

Unit 22: Sampling Distributions



PREREQUISITES

This unit serves as the transition from descriptive statistics to inferential statistics. The concepts in this unit are more sophisticated than in earlier units. Students should have considerable exposure to data before viewing this unit. They need familiarity with the mean and standard deviation (Unit 4, Measures of Center, and Unit 6, Standard Deviation). Most importantly, students need background on the normal distribution (Units 7 – 9).

ADDITIONAL TOPIC COVERAGE

Additional coverage of sampling distributions can be found in *The Basic Practice of Statistics*, Chapter 11, Sampling Distributions. More in-depth information on \bar{x} charts can be found in the Companion Chapters (CD insert), Chapter 17, Statistics for Quality: Control and Capability. Unit 23, Control Charts, continues the discussion of control charts. There are a number of applets that allow students to conduct simulations that give them insight into the Central Limit Theorem. For example, take a look at the Central Limit Theorem applet at www.causeweb.org/repository/statjava/

ACTIVITY DESCRIPTION

This activity supports Learning Objectives for this unit A, B, and C. In this activity, students draw random samples by hand from an approximate normal distribution. After calculating the sample means, students compare a histogram of individual observations to a histogram of the sample means.

Drawing samples by hand makes the process of random sampling more concrete to students. Since drawing 100 samples by hand is time-consuming, after several samples have been selected, students can be working on other material until it is their turn to draw one or more samples. If students have access to statistical software, samples can be generated using the

software's normal random number generator (see Extension for Minitab instructions). If you decide to use technology to create the samples, it is still a good idea to begin the activity by having students draw several samples by hand.

MATERIALS

Prepare 100 identical slips of stiff poster board.

Write the numbers as described in Table T22.1 on the slips.

Write each of these numbers	On this many slips
50	10
49, 51	9
48, 52	9
47, 53	8
46, 54	6
45, 55	5
44, 56	3
43, 57	2
42, 58	1
41, 59	1
40, 60	1

Table T22.1. Distribution table for approximate normal data.

These 100 numbered slips form a population. The distribution of the numbers in this population is roughly normal with mean $\mu = 50$ and standard deviation $\sigma = 4$. Before students begin sampling, they should make a graphic display of the population distribution (question 1).

Before students can answer question 2, they need to collect the data and the instructor needs to provide the instructions. The numbered slips should be put into a container and mixed. A student should be asked to draw a sample of size nine for Sample 1 by drawing a slip of paper, recording its number, and returning it to the container for mixing before drawing the second number, and so on until nine numbers have been drawn. The values should be recorded either in a copy of Table T22.2 or in an Excel spreadsheet. This process should be repeated as many times as convenient – around 100 times if possible. The data should be distributed to students so that they can complete the activity. **Students should save the data in Table T22.2 for use in the activity for Unit 24, Confidence Intervals.**

Extension

Students use technology (such as Excel or Minitab) to generate 100 (or more) samples of size 9 from a uniform distribution on the interval from 0 to 1. A graphic display for this population's distribution is box shaped and centered at 0.5. However, the sampling distribution of \bar{x} appears roughly normal in shape but remains centered at 0.5.

Using Excel to Generate the Samples

- Label the first row of columns A – K with the following headings: Sample, X1, X2, X3, X4, X5, X6, X7, X8, X9, and Mean.
- In the column Sample, generate the numbers 1 – 100 as follows: enter 1 in cell A2. In cell A3, enter the formula =a2+1 and then press Enter. Click cell A2. Then click the bottom right corner of the cell and drag down to row 101.
- In cell B2, enter the formula =rand() and press Enter. Click cell B2 and then click the bottom right corner of the cell and drag across the row to cell J2 to form the first sample.
- With the first sample still highlighted, click the bottom right corner and drag down to form the other 99 samples.
- To find the mean of the first sample, click cell K2 (the first entry in the column labeled Mean). Enter the formula =AVERAGE(B2:J2) and press Enter.
- To calculate the means for the remaining 99 samples, click cell K2. Then click the bottom right corner and drag down to cell K101.

Using Minitab to Generate the Samples

- Label columns C1 – C11 as Sample, X1, X2, X3, X4, X5, X6, X7, X8, X9, and Mean.
- Select Calc > Make Patterned Data > Simple Set of Numbers. Then complete the dialog box as follows:

Store patterned data in C1
From first value 1
To last value 100
In steps of 1
Click OK

You should see consecutive integers 1 – 100 under Sample.

- Generate 100 samples of size 9: Calc > Random Data > Uniform. Then complete the dialog box as follows:

Number of rows of data to generate 100 (or more)

Store in columns: Select C2 – C10 for X1 – X9

Click OK

- To calculate the sample means for each sample: Calc > Row Statistics and then complete as follows:

Select Mean for the statistic

Select C2 – C10 for the input variables

Store results in: C11

Click OK.

Note: The Minitab instructions above can be adapted to replace drawing by hand from a normal population. To do so, adapt the second bullet above as follows:

- Generate 100 samples of size 9: Calc > Random Data > Normal and then complete the dialog box as follows:

Number of rows of data to generate 100 (or more)

Store in columns: Select C2 – C10 for X1 – X9

Mean: 50

Standard deviation: 4

Click OK

Samples of Size 9, page 1

Sample	X1	X2	X3	X4	X5	X6	X7	X8	X9	Mean
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										
32										
33										
34										
35										
36										
37										
38										
39										

Samples of Size 9, page 2

Sample	X1	X2	X3	X4	X5	X6	X7	X8	X9	Mean
40										
41										
42										
43										
44										
45										
46										
47										
48										
49										
50										
51										
52										
53										
54										
55										
56										
57										
58										
59										
60										
61										
62										
63										
64										
65										
66										
67										
68										
69										
70										
71										
72										
73										
74										
75										
76										
77										
78										

Samples of Size 9, page 3

Sample	X1	X2	X3	X4	X5	X6	X7	X8	X9	Mean
79										
80										
81										
82										
83										
84										
85										
86										
87										
88										
89										
90										
91										
92										
93										
94										
95										
96										
97										
98										
99										
100										

Table T22.2. Data table for samples of size 9.

Table T22.3 contains sample data that you might use if you don't want to have students draw all 100 samples. The sample solution will be based on these data.

X1	X2	X3	X4	X5	X6	X7	X8	X9	Sample Mean
40	55	45	53	41	50	54	60	56	
53	55	50	53	51	43	54	53	50	
57	48	46	50	46	48	53	51	47	
44	50	55	47	44	44	47	59	53	
48	59	51	49	51	47	47	45	55	
58	41	46	56	46	55	49	50	51	
51	46	51	40	49	47	48	60	46	
50	47	51	52	51	51	60	53	48	
59	54	49	49	49	41	56	46	56	
52	44	46	57	43	46	50	50	47	
51	45	53	54	46	48	49	46	41	
52	48	58	57	56	44	52	50	49	
47	50	53	58	47	44	48	47	48	
52	54	50	49	53	43	53	59	49	
54	44	51	43	46	52	53	58	47	
51	43	49	51	46	49	48	46	48	
51	58	48	45	51	50	59	53	55	
45	51	45	46	52	48	52	48	54	
52	56	51	52	53	53	43	47	48	
53	47	52	50	55	54	46	49	55	
50	40	57	51	52	52	48	46	49	
51	46	47	47	52	55	46	51	60	
56	41	51	52	48	44	51	48	47	
50	50	49	56	60	42	52	45	57	
48	48	54	49	53	49	51	49	47	
48	50	51	52	51	50	47	50	50	
54	53	55	46	48	48	52	52	47	
45	51	53	52	46	40	49	57	43	
49	56	53	48	50	54	44	52	50	
45	52	44	57	48	41	49	51	46	
50	53	52	49	50	53	53	52	55	
49	52	44	40	43	50	51	50	55	
47	51	50	56	47	54	55	50	46	
52	51	51	53	42	47	47	54	45	
53	49	40	49	49	49	48	52	48	
54	48	57	54	48	55	46	52	55	
50	47	55	46	49	52	45	53	46	

48	48	50	54	54	45	49	51	55	
52	49	49	46	53	48	43	55	54	
46	51	51	53	56	53	60	58	48	
56	56	52	45	46	53	51	59	54	
46	55	50	53	53	54	43	47	43	
54	50	53	48	54	46	47	41	41	
50	46	52	45	50	57	50	51	50	
55	50	58	45	49	51	47	57	46	
47	52	52	54	49	48	51	47	53	
46	49	53	54	53	47	52	55	44	
56	43	47	54	50	50	59	54	52	
41	55	51	51	52	53	48	49	55	
51	54	47	46	50	52	52	50	52	
59	47	42	44	50	44	41	45	53	
55	46	48	44	48	56	48	47	46	
52	54	43	54	43	54	51	42	50	
46	52	46	50	40	51	49	52	53	
53	50	46	48	53	47	52	52	54	
49	56	50	46	58	47	58	55	57	
46	60	45	48	49	48	51	53	54	
46	51	46	54	48	53	49	51	47	
56	47	48	52	48	49	52	55	50	
50	46	57	45	46	46	55	52	45	
52	41	52	46	51	51	48	49	40	
49	46	44	51	58	49	41	48	49	
47	56	52	43	47	50	52	50	48	
45	47	52	49	49	45	54	56	46	
53	50	45	49	51	52	49	51	51	
49	46	44	45	48	45	53	44	49	
51	54	48	47	53	49	59	46	47	
48	51	48	53	54	48	51	59	43	
44	44	46	50	58	52	57	53	56	
52	51	53	51	48	58	51	51	47	
60	40	52	43	48	52	55	43	60	
50	51	51	52	55	50	53	55	48	
52	52	54	49	46	57	48	46	55	
51	49	45	53	51	56	53	52	54	
50	46	45	47	53	50	48	49	47	
51	52	44	46	53	44	47	46	42	

49	50	49	54	52	48	53	51	50	
46	47	48	52	52	48	49	48	55	
52	46	51	50	60	50	47	56	52	
46	52	54	45	45	60	56	50	50	
51	50	54	47	47	45	56	51	54	
52	48	50	49	50	51	44	49	52	
40	48	44	48	47	49	60	47	47	
48	53	55	51	48	52	51	51	41	
49	58	53	47	58	50	53	47	52	
52	46	49	47	51	48	49	44	47	
51	50	53	50	52	52	49	54	43	
49	48	52	48	60	54	49	45	50	
59	53	49	45	46	45	50	42	51	
46	53	51	54	60	54	50	57	54	
49	44	47	48	48	54	51	50	58	
44	51	56	53	52	47	45	48	51	
48	47	44	48	48	51	52	52	53	
49	56	45	51	51	54	53	46	48	
53	52	48	47	45	53	51	48	50	
51	51	52	50	47	44	48	50	49	
56	51	55	53	52	53	49	49	48	
53	48	51	53	49	44	51	55	48	
54	53	54	54	44	48	49	49	51	
42	57	44	50	43	59	51	45	55	

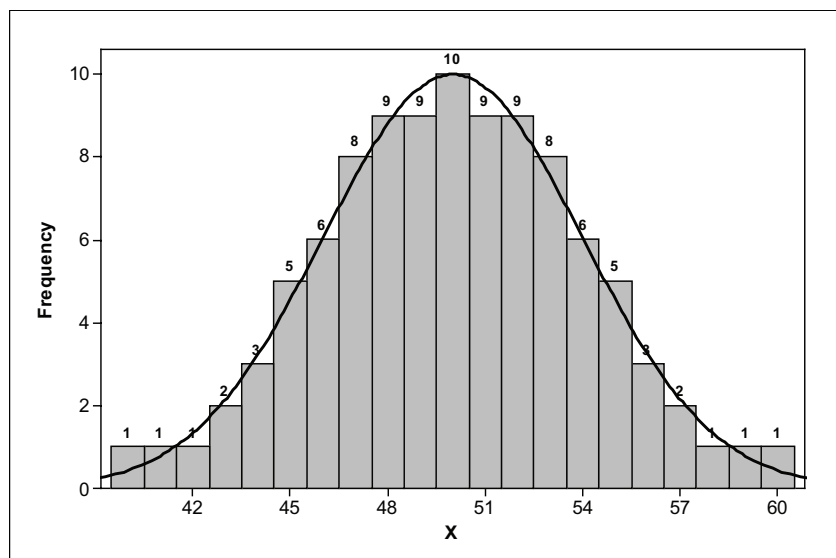
Table T22.3. Sample Data

THE VIDEO SOLUTIONS

1. Parameters describe an entire population and are generally unknown. Statistics are computed from samples.
2. No, it might be too expensive to inspect them all. In addition, it is too costly to wait until the end to examine the finished product. If somewhere in the process things are out of control, it doesn't make sense to put in additional money to finish a defective product.
3. No, there will be variability in the mean scores. Sometimes the mean score will be above 100 and sometimes below 100.
4. The distribution of \bar{x} will be normal with mean 100 and standard deviation $4/\sqrt{5}$.
5. The sample mean is less variable than individual observations. That's because when averaging, high observations will balance out low observations, which makes the mean less variable.
6. The sampling distribution of the sample mean is approximately normally distributed if the sample size is sufficiently large.

ACTIVITY SOLUTIONS

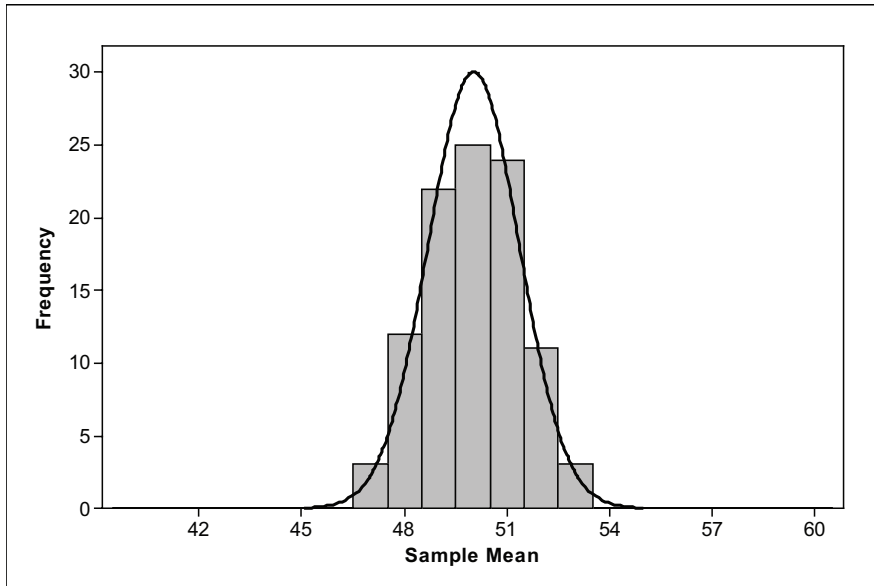
1. The shape is roughly symmetric and mound-shaped. Except for the fact that the population consists of whole numbers so that the distribution can't be normal, it is approximately normally distributed.



2. a. Sample answer (based on sample data):

50.44 51.33 49.56 49.22 50.22 50.22 48.67 51.44 51.00 48.33
48.11 51.78 49.11 51.33 49.78 47.89 52.22 49.00 50.56 51.22
49.44 50.56 48.67 51.22 49.78 49.89 50.56 48.44 50.67 48.11
51.89 48.22 50.67 49.11 48.56 52.11 49.22 50.44 49.89 52.89
52.44 49.33 48.22 50.11 50.89 50.33 50.33 51.67 50.56 50.44
47.22 48.67 49.22 48.78 50.56 52.89 50.44 49.44 50.78 49.11
47.78 48.33 49.44 49.22 50.11 47.00 50.44 50.56 51.11 51.33
50.33 51.67 51.00 51.56 48.33 47.22 50.67 49.44 51.56 50.89
50.56 49.44 47.78 50.00 51.89 48.11 50.44 50.56 48.89 53.22
49.89 49.67 49.22 50.33 49.67 49.11 51.78 50.22 50.67 49.56

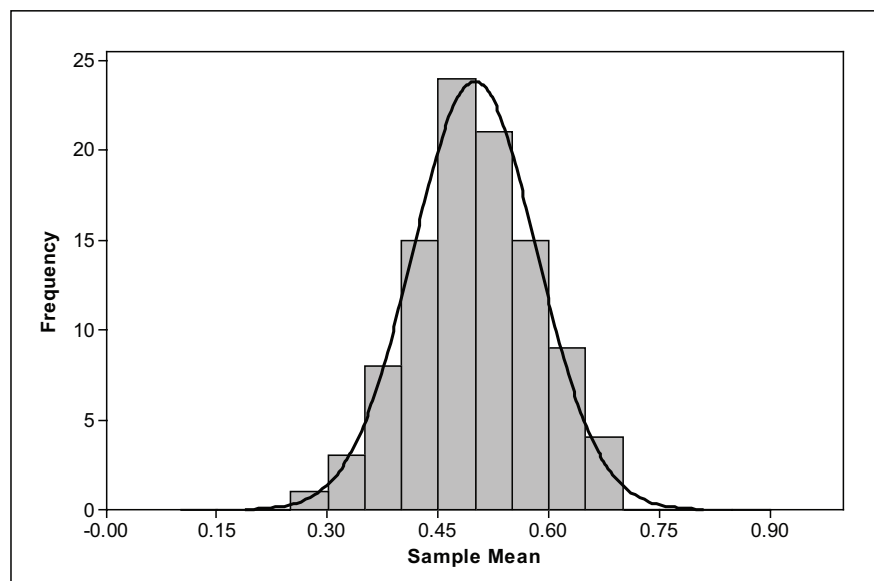
b.



Both the population distribution and sampling distribution of \bar{x} are fairly symmetric and mound-shaped, and appear approximately normal in shape. Both are centered at around 50. However, the sampling distribution is more concentrated about its center, and not as spread out as the population distribution.

Extension

3. c. The distribution will be roughly symmetric and mound-shaped – or approximately normally distributed – similar to the one that follows.



The histogram should be centered close to 0.5, which is the center of the population distribution curve shown in Figure 22.10. However, the sampling distribution will be less spread out than the population distribution, which is evenly spread between 0 and 1. Instead it will be concentrated around 0.5 and trail off on either side of 0.5.

EXERCISE SOLUTIONS

1. a. The mean of the three measurements has standard deviation

$$\frac{\sigma}{\sqrt{n}} = \frac{0.08}{\sqrt{3}} \approx 0.046$$

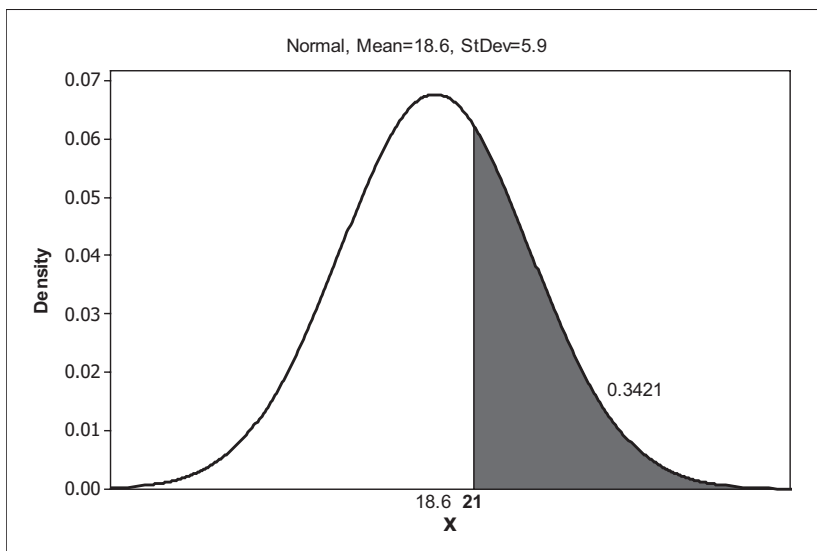
b. The mean of three measurements is less variable than a single measurement. That is why it is good practice to repeat a laboratory measurement several times independently and report the average.

2. This exercise contrasts the distribution of one score with that of the average of 55 scores.

a. We need to calculate the probability that a normal random variable x has value 21 or greater. We first convert to z -scores so that we can use standard normal tables:

$$P(x \geq 21) = P\left(\frac{x - 18.6}{5.9} \geq \frac{21 - 18.6}{5.9}\right) = P(z \geq 0.41) = 1 - 0.6591 = 0.3409$$

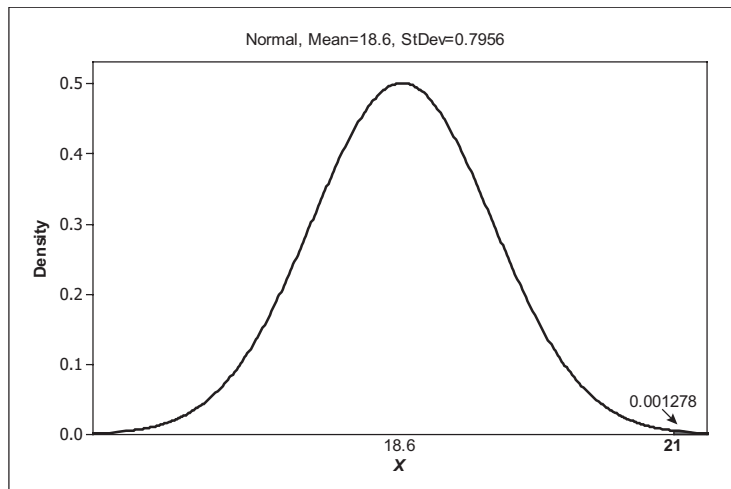
Note: We could also use software such as Minitab or a graphing calculator to compute the probability directly, as shown below. This gives a more accurate answer.



b. The sampling distribution of \bar{x} for samples of size 25 is normally distributed with mean 18.6 and standard deviation $5.9/\sqrt{55} \approx 0.7956$. Below are the calculations for computing the probability using a standard normal table:

$$P(\bar{x} \geq 21) = P\left(\frac{\bar{x} - 18.6}{0.7956} \geq \frac{21 - 18.6}{0.7956}\right) = P(z \geq 3.02) = 1 - 0.9987 = 0.0013$$

We can calculate this probability directly using software:



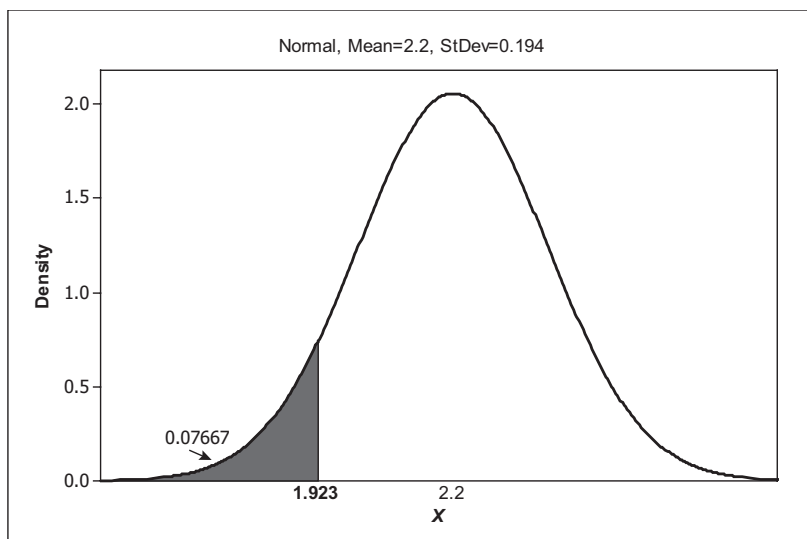
The mean of many scores is less likely to be far away from the population mean than is a single score.

3. a. The distribution is approximately normal with mean 2.2 and standard deviation $1.4/\sqrt{52} \approx 0.194$.

b. Again, transforming in order to use the standard normal table gives:

$$P(\bar{x} < 2) = P\left(\frac{\bar{x} - 2.2}{0.194} < \frac{2 - 2.2}{0.194}\right) = P(z < -1.03) = 0.1515$$

c. To say that the total is less than 100 is exactly the same as saying that the mean per week is less than $100/52$ or 1.923. Now, we can proceed to find $P(\bar{x} < 1.923)$, which we find directly using software:



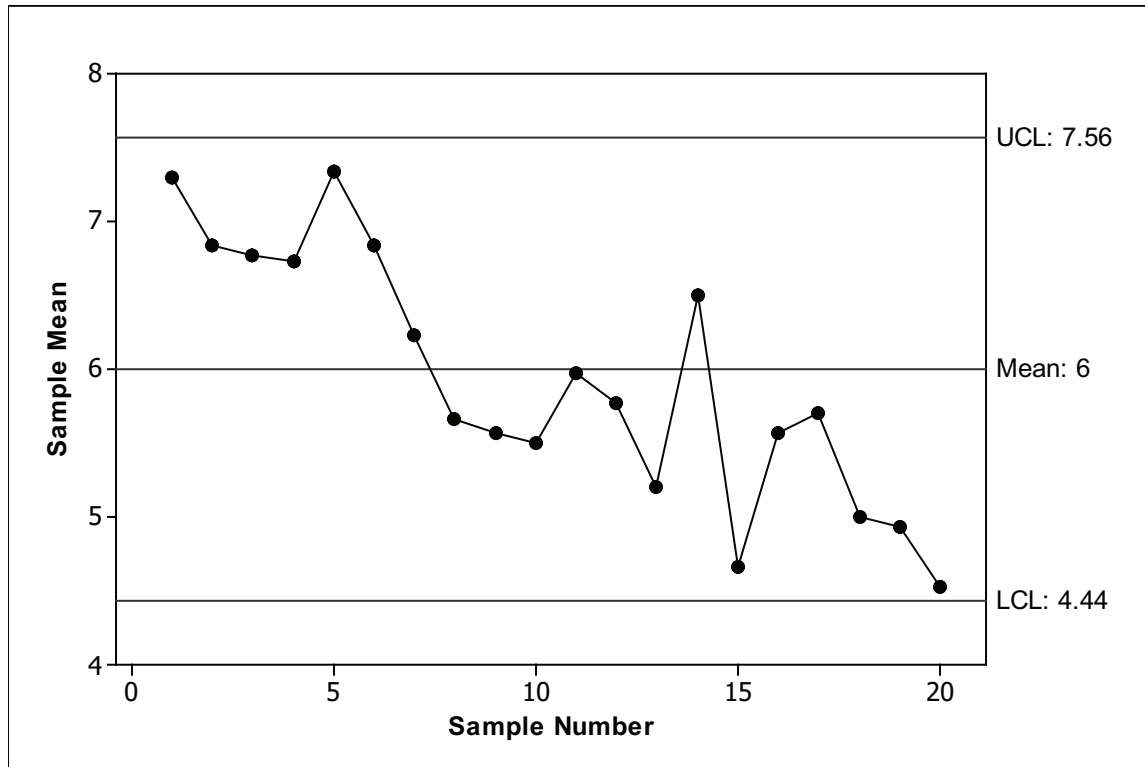
4. a. The lower and upper control limits are $\mu \pm 3\sigma/\sqrt{n}$. In this case, the limits are $6 \pm (3)(0.9/\sqrt{3})$ or LCL ≈ 4.44 and UCL ≈ 7.56 (rounded to two decimals).

b. Samples collected over a 24-hour time period appear in Table 22.3. The sample means appear below.

Sample	pH level			Sample mean
1	5.8	6.2	6.0	6.0
2	6.4	6.9	5.3	6.2
3	5.8	5.2	5.5	5.5
4	5.7	6.4	5.0	5.7
5	6.5	5.7	6.7	6.3
6	5.2	5.2	5.8	5.4
7	5.1	5.2	5.6	5.3
8	5.8	6.0	6.2	6.0
9	4.9	5.7	5.6	5.4
10	6.4	6.3	4.4	5.7
11	6.9	5.2	6.2	6.1
12	7.2	6.2	6.7	6.7
13	6.9	7.4	6.1	6.8
14	5.3	6.8	6.2	6.1
15	6.5	6.6	4.9	6.0
16	6.4	6.1	7.0	6.5
17	6.5	6.7	5.4	6.2
18	6.9	6.8	6.7	6.8
19	6.2	7.1	4.7	6.0
20	5.5	6.7	6.7	6.3

21	6.6	5.2	6.8	6.2
22	6.4	6.0	5.9	6.1
23	6.4	4.6	6.7	5.9
24	7.0	6.3	7.4	6.9

c.



d. No, none of the sample means is above the upper control limit or below the lower control limit.

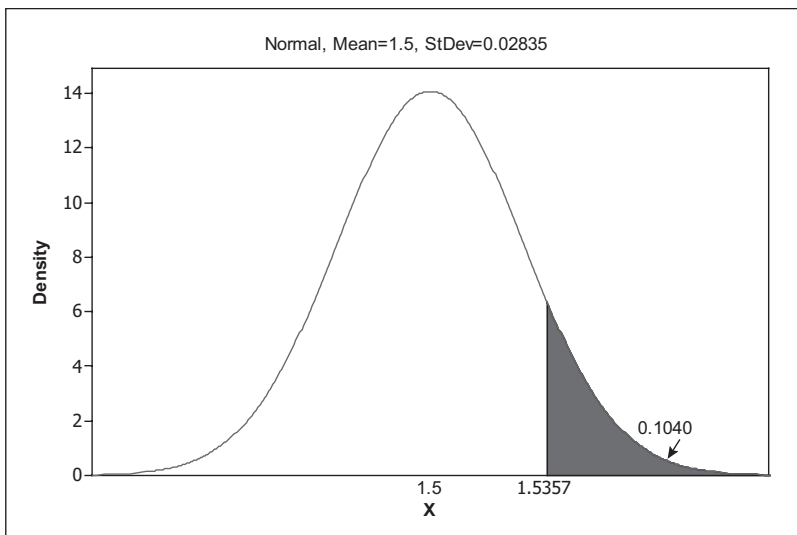
e. Although the sample means lie within the control limits, the process appears to be changing. Initially, the sample means were fairly close to the upper control limit. Over time, the sample means tended to decrease. If this pattern continues, the sample means will begin to fall below the lower control limit. So, this process does not appear to be stable.

REVIEW QUESTIONS SOLUTIONS

1. a. Using the 68-95-99.7 rule, 95% of the caps will have diameters within two standard deviations of the mean – hence, between 0.497 and 0.503 inches. Thus, around 5% of the bottles will have diameters outside of the chemical manufacturer's specification limits.
- b. \bar{x} will have a normal distribution with mean 0.500 inch and standard deviation $0.0015/\sqrt{9} = 0.0005$ inch.
- c. The endpoints of the acceptance interval can be written as 0.500 ± 0.001 , which is equivalent to $0.005 \pm 2(0.0005)$. Hence, 95% of the samples will have means within this interval. The production process will be stopped 5% of the time (or a proportion of 0.05).
2. a. No. The number of people in a car must be a whole number. A normal random variable can take any value, not just whole number values. (Normal distributions are *continuous* distributions.)
- b. The sample mean \bar{x} has an approximately normal distribution with mean 1.5 and standard deviation $0.75/\sqrt{700} \approx 0.02835$
- c. The total in 700 cars exceeds 1075 people exactly when the mean \bar{x} exceeds $1075/700 \approx 1.5357$ persons per car. So, the probability we want is:

$$P(\bar{x} > 1.5357) = P\left(\frac{\bar{x} - 1.5}{0.02835} > \frac{1.5357 - 1.5}{0.02835}\right) = P(z > 1.26) = 1 - 0.8962 = 0.1038$$

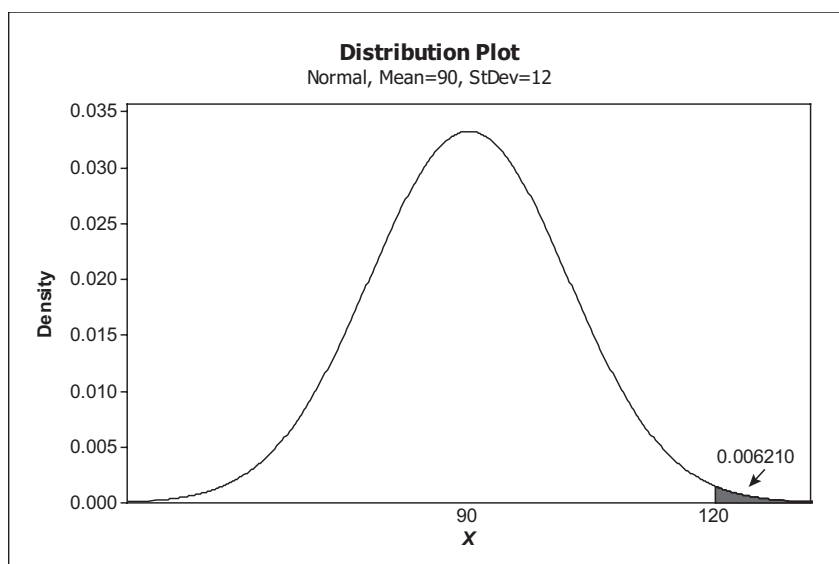
We can compute this directly using software (see chart next page):



3. a. $\mu_{\bar{x}} = 90$ seconds; $\sigma_{\bar{x}} = 120/\sqrt{10} \approx 37.9$ seconds. The shape will be less skewed to the right than the original distribution. Given that the sample size is relatively small, you can't say much more than that.

b. $\mu_{\bar{x}} = 90$ seconds; $\sigma_{\bar{x}} = 120/\sqrt{100} = 12$ seconds. Since the sample size is large, by the Central Limit Theorem we can say that the shape of the distribution will be approximately normal.

c. First, we convert 2 minutes into 120 seconds. We need to calculate $P(\bar{x} > 120)$, which we determine using software, which gives $P(\bar{x} > 120) \approx 0.0062$ as shown below.



Unit 23: Control Charts



PREREQUISITES

Unit 22, Sampling Distributions, is a prerequisite for this unit. Students need to have an understanding of the sampling distribution of the sample mean. Students should be familiar with normal distributions and the 68-95-99.7% Rule (Unit 7: Normal Curves, and Unit 8: Normal Calculations). They should know how to calculate sample means (Unit 4: Measures of Center).

ADDITIONAL TOPIC COVERAGE

Additional coverage of this topic can be found in *The Basic Practice of Statistics*, Chapter 27, Statistical Process Control.

ACTIVITY DESCRIPTION

This activity should be used at the end of the unit and could serve as an assessment of \bar{x} charts. For this activity students will use the Control Chart tool from the Interactive Tools menu. Students can either work individually or in pairs.

MATERIALS

Access to the Control Chart tool. Graph paper (optional).

For the Control Chart tool, students select a mean and standard deviation for the process (from when the process is in control), and then decide on a sample size. After students have determined and entered correct values for the upper and lower control limits, the Control Chart tool will draw the reference lines on the control chart. (Remind students to enter the values for the upper and lower control limits to four decimals.) At that point, students can use the Control Chart tool to generate sample data, compute the sample mean, and then plot the mean

against the sample number. After each sample mean has been plotted, students must decide either that the process is in control and thus should be allowed to continue or that the process is out of control and should be stopped.

In order to verify their work, students need to either make sketches of their control charts or copy screenshots into a Word document. If you want students to sketch their control charts by hand, they will need graph paper. In hand-drawn sketches, students should aim for capturing the approximate pattern of the control chart and not strive for accuracy. The better approach would be to have students do a screen capture and then paste it into a Word document.

Here's how to accomplish this on a PC:

1. Hold the Alt key while pressing the Print-Screen key.
2. In an open Word document, hold the Ctrl key while pressing V to paste the screen capture into their document.

On a Mac:

1. Press "Command" + "Shift" + "3" to capture the entire screen.
2. The file will be saved on your desktop.

THE VIDEO SOLUTIONS

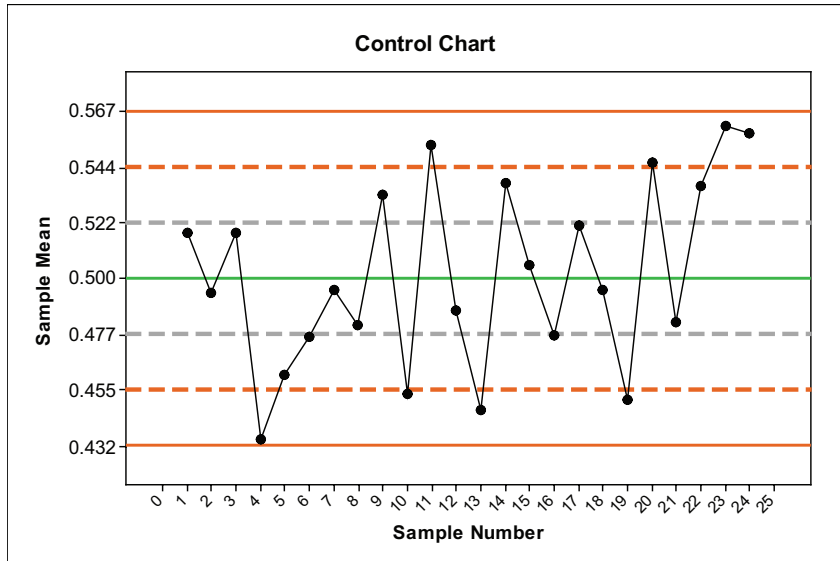
1. He was among the first to champion the idea of statistical process management. He helped Japanese industry rebuild after World War II.
2. A process is a chain of steps that turns inputs into outputs. For example, start with raw foods, prepare them for cooking, combine the prepared ingredients, cook them, serve them on a plate – the served food is the output.
3. When a process is running smoothly, with its variables staying within an expected range.
4. They had a 2:00 a.m. target to get all specimens logged in and ready for processing. They were rarely meeting this target.
5. The control limits were set three standard deviations above and below the center line.
6. Quest remodeled the entire department dividing it into self-contained pods and changing staffing. These changes brought the mean finish time closer to the 2:00 a.m. target.

UNIT ACTIVITY SOLUTIONS

1. Note: Students' control charts will vary even if they make the same choices for mean, standard deviation, and sample size.

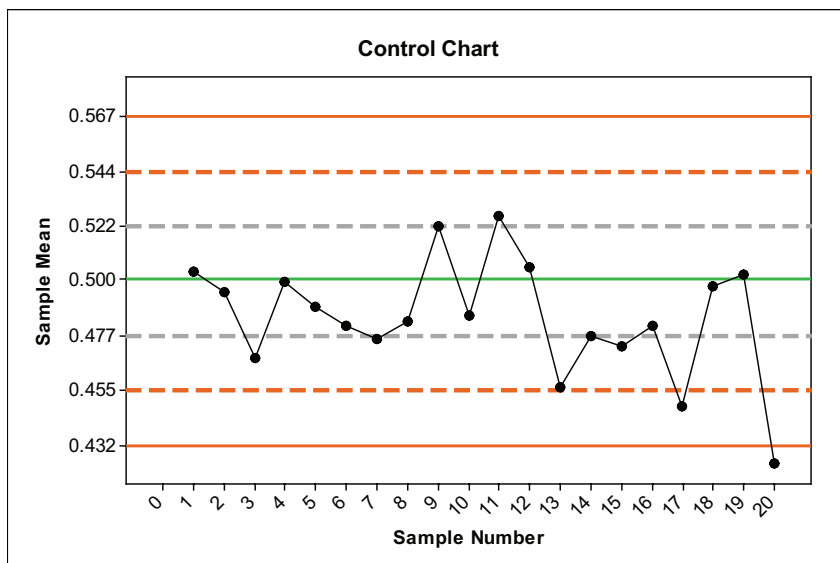
Sample answer: We selected $\mu = 0.5$, $\sigma = 0.05$, and $n = 5$.

Control limits: $0.5 \pm 3(.05/\sqrt{5})$; LCL = 0.4329 and UCL = 0.5671.



We stopped the process at sample 24. Samples 22, 23, and 24 had two of three consecutive points above the $2\sigma/\sqrt{n}$ limit of 0.544; these points were on the same side of the center line. So, using Rule 2, we made the decision to stop the process.

2. Sample answer: See solutions to question 1 for the settings and control limits.

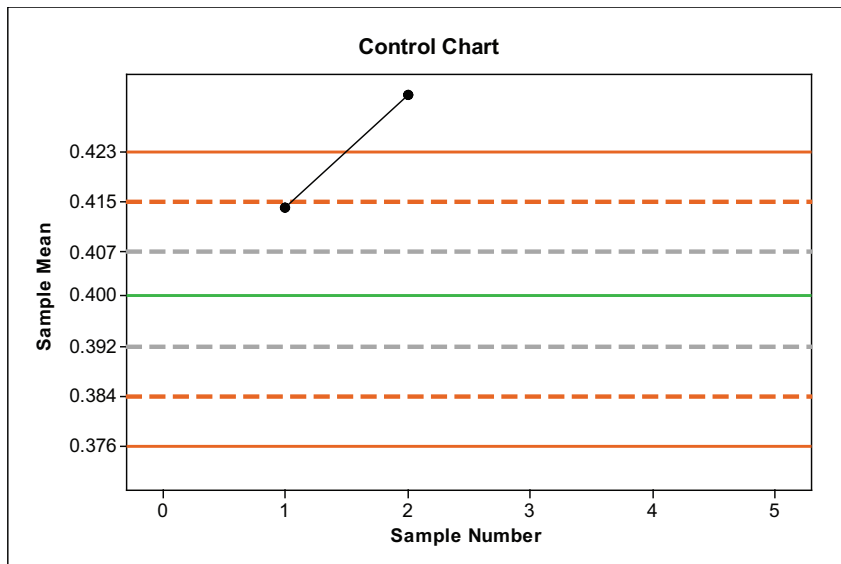


We stopped the process at sample 20. Sample 20 had a sample mean that fell below the lower control limit. So, using Rule 1, we made the decision to stop the process.

3. a. We selected $\mu = 0.4$, $\sigma = 0.025$, and $n = 10$.

Control limits: $0.4 \pm 3(0.025/\sqrt{10})$; LCL = 0.376 and UCL = 0.424.

b.



c. Feedback: We had made the right decision and deserved a bonus.

d. The second sample had a sample mean that was larger than the upper control limit. Based on Rule 1, we made the decision to stop the process.

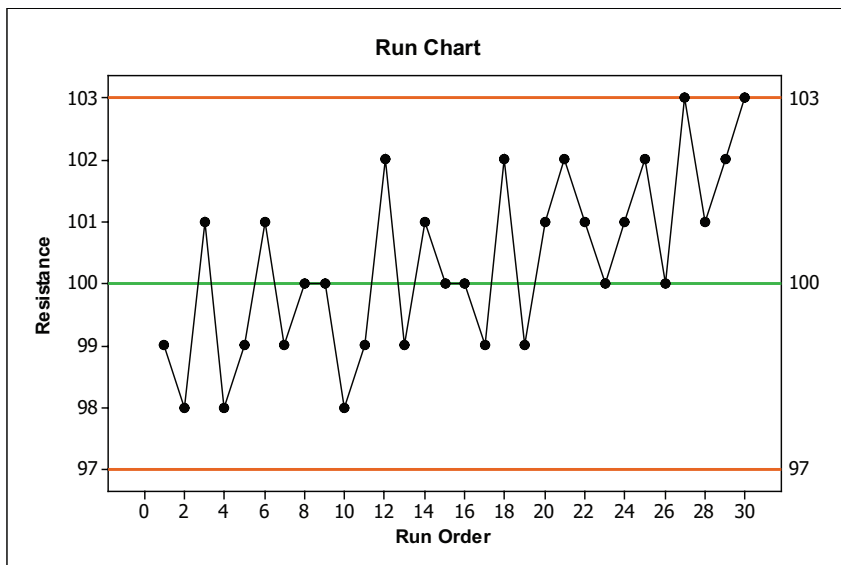
4. See sample answers to questions 1 - 3.

EXERCISE SOLUTIONS

1. a. See run chart for part (c) and refer to the portion of the plot associated with run orders 1 – 15.

b. Sample answer: There is no evidence that the process is out of control. The dots appear randomly scattered above and below the reference line of 100 ohms. No dots are outside of the tolerance interval.

c.



d. Sample answer: Although no dots are outside the tolerance limits, the values for resistance appear to be increasing over run order. The dots associated with run orders 20 to 30 all lie on or above the target value of 100 ohms. In the last four data values, two of the four values were at the upper tolerance level. Management needs to investigate the cause of the changes in the distribution of resistance.

2. a. Using the 68-95-99.7 rule, 95% of the caps will have diameters within two standard deviations of the mean – within 0.497 and 0.503 inch. Thus, around 5% of the bottles will have diameters outside of the chemical manufacturer's specification limits.

b. \bar{x} will have a normal distribution with mean 0.500 inch and standard deviation $0.0015/3 = 0.0005$ inch.

c. The endpoints of the acceptance interval can be written as 0.500 ± 0.001 , which is equivalent to $0.005 \pm 2(0.0005)$. Hence, 95% of the samples will have means within this interval. The production process will be stopped 5% of the time (or a proportion of 0.05).

3. a. The process is not in control. Rule 3 applies. Look at the dots corresponding to Samples 3, 4, 5, 6, and 7. The means of the last four of these samples fall below the σ/\sqrt{n} limit and are on the same side of the center line.

b. The process is in control. None of the decision rules applies.

c. The process is not in control. A run of 9 consecutive points, Samples 7 – 15, appears below the center line. Rule 4 applies after observing the sample mean from Sample 15. At that point the process should be stopped.

4. a. The lower and upper control limits are $\mu \pm 3\sigma/\sqrt{n}$. In this case, the limits are $6 \pm 3(0.9)/\sqrt{3}$ or approximately LCL = 4.44 and UCL = 7.56 (rounded to two decimals).

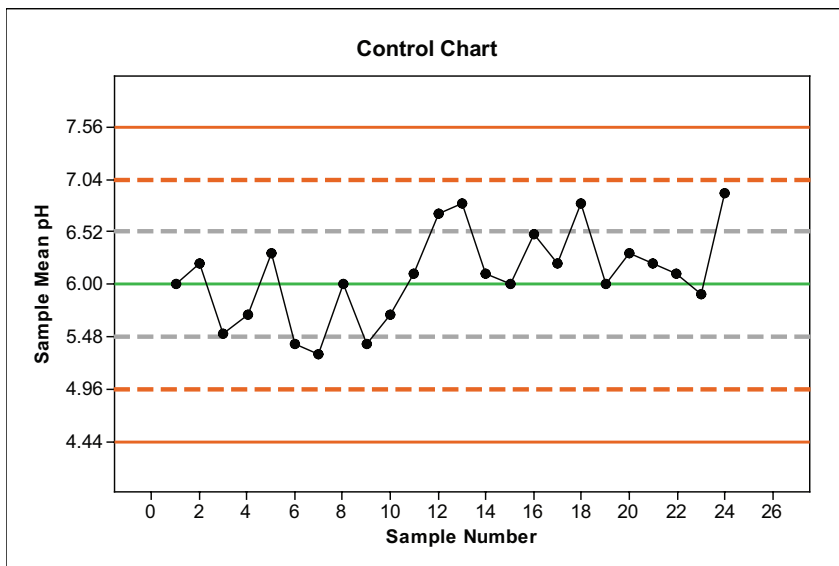
b.

Sample	pH level			Sample mean
1	5.8	6.2	6.0	6.0
2	6.4	6.9	5.3	6.2
3	5.8	5.2	5.5	5.5
4	5.7	6.4	5.0	5.7
5	6.5	5.7	6.7	6.3
6	5.2	5.2	5.8	5.4
7	5.1	5.2	5.6	5.3
8	5.8	6.0	6.2	6.0
9	4.9	5.7	5.6	5.4
10	6.4	6.3	4.4	5.7
11	6.9	5.2	6.2	6.1
12	7.2	6.2	6.7	6.7
13	6.9	7.4	6.1	6.8
14	5.3	6.8	6.2	6.1
15	6.5	6.6	4.9	6.0
16	6.4	6.1	7.0	6.5
17	6.5	6.7	5.4	6.2
18	6.9	6.8	6.7	6.8

(Continued...)

19	6.2	7.1	4.7	6.0
20	5.5	6.7	6.7	6.3
21	6.6	5.2	6.8	6.2
22	6.4	6.0	5.9	6.1
23	6.4	4.6	6.7	5.9
24	7.0	6.3	7.4	6.9

C.

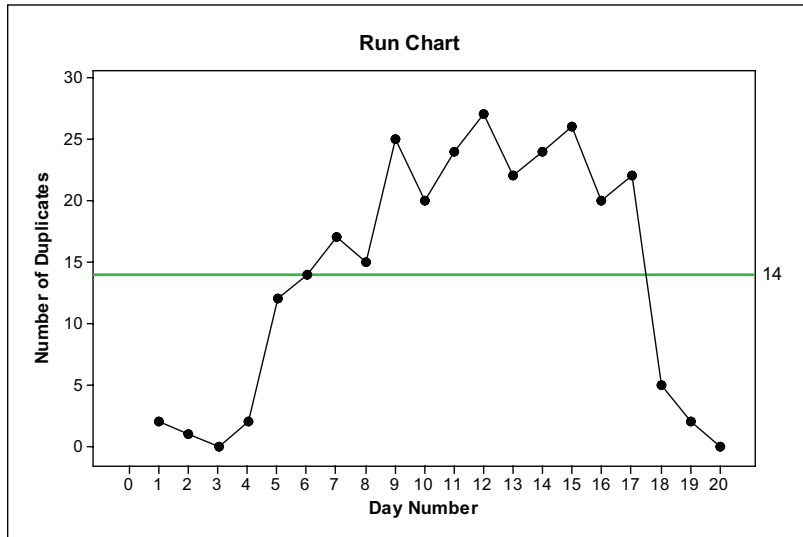


d. The process is in control. None of the decision rules applies.

REVIEW QUESTIONS SOLUTIONS

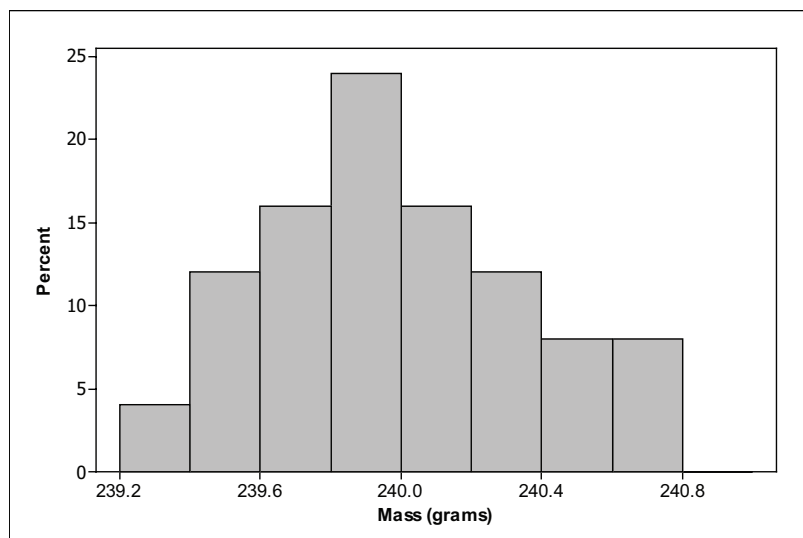
1. a. 14 duplicate e-mails.

b.

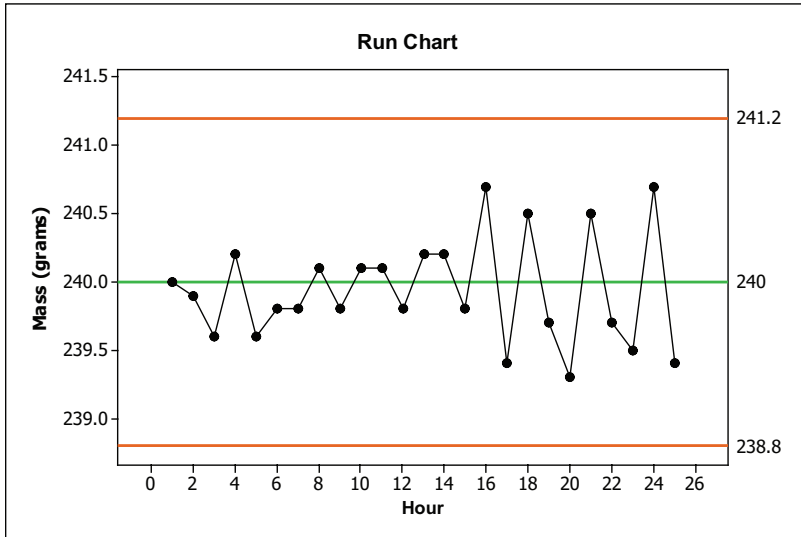


c. Sample answer: Yes – the data points corresponding to Day 7 – Day 17 all lie above the center line. So, this most likely is an example of special cause variation. Perhaps there may be a problem with the spam filter or someone has the manager duplicated on a group list.

2. a. The histogram below gives no evidence that the process is not in control. The histogram appears centered around 240.0 grams. All the data is contained in the interval 240.0 grams \pm 1.2 grams (or from 238.8 grams to 241.2 grams).



b. The upper and lower control limits are at 238.8 and 241.2.



c. Sample answer: The variability of the masses about the center line appears to have increased after hour 15. So, that may indicate there is some problem.

3. a. LCL = $0.91 \mu\text{m}$, L2 = $0.94 \mu\text{m}$, L1 = $0.97 \mu\text{m}$, U1 = $1.03 \mu\text{m}$, U2 = $1.06 \mu\text{m}$, UCL = $1.09 \mu\text{m}$

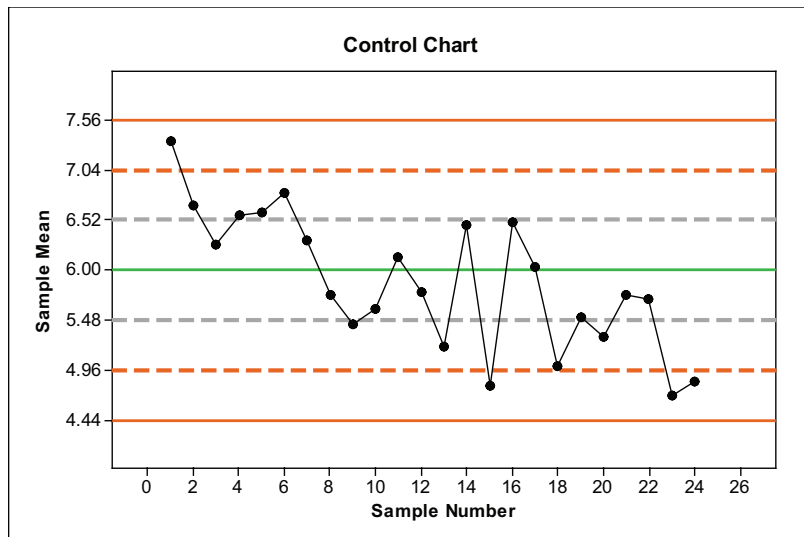
b. The process is not in control. Rule 1 applies. The process should have been shut down after the results from the 9th sample.

4. a. The sample means are as follows, given in the order of the sample numbers:

7.33 6.67 6.27 6.57 6.60 6.80 6.30 5.73 5.43 5.60 6.13 5.77

5.20 6.47 4.80 6.50 6.03 5.00 5.50 5.30 5.73 5.70 4.70 4.83

b.



c. The process is not in control. Rule 3 applies right from the start. Four out of five means from samples 1 – 5 fall above 6.52, and lie on the same side of the center line. The pH level in the liquid is too high at the start of the sampling and appears to be getting lower over time.

Unit 24: Confidence Intervals



PREREQUISITES

Students must have some understanding of the sampling distribution of \bar{x} , which is covered in Unit 22, Sampling Distributions. They must also be familiar with the material on normal distributions from Units 7 – 9.

ADDITIONAL TOPIC COVERAGE

Additional coverage of confidence intervals for the population mean can be found in *The Basic Practice of Statistics*, Chapter 14, Confidence Intervals: The Basics.

ACTIVITY DESCRIPTION

The purpose of this activity is to help students understand the meaning of the confidence level in a 95% confidence interval. In this activity, students will use the simulation data collected in Unit 22's activity. Based on 100 samples of size 9 from a population with known standard deviation, $\sigma = 4$, students will calculate 100 confidence intervals for the mean, μ . Because the data were simulated from a known population, we know that the true value of the population mean is $\mu = 50$. This allows students to find the proportion of confidence intervals that contain the value 50.

MATERIALS

Container of numbered slips of paper from Table 22.2. Data from Unit 22's activity (one copy for each group).

To begin this activity, have students work in groups. Each group should draw a sample of size 9 from the container. Then they should calculate a 95% confidence interval for μ . Groups should share their intervals with the class. Each interval should be classified according to

whether or not it contains the true mean $\mu = 50$.

Students will need the data from 100 size-9 samples simulated for question 2 from Unit 22's activity. As part of that activity, students had already computed the sample means. In question 2 of this activity, students will compute 95% confidence intervals for each of these 100 samples. If the data are distributed in an Excel spreadsheet, then students can enter formulas for the lower and upper endpoints of the interval and very quickly compute all 100 confidence intervals. Otherwise, they should split up the calculations among members of their groups.

Simulation can help students understand the central idea in this unit: A 95% confidence interval is produced by a formula that catches the true population mean 95% of the time *in the long run* when used repeatedly many, many times. More formally, the confidence interval produced from the formula

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

has probability 0.95 of producing an interval that catches the true mean.

The hand simulation uses the same population prepared for the simulation activity in Unit 22. Thus, for random samples of size $n = 9$, the 95% confidence interval for the mean μ is:

$$\bar{x} \pm 1.96 \left(\frac{\sigma}{\sqrt{n}} \right) = \bar{x} \pm 1.96 \left(\frac{4}{\sqrt{9}} \right) \approx \bar{x} \pm 2.61$$

After students have computed the confidence intervals for 100 samples, they should classify each interval by whether or not it traps $\mu = 50$ between its endpoints. Students may be surprised if their success rate differs from 95 out of 100. For example, in the sample solutions the success rate is 96 out of 100 or 96%. This is a good opportunity to discuss what is meant by the confidence level – it is the success rate in the long run, over many, many repeats (far more than 100 repetitions).

THE VIDEO SOLUTIONS

1. Blood pressure readings vary from day to day, and time of day. So, a sample of blood pressure readings is needed to estimate a person's average blood pressure.
2. We need an interval estimate and a level of confidence.
3. (1) Independent observations, (2) data are from a normal distribution or the sample size is large, and (3) the population standard deviation is known.
4. The process used to create the confidence interval is one that gives correct results 95% of the time over the long run.

UNIT ACTIVITY SOLUTIONS

1. a. $\sigma_{\bar{x}} = 4/\sqrt{9} = 4/3$

b. Margin of error = $(1.96)(4/\sqrt{9}) \approx 2.61$

2. a. Sample answer: Sample is 40, 55, 45, 53, 41, 50, 54, 60, 56; $\bar{x} = 50.44$

b. Sample answer: 50.44 ± 2.61 or (47.83,53.05)

c. Sample answer: Yes

3. a. See solution to b.

b. Sample answer based on sample data used in Unit 22's activity. The table that follows contains 100 samples of size 9 drawn from the distribution in Table 22.2. The sample means have been used to calculate 95% confidence intervals for μ . In this simulation 94 of the 100 samples produced intervals that contained $\mu = 50$ between their endpoints. We would expect 95 of the 100 intervals to contain 50. However, the fraction of 95/100 is the long-run results of many, many repeats and not for as few as 100 repeats. So, there is no real discrepancy – we would need to repeat the simulation many more times to get a proportion of Yes outcomes closer to 0.95.

X1	X2	X3	X4	X5	X6	X7	X8	X9	Sample Mean	Lower Endpoint	Upper Endpoint	Contains Mu?
40	55	45	53	41	50	54	60	56	50.444	47.831	53.057	Yes
53	55	50	53	51	43	54	53	50	51.333	48.720	53.946	Yes
57	48	46	50	46	48	53	51	47	49.556	46.943	52.169	Yes
44	50	55	47	44	44	47	59	53	49.222	46.609	51.835	Yes
48	59	51	49	51	47	47	45	55	50.222	47.609	52.835	Yes
58	41	46	56	46	55	49	50	51	50.222	47.609	52.835	Yes
51	46	51	40	49	47	48	60	46	48.667	46.054	51.280	Yes
50	47	51	52	51	51	60	53	48	51.444	48.831	54.057	Yes
59	54	49	49	49	41	56	46	56	51.000	48.387	53.613	Yes
52	44	46	57	43	46	50	50	47	48.333	45.720	50.946	Yes
51	45	53	54	46	48	49	46	41	48.111	45.498	50.724	Yes
52	48	58	57	56	44	52	50	49	51.778	49.165	54.391	Yes
47	50	53	58	47	44	48	47	48	49.111	46.498	51.724	Yes
52	54	50	49	53	43	53	59	49	51.333	48.720	53.946	Yes
54	44	51	43	46	52	53	58	47	49.778	47.165	52.391	Yes
51	43	49	51	46	49	48	46	48	47.889	45.276	50.502	Yes

(Continued...)

X1	X2	X3	X4	X5	X6	X7	X8	X9	Sample Mean	Lower Endpoint	Upper Endpoint	Contains μ ?
51	58	48	45	51	50	59	53	55	52.222	49.609	54.835	Yes
45	51	45	46	52	48	52	48	54	49.000	46.387	51.613	Yes
52	56	51	52	53	53	43	47	48	50.556	47.943	53.169	Yes
53	47	52	50	55	54	46	49	55	51.222	48.609	53.835	Yes
50	40	57	51	52	52	48	46	49	49.444	46.831	52.057	Yes
51	46	47	47	52	55	46	51	60	50.556	47.943	53.169	Yes
56	41	51	52	48	44	51	48	47	48.667	46.054	51.280	Yes
50	50	49	56	60	42	52	45	57	51.222	48.609	53.835	Yes
48	48	54	49	53	49	51	49	47	49.778	47.165	52.391	Yes
48	50	51	52	51	50	47	50	50	49.889	47.276	52.502	Yes
54	53	55	46	48	48	52	52	47	50.556	47.943	53.169	Yes
45	51	53	52	46	40	49	57	43	48.444	45.831	51.057	Yes
49	56	53	48	50	54	44	52	50	50.667	48.054	53.280	Yes
45	52	44	57	48	41	49	51	46	48.111	45.498	50.724	Yes
50	53	52	49	50	53	53	52	55	51.889	49.276	54.502	Yes
49	52	44	40	43	50	51	50	55	48.222	45.609	50.835	Yes
47	51	50	56	47	54	55	50	46	50.667	48.054	53.280	Yes
52	51	51	53	42	47	47	54	45	49.111	46.498	51.724	Yes
53	49	40	49	49	49	48	52	48	48.556	45.943	51.169	Yes
54	48	57	54	48	55	46	52	55	52.111	49.498	54.724	Yes
50	47	55	46	49	52	45	53	46	49.222	46.609	51.835	Yes
48	48	50	54	54	45	49	51	55	50.444	47.831	53.057	Yes
52	49	49	46	53	48	43	55	54	49.889	47.276	52.502	Yes
46	51	51	53	56	53	60	58	48	52.889	50.276	55.502	No
56	56	52	45	46	53	51	59	54	52.444	49.831	55.057	Yes
46	55	50	53	53	54	43	47	43	49.333	46.720	51.946	Yes
54	50	53	48	54	46	47	41	41	48.222	45.609	50.835	Yes
50	46	52	45	50	57	50	51	50	50.111	47.498	52.724	Yes
55	50	58	45	49	51	47	57	46	50.889	48.276	53.502	Yes
47	52	52	54	49	48	51	47	53	50.333	47.720	52.946	Yes
46	49	53	54	53	47	52	55	44	50.333	47.720	52.946	Yes
56	43	47	54	50	50	59	54	52	51.667	49.054	54.280	Yes
41	55	51	51	52	53	48	49	55	50.556	47.943	53.169	Yes
51	54	47	46	50	52	52	50	52	50.444	47.831	53.057	Yes
59	47	42	44	50	44	41	45	53	47.222	44.609	49.835	No
55	46	48	44	48	56	48	47	46	48.667	46.054	51.280	Yes
52	54	43	54	43	54	51	42	50	49.222	46.609	51.835	Yes
46	52	46	50	40	51	49	52	53	48.778	46.165	51.391	Yes
53	50	46	48	53	47	52	52	54	50.556	47.943	53.169	Yes
49	56	50	46	58	47	58	55	57	52.889	50.276	55.502	No

(Continued...)

X1	X2	X3	X4	X5	X6	X7	X8	X9	Sample Mean	Lower Endpoint	Upper Endpoint	Contains μ ?
46	60	45	48	49	48	51	53	54	50.444	47.831	53.057	Yes
46	51	46	54	48	53	49	51	47	49.444	46.831	52.057	Yes
56	47	48	52	48	49	52	55	50	50.778	48.165	53.391	Yes
50	46	57	45	46	46	55	52	45	49.111	46.498	51.724	Yes
52	41	52	46	51	51	48	49	40	47.778	45.165	50.391	Yes
49	46	44	51	58	49	41	48	49	48.333	45.720	50.946	Yes
47	56	52	43	47	50	52	50	48	49.444	46.831	52.057	Yes
45	47	52	49	49	45	54	56	46	49.222	46.609	51.835	Yes
53	50	45	49	51	52	49	51	51	50.111	47.498	52.724	Yes
49	46	44	45	48	45	53	44	49	47.000	44.387	49.613	No
51	54	48	47	53	49	59	46	47	50.444	47.831	53.057	Yes
48	51	48	53	54	48	51	59	43	50.556	47.943	53.169	Yes
44	44	46	50	58	52	57	53	56	51.111	48.498	53.724	Yes
52	51	53	51	48	58	51	51	47	51.333	48.720	53.946	Yes
60	40	52	43	48	52	55	43	60	50.333	47.720	52.946	Yes
50	51	51	52	55	50	53	55	48	51.667	49.054	54.280	Yes
52	52	54	49	46	57	48	46	55	51.000	48.387	53.613	Yes
51	49	45	53	51	56	53	52	54	51.556	48.943	54.169	Yes
50	46	45	47	53	50	48	49	47	48.333	45.720	50.946	Yes
51	52	44	46	53	44	47	46	42	47.222	44.609	49.835	No
49	50	49	54	52	48	53	51	50	50.667	48.054	53.280	Yes
46	47	48	52	52	48	49	48	55	49.444	46.831	52.057	Yes
52	46	51	50	60	50	47	56	52	51.556	48.943	54.169	Yes
46	52	54	45	45	60	56	50	50	50.889	48.276	53.502	Yes
51	50	54	47	47	45	56	51	54	50.556	47.943	53.169	Yes
52	48	50	49	50	51	44	49	52	49.444	46.831	52.057	Yes
40	48	44	48	47	49	60	47	47	47.778	45.165	50.391	Yes
48	53	55	51	48	52	51	51	41	50.000	47.387	52.613	Yes
49	58	53	47	58	50	53	47	52	51.889	49.276	54.502	Yes
52	46	49	47	51	48	49	44	47	48.111	45.498	50.724	Yes
51	50	53	50	52	52	49	54	43	50.444	47.831	53.057	Yes
49	48	52	48	60	54	49	45	50	50.556	47.943	53.169	Yes
59	53	49	45	46	45	50	42	51	48.889	46.276	51.502	Yes
46	53	51	54	60	54	50	57	54	53.222	50.609	55.835	No
49	44	47	48	48	54	51	50	58	49.889	47.276	52.502	Yes
44	51	56	53	52	47	45	48	51	49.667	47.054	52.280	Yes
48	47	44	48	48	51	52	52	53	49.222	46.609	51.835	Yes
49	56	45	51	51	54	53	46	48	50.333	47.720	52.946	Yes
53	52	48	47	45	53	51	48	50	49.667	47.054	52.280	Yes
51	51	52	50	47	44	48	50	49	49.111	46.498	51.724	Yes

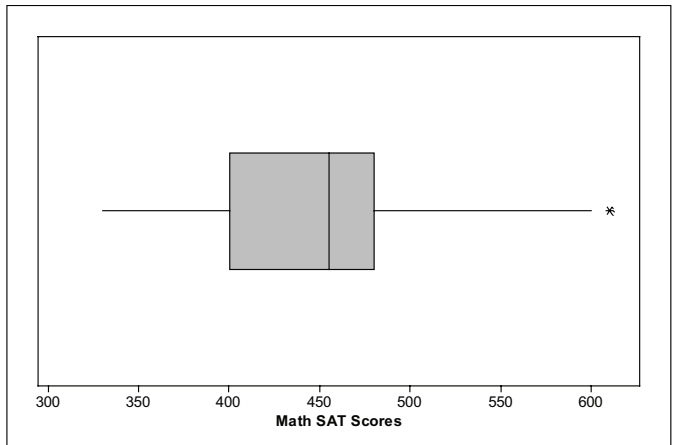
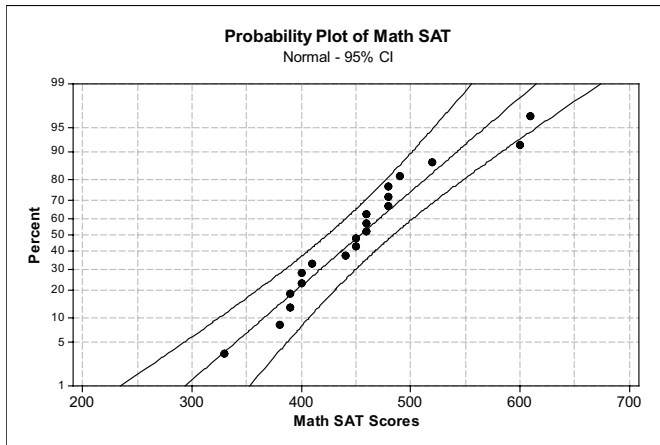
(Continued...)

X1	X2	X3	X4	X5	X6	X7	X8	X9	Sample Mean	Lower Endpoint	Upper Endpoint	Contains μ ?
56	51	55	53	52	53	49	49	48	51.778	49.165	54.391	Yes
53	48	51	53	49	44	51	55	48	50.222	47.609	52.835	Yes
54	53	54	54	44	48	49	49	51	50.667	48.054	53.280	Yes
42	57	44	50	43	59	51	45	55	49.556	46.943	52.169	Yes

EXERCISE SOLUTIONS

1. a. $\sigma_{\bar{x}} = 100/\sqrt{20} \approx 22.36$

b. A normal quantile plot shows most of the dots contained within the 95% bands. A boxplot shows only one outlier, and it does not appear to be extreme. The boxplot is not completely symmetric, but the most severe asymmetry is not out in the whiskers. So, the data do not appear to have severe departures from normality. Even with normal data, it is fairly usual to observe a mild outlier.



c. The sample mean is $\bar{x} = 454$. Endpoints of the 95% interval: $454 \pm 1.96(22.36)$, which gives (410.2, 497.8). To get a smaller margin of error, collect a larger sample. As a point of discussion, note that once we are given $\sigma = 100$, we could calculate the margin of error in advance. So, given any desired margin of error, we could determine the sample size needed for that margin of error.

d. For 99% confidence, replace $z^* = 1.96$ from the 95% confidence interval with $z^* = 2.576$. The endpoints of the 99% confidence interval are $454 \pm 2.576(22.36)$, which gives (396.4, 511.6). An interval that covers the true value of μ 99% of the time must be wider than one that covers the true value of μ only 95% of the time.

2. a. The 95% confidence interval is calculated as $252.53 \pm 1.96\left(\frac{17}{\sqrt{30}}\right) \approx 252.53 \pm 6.08$, or (246.45, 258.61).

b. The 95% confidence interval is calculated as $254.50 \pm 1.96\left(\frac{17}{\sqrt{60}}\right) \approx 254.50 \pm 4.30$, or (250.20, 258.80).

c. The margin of error in (a) was $1.96\left(\frac{17}{\sqrt{30}}\right) \approx 6.08$; the margin of error in (b) was $1.96\left(\frac{17}{\sqrt{60}}\right) \approx 4.30$.

The only difference between these two calculations is the square root of the sample size in the denominator of the standard deviation of $\sigma_{\bar{x}}$. Hence, the margin of error associated with the larger sample size is smaller.

d. We need to solve $1.96\left(\frac{17}{\sqrt{n}}\right) = 3.0$ for n . This gives $n = \left(\frac{1.96 \times 17}{3.0}\right)^2 \approx 123.4$.

Hence, you would need at least 124 observations.

3. a. $875 \pm (1.96)(255)/10 \approx 875 \pm 50$; (825, 925)

b. We cannot say that there is a 95% chance that the true value of μ , the mean living area, is within this interval. Either μ is in the interval (in which case, the chance is 100%) or it is not (in which case, the chance is 0%). The 95% refers to the track record of using this method for computing interval estimates – the process works 95% of the time.

4. a. $\bar{x} = 17.11$ and $s = 7.89$

b. $17.11 \pm 1.96\left(\frac{7.89}{\sqrt{50}}\right) \approx 17.11 \pm 2.19$, or (\$14.92, \$19.30)

c. The confidence interval refutes the politician's claim because the value 20 is not in the confidence interval. The mean rate he is reporting is too high.

d. $17.11 \pm 2.576\left(\frac{7.89}{\sqrt{50}}\right) \approx 17.11 \pm 2.87$, or (\$14.24, \$19.98).

The confidence interval still refutes the politician's claims since the value 20 is not contained in the confidence interval. So, the conclusion is still that the politician is inflating the mean hourly rate.

REVIEW QUESTIONS SOLUTIONS

1. a. The 95% confidence interval is:

$$\bar{x} \pm z * \frac{\sigma}{\sqrt{n}} = 71.1 \pm 1.96 \left(\frac{2.7}{\sqrt{96}} \right) = 71.1 \pm 0.54, \text{ or about } 70.56 \text{ inches to } 71.64 \text{ inches.}$$

b. Sample answer: This is typical of the issues met in applying statistics when it is not possible to take a truly random sample. Julie's 96 players probably should not be treated as a random sample. These 96 players are all the players in one league. Some leagues contain all large schools and others contain small schools. Most likely leagues comprised of larger schools will have taller players. So, players in Julie's league may not be representative of all male high school basketball players.

2. a. $\bar{x} = 448.5 \text{ mm}$; $s = 25.29 \text{ mm}$

b. The 90% would be the narrowest confidence interval and the 99% would be the widest. As confidence increases precision decreases, meaning the intervals get wider.

c. 90% confidence interval: $448.5 \pm 1.645 \left(\frac{25.29}{\sqrt{36}} \right)$, or (441.57 mm, 455.43 mm)

95% confidence interval: $448.5 \pm 1.96 \left(\frac{25.29}{\sqrt{36}} \right)$, or (440.24 mm, 456.76 mm)

99% confidence interval: $448.5 \pm 2.576 \left(\frac{25.29}{\sqrt{36}} \right)$, or (437.64 mm, 459.36 mm)

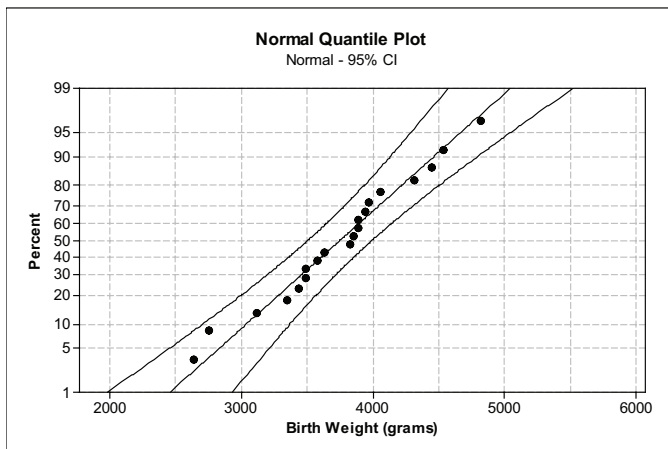
The 99% confidence interval is the widest and the 90% is the narrowest. This confirms the answer to (b).

3. a. Yes all three assumptions are satisfied as outlined below.

(1) Independent observations – Since the sample is a random sample, independence is satisfied.

(2) Normal distribution or n large – Since the sample size $n < 30$, we need to check whether the data follow a normal distribution. Notice that the dots in the normal quantile plot below all

lie within the curved bands. Hence, the normality assumption is reasonably satisfied. (Students could also make a boxplot. The boxplot is roughly symmetric and there is only one outlier – so again, the normality assumption is reasonable.)



(3) The population standard deviation is known – we are given $\sigma = 600$.

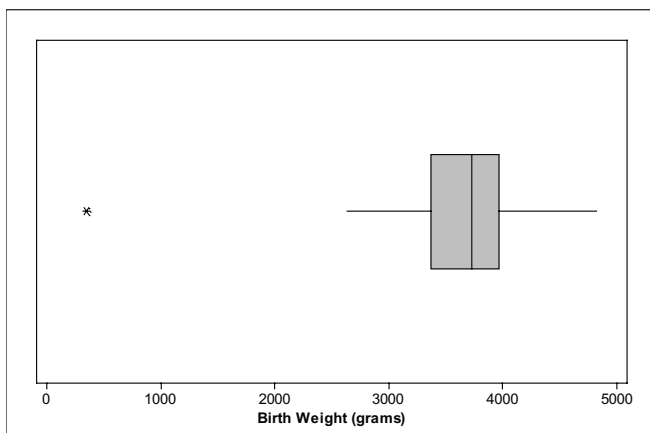
b. The calculations for the 95% confidence interval are:

$$3749 \text{ g} \pm 1.96 \left(\frac{600 \text{ g}}{\sqrt{20}} \right) \approx 3749 \text{ g} \pm 263 \text{ g}, \text{ or } (3486 \text{ g}, 4012 \text{ g})$$

c. All that is needed is to convert the sample mean and standard deviation, 3749 g and 600 g, from grams to ounces by multiplying each by 0.03527 oz/g. This is the same as multiplying the endpoints of the confidence interval by the same conversion factor. The result gives the following 95% confidence interval for μ : (122.95 oz, 141.50 oz).

Sample answer: If we convert the result from pounds to ounces, we get an interval of around 7.7 lb. to 8.8 lb, which seems a normal weight for newborns.

4. a. Yes, the 350 g baby is an outlier – in fact, it is an extreme outlier.



b. The 95% confidence interval based on the modified data is (3301 g, 3827 g). It shifted the confidence interval by 185 g. That shift is larger than half the value of the outlier.

Unit 25: Tests of Significance



PREREQUISITES

In this unit, we use inference about the mean μ of a normal distribution to illustrate the reasoning of significance tests. Hence, students should be familiar with normal distributions (Units 7, 8, and 9). Inference about μ is based on the sample mean \bar{x} . Hence, students should be familiar with facts about the sampling distribution of \bar{x} presented in Unit 22, Sampling Distributions. Unit 24, Confidence Intervals, should be covered before this unit since several of the exercises ask students to compute a confidence interval either before or after performing a significance test.

ADDITIONAL TOPIC COVERAGE

An introduction into significance tests can be found in *The Basic Practice of Statistics*, Chapter 15, Tests of Significance: The Basics.

ACTIVITY DESCRIPTION

In this activity, students will check whether the mean number of chips per cookie in Nabisco's Chips Ahoy chocolate chip cookies has changed from what it was advertised to be in the 1980s. In Unit 27's activity, students will determine whether the mean number of chips in regular Chips Ahoy Chocolate Chip Cookies differs significantly from the mean number of chips in reduced fat Chips Ahoy Chocolate Chip Cookies. So, you may want to collect the data from both types of cookies now.

MATERIALS

One or two bags of Nabisco's Chips Ahoy chocolate chip cookies; paper plates or paper towels. (Include a bag or two of Nabisco's reduced fat Chips Ahoy chocolate chip cookies if you want to also collect the data needed for Unit 27's activity.)

This activity should be done in groups with 2 to 4 students. If you need to have a group of 3, students in the group can trade off being chip counters.

For question 1, students will need to collect data. Before students begin counting the chips in the cookies, hold a discussion to establish rules that students will follow when counting chips. To start the discussion, hand a cookie to one student, and ask him to count the chips but not to reveal his result. Next, the cookie should be passed to two other students who do the same. After all three have counted the chips in the same cookie, they should report on the number of chips they counted. Often the chip counts for the same cookie are very different, which is an indication that the variability due to the counting procedure needs to be controlled. At this point, students should discuss in their groups how they think the counting should be done so that the variability due to the counting process is reduced. Give groups a chance to present their counting plans. After groups have presented their plans, the class should decide on a set of rules. Here is a sample set of rules:

- Count chips that look larger than a half chip and ignore anything smaller.
- Count chips appearing on the top and bottom of the cookie as separate chips. Chips on the side of the cookie only get counted once, even though they might appear both from the top and bottom.
- Two independent counts will be taken on each cookie and the cookie count will be the average of the two independent counts. (This will reduce the variability due to the counting procedure.)
- Give each group a paper plate or paper towel. Place a bunch of cookies on each plate.

Group members will count the number of chips in each cookie – labeling them cookie #1, cookie #2, etc. so that the independent counts get recorded and matched to the same cookie. If students are also collecting data on the reduced fat cookies, make sure they label the cookie type as well. When students have finished counting the number of chips in their bunch of cookies, they should hand in their data so that it can be consolidated into a single data sheet (or spreadsheet). (Once the counting is completed, students can eat the cookies!)

Students will need a copy of the class data for questions 2 – 4. If you decide not to collect the data (a task students really enjoy), use the sample data (see sample answer to question 1), which was collected in two statistics classes (each class got a bag of cookies).

THE VIDEO SOLUTIONS

1. After entering a poem, the program could tell you how many new words there are in the poem that Shakespeare did not use in any of his other writings.
2. The null hypothesis was that Shakespeare was the author of the poem. The alternative hypothesis was that someone else authored the poem.
3. The number of unique words per poem was approximately normally distributed with mean $\mu = 7$ and standard deviation $\sigma = 2.6$.
4. No. Thisted could expect to find a value at least as extreme as 10 unique words about 25% of the time when the poems were Shakespeare's.
5. A small p -value.

UNIT ACTIVITY SOLUTIONS

1. The sample data below are from the non-broken cookies in two bags of Nabisco's Chips Ahoy chocolate chip cookies. Two students counted the chips in each cookie independently and the results were averaged. Sample answers to questions 2 and 3 are based on these data.

18.5 17.0 14.0 14.5 15.0 13.5 16.5 15.5 19.0 16.0
20.5 23.5 17.5 18.5 21.5 18.5 22.5 15.0 16.0 17.5
12.0 14.5 12.5 12.0 20.5 21.5 22.5 24.0 18.5 16.5
22.5 22.0 18.5 18.5 21.5 20.0 17.5 16.5 17.5 17.5
19.5 21.5 24.5 18.0 23.0 20.5 19.5 25.0 19.0 20.0
22.0 19.0 21.5 18.0 14.5 17.0 21.0 10.5 18.0 18.0
20.0 13.5 23.5 16.5 19.5

2. Sample answer: $\bar{x} = 18.462$ and $s = 3.308$

3. a. Let μ be the mean number of chips per cookie in Nabisco Chips Ahoy chocolate chip cookies. The null and alternative hypotheses are:

$$H_0 : \mu = 16$$

$$H_a : \mu \neq 16$$

$$\text{b. } z = \frac{18.462 - 16}{3.308 / \sqrt{65}} \approx 6.00$$

c. The p -value is essentially 0. Hence, we conclude that the mean number of chips in Chips Ahoy chocolate chip cookies has changed since the 1980s.

4. The confidence interval is $18.462 \pm 1.96 \left(\frac{3.308}{\sqrt{65}} \right) = 18.462 \pm 0.804$ or (17.7, 19.3).

Since all the numbers in the interval are larger than 16, it appears that the average number of chips per cookie is greater than 16.

EXERCISE SOLUTIONS

1. a. Let μ be Larry's average miles per gallon when using the new oil. He hopes to show that μ is greater than the 32 mpg he got before switching to the new oil. The hypotheses are:

$$H_0 : \mu = 32$$

$$H_a : \mu > 32$$

The null hypothesis says "no change" and the alternative says that mileage has increased.

b. The question asked is "Do students who get credit by the placement exam differ from the usual level, in either direction?" So the alternative hypothesis is two-sided. Take μ to be the average listening score for all students who get credit by the placement exam.

The hypotheses are:

$$H_0 : \mu = 24$$

$$H_a : \mu \neq 24$$

c. The student hopes to show that the mice take less time when responding to a loud noise. Let μ be the mean time for mice to run the maze after hearing a loud noise.

The hypotheses are:

$$H_0 : \mu = 18$$

$$H_a : \mu < 18$$

2. The null hypothesis states that $\mu = 115$. Because the standard deviation of the population is $\sigma = 30$ the standard deviation of the sample mean for $n = 25$ older students is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{30}{\sqrt{25}} = 6.$$

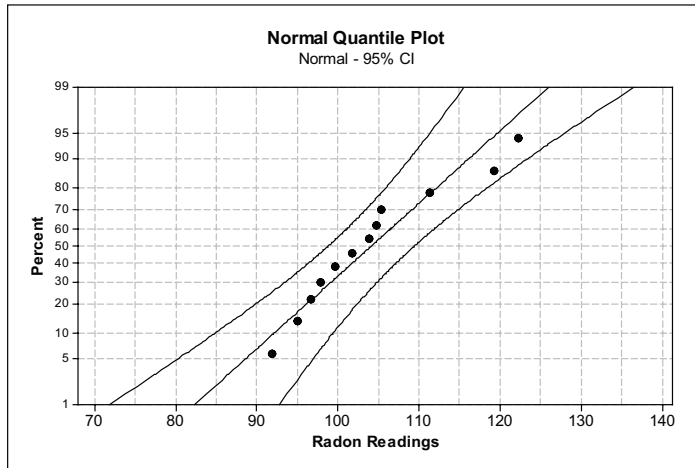
The value of the z-test statistic is $z = \frac{125.2 - 115}{6} = 1.7$.

The alternative is one-sided on the high side. So, the p -value is the area under a standard normal curve to the right of 1.7, which gives a p -value $\approx 1 - 0.9554 = 0.0446$.

A sample of older students would have an average score at least as high as 125.2 in less than 5% of all samples assuming the null hypothesis is true. Because a sample result this high is

unlikely to occur just by chance, it is evidence that the mean for all older students is really higher than 115.

3. a. In the normal quantile plot below, all dots lie between the curved bands. So, the normality assumption is reasonably satisfied.



b. The null and alternative hypotheses are:

$$H_0 : \mu = 105$$

$$H_a : \mu \neq 105$$

The observed sample mean is $\bar{x} = 104.13$. The value of the test statistic is

$$z = \frac{104.13 - 105}{9/\sqrt{12}} \approx -0.335$$

This is a two-sided alternative. Hence the p-value is $2(0.3688) \approx 0.738$. That tells us that if the true mean reading of all radon detectors of this type is 105, we would expect to see an observed z-test value at least as extreme as the one we observed about 74% of the time. Based on these data, we have no evidence that the mean reading of these detectors differs from 105.

4. a. Let μ = the mean BMI for 6-year-old girls. The null and alternative hypotheses are

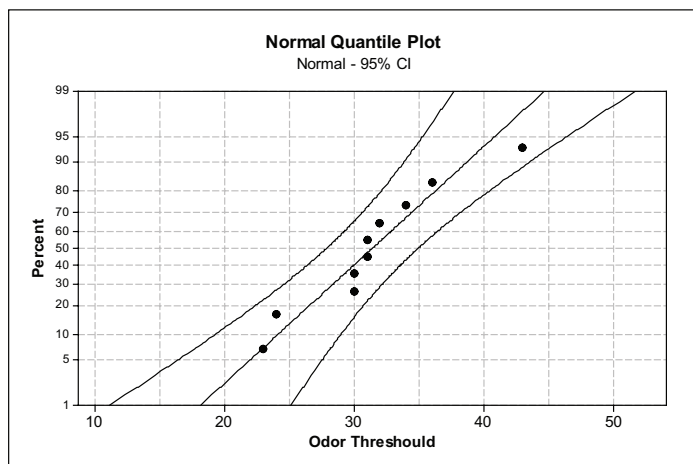
$$H_0 : \mu = 15.2 \text{ versus } H_a : \mu > 15.2.$$

b. $\bar{x} = 16.173 \text{ kg/m}^2$; $s = 2.669 \text{ kg/m}^2$

c. $z = \frac{16.173 - 15.2}{2.669/\sqrt{30}} \approx 2.00$; $p \approx 0.023$

d. Because $p \approx 0.023 < 0.05$, there is sufficient evidence to reject the null hypothesis and

conclude that the mean BMI for 6-year-old girls has increased since the time the CDC collected data to construct its BMI charts.



REVIEW QUESTIONS SOLUTIONS

1. a. The dots in the normal quantile plot are within the curved bands. In addition, the pattern is roughly linear. So, it is reasonable to assume these data are approximately normal.

b. The null and alternative hypotheses are:

$$H_0 : \mu = 25$$

$$H_a : \mu > 25$$

c. The observed sample mean is $\bar{x} = 30.4$. The value of the test statistic is

$$z = \frac{30.4 - 25}{7/\sqrt{10}} \approx 2.44 ; \text{ the } p\text{-value is } 1 - 0.9927 \text{ or } 0.0073.$$

That is, an observed average result at least as high as 30.4 would happen only 7 times in 1000 samples if the true mean were really 25. This is unlikely; hence, we have strong evidence that the mean odor threshold for students is higher than 25.

2. a. $\bar{x} = 516.2$ and $s = 80.7$

b. The null and alternative hypotheses are:

$$H_0 : \mu = 514$$

$$H_a : \mu \neq 514$$

The value of the test statistic is $z = \frac{516.2 - 514}{80.7/\sqrt{50}} \approx 0.19$. Since this is a two-sided alternative, the p -value is $2(0.4247) \approx 0.85$. There is insufficient evidence to reject the null hypothesis.

c. The confidence interval is $516.2 \pm 1.96(80.7/\sqrt{50})$ or (493.8, 538.6). Since 514 is in the confidence interval, it is one of the plausible values for μ .

3. a. The null and alternative hypotheses are:

$$H_0 : \mu = 90$$

$$H_a : \mu > 90$$

b. The sample mean and standard deviation are $\bar{x} = 118.4$ and $s = 186.5$;

$$z = \frac{118.4 - 90}{186.5/\sqrt{100}} \approx 1.52$$

the p -value is 0.064. The p -value is not sufficiently small to reject the null hypothesis.

c. Yes. Since the p -value is less than 0.10, we would conclude that there is sufficient evidence that the mean length of calls coming into the center have increased.

4. a. $H_0 : \mu = 5$ versus $H_a : \mu > 5$. You would want to gather evidence that the mean mercury concentration was above the recommended safe limit before closing the lake to fishing.

b. Sample answer #1: Significance level preference – 0.01. Choosing the smaller significance level means that there must be very strong evidence that the mercury level is unsafe before Lake Natoma is closed to fishing. Closing the lake would anger fishermen and could depress the part of the local economy that depends on fishing.

Sample answer #2: Significance level preference – 0.1. Since eating fish with high mercury concentrations is a serious health risk, it is better to err on the side of caution.

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 σ .
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

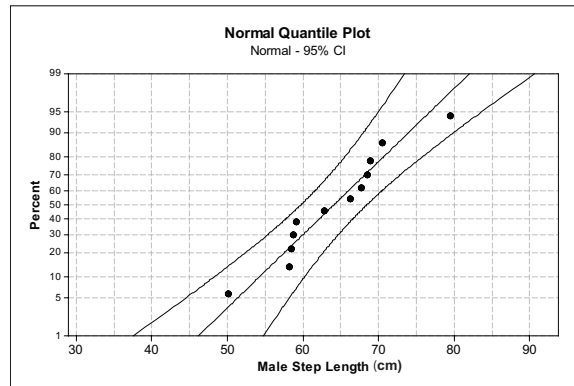
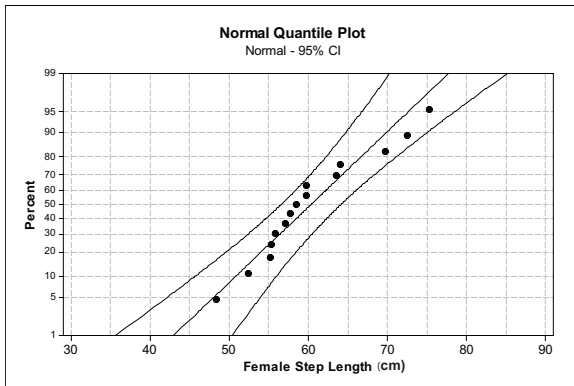
1. Sample answer: Place your heel against the wall. Then take four steps. Measure the distance from the wall to the back of your heel after the fourth step. Divide this distance by four to get the step length.

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: $df_M = 12 - 1 = 11$

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

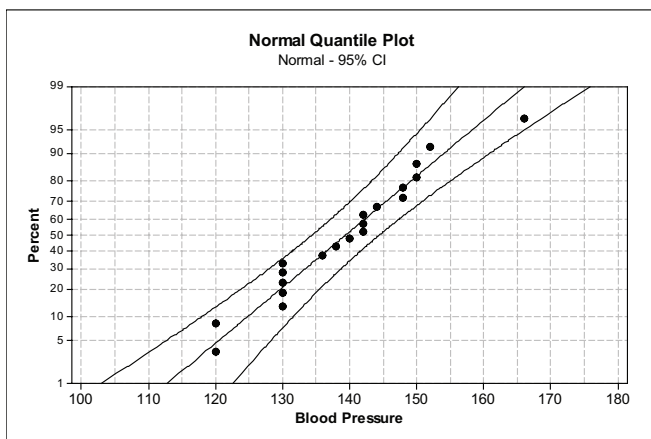
b. Sample answer: $df_F = 15 - 1 = 14$

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

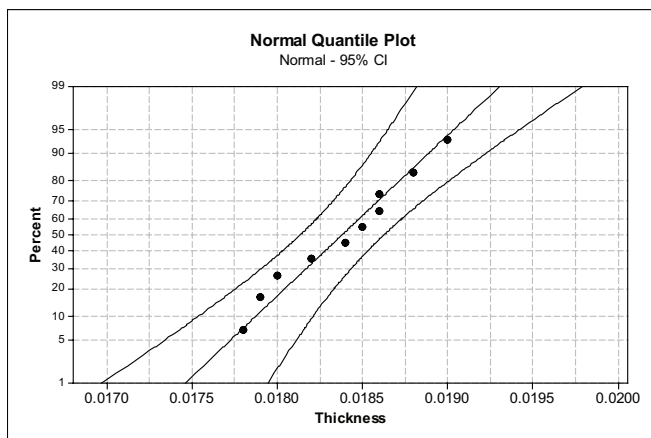
6. Sample answer: The confidence intervals contain the plausible values for the population means. Since the two confidence intervals share some values, you cannot conclude that the mean step length for men differs from the mean step length for women.

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 : $p\text{-value} \approx 2(0.0004)$ 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 μ_{Forearm} and μ_{Foot} .

c. 90% confidence interval for μ_{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 μ_{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 $s = 9.32$

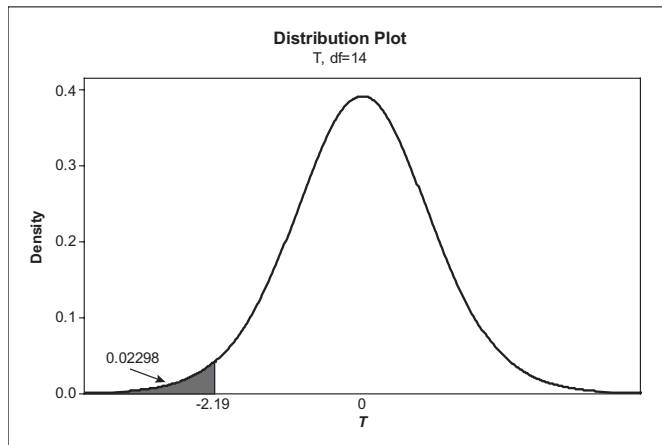
b. The null hypothesis is that there is no difference between Exam 1 and Exam 2 scores. The alternative hypothesis is that scores went down from Exam 1 to Exam 2, hence the mean difference (Exam 2 – Exam 1) is negative. These hypotheses can be expressed symbolically

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

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 μ_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 μ_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.

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

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.

Unit 27: Comparing Two Means



PREREQUISITES

Students should have experience with one-sample t -procedures before they begin this unit. That material is covered in Unit 26, Small Sample Inference for One Mean.

ADDITIONAL TOPIC COVERAGE

Additional coverage of two-sample t -procedures can be found in *The Basic Practice of Statistics*, Chapter 19, Two-Sample Problems.

ACTIVITY DESCRIPTION

In this activity, students will check whether the mean number of chips per cookie in Nabisco's Chips Ahoy regular and reduced fat chocolate chip cookies differ. In Unit 25's activity, students collected data on the number of chips per cookie from Chips Ahoy regular chocolate chip cookies. They can use those data in this activity. In addition, if they have not already done so, they will need to collect data on the number of chips per cookie from one or more bags of Chips Ahoy reduced fat chocolate chip cookies.

MATERIALS

One or two bags each of Nabisco's Chips Ahoy chocolate chip cookies, regular and reduced fat; paper plates or paper towels.

If students have not collected the data as part of Unit 25's activity, then they should be assigned to work in groups of 2 to 4 students to collect the data for this activity. If you need to have a group of 3, students in the group can trade off being chip counters. See Unit 25's Activity Description for suggestions on collecting these data. The sample data, on which sample solutions to the activity are based, are given in Table T27.1. For the sample data, two

independent counts were taken on the number of chips in each cookie, and then the results were averaged.

Original Chip Count (averaged from two independent counts)										
18.5	17.0	14.0	14.5	15.0	13.5	16.5	15.5	19.0	16.0	20.5
23.5	17.5	18.5	21.5	18.5	22.5	15.0	16.0	17.5	12.0	14.5
12.5	12.0	20.5	21.5	22.5	24.0	18.5	16.5	22.5	22.0	18.5
18.5	21.5	20.0	17.5	16.5	17.5	17.5	19.5	21.5	24.5	18.0
23.0	20.5	19.5	25.0	19.0	20.0	22.0	19.0	21.5	18.0	14.5
17.0	21.0	10.5	18.0	18.0	20.0	13.5	23.5	16.5	19.5	
Reduced Chip Count (averaged from two independent counts)										
11.0	7.5	5.5	8.5	9.5	15.5	20.0	14.0	18.5	21.5	18.5
18.0	19.5	16.0	21.0	19.0	17.0	21.5	13.5	14.5	14.0	18.5
11.5	13.0	19.0	20.0	16.0	18.0	14.0	20.0	9.0	13.0	11.5
14.5	13.5	11.5	13.0	18.0	19.0	21.5	16.5	12.0	14.0	14.0
11.0	14.5	15.0	21.0	21.5	17.5	15.5	16.5	14.5	16.0	16.5
10.5	15.5	16.0	15.0	9.0	11.5	8.5	15.0	13.5	17.0	15.0
9.5	5.5	12.5	17.0	19.5	20.5	16.0	20.0	19.0	15.5	19.5

Table T27.1. Sample data.

Students will need a copy of the class data for questions 2 – 5. If you decide not to collect the data (a task students really enjoy), have the class use the sample data in Table T27.1, which was collected in two statistics classes (each class got a bag of each type of cookie).

THE VIDEO SOLUTIONS

1. Sample answer: Take a sample from all licensed drivers in some state and look at the number of tickets the males and females in this group received for moving violations. We could then compare the mean number of tickets for women to the mean number of tickets for men.
2. The Hadza are a group of traditional hunter-gatherers who live in a way that is very similar to our ancestors. The men hunt with bows and arrows and the women forage for berries and root vegetables.
3. The original assumption was that the Hadza would burn more calories than the Westerners do, due to their more active lifestyle.
4. A two-sample t -test was used.
5. It turned out that no significant difference was found between the mean daily total energy expenditures of the Hadza and the Westerners. So, the researchers' original assumption that the Hadza burned more calories than Westerners due to a more active lifestyle was not supported by the data.
6. He placed the blame on people eating too much rather than on a sedentary lifestyle.

UNIT ACTIVITY SOLUTIONS

1. See question 3.

2. a. Sample answer: Since chocolate chips have fat in them, we think that Nabisco may put fewer chips in its reduced fat cookies as a way of reducing the fat content.

b. Sample answer: Given our answer in (a), we decided to use a one-sided alternative. Let μ_1 = the mean number of chips per cookie in regular Chips Ahoy chocolate chip cookies and let μ_2 = the mean number of chips per cookie in reduced fat Chips Ahoy chocolate chip cookies. Below are the null and alternative hypotheses:

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_a : \mu_1 - \mu_2 > 0$$

3. Sample data:

Original Chip Count (averaged from two independent counts)

18.5	17.0	14.0	14.5	15.0	13.5	16.5	15.5	19.0	16.0	20.5
23.5	17.5	18.5	21.5	18.5	22.5	15.0	16.0	17.5	12.0	14.5
12.5	12.0	20.5	21.5	22.5	24.0	18.5	16.5	22.5	22.0	18.5
18.5	21.5	20.0	17.5	16.5	17.5	17.5	19.5	21.5	24.5	18.0
23.0	20.5	19.5	25.0	19.0	20.0	22.0	19.0	21.5	18.0	14.5
17.0	21.0	10.5	18.0	18.0	20.0	13.5	23.5	16.5	19.5	

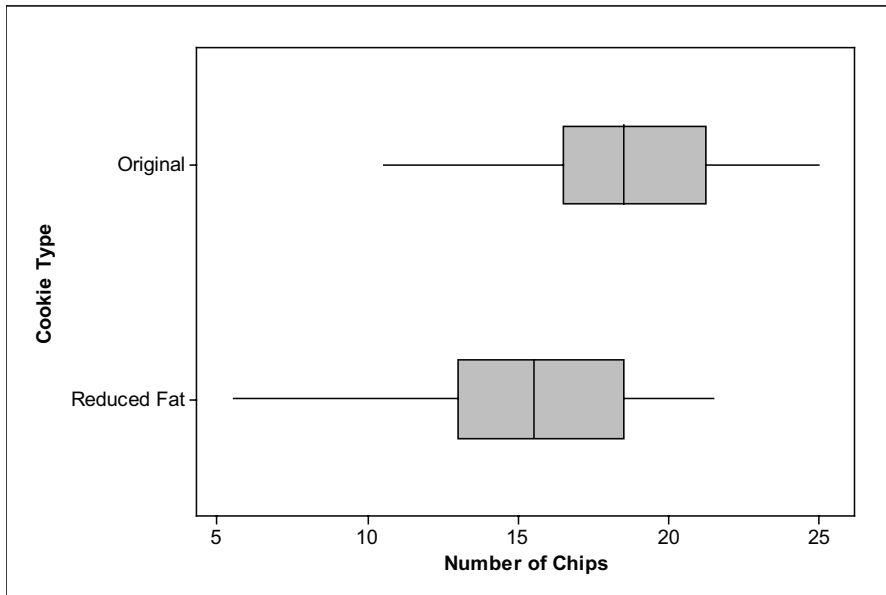
Reduced Chip Count (averaged from two independent counts)

11.0	7.5	5.5	8.5	9.5	15.5	20.0	14.0	18.5	21.5	18.5
18.0	19.5	16.0	21.0	19.0	17.0	21.5	13.5	14.5	14.0	18.5
11.5	13.0	19.0	20.0	16.0	18.0	14.0	20.0	9.0	13.0	11.5
14.5	13.5	11.5	13.0	18.0	19.0	21.5	16.5	12.0	14.0	14.0
11.0	14.5	15.0	21.0	21.5	17.5	15.5	16.5	14.5	16.0	16.5
10.5	15.5	16.0	15.0	9.0	11.5	8.5	15.0	13.5	17.0	15.0
9.5	5.5	12.5	17.0	19.5	20.5	16.0	20.0	19.0	15.5	19.5

Regular: $n_1 = 65$, $\bar{x}_1 = 18.462$, $s_1 = 3.308$

Reduced Fat: $n_2 = 77$, $\bar{x}_2 = 15.266$, $s_2 = 3.941$

4. Based on comparative boxplots, it appears that Chips Ahoy regular has more chips per cookie than Chips Ahoy reduced fat.



5. a. Sample answer:

$$t = \frac{(18.462 - 15.266) - 0}{\sqrt{\frac{3.308^2}{65} + \frac{3.941^2}{77}}} \approx 5.25$$

b. Sample answer: Using a conservative approach, we assume that $df = 65 - 1 = 64$. From software and using a one-sided alternative, we get a p -value that is essentially 0 (at least to 4 decimals).

c. Sample answer: Reject the null hypothesis. Conclude that the difference in the mean number of chips per cookie between the two types of cookies is positive. Hence, the mean number of chips per cookie in Chips Ahoy regular cookies is greater than the mean number of chips per cookie in Chips Ahoy reduced fat cookies.

6. Sample answer: $(18.462 - 15.266) \pm (1.998)\sqrt{\frac{(3.308)^2}{65} + \frac{(3.941)^2}{77}} \approx 3.196 \pm 1.215$, or
(1.981, 4.411)

EXERCISE SOLUTIONS

1. a. Let μ_1 be the mean job performance rating for pregnant employees and μ_2 be the mean job performance rating for non-pregnant female employees.

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_a : \mu_1 - \mu_2 \neq 0$$

$$\text{b. } t = \frac{(2.38 - 2.69) - 0}{\sqrt{\frac{1.10^2}{71} + \frac{0.58^2}{71}}} \approx -2.10$$

Using a two-sided alternative, and letting $df = 71 - 1 = 70$ (the conservative approach), we determine $p = 2(0.01967) \approx 0.039$. (If we use the statistical package Minitab, we get $df = 106$ and $p = 0.038$. The conservative approach yields a slightly higher p -value.) We conclude that there is a statistical difference in the mean job performance ratings for the two groups.

c. Using the conservative approach, we set $df = 70$ in order to determine the t -critical value: $t^* \approx 1.994$.

$$(2.38 - 2.69) \pm (1.994) \sqrt{\frac{1.10^2}{71} + \frac{0.58^2}{71}} \approx -0.31 \pm 0.294, \text{ or } (-0.604, -0.016)$$

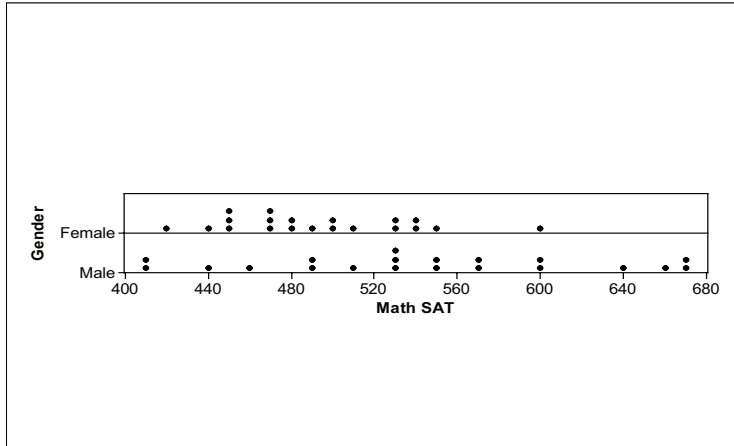
Since both endpoints of the confidence interval are negative, the mean job performance appraisal rating for pregnant employees is lower than for non-pregnant female employees. In this case, lower ratings are associated with better job performance.

2. a. A one-sample paired t -test should be used. The “during” and “after” data involve the same group of women rather than samples from two independent populations.

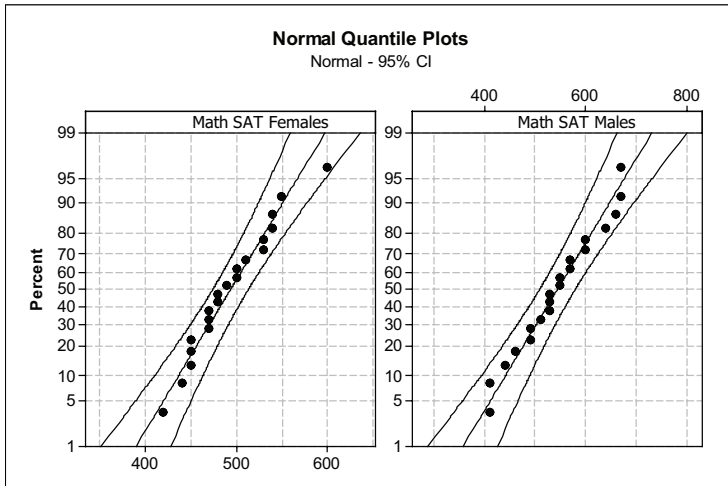
$$\text{b. } t = \frac{-0.27 - 0}{\frac{2.00}{\sqrt{71}}} \approx -1.138 ;$$

based on a t -distribution with $df = 70$ and a two-sided alternative, $p = 2(0.1295) \approx 0.259$. The mean difference in job performance appraisal ratings during pregnancy and after returning from pregnancy leave do not differ significantly.

3. a. Sample answer: It is somewhat difficult to tell if the distribution of SAT Math scores is higher for males than for females. From the dotplots, the distribution of SAT Math scores is more variable for males than for females. The highest SAT Math scores are from males but the lowest are also from males. Nevertheless, the midpoint of the distribution for the male students appears higher than the midpoint of the distribution for the female students.



b. Consider the normal quantile plots below. Given the dots fall within 95% confidence interval bands on the normal quantile plots, it is reasonable to assume the data from both samples come from approximately normal distributions.



c. $\bar{x}_F = 493.5$, $s_F = 44.64$; $\bar{x}_M = 544.0$, $s_M = 80.6$

d. $t = \frac{(544.0 - 493.5) - 0}{\sqrt{\frac{80.6^2}{20} + \frac{44.64^2}{20}}} \approx 2.45$; using $df = 19$, we get the following p -value: $p \approx 0.012$

Therefore, we can conclude that the mean SAT Math scores are significantly higher for first-year male students entering this university than for first-year female students.

4. a. $H_0 : \mu_G - \mu_B = 0$ versus $H_a : \mu_G - \mu_B \neq 0$. (Could also be expressed as $H_0 : \mu_G = \mu_B$ versus $H_a : \mu_G \neq \mu_B$.)

$$b. t = \frac{(494 - 409) - 0}{\sqrt{\frac{(172)^2}{31} + \frac{(148)^2}{27}}} \approx 2.023$$

c. Use $df = 27 - 1$, or 26; $p = 2(0.02673) \approx 0.053$. Since $p > 0.05$, the results are not significant at the 0.05 level.

$$d. df = \frac{\left(\frac{172^2}{31} + \frac{148^2}{27} \right)^2}{\left(\frac{1}{30} \right) \left(\frac{172^2}{31} \right)^2 + \left(\frac{1}{26} \right) \left(\frac{148^2}{27} \right)^2} \approx 55.9948 ; \text{ use } df = 55.$$

$p = 2(0.02397) \approx 0.048$. Since $p < 0.05$, the results are significant at the 0.05 level.

Notice that the p -value in (c) is only slightly larger using the conservative approach than it is here.

REVIEW QUESTIONS SOLUTIONS

1. The men in the sample earned more, on average, than did the women. The difference in the sample means was large enough that it couldn't be attributed to chance variation. If we assume that men and women in the entire student population have the same average earnings, the probability of observing a difference as large as we observed is only $p = 0.038$, making it pretty unlikely (less than a 4% chance). Because this probability is so small, we have good evidence that the average earnings of all male and female students (not just those who happen to be in the samples) differ.

There was also some difference between the average earnings of African- American and white students in the sample. However, a difference at least this large would happen almost half the time ($p = 0.476$) just by chance if African Americans and whites in the entire student population had exactly the same average earnings.

2. a. Sample answer: No mention was made that the samples were randomly selected from each area. The researchers wanted to be sure that the differences in pulmonary function could be attributed to the pollution levels in the two areas and not due to a lurking variable such as age, height, weight or BMI that was different for the men in the two groups of participants in the study.

$$\text{b. } t = \frac{(4.49 - 4.32) - 0}{\sqrt{\frac{0.43^2}{60} + \frac{0.45^2}{60}}} \approx 2.116; \text{ using } df = 59, \text{ we get } p = 2(0.01929) \approx 0.039$$

We conclude that there is a statistically significant difference in mean forced vital capacity (FVC) in healthy, non-smoking, young men in Area 1 and Area 2. (Because the sample size was large, even a relatively small difference in sample means turned out to be statistically different.)

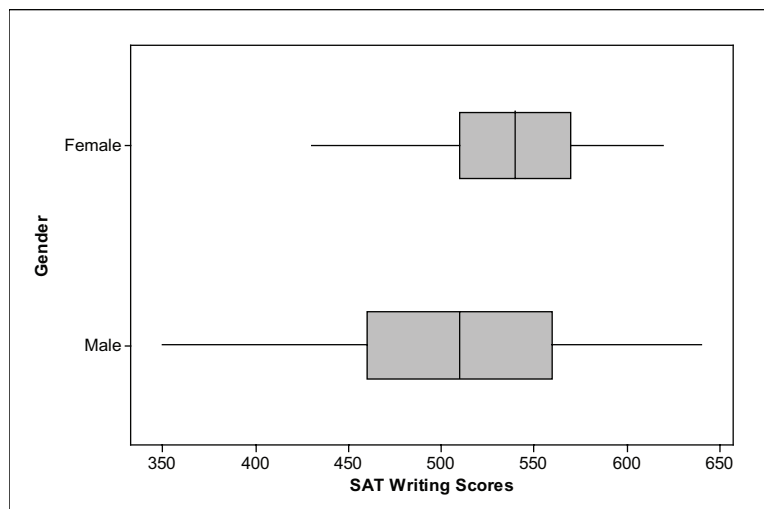
$$\text{c. } t = \frac{(17.17 - 16.28) - 0}{\sqrt{\frac{4.26^2}{60} + \frac{2.39^2}{60}}} \approx 1.411; \text{ using } df = 59, \text{ we get } p = 2(0.08175) \approx 0.164$$

We conclude that there is no significant difference in the mean respiratory rate in healthy, non-smoking, young men in Areas 1 and 2.

d. We construct a 95% confidence interval for the difference in mean FVC between the two areas. We use $df = 59$ (conservative approach) to find the t -critical value: $t^* = 2.001$.

$$(4.49 - 4.32) \pm (2.001) \sqrt{\frac{0.43^2}{60} + \frac{0.45^2}{60}} \approx 0.17 \pm 0.16 \text{ or } (0.01, 0.33).$$

3. a. Both distributions appear reasonably symmetric. Neither data set has outliers. The whiskers on the boxplots appear slightly longer than half the width of the boxes. These are characteristics that you would expect from normal data. So, based on the boxplots it appears reasonable to assume that both data sets are approximately normal. Although the SAT Writing scores for the male students are more spread out than for the female students, it looks as if the SAT Writing scores for the female students tend to be higher than for the male students.



b. Female students: $n_F = 25$, $\bar{x}_F = 541.20$, $s_F = 48.59$

Male students: $n_M = 35$, $\bar{x}_M = 507.10$, $s_M = 73.70$

$$c. t = \frac{(541.20 - 507.1) - 0}{\sqrt{\frac{(48.59)^2}{25} + \frac{(73.7)^2}{35}}} \approx 2.16$$

Adopting a conservative approach, we use $df = 24$ to determine a p -value: We get $p = 2(0.02049) \approx 0.041 < 0.05$. We conclude that there is a significant difference between the mean SAT Writing scores for first-year women at this university compared to first-year men.

$$d. (541.20 - 507.10) \pm 2.064 \sqrt{\frac{48.59^2}{25} + \frac{73.7^2}{35}} \approx 34.10 \pm 32.61, \text{ or } (1.49, 66.71)$$

4. a. $H_0 : \mu_B - \mu_G = 0$ versus $H_a : \mu_B - \mu_G > 0$.

(Could also be expressed as $H_0 : \mu_B = \mu_G$ versus $H_a : \mu_B > \mu_G$.)

$$\text{b. } t = \frac{(92.3 - 80.8) - 0}{\sqrt{\frac{42.0^2}{27} + \frac{41.4^2}{31}}} \approx 1.05; \text{ using } df = 26, p = 0.152$$

There is insufficient evidence to reject the null hypothesis in favor of the alternative. Hence, we are unable to conclude from these data that 4-year-old boys consume more mouthfuls of food during a meal than 4-year-old girls.

Unit 28: Inference for Proportions



PREREQUISITES

Students need the background on z-procedures covered in Unit 24, Confidence Intervals, and Unit 25, Tests of Significance.

ADDITIONAL TOPIC COVERAGE

Additional coverage of inference for proportions can be found in *The Basic Practice of Statistics*, Chapter 20, Inference about a Population Proportion, and Chapter 21, Comparing Two Proportions.

ACTIVITY DESCRIPTION

In this activity, students revisit the data they collected for Unit 21's activity. However, here the focus is on using sample proportions to estimate population proportions. The population is children born to brown-eyed parents, each of whom has a recessive gene for blue eyes. The characteristic of interest is blue eyes. In this case, we know the population proportion of blue-eyed children is $p = 0.25$. In this activity, students use the simulated data collected in Unit 21's activity to make a number of estimates of p using samples of different sizes.

MATERIALS

Completed Table 21.1, simulated data for Unit 21's activity.

In Unit 21, each sample represented a family. Here each size-4 sample can be thought of as a simple random sample from the population of all children born to brown-eyed parents with the recessive gene for blue eyes.

In question 2, students calculate the sample proportions and use them as estimates of the population proportion of blue-eyed children born to brown-eyed parents with recessive genes for blue eyes. Since the data contain 30 size-4 samples, students will have 30 values of \hat{p} . Next, they make a histogram of their sample proportions to get a sense of the shape of the sampling distribution of the sample proportion. (The concept of a sampling distribution of a sample statistic is difficult for students to grasp.) Students' histograms won't closely resemble a normal density curve because the sample size is too small. However, this question can serve as a good springboard to a discussion of the sampling distribution of the sample proportion when the sample size is large.

In questions 4, 5, and 6, students combine data from the first 10 samples, first 20 samples, and then all 30 samples. They use these data to construct three 95% confidence intervals for p . This gives students an opportunity to see that the size of the margin of error shrinks as the sample size increases.

Question 8 involves a little algebra. (Tell students to skip this problem if you don't want to deal with the algebra. In that case, don't assign Review Question 4(d).) In question 8, students find the sample size that guarantees that the margin of error will be less than 0.5. If E represents the margin of error, then the sample size needed to guarantee a margin of error less than E in a 95% confidence interval is:

$$n = \left(\frac{1.96}{E} \right)^2 \hat{p}(1 - \hat{p})$$

The problem is that \hat{p} varies and is not known until after the sample is collected. Since $\hat{p}(1 - \hat{p})$ has a maximum value of 0.25 when $\hat{p} = 0.5$, we can substitute this value into the formula for n , which gives a conservative value for n . The result is the following formula for determining the sample size needed to guarantee that the margin of error is less than E regardless of the value of \hat{p} :

$$n = 0.25 \left(\frac{1.96}{E} \right)^2$$

Finally, in the world of simulation – unlike in the real world – we know that $p = 0.25$ for the sample data collected for this activity. Question 9 gives students an opportunity to see if their confidence intervals got it right – in other words, contained the true population proportion of 0.25.

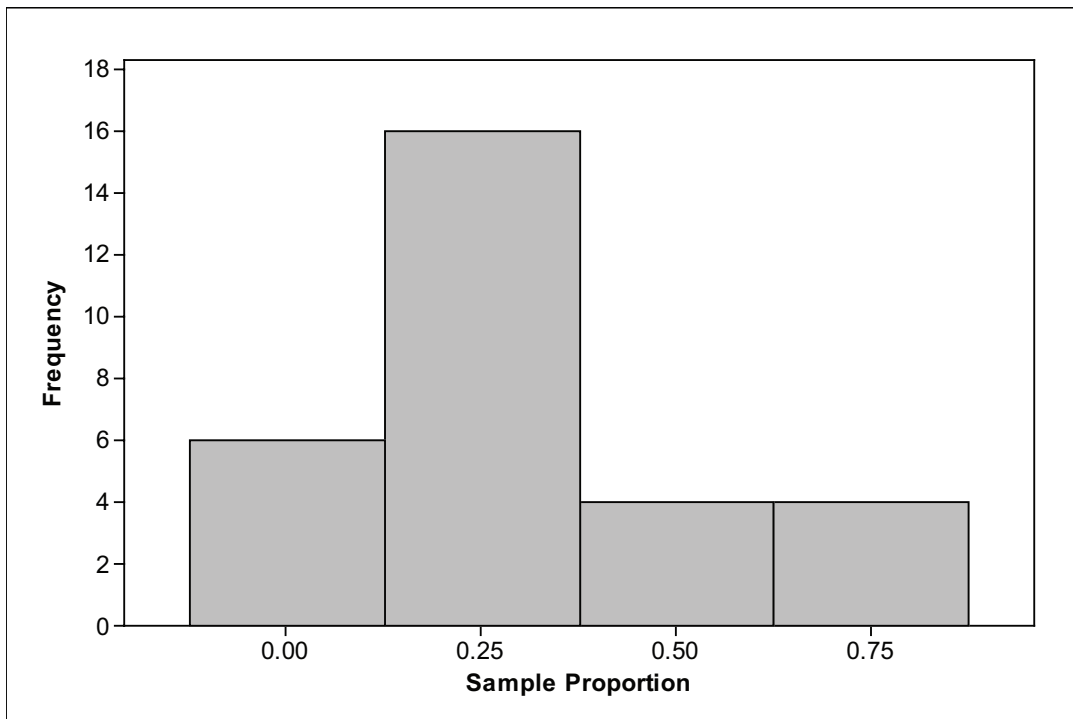
THE VIDEO SOLUTIONS

1. Using information from samples to make inferences about population proportions.
2. People were able to solve problems creatively and come up with new ideas when they felt most motivated and excited about their work.
3. Progress is paramount to people feeling positive and highly motivated about their work.
4. We would expect 20% of the managers to select progress.
5. A z-test statistic was used.
6. The values in the confidence interval all fell below 20%.

UNIT ACTIVITY:

PROPORTIONS OF BLUE EYES SOLUTIONS

1. See sample answer to 3(b). The sample data in the solution to 3(b) will be used for sample answers to this activity.
2. a. See sample answer to 3(b).
b. Sample answer: The smallest sample proportion was 0 and the largest was 0.75.
c. The histogram does not appear to have a normal distribution. There were only four observed values for the sample proportion. However, there is one peak at value 0.25.



3. a. See sample answer to 3(b).
(On next page...)

3. b.

Sample Number	Number of Blue-Eyed Children $n = 4$	Sample Proportion Blue-Eyed Children $n = 4$	Running Total Number of Children	Running Total Number of Blue-Eyed Children
1	1	0.25	4	1
2	0	0	8	1
3	2	0.5	12	3
4	1	0.25	16	4
5	0	0	20	4
6	1	0.25	24	5
7	1	0.25	28	6
8	3	0.75	32	9
9	1	0.25	36	10
10	1	0.25	40	11
11	1	0.25	44	12
12	3	0.75	48	15
13	0	0	52	15
14	3	0.75	56	18
15	1	0.25	60	19
16	1	0.25	64	20
17	2	0.5	68	22
18	1	0.25	72	23
19	2	0.5	76	25
20	1	0.25	80	26
21	1	0.25	84	27
22	1	0.25	88	28
23	1	0.25	92	29
24	2	0.5	96	31
25	1	0.25	100	32
26	0	0	104	32
27	3	0.75	108	35
28	0	0	112	35
29	0	0	116	35
30	1	0.25	120	36

4. a. Sample answer based on data shown in 3(b): $\hat{p} = 11/40 = 0.275$

b. $0.275 \pm (1.96) \sqrt{\frac{(0.275)(1-0.275)}{40}} = 0.275 \pm 0.138$; from 0.137 to 0.413

c. The margin of error is 0.138.

5. a. Sample answer based on data shown in 3(b): $\hat{p} = 26/80 = 0.325$

b. $0.325 \pm (1.96)\sqrt{\frac{(0.325)(1-0.325)}{80}} = 0.325 \pm 0.103$; from 0.222 to 0.428

c. The margin of error is 0.103.

6. a. Sample answer based on data shown in 3(b): $\hat{p} = 36/120 = 0.3$

b. $0.300 \pm (1.96)\sqrt{\frac{(0.300)(1-0.300)}{120}} = 0.300 \pm 0.082$; from 0.218 to 0.382

c. The margin of error is 0.082.

7. The margin of error decreased as the sample size increased.

8. a. $n = (1536.64)(\hat{p})(1 - \hat{p})$

b.

\hat{p}	$\hat{p}(1 - \hat{p})$
0.1	0.09
0.2	0.16
0.3	0.21
0.4	0.24
0.5	0.25
0.6	0.24
0.7	0.21
0.8	0.16
0.9	0.09

c. $n = (1536.64)(0.25) = 384.16$. So, a sample size of at least $n = 385$ will guarantee that the margin of error is less than 0.05.

9. Sample answer: All of our confidence intervals contained the value 0.25 and gave correct results.

EXERCISE SOLUTIONS

1. a. $\hat{p} = \frac{2098}{2454} \approx 0.855$, or around 85.5%

b. $0.855 \pm 1.96 \sqrt{\frac{(0.855)(1-0.855)}{2454}}$

$0.855 \pm (1.96)(0.007) = 0.855 \pm 0.014$; between 0.841 and 0.869 or between 84.1% and 86.9%

c. A 90% confidence interval would be narrower. Using a less reliable procedure gives more precision. Instead of using $z^* = 1.96$, we would use $z^* = 1.654$, which would make for a smaller margin of error.

2. a. $z = \frac{0.466 - 0.50}{\sqrt{\frac{(0.50)(0.50)}{440}}} \approx -1.426$; $p = 2(0.07693) \approx 0.154$.

There is insufficient evidence to reject the null hypothesis.

b. $0.466 \pm 1.96 \sqrt{\frac{(0.466)(1-0.466)}{440}} \approx 0.466 \pm 0.047$; from 0.419 to 0.513, or 41.9% to 51.3%

c. Since 0.5 is in the confidence interval, we cannot reject the null hypothesis. This decision is the same as the decision made using the z-test statistic in (a). (It should be noted that in the case of proportions, results from significance tests and confidence intervals do not always lead to the same results.)

3. a. The population is all American households in which a member of the household owns or uses a computer at home.

b. $H_0 : p = 0.90$

$H_a : p > 0.90$

c. $\hat{p} = 1816/1910 \approx 0.951$

$z = \frac{0.951 - 0.90}{\sqrt{\frac{(0.90)(0.10)}{1910}}} \approx 7.43$

The p -value is essentially 0.

Conclusion: Reject the null hypothesis and conclude the population percentage is greater than 90%.

$$4. 0.951 \pm 1.96 \sqrt{\frac{(.951)(1-0.951)}{1910}} \approx 0.951 \pm 0.010 ; \text{ between } 0.941 \text{ and } 0.961$$

Expressed as a percentage: between 94.1% and 96.1%

REVIEW QUESTIONS SOLUTIONS

1. a. $\hat{p} = \frac{2998}{5462} \approx 0.549$, or around 54.9%

b. $H_0: p = 0.5$
 $H_a: p > 0.5$

c. $z = \frac{0.549 - 0.50}{\sqrt{\frac{(0.50)(1-0.50)}{5462}}} \approx 7.24$; $p \approx 0.000$.

Hence, we can conclude that the physical education teacher is correct.

2. a. The result of the poll was that between 45.4% and 50.6% (that's $48\% \pm 2.6\%$) of the voters intended to vote for Obama. This result was obtained by a method that gives correct results 95% of the time when used repeatedly.

b. The two confidence intervals, Obama's 45.4% to 50.6% and Romney's 44.4% to 49.6%, overlap. If Obama's actual percentage is at the low end of his confidence interval and Romney's actual percentage is at the high end of his confidence interval, then Romney would win. On the other hand, if Obama's actual percentage is at the high end of his confidence interval and Romney's is at the low end of his interval, then Obama would win.

3. a. $\hat{p} = \frac{255}{296} \approx 0.861$

$$0.861 \pm 1.96 \sqrt{\frac{(0.861)(1-0.861)}{296}} \approx 0.861 \pm 0.039;$$

between 0.822 and 0.900, or between 82.2% and 90.0%.

b. $\hat{p} = \frac{234}{298} \approx 0.785$

$$0.785 \pm 1.96 \sqrt{\frac{(0.785)(1-0.785)}{298}} \approx 0.785 \pm 0.047;$$

between 0.738 and 0.832, or between 73.8% and 83.2%.

c. $\hat{p} = \frac{174}{297} \approx 0.586$

$$0.586 \pm 1.96 \sqrt{\frac{(0.586)(1-0.586)}{297}} \approx 0.586 \pm 0.056 ; \text{ between } 0.530 \text{ and } 0.642.$$

4. a. $0.63 \pm 1.96 \sqrt{\frac{0.63(1-0.63)}{1000}} \approx 0.63 \pm 0.03 ; \text{ from } 0.60 \text{ to } 0.66, \text{ or from } 60\% \text{ to } 66\%.$

The margin of error, to two decimals, is 0.03 or $\pm 3\%$.

(It was 2.99%, which we rounded to 3%.)

b. They match.

c. Corresponding to 50%: margin of error is $1.96 \sqrt{\frac{(.5)(.5)}{1000}} \approx 0.031 ; 3.1\%.$

So, the margin of error would be 3% only if we round to the nearest whole percent.

Corresponding to 80%: margin of error is $1.96 \sqrt{\frac{(.8)(.2)}{1000}} \approx 0.025 ; 2.5\%$

d. $n = \left(\frac{1.96}{.03} \right)^2 (0.25) \approx 1067.11.$

In order to guarantee that the margin of error was less than 3%, a sample size of at least 1,068 should have been used.

Unit 29: Inference for Two-Way Tables



PREREQUISITES

Unit 13, Two-Way Tables is a prerequisite for this unit. In addition, students need some background in significance tests, which was introduced in Unit 25.

ADDITIONAL TOPIC COVERAGE

Additional coverage of inference for two-way tables can be found in *The Basic Practice of Statistics*, Chapter 23, Two Categorical Variables: The Chi-Square Test.

ACTIVITY DESCRIPTION

Students should work in small groups on this activity. The activity consists of three parts. The first part provides a justification for the formula for computing the expected cell counts for chi-square tables. Students can work on Part I on their own or it could be part of a lecture/class discussion. Parts II and III involve two different structures for datasets, both of which are appropriate for the chi-square analysis covered in this unit.

Here are the two data structures: (1) subjects from a single sample are classified according to two categorical variables and (2) subjects from multiple samples (drawn from different populations) are classified according to a single categorical variable. In the latter case, “which sample” can be thought of as the second categorical variable. In the first case, a chi-square test for independence is performed; in the second case, a chi-square test for homogeneity is performed. The chi-square test statistics and the analyses are the same for both situations. So, in this unit, we have put little emphasis on distinguishing between these two situations.

MATERIALS

For Part III, bags of at least two different types of M&Ms are needed. Large-sized bags were used for the sample data, with the exception of the M&Ms minis, for which a medium bag was purchased. In addition, students will need paper plates or bowls to contain the M&Ms while they are being counted.

Part I: Introduction – Assumption of Independence and Expected Count Formula

Part I provides an explanation of the expected counts formula used in a chi-square test of independence. Students need to be familiar with the Multiplication Rule from Unit 19, Probability Models. This part could be approached either as an activity or as part of an informal lecture that introduces the topic of this activity. It could also be skipped and students could move directly to Part II.

Part II: Single Sample, Classified on Two Categorical Variables

For this part, students will need to collect data from people. The class could serve as the sample, or perhaps combine this class with another class, or have students add their friends to the sample. Students will need to classify each individual in the sample by gender and eye color. An easy way to collect the data is to draw a table on the board. Each student should come up to the board and put a tally line in the appropriate box for gender and eye color. After students have completed their entries, numbers can replace the tally marks. Students can then copy the table from the board and begin work on Part II.

Part III: Multiple Samples, Classified on One Categorical Variable

Students should work in groups to collect the data on the M&Ms colors. Again, you may want to put a chart on the board and have students enter their results for each color as they finish sorting their M&Ms into colors. Once the data are collected, groups will need a copy of the class data. Since the resulting two-way table is quite large, group members should be encouraged to divide up the work of computing the expected cell counts.

The color distribution of M&Ms differs by types and has changed over the years. You can write to Mars, the makers of M&Ms, for the latest color distribution in its candies.

THE VIDEO SOLUTIONS

1. Dr. Pardis Sabeti investigates the nonstop evolutionary arms race between our bodies and the infectious microorganisms that invade and inhabit them. In other words, she investigates connections between genotypes and protections from infectious diseases. Her work on Lassa fever is still in its early stages.

2. Sickle cell anemia hemoglobin mutation, HbS.

3. H_0 : No association between malaria and HbS.

H_a : Association between malaria and HbS.

4. Expected count = $\frac{(\text{row total})(\text{column total})}{\text{grand total}}$.

5. We reject the null hypothesis and conclude that there is an association between the HbS gene and malaria.

UNIT ACTIVITY:

ASSOCIATIONS WITH COLOR SOLUTIONS

Part I: Introduction – Assumption of Independence and Expected Count Formula

1. a. $P(\text{DEM and female}) = P(\text{DEM}) P(\text{female}) = \left(\frac{196}{500}\right)\left(\frac{246}{500}\right) \approx 0.1929$

b. Expected number = $\left(\frac{196}{500}\right)\left(\frac{246}{500}\right)(500) = \frac{(196)(246)}{500} \approx 96.432$

c. Expected count = $\frac{(196)(246)}{500} \approx 96.432$

d. $P(\text{DEM and male}) = P(\text{DEM})P(\text{male}) = \left(\frac{196}{500}\right)\left(\frac{254}{500}\right) \approx 0.1991$

Expected number = $\left(\frac{196}{500}\right)\left(\frac{254}{500}\right)(500) = \frac{(196)(254)}{500} \approx 99.57$

Expected count = $\frac{(196)(254)}{500} \approx 99.57$

2. a.

Expected		Male	Female	Total
Political	DEM (Blue)	96.43	99.57	196
Preference	GOP (Red)	91.02	93.98	185
Color	IND (White)	58.55	60.45	119
Total		246	254	500

b. $\chi^2 = +\frac{(107 - 96.43)^2}{96.43} + \frac{(89 - 99.57)^2}{99.57} + \dots + \frac{(56 - 60.45)^2}{60.45} \approx 7.825$

$df = (3 - 1)(2 - 1) = 2; p \approx 0.02$

c. There is sufficient evidence to reject the null hypothesis. There is association between these two variables. In other words, they are dependent.

3. a. Sample data will be used to provide sample answers.

		Eye Color			Total
Count		Blue	Brown	Other	
Gender	Male	8	20	6	34
	Female	4	16	12	32
Total		12	36	18	66

b. H_0 : No association between gender and eye color.

H_a : Association between gender and eye color.

c. Sample answer:

		Eye Color			Total
Count		Blue	Brown	Other	
Gender	Male	8 6.18	20 18.55	6 9.27	34
	Female	4 5.82	16 17.45	12 8.73	32
Total		12	36	18	66

d. Sample answer:

$$\chi^2 = \frac{(8 - 6.18)^2}{6.18} + \frac{(20 - 18.55)^2}{18.55} + \dots + \frac{(12 - 8.73)^2}{8.73} \approx 3.72 ; df = 2$$

$p \approx 0.151$. There is insufficient evidence to reject the null hypothesis. In other words, there is no strong evidence to suggest that there is an association between eye color and gender.

4. a. Sample data (will be used for sample answers) (See next page...):

Count		Type 1 Dark	Type 2 Regular	Type 3 Peanut	Type 4 Mini	Total
Color	Green	112	109	41	228	490
	Blue	188	160	39	203	590
	Yellow	75	91	47	210	423
	Orange	141	123	36	187	487
	Red	81	62	20	221	384
	Brown	59	84	30	100	273
Total		656	629	213	1149	2647

b. H_0 : No association between M&M type and color distribution.

H_a : Association between M&M type and color distribution.

c. Sample answer:

Count		Type 1 Dark	Type 2 Regular	Type 3 Peanut	Type 4 Mini	Total
Color	Green	112 121.4	109 116.4	41 39.4	228 212.7	490
	Blue	188 146.2	160 140.2	39 47.5	203 256.1	590
	Yellow	75 104.8	91 100.5	47 34	210 183.6	423
	Orange	141 120.7	123 115.7	36 39.2	187 211.4	487
	Red	81 95.2	62 91.2	20 30.9	221 166.7	384
	Brown	59 67.7	84 64.9	30 22	100 118.5	273
Total		656	629	213	1149	2647

d. $\chi^2 \approx 100.3$; $df = (6 - 1)(4 - 1) = 15$; $p \approx 0$

There is an association between M&Ms type and color distribution. In other words, Different types of M&Ms have different color distributions.

EXERCISE SOLUTIONS

1. a. There were two cells with expected counts less than 1. The guidelines call for all expected counts to be greater than 1. Also, there were 7 cells with expected counts below 5. That means that around 39% of the cells have expected counts under 5. The guidelines state that no more than 20% of the cells should have expected counts less than 5.

b. See solution to (c).

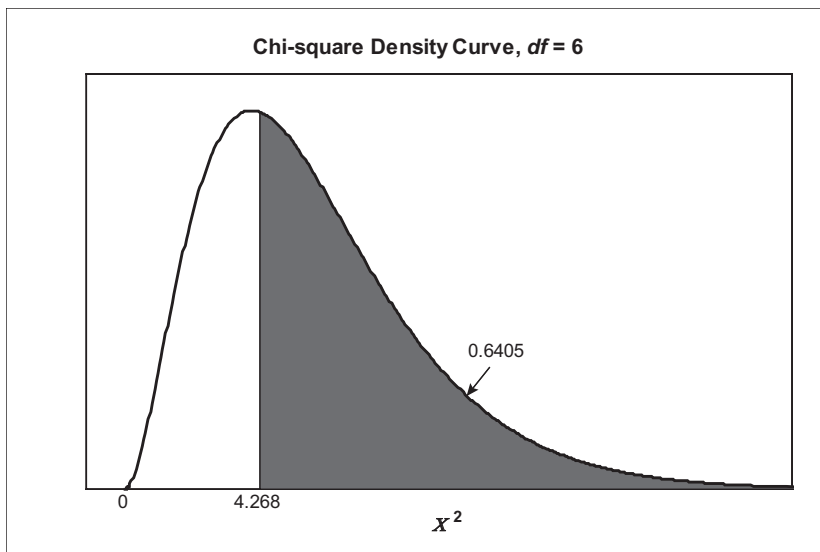
c. Based on the completed table below, all expected counts were greater than 1. Two expected counts were below 5, which is just under 17% of the cells. So, the expected counts in the table below meet the guidelines.

			Environment			Total
			Farm	Country	City	
Energy Drinks	None	Observed	57	144	598	799
		Expected	52.55	150.44	596.01	
	One	Observed	11	44	160	215
		Expected	14.14	40.48	160.38	
	Two	Observed	4	13	36	53
		Expected	3.49	9.98	39.54	
	Three +	Observed	1	8	34	43
		Expected	2.83	8.10	32.08	
Total			73	209	828	1110

d. This is a 4×3 table; $df = (4 - 1)(3 - 1) = 6$. The chi-square test statistic is calculated below:

$$\begin{aligned}
 \chi^2 &= \frac{(57 - 52.55)^2}{52.55} + \frac{(144 - 150.44)^2}{150.44} + \frac{(598 - 596.01)^2}{596.01} \\
 &+ \frac{(11 - 14.14)^2}{14.14} + \frac{(44 - 40.48)^2}{40.48} + \frac{(160 - 160.38)^2}{160.38} \\
 &+ \frac{(4 - 3.49)^2}{3.49} + \frac{(13 - 9.98)^2}{9.98} + \frac{(36 - 39.54)^2}{39.54} \\
 &+ \frac{(1 - 2.83)^2}{2.83} + \frac{(8 - 8.10)^2}{8.10} + \frac{(34 - 32.08)^2}{32.08} \\
 &\approx 4.268
 \end{aligned}$$

e. $p \approx 0.64$. (See area under density curve below.) There is insufficient evidence to reject the null hypothesis. We found no clear evidence of an association between 12th-grade students' consumption of energy drinks and their growing-up environment.



2. a. Gender is the explanatory variable. We would like to use gender to explain how students' rate their intelligence compared to their peers.

b. H_0 : No association between gender and intelligence rating.

H_a : Association between gender and intelligence rating.

c.

		Intelligence			Total
Count		Below Average	Average	Above Average	
Gender	Female	437 448.5	2243 1951.7	4072 4351.8	6752
	Male	456 444.5	1643 1934.3	4593 4343.2	6692
Total		893	3886	8665	13444

d. $df = (2 - 1)(3 - 1) = 2$

$$\begin{aligned} \chi^2 &= \frac{(437 - 448.5)^2}{448.5} + \frac{(2243 - 1951.7)^2}{1951.7} + \frac{(4072 - 4351.8)^2}{4351.8} \\ &+ \frac{(456 - 444.5)^2}{444.5} + \frac{(1643 - 1934.3)^2}{1934.3} + \frac{(4593 - 4313.2)^2}{4313.2} \\ &\approx 124.1 \end{aligned}$$

(Answers may vary somewhat depending on the number of decimals used in the expected cell count.)

e. $p \approx 0$. Reject the null hypothesis. There is a statistically significant difference between how males and females rate their intelligence compared to their peers. (In other words, there is an association between gender and intelligence rating.)

3. a. H_0 : No association between intelligence rating and average grades.
 H_a : Association between intelligence rating and average grades.

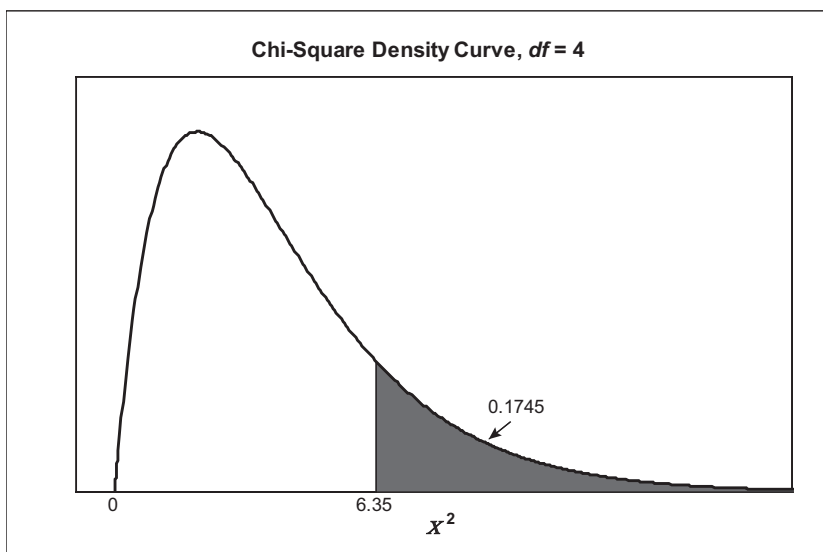
b.

		Average Grade			Total
		A	B	C or Below	
Intelligence	Above	2886 2894.9	4044 4055.8	1387 1366.2	8317
	Average	1335 1323	1881 1853.6	585 624.4	3801
	Below	305 308	416 431.6	164 145.4	885
Total		4526	6341	2136	13003

c. $df = (3 - 1)(3 - 1) = 4$

$$\chi^2 = \frac{(2886 - 2894.9)^2}{2894.9} + \dots + \frac{(164 - 145.4)^2}{145.4} \approx 6.35$$

As shown below, $p \approx 0.174$



- d. We would expect to see a value from a chi-square distribution with $df = 4$ as or more extreme than 6.35 roughly 17.4% of the time. So, this is a somewhat common occurrence. It does not provide strong evidence against the null hypothesis. Generally strong evidence means that the percentage should be below 5%.

4. a. H_0 : No association between gender and hours worked/week.
 H_a : Association between gender and hours worked/week.

- b. $\chi^2 = 12.705$; $p = 0.005 < 0.05$. Therefore, the results are significant. There is an association between gender and hours worked per week. (Note: The practical significance is another matter and cannot be determined by a p -value.)

c. The biggest discrepancy in work patterns is that a higher percentage of males did not work (43.52%) compared to females (40.59%). Furthermore, in every category of hours worked/week, there is a higher percentage of females than males.

REVIEW QUESTIONS SOLUTIONS

1. a. H_0 : No association between habitat use and eel species.

H_a : Association between habitat use and eel species.

b.

Count		Spotted	Purplemouth	Total
Habitat Use	G	127 142.8	116 100.2	243
	S	99 97.5	67 68.5	166
	B	264 249.7	161 175.3	425
Total		490	344	834

c. Here are the calculations for the chi-square test statistic:

$$\begin{aligned}\chi^2 &= \frac{(127 - 142.8)^2}{142.8} + \frac{(116 - 100.2)^2}{100.2} + \frac{(99 - 97.5)^2}{97.5} \\ &\quad + \frac{(67 - 68.5)^2}{68.5} + \frac{(264 - 249.7)^2}{249.7} + \frac{(161 - 175.3)^2}{175.3} \\ &\approx 6.28\end{aligned}$$

The degrees of freedom are: $df = (3 - 1)(2 - 1) = 2$.

Using software, $p \approx 0.043$.

Since $p < 0.05$, we reject the null hypothesis and conclude that there is an association between habitat use and moray eel species.

d. Column percentages are more appropriate. The explanatory variable is the eel species. So, we should compare the conditional distributions of habitat use for each species of moray eel.

		Spotted	Purplemouth
Habitat Use	G	25.9%	33.7%
	S	20.2%	19.5%
	B	53.9%	40.8%
Total		100%	100%

We learn that a majority (53.9%) of the spotted moray eels were found in border habitats compared to only 46.8% of the purplemouth moray eels.

2. a. Educational attainment is the explanatory variable and voting is the response variable. We expect that a person's highest educational attainment will shed light on whether or not they voted in the 2012 elections.

b. H_0 : No association between education and voting.

H_a : Association between education and voting.

c.

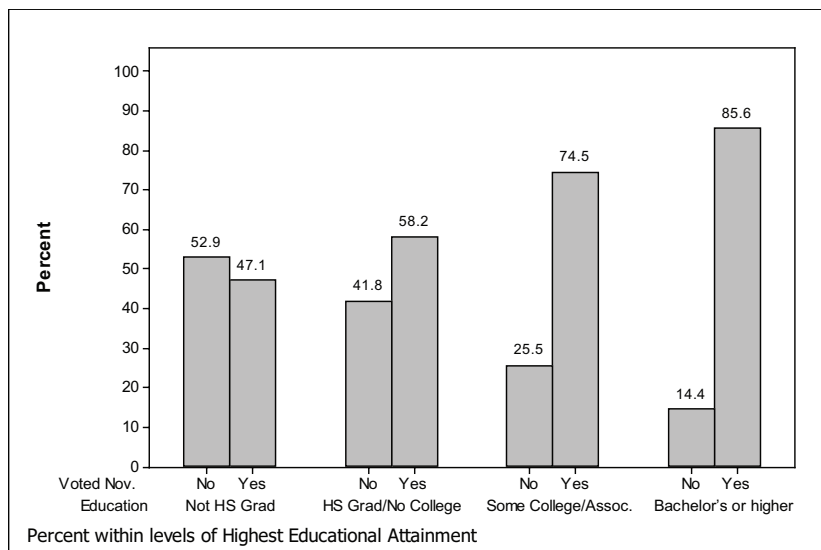
		Voted Nov. 2012		Total
		Yes	No	
Highest Educational Attainment	Not HS Grad	57	64	121
	Expected	84.5	36.5	
	HS Grad/No College	227	163	390
	Expected	272.3	117.7	
	Some College/Associate's	303	51	364
	Expected	254.1	109.9	
	Bachelor's or Higher	303	51	354
	Expected	247.1	106.9	
Total		858	371	1229

$$d. \chi^2 = \frac{(57 - 84.5)^2}{84.5} + \frac{(64 - 36.5)^2}{36.5} + \dots + \frac{(51 - 106.9)^2}{106.9} \approx 100.1$$

$$df = (4 - 1)(2 - 1) = 3; p \approx 0.000$$

Since $p < 0.5$, the results are significant. There is a relationship between these two variables.

e. Since the explanatory variable is highest educational attainment, the chart below represents graphically the conditional distributions of voting for each level of highest educational attainment.



As the level of highest educational attainment increases, so does the participation in voting. More educated people are more likely to vote than those who are not educated.

3. a.

Count		Female	Male	Total
Energy Shots Consumed Per Day	None	896 888.89	938 945.11	1834
	Less than one	63 64.46	70 68.54	133
	One	16 16.96	19 18.04	35
	Two	5 10.18	16 10.82	21
	Three	7 5.82	5 6.18	12
	Four	1 0.48	0 0.52	1
	Five or Six	4 3.88	4 4.12	8
	Seven or more	4 5.33	7 5.67	11
Total		996	1059	2055

b. No, the guidelines are not satisfied. There are two cells that have counts below 1 (0.48 and 0.52). In addition, there are 4 cells with counts less than 5, which is 25% of the cells.

c. Sample answer (students may decide to combine different categories):

Count		Female	Male	Total
Energy Shots Consumed Per Day	None	896 888.9	938 945.1	1834
	One or Less	79 81.4	89 86.6	168
	Two or Three	12 16	21 17	33
	Four or more	9 9.7	11 10.3	20
Total		996	1059	2055

d. Sample answer is based on sample answer to (c): $\chi^2 = 2.282$; $p \approx 0.52$.

There is insufficient evidence to reject the null hypothesis. There is insufficient evidence to indicate that there is a linkage between amounts of energy drink shots consumed and gender.

Unit 30: Inference for Regression



PREREQUISITES

Students should be familiar with the topic of least-squares regression lines, which was covered in Unit 11, Fitting Lines to Data. Students need some background in significance tests, confidence intervals, and the t -distributions. Coverage of the material in Unit 26, Small Sample Inference for One Mean, is a prerequisite for this unit. In addition, students must have some familiarity with material on normal distributions covered in Units 8 and 9, Normal Calculations and Checking Assumptions of Normality, respectively.

ADDITIONAL TOPIC COVERAGE

Additional coverage of inference for simple linear regression can be found in *The Basic Practice of Statistics*, Chapter 24, Inference for Regression. To extend the topic, see Chapter 28, Multiple Regression (in Optional Companion Chapters).

ACTIVITY DESCRIPTION

In the activity, students use clues left by a thief – his/her step length and forearm length – to estimate the height of the thief. But first students must build models, one to estimate height from forearm length and the other to estimate height from step length. Students now have two competing models to predict the thief's height. Their choice will depend on each model's standard error of the estimate, s_e .

MATERIALS

Rulers and/meter stick (optional, if you plan to collect your own data).

In this activity, data on height, step length, and forearm length from 9th and 10th grade students are provided. However, your class can collect data of their own and either add it to the data

provided in the activity or substitute it for the data provided. If your students are older – college students or 11th- or 12th-grade students – your data might have a different pattern than the data contained in the activity (especially for the male students).

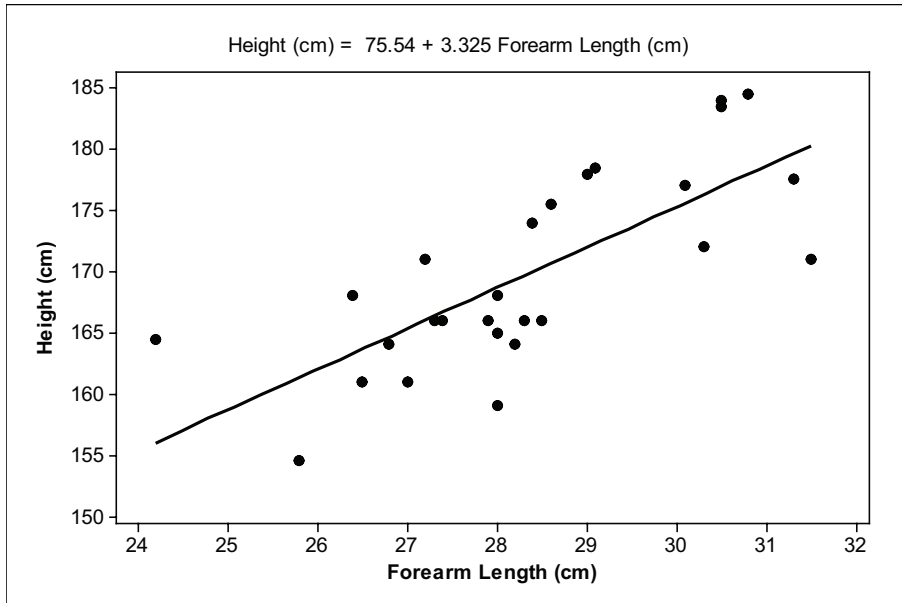
Particularly if you add data, you could extend this project by creating separate models for each gender. The footprints appeared to be from male sneakers (but sometimes females wear male sneakers). So using a model developed from the male student data could have a smaller s_e than a model using all students (both male and female).

THE VIDEO SOLUTIONS

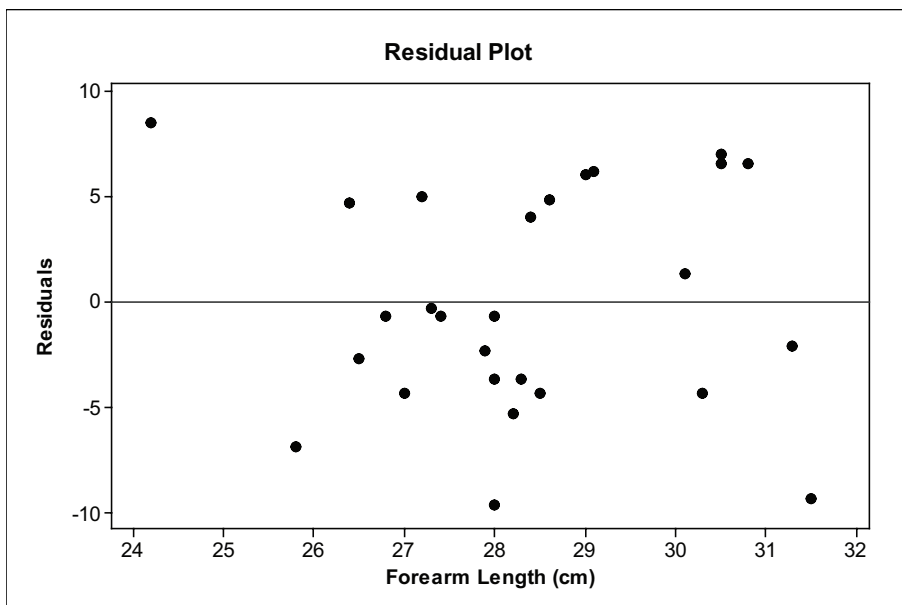
1. Peregrine falcons were not able to hatch their eggs due to eggshell thinning. Scientists believed the cause of eggshell thinning was due to DDT or its derivative DDE.
2. The log-concentration of DDE was the explanatory variable and eggshell thickness was the response variable.
3. The pattern of the data showed a negative, linear relationship.
4. A line fit to data from the entire population.
5. Because they depend on the sample data and can vary from sample to sample.
6. H_o : Amount DDE and eggshell thickness have no linear relationship (or $\beta = 0$).
 H_a : Amount DDE and eggshell thickness have a negative linear relationship (or $\beta < 0$).
7. The null hypothesis was rejected.
8. Yes, their populations have increased since DDT was banned in the U.S. and Western European countries.

UNIT ACTIVITY SOLUTIONS

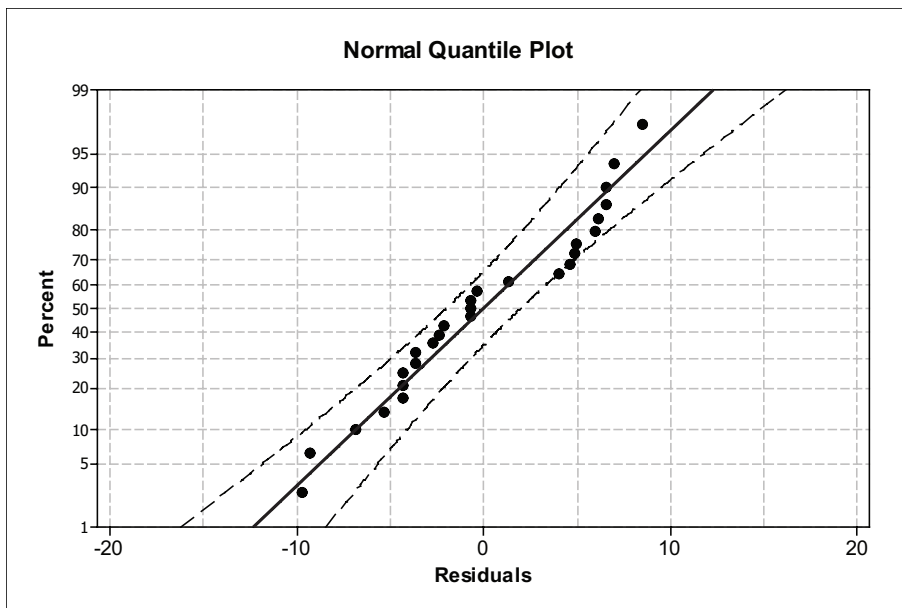
1. a. Equation of least-squares line: Height = 75.53 + 3.325 Forearm Length.



b. The dots in the residual plot below appear randomly scattered with no strong patterns. The vertical spread of the dots appears to stay the same as x increases. So, Conditions 1 and 4 appear to be reasonably satisfied.



The normal quantile plot of the residuals shown below is fairly linear. The dots stay pretty much within the curved bands that Minitab adds to the plot. So, there does not appear to be a strong departure from normality. There is no evidence of an extreme outlier.



Finally, these data were a random sample of 9th- and 10th-grade students. So, the heights for fixed values of the explanatory variables are independent of each other.

In conclusion, all four conditions for inference are reasonably met.

c. $s_e \approx 5.386$ cm

2. a. First, we need to compute the standard error of the slope, $s_b = \frac{s_e}{\sqrt{\sum (x - \bar{x})^2}}$.

We know the numerator but need to determine the denominator.

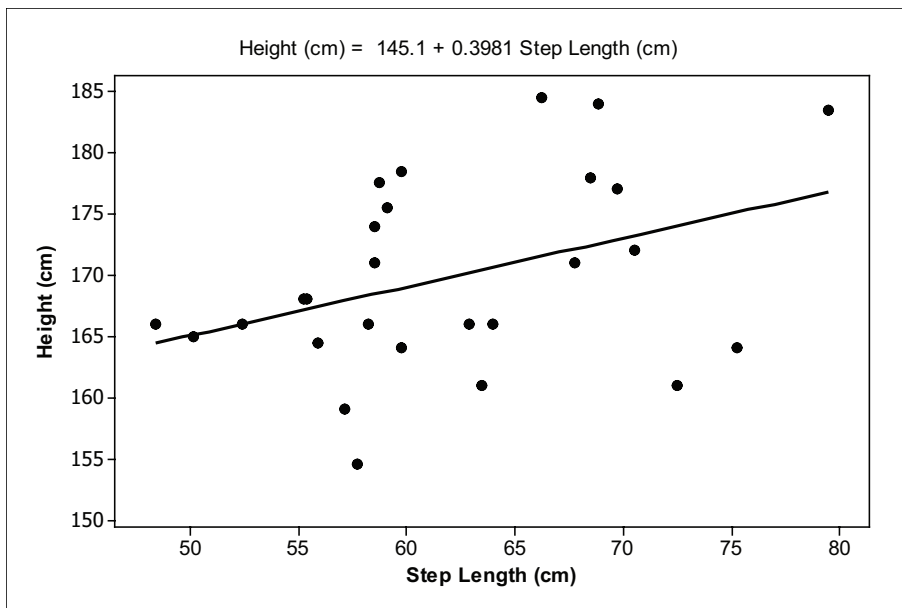
We used Excel: $s_b = \frac{5.386}{\sqrt{80.9067}} \approx 0.599$.

$$t = \frac{3.325 - 0}{0.599} \approx 5.55; df = 25; p = 0.000.$$

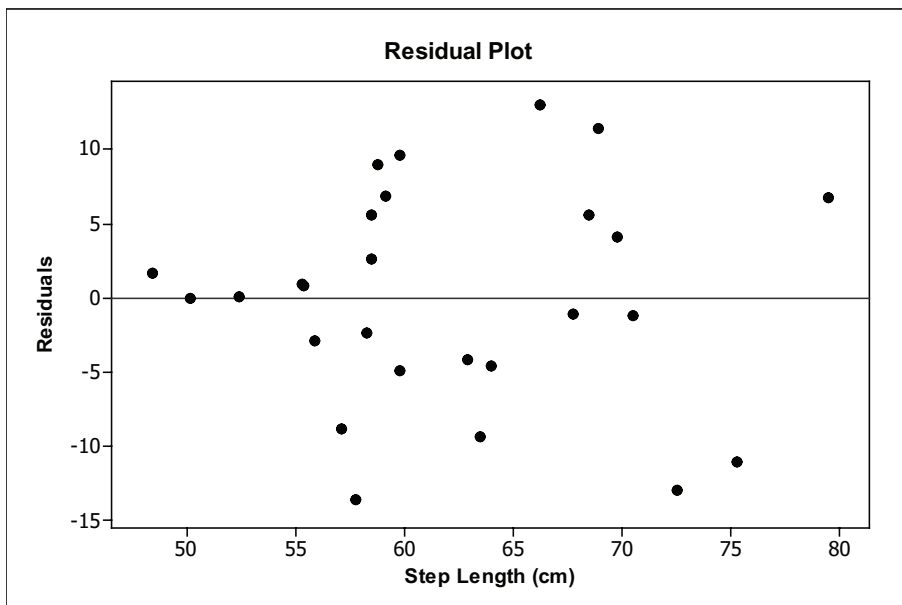
Reject the null hypothesis. Conclude $\beta > 0$; in other words, there is a positive linear relationship between height and forearm length.

b. We use a t -table to find $t^* = 2.060$. Now, we are ready to calculate a 95% confidence interval for β : $3.325 \pm (2.060)(0.599) \approx 3.325 \pm 1.234$, or from 2.091 to 4.559. Therefore, for each 1 centimeter increase in forearm length, we would expect an increase in height of between 2.09 cm and 4.56 cm.

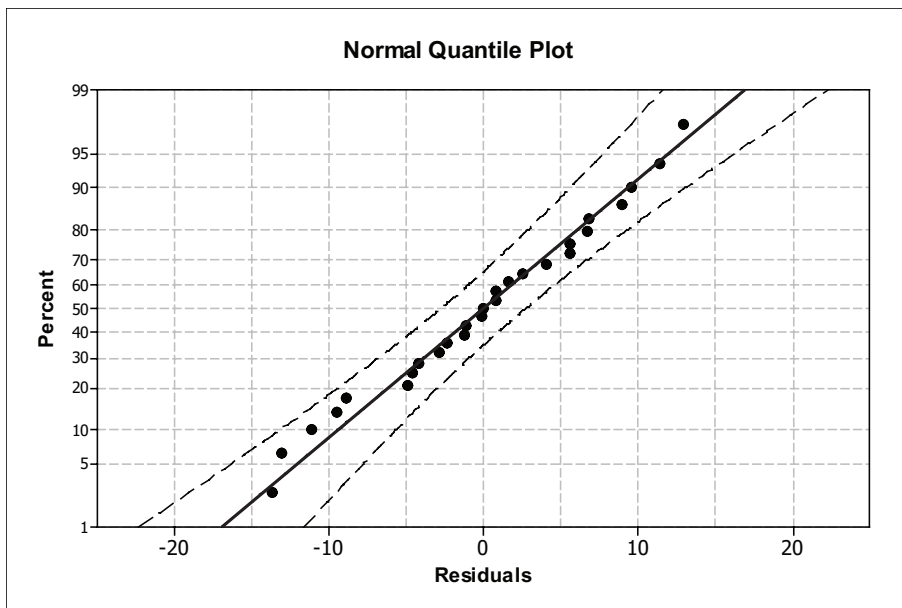
3.a. Equation of least-squares line: Height = 145.1 + 0.3981 Step Length.



b. The dots in the residual plot below appear randomly scattered with no strong patterns. The vertical spread of the dots appears to stay the same as x increases with one exception. The vertical spread of the first three dots is very small – indicating less variability in height for people with the smallest step lengths. However, the vertical spread in the rest of the plot looks good. So, Conditions 1 and 4 appear to be reasonably satisfied.



The normal quantile plot of the residuals shown below is fairly linear. It is a reasonable assumption that the residuals follow an approximately normal distribution.



Finally, these data were a random sample of 9th- and 10th-grade students. So, the heights for fixed values of the explanatory variables are independent of each other.

In conclusion, all four conditions for inference are reasonably met.

c. $s_e \approx 7.424$ cm

4. a. First, we need to compute the standard error of the slope, $s_b = \frac{s_e}{\sqrt{\sum (x - \bar{x})^2}}$.

We know the numerator but need to determine the denominator.

We used Excel to calculate s_b : $s_b = \frac{7.424}{\sqrt{1527.64}} \approx 0.1899$

$t = \frac{0.3981 - 0}{0.1899} \approx 2.096$; $df = 25$; $p = 0.02318$. Reject the null hypothesis.

Conclude that there is a positive linear relationship between height and step length; in other words, $\beta > 0$.

b. We use a t -table to find $t^* = 2.060$. Now, we are ready to calculate the 95% confidence interval for β : $0.3981 \pm (2.060)(0.1899) \approx 0.398 \pm 0.391$, or from 0.007 to 0.789. Therefore, for each 1 centimeter increase in step length, we would expect an increase in height of between 0.007 cm and 0.789 cm.

5. a. The model based on the explanatory variable forearm length will produce more precise estimates. The standard error of the estimate for the forearm length model is $s_e \approx 5.386$ compared to $s_e \approx 7.424$ for the model based on step length.

b. Students may decide to answer this question in several different ways.

Sample answer: Because the least-squares line based on forearm length is associated with a smaller value for s_e , we decided to use that linear model for our prediction. For the point estimate we used the forearm length of 26.5 cm, midway between 26 cm and 27 cm and for the smallest and largest estimate, we used the forearm lengths of 26 cm and 27 cm. These three point estimates allowed us to complete the sentence as follows:

We predict that the thief is 163.7 cm tall. But the thief might be as short as 162.0 cm or as tall as 165.3. (From about 5'3 $\frac{3}{4}$ " to 5'5".)

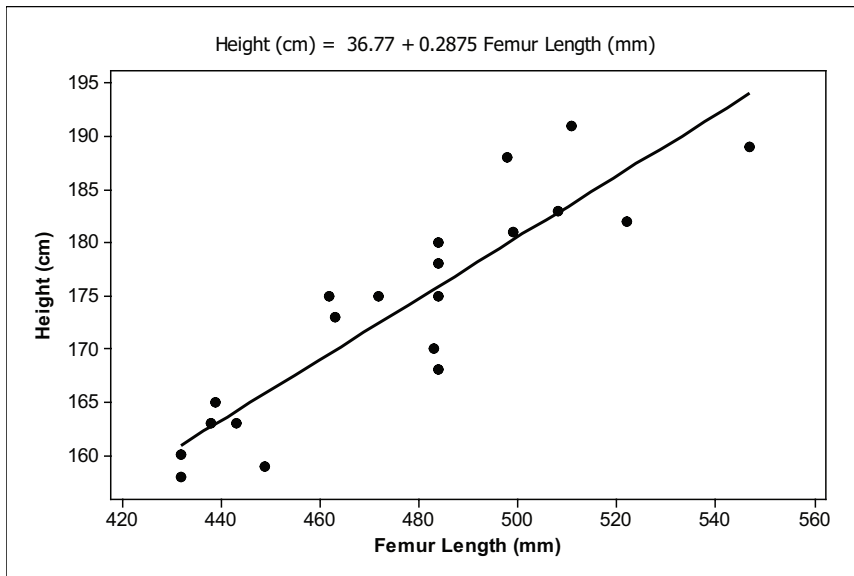
However, in our statement above, we did not use the fact that $s_e \approx 5.386$ (we only used it to select the model). If we want to give more conservative bounds, we would subtract s_e from our lower bound and add it to our upper bound to give from between 156.7 cm to 170.7 cm. (From 5'1 $\frac{5}{8}$ " to 5'6".) However, if we want to give very conservative bounds, we could subtract $2s_e$ from our lower bound and add it to our upper bound, giving 151.9 cm to 176.1 cm. However, that bound is probably too wide to be of much use in finding the thief. (From around 5' to 5'9".)

(Another possible answer might be to give 163.7 cm for the point estimate and 163.7 cm ± 5.386 cm or 163.7 cm ± 10.772 cm for the bounds.)

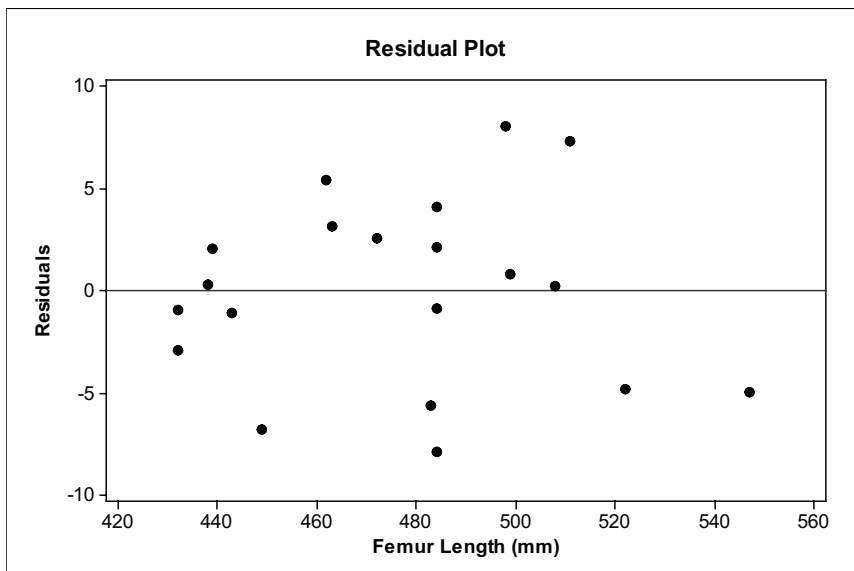
EXERCISE SOLUTIONS

1. a. The pattern appears to be linear and the association to be positive. (See solution to 1(b) for the scatterplot.)

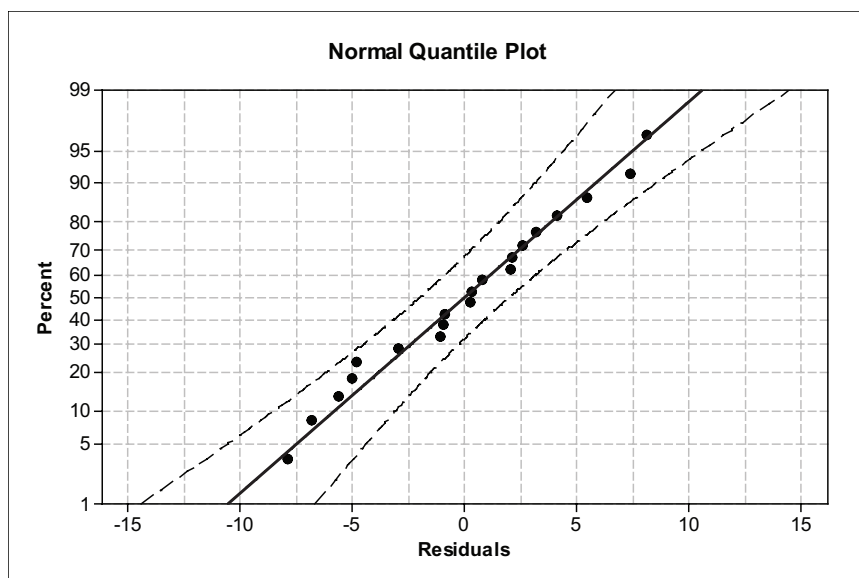
b. Equation of the least-squares line: $y = 36.77 + 0.2875x$



c. Sample answer: The dots in the residual plot appear to be randomly scattered with a good split of points above and below the horizontal axis. There is some evidence that the vertical spread of the residuals is smaller for small femur lengths than for larger femur lengths. However, that difference seems relatively small and could be due to the small sample size. So, conditions 1 and 4 seem to be reasonably satisfied.



A normal quantile plot of the residuals is fairly linear. So, it is reasonable to assume the residuals follow a normal distribution. Condition 2 appears to be satisfied.



Finally, these data are a random sample. So, the heights are independent of each other. We conclude that Conditions 1 – 4 are reasonably satisfied.

2. a. The residuals are listed below:

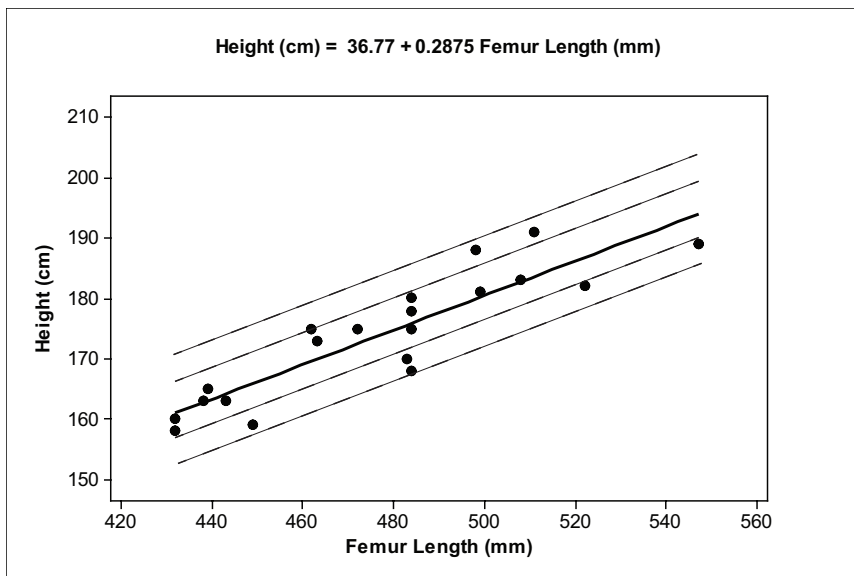
-2.95031	8.07699	3.13827	-1.11242	7.33994
-5.00880	2.10150	-4.82217	0.32490	5.42574
-6.83721	0.78953	-7.89850	2.55109	-0.89850
-0.95031	2.03744	-5.61103	4.10150	0.20234

To find the SSE, we square the residuals and then find the sum:

$$SSE = (-2.9531)^2 + \dots + (0.20234)^2 \approx 391.679$$

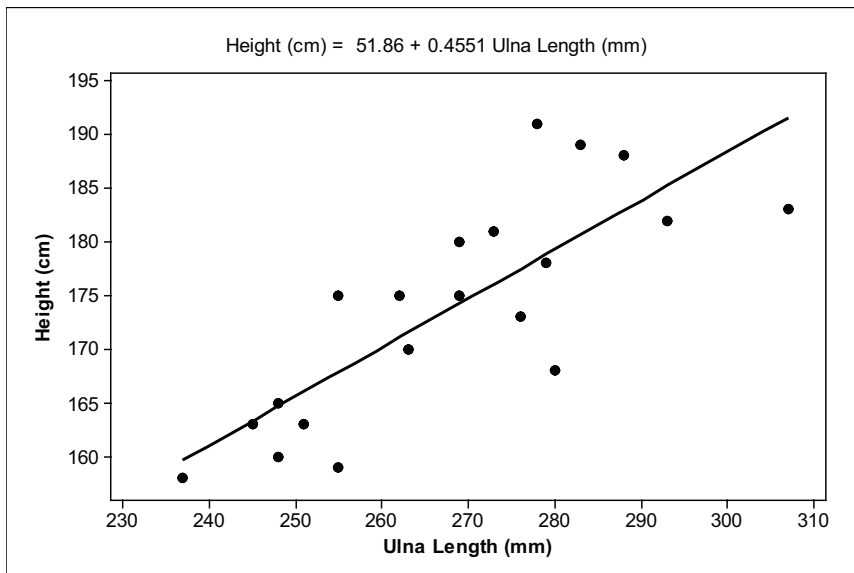
$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{391.679}{20-2}} \approx 4.665 \text{ cm}$$

b. $y = 36.77 + 0.2875x \pm 4.665$; $y = 36.77 + 0.2875x \pm 9.330$



c. In this case, all 20 data points are trapped between the outermost error bands.

3. a. Equation of least-squares line: $y = 51.86 + 0.4551x$



b. The residuals are given below:

-1.7150	5.0755	-4.4635	-0.3557	12.6264
8.3509	-0.8287	-3.2000	-3.0863	3.9078
-8.9066	4.9018	-11.2838	7.0934	0.7222
-4.7210	0.2790	-1.5473	5.7222	-8.5712

SSE = 726.122

$$s_e = \sqrt{\frac{726.112}{20-2}} \approx 6.351 \text{ cm}$$

c. Height prediction from ulna length: $y = 51.86 + 0.4551(287) \approx 182.47$ cm.

Height prediction from femur length: $y = 36.77 + 0.2875(520) \approx 186.27$ cm.

The prediction based on femur length is likely to be more reliable. The s_e for the model based on femur length is 4.665 cm compared to 6.351 cm for the model based on ulna length.

4. a. $t = 8.63$; $df = 18$; $p \approx 0$; Reject the null hypothesis and conclude that there is a positive linear relationship between height and femur length.

b. First, we need to calculate the standard error of the slope, s_b : $s_b = \frac{s_e}{\sqrt{\sum (x - \bar{x})^2}}$

We use Excel to calculate $\sum (x - \bar{x})^2 = \sum (x - 476.70)^2 \approx 19598.2$

$$s_b = \frac{4.665}{\sqrt{19598.2}} \approx 0.0333$$

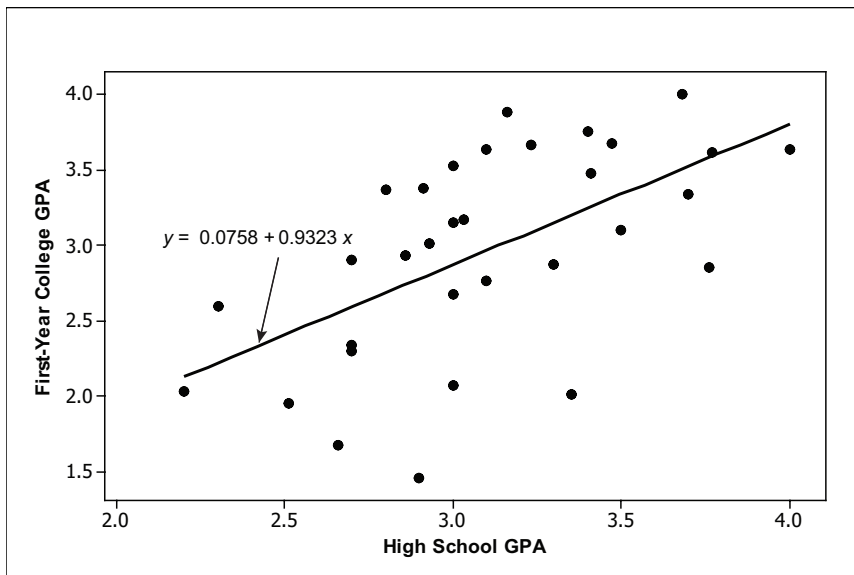
Next, we use a t -table to find the t -critical value for a 95% confidence interval: $t^* = 2.101$.

The 95% confidence interval for β : $0.2875 \pm (2.101)(0.0333) \approx 0.2875 \pm 0.0700$,
or from 0.2175 to 0.3575.

REVIEW QUESTIONS SOLUTIONS

1. From Figure 30.14, we note that the dots appear randomly scattered, roughly half above the horizontal axis and roughly half below. This indicates that Condition 1, linearity, is satisfied. In addition, the vertical spread of the dots on this plot appears roughly the same as x increases. Therefore, Condition 4, equal standard deviations, is satisfied. A plot of the residuals is somewhat linear – at least the dots stay within the curved bands provided by Minitab. So, it is reasonable to assume the residuals follow at least an approximate normal distribution. Finally, the eggs were from a random sample, so the egg thicknesses were independent of each other. Conditions 1 – 4 are reasonably satisfied and therefore, the results of the inference shown in the video are trustworthy.

2. a. The form appears linear. The relationship is positive.

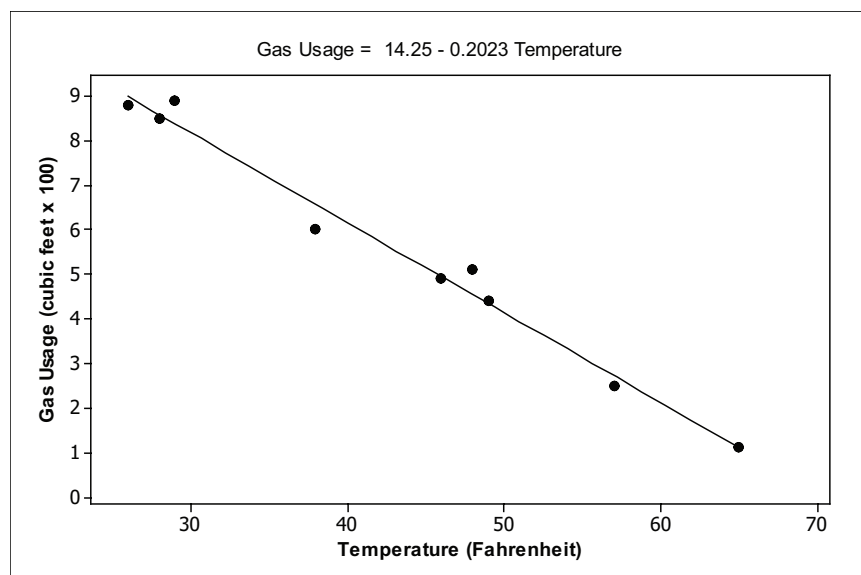


b. See (a) for scatterplot and line: $y = 0.0758 + 0.9323x$

c. $t = \frac{0.9323 - 0}{0.2370} \approx 3.93$; $df = 32 - 2 = 30$

d. $p \approx 0.0002$; reject the null hypothesis and conclude that there is a positive linear relationship between students' high school GPAs and their first-year college GPAs.

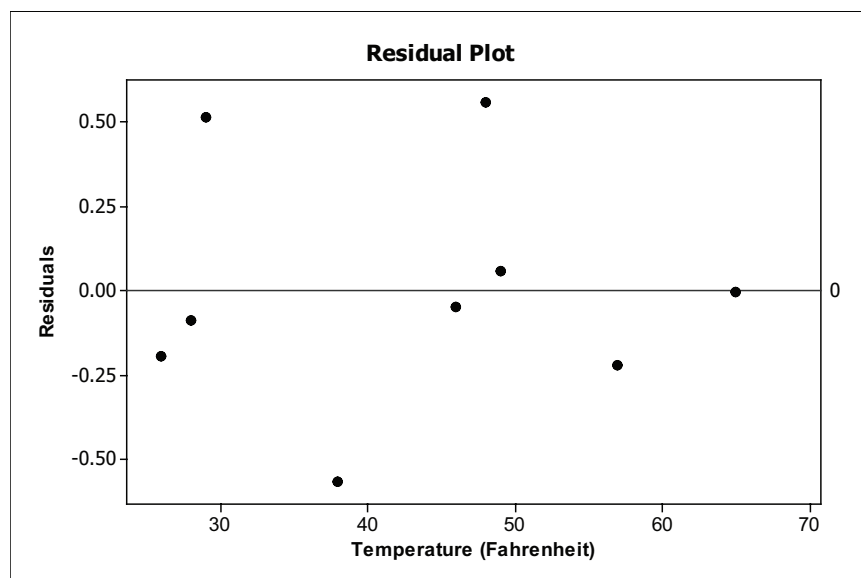
3. a. The relationship appears to be a negative linear relationship.



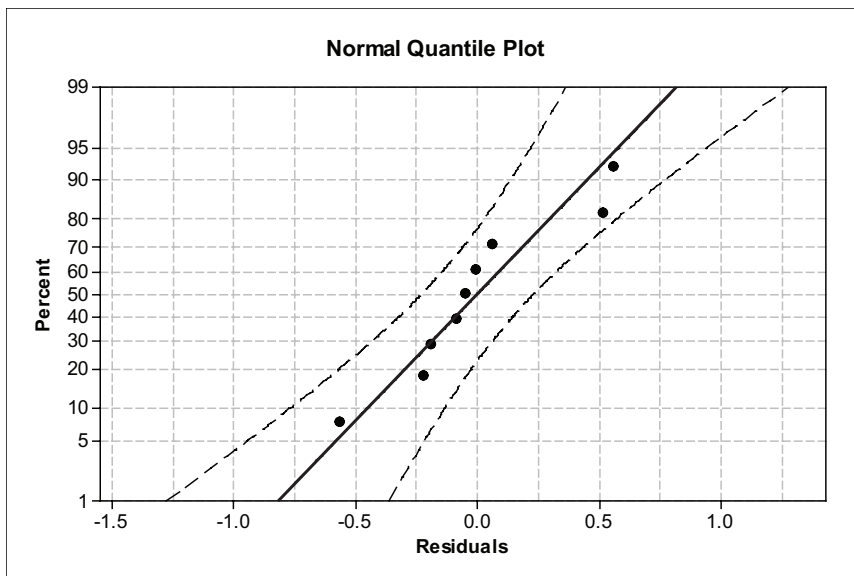
b. See scatterplot in solution to 2(a).

The equation of the least-squares line is $y = 14.25 - 0.2023x$.

c. The dots in the residual plot below appear to be randomly scattered with some above and below the horizontal axis. So, the line is adequate to describe the pattern in the data and Condition 1 is satisfied. In addition, the vertical spread of the dots stays roughly the same as temperature increases. So, Condition 4 is satisfied.



The pattern of the normal probability plot below is roughly linear. Given that the dots stay within the curved bands provided by Minitab, we can say that there are no strong departures from normality in the residuals. So, Condition 4 is reasonably satisfied.



Observed gas usage for different months should depend only on temperature and be independent of each other.

So, given the conditions are satisfied, and we can proceed with inference.

$$d. s_e = \sqrt{\frac{SSE}{n-2}}$$

We begin our calculations with the residuals, which were calculated using technology:

0.555956	-0.048548	-0.566567	0.513162	-0.193595
-0.089090	0.058209	-0.223773	-0.005754	

$$SSE = (0.555956)^2 + \dots + (-0.005754)^2 = 0.994689$$

$$s_e = \sqrt{\frac{0.994689}{9-2}} \approx 0.37696$$

$$s_b = \frac{s_e}{\sqrt{\sum (x - \bar{x})^2}}; \bar{x} = 42.89;$$

the deviations of the temperatures, x , from their mean is given below:

5.11	3.11	-4.89	-13.89	-16.89	-14.89	6.11	14.11	22.11
------	------	-------	--------	--------	--------	------	-------	-------

$$\sum (x - \bar{x})^2 = (5.11)^2 + \dots + (22.11)^2 = 1484.89$$

$$s_b = \frac{0.37696}{\sqrt{1484.89}} \approx 0.00978$$

e. Sample answer: We would expect to need more gas as the temperature gets colder. So, it makes sense to test $H_a : \beta < 0$; in other words, the alternative is for a negative linear relationship.

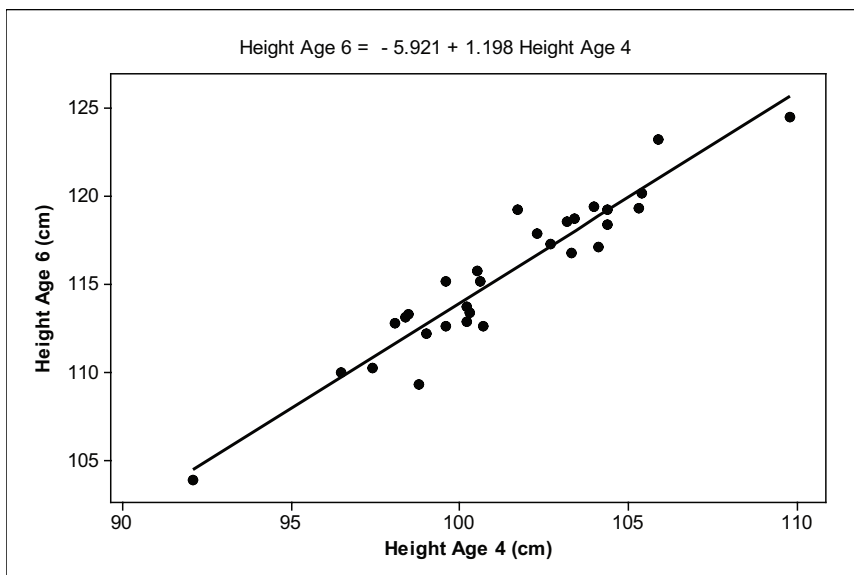
$$t = \frac{-0.2023 - 0}{0.00978} \approx -20.68; df = 7; p \approx 0.$$

Reject the null hypothesis and conclude that there is a negative linear relationship between temperature and gas usage.

f. We use the t -table to find t^* , the t -critical value from a t -distribution with $df = 7$. We find $t^* = 2.365$. Now we have everything that we need to construct a 95% confidence interval for β : $-0.2023 \pm (2.365)(0.00978) \approx -0.202 \pm 0.023$, or from -0.225 to -0.179.

Interpretation: For each 1°F increase in temperature, the average daily gas usage decreases by between 0.179 cubic feet $\times 100$ and 0.225 cubic feet $\times 100$.

4. a. The equation of the least-squares line is $y = -5921 + 1.198x$. The scatterplot and a graph of the least-squares line appear below.



b. First, we use a t -table, $df = 28$, to determine $t^* = 2.048$

95% confidence interval for β : $1.198 \pm (2.048)(0.07437) \approx 1.198 \pm 0.152$, or from 1.046 to 1.350.

Interpretation in context: For each 1 cm increase in height at age 4, we expect an increase of between 1.046 cm and 1.350 cm in height at age 6.

Unit 31: One-Way ANOVA



PREREQUISITES

Students should be familiar with comparative boxplots covered in Unit 5, Boxplots. In addition, students need to have a background on significance tests. At a minimum, they should be familiar with the material covered in Unit 25, Tests of Significance, and Unit 27, Comparing Two Means. An understanding of experiments contained in Unit 15, Designing Experiments, would also be useful.

ADDITIONAL TOPIC COVERAGE

Additional coverage of ANOVA can be found in *The Basic Practice of Statistics*, Chapter 25, One-Way Analysis of Variance: Comparing Several Means.

ACTIVITY DESCRIPTION

In this activity, students conduct three experiments to see if changing control settings in a manufacturing process affects the mean thickness of a product, polished wafers used to make microchips. Students use Wafer Thickness from the Interactive Tools to collect data. In each experiment, two of the three Control settings are fixed at level 2, the middle setting, and the third is varied from level 1, to 2, to 3. Samples of size 10 are collected using Real Time mode so that students can watch the data being collected. The data can be copied by hand and then entered into software or graphing calculators for analysis. Students can also save the data in CVS format and then transfer their individual data sets into statistical or spreadsheet software for analysis.

Because the applet generates random data, each student (or group of students) will be working with different data. However, the results should be similar:

- For Control 1, the mean thickness differs among control levels.
- For Control 2, there is insufficient evidence to conclude that mean thickness differs among the control levels.

- For Control 3, the underlying assumption of equal standard deviations is not satisfied. Hence, the data are not analyzed by ANOVA.

This would be a good opportunity to show students that repeating the same experiment results in a different value for the F -statistic. Generally, the conclusions will be the same. However, it is possible that some students' conclusions will differ from the majority due to sampling variability.

MATERIALS

Students need access to the Wafer Thickness tool.

THE VIDEO SOLUTIONS

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

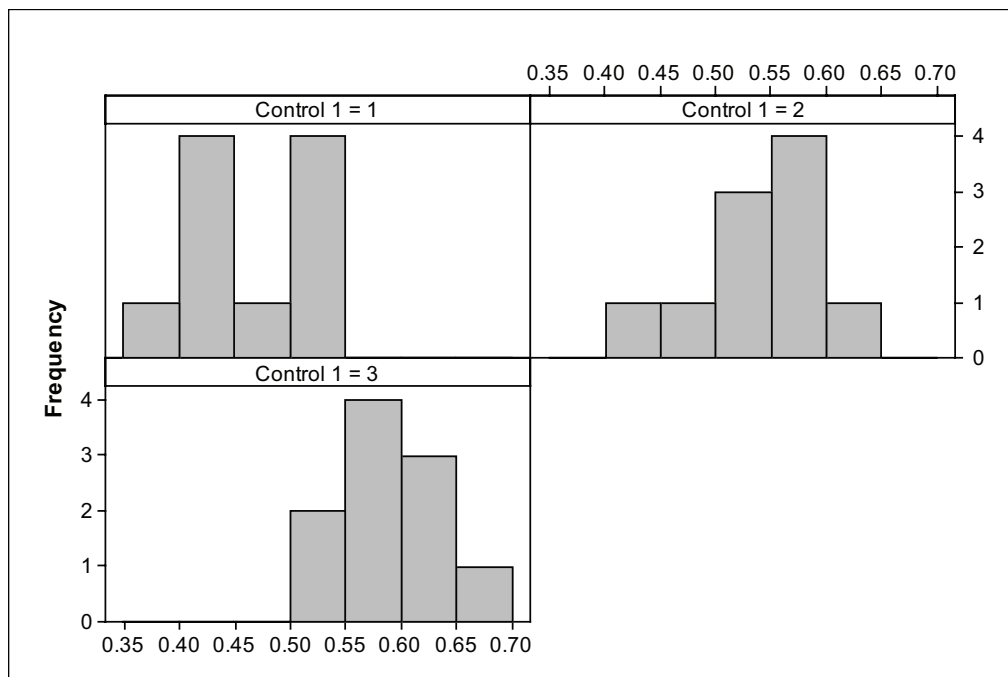
1. The average of all of the guesses will probably be more accurate than most of the individual guesses.
2. The weights of the clipboards differed. One weighed around one pound, another around two pounds, and a third around three pounds.
3. The population mean estimates are the same regardless of which clipboard is being held:
 $H_0 : \mu_1 = \mu_2 = \mu_3$.
4. ANOVA results in an F -statistic.
5. No. The p -value was above 0.05. Hence, he could not reject the null hypothesis based on the data.
6. Yes. The mean guess from the crowd was around \$100 off from the actual amount of money in the jar. But that was closer to the actual amount of money in the jar than the guesses of three-quarters of the students.

UNIT ACTIVITY:

CONTROLLING WAFER THICKNESS SOLUTIONS

1. a. - c. Sample answers based on sample data below.

Control 1 = 1	Control 1 = 2	Control 1 = 3
0.448	0.483	0.54
0.427	0.617	0.596
0.527	0.508	0.577
0.508	0.515	0.609
0.461	0.405	0.632
0.507	0.554	0.631
0.388	0.531	0.66
0.409	0.587	0.503
0.404	0.568	0.583
0.512	0.555	0.583



d. Sample answer: It appears that mean thickness increases as the level of Control 1 is increased. Below are the sample means. The difference among the sample means appears large compared to the variability within each sample. Based on the histograms, although there is overlap between data values collected when Control 1 = 1 and Control 1 = 3, there is a considerable amount of shift.

Control Level	Sample Mean	Standard Deviation
1	0.459	0.0515
2	0.532	0.0596
3	0.591	0.0460

e. Output from Minitab:

Source	DF	SS	MS	F	P
Factor	2	0.08785	0.04392	15.85	0.000
Error	27	0.07487	0.00277		
Total	29	0.16271			

The null hypothesis is that the mean thickness of polished wafers was the same regardless of Control 1's setting. The value of F and its p -value are highlighted in the ANOVA table above. The conclusion is that there are differences among the population mean thickness for wafers produced under the levels 1, 2, and 3 of Control 1.

2. a. Sample answer based on sample data in table below.

Control 2 = 1	Control 2 = 2	Control 2 = 3
0.534	0.521	0.472
0.522	0.450	0.574
0.506	0.521	0.496
0.454	0.495	0.623
0.579	0.514	0.620
0.496	0.481	0.539
0.485	0.414	0.520
0.618	0.497	0.514
0.465	0.553	0.539
0.522	0.603	0.574

b. Sample answer: Yes. The standard deviations for the three samples are 0.050, 0.052, and 0.050, which are very close.

c. Output from Minitab is shown below. Values for F and p are highlighted.

Source	DF	SS	MS	F	P
Factor	2	0.00932	0.00466	1.80	0.185
Error	27	0.06991	0.00259		
Total	29	0.07923			

3. a. Sample answer based on sample data in table below.

Control 3 = 1	Control 3 = 2	Control 3 = 3
0.425	0.494	0.456
0.528	0.525	0.473
0.609	0.573	0.505
0.616	0.493	0.473
0.542	0.558	0.464
0.712	0.472	0.442
0.405	0.444	0.463
0.606	0.44	0.448
0.489	0.491	0.489
0.641	0.565	0.436

b. Sample answer: No. The standard deviations for the three samples are 0.098, 0.048, and 0.021. The ratio $0.098/0.021$ is around 4.7, considerably more than twice as large.

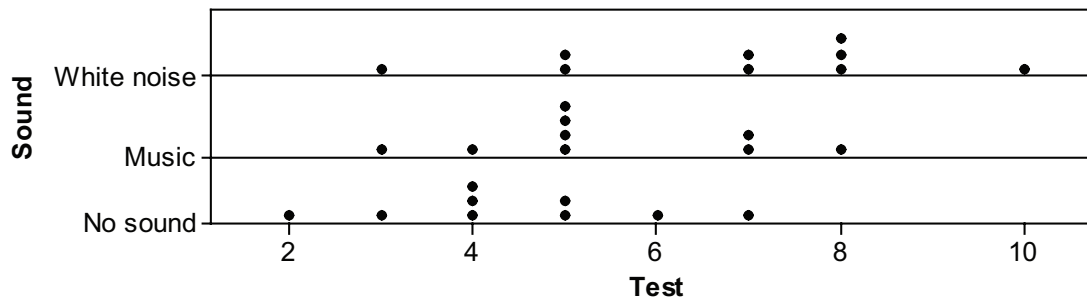
c. The answer to (b) was No – so, we have skipped this part.

EXERCISE SOLUTIONS

1. a.

Group	Mean	Standard Deviation
White Noise	6.778	2.108
Music	5.444	1.59
No Sound	4.444	1.59

b. Sample answer: It looks as if the students who studied with white noise did slightly better than students who studied with no sound. However, it's difficult to tell if that difference is significant. It could be due to chance variation.



c. Hypotheses:

$$H_0 : \mu_{\text{White Noise}} = \mu_{\text{Music}} = \mu_{\text{No Noise}}$$

H_a : There is some difference in the population means.

$$F = \frac{\text{variation among sample means}}{\text{variation among individuals in same sample}} = \frac{MSG}{MSE}$$

To calculate the MSG , we first have to calculate the grand mean, the mean of all the observations: $\bar{x} = 5.556$. Next, we need to calculate the deviations of the group means from the grand mean:

$$6.778 - 5.556 = 1.222; 5.444 - 5.556 = -0.112; 4.444 - 5.556 = -1.112$$

$$MSG = \frac{9(1.222)^2 + 9(-0.112)^2 + 9(-1.112)^2}{3-1} = \frac{24.6813}{2} \approx 12.34$$

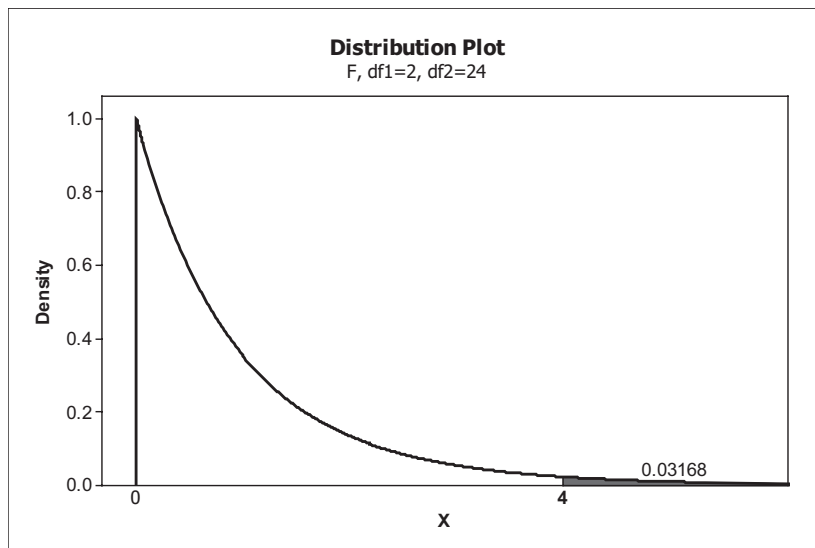
To calculate the MSE , we need the standard deviations for the test scores in each of the groups.

$$MSE = \frac{(9-1)(2.108)^2 + (9-1)(1.590)^2 + (9-1)(1.509)^2}{27-3} = \frac{73.9989}{24} \approx 3.08$$

$$F = 12.34/3.08 \approx 4.01$$

F has numerator degrees of freedom 2 and denominator degrees of freedom 24.

d. Using statistical software, we can calculate the p -value from an F distribution as shown below. This gives a p -value of 0.03168. Conclusion: Reject the null hypothesis; conclude that the mean test scores differ depending on the surrounding sound during the study time.



Compare this value to the output from ANOVA (p -value is highlighted):

Source	DF	SS	MS	F	P
Sound	2	24.67	12.33	4.00	0.032
Error	24	74.00	3.08		
Total	26	98.67			

2. a. Standard deviations: Beef – 31.97; Poultry – 27.16; Veggie – 22.24.

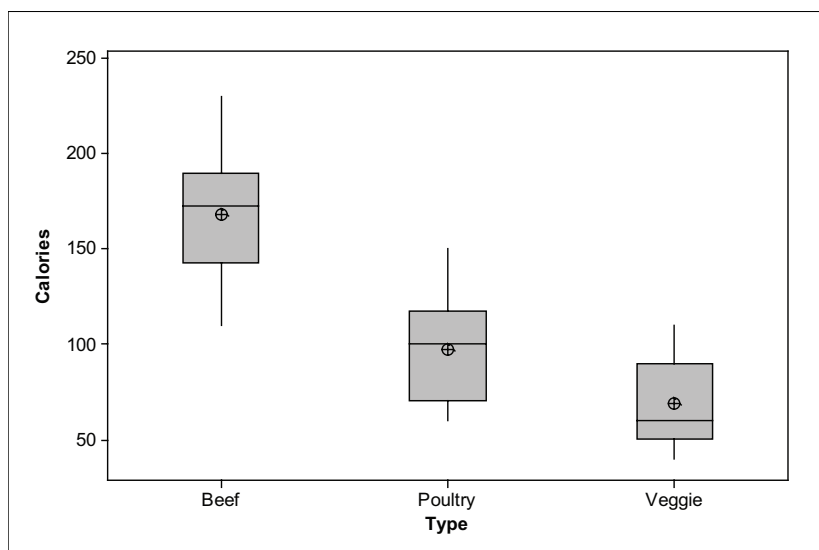
The largest standard deviation is only about 1.4 times the smallest standard deviation. So, it's reasonable to run an ANOVA on these data.

b. Below is the ANOVA table. The value of the F -statistic and p -value are highlighted.

The numerator and denominator degrees of freedom for F are 2 and $19 + 19 + 18 = 56$, respectively. Given the p -value is essentially 0, the conclusion is that there is some difference among the population mean calories of the Beef, Poultry, and Veggie hot dogs.

Source	DF	SS	MS	F	P
Factor	2	101620	50810	67.20	0.000
Error	56	42344	756		
Total	58	143964			

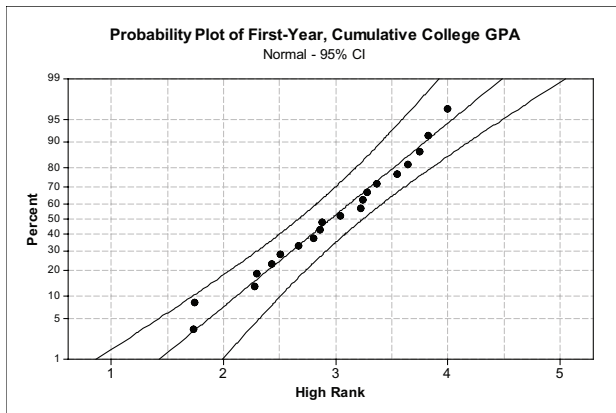
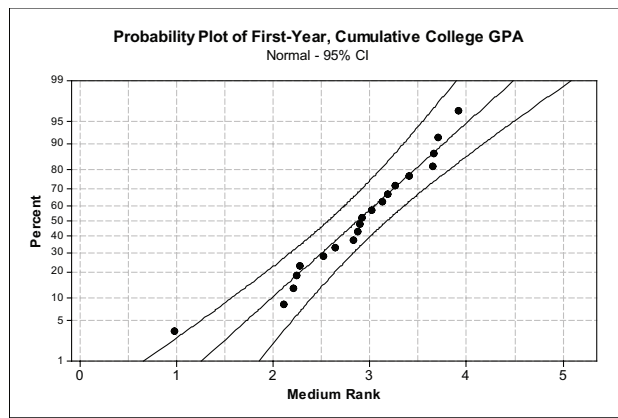
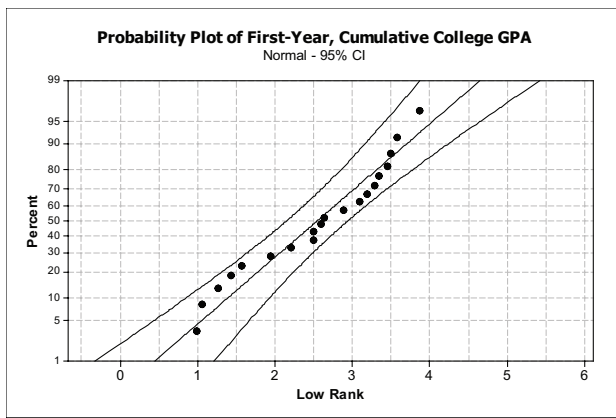
c. From ANOVA, we know there is a significant difference in the mean calorie content among the three types of hot dogs. The boxplot shows that while all three types of hotdogs have some overlap in terms of calorie content, it appears that, on average, beef hotdogs have the highest mean calories, then poultry, and last veggie. There is more overlap in the boxplots between the Poultry and Veggie hotdogs – so, it is not as clear that their mean calorie contents differ significantly. (Keep in mind that ANOVA only says that at least one of the population means differs from the others. It doesn't guarantee that all three population means differ or identify which means differ.)



3. a. Sample means for the three groups are: High – 2.956; Medium – 2.872; Low – 2.546. It appears that as ratings go up, mean GPA goes up as well.

b. The standard deviations for the three groups are: High – 0.657; Medium – 0.695; and Low – 0.904. The highest standard deviation is around 1.38 times the lowest; hence, they are reasonably close. Normal quantile plots for the data in the three groups are shown on the next page. The data in each group appear to be approximately normal; only one data value lies outside of the 95% confidence interval bands.

Students might also make three boxplots and note that there are no outliers and the plots are roughly symmetric (or at least not horribly asymmetric – even though the lower whisker on the Medium rank plot is longer than the upper whisker).



c. $F = 1.63$; p -value = 0.205. There is insufficient evidence to conclude that the mean GPAs differ among the three high school rankings.

4. a. Although the three sample means differ, you need to show that the variability in sample means is large in comparison to the individual variability of scores within each group. So, you cannot conclude that population means differ significantly based only on the three sample means.

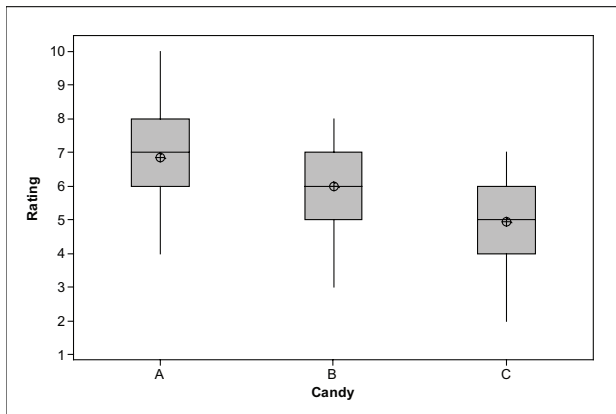
b. The numerator and denominator degrees of freedom are 2 and 2997, respectively.

c. $p = 0.0002344$ or $p \approx 0.000$. There was a significant difference in (population) mean ACL scores among the three majors.

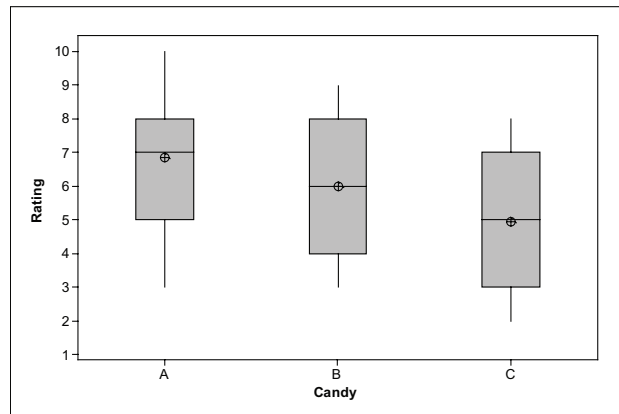
REVIEW QUESTIONS SOLUTIONS

1. a. For both data sets, the mean ratings are 6.87, 6, and 4.93 for candy type A, B, and C, respectively. These means are the same in both data sets. Without knowing anything about the variability of the ratings within each group, it is not possible to determine if the population mean ratings would differ among the three candy types.

b.



Data Set #1



Data Set #2

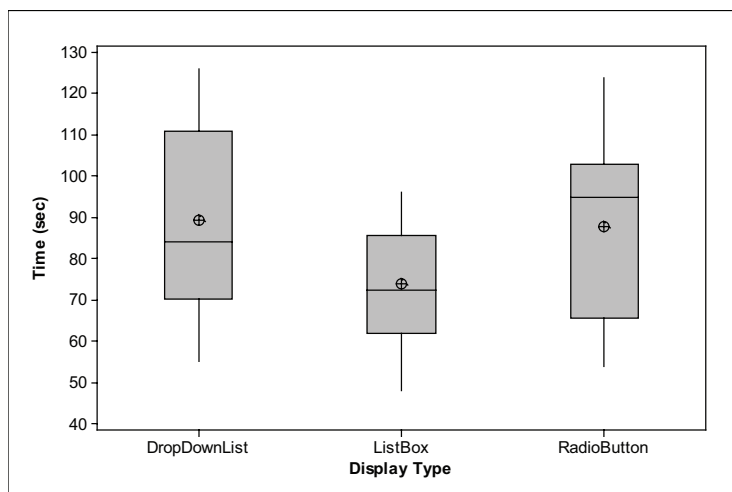
Even though the means and medians for corresponding ratings are the same for both data sets, the difference in means is more likely to be significant based on Data Set #1. In each case the variation in the means is the same. However, the individual variation within each group is larger for Data Set #2. The denominator of the F -statistic will be larger, making the value of F smaller. Hence, it will be less likely that the results based on Data Set #2 will be significant compared to the results based on Data Set #1.

c. Data Set #1: $F(2, 42) = 6.98$; p -value = 0.002. There is a significant difference in the mean ratings based on the type of candy.

Data Set #2: $F(2, 42) = 3.13$; p -value = 0.054, which just misses being significant. There is insufficient evidence to conclude that there are differences in the population mean candy ratings for the three types of candy.

Because the ratings data within each group in Data Set #2 is more spread out than in Data Set #1, it is not surprising that there is insufficient evidence to reject the null hypothesis.

2. b. Sample answer: Although the boxplots are not perfectly symmetric, there is no strong evidence that the times for each display type are strongly skewed. In addition, there are no outliers. The spread of the data set that has the most variability appears to be less than double the spread of the data set with the least variability.



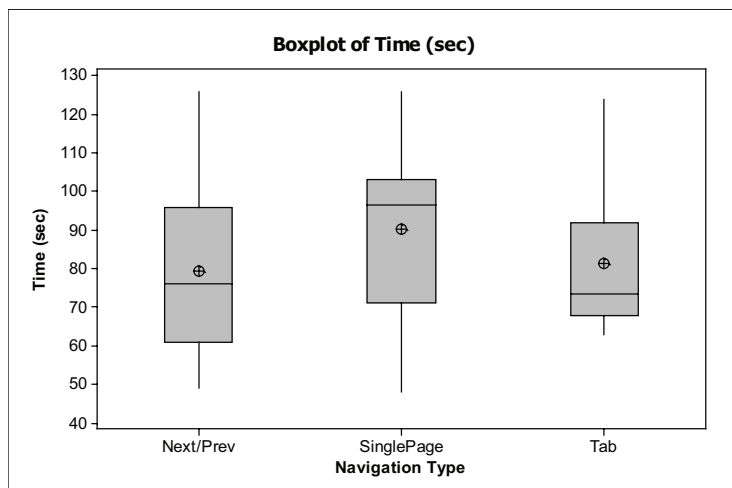
c. The null hypothesis is: $H_0 : \mu_1 = \mu_2 = \mu_3$, that the population mean times do not differ depending on the display type used for answer entry.

Output from Minitab gives $F = 3.26$ and $p = 0.046$ (see below). Hence, we can conclude that there is a significant difference among the mean times.

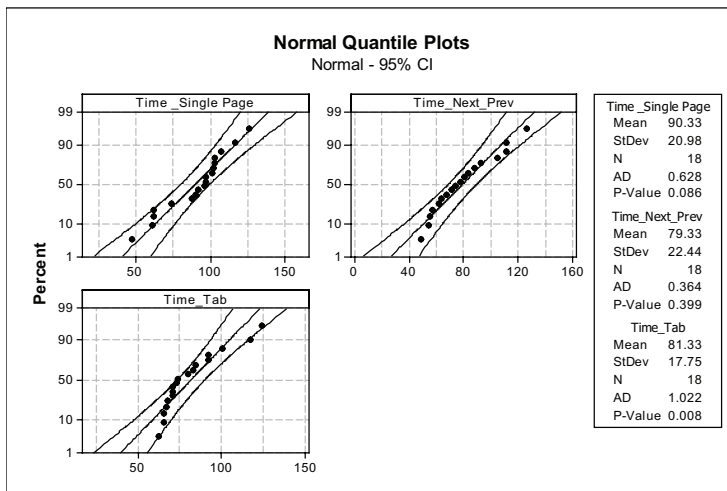
One-way ANOVA: Time (sec) versus Display Type

Source	DF	SS	MS	F	P
Display Type	2	2567	1283	3.26	0.046
Error	51	20071	394		
Total	53	22638			

d. Sample answer: The boxplots don't indicate any outliers. However the times associated with the Tab navigation are rather skewed to the right.



Sample answer: Although the Tab time data appear skewed in the boxplot, the normal quantile plots show all dots within the curved bands. In addition, the ratio of the largest standard deviation to the smallest standard deviation is $20.98/17.75$ or around 1.2, which is quite good. So, it's probably OK to run an ANOVA.



e. $F(2, 51) = 1.47$; $p = 0.239$. Conclusion: There is insufficient evidence to conclude that the three navigation types have an impact on the mean times to complete the questionnaire.

Below is the output from Minitab:

One-way ANOVA: Time (sec) versus Navigation Type

Source	DF	SS	MS	F	P
Navigation Type	2	1236	618	1.47	0.239
Error	51	21402	420		
Total	53	22638			

f. Sample answer: For answer entry, use the List Box for Display Type. The sample mean time to complete the surveys was smallest for the List Box display type. The ANOVA did not show that mean times differed significantly with Navigation Type.

3. a. It is reasonable to assume that the hourly rate standard deviations from the four regions of the country are the same. The ratio of the largest standard deviation to the smallest standard deviation is $9.289/6.381$, which is less than 1.5.

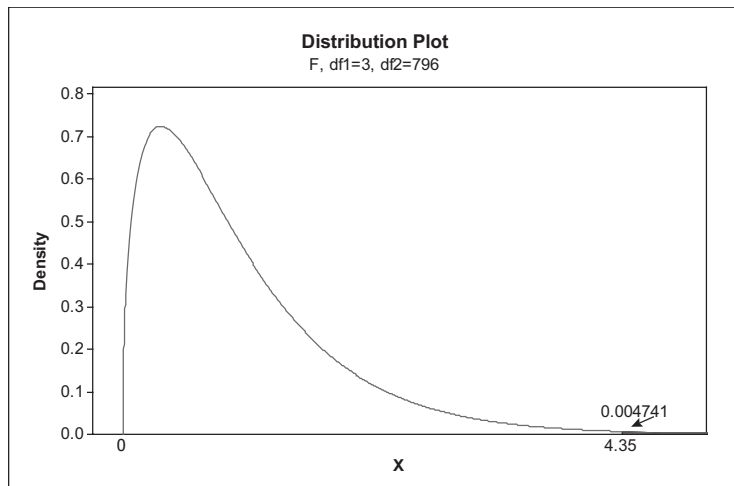
b. $\mu_{\text{Northeast}} = \mu_{\text{Midwest}} = \mu_{\text{South}} = \mu_{\text{West}}$

c. $MSG = [(200)(16.560 - 15.467)^2 + (200)(15.154 - 15.467)^2 + (200)(13.931 - 15.467)^2 + (200)(16.223 - 15.467)^2] / 3 = 844.69 / 3 \approx 281.563$

$$MSE = [(199)(9.164)^2 + (199)(6.381)^2 + (199)(6.933)^2 + (199)(9.289)^2]/(800 - 4) \approx 51550.6/796 \approx 64.762$$

$$F(3, 796) = 281.563/64.762 \approx 4.35$$

d. Using software gives a p -value of around 0.005. (See distribution plot below.)



e. Not all of the population mean hourly pay rates for the four regions of the country are the same. The researchers could conclude that the mean hourly rate for the south is lower than the mean hourly rate for the northeast (since these two sample means are the farthest apart).

4. a.

Occupation	Sample Mean	Standard Deviation
Cashier	424.9	178.1
Customer Service Representative	649.5	333.7
Receptionist	573.2	289.4
Secretary/Administrative Assistant	676.0	319.4

b. The ratio of the largest standard deviation to the smallest standard deviation is $333.7/178.1$, which is under 1.9. Hence, this assumption is reasonably satisfied.

c. Output from Minitab shown below.

Source	DF	SS	MS	F	P
Factor	3	1907179	635726	7.73	0.000
Error	196	16112243	82205		
Total	199	18019422			

d. There are differences among the population mean weekly salaries among these four occupations. Based on the sample means, it appears that there is a difference in mean weekly wages between cashiers and secretaries/administrative assistants.