

Unit 26: Small Sample Inference for One Mean



SUMMARY OF VIDEO

The z -procedures for computing confidence intervals or hypothesis testing work in cases where we know the population's standard deviation. But that's hardly ever the case in real life. For times when we don't know the population standard deviation but still want to figure out confidence intervals and do significance tests, statisticians turn to t -inference procedures. These t -procedures were invented in 1908 by William S. Gosset. Gosset was a chemist at the Guinness Brewery in Ireland. Making ale requires constant sampling of everything from barley to yeast to the beer itself. Gosset wanted to save time and money using small samples and their standard deviations as an estimate of the unknown population standard deviation, σ . Using the standard deviation s derived from only a few data values does not give a sufficiently good estimate of the entire population's standard deviation; and he couldn't proceed with a z -procedure. In his efforts to figure out a way around this issue, Gosset created a new class of distribution called the t -distributions.

The video now turns to a modern day brewery, Pretty Things Beer and Ale Project. Dann Paquette and Martha Holley-Paquette, owners of the operation, take samples at various stages in the brewing process. At one stage, they take a sample and measure its density. They aim for a density reading of at least 19.5 degrees Plato. (Degrees Plato measure how much more dense the liquid is than water.) In one batch of Baby Tree beer, they got a reading of 20.3 degrees Plato – a good sign that this batch will be great.

Let's imagine that Pretty Things wants to see how closely their production of Baby Tree beer is hitting their pre-fermentation density goal of 19.5 degrees Plato. Data collected from 10 batches are given below.

20.2 18.9 19.6 20.6 20.3 18.7 21.0 18.5 20.1 19.3

Since our sample size is small, its standard deviation could be quite far from the population standard deviation of all Baby Tree beer ever brewed. So, z -procedures won't work. Instead, we call on the t -procedure that William Gosset invented.

In Figure 26.1, we compare a t -distribution for a sample of size 3 to the standard normal curve.

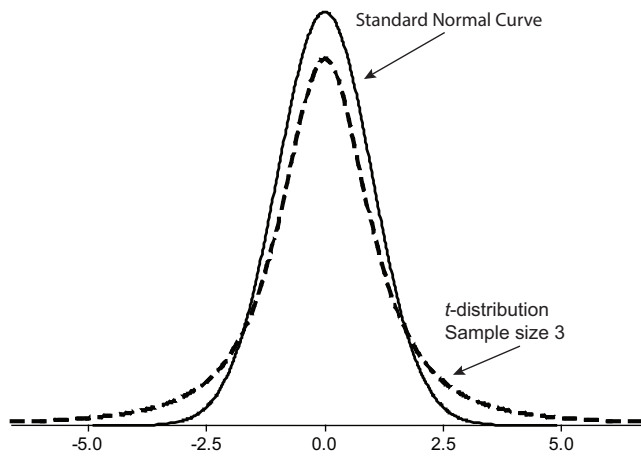


Figure 26.1. Comparing t -distribution to standard normal distribution.

The two curves share certain features. They are both bell-shaped. But the t -density curve is broader with a shorter peak, and its tails are higher. These fatter tails mean there is more probability of getting results far from zero. That's because the sample standard deviation varies from sample to sample, particularly when the sample size is small, adding uncertainty. Another difference is that although there is a single standard normal distribution, there is a different t -distribution for every sample size. As the sample size increases, the sample standard deviation s gets closer and closer to the population standard deviation, σ and the t -density curve gets closer to the standard normal curve.

The different t -distributions are specified by something called degrees of freedom, which are related to the sample size n : degrees of freedom = $n - 1$. For our beer data, the degrees of freedom are $10 - 1 = 9$. Next, we calculate a confidence interval for the population mean of the density of all Baby Tree beer. For these calculations to work, our data need to be from a normal distribution, which is a safe assumption in this case. Here's our formula:

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

Filling in information from our data, we get

$$19.72 \pm t^* \frac{0.85}{\sqrt{10}}$$

In order to determine t^* , we choose whatever confidence level we like – we'll go with 95%. We calculate the value of t^* from software as illustrated in Figure 26.2: $t^* = 2.262$.

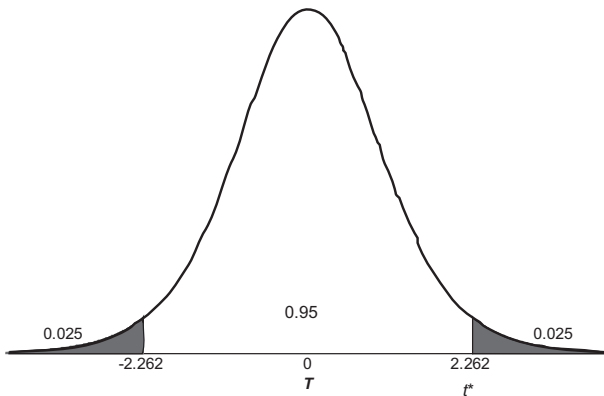


Figure 26.2. t -distribution with $df = 9$.

Plugging in the value of t^* , we get the following confidence interval:

$$19.72 \pm (2.262) \left(\frac{0.85}{\sqrt{10}} \right) \approx 19.72 \pm 0.61$$

So, our confidence interval (19.11, 20.33) gives us a range of plausible values for μ , the mean density of Baby Tree beer.

Take a moment to compare z^* , the z -critical value for a 95% z -confidence interval with our value for t^* :

$$z^* = 1.960$$

$$t^* = 2.262$$

Here, the t -critical value of 2.262 gives us a wider confidence interval. That is the price we pay for having a small sample and for not knowing σ .

Pretty Things' goal for their Baby Tree beer is a density of 19.5 degrees Plato. Using the confidence interval that we have calculated, we can say with 95% confidence that a 19.5 population mean is within our range of plausible values for μ .

STUDENT LEARNING OBJECTIVES

- A. Understand when to use t -procedures for a single sample and how they differ from the z -procedures covered in Units 24 and 25.
- B. Understand what a t -distribution is and how it differs from a normal distribution.
- C. Know how to check whether the underlying assumptions for a t -test or t -confidence interval are reasonably satisfied.
- D. Be able to calculate a t -confidence interval for a population mean.
- E. Be able to test a population mean with a t -test. Be able to calculate the t -test statistic and to determine the p -value as an area under a t -density curve.
- F. Be able to adapt one-sample t -procedures to analyze matched pairs data.

CONTENT OVERVIEW

In Units 24 and 25, we introduced z-procedures for (1) calculating confidence intervals for a population mean and (2) conducting significance tests about a population mean. The confidence interval formula and z-test statistic are as follows:

$$(1) \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} \quad (2) z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

For both procedures we assumed that the population was normally distributed or the sample size n was large, and that the population standard deviation σ was known. However, in real life, σ is generally unknown.

Let's start with an example. The weights (pounds) from a random sample of 16 4-year-old children who took part in a study on childhood obesity appear below.

37.1 26.7 36.1 36.2 40.3 43.9 36.2 40.7
42.5 34.8 37.9 34.5 31.1 36.4 35.7 33.4

From these data, we can compute the sample mean and sample standard deviation: $\bar{x} = 36.47$ lbs and $s = 4.23$ lbs. However, the population standard deviation σ is unknown. Nevertheless, we would like to calculate a confidence interval for μ , the mean weight of 4-year-olds.

It seems reasonable to assume that weights of 4-year-olds are normally distributed and the normal quantile plot in Figure 26.3 confirms this assertion.

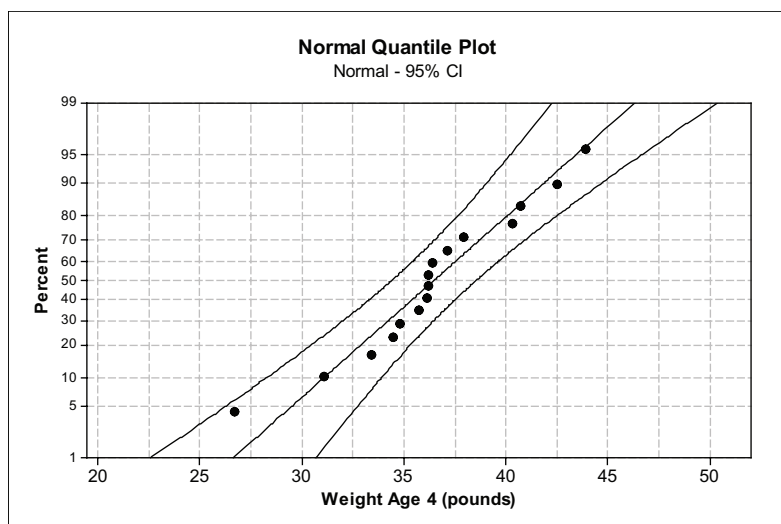
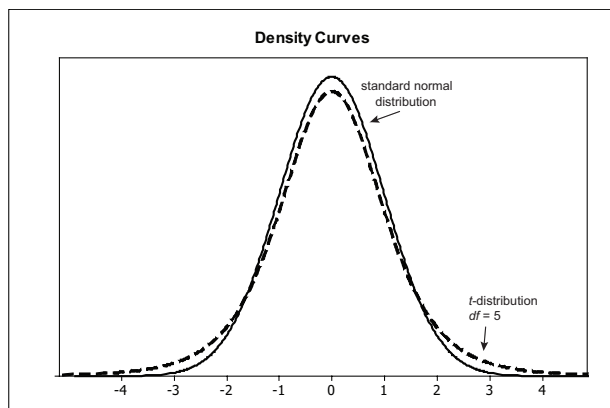


Figure 26.3. Normal quantile plot of children's weights.

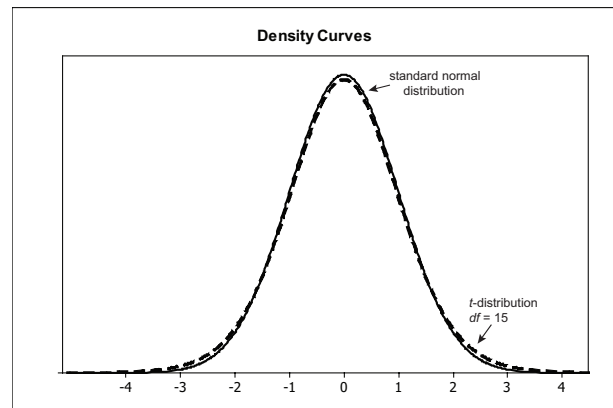
But we still have to deal with the fact that σ is unknown. We know that when the sample size n is large, s will be close to σ . But in this case $n = 16$, which is not large enough. Hence, simply replacing σ by s in the z -confidence interval formula would introduce too much additional variability. To compensate for the additional variability, we also replace z^* , a critical value from a standard normal distribution, with t^* , a critical value from a t -distribution. The result is the formula for computing **t -confidence intervals**:

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

The **t -distributions** have some features in common with the standard normal distribution. Both have density curves that are bell shaped and centered at zero as can be seen in Figure 26.4. However, there is more area in the tails of t -distributions than there is for the standard normal distribution. This difference is particularly noticeable when the degrees of freedom are small (See Figure 26.4(a)).



(a) t -distribution has 5 degrees of freedom.



(b) t -distribution has 15 degrees of freedom.

Figure 26.4. Comparison: two t -distributions with standard normal distribution.

The degrees of freedom (df) associated with t^* , the t -critical value in our confidence interval formula, are related to the size n of the sample:

$$df = n - 1$$

So, for our sample of 16 observations, $df = 15$. The value of t^* for a 95% confidence interval can be determined from a t -table (Figure 26.5) or using statistical software (Figure 26.6). In either case, $t^* = 2.131$.

DEGREES OF FREEDOM	CONFIDENCE LEVEL C		
	90%	95%	96%
15	1.753	2.131	2.249

Figure 26.5. Using a t-table to determine t^*

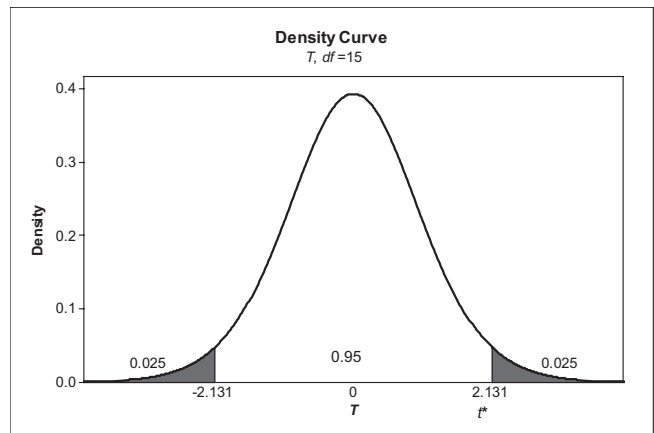


Figure 26.6. Using statistical software to determine t^* .

We now have everything that we need to calculate a 95% confidence interval for μ , the mean weight of 4-year-olds. Here are the calculations:

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}} = 36.47 \pm (2.131) \left(\frac{4.23}{\sqrt{16}} \right) = 36.47 \pm 2.25$$

Hence, we can say that μ is between 34.22 lbs and 38.72 lbs and that we have used a process to calculate this interval that has a 95% track record of giving correct results.

Next, we turn our attention to significance tests about a population mean μ for situations where the sample size is relatively small ($n < 30$), the population has a normal distribution, but σ is unknown. The **t-test statistic** results from replacing σ in the z-test statistic with s :

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

We return to our sample of 16 4-year-olds to see how this works.

Suppose that a height chart listed the average height of 4-year-olds as 39 inches. The heights of the sample of 16 4-year-olds are given below.

39.9 37.4 40.3 39.6 39.2 43.2 40.5 40.6
41.8 39.5 40.9 39.8 40.3 39.4 40.7 39.5

We suspect that children's heights have increased since the time the height chart was created due in part to better nutrition. To test our supposition, we let μ represent the mean height of

4-year-olds. The null and alternative hypotheses are:

$$H_0 : \mu = 39$$

$$H_a : \mu > 39$$

The sample mean and standard deviation for the height data are: $\bar{x} = 40.163$ inches and $s = 1.255$ inches. Now we calculate the t -test statistic, replacing μ_0 with its value from the null hypothesis and substituting in the sample values for \bar{x} and s :

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{40.163 - 39}{1.255/\sqrt{16}} \approx 3.71$$

If the null hypothesis is true, then the t -test statistic will have a t -distribution with $n - 1 = 16 - 1$ or 15 degrees of freedom. All that is left is to determine the p -value, the probability of observing a value at least as extreme as 3.71. Using statistical software we find $p \approx 0.001$ as illustrated in Figure 26.7. Hence, we reject the null hypothesis and conclude the mean height of 4-year-olds has increased since the time that the height chart was created.

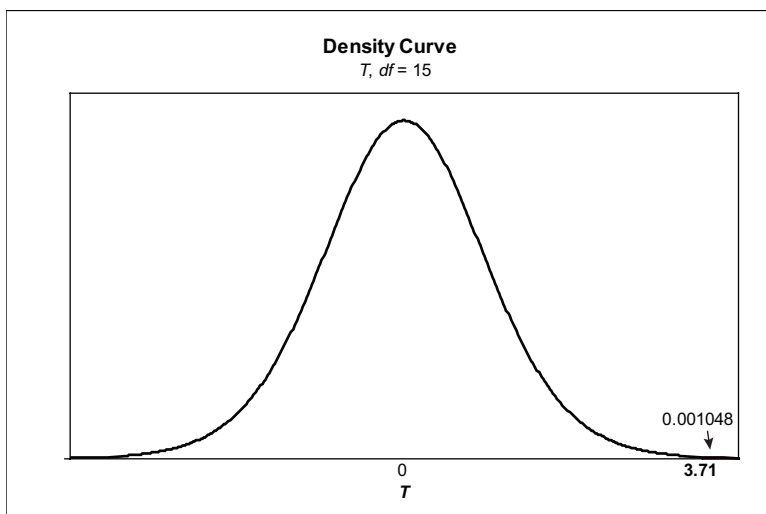


Figure 26.7. Determining the p -value.

The study involving the 16 children followed the children for two years. As part of the study, children were weighed when they were four and again when they were six. Table 26.1 shows the results, including the children's weight gain over the two-year period (Difference).

Weight Age 4 (lbs)	Weight Age 6 (lbs)	Difference (lbs)
37.1	46.9	9.8
26.7	35.9	9.2
36.1	48.3	12.2
36.2	44.2	8.0
40.3	50.6	10.3
43.9	56.4	12.5
36.2	51.8	15.6
40.7	50.2	9.5
42.5	51.6	9.1
34.8	41.4	6.6
37.9	55.3	17.4
34.5	44.5	10.0
31.1	39.9	8.8
36.4	46.3	9.9
35.7	49.6	13.9
33.4	39.0	5.6

Table 26.1. Change in weight from age 4 to age 6.

In the past, children around this age would have been expected to gain 4.5 pounds per year or 9 pounds over the two-year period. However, we suspect that the mean change in weight has increased. To test this assumption, we perform what is called a **matched-pairs t-test**. The parameter μ in a **matched-pairs t-procedure** is the mean difference in observations (or responses) on each individual (or subject in a match pair) – in our case, μ is the mean difference between weight at age 4 and weight at age 6. We set up the null hypothesis (no change from what was expected in the past) and alternative hypothesis (increase from what was expected in the past):

$$H_0 : \mu = 9$$

$$H_a : \mu > 9$$

Now, we calculate the one-sample t -test statistic using the differences. The sample mean and sample standard deviation of the differences are: $\bar{x}_D = 10.525$ lbs and $s_D = 3.122$ lbs. The matched-pairs t -test statistic is computed as follows:

$$t = \frac{\bar{x}_D - \mu_0}{s_D / \sqrt{n}}$$

$$t = \frac{10.525 - 9}{3.122 / \sqrt{16}} \approx 1.95$$

From Figure 26.8 we see that $p \approx 0.035$. Since $p < 0.05$, we reject the null hypothesis and accept the alternative that the mean weight gain in children from age 4 to age 6 is greater than 9 pounds.

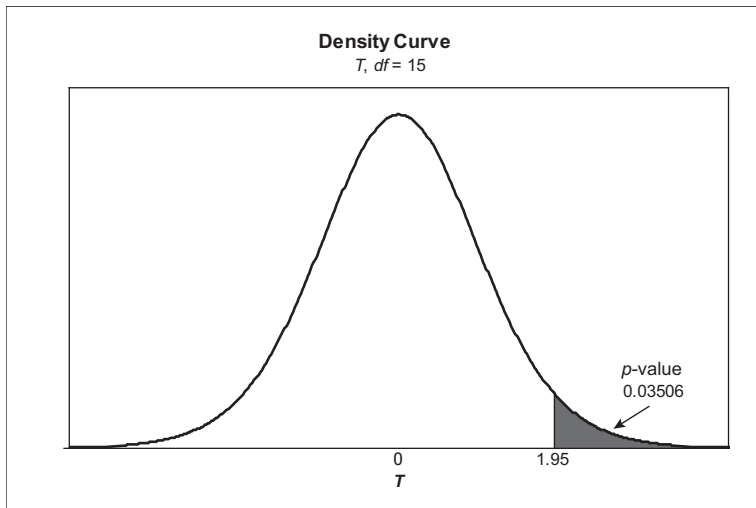


Figure 26.8. Calculating the p -value for matched-pairs t -test.

The last step in this analysis is to use a matched-pairs t -procedure to calculate a 95% confidence interval for μ , the mean weight gain from age four to age six. The matched-pairs t -confidence interval is computed as follows:

$$\bar{x}_D \pm t^* \left(\frac{s_D}{\sqrt{n}} \right)$$

$$10.525 \pm (2.131) \left(\frac{3.122}{\sqrt{16}} \right) = 10.525 \pm 1.663$$

Notice that the t -critical value, t^* , depends only on the sample size and the confidence level and not on whether we are calculating a one-sample or a **matched pairs t -confidence interval**. Our confidence interval for μ is from 8.9 lbs to 12.2 lbs.

Look back at the results from our two matched-pairs t -procedures. We concluded from the t -test that μ , the mean weight gain from age 4 to age 6, was greater than 9 pounds. However, our confidence interval for μ was (8.2, 12.2), an interval that includes values that are below 9 pounds. Results from a two-sided confidence interval do not always match the results from a significance test involving a one-sided alternative.

KEY TERMS

Density curves for ***t*-distributions** are bell-shaped and centered at zero, similar to the standard normal density curve. Compared to the standard normal distribution, a *t*-distribution has more area under its tails. The shape of a *t*-distribution, and how closely it resembles the standard normal distribution, is controlled by a number called its **degrees of freedom (*df*)**. A *t*-distribution with $df > 30$ is very close to a standard normal distribution.

A ***t*-confidence interval for μ** when σ is unknown is calculated from the formula:

$$\bar{x} \pm t^* \left(\frac{s}{\sqrt{n}} \right)$$

where t^* is a *t*-critical value associated with the confidence level and determined from a *t*-distribution with $df = n - 1$.

A ***t*-test statistic** for testing $H_0 : \mu = \mu_0$, where μ is the population mean, is given by:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

The *t*-test is a modification of the *z*-test and is used in situations where the population standard deviation σ is unknown and either the population has a normal distribution or the sample size n is large. The *p*-value is determined from a *t*-distribution with $df = n - 1$.

A **matched-pairs, *t*-confidence interval for μ_D** , the population mean difference, is given by the formula:

$$\bar{x}_D \pm t^* \left(\frac{s_D}{\sqrt{n}} \right)$$

where t^* is a *t*-critical value associated with the confidence level and determined from a *t*-distribution with $df = n - 1$ and \bar{x}_D and s_D are the mean and standard deviation of the sample differences.

A **matched-pairs *t*-test statistic for testing $H_0 : \mu_D = \mu_{D0}$** , where μ_D is the population mean difference, is given by

$$t = \frac{\bar{x}_D - \mu_{D0}}{s_D/\sqrt{n}}$$

where \bar{x}_D and s_D are the mean and standard deviation of the sample differences.

THE VIDEO

Take out a piece of paper and be ready to write down answers to these questions as you watch the video.

1. Why won't the z -procedure work in most cases, particularly if the sample size is small?
2. Who invented t -inference procedures?
3. Compare a normal density curve with a t -distribution for a sample size of 3. How are the two distributions similar and how do they differ?
4. For a t -distribution, how are the degrees of freedom related to sample size?
5. For a 95% confidence interval, which is larger, z^* or t^* ?

UNIT ACTIVITY:

STEP-BY-STEP

Pedometers count the number of steps a person walks. If you want your pedometer to calculate how far you have walked, you need to enter in your step length (distance from heel of one foot to heel of other foot when walking). In this activity, you will collect data on step length for males and females in the class. Assuming that students in your class are representative of the student population, you will calculate confidence intervals for the mean step length of males and females.

1. Discuss methods for getting reliable measurements for step length. After your group discussion, the class must decide on the method that will be used to collect the step-length data. Write a brief description of this method.

2. Collect the step-length data for males and females separately. After the data are collected, the two data sets (male step lengths, female step lengths) should be distributed to the class.

In answering the remaining questions, assume that the class data are representative of the general male and female student populations.

3. Check that the underlying assumption of normality is reasonably satisfied in both data sets.

4. a. Calculate the mean and standard deviation for the male step-length data.

b. Calculate the mean and standard deviation for the female step-length data.

5. a. Calculate a 95% confidence interval for the mean step length of males. What are the degrees of freedom of the t -critical value?

b. Calculate a 95% confidence interval for the mean step lengths of females. What are the degrees of freedom of the t -critical value?

6. Based on your confidence intervals in question 5, can you conclude that the mean step length for males is greater than for females? Explain.

EXERCISES

1. A woman in a nursing home is on medication for high blood pressure. Her blood pressure is taken daily. A sample of 20 blood pressure readings (mmHg) appears below.

150	148	136	120	142	144	130	150	130	142
140	130	148	142	138	130	120	166	130	152

- Calculate the sample mean and standard deviation.
- Determine a 95% confidence interval for μ , her mean systolic blood pressure. Show your calculations.
- Is the underlying assumption that these data come from a normal distribution reasonably satisfied? Explain.

2. A manufacturer of brass washers produces one type of washer that has a target mean thickness of 0.019 inches. After the production process had continued for some time without any adjustment, a random sample of 10 washers was selected and measured for thickness. The data are given below.

0.0185	0.0190	0.0180	0.0184	0.0179
0.0186	0.0188	0.0178	0.0182	0.0186

- Does the assumption that washer thickness is normally distributed seem reasonable given these data? Explain.
- Do you think the production process is still in control or do the data indicate that it is time to make some adjustments? To answer this question, test the hypothesis that the mean thickness equals 0.019 inches against the hypothesis that it does not. Report the value of the test statistic, the p -value and your conclusion.
- Calculate a 95% confidence interval for μ , the mean thickness of washers currently being produced. Show your calculations. Does your interval indicate that the process needs to be adjusted to increase washer thickness or decrease washer thickness? Explain.

3. Students in a statistics class measured their foot lengths and forearm lengths. The data are given in Table 26.2. Assume this sample is representative of the student population.

Forearm Length (inches)		Foot Length (inches)	
8.75	10.00	10.00	9.00
9.00	9.00	9.00	10.50
8.50	10.00	9.50	11.00
10.25	10.00	10.00	11.50
10.25	11.50	12.50	11.25
8.50	9.00	11.50	9.00
9.25	8.50	9.00	10.50
10.50	6.75	9.25	10.50
8.25	10.00	9.50	8.50
9.00	8.25	8.25	10.00
7.00	8.25	9.50	8.75
9.50	9.00	9.50	8.75
9.75	8.00	9.50	10.00

Table 26.2. Data on forearm and foot lengths.

- Calculate a 95% confidence interval for μ_{Forearm} , the mean forearm length of students. Then calculate a 95% confidence interval for μ_{Foot} , the mean foot length of students.
- Do your 95% confidence intervals support the hypothesis that the mean forearm length of students differs from the mean foot length of students? Explain.
- Calculate 90% confidence intervals for μ_{Forearm} and μ_{Foot} . Compare the 95% confidence intervals to the 90% confidence intervals. Which of the two were wider? Explain why that was the case.
- Answer part (b) based on the 90% confidence intervals calculated for (c).

4. A statistics professor was concerned that students were not as successful on the second exam (which was on inference) as they were on the first exam (which was on descriptive statistics). She took a random sample of 15 students enrolled in her introductory statistics courses over the past three years. Their grades on these two exams appear in Table 26.3 (continued on next page).

Exam 1	Exam 2
67	59
74	82
85	96
71	62
78	69
96	83
63	52
91	94
93	82
81	67
84	66
64	66
88	84
89	75
82	90

Table 26.3. Statistics exam scores.

- Compute the differences between Exam 2 and Exam 1. What is the sample mean for the differences? What is the sample standard deviation for the differences?
- Return to the professor's concern that students were not as successful on the material for Exam 2 as they were on the material for Exam 1. Let μ_D be the population mean difference between the two exam scores. Set up null and alternative hypotheses to test the professor's supposition.
- Conduct a significance test. Report the value of the test statistic and the p -value.
- Do the results of your test of hypotheses support the professors' concern? Explain.

REVIEW QUESTIONS

1. Given a simple random sample of size n , you want to compute a confidence interval for μ of the form: $\bar{x} \pm t^* \left(s / \sqrt{n} \right)$. Find the value of t^* for each of the following confidence levels and sample sizes.

- A 95% confidence interval for a sample size $n = 12$.
- A 99% confidence interval for a sample size $n = 15$.
- An 80% confidence interval for a sample size $n = 10$.

2. Supermarket rotisserie chickens have become very popular with the American public. A study was conducted to compare the nutrient composition of commercially-prepared rotisserie chicken to that of roasted chicken, which is listed in the *USDA National Nutrient Database for Standard Reference (SR)*.

- The *Standard Reference (SR)* listed the mean protein content of roasted chicken breast as 31 grams. In the sample of 9 rotisserie chickens, $\bar{x} = 29.86$ grams and $s = 1.95$ grams. Conduct a t -test to see if the mean protein content in rotisserie chicken breasts differs from the SR. Report the value of the test statistic, the p -value, and your conclusion.
- The *SR* listed the mean cholesterol level in roasted chicken thighs as 95 milligrams. In a sample of 9 rotisserie chicken thighs, $\bar{x} = 134$ milligrams and $s = 2.43$ milligrams. Conduct a t -test to see if the mean cholesterol level in rotisserie chicken thighs differs from the SR. Report the value of the test statistic, the p -value, and your conclusion.

3. A researcher studying ocean literacy focused her efforts on the program Ocean Commotion – a one-day ocean/wetlands literacy program that includes hands-on demonstrations about marine environments and products. Prior to attending the program, a sample of 337 students from 6 schools were given a pre-test to measure their attitudes toward the ocean and wetlands. After the program, students were given a post-test. The test was graded on a scale from 1 (lowest) to 5 (highest). The mean score on the pre-test was 4.06. The mean score on the post-test was 4.13. The standard deviation of the pre-post differences was 0.40.

- a. The researcher wanted to test whether attendance at Ocean Commotion had a significant positive effect on students' attitudes toward the ocean and wetlands. Let μ_D be the mean difference in attitude after the program compared to before the program. Write the researcher's null and alternative hypotheses.
- b. Calculate the value of the t -test statistic. What are the degrees of freedom associated with the t -test statistic?
- c. What is the p -value?
- d. The researcher concluded that Ocean Commotion had a significant effect on students' attitudes toward the ocean and wetlands. Given your answer to (c), do you agree with her conclusion? Do you think that the difference is of practical importance? Explain.
4. A psychology class wanted to find out whether attendance at a workshop on happiness can teach people how to be happier. The class arranged to attend a "happiness" workshop. A week prior to going they completed the Oxford Happiness Questionnaire. The questionnaire, developed by psychologists Michael Argyle and Peter Hills, is scored on a scale from 1 (not happy) to 6 (too happy). Students completed the Oxford Happiness Questionnaire a second time two days after the workshop and a third time six weeks after the workshop. Table 26.4 displays their data (See table on next page).

First Questionnaire	Second Questionnaire	Third Questionnaire
2.6	3.1	2.2
4.9	5.0	2.5
6.0	4.8	3.9
3.3	3.5	4.5
4.1	4.2	2.9
3.8	4.0	3.1
5.2	4.6	4.8
4.3	4.7	5.1
3.1	5.3	3.9
5.5	5.2	3.8
4.2	5.4	4.9
2.0	3.2	2.8
4.3	5.1	2.9
3.6	4.6	3.2
3.1	3.8	2.9
5.2	4.7	3.4

Table 26.4. Results from Oxford Happiness Questionnaire

- Students thought that the workshop would have a positive effect on happiness, at least short term. Use the data from the first and second questionnaires to test their hypothesis. State the null and alternative hypotheses, calculate the value of the t -test statistic, determine the p -value and give your conclusion.
- Use the data from the first and third questionnaires to test whether there is any long-term positive effect on students from this type of happiness workshop. State the null and alternative hypotheses, the value of the test statistic, the p -value, and your conclusion.
- Let $\mu_{(\text{Third} - \text{First})}$ be the population mean difference of Oxford Happiness Questionnaire scores: taken before and six weeks after participation in a happiness workshop. To estimate the long-term effect of the workshop on students' happiness, calculate a 95% confidence interval for $\mu_{(\text{Third} - \text{First})}$. Interpret what it tells you about students who participated in happiness workshops before and after participating in the Oxford Happiness Questionnaire.
- The sample for this study consisted of students from one psychology class. Do you think the results are valid for all college students from this university? Explain.