

# Unit 26: Small Sample Inference for One Mean



## PREREQUISITES

Students need the background on confidence intervals and significance tests covered in Units 24 and 25.

## ADDITIONAL TOPIC COVERAGE

Additional coverage of  $t$ -confidence intervals and significance tests can be found in *The Basic Practice of Statistics*, Chapter 18, Inference about a Population Mean.

## ACTIVITY DESCRIPTION

Most pedometers record step lengths. If you want to know how far you have walked, then you have to calibrate the pedometer by entering your step length. In this activity, students will use  $t$ -confidence intervals to estimate the mean step length of male students and female students. They will first need to collect two sets of data: (1) male step lengths and (2) female step lengths.

## MATERIALS

Meter stick or tape measure.

The sample data used in the solutions to this activity were gathered from 10th-grade students. Expect estimates to be larger for college students, especially for males.

In question 1, students are asked to write a plan for collecting the step-length data. This question could also be a class discussion. In gathering the data, students will get better results if they measure the distance of more than one step and then divide the total distance by the number of steps they walked.

In question 4, students are asked to determine whether the mean step length for females differs from the mean step length for males. If there is any overlap in their two confidence intervals, then the answer must be no. However, Unit 27 provides a better way of addressing this question – using a two-sample  $t$ -test or confidence interval.

Unit 27's Content Overview will use the sample data from this activity as part of the discussion of two sample  $t$ -procedures.

# THE VIDEO SOLUTIONS

1. The z-procedure assumes that the population standard deviation is known. In most cases, we don't know  $\sigma$ .
2. William S. Gosset
3. Both density curves have a bell shape and are centered at zero. However, the  $t$ -distribution has a shorter peak but is more spread out than the standard normal density.
4. The degrees of freedom are one less than the sample size:  $df = n - 1$ .
5.  $t^*$  is larger.

# UNIT ACTIVITY SOLUTIONS

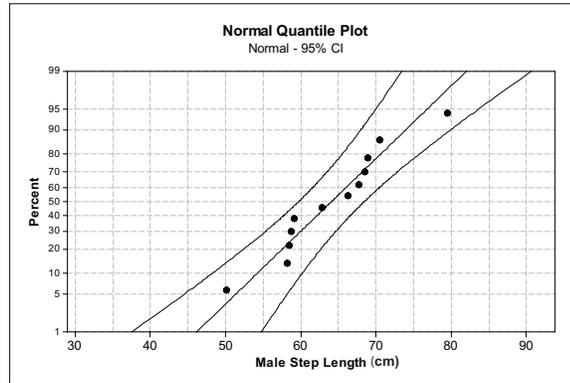
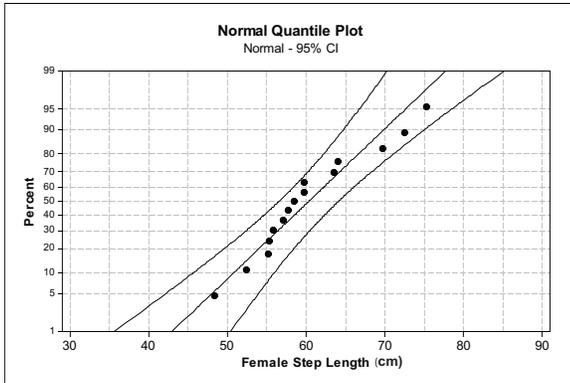
1. Sample answer: Place your heel against the wall. Then take four steps. Measure the distance from the wall to the back of your heel after the fourth step. Divide this distance by four to get the step length.

Note: It is important to measure more than one step since your step length with your right foot might differ from your step length with your left foot.

2. Sample data:

Male Step Length (cm)	Female Step Length (cm)
58.250	55.875
68.500	52.375
58.500	55.375
50.125	59.750
58.750	48.375
62.875	57.125
59.125	64.000
67.750	57.750
68.875	63.500
66.250	69.750
79.500	72.500
70.500	75.250
	58.500
	59.750
	55.250

3. Sample answer (based on sample data from question 2).



Given that the pattern of the dots appears fairly linear in both normal quantile plots, and that all dots remain inside the curved bands, it is reasonable to assume that male and female step lengths are normally distributed.

4. a. Sample answer:  $\bar{x}_M = 64.08$  cm and  $s_M = 7.71$  cm

b. Sample answer:  $\bar{x}_F = 60.34$  cm and  $s_M = 7.47$  cm

5. a. Sample answer:  $df_M = 12 - 1 = 11$

$$64.08 \pm (2.201) \left( \frac{7.71}{\sqrt{12}} \right) \approx 64.08 \pm 4.90 \text{ or } (59.18, 68.98)$$

b. Sample answer:  $df_F = 15 - 1 = 14$

$$60.34 \pm (2.145) \left( \frac{7.47}{\sqrt{15}} \right) \approx 60.34 \pm 4.14 \text{ or } (56.20, 64.48)$$

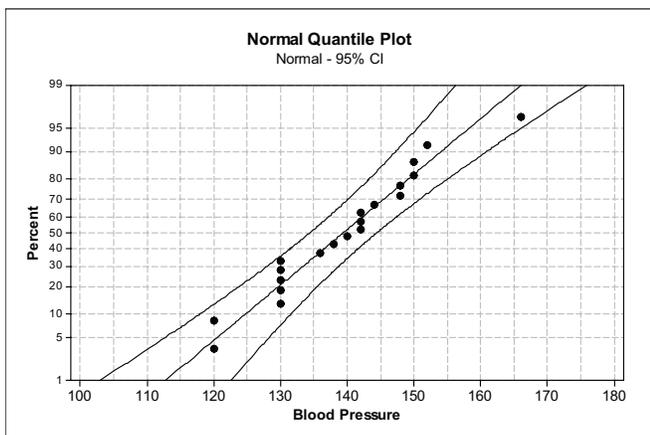
6. Sample answer: The confidence intervals contain the plausible values for the population means. Since the two confidence intervals share some values, you cannot conclude that the mean step length for men differs from the mean step length for women.

# EXERCISE SOLUTIONS

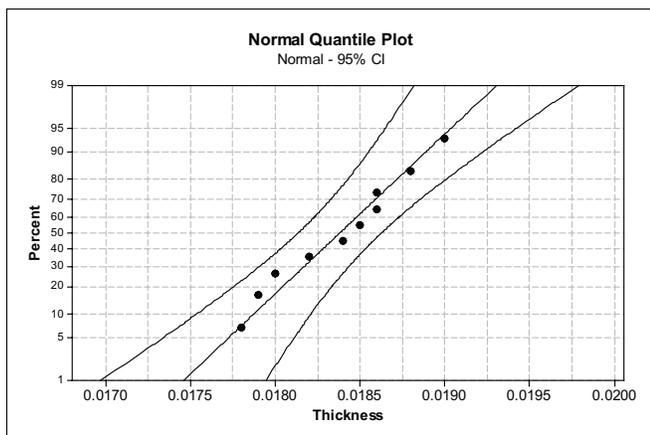
1. a.  $\bar{x} = 139.40$  mmHg and  $s = 11.45$  mmHg

b. The endpoints of the confidence interval are  $139.40 \pm 2.093(11.45/\sqrt{20}) \approx 139.40 \pm 5.36$  or (134.04, 144.76).

c. Although the normal quantile plot below is not completely linear due to the five readings of 130, all the dots stay within the curved bands. In addition, a boxplot shows no outliers. So, the data come from a distribution that is close enough to a normal distribution so that  $t$ -procedures are valid.



2. a. The dots in the normal quantile plot appear roughly linear and all dots stay within the curved bands. It is reasonable to assume the data come from a normal population.



b.  $\bar{x} = 0.01838$  and  $s \approx 0.00040$ . The value of the  $t$  test statistic is:  $t = \frac{0.01838 - 0.019}{0.0004/\sqrt{10}} \approx -4.90$

To determine the  $p$ -value, we consult a  $t$ -distribution with  $df = 9$ .

This is a two-sided test, so we need the area under the  $t$ -density curve to the left of  $-4.90$  and to the right of  $4.90$ :  $p$ -value  $\approx 2(0.0004)$  or slightly less than  $0.001$ . Since this  $p$ -value is so small, we conclude that the mean thickness of the brass washers currently being produced is not  $0.019$ . The manufacturing process needs to be adjusted.

c.  $0.01838 \pm (2.262)(0.0004 / \sqrt{10}) \approx 0.01838 \pm 0.00029$  or  $(0.0181 \text{ inch}, 0.0187 \text{ inch})$ . Since the target mean is  $0.019$  inch, the confidence interval indicates that the mean has fallen below the target. Hence, the process needs to be adjusted to increase washer thickness.

3. a.  $9.106 \pm (2.060)(1.073/\sqrt{26}) \approx 9.106 \pm 0.433$  or  $(8.673, 9.539)$ ;

$9.856 \pm 2.060(1.054/\sqrt{26}) \approx 9.856 \pm 0.426$  or  $(9.430, 10.282)$

b. No. The confidence intervals represent the plausible values for the population means. Since the two confidence intervals overlap, they have some values in common which are plausible values for both  $\mu_{\text{Forearm}}$  and  $\mu_{\text{Foot}}$ .

c. 90% confidence interval for  $\mu_{\text{Forearm}}$ :

$9.106 \pm (1.708)(1.073/\sqrt{26}) \approx 9.106 \pm 0.359$  or  $(8.747, 9.465)$

90% confidence interval for  $\mu_{\text{Foot}}$ :

$9.856 \pm 1.708(1.054/\sqrt{26}) \approx 9.856 \pm 0.353$  or  $(9.503, 10.209)$

The 95% confidence intervals were wider. In order to be more confident, the precision of the interval estimates decreases, which means the intervals get wider.

d. In this case, the confidence intervals support the hypotheses that the two means are different because the two intervals do not overlap.

4. a. The differences are:

-8 8 11 -9 -9 -13 -11 3 -11 -14 -18 2 -4 -14 8

$\bar{x}_D = -5.27$  and  $s = 9.32$

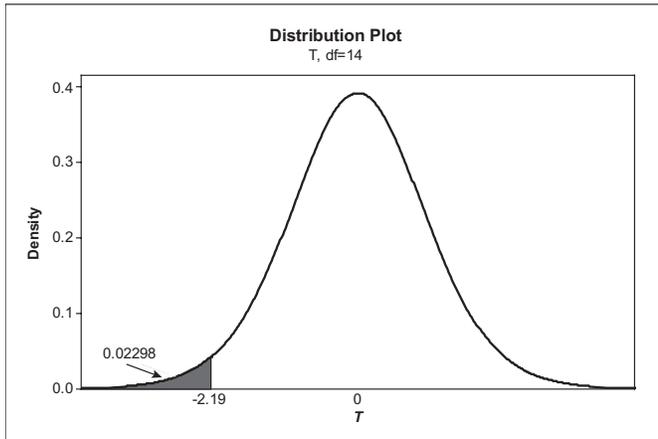
b. The null hypothesis is that there is no difference between Exam 1 and Exam 2 scores. The alternative hypothesis is that scores went down from Exam 1 to Exam 2, hence the mean difference (Exam 2 – Exam 1) is negative. These hypotheses can be expressed symbolically

as follows:  $H_0 : \mu_D = 0$

$H_a : \mu_D < 0$

c.  $t = \frac{-5.27 - 0}{9.32/\sqrt{15}} \approx -2.19$ ;  $p\text{-value} \approx 0.23$ ,

which was computed using software and is illustrated below.



d. Yes the professor's concern is supported by the data. The  $p$ -value indicates that the null hypothesis should be rejected in favor of the alternative. Hence, the data support the hypothesis that the mean difference in exam scores is less than zero. On average, students do worse on the second exam.

# REVIEW QUESTIONS SOLUTIONS

1. a. For the  $t$ -distribution,  $df = 11$ ;  $t^* = 2.201$

b. For the  $t$ -distribution,  $df = 14$ ;  $t^* = 2.977$

c. For the  $t$ -distribution,  $df = 9$ ;  $t^* = 1.383$

2. a.  $t = \frac{29.86 - 31}{1.95 / 3} \approx -1.754$  ;  $p = 2(0.05876) \approx 0.118$ .

The mean protein content of rotisserie chicken breasts does not differ from its listing in the *SR*.

b.  $t = \frac{134 - 95}{2.43 / 3} \approx 48.148$  ;  $p$ -value =  $2(1.915 \times 10^{-11}) \approx 0$ .

The mean cholesterol level of rotisserie chicken thighs does differ significantly from what is listed in the *SR*.

3. a.  $H_0 : \mu_D = 0$   
 $H_a : \mu_D > 0$

b.  $t = \frac{(\bar{x}_{\text{Post-test}} - \bar{x}_{\text{Pre-test}}) - 0}{s / \sqrt{n}} = \frac{(4.13 - 4.06) - 0}{0.40 / \sqrt{337}} \approx 3.21$ ; degrees of freedom = 336

c.  $p = 0.0007$

d. Yes, the  $p$ -value is very small, which provides strong evidence against the null hypothesis in favor of the alternative. In fact,  $p < 0.05$ , the standard cutoff for significance. The difference in sample means is only 0.07. While that difference is significant, it does not represent a large improvement in attitude.

4. a. Let  $\mu_D$  be the mean of the differences, second questionnaire score – first questionnaire score. The null and alternative hypotheses are:

$$H_0 : \mu_D = 0$$

$$H_a : \mu_D > 0$$

The test statistic is  $t = \frac{0.375 - 0}{0.827/\sqrt{16}} \approx 1.814$ ;  $p = 0.045$ .

Given  $p < 0.05$ , we conclude that the workshop had a positive short term effect on students' happiness.

b. Let  $\mu_D$  be the mean of the differences, third questionnaire score – first questionnaire score. The null and alternative hypotheses are:

$$H_0 : \mu_D = 0$$

$$H_a : \mu_D > 0$$

The test statistic is  $t = \frac{-0.525 - 0}{1.159/\sqrt{16}} \approx -1.812$ ;

$p = 0.955$ . We conclude that the workshop had no long term positive effect on students' happiness.

$$c. -0.525 \pm 2.131 \left( \frac{1.159}{\sqrt{16}} \right) \approx -0.525 \pm 0.617 \text{ or } (-1.142, 0.092)$$

Since the confidence interval includes 0, it is unlikely that the workshop had a positive long-term effect.

d. Sample answer: No. Psychology students are probably not representative of all students. For example, biology students or art students might have responded quite differently to the happiness workshop.