7.209	5.715	5.078	4.718	4.484	4.320	4.197	4.102	4.026	3.964
	10	6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779	3.717
	11	6.724	5.256	4.630	4.275	4.044	3.881	3.759	3.664	3.588	3.526
	12	6.554	5.096	4.474	4.121	3.891	3.728	3.607	3.512	3.436	3.374
	13	6.414	4.965	4.347	3.996	3.767	3.604	3.483	3.388	3.312	3.250
	14	6.298	4.857	4.242	3.892	3.663	3.501	3.380	3.285	3.209	3.147
	15	6.200	4.765	4.153	3.804	3.576	3.415	3.293	3.199	3.123	3.060
	16	6.115	4.687	4.077	3.729	3.502	3.341	3.219	3.125	3.049	2.986
	17	6.042	4.619	4.011	3.665	3.438	3.277	3.156	3.061	2.985	2.922
	18	5.978	4.560	3.954	3.608	3.382	3.221	3.100	3.005	2.929	2.866
	19	5.922	4.508	3.903	3.559	3.333	3.172	3.051	2.956	2.880	2.817
	20	5.872	4.461	3.859	3.515	3.289	3.128	3.007	2.913	2.837	2.774
	21	5.827	4.420	3.819	3.475	3.250	3.090	2.969	2.874	2.798	2.735
	22	5.786	4.383	3.783	3.440	3.215	3.055	2.934	2.839	2.763	2.700
	23	5.750	4.349	3.751	3.408	3.184	3.023	2.902	2.808	2.731	2.668
	24	5.717	4.319	3.721	3.380	3.155	2.995	2.874	2.780	2.703	2.640

I will wait while you look for the formula.

I hope you found it. It is written below:

$$\bar{X}_1 - \bar{X}_2 \pm t_\alpha \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$$

You comment that this formula looks familiar.

You are correct. Great observation! The main differences are the use of a t -value and not a Z -value and the pooling of the variances.

Pooling the variances gives me a weighted average of the two individual variances. *The pooled variance will not be less than the smallest variance (3.5) nor higher than the highest variance (3.9).* The pooled variance is a weighted average of the individual sample variances. The sample variances do not equal ($3.5 \neq 3.9$), nor will they ever equal. The variances I assume to be equal are the *true unknown, population variances*.

I have a question, Professor.

Okay, go ahead.

Why do I care about pooling the variances?

Good question. I have briefly addressed that issue but do not mind covering it again, since you are being bombarded by so many new ideas and terms. I am using a t -test. A t -test requires I know the degrees of freedom in order to look up the critical value. By pooling the variances, I can easily identify the degrees of freedom as $(n_1 + n_2) - 2$; $(10 + 15) - 2 = 23$ degrees of freedom.

Okay, you respond, but why did you subtract 2? Degrees of freedom in previous chapters has been $n - 1$.

Yes, but that was for one population. Now I have two populations, so I will subtract 2 rather than 1. I lose a degree of freedom for each population.

My first calculation is of the pooled variances. That formula and calculations are as follows:

$$S_p^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

$$S_p^2 = \frac{(3.5)(15 - 1) + (3.9)(10 - 1)}{15 + 10 - 2}$$

$$S_p^2 = \frac{49 + 35.1}{23}$$

$$S_p^2 = 3.66 \text{ (Falls between 3.5 and 3.9 and it must).}$$

Notice that I used 3.5 and 3.9 without squaring them. Why do you suppose I did not square the variances?

You respond, that's obvious, you gave us the variance. Variance is already squared. To square it again would not make it the variance, but variance squared.

My head is spinning, but excellent response and a bit scary! Could it be you are understanding this stuff? (Okay, I admit, maybe that is stretching things a bit.)

The next step is to determine the confidence interval using the pooled variance and the t -value. I need a critical value of t using an alpha value of 0.05 (95% CL). I have been able to pool the variances. With degrees of freedom of 23 ($n_1 + n_2 - 2$, which is $15 + 10 - 2 = 23$ degrees of freedom), I look up the t -value of 2.06866 (Round to 2.069).

Try it and make sure it does. The t -value table is Table 7.7 in Chapter 7 or is at the back of this textbook. Take a minute or two to check this out or use the Excel function to confirm the critical t -value.

The calculations are as follows:

$$\begin{aligned} \bar{X}_1 - \bar{X}_2 \pm t_\alpha \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}} \\ 15.3 - 17.1 \pm (2.069) \sqrt{\frac{3.66}{15} + \frac{3.66}{10}} \\ -1.8 \pm (2.069) \sqrt{0.244 + 0.366} \\ -1.8 \pm (2.069) \sqrt{0.610} \\ -1.8 \pm (2.069) (0.781) \\ -1.8 \pm 1.616 \\ -0.184 \text{ to } -3.416 \text{ (The values are both negative numbers;} \\ \text{same sign).} \end{aligned}$$

Interpretation

From observation I do not find a zero in the interval. Both signs are negative, thus zero cannot be in the interval. Not finding a zero means that both fill weights are not statistical equivalents. I must, therefore, conclude there is a statistical difference in the fill weights before and after the adjustment. I wanted to make a slight adjustment in the fill level but apparently made a bigger one than I wanted. I wanted the difference before and after the adjustment to be explained by random sampling error. It is not.

What I must conclude is that there is a statistically-significant difference between the 17.1 ounces and the 15.3 ounces. The difference cannot be explained by random sampling error alone.

Statement #1: I am 95% confident the fill level of the cups, *after the adjustment*, is greater by between 0.184 to 3.416 ounces.

Statement #2: I can also say I am 95% confident the fill level of the cups, *before the adjustment*, is less by between 0.184 to 3.416 ounces.

Either statement is correct. It depends on your point of reference—before or after the adjustment. The sign really does not affect the absolute value of the concluding statement.

Approach Two: Unequal Variances (Small Sample)–CI

Working with the assumption of unequal variances presents an entirely different problem. Statisticians have concluded that the distribution of unequal population variances does not fit a conventional t -distribution like equal variances do. *With unequal variances, I must calculate the degrees of freedom* rather than take the sum of the two variances less 2 $(n_1 + n_2) - 2$. Calculation of the degrees of freedom is done by using the following formula. Complicated as it looks, it is not. Tedious it is, but it is not complicated. You can make errors if you do not set the calculations up carefully and watch your order of operation.

$$\text{d.f.} = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$$

When I make this calculation the t -value is referred to as t' -value, which is called t -prime. I am still going to use the t -table (Table 7.7 in Chapter 7 or back of the textbook). Excel is also helpful as I have pointed out before.

The confidence interval formula for unequal variances is similar to the formula for equal variances. I have shown both below so you can visually compare them.

Notice the unequal variance formula uses the sample variance representing each population whereas the equal variance formula uses the pooled variance. The other difference is the manner in which degrees of freedom are determined. With the unequal variance approach, you calculate the degrees of freedom using the very tedious formula shown just above. With the equal variance approach, you determine degrees of freedom by $(n_1 + n_2) - 2$.

$$\begin{aligned} \bar{X}_1 - \bar{X}_2 \pm t'_\alpha \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} & \text{ unequal variances} \\ \bar{X}_1 - \bar{X}_2 \pm t_\alpha \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}} & \text{ equal variances} \end{aligned}$$

Also notice the unequal variances approach inserts a value called t -prime whereas the equal variances approach inserts a value called t . Fortunately, the use of t or t' is looked up the same way in the same t -table. I am personally not sure why the distinction needs to be made because they look up in the same t -table in the same manner. However, they are designated differently so just live with it and move on.

One word of caution is appropriate here. Let's say that you calculate the degrees of freedom as 24.65 (unequal variance approach). The rule is you round down to 24 degrees of freedom, never up to 25 degrees of freedom.

Let's look at an example. Babies To Go, Inc. is the manufacturer of car seats for children. They manufacture two distinct types of car seats. They want to test the car seats for durability. They

take a sample of 15 car seats of Type 1 design and find the average life to be 11.4 months with a standard deviation of 1.2 months. A durability test of 10 car seats of Type 2 design yields an average life of 7.5 months with a standard deviation of 0.8 months. Here, for illustration purposes, I will assume there is no evidence to suggest that the variances in the wear of Type 1 versus Type 2 are equal. For illustration purposes, this assumption overrides any results obtained using the F -test. The CEO wants to test the durability of each design using an alpha of 0.05.

Let's recap the data associated with the two samples:

Table 9.4 Babies To Go, Inc. data.

	Type 1	Type 2
Mean Life	11.4 months	7.5 months
Sample Size	15	10
Standard Deviation	1.2 months	0.8 months
Variance	1.44	0.64

Again the formula for determining degrees of freedom is as follows:

$$\text{d.f.} = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Let's first calculate the degrees of freedom. You are going to substitute the values you know into the formula and *CAREFULLY* make the calculation as follows:

Numerator Calculation

$$\text{Numerator} = \left(\frac{1.44}{15} + \frac{0.64}{10}\right)^2$$

This is the numerator of the equation above. I will work the numerator then the denominator to keep the process from being too confusing.

Solving the equation:

$$\text{Numerator} = (0.096 + 0.064)^2$$

$$\text{Numerator} = (0.1600)^2$$

$$\text{Numerator} = 0.0256$$

Denominator Calculation

$$\text{Denominator} = \frac{(1.44 \div 15)^2}{15 - 1} + \frac{(0.64 \div 10)^2}{10 - 1}$$

This is the denominator calculation for the formula just above.
Solving the equation:

$$\text{Denominator} = \frac{(0.0960)^2}{15 - 1} + \frac{(0.0640)^2}{10 - 1}$$

$$\text{Denominator} = \frac{(0.0092)}{15 - 1} + \frac{(0.0041)}{10 - 1}$$

$$\text{Denominator} = 0.0007 + 0.0005$$

$$\text{Denominator} = 0.0012$$

Putting Them Together

$$\text{Degrees of freedom} = \frac{0.0256}{0.0012} = 21.33$$

d.f. = 21 rounded down per the rule.

Okay, that tedious calculation is done, so what is next, you ask?

Well, since you have done such a great job with this one, let's push on.

Next, I need to look up the value of t' at alpha 0.05 and the degrees of freedom of 21. The corresponding t' -value is 2.080. Look it up yourself and make sure I am correct. This value is found in the t -table not the Z -table.

Let's look at the confidence interval formula once again.

$$\bar{X}_1 - \bar{X}_2 \pm t'_\alpha \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \quad \text{unequal variances}$$

Okay, now let's insert the numbers. I have been given in Table 9.4 or have developed by looking up the t -value (2.080).

$$11.4 - 7.5 \pm 2.080 \sqrt{\frac{1.44}{15} + \frac{0.64}{10}}$$

$$3.9 \pm 2.080 \sqrt{0.096 + 0.064}$$

$$3.9 \pm 2.080\sqrt{0.160}$$

$$3.9 \pm 2.080(0.400)$$

$$3.9 \pm 0.8320$$

3.068 to 4.732 months (The 95% confidence interval)

Interpretation

Both values are positive, so the interval does NOT contain zero. I can conclude the two types are not statistical equivalents. *If I go back and look at the sampling data (Table 9.4), I can conclude that I am 95% confident that Type 1 lasts from 3.068 to 4.732 months longer than Type 2.* If the costs of manufacturing the two types are equal to each other, the CEO should prefer Type 1.

Let's change the problem slightly. Let's assume the CEO found that Type 1 was a lot more expensive to manufacture. The increased cost will lead to an increased sales price. Can the higher sales price be justified based on durability? Let us further assume that the CEO has determined if Type 1 lasts at least five months longer than Type 2, the cost of manufacture can be justified, and the Type 1 unit can be sold for enough to more than make up the additional costs. What decision do you make?

The numbers do not change. The calculations I have made remain the same. From those calculations, I have determined there is a statistically-significant difference in the two types. Type 1 has a longer life than Type 2, but is it at least five months longer?

The answer is no. The calculated interval of 3.068 months to 4.732 months does not contain the five months. Therefore, I cannot conclude that the cost of manufacturing Type 1 can be justified at an increased sales price. Under these circumstances, I would opt for Type 2.

Wow, Professor. I can see this testing stuff as being very useful in making informed decisions rather than flying by the seat of your pants. I am beginning to understand some of this stuff.

I look at you in amazement, but say nothing, assuming your comment is rhetorical.

Let's move on with another helpful technique.

Dependent Paired Sampling—CI

As a recap, I have listed the main types of two-population tests.

Independent Testing:

Large size samples

Small samples

Equal variances

Unequal variances

Dependent Testing:

Matched pairs

Large size samples, small samples with equal variances, and small samples of unequal variances have been examined. These three are associated with independent testing. Matched pairs (dependent testing) present a different challenge.

Suppose I want to evaluate *one relevant aspect* among two similar data sets. My method of testing choice would be paired sampling.

Paired Sampling offers several advantages.

- First, smaller samples may be used.
- Second, smaller variances are encountered.
- Third, fewer degrees of freedom are lost.
- Fourth, a smaller sampling error will occur.

Several examples will serve to help you understand the concept of paired sampling or matched pairs. The groups or pairs will have some common relationship. There will be an attempt to match the pairs as closely as possible.

Example 1

In medical research it is not uncommon to have control groups when a new medicine is tested. The groups will be matched as closely as possible. Let's say a new blood pressure medicine needs to be tested. Two groups of subjects are chosen. The groups are matched as closely as possible in such characteristics as gender, weight, age, activity level, cholesterol, and any other factor affecting blood pressure. The groups are essentially twins or mirrors of each other except for one relevant factor—the medicine given. Group 1 is given Medicine A and Group 2 is given a placebo (non-medication). The difference in the blood pressure after the study time period has lapsed will be due solely to the medicine or the lack of medication. I want to know if there is a difference in blood pressure between groups.

Example 2

Let's assume I have a product that I want to market. I have been presented with a choice of two different marketing programs by my advertising agency. I would select a sample from two matched test markets, which would be matched as to population size, socio-economic characteristics, demographic characteristics, etc. The purpose of the test is to determine which marketing program will yield the best sales results.

Example 3

End cap displays are generally sought by most vendors, since they tend to increase sales significantly. Often a vendor will pay a premium to the drug store or grocery store for a preferred display position. However, you want to know if you can justify the premium. Weekly sales for a soft drink product at a local supermarket are recorded before and after an end cap display is installed. Did the end cap display make enough of a difference in sales after the display is installed to justify paying a premium for the location?

The purpose of matched pairs is to control as many factors in the testing process as possible and then introduce an uncommon variable to determine the effect of the uncommon variable on the two groups.

Let's work through an example.

I want to test the effect of gender on starting salaries of graduates from a certain college in Texas. I select two groups—one male and one female. I gather the information on a large number of subjects by compiling a pool of males and a pool of females. I record the GPA and the major for each member of the pool. I will select all marketing majors or all economics majors. I would never mix majors such as marketing and education. I next select at random one male from the larger pool of males. I then select one female from the larger pool of females. I record this as matched pair number one (Table 9.5). I then continue this selection process nine additional times. The result is 10 matched pairs of males and females.

The confidence interval formulas I will use for matched pairs are as follows:

$$\bar{d} \pm t_{\alpha} \frac{S_d}{\sqrt{n}}$$

where \bar{d} is the average difference in the matched pairs.

S_d is the standard deviation of the difference.

t_{α} is the critical t -value looked up in the t -table or found by Excel.

n is the sample size.

The two calculations are from the following formulas:

$$\bar{d} = \frac{\sum d_i}{n}$$

$$S_d = \sqrt{\frac{\sum d_i^2 - n(\bar{d})^2}{n - 1}}$$

The results of the study are recapped in Table 9.5 below.

Table 9.5 Matched pairs study of starting salaries.

Pair	Male Salary	Female Salary	d_i	d_i^2
1	\$39,500	\$39,000	500	250,000
2	\$41,800	\$41,900	-100	10,000
3	\$40,200	\$39,600	600	360,000
4	\$38,500	\$38,500	0	0
5	\$43,400	\$42,500	900	810,000
6	\$37,800	\$38,000	-200	40,000
7	\$39,500	\$39,200	300	90,000
8	\$41,100	\$40,000	1,100	1,210,000
9	\$38,300	\$38,100	200	40,000
10	\$39,200	\$38,500	700	490,000
Totals	\$399,300	\$395,300	\$4,000	\$3,300,000
Means	\$39,930	\$39,530	\$400	\$330,000

If you simply look at the data, the mean starting salary for males is \$39,930 and for females is \$39,530. This is a difference of \$400.

I want to know if there is a statistically-significant difference in the starting salaries of males and females with a certain GPA and a certain major or is the difference in \$39,930 and \$39,530 due to random sampling error alone? The test will be conducted at the alpha level of 0.05. This is a *t*-test, since the sample size is less than 30.

From observation, it appears as if there is no difference (\$400 is not much). However, one of the things I hope you learn is to never trust your first impression without applying some sort of statistical procedure to test it.

I begin my calculations by determining the mean of the difference. That value is the sum of the column headed d_i divided by the number of matched pairs, which is 10 in this instance.

$$\bar{d} = \frac{\sum d_i}{n}$$

$$\bar{d} = \frac{4,000}{10} = 400 \text{ (This is also the difference in the \$39,930 and the \$39,530 starting salaries).}$$

Next, I calculate the standard deviation of the difference.

$$S_d = \sqrt{\frac{\sum d_i^2 - n(\bar{d})^2}{n - 1}}$$

$$S_d = \sqrt{\frac{3,300,000 - (10)(400)^2}{10 - 1}}$$

$$S_d = \sqrt{\frac{3,300,000 - (10)(160,000)}{9}}$$

$$S_d = \sqrt{\frac{3,300,000 - (1,600,000)}{9}}$$

$$S_d = \sqrt{\frac{1,700,000}{9}}$$

$$S_d = \sqrt{188,889}$$

$$S_d = 434.61$$

The 95% confidence interval for the difference in the matched pairs is shown below. The *t*-value (2.262 with d.f. of 9) is determined from table look up (Chapter 7 table or back of the textbook). Check it out by looking it up yourself.

$$\bar{d} \pm t_\alpha \frac{S_d}{\sqrt{n}}$$

$$400 \pm (2.262)\left(\frac{434.61}{\sqrt{10}}\right)$$

$$400 \pm (2.262)\left(\frac{434.61}{3.1623}\right)$$

$$400 \pm (2.262)(137.435)$$

$$400 \pm 310.878$$

$$\$89.12 \text{ to } \$710.88$$

Interpretation

First I observe, there is no zero in the interval. This means there is a difference in the starting salaries that is significant. The difference in the starting salaries for men of \$39,930 and women of \$39,530 cannot be explained by sampling error alone. I am 95% confident the mean starting salaries for males are higher than the mean starting salaries for women (\$39,930 is higher than \$39,530). Were the interval to contain zero, I would have to conclude there is no statistically significant difference in the starting salaries—men are paid the same as women. Since this is not the case, do you have any idea *how much more* the men are paid than women?

Don't read the next sentence until you think about it.

On average, the starting salary for men is from \$89.12 to \$710.88 per year higher than the starting salary for women.

Test

Wow, you quickly say. I am not ready for a test.

Not so, I say.

Anyway, this is not really a test but a check of your understanding of when to apply an independent test or when to apply a matched pairs test. Let's look at a couple of examples.

Let's assume the following:

I am a manufacturer of tires. I would like to produce tires that last longer, so I conduct a test of the mean life of a certain tire. I have concluded that my tires should not be tested on the race-track at Daytona by professional drivers. I want my test to include only the everyday, ordinary drivers driving on neighborhood streets. After all, the ordinary driver is my target market, so I better use him or her in my test.

The new tire is a steel-belted radial. I have two designs in mind, so I produce enough tires with both designs to accomplish the testing. The experiment is set up as follows. Twenty new cars were selected at random and two tires of the *same design* were installed on the rear wheels of those 20 cars. Twenty more new cars were selected at random and two tires of the *alternate design* were installed on the rear wheels of those 20 cars. The test as designed thus far includes 40 new cars split into two groups with 20 cars having one design on the rear wheels and the other 20 cars having the alternate design on the rear wheels. The drivers were told to drive their car as usual. Get the picture?

Okay, let's now say I conduct the test. How should I measure the results—independent sampling or paired sampling?

Think about it before reading on.

This experiment as set forth above would be the comparison of two independent populations. The experimental design as set forth above does not consider nor control the wear due to each driver. Each driver's driving style and driving habits provide an uncontrolled wild card in this experiment.

How could I have changed the experiment to better control this wild card?

Think about it before reading on.

I could have selected 20 new cars randomly and placed one design of each type on the same car. In this instance, the rear wheels would contain design A on one side and design B on the other side. Here the driving habits of the drivers are minimized. The "mean life to wear-out" is better established when the wild card is better controlled. This latter experimental design would use the matched pairs approach.

Do you see the difference? In the former, I leave the difference in driving skills and habits of each of the drivers as a wild card in the experiment. This allows me to designate each of the populations as independent. In the latter design, I tend to minimize the driving skills and habits of each of the drivers by matching the pairs of tires to the drivers.

Now that wasn't so bad was it? Design of the experiment is very critical as this example illustrates.

Okay, that wraps up confidence intervals, so let's move to hypothesis testing for two populations.

Hypothesis Testing versus Confidence Intervals

With hypothesis testing I have the same types of tests available to me.

Independent Testing:

Large size samples

Small samples

Equal variances

Unequal variances

Dependent Testing:

Matched pairs

Go to Table 9.1. I will wait.

Okay, got it? I have been addressing confidence intervals. The formulas I have been using are shown in the left-hand column. There is a mirrored image formula for each left-hand test in the right-hand column. The right-hand column is the hypothesis test approach. The left-hand column is the confidence interval approach. The only difference between the formulas in the left column and the right column is one of solving for a different term. The formulas are essentially the same.

Both hypothesis testing and confidence intervals will reveal if there is a statistically-significant difference between the means of the two populations. Confidence intervals will tell

you the amount of the difference. Hypothesis testing can only reveal whether the difference is or is not statistically significant. If you have a problem that requires a conclusion about the amount of the difference, then you must use confidence intervals. If you are not interested in the amount of the difference and just want to know if the two population means are different, then use either.

Let's take a couple of examples of hypothesis testing.

Hypothesis Testing

As you can see from Table 9.1 created early in this chapter, the formulas for confidence interval and hypothesis testing are closely related. Remember, hypothesis testing can be done using a Form I test (left-tail rejection region) or Form II test (right-tail rejection region) or Form III test (two-tailed rejection region). If you need to review these three forms, you might return to Chapter 8.

In any hypothesis test, you compare a critical value of either Z or t to a calculated value of the corresponding Z or t . Let's work through an example of a hypothesis test using the following example.

A Texas study was commissioned to determine if there is a statistical difference between employees' contributions to two different types of retirement plans. Two sets of data are evaluated. One group of employees contributes to a tax-deferred annuity (TDA) while a second group of employees contributes to a 401(K) plan.

The TDA plan places deposits with an investment bank which invests the money in stock or bond portfolios. A second alternative is a standard 401(K) plan where the company matches all or a portion of the contribution by the employee. If the portfolios perform well, the retirement funds grow. Two groups are involved in making contributions.

The objective of the study is to determine if the two groups contribute the same or different amounts. It is felt that employees will contribute more to what is perceived as a favorable plan. The mean contribution for the TDA is \$2,155 and the mean to the 401(K) is \$2,040. Is there a statistical difference between the two means or does sampling error explain the difference? If there is a difference, it might be argued that employees see the TDA as a better plan than the 401(K).

A random sample of 12 people from the TDA plan is selected and a sample of 15 people from the 401(K) plan is selected. The results are shown in Table 9.6 below:

Table 9.6 Comparison of two plans.

Results	TDA Eligible	401(K) Eligible
Sample Size	12	15
Sample Mean	\$2,155	\$2,040
Sample Standard Deviation	\$612	\$695

This is an independent sample, since neither group influences the contribution of the other. Actually, the null hypothesis can take on any of the three forms of hypothesis testing.

Remember, Form I is a left-tail rejection region. Form II is a right-tail rejection region. Form III is a two-tail rejection region. Initially let's opt for a two-tail test—Form III.

$$H_0: \mu_{TDA} = \mu_{401(k)}$$

$$H_a: \mu_{TDA} \neq \mu_{401(k)}$$

This is a t -test since at least one of the samples is less than 30 (actually both are less than 30). Should you use the assumption of equal or unequal variances? In this instance, assume the variances are equal. The F -test supports this assumption (1.2896 if you care to calculate it) and there is no overriding statement to assume otherwise. The t -test formula can be found in Table 9.1 but is shown below.

$$t_{\text{TEST}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

where $\bar{X}_1 - \bar{X}_2$ is the difference between the two means.

S_p^2 is the pooled variance.

$\mu_1 - \mu_2 = 0$ and is always zero.

n is the sample size.

Okay, Professor, why do you include $\mu_1 - \mu_2 = 0$, if it is always zero?

The only reason for including this in the formula is a conceptual one. The idea is associated with zero being important. If the means of the populations are equal, the difference in the means will equal zero (statistically speaking). Zero in the interval tells me everything is okay and the difference in the means is explained by sampling error.

The first step is to determine the pooled variance using the information given in Table 9.6. This formula and solution is as follows:

$$S_p^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

$$S_p^2 = \frac{(612)^2(12 - 1) + (695)^2(15 - 1)}{12 + 15 - 2}$$

$$S_p^2 = \frac{(374,544)(11) + (483,025)(14)}{25}$$

$$S_p^2 = \frac{(4,119,984) + (6,762,350)}{25}$$

$$S_p^2 = \frac{10,882,334}{25}$$

$$S_p^2 = 435,293.4$$

Using the information in Table 9.6 and the pooled variance calculated just above, the calculated t -test gives us the following result:

$$t_{\text{test}} = \frac{(2,155 - 2,040) - 0}{\sqrt{\frac{435,293.4}{12} + \frac{435,293.4}{15}}}$$

$$t_{\text{test}} = \frac{115}{\sqrt{36,274.5 + 29,019.6}}$$

$$t_{\text{test}} = \frac{115}{\sqrt{65,294.1}}$$

$$t_{\text{test}} = \frac{115}{255.53}$$

$$t_{\text{test}} = 0.4500 \text{ This is the calculated } t\text{-value.}$$

At the alpha level of 0.05, the *critical* t -value from Chapter 7 or the back of the textbook look up at 25 degrees of freedom ($n_1 + n_2 - 2$, which is $12 + 15 - 2 = 25$) is 2.05954 (rounded to 2.060). Excel will give you the same result. Look it up or use Excel to make sure I am correct before you move on with the rest of the problem. This is a two-tailed rejection region test.

Now you are ready to compare the calculated t -test value (0.4500) with the table look up critical value (2.060). The DNR (do not reject) range is from -2.060 to $+2.060$. The critical value is 0.4500. This value is less than $+2.060$ but greater than -2.060 ; therefore, it falls in the DNR region. I can now assert that there is no statistically-significant difference at the 0.05 alpha level (95% level of significance) between the contributions of the employees to the TDA plan or the employees to the 401(K) plan. The contributions are statistically the same. The differences are due to sampling error and sampling error alone.

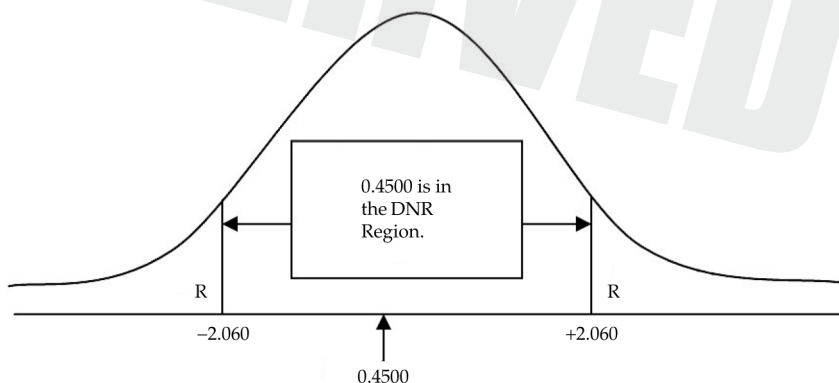


Figure 9.3 Two tail rejection region (TDA & 401k).

One-Tail Rejection Region

In hypothesis testing for two populations, I do have some latitude in using Form I, Form II, or Form III. I could have set this problem up as a right-tail rejection region test or as a left-tail rejection region test. How I state the null and alternate hypotheses would determine whether I use Form I, II, or III. For example, if I were to test that the TDA contributions are more than the 401(K) contributions, the hypothesis would be as follows.

$$H_o: \mu_{TDA} \leq \mu_{401(k)}$$

$$H_a: \mu_{TDA} > \mu_{401(k)}$$

In this case, the critical t -value becomes (one-tailed, right test rejection region) +1.708141 (Excel) (rounded to 1.708). The calculated t -value remains the same at 0.4500. You would reach the same conclusion. Because 0.4500 is less than 1.708, you would DNR the null hypothesis and conclude that the contributions to the TDA plan are less than or equal to the mean of the 401(K) plan.

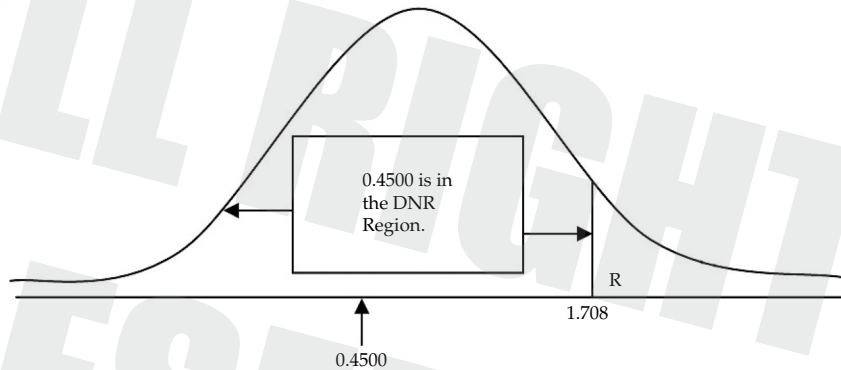


Figure 9.4 One tail rejection region (TDA & 401k).

What do you think would happen if I applied the confidence interval approach rather than the hypothesis testing approach? While Confidence Intervals are generally two-tailed rejection regions, here I am only interested in the right hand side rejection region.

I would hope I would reach the same conclusion, you say.

That would be a correct statement, I respond.

Let's go through that process and see what would happen. I am assuming the variances are equal, so the variances must be pooled. I will use the pooled variance, confidence interval formula for a small sized sample as follows:

$$\begin{aligned} \bar{X}_1 - \bar{X}_2 \pm t_\alpha \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}} \\ 2,155 - 2,040 \pm (1.708) \sqrt{\frac{435,293.4}{12} + \frac{435,293.4}{15}} \\ 115 \pm (1.708)(255.53) \end{aligned}$$

$$115 \pm 436.4452$$

$$-321.4452 \text{ to } 551.4452$$

$$-321.45 \text{ to } 551.45 \text{ (Rounded)}$$

The interval contains zero. If zero is in the interval, you can assert that there is no statistically significant difference between the contributions of the two groups of employees. The difference can be explained by sampling error. I can conclude TDA contributions are from minus \$321.45 less than the 401(K) contributions to plus \$551.45 more than the 401(K) contributions, but the overall difference is not significant.

Strange how that works, isn't it? You get the same conclusion by using the hypothesis testing or the confidence interval process.

With confidence intervals I can determine how much the difference is, but with either confidence intervals or hypothesis testing, I can determine if the difference is significant.



Figure 9.5 Confidence Interval (TDA & 401k).

Proportions

As you might suspect, you not only can determine if the means of two populations are statistically different, but you can test proportions in the same manner.

For example, let's suppose you wanted to determine the percentage (proportion) associated with two different manufacturing techniques. In this instance, I want to know if there is a difference in the defects produced by Process A versus the defects produced by Process B. I want to know which one yields the better result. This is a proportion problem. The only two outcomes are defective and non-defective.

The procedures associated with proportion and mean testing are essentially the same. The formulas are, however, different, as you might imagine.

The proportion formula for a confidence interval is as follows:

$$p_1 - p_2 \pm Z_{\alpha} S_{p_1 - p_2}$$

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

Compare the proportion formula to the mean formula.

$$\bar{X}_1 - \bar{X}_2 \pm Z_\alpha S_{\bar{X}_1 - \bar{X}_2}$$

Notice any similarity?

Yep, you say. They look like the same formula, but one is for proportions and one is for the means.

Actually, proportions work the same way as mean testing.

Let's work through an example. Assume I have a manufacturing plant and want to determine if shift 1 has more tardiness than shift 2. I take a sample of 100 workers from shift 1 and 125 from shift 2. The results show that 45 workers are tardy on shift 1 and 49 workers are tardy on shift 2. I want to determine a 95% confidence interval for the population proportion.

The proportions:

$$45 \div 100 = 0.45 \text{ for shift 1}$$

$$49 \div 125 = 0.39 \text{ for shift 2}$$

First, calculate the standard deviation of the difference between the proportions using the information given above.

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

$$S_{p_1 - p_2} = \sqrt{\frac{0.45(0.55)}{100} + \frac{0.39(0.61)}{125}}$$

$$S_{p_1 - p_2} = \sqrt{\frac{0.2475}{100} + \frac{0.2379}{125}}$$

$$S_{p_1 - p_2} = \sqrt{0.0025 + 0.0019}$$

$$S_{p_1 - p_2} = \sqrt{0.0044}$$

$$S_{p_1 - p_2} = 0.0663$$

Next develop the confidence interval.

$$p_1 - p_2 \pm Z_\alpha S_{p_1 - p_2}$$

$$0.45 - 0.39 \pm (1.96)(0.0663)$$

$$0.06 \pm 0.1299$$

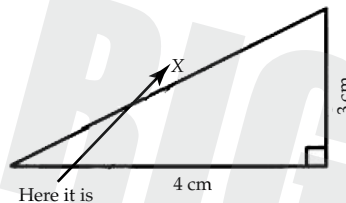
$$-0.0699 \text{ to } 0.1899$$

Interpretation

Zero is in the interval, so I must conclude that there is no statistical significance between the two population proportions. The difference can be explained by sampling error. I am 95% confident the true, unknown population proportion is between -0.0699 and 0.1899 . There is no difference between the tardiness of shift 1 and shift 2 employees. This does not mean the level of tardiness is acceptable for either shift, just the difference is explained by sampling error.

Use of Excel

I have to share with you an answer one student gave on the exit examination for high school students. I am told the student was asked to *find "X" given certain information*. The answer was given by a blonde (not a slam at blondes since I know a number of really intelligent ones), female student as follows:



I personally love the answer. I understand she got credit for the problem on the examination. Do you suppose the testing bureau should be more specific?

Life with Excel is not quite that simple, but as you might expect, you can use Excel to make many of the computations for two-population testing.

For example, open an Excel worksheet > Tools > Data Analysis. Scroll down and find the following four possible Excel solution helps.

- *t*-test: Paired Two Sample for Means
- *t*-test: Two Sample Assuming Equal Variances
- *t*-test: Two Sample Assuming Unequal Variances
- *Z*-test: Two Sample for Mean

t-test: Paired Two Sample Means

The paired two sample *t*-test for means is the same as the dependent sample test shown to you above. As a reminder, I selected 10 males and 10 females who were matched as to GPA and major. I wanted to determine if the starting salary of males was equal to or greater than the starting salary of females.

To use Excel, I will enter all of my data from Table 9.5 into an Excel worksheet. Next, I select the *t*-test, paired two sample for means. I am asked to enter the range for variable one (male salary) and the range for variable two (female salary). I place my cursor in the variable one box

and then, using the mouse, highlight the entire data column for male salaries. Since I have a heading “Male Salary” for my column heading I include it in my variable one data range.

I now move the cursor to the second box and repeat the process for the column headed “Female Salary.” I next input the difference in the hypothesized mean, which I want to be zero. In other words, I am asserting the difference in the means is zero. I check the box marked “Labels in the First Row.” I choose the alpha level of 0.05. I will let the output be displayed in a new tab by selecting the second output choice. Then I click okay. The result is displayed in the following table (Table 9.7):

Table 9.7 *t*-test: Paired two sample test for means from Excel.

t-Test: Paired Two Sample Test for Means		
	<i>Male Salary</i>	<i>Female Salary</i>
Mean	39,930	39,530
Variance	3013444.444	2395666.667
Observations	10	10
Pearson Correlation	0.971436027	
Hypothesized Mean Difference	0	
Df	9	
t Stat	2.9104275	
<i>P</i> (<i>T</i> ≤ <i>t</i>) one-tail	0.008649604	
<i>t</i> Critical one-tail	1.833112923	
<i>P</i>(<i>T</i> ≤ <i>t</i>) two-tail	0.017299209	
<i>t</i> Critical two-tail	2.262157158	

If you go back and review the manual calculations associated with this problem, you will notice the *t*-value of 2.262 is the same. This approach determines a *p*-value, which for a two-tailed test is 0.0173 (rounded). Were this a Form III (two-tailed rejection region) test, the hypothesis that would be set up as follows:

$$H_0: \mu_1 = \mu_2 \text{ or } \mu_1 - \mu_2 = 0 \text{ (These have the same meaning.)}$$

$$H_a: \mu_1 \neq \mu_2 \text{ or } \mu_1 - \mu_2 \neq 0$$

Since $0.0173 < 0.05$ (*p*-value < alpha value), the null hypothesis is rejected. The alternate hypothesis will be accepted, thus telling you the starting salaries of the males and females are unequal. The males have a higher starting salary. Using the Excel approach will tell you if there

is a difference, but you cannot determine the amount of the difference like you can with confidence intervals. You would have to manually determine the confidence intervals as previously shown. Shown in the Table 9.7 is something referred to as a *t*-Stat (2.91). This is what I have referred to as the calculated *t*-value. Since this is a two-tailed confidence interval, the lower and upper *critical limits* are -2.262 and $+2.262$. Since $+2.91$ (*calculated value*) is greater than the upper *critical limit* of $+2.262$, you would reject the null hypothesis, just as I did using the *p*-value.

Just as a check, you can determine the *t*-test value by using the manual *t*-test formula.

$$t_{\text{test}} = \frac{\bar{X}_1 - \bar{X}_2}{\frac{S_d}{\sqrt{n}}} = \frac{35,930 - 35,530}{\frac{434.61}{\sqrt{10}}} = \frac{400}{137.535} = 2.9084 \text{ (Round to 2.91)}$$

You will have to refer back to the problem itself and the manual calculations to see where I got some of the numbers I am using. Using Excel is much faster than doing manual calculations, but as you can see the results are the same.

To use Excel, you need the entire, original data set. Without the original data set, you do not have anything to input. The other three processes using Excel ask about the same questions and work essentially the same way as demonstrated above. Additionally, if you have problems involving proportions, Excel works with their values as well as long as you have the entire data set from which the proportions are taken. Excel is helpful and much faster, so consider learning it and using it.

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

- You see people wear bib overalls at funerals.
- You refer to the capital of Texas as the “home of the Longhorns”.
- You can actually burn your hand opening the car door.

In-Class Problems Chapter 9

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.

Work the following problems in class if your professor allows or tear them out of the book and work them and turn them in to the professor.

Confidence Intervals

1. A sample of 35 textbooks was taken from the TCU bookstore. The average price of the books was \$52.50 with a *standard deviation* (S) of \$5.10. Forty textbooks were sampled from the SMU bookstore. The average price of those 40 books was \$65.00 with a *variance* (S^2) of \$18.06. Calculate a 90% confidence interval for the difference in the average prices for the two stores. Explain your findings.
2. Two car models are tested for a difference in average gas mileage. Ten cars of Model A averaged 32.1 miles per gallon with a standard deviation of 5.2 mpg. Twelve cars of Model B averaged 35.5 mpg with a standard deviation of 4.4 mpg. Calculate a 95% confidence interval for the difference between the two cars' mileage. There is no evidence to suggest the variances of the populations are equal. Explain your findings.
3. Bendix Fertilizer Company has two plants—one in LA and the other in Atlanta. Recent customer complaints suggest that the Atlanta shipments are underweight as compared to the LA shipments. Ten (10) boxes from LA average 96.3 pounds with a standard deviation (S) of 12.5 pounds and fifteen (15) boxes are selected from Atlanta with an average weight of 101.1 pounds with a standard deviation (S) of 10.3 pounds. Does a 99% confidence interval support those complaining? Assume equal population variances. Explain your findings.
4. The monthly starting salaries in thousands of dollars of 12 maintenance personnel at TCU are compared to those from SMU using the data in the table shown below. Develop and interpret a 95% confidence interval for the difference in mean starting salaries at TCU and SMU.

TCU	SMU
15.6	13.7
16.8	13.6
18.5	15.2
16.5	11.2

(Continued)

TCU	SMU
15.5	11.6
14.8	15.2
18.8	12.5
19.5	13.5
17.5	13.9
16.5	18.2
14.5	14.5
18.7	11.2

Hypothesis Testing

5. Each patient in Farmer's Hospital is asked to evaluate the service at the time of discharge. Recently there have been complaints that the resident physicians and nurses on the surgical wing responded too slowly to calls from the senior citizens. Other patients seem to be receiving faster service. Mr. Bob Anderson, Hospital Administrator, asked for a study by the quality assurance team. After studying the problem, the QA team, headed by Sharon Schwartz, decided to collect the following sample information and to test the claim at the alpha 0.01 level of significance. The question to be answered was: Are the other patients receiving faster service than the senior citizens? As a side calculation, Mrs. Schwartz decided to determine the p -value even though she knew she would have to explain its meaning to Mr. Anderson.

Conduct the study by choosing the appropriate test statistic and explain your results.

The data sets are as follows:

Patient Type	Sample Mean	Sample Standard Deviation	Sample Size
Senior Citizen	5.50 minutes	0.40 minutes	50
Other	5.30 minutes	0.30 minutes	100

6. Owens Lawn Care, Inc. manufactures and assembles lawnmowers, which are shipped to dealers throughout the United States and Canada. Two different procedures have been proposed for mounting the engine on the frame of the lawnmower. The question is: Is there a difference in the mean time to mount the engines on the frames of the lawnmowers? The first procedure was developed by Welles (#1) and the second was developed by Atkins (#2). To evaluate the two methods, it was decided to conduct a time

and motion study. A sample of five employees was timed using procedure #1 and six different employees were timed using procedure #2. The results in minutes are shown below. Is there a difference in the mean mounting times? There is reason to believe the population variances are equal. Use the 0.10 level of significance.

Procedure #1 in Minutes	Procedure #2 in Minutes
2	3
4	7
9	5
3	8
2	4
	3
Average = 4.0 minutes	Average = 5.0 minutes

ALL RIGHTS RESERVED

KH

**ALL RIGHTS
RESERVED**

Homework Problems Chapter 9

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:

Large-Size Samples—Confidence Intervals

- 1. Testing the Mean:** American Hallmark Insurance sells policies to residents throughout the DFW Metroplex area. Claims are certain to be filed. The company wants to estimate the difference in the mean claim costs (in dollars) between people living in urban areas and those residing in the suburbs. Of the 180 urban policies selected as a sample, the mean insurance claim of \$2,025 was reported with $s = \$918$. The sample of 200 suburban policies revealed a mean insurance claim of \$1,802 and $s = \$512$. What does a 95 percent confidence interval tell the company about the claims filed by these two groups? Interpret your results.

- 2. Testing the Mean:** Two production designs are used to manufacture a certain product. The mean time required to produce the product using design A was 3.51 days with $s = 0.79$ days. Design B required a mean time of 3.32 days with $s = 0.73$ days. Equal size samples of 150 were used for both designs. What would a 99 percent confidence interval reveal about which design should be used? Interpret your results.

3. **Testing the Mean:** An accountant for a large corporation in the Midwest must decide whether to select MCI or Sprint to handle the firm's long-distance telephone service. Data collected for calls using both services are reported here. See Data Table 9.1.

Data Table 9.1 MCI Compared to Sprint.

	MCI	Sprint
Number of calls	145	102
Mean cost	\$4.07	\$3.89
Standard deviation	\$0.97	\$0.85

What does a 95 percent confidence interval reveal about the difference in the population means? Interpret the results.

Large-Size Sample—Hypothesis Testing

4. **Testing the Mean:** Samples of size 50 and 60 reveal means of 512 and 587 and standard deviations of 125 and 145, respectively. At the 2 percent alpha level, test the hypothesis that $\mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$.

Small-Size Sample—Equal Variances (Confidence Intervals)

5. **Testing the Mean:** Seventeen cans of *Energy Aid* report a mean fill level of 17.2 ounces with a standard deviation of 3.2 ounces, and 13 cans of *Charged* produce a mean fill level of 18.1 ounces and $s = 2.7$ ounces. Assuming equal variances and normal distributions in population weights, what conclusion can you draw regarding the difference in mean weights based on a 98 percent confidence interval? Interpret your results.

- 6. Testing the Mean:** Grant requests must be submitted to either the National Science Foundation (NSF) or Health and Human Services (HHS), but the TCU Foundation for Health Studies prefers to have an answer as soon as possible. A study of both organizations has yielded the following results. Fourteen grant requests to NSF took an average of 45.7 weeks with a standard deviation of 12.6 weeks. Twelve grant requests to HHS took an average of 32.9 weeks with a standard deviation of 16.8 weeks. If the NSF takes more than five weeks longer than HHS, the TCU foundation will submit the grant requests to HHS. At the 90% confidence level, test the difference in the process time for the two organizations. Assume the population variances are equal. Should the foundation bother submitting to HHS?

Small-Size Sample—Equal Variances (Hypothesis Testing)

- 7. Testing the Mean:** Using the information in problem #6 mentioned above, test the hypothesis that the two foundations process the grants with the same efficiency and speed. Use the alpha level of 0.10. Assume equal population variances. How does the confidence interval approach used in #6 differ from this approach or does it?

Small-Size Sample—Unequal Variances (Confidence Interval)

8. **Testing the Mean:** Twenty-six mutual funds, each with \$5,000 invested in them, are selected for comparison. Of the 26 funds, 12 are income oriented and yield a mean return of \$1,098.60 with a standard deviation of \$43.20. The remaining 14 funds are growth oriented and yield a mean return of \$987.60 with a standard deviation of \$53.40.
- Calculate and interpret the **80 percent confidence interval** for the difference between the population mean returns. There is no reason to believe that the population variances are equal.
 - Sample Size for the Mean:** What sample size is necessary to be 95 percent certain that the error does not exceed \$10.00?

Small-Size Sample—Unequal Variances (Hypothesis Testing)

9. **Testing the Mean:** At the 1 percent alpha level, test for the equality of means if samples of 10 and 8 yield means of 36 and 49 and standard deviations of 12 and 18, respectively. Assume unequal population variances.

Matched Pairs—Confidence Interval

10. **Testing the Mean:** Rankin Associates will accept bids from two construction companies on a remodeling job at the home office. The decision about which offers to accept depends in part on the mean completion times of similar jobs by each company. Data are collected and paired from several previous remodeling jobs done by each company. Based on a 99 percent level of confidence, which company would receive the contract from Rankin Associates? See Data Table 9.2.

Data Table 9.2 Contract Selection Process for Ranking Associates.

Pair Number	Company 1	Company 2
1	10.0	9.2
2	12.2	10.0
3	15.3	9.2
4	9.6	10.5
5	8.6	9.5
6	9.4	8.4
7	12.5	7.2
8	7.3	8.4
9	9.4	10.5
10	8.7	6.2
11	9.1	8.1

11. Costs of services at two local car dealerships tend to confuse many consumers. In an attempt to understand which might be the most costly, the statistics department at TCU conducts a study. The services are matched as closely as possible and the results are shown in Data Table 9.3.

Data Table 9.3 Service Department Cost Study.

Service	Dealer 1-\$	Dealer 2-\$
1	54	36
2	56	35
3	59	34
4	65	39
5	62	37
6	43	32
7	38	31
8	48	30
9	46	29
10	59	45

Calculate and interpret the 90% confidence interval for the difference between the population means. Assuming the quality of service is the same, which dealer should be used or does it make a difference?

Matched Pairs—Hypothesis Testing

12. **Testing the Mean:** Two major automobile manufacturers have produced compact cars with the same size engines. Is there a statistically-significant difference in the miles per gallon (MPG) of the two brands? A random sample of 8 cars is selected from each manufacturer. A specified driving distance is established and 8 drivers are selected. The number of drivers is not 16, but 8 who will drive car A and then drive car B. The results of the test are tabulated below in Data Table 9.4. Set $\alpha = 0.05$.

Data Table 9.4 MPG Study for Two Manufacturers.

Driver	MPG MFG A	MPG MFG B
1	29	27
2	24	23
3	26	28
4	24	23
5	25	24
6	27	26
7	30	28
8	25	27

13. **Testing the Mean:** Snow White buys her seven dwarfs (bet you can't name them) new shovels for Christmas. The tons dug with the old and new shovels are shown in Data Table 9.5. At alpha 0.10, did Snow White's gift to her seven buddies change the output or did it remain the same?

Data Table 9.5 Output in Tons of Old Shovels and New Shovels.

Dwarf	Daily Output in Tons	
	Old Shovels	New Shovels
Doc	1.7	1.9
Happy	1.4	1.5
Grumpy	2.1	2.2
Bashful	1.9	2.0
Sleepy	2.2	2.2
Dopey	1.4	1.5
Sneezy	1.9	1.8

Proportions—Confidence Intervals and Hypothesis Testing

14. **Confidence Interval:** A study is conducted to determine if there is a difference in the use of credit cards between men and women when Christmas gifts are purchased. The test included a sample of 150 men with 27% using credit cards and a sample of 130 women with 35% using credit cards. At the confidence level of 99%, develop and interpret the confidence interval for the proportion of men and women who rely on credit cards.

15. **Confidence Interval:** Of the 40 Dallas accounting firms surveyed, six indicated they were advertising. A similar survey was conducted in Phoenix where 8 out of 50 firms reported they were advertising. Determine a 90% confidence interval (alpha 0.10) to estimate the difference between the proportions of the firms who are advertising in the two cities. Interpret your results.

- 16. Hypothesis Testing:** Using the information given in problem #15, test the hypothesis that the proportions of those advertising in the two cities are the same. Use the alpha level of 0.10. Interpret your results. Explain the difference between the results determined with the confidence interval solution in #15 and this hypothesis testing approach.

KH

ALL RIGHTS

RESERVED

Extra Problems Chapter 9

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. Large Sample, Z-Solution:** Texas Health Care operates 33 medical clinics in the Dallas-Fort Worth Metroplex area. A study was initiated to determine the difference in the mean time spent per visit for men and women patients. Previous studies indicated the standard deviation is 11 minutes for men and 16 minutes for women. A random sample of 100 males and 100 females was selected and the sample means were 34.5 minutes for the males and 42.4 minutes for the females. Determine a 95% confidence interval and interpret the result. If there is a difference, how much is the difference?
- 2. Small Sample, t -solution, Equal Variances:** Baylor Hospital in Grapevine, TX is working with a medication they believe will help patients recover from certain illnesses more quickly. They have noticed that there seems to be a difference in response time in patients 50 and under and those patients over 50. To test this theory, they want to know if there is a difference in time when the medication actually reaches the blood stream. A random sample of six (6) people, ages 50 or under, yields a sample mean of 13.6 minutes with a standard deviation of 3.1 minutes. A random sample of eight (8) people over 50 yields a sample mean of 11.2 minutes with a standard deviation of 5.0 minutes. Determine the confidence interval at an alpha level of 0.05. Can the claim that the older patients process the medication more rapidly than the younger patients be supported?
- 3. Small Sample, t -solution, Equal Variances:** Apex Industries has two manufacturing plants which build sub-assembly parts for shipment to Lockheed-Martin. Plant A is reasonably well automated and Plant B still has a number of manual operations in place. In order to determine if the assembly times are different between the automated plant and the manual plant, a random sample of 15 parts for each plant is tracked. The results were that Plant A (automated) processed the parts in 56.7 hours with a standard deviation of 7.1 hours. Plant B (manual) processed the parts in 70.4 hours with a standard deviation of 8.3 hours. Determine and interpret a 95% confidence interval to see if the automated plant is more efficient than the manual plant.
- 4. Hypothesis Test, Large Sample, Z-Solution:** Bates Manufacturing makes parts for equipment used in the oil and gas business. One part is a coupling, useful on natural gas drilling equipment. Two different machines are used in manufacturing the part. The standard deviation for Machine 1 has been established at 0.025 inches and the standard deviation for Machine 2 has been established at 0.034 inches. The question is does Machine 2 produce a part with higher average diameters than Machine 1? The alternate would be, "is the Machine 1 diameter less than the Machine 2 diameter?" A random sample of 100 parts from each machine is selected. The results yielded a mean of 0.501 inches for Machine 1 and a mean of 0.509 for Machine 2. The claim is that Machine 1 produces a part that is less in diameter than Machine 2. Set up the null and alternate hypothesis and test the parts diameter at an alpha level of 0.05.

5. **Confidence Interval: Large Sample, Z-Solution.** Use the information from the problem just above (Bates Manufacturing) and calculate a 95% Confidence Interval. Did you find the same answer? How much is the difference, if one exists?
6. **P-Value:** Using the data from the problem just above (Bates Manufacturing), determine the p-value and interpret the results? Did you get the same answer? Why or Why Not?
7. **Hypothesis Testing: Small Sample, t -Solution, Equal Variances:** To determine if one retirement plan is more popular than another, Wilson Investments decided to evaluate the plans based on average contribution. Wilson's management wants to test the theory that the plan yielding the better result is the one into which more people will make larger contributions. A random sample of 15 people for Plan A and 15 people for Plan B is taken. The mean contribution of Plan A is \$2,119.70 with a standard deviation of \$709.70. The mean contribution of Plan B is \$1,777.70 with a standard deviation of \$593.90. Wilson's management wants to know at the alpha 0.05 level if the means can be considered to be equal. Assume equal population variances. Set up the null and alternate hypothesis and interpret the results.
8. **Hypothesis Testing: Small Sample, t -Solution, Equal Variances:** Budget Car Rental wants to see if there is a difference in the mean mileage per gallon for their most popular SUV. Budget believes that the mean mileage for driving on the highway will exceed that of driving in town. Budget randomly selects 15 SUVs to be driven on the highway and another 15 SUVs to be driven in the city. The vehicles are filled with exactly 14 gallons of gasoline. The test driver is asked to drive the vehicle until the car runs out of gas. The sample yielded the following results: Highway average miles per gallon was 19.6468 and City average miles per gallon was 16.1460. The variance for the Highway study was 18.4637 miles squared and for the City study was 29.6064 miles squared. At the alpha 0.05 level, do the Highway miles per gallon exceed the City miles per gallon?
9. **Matched Pairs:** Students at TCU are curious to see if the cost of books purchased through the bookstore are the same as the cost of those same books purchased online. Nineteen courses are selected and the prices of the textbooks are compared. The results of the study revealed that the mean cost of the textbooks through the bookstore was \$139.40 and online was \$126.70 and the standard deviation of the difference (S_d) is 30.4488. Using Hypothesis testing and the t -test, determine if there is a difference in the prices through the bookstore and online at the alpha level of 0.05.
10. **Confidence Interval:** Work the problem just above using a 95% confidence interval. Do the solutions yield the same result? Why or Why Not?
11. **Hypothesis Testing: Large Sample, Z-Solution.** A think-tank in Washington, DC wants to find out if there is a difference between the salaries of federal workers and those in the private sector. A random sample of 35 federal jobs and 32 private sector jobs is selected. The average salary for the federal jobs was \$66,700 and for the private sector jobs was \$60,400 with a standard deviation of \$12,000 and \$11,000, respectively. At the alpha level of 0.05, test to see if there is any difference.
12. **P-value:** Work the problem just above using the p-value approach (Federal and Private Sector Jobs).