Solutions Manual

to accompany

PRINCIPLES OF STATISTICS FOR ENGINEERS AND SCIENTISTS

by

William Navidi

Table of Contents

Chapter 1

Section 1.1

1. (a) The population consists of all the bolts in the shipment. It is tangible.

   (b) The population consists of all measurements that could be made on that resistor with that ohmmeter. It is conceptual.

   (c) The population consists of all residents of the town. It is tangible.

   (d) The population consists of all welds that could be made by that process. It is conceptual.

   (e) The population consists of all parts manufactured that day. It is tangible.

3. (a) False

   (b) True

5. (a) No. What is important is the population proportion of defectives; the sample proportion is only an approximation. The population proportion for the new process may in fact be greater or less than that of the old process.

   (b) No. The population proportion for the new process may be 0.12 or more, even though the sample proportion was only 0.11.

   (c) Finding 2 defective circuits in the sample.

7. A good knowledge of the process that generated the data.

Section 1.2

1. (a) The mean will be divided by 2.2.

   (b) The standard deviation will be divided by 2.2.

3. False

5. No. In the sample 1, 2, 4 the mean is 7/3, which does not appear at all.

7. The sample size can be any odd number.

9. Yes. If all the numbers in the list are the same, the standard deviation will equal 0.

11. The sum of the mens' heights is $20 \times 178 = 3560$. The sum of the womens' heights is $30 \times 164 = 4920$. The sum of all 50 heights is $3560+4920 = 8480$. Therefore the mean score for the two classes combined is $8480/50 = 169.6$.

13. (a) All would be divided by 2.54.

    (b) Not exactly the same, because the measurements would be a little different the second time.

15. (a) The sample size is $n = 16$. The tertiles have cutpoints $(1/3)(17) = 5.67$ and $(2/3)(17) = 11.33$. The first tertile is therefore the average of the sample values in positions 5 and 6, which is $(44+46)/2 = 45$. The second tertile is the average of the sample values in positions 11 and 12, which is $(76+79)/2 = 77.5$.

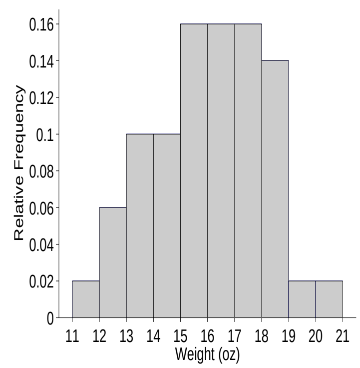
    (b) The sample size is $n = 16$. The quintiles have cutpoints $(i/5)(17)$ for $i = 1, 2, 3, 4$. The quintiles are therefore the averages of the sample values in positions 3 and 4, in positions 6 and 7, in positions 10 and 11, and in positions 13 and 14. The quintiles are therefore $(23 + 41)/2 = 32$, $(46 + 49)/2 = 47.5$, $(74 + 76)/2 = 75$, and $(82 + 89)/2 = 85.5$.

Section 1.3

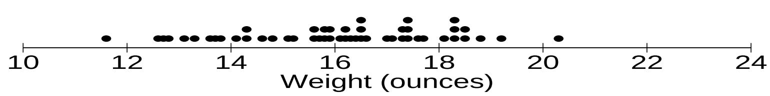
1. (a)

Stem	Leaf
11	6
12	678
13	13678
14	13368
15	126678899
16	122345556
17	013344467
18	1333558
19	2
20	3

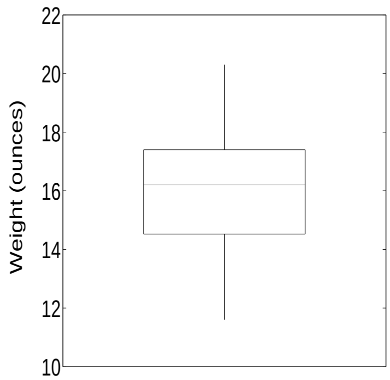
(b) Here is one histogram. Other choices for the endpoints are possible.



(c)



(d)



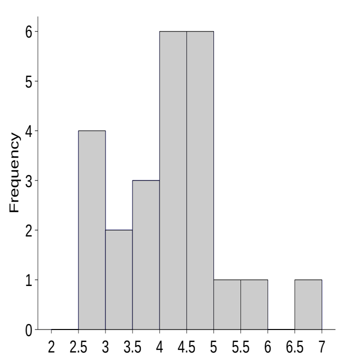
The boxplot shows no outliers.

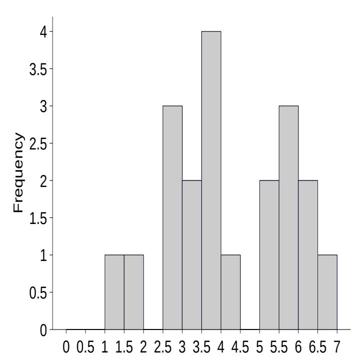
3.

Stem	Leaf
1	1588
2	00003468
3	0234588
4	0346
5	2235666689
6	00233459
7	113558
8	568
9	1225
10	1
11	
12	2
13	06
14	
15	
16	
17	1
18	6
19	9
20	
21	
22	
23	3

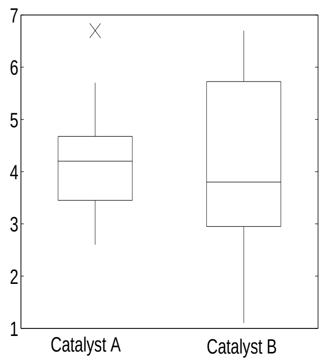
There are 23 stems in this plot. An advantage of this plot over the one in Figure 1.6 is that the values are given to the tenths digit instead of to the ones digit. A disadvantage is that there are too many stems, and many of them are empty.

5. (a) Here are histograms for each group. Other choices for the endpoints are possible.





(b)



(c) The results for Catalyst B are noticeably more spread out than those for Catalyst A. The median yield for catalyst A is greater than the median for catalyst B. The median yield for B is closer to the first quartile than the third, but the lower whisker is longer than the upper one, so the median is approximately equidistant from the extremes of the data. The largest result for Catalyst A is an outlier; the remaining yields for catalyst A are approximately symmetric.

7. (a) The proportion is the sum of the relative frequencies (heights) of the rectangles above 130. This sum is approximately $0.12 + 0.045 + 0.045 + 0.02 + 0.005 + 0.005 = 0.24$. This is closest to 25%.

(b) The height of the rectangle over the interval 130–135 is greater than the sum of the heights of the rectangles over the interval 140–150. Therefore there are more women in the interval 130–135 mm.

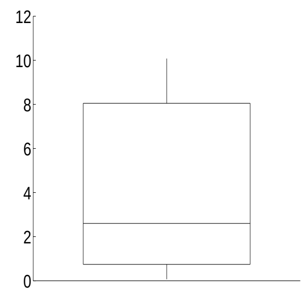
9. Any point more than 1.5 IQR (interquartile range) below the first quartile or above the third quartile is labeled an outlier. To find the IQR, arrange the values in order: 4, 10, 20, 25, 31, 36, 37, 41, 44, 68, 82. There are $n = 11$ values. The first quartile is the value in position $0.25(n + 1) = 3$, which is 20. The third quartile is the value in position $0.75(n + 1) = 9$, which is 44. The interquartile range is $44 - 20 = 24$. So 1.5 IQR is equal to $(1.5)(24) = 36$. There are no points less than $20 - 36 = -16$, so there are no outliers on the low side. There is one point, 82, that is greater than $44 + 36 = 80$. Therefore 82 is the only outlier.

11. The figure on the left is a sketch of separate histograms for each group. The histogram on the right is a sketch of a histogram for the two groups combined. There is more spread in the combined histogram than in either of the separate ones. Therefore the standard deviation of all 200 heights is greater than

2.5 in. The answer is (ii).

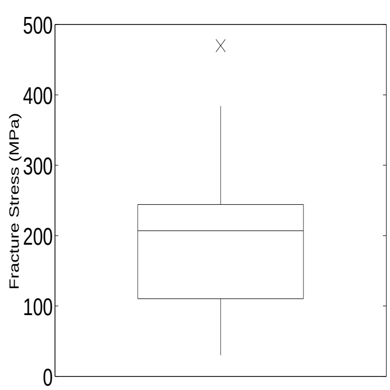
13. (a) IQR = 3rd quartile − 1st quartile. A: IQR = 6.02 − 1.42 = 4.60, B: IQR = 9.13 − 5.27 = 3.86

(b) Yes, since the minimum is within 1.5 IQR of the first quartile and the maximum is within 1.5 IQR of the third quartile, there are no outliers, and the given numbers specify the boundaries of the box and the ends of the whiskers.

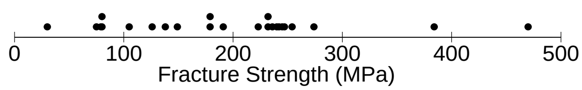


(c) No. The minimum value of −2.235 is an "outlier," since it is more than 1.5 times the interquartile range below the first quartile. The lower whisker should extend to the smallest point that is not an outlier, but the value of this point is not given.

15. (a)



(b) The boxplot indicates that the value 470 is an outlier.

(c) 

(d) The dotplot indicates that the value 384 is detached from the bulk of the data, and thus could be considered to be an outlier.

Supplementary Exercises for Chapter 1

1. The mean and standard deviation both increase by 5%.

3. (a) False. The true percentage could be greater than 5%, with the observation of 4 out of 100 due to sampling variation.

(b) True

(c) False. If the result differs greatly from 5%, it is unlikely to be due to sampling variation.

(d) True. If the result differs greatly from 5%, it is unlikely to be due to sampling variation.

5. (a) It is not possible to tell by how much the mean changes, because the sample size is not known.

(b) If there are more than two numbers on the list, the median is unchanged. If there are only two numbers on the list, the median is changed, but we cannot tell by how much.

(c) It is not possible to tell by how much the standard deviation changes, both because the sample size is unknown and because the original standard deviation is unknown.

7. (a) The mean decreases by 0.774.

   (b) The value of the mean after the change is $25 - 0.774 = 24.226$.

   (c) The median is unchanged.

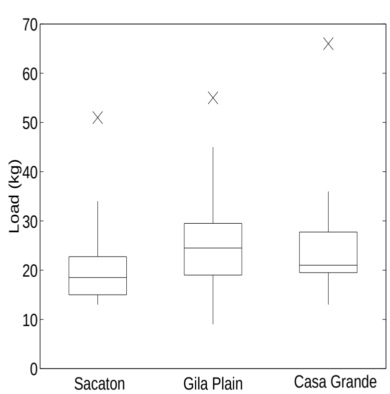
   (d) It is not possible to tell by how much the standard deviation changes, because the original standard deviation is unknown.

9. Statement (i) is true. The sample is skewed to the right.

11. (a) Skewed to the left. The 85th percentile is much closer to the median (50th percentile) than the 15th percentile is. Therefore the histogram is likely to have a longer left-hand tail than right-hand tail.

    (b) Skewed to the right. The 15th percentile is much closer to the median (50th percentile) than the 85th percentile is. Therefore the histogram is likely to have a longer right-hand tail than left-hand tail.

13. (a)



    (b) Each sample contains one outlier.

(c) In the Sacaton boxplot, the median is about midway between the first and third quartiles, suggesting that the data between these quartiles are fairly symmetric. The upper whisker of the box is much longer than the lower whisker, and there is an outlier on the upper side. This indicates that the data as a whole are skewed to the right. In the Gila Plain boxplot data, the median is about midway between the first and third quartiles, suggesting that the data between these quartiles are fairly symmetric. The upper whisker is slightly longer than the lower whisker, and there is an outlier on the upper side. This suggest that the data as a whole are somewhat skewed to the right. In the Casa Grande boxplot, the median is very close to the first quartile. This suggests that there are several values very close to each other about one-fourth of the way through the data. The two whiskers are of about equal length, which suggests that the tails are about equal, except for the outlier on the upper side.

Chapter 2

Section 2.1

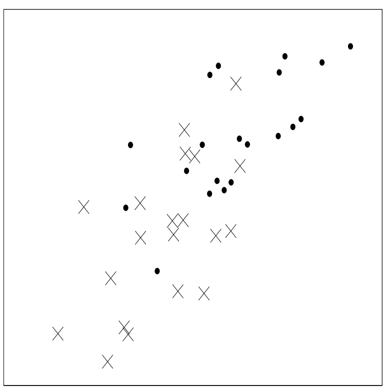
1. $\overline{x} = 3.0$, $\overline{y} = 3.4$, $\sum_{i=1}^n (x_i - \overline{x})^2 = 10$, $\sum_{i=1}^n (y_i - \overline{y})^2 = 21.2$, $\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = 12$.

$$r = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}\sqrt{\sum_{i=1}^n (y_i - \overline{y})^2}} = 0.8242.$$

3. (a) The correlation coefficient is appropriate. The points are approximately clustered around a line.

(b) The correlation coefficient is not appropriate. The relationship is curved, not linear.

(c) The correlation coefficient is not appropriate. The plot contains outliers.

5. 

The heights and weights for the men (dots) are on the whole greater than those for the women (xs). Therefore the scatterplot for the men is shifted up and to the right. The overall plot exhibits a higher correlation than either plot separately. The correlation between heights and weights for men and women taken together will be more than 0.6.

7. (a) Let x represent temperature, y represent stirring rate, and z represent yield.
Then $\overline{x} = 119.875$, $\overline{y} = 45$, $\overline{z} = 75.590625$, $\sum_{i=1}^n (x_i - \overline{x})^2 = 1845.75$,
$\sum_{i=1}^n (y_i - \overline{y})^2 = 1360$, $\sum_{i=1}^n (z_i - \overline{z})^2 = 234.349694$, $\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = 1436$,
$\sum_{i=1}^n (x_i - \overline{x})(z_i - \overline{z}) = 481.63125$, $\sum_{i=1}^n (y_i - \overline{y})(z_i - \overline{z}) = 424.15$.

The correlation coefficient between temperature and yield is

$$r = \frac{\sum_{i=1}^n (x_i - \overline{x})(z_i - \overline{z})}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}\sqrt{\sum_{i=1}^n (z_i - \overline{z})^2}} = 0.7323.$$

The correlation coefficient between stirring rate and yield is

$$r = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(z_i - \overline{z})}{\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}\sqrt{\sum_{i=1}^{n}(z_i - \overline{z})^2}} = 0.7513.$$

The correlation coefficient between temperature and stirring rate is

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} = 0.9064.$$

(b) No, the result might be due to confounding, since the correlation between temperature and stirring rate is far from 0.

(c) No, the result might be due to confounding, since the correlation between temperature and stirring rate is far from 0.

Section 2.2

1. (a) $111.74 + 0.51(65) = 144.89$ kg.

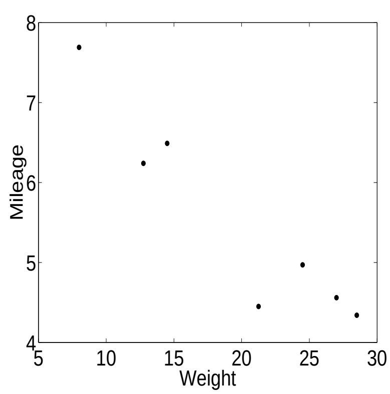
(b) The difference in y predicted from a one-unit change in x is the slope $\widehat{\beta}_1 = 0.51$. Therefore the change in the number of lbs of steam predicted from a change of 5°C is $0.51(5) = 2.55$ kg.

3. (a) $-0.2967 + 0.2738(70) = 18.869$ in.

(b) Let x be the required height. Then $19 = -0.2967 + 0.2738x$, so $x = 70.477$ in.

(c) No, some of the men whose points lie below the least-squares line will have shorter arms.

5. (a)



The linear model is appropriate.

(b) $\overline{x} = 19.5$, $\overline{y} = 5.534286$, $\sum_{i=1}^n (x_i - \overline{x})^2 = 368.125$, $\sum_{i=1}^n (y_i - \overline{y})^2 = 9.928171$, $\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = -57.1075$.

$\widehat{\beta}_1 = \dfrac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = -0.1551307$ and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} = 8.55933$.

The equation of the least-squares line is $y = 8.55933 - 0.1551307x$.

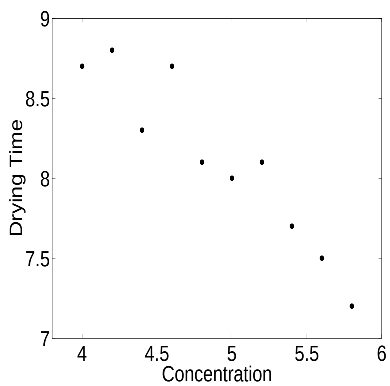
(c) By $0.1551307(5) = 0.776$ miles per gallon.

(d) $8.55933 - 0.1551307(15) = 6.23$ miles per gallon.

(e) miles per gallon per ton

(f) miles per gallon

7. (a)



The linear model is appropriate.

(b) $\overline{x} = 4.9$, $\overline{y} = 8.11$, $\sum_{i=1}^n (x_i - \overline{x})^2 = 3.3$, $\sum_{i=1}^n (y_i - \overline{y})^2 = 2.589$, $\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = -2.75$.

$\widehat{\beta}_1 = \dfrac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = -0.833333$ and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} = 12.19333$.

The equation of the least-squares line is $y = 12.19333 - 0.833333x$.

(c) The fitted values are the values $\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$, and the residuals are the values $e_i = y_i - \widehat{y}_i$, for each value x_i. They are shown in the following table.

x	y	Fitted Value $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x$	Residual $e = y - \widehat{y}$
4.0	8.7	8.860	−0.160
4.2	8.8	8.693	0.107
4.4	8.3	8.527	−0.227
4.6	8.7	8.360	0.340
4.8	8.1	8.193	−0.093
5.0	8.0	8.027	−0.027
5.2	8.1	7.860	0.240
5.4	7.7	7.693	0.007
5.6	7.5	7.527	−0.027
5.8	7.2	7.360	−0.160

(d) $-0.833333(0.1) = -0.0833$. Decrease by 0.0833 hours.

(e) $12.19333 - 0.833333(4.4) = 8.53$ hours.

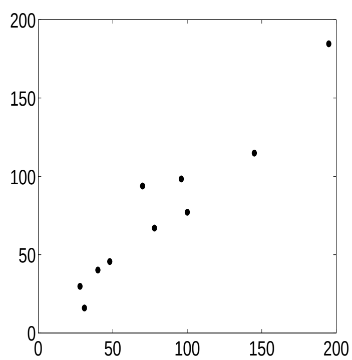
(f) Let x be the required concentration. Then $8.2 = 12.19333 - 0.833333x$, so $x = 4.79\%$.

9. (a) $n = 5$, $\sum_{i=1}^n (x_i - \overline{x})^2 = 0.628$, $\sum_{i=1}^n (y_i - \overline{y})^2 = 0.65612$, $\overline{x} = 1.48$, $\overline{y} = 1.466$, $\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = 0.6386$.

$\widehat{\beta}_1 = \dfrac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = 1.0169$ and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} = -0.038981$.

(b) $-0.038981 + 1.0169(1.3) = 1.283$.

11. (a)



(b) $n = 10$, $\sum_{i=1}^n (x_i - \overline{x})^2 = 25843$, $\sum_{i=1}^n (y_i - \overline{y})^2 = 22131$, $\overline{x} = 83.10$, $\overline{y} = 76.72$, $\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = 22955$.

$\widehat{\beta}_1 = \dfrac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = 0.88824$ and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} = 2.9073$.

The equation of the least-squares line is $y = 2.9073 + 0.88824x$.

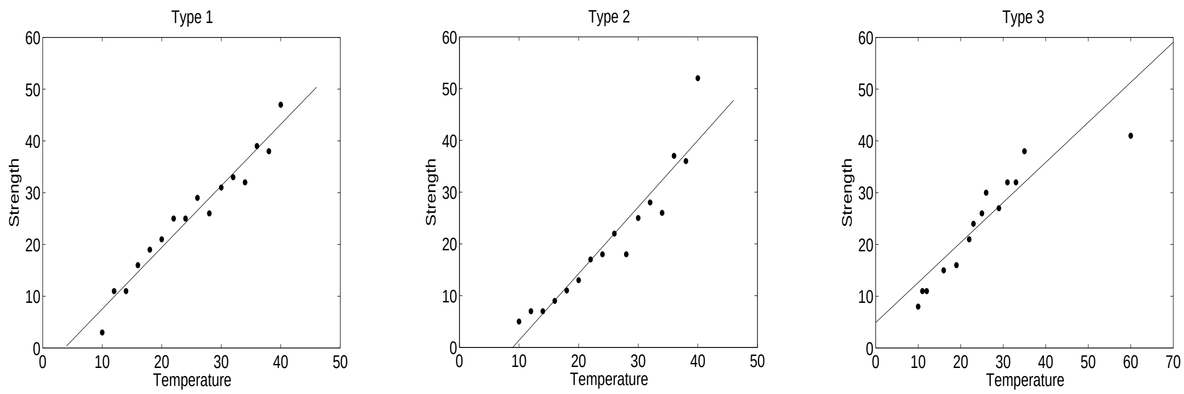
(c) $0.88824(12) = 10.659$. By 10.659×10^{10} joules.

(d) $2.9073 + 0.88824(50) = 47.319 \times 10^{10}$ joules.

(e) Let x be the required income. Then $2.9073 + 0.88824x = 100$, so $x = 109.31$.

Section 2.3

1. (a)



(b) It is appropriate for Type I, as the scatterplot shows a clear linear trend. It is not appropriate for Type II, since the scatterplot has a curved pattern. It is not appropriate for Type III, as the scatterplot contains an outlier.

3. (a) $n = 6$, $\sum_{i=1}^{n}(x_i - \overline{x})^2 = 70$, $\sum_{i=1}^{n}(y_i - \overline{y})^2 = 297.38$, $\overline{x} = 7$, $\overline{y} = 10.9$, $\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = 141.6$.

$\widehat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = 2.0229$ and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1\overline{x} = -3.26$.

The equation of the least-squares line is $y = -3.26 + 2.0229x$.

(b) No, the smallest observed value of time is 2 hours, so this requires extrapolation.

(c) Yes, the prediction is $-3.26 + 2.0229(5) = 6.8545$.

(d) No, the largest observed value of time is 12 hours, so this requires extrapolation.

5. $$r^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = 1 - \frac{1450}{9615} = 0.8492.$$

7. $\widehat{\beta}_1 = rs_y/s_x = (0.85)(1.9)/1.2 = 1.3458$. $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1\overline{x} = 30.4 - 1.3458(8.1) = 19.499$.
The equation of the least-squares line is $y = 19.499 + 1.3458x$.

9. $\widehat{\beta}_1 = rs_y/s_x = 0.5(10/0.5) = 10$. $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1\overline{x} = 50 - 10(3) = 20$.
The equation of the least-squares line is $y = 20 + 10x$.

Supplementary Exercises for Chapter 2

1. (iii) equal to \$47,500. The least-squares line goes through the point $(\overline{x}, \overline{y})$, so when height is equal to its average of 70, income is equal to its average of \$47,500.

3. Closest to -1. If two people differ in age by x years, the graduation year of the older one will be approximately x years less than that of the younger one. Therefore the points on a scatterplot of age vs. graduation year would lie very close to a straight line with negative slope.

5. $\overline{x} = 0.5$ and $\sum_{i=1}^{n}(x_i - \overline{x})^2 = 5$.
$\overline{y} = [-1 + 0 + 1 + y]/4 = y/4$, so $y = 4\overline{y}$.
Express $\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$, $\sum_{i=1}^{n}(y_i - \overline{y})^2$, and r in terms of $\overline{y}$:
$\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = (-1)(-1 - \overline{y}) + 0(0 - \overline{y}) + 1(1 - \overline{y}) + 2(4\overline{y} - \overline{y}) = 6\overline{y} + 2$.
$\sum_{i=1}^{n}(y_i - \overline{y})^2 = (-1 - \overline{y})^2 + (0 - \overline{y})^2 + (1 - \overline{y})^2 + (4\overline{y} - \overline{y})^2 = 12\overline{y}^2 + 2$.
Now $r = \dfrac{6\overline{y} + 2}{\sqrt{5}\sqrt{12\overline{y}^2 + 2}}$.

(a) To obtain $r = 1$, set $6\overline{y} + 2 = \sqrt{5}\sqrt{12\overline{y}^2 + 2}$ so
$36\overline{y}^2 + 24\overline{y} + 4 = 60\overline{y}^2 + 10$, or $4\overline{y}^2 - 4\overline{y} + 1 = 0$. Solving for $\overline{y}$ yields $\overline{y} = 1/2$, so $y = 4\overline{y} = 2$.

(b) To obtain $r = 0$, set $6\overline{y} + 2 = 0$. Then $\overline{y} = -1/3$ so $y = 4\overline{y} = -4/3$.

(c) For the correlation to be equal to -1, the points would have to lie on a straight line with negative slope. There is no value for y for which this is the case.

7. (a) Let x represent area and y represent population

$\overline{x} = 2812.1$, $\overline{y} = 612.74$, $\sum_{i=1}^{n}(x_i - \overline{x})^2 = 335069724.6$, $\sum_{i=1}^{n}(y_i - \overline{y})^2 = 18441216.4$, $\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = 32838847.8$.

$\widehat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = 0.098006$ and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1\overline{x} = 337.13$.

The equation of the least squares line is $y = 337.13 + 0.098006x$.

(b) $337.13 + 0.098006(5000) = 827$.

(c) Let x represent ln area and y represent ln population.

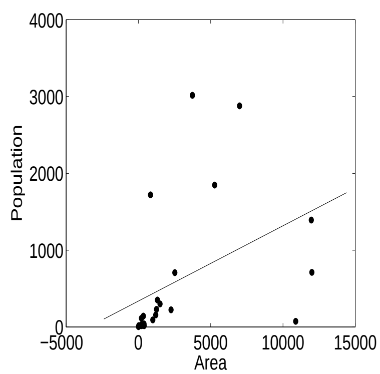
$\overline{x} = 6.7769$, $\overline{y} = 5.0895$, $\sum_{i=1}^{n}(x_i - \overline{x})^2 = 76.576$, $\sum_{i=1}^{n}(y_i - \overline{y})^2 = 78.643$, $\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = 62.773$.

$\widehat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = 0.8198$ and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1\overline{x} = -0.4658$.

The equation of the least squares line is $\ln \text{population} = -0.4658 + 0.8198 \ln \text{area}$.

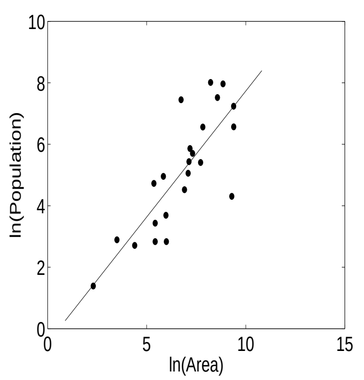
(d) $\ln \text{population} = -0.4658 + 0.8198(\ln 5000) = 6.5166$. The predicted population is $e^{6.5166} \approx 676$.

(e)



The linear model is not appropriate.

(f)



The linear model is appropriate.

(g) The scatterplot of ln population versus ln area.

(h) The prediction in part (d) is more reliable, since it is based on a scatterplot that is more linear.

Chapter 3

Section 3.1

1. $P(\text{does not fail}) = 1 - P(\text{fails}) = 1 - 0.12 = 0.88$.

3. Let A denote the event that the resistance is above specification, and let B denote the event that the resistance is below specification. Then A and B are mutually exclusive.

(a) $P(\text{doesn't meet specification}) = P(A \cup B) = P(A) + P(B) = 0.05 + 0.10 = 0.15$

(b) $P[B \mid (A \cup B)] = \dfrac{P[(B \cap (A \cup B)]}{P(A \cup B)} = \dfrac{P(B)}{P(A \cup B)} = \dfrac{0.10}{0.15} = 0.6667$

5. $$\begin{aligned} P(A \cap B) &= P(A) + P(B) - P(A \cup B) \\ &= 0.98 + 0.95 - 0.99 \\ &= 0.94 \end{aligned}$$

7. (a) 0.6

(b) $$\begin{aligned} P(\text{personal computer or laptop computer}) &= P(\text{personal computer}) + P(\text{laptop computer}) \\ &= 0.6 + 0.3 \\ &= 0.9 \end{aligned}$$

Section 3.2

1. Let A represent the event that the biotechnology company is profitable, and let B represent the event that the information technology company is profitable. Then $P(A) = 0.2$ and $P(B) = 0.15$.

(a) $P(A \cap B) = P(A)P(B) = (0.2)(0.15) = 0.03$.

(b) $P(A^c \cap B^c) = P(A^c)P(B^c) = (1-0.2)(1-0.15) = 0.68$.

(c)
$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(A)P(B) \\ &= 0.2 + 0.15 - (0.2)(0.15) \\ &= 0.32 \end{aligned}$$

3. (a) $\dfrac{88}{12+88} = 0.88$.

(b) $\dfrac{88}{88+165+260} = 0.1715$.

(c) $\dfrac{88+165}{88+165+260} = 0.4932$.

(d) $\dfrac{88+165}{88+165+12+35} = 0.8433$.

5. (a) $\dfrac{56+24}{100} = 0.80$

(b) $\dfrac{56+14}{100} = 0.70$

(c)
$$\begin{aligned} P(\text{Gene 2 dominant} \mid \text{Gene 1 dominant}) &= \frac{P(\text{Gene 1 dominant} \cap \text{Gene 2 dominant})}{P(\text{Gene 1 dominant})} \\ &= \frac{56/100}{0.8} \\ &= 0.7 \end{aligned}$$

(d) Yes. $P(\text{Gene 2 dominant} \mid \text{Gene 1 dominant}) = P(\text{Gene 2 dominant})$

7. Let A be the event that component A functions, let B be the event that component B functions, let C be the event that component C functions, and let D be the event that component D functions. Then $P(A) = 1-0.1 = 0.9$, $P(B) = 1-0.2 = 0.8$, $P(C) = 1-0.05 = 0.95$, and $P(D) = 1-0.3 = 0.7$. The event that the system functions is $(A \cup B) \cup (C \cup D)$.

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A)P(B) = 0.9 + 0.8 - (0.9)(0.8) = 0.98$.

$P(C \cup D) = P(C) + P(D) - P(C \cap D) = P(C) + P(D) - P(C)P(D) = 0.95 + 0.7 - (0.95)(0.7) = 0.985$.

$P[(A \cup B) \cup (C \cup D)] = P(A \cup B) + P(C \cup D) - P(A \cup B)P(C \cup D) = 0.98 + 0.985 - (0.98)(0.985) = 0.9997$.

9. Let C denote the event that component C functions, and let D denote the event that component D functions.

(a)
$$\begin{aligned} P(\text{system functions}) &= P(C \cup D) \\ &= P(C) + P(D) - P(C \cap D) \\ &= (1-0.08) + (1-0.12) - (1-0.08)(1-0.12) \\ &= 0.9904 \end{aligned}$$

Alternatively,

$$\begin{aligned} P(\text{system functions}) &= 1 - P(\text{system fails}) \\ &= 1 - P(C^c \cap D^c) \\ &= 1 - P(C^c)P(D^c) \\ &= 1 - (0.08)(0.12) \\ &= 0.9904 \end{aligned}$$

(b) $P(\text{system functions}) = 1 - P(C^c \cap D^c) = 1 - p^2 = 0.99$. Therefore $p = \sqrt{1-0.99} = 0.1$.

(c) $P(\text{system functions}) = 1 - p^3 = 0.99$. Therefore $p = (1-0.99)^{1/3} = 0.2154$.

(d) Let n be the required number of components. Then n is the smallest integer such that $1 - 0.5^n \geq 0.99$. It follows that $n \ln(0.5) \leq \ln 0.01$, so $n \geq (\ln 0.01)(\ln 0.5) = 6.64$. Since n must be an integer, $n = 7$.

11. Let A be the event that the bit is reversed at the first relay, and let B be the event that the bit is reversed at the second relay. Then $P(\text{bit received is the same as the bit sent}) = P(A^c \cap B^c) + P(A \cap B) = P(A^c)P(B^c) + P(A)P(B) = 0.9^2 + 0.1^2 = 0.82$.

Section 3.3

1. (a) Discrete

(b) Continuous

(c) Discrete

(d) Continuous

(e) Discrete

3. (a) $\mu_X = 1(0.4) + 2(0.2) + 3(0.2) + 4(0.1) + 5(0.1) = 2.3$

(b) $\sigma_X^2 = (1-2.3)^2(0.4) + (2-2.3)^2(0.2) + (3-2.3)^2(0.2) + (4-2.3)^2(0.1) + (5-2.3)^2(0.1) = 1.81$
Alternatively, $\sigma_X^2 = 1^2(0.4) + 2^2(0.2) + 3^2(0.2) + 4^2(0.1) + 5^2(0.1) - 2.3^2 = 1.81$

(c) $\sigma_X = \sqrt{1.81} = 1.345$

(d) $Y = 10X$. Therefore the probability density function is as follows.

y	10	20	30	40	50
$p(y)$	0.4	0.2	0.2	0.1	0.1

(e) $\mu_Y = 10(0.4) + 20(0.2) + 30(0.2) + 40(0.1) + 50(0.1) = 23$

(f) $\sigma_Y^2 = (10-23)^2(0.4) + (20-23)^2(0.2) + (30-23)^2(0.2) + (40-23)^2(0.1) + (50-23)^2(0.1) = 181$
Alternatively, $\sigma_Y^2 = 10^2(0.4) + 20^2(0.2) + 30^2(0.2) + 40^2(0.1) + 50^2(0.1) - 23^2 = 181$

(g) $\sigma_Y = \sqrt{181} = 13.45$

5. (a) $\sum_{x=1}^{4} cx = 1$, so $c(1+2+3+4) = 1$, so $c = 0.1$.

(b) $P(X = 2) = c(2) = 0.1(2) = 0.2$

(c) $\mu_X = \sum_{x=1}^{4} xP(X = x) = \sum_{x=1}^{4} 0.1x^2 = (0.1)(1^2 + 2^2 + 3^2 + 4^2) = 3.0$

(d) $\sigma_X^2 = \sum_{x=1}^{4} (x - \mu_X)^2 P(X = x) = \sum_{x=1}^{4} (x-3)^2(0.1x) = 4(0.1) + 1(0.2) + 0(0.3) + 1(0.4) = 1$
Alternatively, $\sigma_X^2 = \sum_{x=1}^{4} x^2 P(X = x) - \mu_X^2 = \sum_{x=1}^{4} 0.1x^3 - 3^2 = 0.1(1^3 + 2^3 + 3^3 + 4^3) - 3^2 = 1$

(e) $\sigma_X = \sqrt{1} = 1$

7. (a) $\displaystyle\int_{80}^{90} \frac{x-80}{800}\,dx = \left.\frac{x^2-160x}{1600}\right|_{80}^{90} = 0.0625$

(b) $\displaystyle\int_{80}^{120} x\frac{x-80}{800}\,dx = \left.\frac{x^3-120x}{2400}\right|_{80}^{120} = 320/3 = 106.67$

(c) $\displaystyle\sigma_X^2 = \int_{80}^{120} x^2\frac{x-80}{800}\,dx - (320/3)^2 = \left.\frac{x^4}{3200} - \frac{x^3}{30}\right|_{80}^{120} - (320/3)^2 = 800/9$

$\sigma_X = \sqrt{800/9} = 9.428$

(d) $\displaystyle F(x) = \int_{-\infty}^{x} f(t)\,dt$

If $x < 80$, $\displaystyle F(x) = \int_{-\infty}^{80} 0\,dt = 0$

If $80 \le x < 120$, $\displaystyle F(x) = \int_{-\infty}^{80} 0\,dt + \int_{80}^{x} \frac{t-80}{800}\,dt = x^2/1600 - x/10 + 4.$

If $X \ge 120$, $\displaystyle F(x) = \int_{-\infty}^{80} 0\,dt + \int_{80}^{120} \frac{t-80}{800}\,dt + \int_{120}^{x} 0\,dt = 1.$

9. (a)
$$\begin{aligned}
\mu &= \int_0^\infty 0.1te^{-0.1t}\,dt \\
&= \left.-te^{-0.1t}\right|_0^\infty - \int_0^\infty -e^{-0.1t}\,dt \\
&= \left.0 - 10e^{-0.1t}\right|_0^\infty \\
&= 10
\end{aligned}$$

(b)
$$\begin{aligned}
\sigma^2 &= \int_0^\infty 0.1t^2e^{-0.1t}\,dt - \mu^2 \\
&= \left.-t^2e^{-0.1t}\right|_0^\infty - \int_0^\infty -2te^{-0.1t}\,dt - 100 \\
&= 0 + 20\int_0^\infty 0.1te^{-0.1t}\,dt - 100 \\
&= 0 + 20(10) - 100 \\
&= 100
\end{aligned}$$

$$\sigma_X = \sqrt{100} = 10$$

(c) $F(x) = \int_{-\infty}^{x} f(t)\,dt$.

If $x \le 0$, $F(x) = \int_{-\infty}^{x} 0\,dt = 0$.

If $x > 0$, $F(x) = \int_{-\infty}^{0} 0\,dt + \int_{0}^{x} 0.1e^{-0.1t}\,dt = 1 - e^{-0.1x}$.

(d) Let T represent the lifetime. $P(T < 12) = P(T \le 12) = F(12) = 1 - e^{-1.2} = 0.6988$.

11. (a) $P(X > 0.5) = \int_{0.5}^{1} 1.2(x + x^2)\,dx = 0.6x^2 + 0.4x^3 \Big|_{0.5}^{1} = 0.8$

(b) $\mu = \int_{0}^{1} 1.2x(x + x^2)\,dx = 0.4x^3 + 0.3x^4 \Big|_{0}^{1} = 0.7$

(c) X is within ± 0.1 of the mean if $0.6 < X < 0.8$.

$P(0.6 < X < 0.8) = \int_{0.6}^{0.8} 1.2(x + x^2)\,dx = 0.6x^2 + 0.4x^3 \Big|_{0.6}^{0.8} = 0.2864$

(d) The variance is

$$\begin{aligned}
\sigma^2 &= \int_{0}^{1} 1.2x^2(x + x^2)\,dx - \mu^2 \\
&= 0.3x^4 + 0.24x^5 \Big|_{0}^{1} - 0.7^2 \\
&= 0.05
\end{aligned}$$

The standard deviation is $\sigma = \sqrt{0.05} = 0.2236$.

(e) X is within $\pm 2\sigma$ of the mean if $0.2528 < X < 1.1472$. Since $P(X > 1) = 0$, X is within $\pm 2\sigma$ of the mean if $0.2528 < X < 1$.

$P(0.2528 < X > 1) = \int_{0.2528}^{1} 1.2(x + x^2)\,dx = 0.6x^2 + 0.4x^3 \Big|_{0.2528}^{1} = 0.9552$

(f) $F(x) = \int_{-\infty}^{x} f(t)\,dt$

If $x \le 0$, $F(x) = \int_{-\infty}^{x} 0\,dt = 0$

If $0 < x < 1$, $F(x) = \int_{0}^{x} 1.2(t + t^2)\,dt = 0.6x^2 + 0.4x^3$

If $x > 1$, $F(x) = \int_{0}^{1} 1.2(t + t^2)\,dt = 1$

13. (a) $P(X < 2.5) = \int_{2}^{2.5} (3/52)x(6 - x)\,dx = (9x^2 - x^3)/52 \Big|_{2}^{2.5} = 0.2428$

(b) $P(2.5 < X < 3.5) = \int_{2.5}^{3.5} (3/52)x(6 - x)\,dx = \dfrac{9x^2 - x^3}{52}\Bigg|_{2.5}^{3.5} = 0.5144$

(c) $\mu = \int_{2}^{4} (3/52)x^2(6 - x)\,dx = \dfrac{24x^3 - 3x^4}{208}\Bigg|_{2}^{4} = 3$

(d) The variance is

$$
\begin{aligned}
\sigma^2 &= \int_{2}^{4} (3/52)x^3(6 - x)\,dx - \mu^2 \\
&= \frac{9x^4}{104} - \frac{3x^5}{260}\Bigg|_{2}^{4} - 3^2 \\
&= 0.3230769
\end{aligned}
$$

The standard deviation is $\sigma = \sqrt{0.3230769} = 0.5684$.

(e) Let X represent the thickness. Then X is within $\pm\sigma$ of the mean if $2.4316 < X < 3.5684$.

$P(2.4316 < X < 3.5684) = \int_{2.4316}^{3.5684} (3/52)x(6 - x)\,dx = \dfrac{9x^2 - x^3}{52}\Bigg|_{2.4316}^{3.5684} = 0.5832$

(f) $F(x) = \int_{-\infty}^{x} f(t)\,dt$

If $x \le 2$, $F(x) = \int_{-\infty}^{x} 0\,dt = 0$.

If $2 < x < 4$, $F(x) = \int_{-\infty}^{2} 0\,dt + \int_{2}^{x} (3/52)t(6 - t)\,dt = \dfrac{9x^2 - x^3 - 28}{52}$.

If $x \geq 4$, $F(x) = \int_{-\infty}^{2} 0\,dt + \int_{2}^{4} (3/52)t(6-t)\,dt + \int_{4}^{x} 0\,dt = 1.$

15. (a) $P(X > 3) = \int_{3}^{4} (3/64)x^2(4-x)\,dx = \left.\dfrac{x^3}{16} - \dfrac{3x^4}{256}\right|_{3}^{4} = 67/256$

(b) $P(2 < X < 3) = \int_{2}^{3} (3/64)x^2(4-x)\,dx = \left.\dfrac{x^3}{16} - \dfrac{3x^4}{256}\right|_{2}^{3} = 109/256$

(c) $\mu_X = \int_{0}^{4} (3/64)x^3(4-x)\,dx = \int_{0}^{4} \left.\dfrac{3x^3}{64} - \dfrac{3x^5}{320}\right|_{0}^{4} = 2.4$

(d) The variance is

$$\begin{aligned}
\sigma^2 &= \int_{0}^{4} (3/64)x^4(4-x)\,dx - \mu^2 \\
&= \left.\frac{3x^5}{80} - \frac{x^6}{128}\right|_{0}^{4} - 2.4^2 \\
&= 0.64
\end{aligned}$$

(e) $F(x) = \int_{-\infty}^{x} f(t)\,dt$

If $x < 0$, $F(x) = \int_{-\infty}^{0} 0\,dt = 0.$

If $0 \leq x < 4$, $F(x) = \int_{-\infty}^{0} 0\,dt + \int_{0}^{x} (3/64)t^2(4-t)\,dt = \left.\left(\dfrac{t^3}{16} - \dfrac{3t^4}{256}\right)\right|_{0}^{x} = \dfrac{x^3}{16} - \dfrac{3x^4}{256}$

If $x \geq 4$, $F(x) = \int_{0}^{4} (3/64)t^2(4-t)\,dt = 1.$

Section 3.4

1. (a) $\mu_{3X} = 3\mu_X = 3(9.5) = 28.5$
   $\sigma_{3X} = 3\sigma_X = 3(0.4) = 1.2$

(b) $\mu_{Y-X} = \mu_Y - \mu_X = 6.8 - 9.5 = -2.7$
$\sigma_{Y-X} = \sqrt{\sigma_Y^2 + \sigma_X^2} = \sqrt{0.1^2 + 0.4^2} = 0.412$

(c) $\mu_{X+4Y} = \mu_X + 4\mu_Y = 9.5 + 4(6.8) = 36.7$
$\sigma_{X+4Y} = \sqrt{\sigma_X^2 + 4^2\sigma_Y^2} = \sqrt{0.4^2 + 4^2(0.1^2)} = 0.566$

3. Let $X_1, ..., X_{24}$ denote the volumes of the bottles in a case. Then $\overline{X}$ is the average volume.
$\mu_{\overline{X}} = \mu_{X_i} = 2.013$, and $\sigma_{\overline{X}} = \sigma_{X_i}/\sqrt{24} = 0.005/\sqrt{24} = 0.00102$.

5. Let $X_1, ..., X_5$ denote the thicknesses of the layers, and let $S = X_1 + X_2 + X_3 + X_4 + X_5$ denote the thickness of the piece of plywood.

(a) $\mu_S = \sum \mu_{X_i} = 5(3.5) = 17.5$

(b) $\sigma_S = \sqrt{\sum \sigma_{X_i}^2} = \sqrt{5(0.1^2)} = 0.224$

7. (a) $\mu_M = \mu_{X+1.5Y} = \mu_X + 1.5\mu_Y = 0.125 + 1.5(0.350) = 0.650$

(b) $\sigma_M = \sigma_{X+1.5Y} = \sqrt{\sigma_X^2 + 1.5^2\sigma_Y^2} = \sqrt{0.05^2 + 1.5^2(0.1^2)} = 0.158$

9. Let X_1 and X_2 denote the lengths of the pieces chosen from the population with mean 30 and standard deviation 0.1, and let Y_1 and Y_2 denote the lengths of the pieces chosen from the population with mean 45 and standard deviation 0.3.

(a) $\mu_{X_1+X_2+Y_1+Y_2} = \mu_{X_1} + \mu_{X_2} + \mu_{Y_1} + \mu_{Y_2} = 30 + 30 + 45 + 45 = 150$

(b) $\sigma_{X_1+X_2+Y_1+Y_2} = \sqrt{\sigma_{X_1}^2 + \sigma_{X_2}^2 + \sigma_{Y_1}^2 + \sigma_{Y_2}^2} = \sqrt{0.1^2 + 0.1^2 + 0.3^2 + 0.3^2} = 0.447$

11. The tank holds 20 gallons of gas. Let Y_1 be the number of miles traveled on the first gallon, let Y_2 be the number of miles traveled on the second gallon, and so on, with Y_{20} being the number of miles traveled on the 20th gallon. Then $\mu_{Y_i} = 25$ miles and $\sigma_{Y_i} = 2$ miles. Let $X = Y_1 + Y_2 + \cdots Y_{20}$ denote the number of miles traveled on one tank of gas.

(a) $\mu_X = \mu_{Y_1} + \cdots + \mu_{Y_{20}} = 20(25) = 500$ miles.

(b) $\sigma_X^2 = \sigma_{Y_1}^2 + \cdots + \sigma_{Y_{20}}^2 = 20(2^2) = 80$. So $\sigma_X = \sqrt{80} = 8.944$.

(c) $\mu_{X/20} = (1/20)\mu_X = (1/20)(500) = 25$.

(d) $\sigma_{X/20} = (1/20)\sigma_X = (1/20)(8.944) = 0.4472$

13. $s = 5$, $t = 1.01$, $\sigma_t = 0.02$, $g = 2st^{-2} = 9.80$

$\dfrac{dg}{dt} = -4st^{-3} = -19.4118 \qquad \sigma_g = \left|\dfrac{dg}{dt}\right| \sigma_t = 0.39$

$g = 9.80 \pm 0.39$ m/s^2

Supplementary Exercises for Chapter 3

1. (a) The events of having a major flaw and of having only minor flaws are mutually exclusive. Therefore

$$P(\text{major flaw or minor flaw}) = P(\text{major flaw}) + P(\text{only minor flaws}) = 0.15 + 0.05 = 0.20.$$

(b) $P(\text{no major flaw}) = 1 - P(\text{major flaw}) = 1 - 0.05 = 0.95$.

3. (a) That the gauges fail independently.

(b) One cause of failure, a fire, will cause both gauges to fail. Therefore, they do not fail independently.

(c) Too low. The correct calculation would use $P(\text{second gauge fails}|\text{first gauge fails})$ in place of $P(\text{second gauge fails})$. Because there is a chance that both gauges fail together in a fire, the condition that the first gauge fails makes it more likely that the second gauge fails as well.
Therefore $P(\text{second gauge fails}|\text{first gauge fails}) > P(\text{second gauge fails})$.

5. $P(\text{system functions}) = P[(A \cap B) \cap (C \cup D)]$. Now $P(A \cap B) = P(A)P(B) = (1 - 0.05)(1 - 0.03) = 0.9215$, and $P(C \cup D) = P(C) + P(D) - P(C \cap D) = (1-0.07) + (1-0.14) - (1-0.07)(1-0.14) = 0.9902$. Therefore

$$\begin{aligned} P[(A \cap B) \cap (C \cup D)] &= P(A \cap B)P(C \cup D) \\ &= (0.9215)(0.9902) \\ &= 0.9125 \end{aligned}$$

7. (a) $\mu_{3X} = 3\mu_X = 3(2) = 6, \quad \sigma^2_{3X} = 3^2\sigma^2_X = (3^2)(1^2) = 9$

(b) $\mu_{X+Y} = \mu_X + \mu_Y = 2 + 2 = 4, \quad \sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y = 1^2 + 3^2 = 10$

(c) $\mu_{X-Y} = \mu_X - \mu_Y = 2 - 2 = 0, \quad \sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y = 1^2 + 3^2 = 10$

(d) $\mu_{2X+6Y} = 2\mu_X + 6\mu_Y = 2(2) + 6(2) = 16, \quad \sigma^2_{2X+6Y} = 2^2\sigma^2_X + 6^2\sigma^2_Y = (2^2)(1^2) + (6^2)(3^2) = 328$

9. $g = 9.80$, $L = 0.742$, $\sigma_L = 0.005$, $T = 2\pi\sqrt{L/g} = 2.00709\sqrt{L} = 1.7289$

$\dfrac{dT}{dL} = 1.003545L^{-1/2} = 1.165024 \quad \sigma_T = \left|\dfrac{dT}{dL}\right|\sigma_L = 0.0058$

$T = 1.7289 \pm 0.0058$ s

11. (a) $P(X < 2) = \displaystyle\int_0^2 xe^{-x}\,dx = \left(-xe^{-x}\Big|_0^2 + \int_0^2 e^{-x}\,dx\right) = \left(-2e^{-2} - e^{-x}\Big|_0^2\right) = 1 - 3e^{-2} = 0.5940$

(b) $P(1.5 < X < 3) = \displaystyle\int_{1.5}^3 xe^{-x}\,dx = \left(-xe^{-x}\Big|_{1.5}^3 + \int_{1.5}^3 e^{-x}\,dx\right) = \left(-3e^{-3} + 1.5e^{-1.5} - e^{-x}\Big|_{1.5}^3\right)$

$= 2.5e^{-1.5} - 4e^{-3} = 0.3587$

(c) $\mu = \displaystyle\int_0^\infty x^2e^{-x}\,dx = -x^2e^{-x}\Big|_0^\infty + \int_0^\infty 2xe^{-x}\,dx = 0 + 2xe^{-x}\Big|_0^\infty = 2$

(d) $F(x) = \displaystyle\int_{-\infty}^x f(t)\,dt$

If $x < 0$, $F(x) = \displaystyle\int_{-\infty}^x 0\,dt = 0$

If $x > 0$, $F(x) = \int_0^x te^{-t}\,dt = 1 - (x+1)e^{-x}$

13. With this process, the probability that a ring meets the specification is

$$\int_{9.9}^{10.1} 15[1-25(x-10.05)^2]/4\,dx = \int_{-0.15}^{0.05} 15[1-25x^2]/4\,dx = 0.25(15x - 125x^3)\Bigg|_{-0.15}^{0.05} = 0.641.$$

With the process in Exercise 12, the probability is

$$\int_{9.9}^{10.1} 3[1-16(x-10)^2]\,dx = \int_{-0.1}^{0.1} 3[1-16x^2]\,dx = 3x - 16x^3\Bigg|_{-0.1}^{0.1} = 0.568.$$

Therefore this process is better than the one in Exercise 12.

15. (a) $\mu = 0.0695 + \dfrac{1.0477}{20} + \dfrac{0.8649}{20} + \dfrac{0.7356}{20} + \dfrac{0.2171}{30} + \dfrac{2.8146}{60} + \dfrac{0.5913}{15} + \dfrac{0.0079}{10} + 5(0.0006) = 0.2993$

(b) $\sigma = \sqrt{0.0018^2 + (\frac{0.0269}{20})^2 + (\frac{0.0225}{20})^2 + (\frac{0.0113}{20})^2 + (\frac{0.0185}{30})^2 + (\frac{0.0284}{60})^2 + (\frac{0.0031}{15})^2 + (\frac{0.0006}{10})^2 + 5^2(0.0002)^2} = 0.00288$

Chapter 4

Section 4.1

1. (a) $P(X = 3) = \dfrac{10!}{3!(10-3)!}(0.6)^3(1-0.6)^{10-3} = 0.0425$

   (b) $P(X = 6) = \dfrac{10!}{6!(10-6)!}(0.6)^6(1-0.6)^{10-6} = 0.2508$

   (c)
$$\begin{aligned} P(X \le 4) &= P(X=0) + P(X=1) + P(X=2) + P(X=3) \\ &= \frac{10!}{0!(10-0)!}(0.6)^0(1-0.6)^{10-0} + \frac{10!}{1!(10-1)!}(0.6)^1(1-0.6)^{10-1} + \frac{10!}{2!(10-2)!}(0.6)^2(1-0.6)^{10-2} \\ &\quad + \frac{10!}{3!(10-3)!}(0.6)^3(1-0.6)^{10-3} \\ &= 0.1662 \end{aligned}$$

   (d)
$$\begin{aligned} P(X > 8) &= P(X=9) + P(X=10) \\ &= \frac{10!}{9!(10-9)!}(0.6)^9(1-0.6)^{10-9} + \frac{10!}{10!(10-10)!}(0.6)^{10}(1-0.6)^{10-10} \\ &= 0.0464 \end{aligned}$$

   (e) $\mu_X = (10)(0.6) = 6$

   (f) $\sigma_X^2 = (10)(0.6)(0.4) = 2.4$

3. (a) $P(X = 7) = \dfrac{13!}{7!(13-7)!}(0.4)^7(1-0.4)^{13-7} = 0.1312$

   (b)
$$\begin{aligned} P(X \ge 2) &= 1 - P(X=0) - P(X=1) \\ &= 1 - \frac{8!}{0!(8-0)!}(0.4)^0(1-0.4)^{8-0} - \frac{8!}{1!(8-1)!}(0.4)^1(1-0.4)^{8-1} \\ &= 0.8936 \end{aligned}$$

   (c)
$$\begin{aligned} P(X < 5) &= 1 - P(X=5) - P(X=6) \\ &= 1 - \frac{6!}{5!(6-5)!}(0.7)^5(1-0.7)^{6-5} - \frac{6!}{6!(6-6)!}(0.7)^6(1-0.7)^{6-6} \\ &= 0.5798 \end{aligned}$$

(d) $$\begin{aligned} P(2 \le X \le 4) &= P(X=2) + P(X=3) + P(X=4) \\ &= \frac{7!}{2!(7-2)!}(0.1)^2(1-0.1)^{7-2} - \frac{7!}{3!(7-3)!}(0.1)^3(1-0.1)^{7-3} + \frac{7!}{4!(7-4)!}(0.1)^4(1-0.1)^{7-4} \\ &= 0.1495 \end{aligned}$$

5. Let X be the number of failures that occur in the base metal. Then $X \sim \text{Bin}(20, 0.15)$.

(a) $P(X=5) = \dfrac{20!}{5!(20-5)!}(0.15)^5(1-0.15)^{20-5} = 0.1028$

(b) $$\begin{aligned} P(X < 4) &= P(X=0) + P(X=1) + P(X=2) + P(X=3) \\ &= \frac{20!}{0!(20-0)!}(0.15)^0(1-0.15)^{20-0} + \frac{20!}{1!(20-1)!}(0.15)^1(1-0.15)^{20-1} + \frac{20!}{2!(20-2)!}(0.15)^2(1-0.15)^{20-} \\ &\quad + \frac{20!}{3!(20-3)!}(0.15)^3(1-0.15)^{20-3} \\ &= 0.6477 \end{aligned}$$

(c) $P(X=0) = \dfrac{20!}{0!(20-0)!}(0.15)^0(1-0.15)^{20-0} = 0.0388$

(d) $\mu_X = (20)(0.15) = 3$

(e) $\sigma_X = \sqrt{(20)(0.15)(0.85)} = 1.5969$

7. Let X be the number of heads. Then $X \sim \text{Bin}(8, 0.5)$.

(a) $P(X=5) = \dfrac{8!}{5!(8-5)!}(0.5)^5(1-0.5)^{8-5} = 0.21875$

(b) $\mu_X = (8)(0.5) = 4$

(c) $\sigma_X^2 = (8)(0.5)(0.5) = 2$

(d) $\sigma_X = \sqrt{(8)(0.5)(0.5)} = 1.4142$

9. (a) The probability that a bolt can be used, either immediately or after being cut, is $0.85 + 0.10 = 0.95$.

(b) Let X be the number of bolts out of 10 that can be used, either immediately or after being cut. Then $X \sim \text{Bin}(10, 0.95)$.

$$\begin{aligned}
P(X < 9) &= 1 - P(X = 9) - P(X = 10) \\
&= 1 - \frac{10!}{9!(10-9)!}(0.95)^9(1-0.95)^{10-9} + \frac{10!}{10!(10-10)!}(0.95)^{10}(1-0.95)^{10-10} \\
&= 0.0861
\end{aligned}$$

11. (a) Let X be the number of components that function. Then $X \sim \text{Bin}(5, 0.9)$.

$$\begin{aligned}
P(X \geq 3) = &\frac{5!}{3!(5-3)!}(0.9)^3(1-0.9)^{5-3} + \frac{5!}{4!(5-4)!}(0.9)^4(1-0.9)^{5-4} \\
&+ \frac{5!}{5!(5-5)!}(0.9)^5(1-0.9)^{5-5} = 0.9914
\end{aligned}$$

(b) We need to find the smallest value of n so that $P(X \leq 2) < 0.10$ when $X \sim \text{Bin}(n, 0.9)$. Consulting Table A.1, we find that if $n = 3$, $P(X \leq 2) = 0.271$, and if $n = 4$, $P(X \leq 2) = 0.052$. The smallest value of n is therefore $n = 4$.

13. (a) $X \sim \text{Bin}(10, 0.15)$.

$$\begin{aligned}
P(X \geq 7) &= P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10) \\
&= \frac{10!}{7!(10-7)!}(0.15)^7(1-0.15)^{10-7} + \frac{10!}{8!(10-8)!}(0.15)^8(1-0.15)^{10-8} \\
&\quad + \frac{10!}{9!(10-9)!}(0.15)^9(1-0.15)^{10-9} + \frac{10!}{10!(10-10)!}(0.15)^{10}(1-0.15)^{10-10} \\
&= 1.2591 \times 10^{-4} + 8.3326 \times 10^{-6} + 3.2677 \times 10^{-7} + 5.7665 \times 10^{-9} \\
&= 1.346 \times 10^{-4}
\end{aligned}$$

(b) Yes, only about 13 or 14 out of every 100,000 samples of size 10 would have 7 or more defective items.

(c) Yes, because 7 defectives in a sample of size 10 is an unusually large number for a good shipment.

(d)
$$\begin{aligned}
P(X \geq 2) &= 1 - P(X < 2) \\
&= 1 - P(X = 0) - P(X = 1) \\
&= 1 - \frac{10!}{0!(10-0)!}(0.15)^0(1-0.15)^{10-0} - \frac{10!}{1!(10-1)!}(0.15)^1(1-0.15)^{10-1} \\
&= 1 - 0.19687 - 0.34743 \\
&= 0.4557
\end{aligned}$$

(e) No, in about 45% of the samples of size 10, 2 or more items would be defective.

(f) No, because 2 defectives in a sample of size 10 is not an unusually large number for a good shipment.

15. (a) Let X be the number of bits that are reversed. Then $X \sim \text{Bin}(5, 0.3)$. The correct value is assigned if $X \leq 2$.

$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) = \dfrac{5!}{0!(5-0)!}(0.3)^0(1-0.3)^{5-0} + \dfrac{5!}{1!(5-1)!}(0.3)^1(1-0.3)^{5-1}$

$$+\frac{5!}{2!(5-2)!}(0.3)^2(1-0.3)^{5-2} = 0.8369$$

(b) We need to find the smallest odd value of n so that $P(X \leq (n-1)/2) \geq 0.90$ when $X \sim \text{Bin}(n, 0.3)$. Consulting Table A.1, we find that if $n = 3$, $P(X \leq 1) = 0.784$, if $n = 5$, $P(X \leq 2) = 0.837$, if $n = 7$, $P(X \leq 3) = 0.874$, and if $n = 9$, $P(X \leq 4) = 0.901$. The smallest value of n is therefore $n = 9$.

Section 4.2

1. (a) $P(X = 2) = e^{-3}\dfrac{3^2}{2!} = 0.2240$

(b) $P(X = 0) = e^{-3}\dfrac{3^0}{0!} = 0.0498$

(c)
$$\begin{aligned} P(X < 3) &= P(X=0) + P(X=1) + P(X=2) \\ &= e^{-3}\frac{3^0}{0!} + e^{-3}\frac{3^1}{1!} + e^{-3}\frac{3^2}{2!} \\ &= 0.049787 + 0.14936 + 0.22404 \\ &= 0.4232 \end{aligned}$$

(d)
$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) \\ &= 1 - P(X=0) - P(X=1) - P(X=2) \\ &= 1 - e^{-3}\frac{3^0}{0!} - e^{-3}\frac{3^1}{1!} - e^{-3}\frac{3^2}{2!} \\ &= 1 - 0.049787 - 0.14936 - 0.22404 \\ &= 0.5768 \end{aligned}$$

(e) Since $X \sim \text{Poisson}(3)$, $\mu_X = 3$.

(f) Since $X \sim \text{Poisson}(3)$, $\sigma_X = \sqrt{3} = 1.7321$.

3. X is the number of successes in $n = 1000$ independent Bernoulli trials, each of which has success probability $p = 0.002$. The mean of X is $np = (1000)(0.002) = 2$. Since n is large and p is small, $X \sim \text{Poisson}(2)$ to a very close approximation.

(a) $P(X = 4) = e^{-2}\dfrac{2^4}{4!} = 0.0902$.

(b)
$$\begin{aligned} P(X \le 1) &= P(X = 0) + P(X = 1) \\ &= e^{-2}\frac{2^0}{0!} + e^{-2}\frac{2^1}{1!} \\ &= 0.13534 + 0.27067 \\ &= 0.4060 \end{aligned}$$

(c)
$$\begin{aligned} P(1 \le X < 4) &= P(X = 1) + P(X = 2) + P(X = 3) \\ &= e^{-2}\frac{2^1}{1!} + e^{-2}\frac{2^2}{2!} + e^{-2}\frac{2^3}{3!} \\ &= 0.27067 + 0.27067 + 0.18045 \\ &= 0.7218 \end{aligned}$$

(d) Since $X \sim \text{Poisson}(2)$, $\mu_X = 2$.

(e) Since $X \sim \text{Poisson}(2)$, $\sigma_X = \sqrt{2} = 1.4142$.

5. (a) Let X be the number of hits in one minute. Since the mean rate is 4 messages per minute, $X \sim \text{Poisson}(4)$.

$P(X = 5) = e^{-4}\dfrac{4^5}{5!} = 0.1563$

(b) Let X be the number of hits in 1.5 minutes. Since the mean rate is 4 messages per minute, $X \sim \text{Poisson}(6)$.

$P(X = 9) = e^{-6}\dfrac{6^9}{9!} = 0.0688$

(c) Let X be the number of hits in 30 seconds. Since the mean rate is 4 messages per minute, $X \sim$ Poisson(2).

$$\begin{aligned} P(X < 3) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= e^{-2}\frac{2^0}{0!} + e^{-2}\frac{2^1}{1!} + e^{-2}\frac{2^2}{2!} \\ &= 0.13534 + 0.27067 + 0.27067 \\ &= 0.6767 \end{aligned}$$

7. Let X be the number of messages that fail to reach the base station. Then X is the number of successes in
$n = 1000$ Bernoulli trials, each of which has success probability $p = 0.005$.
The mean of X is $np = (1000)(0.005) = 5$.
Since n is large and p is small, $X \sim$ Poisson(5) to a very close approximation.

(a) $P(X = 3) = e^{-5}\frac{5^3}{3!} = 0.14037$

(b) The event that fewer than 994 messages reach the base station is the same as the event that more than 6 messages fail to reach the base station, or equivalently, that $X > 6$.

$$\begin{aligned} P(X > 6) = 1 - P(X \le 6) &= 1 - e^{-5}\frac{5^0}{0!} - e^{-5}\frac{5^1}{1!} - e^{-5}\frac{5^2}{2!} - e^{-5}\frac{5^3}{3!} - e^{-5}\frac{5^4}{4!} - e^{-5}\frac{5^5}{5!} - e^{-5}\frac{5^6}{6!} \\ &= 1 - 0.00674 - 0.03369 - 0.08422 - 0.14037 - 0.17547 - 0.17547 - 0.14622 = 0.2378 \end{aligned}$$

(c) $\mu_X = 5$

(d) $\sigma_X = \sqrt{5} = 2.2361$

9. (ii). Let $X \sim \text{Bin}(n, p)$ where $\mu_X = np = 3$. Then $\sigma_X^2 = np(1-p)$, which is less than 3 because $1 - p < 1$. Now let Y have a Poisson distribution with mean 3. The variance of Y is also equal to 3, because the variance of a Poisson random variable is always equal to its mean. Therefore Y has a larger variance than X.

11. If the mean number of particles is exactly 7 per mL, then $X \sim$ Poisson(7).

(a) $P(X \le 1) = P(X = 0) + P(X = 1) = e^{-7}\frac{7^0}{0!} + e^{-7}\frac{7^1}{1!} = 7.295 \times 10^{-3}$

(b) Yes. If the mean concentration is 7 particles per mL, then only about 7 in every thousand 1 mL samples will contain 1 or fewer particles.

(c) Yes, because 1 particle in a 1 mL sample is an unusually small number if the mean concentration is 7 particles per mL.

(d) $P(X \leq 6) = \sum_{x=0}^{6} P(X = x) = \sum_{x=0}^{6} e^{-7}\dfrac{7^x}{x!} = 0.4497$

(e) No. If the mean concentration is 7 particles per mL, then about 45% of all 1 mL samples will contain 6 or fewer particles.

(f) No, because 6 particles in a 1 mL sample is not an unusually small number if the mean concentration is 7 particles per mL.

Section 4.3

1. (a) Using Table A.2: 0.7734

(b) Using Table A.2: $0.8749 - 0.6554 = 0.2195$

(c) Using Table A.2: $0.7580 - 0.2743 = 0.4837$

(d) Using Table A.2: $0.7734 + (1 - 0.9032) = 0.8702$

3. (a) Using Table A.2: $c = 1$

(b) Using Table A.2: $c = 0.86$

(c) Using Table A.2: $c = 1.50$

(d) Using Table A.2: $c = -1.70$

(e) Using Table A.2: $c = 1.45$

5. (a) $z = (550 - 460)/80 = 1.13$. The area to the right of $z = 1.13$ is 0.1292.

   (b) The z-score of the 35th percentile is ≈ -0.39.
   The 10th percentile is therefore $\approx 460 - 0.39(80) = 428.8$.

   (c) $z = (600 - 460)/80 = 1.75$. The area to the left of $z = 1.75$ is 0.9599.
   Therefore a score of 600 is on the 96th percentile, approximately.

   (d) For 420, $z = (420 - 460)/80 = -0.50$. For 520, $z = (520 - 460)/80 = 0.75$.
   The area between $z = -0.50$ and $z = 0.75$ is $0.7734 - 0.3085 = 0.4649$.

7. (a) $z = (1800 - 1400)/200 = 2.00$. The area to the right of $z = 2.00$ is 0.0228.

   (b) The z-score of the 10th percentile is ≈ -1.28.
   The 10th percentile is therefore $\approx 1400 - 1.28(200) = 1144$.

   (c) $z = (1645 - 1400)/200 = 1.23$. The area to the left of $z = 1.23$ is 0.8907.
   Therefore a score of 1645 is on the 89th percentile, approximately.

   (d) For 1350, $z = (1350 - 1400)/200 = -0.25$. For 1550, $z = (1550 - 1400)/200 = 0.75$.
   The area between $z = -0.25$ and $z = 0.75$ is $0.7734 - 0.4013 = 0.3721$.

9. (a) $z = (12 - 10)/1.4 = 1.43$. The area to the right of $z = 1.43$ is $1 - 0.9236 = 0.0764$.

   (b) The z-score of the 25th percentile is ≈ -0.67.
   The 25th percentile is therefore $\approx 10 - 0.67(1.4) = 9.062$ GPa.

   (c) The z-score of the 95th percentile is ≈ 1.645.
   The 25th percentile is therefore $\approx 10 + 1.645(1.4) = 12.303$ GPa.

11. (a) $z = (6 - 4.9)/0.6 = 1.83$. The area to the right of $z = 1.83$ is $1 - 0.9664 = 0.0336$.
   The process will be shut down on 3.36% of days.

(b) $z = (6 - 5.2)/0.4 = 2.00$. The area to the right of $z = 2.00$ is $1 - 0.9772 = 0.0228$.

Since a process with this broth will be shut down on 2.28% of days, this broth will result in fewer days of production lost.

Section 4.4

1. Let Y be the lifetime of the component.

(a) $E(Y) = e^{\mu+\sigma^2/2} = e^{1.2+(0.4)^2/2} = 3.5966$

(b) $P(3 < Y < 6) = P(\ln 3 < \ln Y < \ln 6) = P(1.0986 < \ln Y < 1.7918)$. $\ln Y \sim N(1.2, 0.4^2)$.

The z-score of 1.0986 is $(1.0986 - 1.2)/0.4 = -0.25$.

The z-score of 1.7918 is $(1.7918 - 1.2)/0.4 = 1.48$.

The area between $z = -0.25$ and $z = 1.48$ is $0.9306 - 0.4013 = 0.5293$.

Therefore $P(3 < Y < 6) = 0.5293$.

(c) Let m be the median of Y. Then $P(Y \leq m) = P(\ln Y \leq \ln m) = 0.5$.

Since $\ln Y \sim N(1.2, 0.4^2)$, $P(\ln Y < 1.2) = 0.5$. Therefore $\ln m = 1.2$, so $m = e^{1.2} = 3.3201$.

(d) Let y_{90} be the 90th percentile of Y. Then $P(Y \leq y_{90}) = P(\ln Y \leq \ln y_{90}) = 0.90$.

The z-score of the 90th percentile is approximately $z = 1.28$.

Therefore the z-score of $\ln y_{90}$ must be 1.28, so $\ln y_{90}$ satisfies the equation $1.28 = (\ln y_{90} - 1.2)/0.4$.

$\ln y_{90} = 1.712$, so $y_{90} = e^{1.712} = 5.540$.

3. Let Y represent the BMI for a randomly chosen man aged 25–34.

(a) $E(Y) = e^{\mu+\sigma^2/2} = e^{3.215+(0.157)^2/2} = 25.212$

(b) $V(Y) = e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2} = e^{2(3.215)+2(0.157)^2} - e^{2(3.215)+(0.157)^2} = 15.86285$.

The standard deviation is $\sqrt{V(Y)} = \sqrt{e^{2(3.215)+2(0.157)^2} - e^{2(3.215)+(0.157)^2}} = \sqrt{15.86285} = 3.9828$.

(c) Let m be the median of Y. Then $P(Y \leq m) = P(\ln Y \leq \ln m) = 0.5$.

Since $\ln Y \sim N(3.215, 0.157^2)$, $P(\ln Y < 3.215) = 0.5$.

Therefore $\ln m = 3.215$, so $m = e^{3.215} = 24.903$.

(d) $P(Y < 22) = P(\ln Y < \ln 22) = P(\ln Y < 3.0910)$.
The z-score of 3.0910 is $(3.0910 - 3.215)/0.157 = -0.79$.
The area to the left of $z = -0.79$ is 0.2148.
Therefore $P(Y < 22) = 0.2148$.

(e) Let y_{75} be the 75th percentile of Y. Then $P(Y \le y_{75}) = P(\ln Y \le \ln y_{75}) = 0.75$.
The z-score of the 75th percentile is approximately $z = 0.67$.
Therefore the z-score of $\ln y_{75}$ must be 0.67, so $\ln y_{75}$ satisfies the equation $0.67 = (\ln y_{75} - 3.215)/0.157$.
$\ln y_{75} = 3.3202$, so $y_{75} = e^{3.3202} = 27.666$.

5. Let X represent the price of a share of company A one year from now. Let Y represent the price of a share of company B one year from now.

(a) $E(X) = e^{0.05+(0.1)^2/2} = \$1.0565$

(b) $P(X > 1.20) = P(\ln X > \ln 1.20) = P(\ln X > 0.1823)$.
The z-score of 0.1823 is $(0.1823 - 0.05)/0.1 = 1.32$.
The area to the right of $z = 1.32$ is $1 - 0.9066 = 0.0934$.
Therefore $P(X > 1.20) = 0.0934$.

(c) $E(Y) = e^{0.02+(0.2)^2/2} = \$1.0408$

(d) $P(Y > 1.20) = P(\ln Y > \ln 1.20) = P(\ln Y > 0.1823)$.
The z-score of 0.1823 is $(0.1823 - 0.02)/0.2 = 0.81$.
The area to the right of $z = 0.81$ is $1 - 0.7910 = 0.2090$.
Therefore $P(Y > 1.20) = 0.2090$.

Section 4.5

1. (a) $\mu_T = 1/0.5 = 2$

(b) $\sigma_T^2 = 1/(0.5^2) = 4$

(c) $P(T > 5) = 1 - P(T \le 5) = 1 - (1 - e^{-0.5(5)}) = 0.0821$

(d) Let m be the median. Then $P(T \leq m) = 0.5$.
$P(T \leq m) = 1 - e^{-0.5m} = 0.5$, so $e^{-0.5m} = 0.5$.
Solving for m yields $m = 1.3863$.

3. Let X be the diameter in microns.

(a) $\mu_X = 1/\lambda = 1/0.25 = 4$ microns

(b) $\sigma_X = 1/\lambda = 1/0.25 = 4$ microns

(c) $P(X < 3) = 1 - e^{-0.25(3)} = 0.5276$

(d) $P(X > 11) = 1 - (1 - e^{-0.25(11)}) = 0.0639$

(e) Let m be the median. Then $P(T \leq m) = 0.5$.
$P(T \leq m) = 1 - e^{-0.25m} = 0.5$, so $e^{-0.25m} = 0.5$.
Solving for m yields $m = 2.7726$ microns.

(f) Let x_{75} be the 75th percentile, which is also the third quartile. Then $P(T \leq x_{75}) = 0.75$.
$P(T \leq x_{75}) = 1 - e^{-0.25x_{75}} = 0.75$, so $e^{-0.25x_{75}} = 0.25$.
Solving for x_{75} yields $x_{75} = 5.5452$ microns.

(g) Let x_{99} be the 99th percentile. Then $P(T \leq x_{99}) = 0.75$.
$P(T \leq x_{99}) = 1 - e^{-0.25x_{99}} = 0.99$, so $e^{-0.25x_{99}} = 0.01$.
Solving for x_{99} yields $x_{99} = 18.4207$ microns.

5. No. If the lifetimes were exponentially distributed, the proportion of used components lasting longer than 5 years would be the same as the proportion of new components lasting longer than 5 years, because of the lack of memory property.

Section 4.6

1. Let T be the waiting time.

(a) $\mu_T = (0 + 10)/2 = 5$ minutes.

(b) $\sigma_T = \sqrt{\dfrac{(10-0)^2}{12}} = 2.8868$ minutes

3. (a) $\mu_T = 6/2 = 3$

(b) $\sigma_T = \sqrt{6/2^2} = 1.2247$

5. (a) $\alpha = 0.5$, $\beta = 2$, so $1/\alpha = 2$ which is an integer.
$\mu_T = (1/2)2! = 2/2 = 1$.

(b) $\sigma_T^2 = (1/2^2)[4! - (2!)^2] = (1/4)(24 - 4) = 5$

(c) $P(T \leq 2) = 1 - e^{-[(2)(2)]^{0.5}} = 1 - e^{-2} = 0.8647$

(d) $P(T > 3) = 1 - P(T \leq 3) = 1 - \left(1 - e^{-[(2)(3)]^{0.5}}\right) = e^{-6^{0.5}} = 0.0863$

(e)
$$\begin{aligned}
P(1 < T \leq 2) &= P(T \leq 2) - P(T \leq 1) \\
&= \left(1 - e^{-[(2)(2)]^{0.5}}\right) - \left(1 - e^{-[(2)(1)]^{0.5}}\right) \\
&= e^{-[(2)(1)]^{0.5}} - e^{-[(2)(2)]^{0.5}} \\
&= e^{-2^{0.5}} - e^{-4^{0.5}} \\
&= 0.24312 - 0.13534 \\
&= 0.1078
\end{aligned}$$

7. Let T be the lifetime in hours of a bearing.

(a) $P(T > 1000) = 1 - P(T \leq 1000) = 1 - (1 - e^{-[(0.0004474)(1000)]^{2.25}}) = e^{-[(0.0004474)(1000)]^{2.25}} = 0.8490$

(b) $P(T < 2000) = P(T \leq 2000) = 1 - e^{-[(0.0004474)(2000)]^{2.25}} = 0.5410$

(c) Let m be the median.
Then $P(T \leq m) = 0.5$, so $1 - e^{-[(0.0004474)(m)]^{2.25}} = 0.5$, and $e^{-[(0.0004474)(m)]^{2.25}} = 0.5$.
$(0.0004474m)^{2.25} = -\ln 0.5 = 0.693147$
$0.0004474m = (0.693147)^{1/2.25} = 0.849681$
$m = 0.849681/0.0004474 = 1899.2$ hours

(d) $h(t) = \alpha\beta^\alpha t^{\alpha-1} = 2.25(0.0004474^{2.25})(2000^{2.25-1}) = 8.761 \times 10^{-4}$

9. Let T be the lifetime of a fan.

(a) $P(T > 10,000) = 1 - (1 - e^{-[(0.0001)(10,000)]^{1.5}}) = e^{-[(0.0001)(10,000)]^{1.5}} = 0.3679$

(b) $P(T < 5000) = P(T \leq 5000) = 1 - e^{-[(0.0001)(5000)]^{1.5}} = 0.2978$

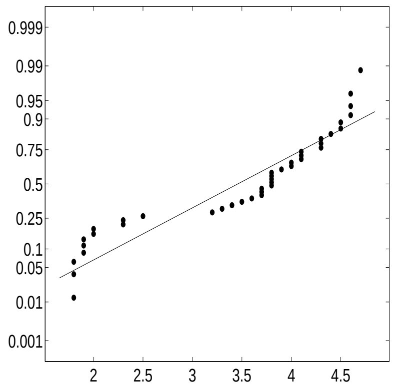
(c)
$$\begin{aligned} P(3000 < T < 9000) &= P(T \leq 9000) - P(T \leq 3000) \\ &= (1 - e^{-[(0.0001)(9000)]^{1.5}}) - (1 - e^{-[(0.0001)(3000)]^{1.5}}) \\ &= 0.4227 \end{aligned}$$

Section 4.7

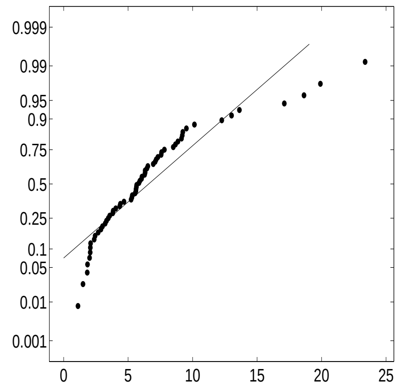
1. (a) No

(b) No

(c) Yes

3. 

These data do not appear to come from an approximately normal distribution.

5. 

The PM data do not appear to come from an approximately normal distribution.

7. Yes. If the logs of the PM data come from a normal population, then the PM data come from a lognormal population, and vice versa.

Section 4.8

1. (a) Let $X_1, ..., X_{144}$ be the volumes in the 144 bottles.

   Then $\overline{X}$ is approximately normally distributed with mean $\mu_{\overline{X}} = 12.01$ and $\sigma_{\overline{X}} = 0.2/\sqrt{144} = 0.01667$.

   The z-score of 12.00 is $(12.00 - 12.01)/0.01667 = -0.60$.

   The area to the left of $z = -0.60$ is 0.2743.

   $P(\overline{X} < 12) = 0.2743$.

   (b) $\overline{X}$ is approximately normally distributed with mean $\mu_{\overline{X}} = 12.03$ and $\sigma_{\overline{X}} = 0.2/\sqrt{144} = 0.01667$.

   The z-score of 12.00 is $(12.00 - 12.03)/0.01667 = -1.80$.

   The area to the left of $z = -1.80$ is 0.0359.

   $P(\overline{X} < 12) = 0.0359$.

3. (a) Let $X_1, ..., X_{50}$ be the weights of the 50 coatings.

   Then $\overline{X}$ is approximately normally distributed with mean $\mu_{\overline{X}} = 125$ and $\sigma_{\overline{X}} = 10/\sqrt{50} = 1.41421$.

   The z-score of 128 is $(128 - 125)/1.41421 = 2.12$.

   The area to the right of $z = 2.12$ is 0.0170.

   $P(\overline{X} > 128) = 0.0170$.

   (b) Let s_{90} denote the 90th percentile

   The z-score of the 90th percentile is approximately $z = 1.28$.

   Therefore s_{90} satisfies the equation $1.28 = (s_{90} - 125)/1.41421$.

$s_{90} = 126.81$.

(c) Let n be the necessary sample size. Then $\overline{X}$ is approximately normally distributed with mean $\mu_{\overline{X}} = 125$ and $\sigma_{\overline{X}} = 10/\sqrt{n}$.

Since $P(\overline{X} < 123) = 0.02$, 123 is the 2nd percentile of the distribution of $\overline{X}$.

The z-score of the 2nd percentile is approximately $z = -2.05$.

Therefore $123 = 125 - 2.05(10/\sqrt{n})$. Solving for n yields $n \approx 105$.

5. Let X represent the number of bearings that meet the specification.

Then $X \sim \text{Bin}(500, 0.90)$, so X is approximately normal with mean $\mu_X = 500(0.90) = 450$ and standard deviation $\sigma_X = \sqrt{500(0.9)(0.1)} = 6.7082$.

To find $P(X > 440)$, use the continuity correction and find the z-score of 440.5.

The z-score of 440.5 is $(440.5 - 450)/6.7082 = -1.42$.

The area to the right of $z = -1.42$ is $1 - 0.0778 = 0.9222$.

$P(X > 90) = 0.9222$.

7. (a) Let $X_1, ..., X_{80}$ be the breaking strengths of the 80 fabric pieces.

Then $\overline{X}$ is approximately normally distributed with mean $\mu_{\overline{X}} = 1.86$ and $\sigma_{\overline{X}} = 0.27/\sqrt{80} = 0.030187$.

The z-score of 1.8 is $(1.8 - 1.86)/0.030187 = -1.99$.

The area to the left of $z = -1.99$ is 0.0233.

$P(\overline{X} < 1.8) = 0.0233$.

(b) Let x_{80} denote the 80th percentile

The z-score of the 80th percentile is approximately $z = 0.84$.

Therefore x_{80} satisfies the equation $0.84 = (x_{80} - 1.86)/0.030187$.

$x_{80} = 1.8854$ mm.

(c) Let n be the necessary sample size. Then $\overline{X}$ is approximately normally distributed with mean $\mu_{\overline{X}} = 1.86$ and $\sigma_{\overline{X}} = 0.27/\sqrt{n}$.

Since $P(\overline{X} < 1.8) = 0.01$, 1.8 is the 1st percentile of the distribution of $\overline{X}$.

The z-score of the 1st percentile is approximately $z = -2.33$.

Therefore $1.8 = 1.86 - 2.33(0.27/\sqrt{n})$. Solving for n yields $n \approx 110$.

9. From the results of Example 4.30, the probability that a randomly chosen wire has no flaws is 0.48.

Let X be the number of wires in a sample of 225 that have no flaws.

Then $X \sim \text{Bin}(225, 0.48)$, so $\mu_X = 225(0.48) = 108$, and $\sigma_X^2 = 225(0.48)(0.52) = 56.16$.

To find $P(X < 110)$, use the continuity correction and find the z-score of 109.5.

The z-score of 109.5 is $(109.5 - 108)/\sqrt{56.16} = 0.20$.

The area to the left of $z = 0.20$ is 0.5793.

$P(X < 110) = 0.5793$.

11. (a) If the claim is true, then $X \sim \text{Bin}(1000, 0.05)$, so X is approximately normal with mean $\mu_X = 1000(0.05) = 50$ and $\sigma_X = \sqrt{1000(0.05)(0.95)} = 6.89202$.

To find $P(X \geq 75)$, use the continuity correction and find the z-score of 74.5.

The z-score of 74.5 is $(74.5 - 50)/6.89202 = 3.55$.

The area to the right of $z = 3.55$ is $1 - 0.9998 = 0.0002$.

$P(X \geq 75) = 0.0002$.

(b) Yes. Only about 2 in 10,000 samples of size 1000 will have 75 or more nonconforming tiles if the goal has been reached.

(c) No, because 75 nonconforming tiles in a sample of 1000 is an unusually large number if the goal has been reached.

(d) If the claim is true, then $X \sim \text{Bin}(1000, 0.05)$, so X is approximately normal with mean $\mu_X = 1000(0.05) = 50$ and $\sigma_X = \sqrt{1000(0.05)(0.95)} = 6.89202$.

To find $P(X \geq 53)$, use the continuity correction and find the z-score of 52.5.

The z-score of 52.5 is $(52.5 - 50)/6.89202 = 0.36$.

The area to the right of $z = 0.36$ is $1 - 0.6406 = 0.3594$.

$P(X \geq 53) = 0.3594$.

(e) No. More than 1/3 of the samples of size 1000 will have 53 or more nonconforming tiles if the goal has been reached.

(f) Yes, because 53 nonconforming tiles in a sample of 1000 is not an unusually large number if the goal has been reached.

Supplementary Exercises for Chapter 4

1. Let X be the number of people out of 105 who appear for the flight.

Then $X \sim \text{Bin}(105, 0.9)$, so X is approximately normal with mean $\mu_X = 105(0.9) = 94.5$ and standard deviation $\sigma_X = \sqrt{105(0.9)(0.1)} = 3.0741$.

To find $P(X \leq 100)$, use the continuity correction and find the z-score for 100.5.

The z-score of 100.5 is $(100.5 - 94.5)/3.0741 = 1.95$.

The area to the left of $z = 1.95$ is 0.9744.

$P(X \leq 100) = 0.9744$.

3. (a) Let X be the number of plants out of 10 that have green seeds. Then $X \sim \text{Bin}(10, 0.25)$.

$$P(X = 3) = \frac{10!}{3!(10-3)!}(0.25)^3(1-0.25)^{10-3} = 0.2503.$$

(b)
$$\begin{aligned}
P(X > 2) &= 1 - P(X \leq 2) \\
&= 1 - P(X = 0) - P(X = 1) - P(X = 2) \\
&= 1 - \frac{10!}{0!(10-0)!}(0.25)^0(1-0.25)^{10-0} - \frac{10!}{1!(10-1)!}(0.25)^1(1-0.25)^{10-1} \\
&\quad - \frac{10!}{2!(10-2)!}(0.25)^2(1-0.25)^{10-2} \\
&= 0.4744
\end{aligned}$$

(c) Let Y be the number of plants out of 100 that have green seeds.

Then $Y \sim \text{Bin}(100, 0.25)$ so Y is approximately normal with mean $\mu_Y = 100(0.25) = 25$ and standard deviation $\sigma_Y = \sqrt{100(0.25)(0.75)} = 4.3301$.

To find $P(Y > 30)$, use the continuity correction and find the z-score for 30.5.

The z-score of 30.5 is $(30.5 - 25)/4.3301 = 1.27$.

The area to the right of $z = 1.27$ is $1 - 0.8980 = 0.1020$.

$P(Y > 30) = 0.1020$.

(d) To find $P(30 \leq Y \leq 35)$, use the continuity correction and find the z-scores for 29.5 and 35.5.

The z-score of 29.5 is $(29.5 - 25)/4.3301 = 1.04$.

The z-score of 35.5 is $(35.5 - 25)/4.3301 = 2.42$.

The area to between $z = 1.04$ and $z = 2.42$ is $0.9922 - 0.8508 = 0.1414$.

$P(30 \leq Y \leq 35) = 0.1414$.

(e) Fewer than 80 have yellow seeds if more than 20 have green seeds.

To find $P(Y > 20)$, use the continuity correction and find the z-score for 20.5.

The z-score of 20.5 is $(20.5 - 25)/4.3301 = -1.04$.

The area to the right of $z = -1.04$ is $1 - 0.1492 = 0.8508$.

$P(Y > 20) = 0.8508$.

5. Let X denote the number of devices that fail. Then $X \sim \text{Bin}(10, 0.01)$.

(a) $P(X=0) = \dfrac{10!}{0!(10-0)!}(0.01)^0(1-0.01)^{10-0} = 0.99^{10} = 0.9044$

(b)
$$\begin{aligned} P(X \geq 2) &= 1 - P(X \leq 1) \\ &= 1 - P(X=0) - P(X=1) \\ &= 1 - \frac{10!}{0!(10-0)!}(0.01)^0(1-0.01)^{10-0} - \frac{10!}{1!(10-1)!}(0.01)^1(1-0.01)^{10-1} \\ &= 0.00427 \end{aligned}$$

(c) Let p be the required probability. Then $X \sim \text{Bin}(10, p)$.

$P(X=0) = \dfrac{10!}{0!(10-0)!}p^0(1-p)^{10-0} = (1-p)^{10} = 0.95.$

Solving for p yields $p = 0.00512$.

7. (a) The probability that a normal random variable is within one standard deviation of its mean is the area under the normal curve between $z=-1$ and $z=1$. This area is $0.8413 - 0.1587 = 0.6826$.

(b) The quantity $\mu + z\sigma$ is the 90th percentile of the distribution of X. The 90th percentile of a normal distribution is 1.28 standard deviations above the mean. Therefore $z = 1.28$.

(c) The z-score of 15 is $(15-10)/\sqrt{2.6} = 3.10$.

The area to the right of $z = 3.10$ is $1 - 0.9990 = 0.0010$.

$P(X > 15) = 0.0010$.

9. (a) The z-score of 215 is $(215-200)/10 = 1.5$.

The area to the right of $z = 1.5$ is $1 - 0.9332 = 0.0668$.

The probability that the clearance is greater than 215 μm is 0.0668.

(b) The z-score of 180 is $(180-200)/10 = -2.00$.

The z-score of 205 is $(205-200)/10 = 0.50$.

The area between $z = -2.00$ and $z = 0.50$ is $0.6915 - 0.0228 = 0.6687$.

The probability that the clearance is between 180 and 205 μm is 0.6687.

(c) Let X be the number of valves whose clearances are greater than 215 μm.

From part (a), the probability that a valve has a clearance greater than 215 μm is 0.0668, so

$X \sim \text{Bin}(6, 0.0668)$.

$P(X = 2) = \dfrac{6!}{2!(6-2)!}(0.0668)^2(1 - 0.0668)^{6-2} = 0.0508$.

11. (a) Let X be the number of assemblies in a sample of 300 that are oversize.

Then $X \sim \text{Bin}(300, 0.05)$, so X is approximately normal with mean $\mu_X = 300(0.05) = 15$ and standard deviation $\sigma_X = \sqrt{300(0.05)(0.95)} = 3.7749$.

To find $P(X < 20)$, use the continuity correction and find the z-score of 19.5.

The z-score of 19.5 is $(19.5 - 15)/3.7749 = 1.19$.

The area to the left of $z = 1.19$ is 0.8830.

$P(X < 20) = 0.8830$.

(b) Let Y be the number of assemblies in a sample of 10 that are oversize. Then $X \sim \text{Bin}(10, 0.05)$.

$P(X \geq 1) = 1 - P(X = 0) = 1 - \dfrac{10!}{0!(10-0)!}(0.05)^0(1 - 0.05)^{10-0} = 0.4013$.

(c) Let p be the required probability, and let X represent the number of assemblies in a sample of 300 that are oversize.

Then $X \sim \text{Bin}(300, p)$, so X is approximately normal with mean $\mu_X = 300p$ and standard deviation $\sigma_X = \sqrt{300p(1-p)}$.

$P(X \geq 20) = 0.01$.

Using the continuity correction, 19.5 is the 1st percentile of the distribution of X.

The z-score of the 1st percentile is approximately $z = -2.33$.

The z-score can be expressed in terms of p by $-2.33 = (19.5 - 300p)/\sqrt{300p(1-p)}$.

This equation can be rewritten as $91,628.67p^2 - 13,328.67p + 380.25 = 0$.

Solving for p yields $p = 0.0390$. (0.1065 is a spurious root.)

13. (a) Let T represent the lifetime of a bearing.

$P(T > 1) = 1 - P(T \leq 1) = 1 - (1 - e^{-[(0.8)(1)]^{1.5}}) = 0.4889$

(b) $P(T \leq 2) = 1 - e^{-[(0.8)(2)]^{1.5}} = 0.8679$

15. (a) S is approximately normal with mean $\mu_S = 75(12.2) = 915$ and $\sigma_S = 0.1\sqrt{75} = 0.86603$.

The z-score of 914.8 is $(914.8 - 915)/0.86603 = -0.23$.

The area to the left of $z = -0.23$ is 0.4090.

$P(S < 914.8) = 0.4090$.

(b) No. More than 40% of the samples will have a total weight of 914.8 ounces or less if the claim is true.

(c) No, because a total weight of 914.8 ounces is not unusually small if the claim is true.

(d) S is approximately normal with mean $\mu_S = 75(12.2) = 915$ and $\sigma_S = 0.1\sqrt{75} = 0.86603$.
The z-score of 910.3 is $(910.3 - 915)/0.86603 = -5.43$.
The area to the left of $z = -5.43$ is negligible.
$P(S < 910.3) \approx 0$.

(e) Yes. Almost none of the samples will have a total weight of 910.3 ounces or less if the claim is true.

(f) Yes, because a total weight of 910.3 ounces is unusually small if the claim is true.

Chapter 5

Section 5.1

1. iii. By definition, an estimator is unbiased if its mean is equal to the true value.

3. (a) We denote the mean of $\widehat{\mu}_1$ by $E(\widehat{\mu}_1)$ and the variance of $\widehat{\mu}_1$ by $V(\widehat{\mu}_1)$.

$E(\widehat{\mu}_1) = \dfrac{\mu_{X_1} + \mu_{X_2}}{2} = \dfrac{\mu + \mu}{2} = \mu.$

The bias of $\widehat{\mu}_1$ is $E(\widehat{\mu}_1) - \mu = \mu - \mu = 0.$

The variance of $\widehat{\mu}_1$ is $V(\widehat{\mu}_1) = \dfrac{\sigma^2 + \sigma^2}{4} = \dfrac{\sigma^2}{2} = \dfrac{1}{2}.$

The mean squared error of $\widehat{\mu}_1$ is the sum of the variance and the square of the bias, so $MSE(\widehat{\mu}_1) = \dfrac{1}{2} + 0^2 = \dfrac{1}{2}.$

(b) We denote the mean of $\widehat{\mu}_2$ by $E(\widehat{\mu}_2)$ and the variance of $\widehat{\mu}_2$ by $V(\widehat{\mu}_2)$.

$E(\widehat{\mu}_2) = \dfrac{\mu_{X_1} + 2\mu_{X_2}}{3} = \dfrac{\mu + 2\mu}{3} = \mu.$

The bias of $\widehat{\mu}_2$ is $E(\widehat{\mu}_2) - \mu = \mu - \mu = 0.$

The variance of $\widehat{\mu}_2$ is $V(\widehat{\mu}_2) = \dfrac{\sigma^2 + 4\sigma^2}{9} = \dfrac{5\sigma^2}{9} = \dfrac{5}{9}.$

The mean squared error of $\widehat{\mu}_2$ is the sum of the variance and the square of the bias, so $MSE(\widehat{\mu}_2) = \dfrac{5}{9} + 0^2 = \dfrac{5}{9}.$

(c) We denote the mean of $\widehat{\mu}_3$ by $E(\widehat{\mu}_3)$ and the variance of $\widehat{\mu}_3$ by $V(\widehat{\mu}_3)$.

$E(\widehat{\mu}_3) = \dfrac{\mu_{X_1} + \mu_{X_2}}{4} = \dfrac{\mu + \mu}{4} = \dfrac{\mu}{2}.$

The bias of $\widehat{\mu}_3$ is $E(\widehat{\mu}_3) - \mu = \dfrac{\mu}{2} - \mu = -\dfrac{\mu}{2}.$

The variance of $\widehat{\mu}_3$ is $V(\widehat{\mu}_3) = \dfrac{\sigma^2 + \sigma^2}{16} = \dfrac{\sigma^2}{8}.$

The mean squared error of $\widehat{\mu}_3$ is the sum of the variance and the square of the bias, so $MSE(\widehat{\mu}_3) = \dfrac{\sigma^2}{8} + \left(-\dfrac{\mu}{2}\right)^2 = \dfrac{2\mu^2 + 1}{8}.$

5. $\widehat{\mu}_3$ has smaller mean squared error than $\widehat{\mu}_2$ whenever $\dfrac{2\mu^2 + 1}{8} < \dfrac{5}{9}.$

Solving for μ yields $-1.3123 < \mu < 1.3123.$

Section 5.2

1. (a) 1.645

   (b) 1.37

   (c) 2.81

   (d) 1.15

3. The level is the proportion of samples for which the confidence interval will cover the true value. Therefore as the level goes up, the reliability goes up. This increase in reliability is obtained by increasing the width of the confidence interval. Therefore as the level goes up the precision goes down.

5. (a) $\overline{X} = 178$, $s = 14$, $n = 120$, $z_{.025} = 1.96$.
   The confidence interval is $178 \pm 1.96(14/\sqrt{120})$, or (175.495, 180.505).

   (b) $\overline{X} = 178$, $s = 14$, $n = 120$, $z_{.005} = 2.58$.
   The confidence interval is $178 \pm 2.58(14/\sqrt{120})$, or (174.703, 181.297).

   (c) $\overline{X} = 178$, $s = 14$, $n = 120$, so the upper confidence bound 180 satisfies $180 = 178 + z_{\alpha/2}(14/\sqrt{120})$.
   Solving for $z_{\alpha/2}$ yields $z_{\alpha/2} = 1.56$.
   The area to the right of $z = 1.56$ is $1 - 0.9406 = 0.0594$, so $\alpha/2 = 0.0594$.
   The level is $1 - \alpha = 1 - 2(0.0594) = 0.8812$, or 88.12%.

   (d) $z_{.025} = 1.96$. $1.96(14/\sqrt{n}) = 2$, so $n = 189$.

   (e) $z_{.005} = 2.58$. $2.58(14/\sqrt{n}) = 2$, so $n = 327$.

7. (a) $\overline{X} = 6230$, $s = 221$, $n = 100$, $z_{.025} = 1.96$.
   The confidence interval is $6230 \pm 1.96(221/\sqrt{100})$, or (6186.7, 6273.3).

   (b) $\overline{X} = 6230$, $s = 221$, $n = 100$, $z_{.005} = 2.58$.
   The confidence interval is $6230 \pm 2.58(221/\sqrt{100})$, or (6173.0, 6287.0).

(c) $\overline{X} = 6230$, $s = 221$, $n = 100$, so the upper confidence bound 6255 satisfies $6255 = 6230 + z_{\alpha/2}(221/\sqrt{100})$.
Solving for $z_{\alpha/2}$ yields $z_{\alpha/2} = 1.13$.
The area to the right of $z = 1.13$ is $1 - 0.8708 = 0.1292$, so $\alpha/2 = 0.1292$.
The level is $1 - \alpha = 1 - 2(0.1292) = 0.7416$, or 74.16%.

(d) $z_{.025} = 1.96$. $1.96(221/\sqrt{n}) = 25$, so $n = 301$.

(e) $z_{.005} = 2.58$. $2.58(221/\sqrt{n}) = 25$, so $n = 521$.

9. (a) $\overline{X} = 1.56$, $s = 0.1$, $n = 80$, $z_{.025} = 1.96$.
The confidence interval is $1.56 \pm 1.96(0.1/\sqrt{80})$, or (1.5381, 1.5819).

(b) $\overline{X} = 1.56$, $s = 0.1$, $n = 80$, $z_{.01} = 2.33$.
The confidence interval is $1.56 \pm 2.33(0.1/\sqrt{80})$, or (1.5339, 1.5861).

(c) $\overline{X} = 1.56$, $s = 0.1$, $n = 80$, so the upper confidence bound 1.58 satisfies $1.58 = 1.56 + z_{\alpha/2}(0.1/\sqrt{80})$.
Solving for $z_{\alpha/2}$ yields $z_{\alpha/2} = 1.79$.
The area to the right of $z = 1.79$ is $1 - 0.9633 = 0.0367$, so $\alpha/2 = 0.0367$.
The level is $1 - \alpha = 1 - 2(0.0367) = 0.9266$, or 92.66%.

(d) $z_{.025} = 1.96$. $1.96(0.1/\sqrt{n}) = 0.01$, so $n = 385$.

(e) $z_{.01} = 2.33$. $2.33(0.1/\sqrt{n}) = 0.01$, so $n = 543$.

11. (a) $\overline{X} = 29$, $s = 9$, $n = 81$, $z_{.025} = 1.96$.
The confidence interval is $29 \pm 1.96(9/\sqrt{81})$, or (27.04, 30.96).

(b) $\overline{X} = 29$, $s = 9$, $n = 81$, $z_{.005} = 2.58$.
The confidence interval is $29 \pm 2.58(9/\sqrt{81})$, or (26.42, 31.58).

(c) $\overline{X} = 29$, $s = 9$, $n = 81$, so the upper confidence bound 30.5 satisfies $30.5 = 29 + z_{\alpha/2}(9/\sqrt{81})$.
Solving for $z_{\alpha/2}$ yields $z_{\alpha/2} = 1.50$.
The area to the right of $z = 1.50$ is $1 - 0.9332 = 0.0668$, so $\alpha/2 = 0.0668$.
The level is $1 - \alpha = 1 - 2(0.0668) = 0.8664$, or 86.64%.

(d) $z_{.025} = 1.96$. $1.96(9/\sqrt{n}) = 1$, so $n = 312$.

(e) $z_{.005} = 2.58$. $2.58(9/\sqrt{n}) = 1$, so $n = 540$.

13. (a) $\overline{X} = 17.3$, $s = 1.2$, $n = 81$, $z_{.02} = 2.05$.
The lower confidence bound is $17.3 - 2.05(1.2/\sqrt{81}) = 17.027$.

(b) The lower confidence bound 17 satisfies $17 = 17.3 - z_\alpha(1.2/\sqrt{81})$.
Solving for z_α yields $z_\alpha = 2.25$.
The area to the left of $z = 2.25$ is $1 - \alpha = 0.9878$.
The level is 0.9878, or 98.78%.

15. (a) $\overline{X} = 1.69$, $s = 0.25$, $n = 63$, $z_{.01} = 2.33$.
The upper confidence bound is $1.69 + 2.33(0.25/\sqrt{63}) = 1.7634$.

(b) The upper confidence bound 1.75 satisfies $1.75 = 1.69 + z_\alpha(0.25/\sqrt{63})$.
Solving for z_α yields $z_\alpha = 1.90$.
The area to the left of $z = 1.90$ is $1 - \alpha = 0.9713$.
The level is 0.9713, or 97.13%.

17. (a) $\overline{X} = 72$, $s = 10$, $n = 150$, $z_{.02} = 2.05$.
The lower confidence bound is $72 - 2.05(10/\sqrt{150}) = 70.33$.

(b) The lower confidence bound 70 satisfies $70 = 72 - z_\alpha(10/\sqrt{150})$.
Solving for z_α yields $z_\alpha = 2.45$.
The area to the left of $z = 2.45$ is $1 - \alpha = 0.9929$.
The level is 0.9929, or 99.29%.

Section 5.3

1. (a) The proportion is $28/70 = 0.4$.

(b) $X = 28$, $n = 70$, $\tilde{p} = (28 + 2)/(70 + 4) = 0.405405$, $z_{.025} = 1.96$.
The confidence interval is $0.405405 \pm 1.96\sqrt{0.405405(1 - 0.405405)/(70 + 4)}$, or (0.294, 0.517).

(c) $X = 28$, $n = 70$, $\tilde{p} = (28 + 2)/(70 + 4) = 0.405405$, $z_{.01} = 2.33$.
The confidence interval is $0.405405 \pm 2.33\sqrt{0.405405(1 - 0.405405)/(70 + 4)}$, or (0.272, 0.538).

(d) Let n be the required sample size.
Then n satisfies the equation $0.10 = 1.96\sqrt{\tilde{p}(1 - \tilde{p})/(n + 4)}$.
Replacing $\tilde{p}$ with 0.405405 and solving for n yields $n = 89$.

(e) Let n be the required sample size.
Then n satisfies the equation $0.10 = 2.33\sqrt{\tilde{p}(1 - \tilde{p})/(n + 4)}$.
Replacing $\tilde{p}$ with 0.405405 and solving for n yields $n = 127$.

3. (a) $X = 13$, $n = 87$, $\tilde{p} = (13 + 2)/(87 + 4) = 0.16484$, $z_{.025} = 1.96$.
The confidence interval is $0.16484 \pm 1.96\sqrt{0.16484(1 - 0.16484)/(87 + 4)}$, or (0.0886, 0.241).

(b) $X = 13$, $n = 87$, $\tilde{p} = (13 + 2)/(87 + 4) = 0.16484$, $z_{.05} = 1.645$.
The confidence interval is $0.16484 \pm 1.645\sqrt{0.16484(1 - 0.16484)/(87 + 4)}$, or (0.101, 0.229).

(c) Let n be the required sample size.
Then n satisfies the equation $0.04 = 1.96\sqrt{\tilde{p}(1 - \tilde{p})/(n + 4)}$.
Replacing $\tilde{p}$ with 0.16484 and solving for n yields $n = 327$.

(d) Let n be the required sample size.
Then n satisfies the equation $0.04 = 1.645\sqrt{\tilde{p}(1 - \tilde{p})/(n + 4)}$.
Replacing $\tilde{p}$ with 0.16484 and solving for n yields $n = 229$.

5. (a) $X = 859$, $n = 10501$, $\tilde{p} = (859 + 2)/(10501 + 4) = 0.081961$, $z_{.025} = 1.96$.
The confidence interval is $0.081961 \pm 1.96\sqrt{0.081961(1 - 0.081961)/(10501 + 4)}$, or (0.0767, 0.0872).

(b) $X = 859$, $n = 10501$, $\tilde{p} = (859 + 2)/(10501 + 4) = 0.081961$, $z_{.005} = 2.58$.
The confidence interval is $0.081961 \pm 2.58\sqrt{0.081961(1 - 0.081961)/(10501 + 4)}$, or (0.0751, 0.0889).

(c) The upper confidence bound 0.085 satisfies the equation $0.085 = 0.081961 + z_\alpha\sqrt{0.081961(1 - 0.081961)/(10501 + 4)}$
Solving for z_α yields $z_\alpha = 1.14$. The area to the left of $z = 1.14$ is $1 - \alpha = 0.8729$.

The level is 0.8729, or 87.29%.

7. $X = 73$, $n = 100$, $\tilde{p} = (73+2)/(100+4) = 0.72115$, $z_{.02} = 2.05$.
The upper confidence bound is $0.72115 + 2.05\sqrt{0.72115(1-0.72115)/(100+4)}$, or 0.811.

9. (a) $X = 26$, $n = 42$, $\tilde{p} = (26+2)/(42+4) = 0.60870$, $z_{.05} = 1.645$.
The confidence interval is $0.60870 \pm 1.645\sqrt{0.60870(1-0.60870)/(42+4)}$, or (0.490, 0.727).

(b) $X = 41$, $n = 42$, $\tilde{p} = (41+2)/(42+4) = 0.93478$, $z_{.025} = 1.96$.
The expression for the confidence interval yields $0.93478 \pm 1.96\sqrt{0.93478(1-0.93478)/(42+4)}$, or (0.863, 1.006).
Since the upper limit is greater than 1, replace it with 1.
The confidence interval is (0.863, 1).

(c) $X = 32$, $n = 42$, $\tilde{p} = (32+2)/(42+4) = 0.73913$, $z_{.005} = 2.58$.
The confidence interval is $0.73913 \pm 2.58\sqrt{0.73913(1-0.73913)/(42+4)}$, or (0.572, 0.906).

11. (a) Let n be the required sample size.
Then n satisfies the equation $0.05 = 1.96\sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$.
Since there are no preliminary data, we replace $\tilde{p}$ with 0.5. Solving for n yields $n = 381$.

(b) $X = 20$, $n = 100$, $\tilde{p} = (20+2)/(100+4) = 0.21154$, $z_{.025} = 1.96$.
The confidence interval is $0.21154 \pm 1.96\sqrt{0.21154(1-0.21154)/(100+4)}$, or (0.133, 0.290).

(c) Let n be the required sample size.
Then n satisfies the equation $0.05 = 1.96\sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$.
Replacing $\tilde{p}$ with 0.21154 and solving for n yields $n = 253$.

13. (a) $X = 61$, $n = 189$, $\tilde{p} = (61+2)/(189+4) = 0.32642$, $z_{.05} = 1.645$.
The confidence interval is $0.32642 \pm 1.645\sqrt{0.32642(1-0.32642)/(189+4)}$, or (0.271, 0.382).

(b) Let n be the required sample size. Then n satisfies the equation $0.03 = 1.645\sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$.
Replacing $\tilde{p}$ with 0.32642 and solving for n yields $n = 658$.

(c) Let n be the required sample size. Then n satisfies the equation $0.05 = 1.645\sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$.
Since there is no preliminary estimate of $\tilde{p}$ available, replace $\tilde{p}$ with 0.5.
The equation becomes $0.03 = 1.645\sqrt{0.5(1-0.5)/(n+4)}$. Solving for n yields $n = 748$.

Section 5.4

1. (a) 1.860

   (b) 2.776

   (c) 2.763

   (d) 12.706

3. (a) 95%

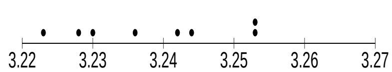
   (b) 98%

   (c) 90%

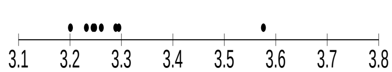
   (d) 99%

   (e) 99.9%

5. $\overline{X} = 13.040$, $s = 1.0091$, $n = 10$, $t_{10-1,.025} = 2.262$.
   The confidence interval is $13.040 \pm 2.262(1.0091/\sqrt{10})$, or (12.318, 13.762).

7. (a) 

   (b) Yes it is appropriate, since there are no outliers.
   $\overline{X} = 3.2386$, $s = 0.011288$, $n = 8$, $t_{8-1,.005} = 3.499$.
   The confidence interval is $3.2386 \pm 2.499(0.011288/\sqrt{8})$, or (3.225, 3.253).

   (c) 

(d) No, the value 3.576 is an outlier.

9. $\overline{X} = 5.900$, $s = 0.56921$, $n = 6$, $t_{6-1,.025} = 2.571$.
The confidence interval is $5.9 \pm 2.571(0.56921/\sqrt{6})$, or (5.303, 6.497).

11. Yes it is appropriate, since there are no outliers.
$\overline{X} = 205.1267$, $s = 1.7174$, $n = 9$, $t_{9-1,.025} = 2.306$.
The confidence interval is $205.1267 \pm 2.306(1.7174/\sqrt{9})$, or (203.81, 206.45).

13. $\overline{X} = 1.250$, $s = 0.6245$, $n = 4$, $t_{4-1,.05} = 2.353$.
The confidence interval is $1.250 \pm 2.353(0.6245/\sqrt{4})$, or (0.515, 1.985).

15. (a) SE Mean is StDev/$\sqrt{\text{N}}$, so $0.52640 = \text{StDev}/\sqrt{20}$, so StDev = 2.3541.

(b) $\overline{X} = 2.39374$, $s = 2.3541$, $n = 20$, $t_{20-1,.005} = 2.861$.
The lower limit of the 99% confidence interval is $2.39374 - 2.861(2.3541/\sqrt{20}) = 0.888$.
Alternatively, one may compute $2.39374 - 2.861(0.52640)$.

(c) $\overline{X} = 2.39374$, $s = 2.3541$, $n = 20$, $t_{20-1,.005} = 2.861$.
The upper limit of the 99% confidence interval is $2.39374 + 2.861(2.3541/\sqrt{20}) = 3.900$.
Alternatively, one may compute $2.39374 + 2.861(0.52640)$.

17. (a) $\overline{X} = 21.7$, $s = 9.4$, $n = 5$, $t_{5-1,.025} = 2.776$.
The confidence interval is $21.7 \pm 2.776(9.4/\sqrt{5})$, or (10.030, 33.370).

(b) No. The minimum possible value is 0, which is less than two sample standard deviations below the sample mean. Therefore it is impossible to observe a value that is two or more sample standard deviations below the sample mean. This suggests that the sample may not come from a normal population.

Section 5.5

1. (a) $\overline{X} = 101.4$, $s = 2.3$, $n = 25$, $t_{25-1,.025} = 2.064$.
The prediction interval is $101.4 \pm 2.064(2.3\sqrt{1 + 1/25})$, or (96.559, 106.241).

(b) $\overline{X} = 101.4$, $s = 2.3$, $n = 25$, $k_{25,.05,.10} = 2.2083$
The tolerance interval is $101.4 \pm 2.2083(2.3)$, or $(96.321, 106.479)$.

3. (a) $\overline{X} = 5.9$, $s = 0.56921$, $n = 6$, $t_{6-1,.01} = 3.365$.
The prediction interval is $5.9 \pm 3.365(0.56921\sqrt{1 + 1/6})$, or (3.8311, 7.9689).

(b) $\overline{X} = 5.9$, $s = 0.56921$, $n = 6$, $k_{6,.05,.05} = 4.4140$
The tolerance interval is $5.9 \pm 4.4140(0.56921)$, or $(3.3875, 8.4125)$.

5. (a) $\overline{X} = 86.56$, $s = 1.02127$, $n = 5$, $t_{5-1,.025} = 2.776$.
The prediction interval is $86.56 \pm 2.776(1.02127\sqrt{1 + 1/5})$, or (83.454, 89.666).

(b) $\overline{X} = 86.56$, $s = 1.02127$, $n = 5$, $k_{5,.01,.10} = 6.6118$
The tolerance interval is $86.56 \pm 6.6118(1.02127)$, or $(79.808, 93.312)$.

Supplementary Exercises for Chapter 5

1. (a) $\widehat{p} = 37/50 = 0.74$

(b) $X = 37$, $n = 50$, $\tilde{p} = (37 + 2)/(50 + 4) = 0.72222$, $z_{.025} = 1.96$.
The confidence interval is $0.72222 \pm 1.96\sqrt{0.72222(1 - 0.72222)/(50 + 4)}$, or (0.603, 0.842).

(c) Let n be the required sample size.
Then n satisfies the equation $0.10 = 1.96\sqrt{\tilde{p}(1 - \tilde{p})/(n + 4)}$.
Replacing $\tilde{p}$ with 0.72222 and solving for n yields $n = 74$.

(d) $X = 37$, $n = 50$, $\tilde{p} = (37 + 2)/(50 + 4) = 0.72222$, $z_{.005} = 2.58$.
The confidence interval is $0.72222 \pm 2.58\sqrt{0.72222(1 - 0.72222)/(50 + 4)}$, or (0.565, 0.879).

(e) Let n be the required sample size.
Then n satisfies the equation $0.10 = 2.58\sqrt{\tilde{p}(1 - \tilde{p})/(n + 4)}$.
Replacing $\tilde{p}$ with 0.72222 and solving for n yields $n = 130$.

3. The higher the level, the wider the confidence interval. Therefore the narrowest interval, (4.20, 5.83), is the 90% confidence interval, the widest interval, (3.57, 6.46), is the 99% confidence interval, and (4.01, 6.02) is the 95% confidence interval.

5. Let n be the required sample size. Then n satisfies the equation $0.04 = 2.33\sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$.

   Since there is no preliminary estimate of $\tilde{p}$ available, replace $\tilde{p}$ with 0.5.

   The equation becomes $0.04 = 2.33\sqrt{0.5(1-0.5)/(n+4)}$. Solving for n yields $n = 845$.

7. Let n be the required sample size.

   The 90% confidence interval based on 144 observations has width ± 0.35.

   Therefore $0.35 = 1.645\sigma/\sqrt{144}$, so $1.645\sigma = 4.2$.

   Now n satisfies the equation $0.2 = 1.645\sigma/\sqrt{n} = 4.2/\sqrt{n}$. Solving for n yields $n = 441$.

9. (a) $X = 30$, $n = 400$, $\tilde{p} = (30+2)/(400+4) = 0.079208$, $z_{.025} = 1.96$.

   The confidence interval is $0.079208 \pm 1.96\sqrt{0.079208(1-0.079208)/(400+4)}$, or (0.0529, 0.1055).

   (b) Let n be the required sample size.

   Then n satisfies the equation $0.02 = 1.96\sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$.

   Replacing $\tilde{p}$ with 0.079208 and solving for n yields $n = 697$.

   (c) Let X be the number of defective components in a lot of 200.

   Let p be the population proportion of components that are defective. Then $X \sim \text{Bin}(200, p)$, so X is approximately normally distributed with mean $\mu_X = 200p$ and $\sigma_X = \sqrt{200p(1-p)}$.

   Let r represent the proportion of lots that are returned.

   Using the continuity correction, $r = P(X > 20.5)$.

   To find a 95% confidence interval for r, express the z-score of $P(X > 20.5)$ as a function of p and substitute the upper and lower confidence limits for p.

   The z-score of 20.5 is $(20.5 - 200p)/\sqrt{200p(1-p)}$. Now find a 95% confidence interval for z by substituting the upper and lower confidence limits for p.

   From part (a), the 95% confidence interval for p is (0.052873, 0.10554). Extra precision is used for this confidence interval to get good precision in the final answer.

   Substituting 0.052873 for p yields $z = 3.14$. Substituting 0.10554 for p yields $z = -0.14$.

   Since we are 95% confident that $0.052873 < p < 0.10554$, we are 95% confident that $-0.14 < z < 3.14$.

   The area to the right of $z = -0.14$ is $1 - 0.4443 = 0.5557$. The area to the right of $z = 3.14$ is $1 - 0.9992 = 0.0008$.

   Therefore we are 95% confident that $0.0008 < r < 0.5557$.

   The confidence interval is (0.0008, 0.5557).

11. (a) False. This a specific confidence interval that has already been computed. The notion of probability does not apply.

    (b) False. The confidence interval specifies the location of the population mean. It does not specify the location of a sample mean.

    (c) True. This says that the method used to compute a 95% confidence interval succeeds in covering the true mean 95% of the time.

    (d) False. The confidence interval specifies the location of the population mean. It does not specify the location of a future measurement.

13. With a sample size of 70, the standard deviation of $\overline{X}$ is $\sigma/\sqrt{70}$. To make the interval half as wide, the standard deviation of $\overline{X}$ will have to be $\sigma/(2\sqrt{70}) = \sigma/\sqrt{280}$. The sample size needs to be 280.

15. The sample mean $\overline{X}$ is the midpoint of the interval, so $\overline{X} = 0.227$. The upper confidence bound 0.241 satisfies $0.241 = 0.227 + 1.96(s/\sqrt{n})$.

    Therefore $s/\sqrt{n} = 0.00714286$. A 90% confidence interval is $0.227 \pm 1.645(s/\sqrt{n}) = 0.227 \pm 1.645(0.00714286)$, or (0.21525, 0.23875).

17. (a) False. The confidence interval is for the population mean, not the sample mean. The sample mean is known, so there is no need to construct a confidence interval for it.

    (b) True. This results from the expression $\overline{X} \pm 1.96(s/\sqrt{n})$, which is a 95% confidence interval for the population mean.

    (c) False. The standard deviation of the mean involves the square root of the sample size, not of the population size.

19. (a) $\overline{X} = 37$, and the uncertainty is $\sigma_{\overline{X}} = s/\sqrt{n} = 0.1$. A 95% confidence interval is $37 \pm 1.96(0.1)$, or (36.804, 37.196).

    (b) Since $s/\sqrt{n} = 0.1$, this confidence interval is of the form $\overline{X} \pm 1(s/\sqrt{n})$. The area to the left of $z = 1$ is approximately 0.1587. Therefore $\alpha/2 = 0.1587$, so the level is $1 - \alpha = 0.6826$, or approximately 68%.

    (c) The measurements come from a normal population.

(d) $t_{9,.025} = 2.262$. A 95% confidence interval is therefore $37 \pm 2.262(s/\sqrt{n}) = 37 \pm 2.262(0.1)$, or (36.774, 37.226).

21. (a) Since X is normally distributed with mean $n\lambda$, it follows that for a proportion $1-\alpha$ of all possible samples, $-z_{\alpha/2}\sigma_X < X - n\lambda < z_{\alpha/2}\sigma_X$.

    Multiplying by -1 and adding X across the inequality yields $X - z_{\alpha/2}\sigma_X < n\lambda < X + z_{\alpha/2}\sigma_X$, which is the desired result.

(b) Since n is a constant, $\sigma_{X/n} = \sigma_X/n = \sqrt{n\lambda}/n = \sqrt{\lambda/n}$.

Therefore $\sigma_{\widehat{\lambda}} = \sigma_X/n$.

(c) Divide the inequality in part (a) by n.

(d) Substitute $\sqrt{\widehat{\lambda}/n}$ for $\sigma_{\widehat{\lambda}}$ in part (c) to show that for a proportion $1-\alpha$ of all possible samples, $\widehat{\lambda} - z_{\alpha/2}\sqrt{\widehat{\lambda}/n} < \lambda < \widehat{\lambda} + z_{\alpha/2}\sqrt{\widehat{\lambda}/n}$.

The interval $\widehat{\lambda} \pm z_{\alpha/2}\sqrt{\widehat{\lambda}/n}$ is therefore a level $1-\alpha$ confidence interval for λ.

(e) $n = 5$, $\widehat{\lambda} = 300/5 = 60$, $\sigma_{\widehat{\lambda}} = \sqrt{60/5} = 3.4641$. A 95% confidence interval for λ is therefore $60 \pm 1.96(3.4641)$, or (53.210, 66.790).

Chapter 6

Section 6.1

1. (a) $\overline{X} = 4.5$, $s = 2.7$, $n = 80$. The null and alternate hypotheses are $H_0\colon \mu \geq 5.4$ versus $H_1\colon \mu < 5.4$.
   $z = (4.5 - 5.4)/(2.7/\sqrt{80}) = -2.98$. Since the alternate hypothesis is of the form $\mu < \mu_0$, the P-value is the area to the left of $z = -2.98$.
   Thus $P = 0.0014$.

   (b) The P-value is 0.0014, so if H_0 is true then the sample is in the most extreme 0.14% of its distribution.

3. (a) $\overline{X} = 1.90$, $s = 21.20$, $n = 160$. The null and alternate hypotheses are $H_0\colon \mu = 0$ versus $H_1\colon \mu \neq 0$.
   $z = (1.90 - 0)/(21.20/\sqrt{160}) = 1.13$. Since the alternate hypothesis is of the form $\mu \neq \mu_0$, the P-value is the sum of the areas to the right of $z = 1.13$ and to the left of $z = -1.13$.
   Thus $P = 0.1292 + 0.1292 = 0.2584$.

   (b) The P-value is 0.2584, so if H_0 is true then the sample is in the most extreme 25.84% of its distribution.

5. (a) $\overline{X} = 5.4$, $s = 1.2$, $n = 85$. The null and alternate hypotheses are $H_0\colon \mu \leq 5$ versus $H_1\colon \mu > 5$.
   $z = (5.4 - 5)/(1.2/\sqrt{85}) = 3.07$. Since the alternate hypothesis is of the form $\mu > \mu_0$, the P-value is the area to the right of $z = 3.07$.
   Thus $P = 0.0011$.

   (b) If the mean silicon content 5 mg/L or less, the probability of observing a sample mean as large as the value of 5.4 that was actually observed would be 0.0011. Therefore we are convinced that the mean daily output is not 5 mg/L or less, but is instead greater than 5 mg/L.

7. (a) $\overline{X} = 715$, $s = 24$, $n = 60$. The null and alternate hypotheses are $H_0\colon \mu \geq 740$ versus $H_1\colon \mu < 740$.
   $z = (715 - 740)/(24/\sqrt{60}) = -8.07$. Since the alternate hypothesis is of the form $\mu < \mu_0$, the P-value is the area to the left of $z = -8.07$.
   Thus $P \approx 0$.

   (b) If the mean daily output were 740 tons or more, the probability of observing a sample mean as small as the value of 715 that was actually observed would be nearly 0. Therefore we are convinced that the mean daily output is not 740 tons or more, but is instead less than 740 tons.

9. (a) $\overline{X} = 763$, $s = 120$, $n = 67$. The null and alternate hypotheses are $H_0\colon \mu \leq 750$ versus $H_1\colon \mu > 750$.
   $z = (763 - 750)/(120/\sqrt{67}) = 0.89$. Since the alternate hypothesis is of the form $\mu > \mu_0$, the P-value is the area to the right of $z = 0.89$.

   Thus $P = 0.1867$.

   (b) If the mean number of kilocycles to failure were 750, the probability of observing a sample mean as far from 750 as the value of 763 that was actually observed would be 0.1867. Since 0.1867 is not a small probability, it is plausible that the mean is 750.

11. (ii) 4. The null distribution specifies that the population mean, which is also the mean of $\overline{X}$, is the value on the boundary between the null and alternate hypotheses.

13. $\overline{X} = 11.98$ and $\sigma_{\overline{X}} = \sigma/\sqrt{n} = 0.02$. The null and alternate hypotheses are $H_0\colon \mu = 12.0$ versus $H_1\colon \mu \neq 12.0$. $z = (11.98 - 12.0)/0.02 = -1.00$. Since the alternate hypothesis is of the form $\mu = \mu_0$, the P-value is is the sum of the areas to the right of $z = -1.00$ and to the left of $z = 1.00$.

    Thus $P = 0.1587 + 0.1587 = 0.3174$.

15. (a) SE Mean $= s/\sqrt{n} = 2.00819/\sqrt{87} = 0.2153$.

    (b) $\overline{X} = 4.07114$. From part (a), $s/\sqrt{n} = 0.2153$. The null and alternate hypotheses are $H_0\colon \mu \leq 3.5$ versus
    $H_1\colon \mu > 3.5$.
    $z = (4.07114 - 3.5)/0.2153 = 2.65$.

    (c) Since the alternate hypothesis is of the form $\mu > \mu_0$, the P-value is the area to the right of $z = 2.65$.

    Thus $P = 0.0040$.

Section 6.2

1. $P = 0.10$. The larger the P-value, the more plausible the null hypothesis.

3. (iv). A P-value of 0.01 means that if H_0 is true, then the observed value of the test statistic was in the most extreme 1% of its distribution. This is unlikely, but not impossible.

5. (a) True. The result is statistically significant at any level greater than or equal to 5%.

   (b) True. The result is statistically significant at any level greater than or equal to 5%.

   (c) False. The result is not statistically significant at any level less than 5%.

7. (a) No. The P-value is 0.177. Since this is greater than 0.05, H_0 is not rejected at the 5% level.

   (b) The value 36 is contained in the 95% confidence interval for μ. Therefore the hypothesis $H_0\colon \mu = 36$ versus $H_1\colon \mu \neq 36$ cannot be rejected at the 5% level.

9. (a) $H_0\colon \mu \leq 10$. If H_0 is rejected we conclude that $\mu > 10$, and that the new type of epoxy should be used.

   (b) $H_0\colon \mu = 20$. If H_0 is rejected we conclude that $\mu \neq 20$, and that the flowmeter should be recalibrated.

   (c) $H_0\colon \mu \leq 8$. If H_0 is rejected we conclude that $\mu > 8$, and that the new type of battery should be used.

11. (a) (ii) The scale is out of calibration. If H_0 is rejected, we conclude that H_0 is false, so $\mu \neq 10$.

    (b) (iii) The scale might be in calibration. If H_0 is not rejected, we conclude that H_0 is plausible, so μ might be equal to 10.

    (c) No. The scale is in calibration only if $\mu = 10$. The strongest evidence in favor of this hypothesis would occur if $\overline{X} = 10$. But since there is uncertainty in $\overline{X}$, we cannot be sure even then that $\mu = 10$.

13. No, she cannot conclude that the null hypothesis is true, only that it is plausible.

15. (i) $H_0\colon \mu = 1.2$. For either of the other two hypotheses, the P-value would be 0.025.

17. (a) Yes. The value 3.5 is greater than the upper confidence bound of 3.45. Quantities greater than the upper confidence bound will have P-values less than 0.05. Therefore $P < 0.05$.

    (b) No, we would need to know the 99% upper confidence bound to determine whether $P < 0.01$.

19. Yes, we can compute the P-value exactly. Since the 95% upper confidence bound is 3.45, we know that $3.40 + 1.645s/\sqrt{n} = 3.45$. Therefore $s/\sqrt{n} = 0.0304$. The z-score is $(3.40 - 3.50)/0.0304 = -3.29$. The P-value is 0.0005, which is less than 0.01.

Section 6.3

1. $X = 130$, $n = 1600$, $\widehat{p} = 130/1600 = 0.08125$.

   The null and alternate hypotheses are $H_0\colon p \geq 0.10$ versus $H_1\colon p < 0.10$.

   $z = (0.08125 - 0.10)/\sqrt{0.10(1 - 0.10)/1600} = -2.50$.

   Since the alternate hypothesis is of the form $p < p_0$, the P-value is the area to the left of $z = -2.50$, so $P = 0.0062$. There is sufficient evidence to reject the claim.

3. $X = 29$, $n = 50$, $\widehat{p} = 29/50 = 0.58$.

   The null and alternate hypotheses are $H_0\colon p \leq 0.50$ versus $H_1\colon p > 0.50$.

   $z = (0.58 - 0.50)/\sqrt{0.50(1 - 0.50)/50} = 1.13$.

   Since the alternate hypothesis is of the form $p > p_0$, the P-value is the area to the right of $z = 1.13$, so $P = 0.1292$. We cannot conclude that more than half of bathroom scales underestimate weight.

5. $X = 274$, $n = 500$, $\widehat{p} = 274/500 = 0.548$.

   The null and alternate hypotheses are $H_0\colon p \leq 0.50$ versus $H_1\colon p > 0.50$.

   $z = (0.548 - 0.50)/\sqrt{0.50(1 - 0.50)/500} = 2.15$.

   Since the alternate hypothesis is of the form $p > p_0$, the P-value is the area to the right of $z = 2.15$, so $P = 0.0158$. We can conclude that more than half of residents are opposed to building a new shopping mall.

7. $X = 110$, $n = 150$, $\widehat{p} = 110/150 = 0.733$.

   The null and alternate hypotheses are $H_0\colon p \leq 0.70$ versus $H_1\colon p > 0.70$.

   $z = (0.733 - 0.70)/\sqrt{0.70(1 - 0.70)/150} = 0.89$.

   Since the alternate hypothesis is of the form $p > p_0$, the P-value is the area to the right of $z = 0.89$, so $P = 0.1867$. We cannot conclude that more than 70% of the households have high-speed Internet access.

9. $X = 470$, $n = 500$, $\widehat{p} = 470/500 = 0.94$.

   The null and alternate hypotheses are $H_0\colon p \geq 0.95$ versus $H_1\colon p < 0.95$.

   $z = (0.94 - 0.95)/\sqrt{0.95(1 - 0.95)/500} = -1.03$.

   Since the alternate hypothesis is of the form $p < p_0$, the P-value is the area to the left of $z = -1.03$, so $P = 0.1515$. The claim cannot be rejected.

11. $X = 73$, $n = 100$, $\widehat{p} = 73/100 = 0.73$.

    The null and alternate hypotheses are $H_0\colon p \leq 0.60$ versus $H_1\colon p > 0.60$.

$z = (0.73 - 0.60)/\sqrt{0.60(1 - 0.60)/100} = 2.65$.

Since the alternate hypothesis is of the form $p > p_0$, the P-value is the area to the right of $z = 2.65$, so $P = 0.0040$. We can conclude that more than 60% of the residences have reduced their water consumption.

13. (a) Sample p $= \widehat{p} = 345/500 = 0.690$.

(b) The null and alternate hypotheses are $H_0\colon p \geq 0.7$ versus $H_1\colon \mu < 0.7$. $n = 500$. From part (a), $\widehat{p} = 0.690$. $z = (0.690 - 0.700)/\sqrt{0.7(1 - 0.7)/500} = -0.49$.

(c) Since the alternate hypothesis is of the form $p < p_0$, the P-value is the area to the left of $z = -0.49$. Thus $P = 0.3121$.

Section 6.4

1. (a) $\overline{X} = 60.01$, $s = 0.026458$, $n = 3$. There are $3 - 1 = 2$ degrees of freedom.

The null and alternate hypotheses are $H_0\colon \mu = 60$ versus $H_1\colon \mu \neq 60$.

$t = (60.01 - 60)/(0.026458/\sqrt{3}) = 0.655$.

Since the alternate hypothesis is of the form $\mu \neq \mu_0$, the P-value is the sum of the areas to the right of $t = 0.655$ and to the left of $t = -0.655$.

From the t table, $0.50 < P < 0.80$. A computer package gives $P = 0.580$.

We cannot conclude that the machine is out of calibration.

(b) The t-test cannot be performed, because the sample standard deviation cannot be computed from a sample of size 1.

3. (a) $H_0\colon \mu \geq 16$ versus $H_1\colon \mu < 16$

(b) $\overline{X} = 15.887$, $s = 0.13047$, $n = 10$. There are $10 - 1 = 9$ degrees of freedom.

The null and alternate hypotheses are $H_0\colon \mu \geq 16$ versus $H_1\colon \mu < 16$.

$t = (15.887 - 16)/(0.13047/\sqrt{10}) = -2.739$.

(c) Since the alternate hypothesis is of the form $\mu < \mu_0$, the P-value is the area to the left of $t = -2.739$.

From the t table, $0.01 < P < 0.025$. A computer package gives $P = 0.011$.

We conclude that the mean fill weight is less than 16 oz.

5. (a) $\overline{X} = 22.571$, $s = 5.28700$, $n = 7$. There are $7 - 1 = 6$ degrees of freedom.

The null and alternate hypotheses are $H_0\colon \mu \leq 20$ versus $H_1\colon \mu > 20$.

$t = (22.571 - 20)/(5.28700/\sqrt{7}) = 1.287$.

Since the alternate hypothesis is of the form $\mu > \mu_0$, the P-value is the area to the right of $t = 1.287$

From the t table, $0.10 < P < 0.25$. A computer package gives $P = 0.123$.

We cannot conclude that the mean amount of solids is greater than 20 g.

(b) $\overline{X} = 22.571$, $s = 5.28700$, $n = 7$. There are $7 - 1 = 6$ degrees of freedom.

The null and alternate hypotheses are $H_0\colon \mu \geq 30$ versus $H_1\colon \mu < 30$.

$t = (22.571 - 30)/(5.28700/\sqrt{7}) = -3.717$.

Since the alternate hypothesis is of the form $\mu > \mu_0$, the P-value is the area to the right of $t = -3.717$

From the t table, $0.001 < P < 0.005$. A computer package gives $P = 0.00494$.

We can conclude that the mean amount of solids is less than 30 g.

(c) $\overline{X} = 22.571$, $s = 5.28700$, $n = 7$. There are $7 - 1 = 6$ degrees of freedom.

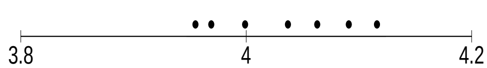
The null and alternate hypotheses are $H_0\colon \mu = 25$ versus $H_1\colon \mu \neq 25$.

$t = (22.571 - 25)/(5.28700/\sqrt{7}) = -1.215$.

Since the alternate hypothesis is of the form $\mu \neq \mu_0$, the P-value is the sum of the areas to the right of $t = 1.215$ and to the left of $t = -1.215$.

From the t table, $0.20 < P < 0.50$. A computer package gives $P = 0.270$.

We cannot conclude that the mean amount of solids differs from 25 g.

7. (a) 

(b) Yes, the sample contains no outliers.

$\overline{X} = 4.032857$, $s = 0.061244$, $n = 7$. There are $7 - 1 = 6$ degrees of freedom.

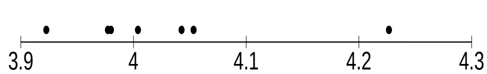
The null and alternate hypotheses are $H_0\colon \mu = 4$ versus $H_1\colon \mu \neq 4$.

$t = (4.032857 - 4)/(0.061244/\sqrt{7}) = 1.419$.

Since the alternate hypothesis is of the form $\mu \neq \mu_0$, the P-value is the sum of the areas to the right of $t = 1.419$ and to the left of $t = -1.419$.

From the t table, $0.20 < P < 0.50$. A computer package gives $P = 0.2056$.

It cannot be concluded that the mean thickness differs from 4 mils.

(c) 

(d) No, the sample contains an outlier.

9. (a) $\overline{X} = 45.2$, $s = 11.3$, $n = 16$. There are $16 - 1 = 15$ degrees of freedom.

   The null and alternate hypotheses are $H_0\colon \mu \leq 35$ versus $H_1\colon \mu > 35$.

   $t = (45.2 - 35)/(11.3/\sqrt{16}) = 3.611$.

   Since the alternate hypothesis is of the form $\mu > \mu_0$, the P-value is the area to the right of $t = 3.611$

   From the t table, $0.001 < P < 0.005$. A computer package gives $P = 0.00128$.

   We can conclude that the mean conversion is greater than 35.

   (b) $\overline{X} = 45.2$, $s = 11.3$, $n = 16$. There are $16 - 1 = 15$ degrees of freedom.

   The null and alternate hypotheses are $H_0\colon \mu = 50$ versus $H_1\colon \mu \neq 50$.

   $t = (45.2 - 50)/(11.3/\sqrt{16}) = -1.699$.

   Since the alternate hypothesis is of the form $\mu \neq \mu_0$, the P-value is the sum of the areas to the right of $t = 1.699$ and to the left of $t = -1.699$.

   From the t table, $0.10 < P < 0.20$. A computer package gives $P = 0.1099$.

   We cannot conclude that the mean conversion differs from 50.

11. $\overline{X} = 1.25$, $s = 0.624500$, $n = 4$. There are $4 - 1 = 3$ degrees of freedom.

   The null and alternate hypotheses are $H_0\colon \mu \geq 2.5$ versus $H_1\colon \mu < 2.5$.

   $t = (1.25 - 2.5)/(0.624500/\sqrt{4}) = -4.003$.

   Since the alternate hypothesis is of the form $\mu < \mu_0$, the P-value is the area to the left of $t = -4.003$.

   From the t table, $0.01 < P < 0.025$. A computer package gives $P = 0.014$.

   We can conclude that the mean amount absorbed is less than 2.5%.

13. (a) StDev = (SE Mean)$\sqrt{\mathrm{N}} = 1.8389\sqrt{11} = 6.0989$.

   (b) $t_{10,.025} = 2.228$. The lower 95% confidence bound is $13.2874 - 2.228(1.8389) = 9.190$.

   (c) $t_{10,.025} = 2.228$. The upper 95% confidence bound is $13.2874 + 2.228(1.8389) = 17.384$.

   (d) $t = (13.2874 - 16)/1.8389 = -1.475$.

Section 6.5

1. (a) Let p_1 represent the probability that a randomly chosen fastener is conforming, let p_2 represent the probability that it is downgraded, and let p_3 represent the probability that it is scrap. Then the null hypothesis is $H_0\colon p_1 = 0.85,\ p_2 = 0.10,\ p_3 = 0.05$

(b) The total number of observation is $n = 500$. The expected values are np_1, np_2 and np_3, or 425, 50, and 25.

(c) The observed values are 405, 55, and 40.
$\chi^2 = (405 - 450)^2/450 + (55 - 50)^2/50 + (40 - 25)^2/25 = 10.4412$.

(d) There are $3-1 = 2$ degrees of freedom. From the χ^2 table, $0.005 < P < 0.01$. A computer package gives $P = 0.00540$.

The true percentages differ from 85%, 10%, and 5%.

3. The row totals are $O_{1.} = 173$ and $O_{2.} = 210$. The column totals are $O_{.1} = 181$, $O_{.2} = 99$, $O_{.3} = 31$, $O_{.4} = 11$, $O_{.5} = 61$. The grand total is $O_{..} = 383$.

The expected values are $E_{ij} = O_{i.}O_{.j}/O_{..}$, as shown in the following table.

Net Excess Capacity

	$< 0\%$	$0 - 10\%$	$11 - 20\%$	$21 - 30\%$	$> 30\%$
Small	81.7572	44.7180	14.0026	4.9687	27.5535
Large	99.2428	54.2820	16.9974	6.0313	33.4465

There are $(2 - 1)(5 - 1) = 4$ degrees of freedom.

$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{5}(O_{ij} - E_{ij})^2/E_{ij} = 12.945$.

From the χ^2 table, $0.01 < P < 0.05$. A computer package gives $P = 0.012$.

It is reasonable to conclude that the distributions differ.

5. The row totals are $O_{1.} = 41$, $O_{2.} = 39$, and $O_{3.} = 412$. The column totals are $O_{.1} = 89$, $O_{.2} = 163$, $O_{.3} = 240$. The grand total is $O_{..} = 492$.

The expected values are $E_{ij} = O_{i.}O_{.j}/O_{..}$, as shown in the following table.

	< 1	1 to < 5	≥ 5
Diseased	7.4167	13.583	20.000
Sensitized	7.0549	12.921	19.024
Normal	74.528	136.50	200.98

There are $(3 - 1)(3 - 1) = 4$ degrees of freedom.

$\chi^2 = \sum_{i=1}^{3}\sum_{j=1}^{3}(O_{ij} - E_{ij})^2/E_{ij} = 10.829$.

From the χ^2 table, $0.025 < P < 0.05$. A computer package gives $P = 0.0286$.

There is some evidence that the proportions of workers in the various disease categories differ among exposure levels.

7. (a) The row totals are $O_{1.} = 37$, $O_{2.} = 25$, and $O_{3.} = 35$. The column totals are $O_{.1} = 27$, $O_{.2} = 35$, $O_{.3} = 35$. The grand total is $O_{..} = 97$.

The expected values are $E_{ij} = O_{i.}O_{.j}/O_{..}$, as shown in the following table.

10.2990	13.3505	13.3505
6.9588	9.0206	9.0206
9.7423	12.6289	12.6289

(b) The χ^2 test is appropriate here, since all the expected values are greater than 5.

There are $(3-1)(3-1) = 4$ degrees of freedom.

$\chi^2 = \sum_{i=1}^{3} \sum_{j=1}^{3} (O_{ij} - E_{ij})^2/E_{ij} = 6.4808$.

From the χ^2 table, $P > 0.10$. A computer package gives $P = 0.166$.

There is no evidence that the rows and columns are not independent.

9. (iii) Both row totals and column totals in the observed table must be the same as the row and column totals, respectively, in the expected table.

11. Let p_1 represent the probability that a randomly chosen plate is classifed as premium, let p_2 represent the probability that it is conforming, let p_3 represent the probability that it is downgraded, and let p_4 represent the probability that it is unacceptable. Then the null hypothesis is $H_0\colon p_1 = 0.10,\ p_2 = 0.70,\ p_3 = 0.15,\ p_4 = 0.05$

The total number of observation is $n = 200$. The expected values are np_1, np_2, np_3, and np_4, or 20, 140, 30, and 10.

The observed values are 19, 133, 35, and 13.

$\chi^2 = (19-20)^2/20 + (133-140)^2/140 + (35-30)^2/30 + (13-10)^2/10 = 2.133$.

There are $4 - 1 = 3$ degrees of freedom. From the χ^2 table, $P > 0.10$. A computer package gives $P = 0.545$.

We cannot conclude that the engineer's claim is incorrect.

13. The row totals are $O_{1.} = 217$ and $O_{2.} = 210$. The column totals are $O_{.1} = 32$, $O_{.2} = 15$, $O_{.3} = 37$, $O_{.4} = 38$, $O_{.5} = 45$, $O_{.6} = 48$, $O_{.7} = 46$, $O_{.8} = 42$, $O_{.9} = 34$, $O_{.10} = 36$, $O_{.11} = 28$, $O_{.12} = 26$. The grand total is $O_{..} = 427$.

The expected values are $E_{ij} = O_{i.}O_{.j}/O_{..}$, as shown in the following table.

	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
Known	16.26	7.62	18.80	19.31	22.87	24.39	23.38	21.34	17.28	18.30	14.23	13.21
Unknown	15.74	7.38	18.20	18.69	22.13	23.61	22.62	20.66	16.72	17.70	13.77	12.79

There are $(2-1)(12-1) = 11$ degrees of freedom.

$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{12} (O_{ij} - E_{ij})^2 / E_{ij} = 41.33$.

From the χ^2 table, $P < 0.005$. A computer package gives $P = 2.1 \times 10^{-5}$.

We can conclude that the proportion of false alarms whose cause is known differs from month to month.

Section 6.6

1. (a) False. H_0 is rejected at any level greater than or equal to 0.07, but 5% is less than 0.07.

   (b) True. 2% is less than 0.07.

   (c) True. 10% is greater than 0.07.

3. The costly error is to reject H_0 when it is true. This is a type I error. The smaller the level we test at, the smaller the probability of a type I error. Therefore a smaller probability is obtained by testing at the 1% level.

5. (a) Type I error. H_0 is true and was rejected.

   (b) Correct decision. H_0 is false and was rejected.

   (c) Correct decision. H_0 is true and was not rejected.

   (d) Type II error. H_0 is false and was not rejected.

7. (a) The null distribution of $\overline{X}$ is normal with mean $\mu = 100$ and standard deviation $\sigma_{\overline{X}} = 0.1/\sqrt{100} = 0.01$.

   Since the alternate hypothesis is of the form $\mu \neq \mu_0$, the rejection region will consist of both the upper and lower 2.5% of the null distribution.

   The z-scores corresponding to the boundaries of upper and lower 2.5% are $z = 1.96$ and $z = -1.96$, respectively.

   Therefore the boundaries are $100 + 1.96(0.01) = 100.0196$ and $100 - 1.96(0.01) = 99.9804$.

   Reject H_0 if $\overline{X} \geq 100.0196$ or if $\overline{X} \leq 99.9804$.

   (b) The null distribution of $\overline{X}$ is normal with mean $\mu = 100$ and standard deviation $\sigma_{\overline{X}} = 0.1/\sqrt{100} = 0.01$.

   Since the alternate hypothesis is of the form $\mu \neq \mu_0$, the rejection region will consist of both the upper and lower 5% of the null distribution.

The z-scores corresponding to the boundaries of upper and lower 5% are $z = 1.645$ and $z = -1.645$, respectively.

Therefore the boundaries are $100 + 1.645(0.01) = 100.01645$ and $100 - 1.645(0.01) = 99.98355$.

Reject H_0 if $\overline{X} \geq 100.01645$ or if $\overline{X} \leq 99.98355$.

(c) Yes

(d) No

(e) Since this is a two-tailed test, there are two critical points, equidistant from the null mean of 100. Since one critical point is 100.015, the other is 99.985.

The level of the test is the sum $P(\overline{X} \leq 99.985) + P(\overline{X} \geq 100.015)$, computed under the null distribution.

The null distribution is normal with mean $\mu = 100$ and standard deviation $\sigma_{\overline{X}} = 0.01$.

The z-score of 100.015 is $(100.015 - 100)/0.01 = 1.5$. The z-score of 99.985 is $(99.985 - 100)/0.01 = -1.5$.

The level of the test is therefore $0.0668 + 0.0668 = 0.1336$, or 13.36%.

Section 6.7

1. (a) True. This is the definition of power.

(b) True. When H_0 is false, making a correct decision means rejecting H_0.

(c) False. The power is 0.85, not 0.15.

(d) False. H_0 does not have a probability of being true.

3. increase. If the level increases, the probability of rejecting H_0 increases, so in particular, the probability of rejecting H_0 when it is false increases.

5. ii. Since 12 is farther from the null mean of 8 than 10 is, the power against the alternative $\mu = 12$ will be greater than the power against the alternative $\mu = 10$.

7. (a) $H_0 : \mu \geq 50,000$ versus $H_1 : \mu < 50,000$. H_1 is true, since the true value of μ is 49,500.

(b) The level is the probability of rejecting H_0 when it is true.

Under H_0, $\overline{X}$ is approximately normally distributed with mean 50,000 and standard deviation $\sigma_{\overline{X}} = 5000/\sqrt{100} = 500$.

The probability of rejecting H_0 is $P(\overline{X} \leq 49,400)$.

Under H_0, the z-score of 49,400 is $z = (49,400 - 50,000)/500 = -1.20$.
The level of the test is the area under the normal curve to the left of $z = -1.20$.

Therefore the level is 0.1151.

The power is the probability of rejecting H_0 when $\mu = 49,500$.

$\overline{X}$ is approximately normally distributed with mean 49,500 and standard deviation
$\sigma_{\overline{X}} = 5000/\sqrt{100} = 500$.

The probability of rejecting H_0 is $P(\overline{X} \leq 49,400)$.

The z-score of 49,400 is $z = (49,400 - 49,500)/500 = -0.20$.
The power of the test is thus the area under the normal curve to the left of $z = -0.20$.

Therefore the power is 0.4207.

(c) Since the alternate hypothesis is of the form $\mu < \mu_0$, the 5% rejection region will be the region $\overline{X} \leq x_5$, where x_5 is the 5th percentile of the null distribution.

The z-score corresponding to the 5th percentile is $z = -1.645$.

Therefore $x_5 = 50,000 - 1.645(500) = 49,177.5$.

The rejection region is $\overline{X} \leq 49,177.5$.

The power is therefore $P(\overline{X} \leq 49,177.5)$ when $\mu = 49,500$.

The z-score of 49,177.5 is $z = (49,177.5 - 49,500)/500 = -0.645$. We will use $z = -0.65$.

The power is therefore the area to the left of $z = -0.65$.

Thus the power is 0.2578.

(d) For the power to be 0.80, the rejection region must be $\overline{X} \leq x_0$ where $P(\overline{X} \leq x_0) = 0.80$ when $\mu = 49,500$.

Therefore x_0 is the 80th percentile of the normal curve when $\mu = 49,500$. The z-score corresponding to the 80th percentile is $z = 0.84$.

Therefore $x_0 = 49,500 + 0.84(500) = 49,920$.

Now compute the level of the test whose rejection region is $\overline{X} \leq 49,920$.

The level is $P(\overline{X} \leq 49,920)$ when $\mu = 50,000$.

The z-score of 49,920 is $z = (49,920 - 50,000)/500 = -0.16$.

The level is the area under the normal curve to the left of $z = -0.16$.

Therefore the level is 0.4364.

(e) Let n be the required number of tires.

The null distribution is normal with $\mu = 50,000$ and $\sigma_{\overline{X}} = 5000/\sqrt{n}$. The alternate distribution is normal with $\mu = 49,500$ and $\sigma_{\overline{X}} = 5000/\sqrt{n}$.

Let x_0 denote the boundary of the rejection region.

Since the level is 5%, the z-score of x_0 is $z = -1.645$ under the null distribution.
Therefore $x_0 = 50,000 - 1.645(5000/\sqrt{n})$.
Since the power is 0.80, the z-score of x_0 is $z = 0.84$ under the alternate distribution.
Therefore $x_0 = 49,500 + 0.84(5000/\sqrt{n})$.
It follows that $50,000 - 1.645(5000/\sqrt{n}) = 49,500 + 0.84(5000/\sqrt{n})$.
Solving for n yields $n = 618$.

9. (a) Two-tailed. The alternate hypothesis is of the form $p \neq p_0$.

(b) $p = 0.5$

(c) $p = 0.4$

(d) Less than 0.7. The power for a sample size of 150 is 0.691332, and the power for a smaller sample size of 100 would be less than this.

(e) Greater than 0.6. The power for a sample size of 150 is 0.691332, and the power for a larger sample size of 200 would be greater than this.

(f) Greater than 0.65. The power against the alternative $p = 0.4$ is 0.691332, and the alternative $p = 0.3$ is farther from the null than $p = 0.4$. So the power against the alternative $p = 0.3$ is greater than 0.691332.

(g) It's impossible to tell from the output. The power against the alternative $p = 0.45$ will be less than the power against $p = 0.4$, which is 0.691332. But we cannot tell without calculating whether the power will be less than 0.65.

11. (a) Two-tailed. The alternate hypothesis is of the form $\mu_1 - \mu_2 \neq \Delta$.

(b) Less than 0.9. The sample size of 60 is the smallest that will produce power greater than or equal to the target power of 0.9.

(c) Greater than 0.9. The power is greater than 0.9 against a difference of 3, so it will be greater than 0.9 against any difference greater than 3.

Section 6.8

1. (a) There are six tests, so the Bonferroni-adjusted P-values are found by multiplying the original P-values by 6. For the setting whose original P-value is 0.002, the Bonferroni-adjusted P-value is therefore 0.012. Since this value is small, we can conclude that this setting reduces the proportion of defective parts.

(b) The Bonferroni-adjusted P-value is $6(0.03) = 0.18$. Since this value is not so small, we cannot conclude that this setting reduces the proportion of defective parts.

3. The original P-value must be $0.05/20 = 0.0025$.

5. (a) No. Let X represent the number of times in 200 days that H_0 is rejected. If the mean burn-out amperage is equal to 15 A every day, the probability of rejecting H_0 is 0.05 each day, so $X \sim \text{Bin}(200, 0.05)$.

The probability of rejecting H_0 10 or more times in 200 days is then $P(X \geq 10)$, which is approximately equal to 0.5636. So it would not be unusual to reject H_0 10 or more times in 200 trials if H_0 is always true.

Alternatively, note that if the probability of rejecting H_0 is 0.05 each day, the mean number of times that H_0 will be rejected in 200 days is $(200)(0.05) = 10$. Therefore observing 10 rejections in 200 days is consistent with the hypothesis that the mean burn-out amperage is equal to 15 A every day.

(b) Yes. Let X represent the number of times in 200 days that H_0 is rejected. If the mean burn-out amperage is equal to 15 A every day, the probability of rejecting H_0 is 0.05 each day, so $X \sim \text{Bin}(200, 0.05)$.

The probability of rejecting H_0 20 or more times in 200 days is then $P(X \geq 20)$ which is approximately equal to 0.0010.

So it would be quite unusual to reject H_0 20 times in 200 trials if H_0 is always true.

We can conclude that the mean burn-out amperage differed from 15 A on at least some of the days.

Supplementary Exercises for Chapter 6

1. $\overline{X} = 51.2$, $s = 4.0$, $n = 110$. The null and alternate hypotheses are $H_0\colon \mu \leq 50$ versus $H_1\colon \mu > 50$.

$z = (51.2 - 50)/(4.0/\sqrt{110}) = 3.15$. Since the alternate hypothesis is of the form $\mu > \mu_0$, the P-value is the area to the right of $z = 3.15$. Thus $P = 0.0008$.

We can conclude that the mean strength is greater than 50 psi.

3. (a) $H_0\colon \mu \geq 90$ versus $H_1\colon \mu < 90$

(b) Let $\overline{X}$ be the sample mean of the 150 times.

Under H_0, the population mean is $\mu = 90$, and the population standard deviation is $\sigma = 5$.

The null distribution of $\overline{X}$ is therefore normal with mean 90 and standard deviation $5/\sqrt{150} = 0.408248$.

Since the alternate hypothesis is of the form $\mu < \mu_0$, the rejection region for a 5% level test consists of the lower 5% of the null distribution.

The z-score corresponding to the lower 5% of the normal distribution is $z = -1.645$.

Therefore the rejection region consists of all values of $\overline{X}$ less than or equal to $90 - 1.645(0.408248) = 89.3284$.

H_0 will be rejected if $\overline{X} < 89.3284$.

(c) This is not an appropriate rejection region. The rejection region should consist of values for $\overline{X}$ that will make the P-value of the test less than or equal to a chosen threshold level. Therefore the rejection region must be of the form $\overline{X} \leq x_0$. This rejection region is of the form $\overline{X} \geq x_0$, and so it consists of values for which the P-value will be greater than some level.

(d) This is an appropriate rejection region.

Under H_0, the z-score of 89.4 is $(89.4 - 90)/0.408248 = -1.47$.

Since the alternate hypothesis is of the form $\mu < \mu_0$, the level is the area to the left of $z = -1.47$.

Therefore the level is $\alpha = 0.0708$.

(e) This is not an appropriate rejection region. The rejection region should consist of values for $\overline{X}$ that will make the P-value of the test less than a chosen threshold level. This rejection region contains values of $\overline{X}$ greater than 90.6, for which the P-value will be large.

5. (a) The null hypothesis specifies a single value for the mean: $\mu = 3$. The level, which is 5%, is therefore the probability that the null hypothesis will be rejected when $\mu = 3$. The machine is shut down if H_0 is rejected at the 5% level. Therefore the probability that the machine will be shut down when $\mu = 3$ is 0.05.

(b) First find the rejection region.

The null distribution of $\overline{X}$ is normal with mean $\mu = 3$ and standard deviation $\sigma_{\overline{X}} = 0.10/\sqrt{50} = 0.014142$.

Since the alternate hypothesis is of the form $\mu \neq \mu_0$, the rejection region will consist of both the upper and lower 2.5% of the null distribution.

The z-scores corresponding to the boundaries of upper and lower 2.5% are $z = 1.96$ and $z = -1.96$, respectively.

Therefore the boundaries are $3 + 1.96(0.014142) = 3.0277$ and $3 - 1.96(0.014142) = 2.9723$.

H_0 will be rejected if $\overline{X} \geq 3.0277$ or if $\overline{X} \leq 2.9723$.

The probability that the equipment will be recalibrated is therefore equal to $P(\overline{X} \geq 3.0277) + P(\overline{X} \leq 2.9723)$, computed under the assumption that $\mu = 3.01$.

The z-score of 3.0277 is $(3.0277 - 3.01)/0.014142 = 1.25$.

The z-score of 2.9723 is $(2.9723 - 3.01)/0.014142 = -2.67$.

Therefore $P(\overline{X} \geq 3.0277) = 0.1056$, and $P(\overline{X} \leq 2.9723) = 0.0038$.

The probability that the equipment will be recalibrated is equal to $0.1056 + 0.0038 = 0.1094$.

7. $X = 37$, $n = 37 + 458 = 495$, $\widehat{p} = 37/495 = 0.074747$.

   The null and alternate hypotheses are $H_0: p \geq 0.10$ versus $H_1: p < 0.10$.

   $z = (0.074747 - 0.10)/\sqrt{0.10(1 - 0.10)/495} = -1.87$.

   Since the alternate hypothesis is of the form $p < p_0$, the P-value is the area to the left of $z = -1.87$, so $P = 0.0307$.

   Since there are four samples altogether, the Bonferroni-adjusted P-value is $4(0.0307) = 0.1228$.

   We cannot conclude that the failure rate on line 3 is less than 0.10.

9. The row totals are $O_{1.} = 214$ and $O_{2.} = 216$. The column totals are $O_{.1} = 65$, $O_{.2} = 121$, $O_{.3} = 244$. The grand total is $O_{..} = 430$.

   The expected values are $E_{ij} = O_{i.}O_{.j}/O_{..}$, as shown in the following table.

	Numbers of Skeletons		
Site	0–4 years	5–19 years	20 years or more
Casa da Moura	32.349	60.219	121.43
Wandersleben	32.651	60.781	122.57

   There are $(2 - 1)(3 - 1) = 2$ degrees of freedom.

   $\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{3}(O_{ij} - E_{ij})^2/E_{ij} = 2.1228$.

   From the χ^2 table, $P > 0.10$. A computer package gives $P = 0.346$.

   We cannot conclude that the age distributions differ between the two sites.

Chapter 7

Section 7.1

1. $\overline{X} = 26.50$, $s_X = 2.37$, $n_X = 39$, $\overline{Y} = 37.14$, $s_Y = 3.66$, $n_Y = 142$, $z_{.025} = 1.96$.
   The confidence interval is $37.14 - 26.50 \pm 1.96\sqrt{2.37^2/39 + 3.66^2/142}$, or (9.683, 11.597).

3. $\overline{X} = 517.0$, $s_X = 2.4$, $n_X = 35$, $\overline{Y} = 510.1$, $s_Y = 2.1$, $n_Y = 47$, $z_{.005} = 2.58$.
   The confidence interval is $517.0 - 510.1 \pm 2.58\sqrt{2.4^2/35 + 2.1^2/47}$, or (5.589, 8.211).

5. $\overline{X} = 8.5$, $s_X = 1.9$, $n_X = 58$, $\overline{Y} = 11.9$, $s_Y = 3.6$, $n_Y = 58$, $z_{.005} = 2.58$
   The confidence interval is $11.9 - 8.5 \pm 2.58\sqrt{1.9^2/58 + 3.6^2/58}$, or (2.0210, 4.7790).

7. It is not possible. The amounts of time spent in bed and spent asleep in bed are not independent.

9. $\overline{X} = 5.92$, $s_X = 0.15$, $n_X = 42$, $\overline{Y} = 6.05$, $s_Y = 0.16$, $n_Y = 37$, $z_{0.025} = 1.96$
   The confidence interval is $6.05 - 5.92 \pm 1.96\sqrt{0.15^2/42 + 0.16^2/37}$, or $(0.06133, 0.19867)$

11. $\overline{X} = 242$, $s_X = 20$, $n_X = 38$, $\overline{Y} = 180$, $s_Y = 31$, $n_Y = 42$.
    The null and alternate hypotheses are $H_0\colon \mu_X - \mu_Y \leq 50$ versus $H_1\colon \mu_X - \mu_Y > 50$.
    $z = (242 - 180 - 50)/\sqrt{20^2/38 + 31^2/42} = 2.08$. Since the alternate hypothesis is of the form $\mu_X - \mu_Y > \Delta$, the P-value is the area to the right of $z = 2.08$.
    Thus $P = 0.0188$.
    The more expensive inhibitor should be used.

13. $\overline{X} = 92.3$, $s_X = 6.2$, $n_X = 70$, $\overline{Y} = 90.2$, $s_Y = 4.4$, $n_Y = 60$.
    The null and alternate hypotheses are $H_0\colon \mu_X - \mu_Y \leq 0$ versus $H_1\colon \mu_X - \mu_Y > 0$.
    $z = (92.3 - 90.2 - 0)/\sqrt{6.2^2/70 + 4.4^2/60} = 2.25$. Since the alternate hypothesis is of the form $\mu_X - \mu_Y > \Delta$, the P-value is the area to the right of $z = 2.25$.
    Thus $P = 0.0122$.
    We can conclude that the mean hardness of welds cooled at 10°C/s is greater than that of welds cooled at 40°C/s.

15. $\overline{X} = 0.67$, $s_X = 0.46$, $n_X = 80$, $\overline{Y} = 0.59$, $s_Y = 0.38$, $n_Y = 60$.

The null and alternate hypotheses are $H_0\colon \mu_X - \mu_Y \leq 0$ versus $H_1\colon \mu_X - \mu_Y > 0$.

$z = (0.67 - 0.59 - 0)/\sqrt{0.46^2/80 + 0.38^2/60} = 1.13$. Since the alternate hypothesis is of the form $\mu_X - \mu_Y > \Delta$, the P-value is the area to the right of $z = 1.13$.

Thus $P = 0.1292$.

We cannot conclude that the mean proportoin of heat recovered is greater at the lower flow speed.

17. (a) $\overline{X} = 7.79$, $s_X = 1.06$, $n_X = 80$, $\overline{Y} = 7.64$, $s_Y = 1.31$, $n_Y = 80$.

Here $\mu_1 = \mu_X$ and $\mu_2 = \mu_Y$. The null and alternate hypotheses are $H_0\colon \mu_X - \mu_Y \leq 0$ versus $H_1\colon \mu_X - \mu_Y > 0$.

$z = (7.79 - 7.64 - 0)/\sqrt{1.06^2/80 + 1.31^2/80} = 0.80$. Since the alternate hypothesis is of the form $\mu_X - \mu_Y > \Delta$, the P-value is the area to the right of $z = 0.80$.

Thus $P = 0.2119$.

We cannot conclude that the mean score on one-tailed questions is greater.

(b) The null and alternate hypotheses are $H_0\colon \mu_X - \mu_Y = 0$ versus $H_1\colon \mu_X - \mu_Y \neq 0$.

The z-score is computed as in part (a): $z = 0.80$.

Since the alternate hypothesis is of the form $\mu_X - \mu_Y \neq \Delta$, the P-value is the sum of the areas to the right of $z = 0.80$ and to the left of $z = 0.80$.

Thus $P = 0.2119 + 0.2119 = 0.4238$.

We cannot conclude that the mean score on one-tailed questions differs from the mean score on two-tailed questions.

19. (a) $\overline{X} = 645$, $s_X = 50$, $n_X = 64$, $\overline{Y} = 625$, $s_Y = 40$, $n_Y = 100$.

The null and alternate hypotheses are $H_0\colon \mu_X - \mu_Y \leq 0$ versus $H_1\colon \mu_X - \mu_Y > 0$.

$z = (645 - 625 - 0)/\sqrt{50^2/64 + 40^2/100} = 2.7$. Since the alternate hypothesis is of the form $\mu_X - \mu_Y > \Delta$, the P-value is the area to the right of $z = 2.7$.

Thus $P = 0.0035$.

We can conclude that the second method yields the greater mean daily production.

(b) $\overline{X} = 645$, $s_X = 50$, $n_X = 64$, $\overline{Y} = 625$, $s_Y = 40$, $n_Y = 100$.

The null and alternate hypotheses are $H_0\colon \mu_X - \mu_Y \leq 5$ versus $H_1\colon \mu_X - \mu_Y > 5$.

$z = (645 - 625 - 5)/\sqrt{50^2/64 + 40^2/100} = 2.02$. Since the alternate hypothesis is of the form $\mu_X - \mu_Y > \Delta$, the P-value is the area to the right of $z = 2.02$.

Thus $P = 0.0217$.

We can conclude that the mean daily production for the second method exceeds that of the first by more than 5 tons.

21. (a) (i) StDev = (SE Mean)$\sqrt{\text{N}} = 1.26\sqrt{78} = 11.128$.

(ii) SE Mean = StDev/$\sqrt{\text{N}}$ = $3.02/\sqrt{63} = 0.380484$.

(b) $z = (23.3 - 20.63 - 0)/\sqrt{1.26^2 + 0.380484^2} = 2.03$. Since the alternate hypothesis is of the form $\mu_X - \mu_Y \neq \Delta$, the P-value is the sum of the areas to the right of $z = 2.03$ and to the left of $z = -2.03$. Thus $P = 0.0212 + 0.0212 = 0.0424$, and the result is similar to that of the t test.

(c) $\overline{X} = 23.3$, $s_X/\sqrt{n_X} = 1.26$, $\overline{Y} = 20.63$, $s_Y/\sqrt{n_Y} = 0.380484$, $z_{.01} = 2.33$.
The confidence interval is $23.3 - 20.63 \pm 2.33\sqrt{1.26^2 + 0.380484^2}$, or $(-0.3967, 5.7367)$.

Section 7.2

1. $X = 1919$, $n_X = 1985$, $\tilde{p}_X = (1919 + 1)/(1985 + 2) = 0.966281$,
$Y = 4561$, $n_Y = 4988$, $\tilde{p}_Y = (4561 + 1)/(4988 + 2) = 0.914228$, $z_{.005} = 2.58$.
The confidence interval is $0.966281 - 0.914228 \pm 2.58\sqrt{\dfrac{0.966281(1 - 0.966281)}{1985 + 2} + \dfrac{0.914228(1 - 0.914228)}{4988 + 2}}$, or (0.0374, 0.0667).

3. $X = 32$, $n_X = 1000$, $\tilde{p}_X = (32 + 1)/(1000 + 2) = 0.032934$,
$Y = 15$, $n_Y = 1000$, $\tilde{p}_Y = (15 + 1)/(1000 + 2) = 0.01597$, $z_{.025} = 1.96$.
The confidence interval is $0.032934 - 0.01597 \pm 1.96\sqrt{\dfrac{0.032934(1 - 0.032934)}{1000 + 2} + \dfrac{0.01597(1 - 0.01597)}{1000 + 2}}$, or (0.00346, 0.03047).

5. $X = 92$, $n_X = 500$, $\tilde{p}_X = (92 + 1)/(500 + 2) = 0.18526$,
$Y = 65$, $n_Y = 500$, $\tilde{p}_Y = (65 + 1)/(500 + 2) = 0.13147$, $z_{.005} = 2.58$.
The confidence interval is $0.18526 - 0.13147 \pm 2.58\sqrt{\dfrac{0.18526(1 - 0.18526)}{500 + 2} + \dfrac{0.13147(1 - 0.13147)}{500 + 2}}$, or $(-0.0055, 0.1131)$.

7. No. The sample proportions come from the same sample rather than from two independent samples.

9. (a) $H_0\colon p_X - p_Y \geq 0$ versus $H_1\colon p_X - p_Y < 0$

(b) $X = 960$, $n_X = 1000$, $\widehat{p}_X = 960/1000 = 0.960$, $Y = 582$, $n_Y = 600$, $\widehat{p}_Y = 582/600 = 0.970$,
$\widehat{p} = (960 + 582)/(1000 + 600) = 0.96375$.
The null and alternate hypotheses are $H_0\colon p_X - p_Y \geq 0$ versus $H_1\colon p_X - p_Y < 0$.

$$z = \frac{0.960 - 0.970}{\sqrt{0.96375(1 - 0.96375)(1/1000 + 1/600)}} = -1.04.$$

Since the alternate hypothesis is of the form $p_X - p_Y < 0$, the P-value is the area to the left of $z = -1.04$.

Thus $P = 0.1492$.

(c) Since $P = 0.1492$, we cannot conclude that machine 2 is better. Therefore machine 1 should be used.

11. $X = 133$, $n_X = 400$, $\widehat{p}_X = 133/400 = 0.3325$, $Y = 50$, $n_Y = 100$, $\widehat{p}_Y = 50/100 = 0.5$,
$\widehat{p} = (133 + 50)/(400 + 100) = 0.366$.

The null and alternate hypotheses are $H_0\colon p_X - p_Y = 0$ versus $H_1\colon p_X - p_Y \neq 0$.

$$z = \frac{0.3325 - 0.5}{\sqrt{0.366(1 - 0.366)(1/400 + 1/100)}} = -3.11.$$

Since the alternate hypothesis is of the form $p_X - p_Y \neq 0$, the P-value is the sum of the areas to the right of $z = 3.11$ and to the left of $z = -3.11$.

Thus $P = 0.0009 + 0.0009 = 0.0018$.

We can conclude that the response rates differ between public and private firms.

13. $X = 285$, $n_X = 500$, $\widehat{p}_X = 285/500 = 0.57$, $Y = 305$, $n_Y = 600$, $\widehat{p}_Y = 305/600 = 0.50833$,
$\widehat{p} = (285 + 305)/(500 + 600) = 0.53636$.

The null and alternate hypotheses are $H_0\colon p_X - p_Y \leq 0$ versus $H_1\colon p_X - p_Y > 0$.

$$z = \frac{0.57 - 0.50833}{\sqrt{0.53636(1 - 0.53636)(1/500 + 1/600)}} = 2.04.$$

Since the alternate hypothesis is of the form $p_X - p_Y > 0$, the P-value is the area to the right of $z = 2.04$.

Thus $P = 0.0207$.

We can conclude that the proportion of voters favoring the proposal is greater in county A than in county B.

15. $X = 18$, $n_X = 77$, $\widehat{p}_X = 18/77 = 0.23377$, $Y = 38$, $n_Y = 280$, $\widehat{p}_Y = 38/280 = 0.13571$,
$\widehat{p} = (18 + 38)/(77 + 280) = 0.15686$.

The null and alternate hypotheses are $H_0\colon p_X - p_Y \leq 0$ versus $H_1\colon p_X - p_Y > 0$.

$$z = \frac{0.23377 - 0.13571}{\sqrt{0.15686(1 - 0.15686)(1/77 + 1/280)}} = 2.10.$$

Since the alternate hypothesis is of the form $p_X - p_Y > 0$, the P-value is the area to the right of $z = 2.10$.

Thus $P = 0.0179$.

We can conclude that the proportion is greater at the higher elevation.

17. $X = 22$, $n_X = 41$, $\widehat{p}_X = 22/41 = 0.53659$, $Y = 18$, $n_Y = 31$, $\widehat{p}_Y = 18/31 = 0.58065$,
$\widehat{p} = (22+18)/(41+31) = 0.55556$.

The null and alternate hypotheses are $H_0\colon p_X - p_Y = 0$ versus $H_1\colon p_X - p_Y \neq 0$.

$$z = \frac{0.53659 - 0.58065}{\sqrt{0.55556(1-0.55556)(1/41+1/31)}} = -0.37.$$

Since the alternate hypothesis is of the form $p_X - p_Y \neq 0$, the P-value is the sum of the areas to the right of $z = 0.37$ and to the left of $z = -0.37$.

Thus $P = 0.3557 + 0.3557 = 0.7114$.

We cannot conclude that the proportion of wells that meet the standards differs between the two areas.

19. No, these are not simple random samples.

21. (a) $101/153 = 0.660131$.

(b) $90(0.544444) = 49$.

(c) $X_1 = 101$, $n_1 = 153$, $\widehat{p}_1 = 101/153 = 0.660131$, $X_2 = 49$, $n_2 = 90$, $\widehat{p}_2 = 49/90 = 0.544444$,
$\widehat{p} = (101+49)/(153+90) = 0.617284$.

$$z = \frac{0.660131 - 0.544444}{\sqrt{0.617284(1-0.617284)(1/153+1/90)}} = 1.79.$$

(d) Since the alternate hypothesis is of the form $p_X - p_Y \neq 0$, the P-value is the sum of the areas to the right of $z = 1.79$ and to the left of $z = -1.79$.

Thus $P = 0.0367 + 0.0367 = 0.0734$.

Section 7.3

1. $\overline{X} = 25.286$, $s_X = 7.8042$, $n_X = 7$, $\overline{Y} = 16.714$, $s_Y = 3.4983$, $n_Y = 7$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{7.8042^2}{7} + \dfrac{3.4983^2}{7}\right]^2}{\dfrac{(7.8042^2/7)^2}{7-1} + \dfrac{(3.4983^2/7)^2}{7-1}} = 8, \text{ rounded down to the nearest integer.}$$

$t_{8,.025} = 2.306$, so the confidence interval is $25.286 - 16.714 \pm 2.306\sqrt{\dfrac{7.8042^2}{7} + \dfrac{3.4983^2}{7}}$,
or $(1.117, 16.026)$.

3. $\overline{X} = 73.1$, $s_X = 9.1$, $n_X = 10$, $\overline{Y} = 53.9$, $s_Y = 10.7$, $n_Y = 10$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{9.1^2}{10} + \dfrac{10.7^2}{10}\right]^2}{\dfrac{(9.1^2/10)^2}{10-1} + \dfrac{(10.7^2/10)^2}{10-1}} = 17\text{, rounded down to the nearest integer.}$$

$t_{17,.01} = 2.567$, so the confidence interval is $73.1 - 53.9 \pm 2.567\sqrt{\dfrac{9.1^2}{10} + \dfrac{10.7^2}{10}}$, or $(7.798, 30.602)$.

5. $\overline{X} = 33.8$, $s_X = 0.5$, $n_X = 4$, $\overline{Y} = 10.7$, $s_Y = 3.3$, $n_Y = 8$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{0.5^2}{4} + \dfrac{3.3^2}{8}\right]^2}{\dfrac{(0.5^2/4)^2}{4-1} + \dfrac{(3.3^2/8)^2}{8-1}} = 7\text{, rounded down to the nearest integer.}$$

$t_{7,.025} = 2.365$, so the confidence interval is $33.8 - 10.7 \pm 2.365\sqrt{\dfrac{0.5^2}{4} + \dfrac{3.3^2}{8}}$, or $(20.278, 25.922)$.

7. $\overline{X} = 0.498$, $s_X = 0.036$, $n_X = 5$, $\overline{Y} = 0.389$, $s_Y = 0.049$, $n_Y = 5$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{0.036^2}{5} + \dfrac{0.049^2}{5}\right]^2}{\dfrac{(0.036^2/5)^2}{5-1} + \dfrac{(0.049^2/5)^2}{5-1}} = 7\text{, rounded down to the nearest integer.}$$

$t_{7,.025} = 2.365$, so the confidence interval is $0.498 - 0.389 \pm 2.365\sqrt{\dfrac{0.036^2}{5} + \dfrac{0.049^2}{5}}$, or $(0.0447, 0.173)$.

9. $\overline{X} = 229.5429$, $s_X = 14.169$, $n_X = 7$, $\overline{Y} = 143.9556$, $s_Y = 59.757$, $n_Y = 9$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{14.169^2}{7} + \dfrac{59.757^2}{9}\right]^2}{\dfrac{(14.169^2/7)^2}{7-1} + \dfrac{(59.757^2/9)^2}{9-1}} = 9\text{, rounded down to the nearest integer.}$$

$t_{9,.025} = 2.262$, so the confidence interval is $229.5429 - 143.9556 \pm 2.262\sqrt{\dfrac{14.169^2}{7} + \dfrac{59.757^2}{9}}$, or $(38.931, 132.24)$.

11. $\overline{X} = 482.79$, $s_X = 13.942$, $n_X = 14$, $\overline{Y} = 464.7$, $s_Y = 14.238$, $n_Y = 9$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{13.942^2}{14} + \dfrac{14.238^2}{9}\right]^2}{\dfrac{(13.942^2/14)^2}{14-1} + \dfrac{(14.238^2/9)^2}{9-1}} = 16\text{, rounded down to the nearest integer.}$$

$t_{16} = (482.79 - 464.7 - 0)/\sqrt{13.942^2/14 + 14.238^2/9} = 2.9973$.

The null and alternate hypotheses are $H_0\colon \mu_X - \mu_Y \le 0$ versus $H_1\colon \mu_X - \mu_Y > 0$.

Since the alternate hypothesis is of the form $\mu_X - \mu_Y > \Delta$, the P-value is the area to the right of $t = 2.9973$.

From the t table, $0.001 < P < 0.005$. A computer package gives $P = 0.00426$.

We can conclude that hockey sticks made from composite B have greater mean breaking strength.

13. $\overline{X} = 60.4$, $s_X = 5.461$, $n_X = 10$, $\overline{Y} = 62.714$, $s_Y = 3.8607$, $n_Y = 7$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{5.461^2}{10} + \dfrac{3.8607^2}{7}\right]^2}{\dfrac{(5.461^2/10)^2}{10-1} + \dfrac{(3.8607^2/7)^2}{7-1}} = 14\text{, rounded down to the nearest integer.}$$

$t_{14} = (60.4 - 62.714 - 0)/\sqrt{5.461^2/10 + 3.8607^2/7} = -1.0236$. The null and alternate hypotheses are $H_0\colon \mu_X - \mu_Y = 0$ versus $H_1\colon \mu_X - \mu_Y \ne 0$.

Since the alternate hypothesis is of the form $\mu_X - \mu_Y \ne \Delta$, the P-value is the sum of the areas to the right of $t = 1.0236$ and to the left of $t = -1.0236$.

From the t table, $0.20 < P < 0.50$. A computer package gives $P = 0.323$.

We cannot conclude that the mean permeability coefficients differ.

15. $\overline{X} = 1.51$, $s_X = 0.38678$, $n_X = 5$, $\overline{Y} = 2.2544$, $s_Y = 0.66643$, $n_Y = 9$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{0.38678^2}{5} + \dfrac{0.66643^2}{9}\right]^2}{\dfrac{(0.38678^2/5)^2}{5-1} + \dfrac{(0.66643^2/9)^2}{9-1}} = 11\text{, rounded down to the nearest integer.}$$

$t_{11} = (1.51 - 2.2544 - 0)/\sqrt{0.38678^2/5 + 0.66643^2/9} = -2.6441$. The null and alternate hypotheses are $H_0\colon \mu_X - \mu_Y = 0$ versus $H_1\colon \mu_X - \mu_Y \ne 0$.

Since the alternate hypothesis is of the form $\mu_X - \mu_Y \ne \Delta$, the P-value is the sum of the areas to the right of $t = 2.6441$ and to the left of $t = -2.6441$.

From the t table, $0.02 < P < 0.05$. A computer package gives $P = 0.0228$.

It is not plausible that the mean resilient modulus is the same for rutted and nonrutted pavement.

17. $\overline{X} = 2.1062$, $s_X = 0.029065$, $n_X = 5$, $\overline{Y} = 2.0995$, $s_Y = 0.033055$, $n_Y = 5$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{0.029065^2}{5} + \dfrac{0.033055^2}{5}\right]^2}{\dfrac{(0.029065^2/5)^2}{5-1} + \dfrac{(0.033055^2/5)^2}{5-1}} = 7\text{, rounded down to the nearest integer.}$$

$t_7 = (2.1062 - 2.0995 - 0)/\sqrt{0.029065^2/5 + 0.033055^2/5} = 0.3444$.

The null and alternate hypotheses are $H_0\colon \mu_X - \mu_Y = 0$ versus $H_1\colon \mu_X - \mu_Y \neq 0$.

Since the alternate hypothesis is of the form $\mu_X - \mu_Y \neq \Delta$, the P-value is the sum of the areas to the right of $t = 0.3444$ and to the left of $t = -0.3444$.

From the t table, $0.50 < P < 0.80$. A computer package gives $P = 0.741$.

We cannot conclude that the calibration has changed from the first to the second day.

19. $\overline{X} = 22.1$, $s_X = 4.09$, $n_X = 11$, $\overline{Y} = 20.4$, $s_Y = 3.08$, $n_Y = 7$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{4.09^2}{11} + \dfrac{3.08^2}{7}\right]^2}{\dfrac{(4.09^2/11)^2}{11-1} + \dfrac{(3.08^2/7)^2}{7-1}} = 15\text{, rounded down to the nearest integer.}$$

$t_{15} = (22.1 - 20.4 - 0)/\sqrt{4.09^2/11 + 3.08^2/7} = 1.002$.

The null and alternate hypotheses are $H_0\colon \mu_X - \mu_Y \leq 0$ versus $H_1\colon \mu_X - \mu_Y > 0$.

Since the alternate hypothesis is of the form $\mu_X - \mu_Y > \Delta$, the P-value is the area to the right of $t = 1.002$.

From the t table, $0.10 < P < 0.25$. A computer package gives $P = 0.166$.

We cannot conclude that the mean compressive stress is greater for no. 1 grade lumber than for no. 2 grade.

21. (a) $\overline{X} = 77.74$, $s_X = 1.6072$, $n_X = 5$, $\overline{Y} = 72.86$, $s_Y = 2.9091$, $n_Y = 5$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{1.6072^2}{5} + \dfrac{2.9091^2}{5}\right]^2}{\dfrac{(1.6072^2/5)^2}{5-1} + \dfrac{(2.9091^2/5)^2}{5-1}} = 6\text{, rounded down to the nearest integer.}$$

$t_6 = (77.74 - 72.86 - 0)/\sqrt{1.6072^2/5 + 2.9091^2/5} = 3.2832$. The null and alternate hypotheses are $H_0\colon \mu_X - \mu_Y \leq 0$ versus $H_1\colon \mu_X - \mu_Y > 0$.

Since the alternate hypothesis is of the form $\mu_X - \mu_Y > \Delta$, the P-value is the area to the right of $t = 3.2832$.

From the t table, $0.005 < P < 0.01$. A computer package gives $P = 0.00838$.

We can conclude that the mean yield for method B is greater than that of method A.

(b) $\overline{X} = 77.74$, $s_X = 1.6072$, $n_X = 5$, $\overline{Y} = 72.86$, $s_Y = 2.9091$, $n_Y = 5$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{1.6072^2}{5} + \dfrac{2.9091^2}{5}\right]^2}{\dfrac{(1.6072^2/5)^2}{5-1} + \dfrac{(2.9091^2/5)^2}{5-1}} = 6, \text{ rounded down to the nearest integer.}$$

$t_6 = (77.74 - 72.86 - 3)/\sqrt{1.6072^2/5 + 2.9091^2/5} = 1.2649$. The null and alternate hypotheses are $H_0\colon \mu_X - \mu_Y \leq 3$ versus $H_1\colon \mu_X - \mu_Y > 3$.

Since the alternate hypothesis is of the form $\mu_X - \mu_Y > \Delta$, the P-value is the area to the right of $t = 1.2649$.

From the t table, $0.10 < P < 0.25$. A computer package gives $P = 0.126$.

We cannot conclude that the mean yield for method B exceeds that of method A by more than 3.

23. (a) SE Mean $=$ StDev$/\sqrt{\text{N}} = 0.482/\sqrt{6} = 0.197$.

(b) StDev $=$ (SE Mean)$\sqrt{\text{N}} = 0.094\sqrt{13} = 0.339$.

(c) $\overline{X} - \overline{Y} = 1.755 - 3.239 = -1.484$.

(d) $t = \dfrac{1.755 - 3.239}{\sqrt{0.482^2/6 + 0.094^2}} = -6.805$.

Section 7.4

1. $\overline{D} = 0.40929$, $s_D = 0.17604$, $n = 14$, $t_{14-1,.025} = 2.160$.

The confidence interval is $0.40929 \pm 2.160(0.17604/\sqrt{14})$, or (0.308, 0.511).

3. The differences are: $2.96, 3.33, 3.17, 2.81, 3.73, 2.42, 5.18$.

$\overline{D} = 3.3714$, $s_D = 0.89736$, $n = 7$, $t_{7-1,.05} = 1.943$.

The confidence interval is $3.3714 \pm 1.943(0.89736/\sqrt{7})$, or (2.712, 4.030).

5. $\overline{D} = 6.736667$, $s_D = 6.045556$, $n = 9$, $t_{9-1,.025} = 2.306$.

The confidence interval is $6.736667 \pm 2.306(6.045556/\sqrt{9})$, or (2.090, 11.384).

7. (a) The differences are: $3.8, 2.6, 2.0, 2.9, 2.2, -0.2, 0.5, 1.3, 1.3, 2.1, 4.8, 1.5, 3.4, 1.4, 1.1, 1.9, -0.9, -0.3$.
$\overline{D} = 1.74444$, $s_D = 1.46095$, $n = 18$, $t_{18-1,.005} = 2.898$.
The confidence interval is $1.74444 \pm 2.898(1.46095/\sqrt{18})$, or (0.747, 2.742).

(b) The level $100(1-\alpha)\%$ can be determined from the equation $t_{17,\alpha/2}(1.46095/\sqrt{18}) = 0.5$.
From this equation, $t_{17,\alpha/2} = 1.452$. The t table indicates that the value of $\alpha/2$ is between 0.05 and 0.10, and closer to 0.10. Therefore the level $100(1-\alpha)\%$ is closest to 80%.

9. (a) $H_0\colon \mu_1 - \mu_2 = 0$ versus $H_1\colon \mu_1 - \mu_2 \neq 0$

(b) $\overline{D} = 1608.143$, $s_D = 2008.147$, $n = 7$. There are $7 - 1 = 6$ degrees of freedom. The null and alternate hypotheses are $H_0\colon \mu_D = 0$ versus $H_1\colon \mu_D \neq 0$.
$t = (1608.143 - 0)/(2008.147/\sqrt{7}) = 2.119$.

(c) Since the alternate hypothesis is of the form $\mu_D \neq \mu_0$, the P-value is the sum of the areas to the right of $t = 2.119$ and to the left of $t = -2.119$.
From the t table, $0.05 < P < 0.10$. A computer package gives $P = 0.078$.
The null hypothesis is suspect, but one would most likely not firmly conclude that it is false.

11. $\overline{D} = 0.58$, $s_D = 0.23358$, $n = 6$. There are $6 - 1 = 5$ degrees of freedom.
The null and alternate hypotheses are $H_0\colon \mu_D = 0$ versus $H_1\colon \mu_D \neq 0$.
$t = (0.58 - 0)/(0.23358/\sqrt{6}) = 6.0823$.
Since the alternate hypothesis is of the form $\mu_D \neq \mu_0$, the P-value is the sum of the areas to the right of $t = 6.0823$ and to the left of $t = -6.0823$.
From the t table, $0.001 < P < 0.002$. A computer package gives $P = 0.00174$.
We can conclude that there is a difference in the mean concentration between the shoot and the root.

13. $\overline{D} = 4.2857$, $s_D = 1.6036$, $n = 7$. There are $7 - 1 = 6$ degrees of freedom.
The null and alternate hypotheses are $H_0\colon \mu_D = 0$ versus $H_1\colon \mu_D \neq 0$.
$t = (4.2857 - 0)/(1.6036/\sqrt{7}) = 7.071$.
Since the alternate hypothesis is of the form $\mu_D \neq \Delta$, the P-value is the sum of the areas to the right of $t = 7.071$ and to the left of $t = -7.071$.
From the t table, $P < 0.001$. A computer package gives $P = 0.00040$.
We can conclude that there is a difference in latency between motor point and nerve stimulation.

15. $\overline{D} = 0.17625$, $s_D = 0.48432$, $n = 8$. There are $8 - 1 = 7$ degrees of freedom.

The null and alternate hypotheses are $H_0\colon \mu_D = 0$ versus $H_1\colon \mu_D \neq 0$.

$t = (0.17625 - 0)/(0.48432/\sqrt{8}) = 1.0293$.

Since the alternate hypothesis is of the form $\mu_D \neq \Delta$, the P-value is the sum of the areas to the right of $t = 1.0293$ and to the left of $t = -1.0293$.

From the t table, $0.20 < P < 0.50$. A computer package gives $P = 0.338$.

We cannot conclude that there is a difference in mean weight loss between specimens cured at the two temperatures.

17. (a) The differences are $5.0, 4.6, 1.9, 2.6, 4.4, 3.2, 3.2, 2.8, 1.6, 2.8$.

Let μ_R be the mean number of miles per gallon for taxis using radial tires, and let μ_B be the mean number of miles per gallon for taxis using bias tires. The appropriate null and alternate hypotheses are $H_0\colon \mu_R - \mu_B \leq 0$ versus $H_1\colon \mu_R - \mu_B > 0$.

$\overline{D} = 3.21$ $s_D = 1.1338$, $n = 10$. There are $10 - 1 = 9$ degrees of freedom.

$t = (3.21 - 0)/(1.1338/\sqrt{10}) = 8.953$.

Since the alternate hypothesis is of the form $\mu_D > \Delta$, the P-value is the area to the right of $t = 8.953$.

From the t table, $P < 0.0050$. A computer package gives $P = 4.5 \times 10^{-6}$.

We can conclude that the mean number of miles per gallon is higher with radial tires.

(b) The appropriate null and alternate hypotheses are $H_0\colon \mu_R - \mu_B \leq 2$ vs. $H_1\colon \mu_R - \mu_B > 2$.

$\overline{D} = 3.21$ $s_D = 1.1338$, $n = 10$. There are $10 - 1 = 9$ degrees of freedom.

$t = (3.21 - 2)/(1.1338/\sqrt{10}) = 3.375$.

Since the alternate hypothesis is of the form $\mu_D > \Delta$, the P-value is the area to the right of $t = 3.375$.

From the t table, $0.001 < P < 0.005$. A computer package gives $P = 0.0041$. We can conclude that the mean mileage with radial tires is more than 2 miles per gallon higher than with bias tires.

19. (a) SE Mean = StDev/$\sqrt{\text{N}}$ = $2.9235/\sqrt{7} = 1.1050$.

(b) StDev = (SE Mean)$\sqrt{\text{N}}$ = $1.0764\sqrt{7} = 2.8479$.

(c) $\mu_D = \mu_X - \mu_Y = 12.4141 - 8.3476 = 4.0665$.

(d) $t = (4.0665 - 0)/1.19723 = 3.40$.

Section 7.5

1. $\nu_1 = 7$, $\nu_2 = 20$. From the F table, the upper 5% point is 2.51.

3. (a) The upper 1% point of the $F_{5,\,7}$ distribution is 7.46. Therefore the P-value is 0.01.

   (b) The P-value for a two-tailed test is twice the value for the corresponding one-tailed test. Therefore $P = 0.02$.

5. The sample variance of the breaking strengths for composite A is $\sigma_1^2 = 202.7175$. The sample size is $n_1 = 9$.

   The sample variance of the breaking strengths for composite B is $\sigma_2^2 = 194.3829$. The sample size is $n_2 = 14$.

   The null and alternate hypotheses are $H_0\colon \sigma_1^2/\sigma_2^2 = 1$ versus $H_1\colon \sigma_1^2/\sigma_2^2 \neq 1$.

   The test statistic is $F = \sigma_1^2/\sigma_2^2 = 1.0429$. The numbers of degrees of freedom are 8 and 13.

   Since this is a two-tailed test, the P-value is twice the area to the right of 1.0429 under the $F_{8,\,13}$ probability density function.

   From the F table, $P > 0.2$. A computer package gives $P = 0.91$.

   We cannot conclude that the variance of the breaking strength varies between the composites.

Supplementary Exercises for Chapter 7

1. $\overline{X} = 40$, $s_X = 12$, $n_X = 75$, $\overline{Y} = 42$, $s_Y = 15$, $n_Y = 100$.

   The null and alternate hypotheses are $H_0\colon \mu_X - \mu_Y > 0$ versus $H_1\colon \mu_X - \mu_Y \leq 0$.

   $z = (40-42-0)/\sqrt{12^2/75 + 15^2/100} = -0.98$. Since the alternate hypothesis is of the form $\mu_X - \mu_Y \leq \Delta$, the P-value is the area to the left of $z = -0.98$.

   Thus $P = 0.1635$.

   We cannot conclude that the mean reduction from drug B is greater than the mean reduction from drug A.

3. $\overline{X}_1 = 4387$, $s_1 = 252$, $n_1 = 75$, $\overline{X}_2 = 4260$, $s_2 = 231$, $n_2 = 75$.

   The null and alternate hypotheses are $H_0\colon \mu_1 - \mu_2 \leq 0$ versus $H_1\colon \mu_1 - \mu_2 > 0$.

   $z = (4387 - 4260 - 0)/\sqrt{252^2/75 + 231^2/75} = 3.22$. Since the alternate hypothesis is of the form $\mu_X - \mu_Y > \Delta$, the P-value is the area to the right of $z = 3.22$.

   Thus $P = 0.0006$. We can conclude that new power supplies outlast old power supplies.

5. $X = 20$, $n_X = 100$, $\tilde{p}_X = (20+1)/(100+2) = 0.205882$,
$Y = 10$, $n_Y = 150$, $\tilde{p}_Y = (10+1)/(150+2) = 0.072368$, $z_{.05} = 1.645$.

The confidence interval is $0.205882 - 0.072368 \pm 1.645\sqrt{\dfrac{0.205882(1-0.205882)}{100+2} + \dfrac{0.072368(1-0.072368)}{150+2}}$,
or $(0.0591,\ 0.208)$.

7. (a) $X = 62$, $n_X = 400$, $\tilde{p}_X = (62+1)/(400+2) = 0.15672$,
$Y = 12$, $n_Y = 100$, $\tilde{p}_Y = (12+1)/(100+2) = 0.12745$, $z_{.025} = 1.96$.

The confidence interval is $0.15672 - 0.12745 \pm 1.96\sqrt{\dfrac{0.15672(1-0.15672)}{400+2} + \dfrac{0.12745(1-0.12745)}{100+2}}$,
or $(-0.0446, 0.103)$.

(b) The width of the confidence interval is $\pm 1.96\sqrt{\dfrac{\tilde{p}_X(1-\tilde{p}_X)}{n_X+2} + \dfrac{\tilde{p}_Y(1-\tilde{p}_Y)}{n_Y+2}}$.

Estimate $\tilde{p}_X = 0.15672$ and $\tilde{p}_Y = 0.12745$.

Then if 100 additional chips were sampled from the less expensive process, $n_X = 500$ and $n_Y = 100$, so the width of the confidence interval would be approximately

$$\pm 1.96\sqrt{\frac{0.15672(1-0.15672)}{502} + \frac{0.12745(1-0.12745)}{102}} = \pm 0.0721.$$

If 50 additional chips were sampled from the more expensive process, $n_X = 400$ and $n_Y = 150$, so the width of the confidence interval would be approximately

$$\pm 1.96\sqrt{\frac{0.15672(1-0.15672)}{402} + \frac{0.12745(1-0.12745)}{152}} = \pm 0.0638.$$

If 50 additional chips were sampled from the less expensive process and 25 additional chips were sampled from the more expensive process, $n_X = 450$ and $n_Y = 125$, so the width of the confidence interval would be approximately

$$\pm 1.96\sqrt{\frac{0.15672(1-0.15672)}{452} + \frac{0.12745(1-0.12745)}{127}} = \pm 0.0670.$$

Therefore the greatest increase in precision would be achieved by sampling 50 additional chips from the more expensive process.

9. No, because the two samples are not independent.

11. $X = 57$, $n_X = 100$, $\widehat{p}_X = 57/100 = 0.57$, $Y = 135$, $n_Y = 200$, $\widehat{p}_Y = 135/200 = 0.675$,
$\widehat{p} = (57+135)/(100+200) = 0.64$.

The null and alternate hypotheses are $H_0\colon p_X - p_Y \geq 0$ versus $H_1\colon p_X - p_Y < 0$.

$$z = \frac{0.57 - 0.675}{\sqrt{0.64(1-0.64)(1/100+1/200)}} = -1.79.$$

Since the alternate hypothesis is of the form $p_X - p_Y < 0$, the P-value is the area to the left of $z = -1.79$.

Thus $P = 0.0367$.

We can conclude that awareness of the benefit increased after the advertising campaign.

13. The differences are $21, 18, 5, 13, -2, 10$.

$\overline{D} = 10.833$, $s_D = 8.471521$, $n = 6$, $t_{6-1,.025} = 2.571$.

The confidence interval is $10.833 \pm 2.571(8.471521/\sqrt{6})$, or $(1.942,\ 19.725)$.

15. $\overline{X} = 7.909091$, $s_X = 0.359039$, $n_X = 11$, $\overline{Y} = 8.00000$, $s_Y = 0.154919$, $n_Y = 6$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{0.359039^2}{11} + \dfrac{0.154919^2}{6}\right]^2}{\dfrac{(0.359039^2/11)^2}{11-1} + \dfrac{(0.154919^2/6)^2}{6-1}} = 14, \text{ rounded down to the nearest integer.}$$

$t_{14,.01} = 2.624$, so the confidence interval is $7.909091 - 8.00000 \pm 2.624\sqrt{\dfrac{0.359039^2}{11} + \dfrac{0.154919^2}{6}}$, or $(-0.420, 0.238)$.

17. This requires a test for the difference between two means. The data are unpaired. Let μ_1 represent the population mean annual cost for cars using regular fuel, and let μ_2 represent the population mean annual cost for cars using premium fuel. Then the appropriate null and alternate hypotheses are $H_0\colon \mu_1 - \mu_2 \geq 0$ versus $H_1\colon \mu_1 - \mu_2 < 0$. The test statistic is the difference between the sample mean costs between the two groups. The z table should be used to find the P-value.

19. $\overline{X} = 1.5$, $s_X = 0.25$, $n_X = 7$, $\overline{Y} = 1.0$, $s_Y = 0.15$, $n_Y = 5$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{0.25^2}{7} + \dfrac{0.15^2}{5}\right]^2}{\dfrac{(0.25^2/7)^2}{7-1} + \dfrac{(0.15^2/5)^2}{5-1}} = 9, \text{ rounded down to the nearest integer.}$$

$t_{9,.005} = 3.250$, so the confidence interval is $1.5 - 1.0 \pm 3.250\sqrt{\dfrac{0.25^2}{7} + \dfrac{0.15^2}{5}}$, or $(0.1234, 0.8766)$.

21. The differences are $-7, -21, 4, -16, 2, -9, -20, -13$.

$\overline{D} = -10$, $s_D = 9.3808$, $n = 8$. There are $8 - 1 = 7$ degrees of freedom.

The null and alternate hypotheses are $H_0\colon \mu_D = 0$ versus $H_1\colon \mu_D \neq 0$.

$t = (-10 - 0)/(9.3808/\sqrt{8}) = -3.015$.

Since the alternate hypothesis is of the form $\mu_D \neq \Delta$, the P-value is the sum of the areas to the right of $t = 3.015$ and to the left of $t = -3.015$.

From the t table, $0.01 < P < 0.02$. A computer package gives $P = 0.0195$.

We can conclude that the mean amount of corrosion differs between the two formulations.

23. (a) Let μ_A be the mean thrust/weight ratio for fuel A and let μ_B be the mean thrust/weight ratio for fuel B. The null and alternate hypotheses are

$H_0\colon \mu_A - \mu_B \leq 0$ versus $H_1\colon \mu_A - \mu_B > 0$.

(b) $\overline{X} = 54.919$, $s_X = 2.5522$, $n_X = 16$, $\overline{Y} = 53.019$, $s_Y = 2.7294$, $n_Y = 16$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{2.5522^2}{16} + \dfrac{2.7294^2}{16}\right]^2}{\dfrac{(2.5522^2/16)^2}{16-1} + \dfrac{(2.7294^2/16)^2}{16-1}} = 29, \text{ rounded down to the nearest integer.}$$

$t_{29} = (54.919 - 53.019 - 0)/\sqrt{2.5522^2/16 + 2.7294^2/16} = 2.0339$.

The null and alternate hypotheses are $H_0\colon \mu_X - \mu_Y \leq 0$ versus $H_1\colon \mu_X - \mu_Y > 0$.

Since the alternate hypothesis is of the form $\mu_X - \mu_Y > \Delta$, the P-value is the area to the right of $t = 2.0339$.

From the t table, $0.025 < P < 0.05$. A computer package gives $P = 0.0256$.

We can conclude that the mean thrust/weight ratio is greater for fuel A than for fuel B.

Chapter 8

Section 8.1

1. (a) $\overline{x} = 65.0$, $\overline{y} = 29.05$, $\sum_{i=1}^{n}(x_i - \overline{x})^2 = 6032.0$, $\sum_{i=1}^{n}(y_i - \overline{y})^2 = 835.42$, $\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = 1988.4$, $n = 12$.

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = 0.329642 \text{ and } \widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1\overline{x} = 7.623276.$$

(b) $r^2 = \dfrac{[\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})]^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2} = 0.784587.$ $\quad s^2 = \dfrac{(1 - r^2)\sum_{i=1}^{n}(y_i - \overline{y})^2}{n - 2} = 17.996003.$

(c) $s = \sqrt{17.996003} = 4.242170.$ $\quad s_{\widehat{\beta}_0} = s\sqrt{\dfrac{1}{n} + \dfrac{\overline{x}^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}} = 3.755613.$

$s_{\widehat{\beta}_1} = \dfrac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} = 0.0546207.$

There are $n - 2 = 10$ degrees of freedom. $t_{10,.025} = 2.228$.

Therefore a 95% confidence interval for β_0 is $7.623276 \pm 2.228(3.755613)$, or $(-0.744, 15.991)$.

The 95% confidence interval for β_1 is $0.329642 \pm 2.228(0.0546207)$, or (0.208, 0.451).

(d) $\widehat{\beta}_1 = 0.329642$, $s_{\widehat{\beta}_1} = 0.0546207$, $n = 12$. There are $12 - 2 = 10$ degrees of freedom.

The null and alternate hypotheses are $H_0\colon \beta_1 \geq 0.5$ versus $H_1\colon \beta_1 < 0.5$.

$t = (0.329642 - 0.5)/0.0546207 = -3.119$.

Since the alternate hypothesis is of the form $\beta_1 < b$, the P-value is the area to the left of $t = -3.119$.

From the t table, $0.005 < P < 0.01$. A computer package gives $P = 0.00545$.

We can conclude that the claim is false.

(e) $x = 40$, $\widehat{y} = 7.623276 + 0.329642(40) = 20.808952$.

$s_{\widehat{y}} = s\sqrt{\dfrac{1}{n} + \dfrac{(x - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}} = 1.834204$. There are 10 degrees of freedom. $t_{10,.025} = 2.228$.

Therefore a 95% confidence interval for the mean response is $20.808952 \pm 2.228(1.834204)$, or (16.722, 24.896).

(f) $x = 40$, $\widehat{y} = 7.623276 + 0.329642(40) = 20.808952$.

$s_{\text{pred}} = s\sqrt{1 + \dfrac{1}{n} + \dfrac{(x - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}} = 4.621721$. There are 10 degrees of freedom. $t_{10,.025} = 2.228$.

Therefore a 95% prediction interval is $20.808952 \pm 2.228(4.621721)$, or (10.512, 31.106).

3. (a) The slope is -0.7524; the intercept is 88.761.

   (b) Yes, the P-value for the slope is ≈ 0, so ozone level is related to humidity.

   (c) $88.761 - 0.7524(50) = 51.141$ ppb.

   (d) Since $\widehat{\beta}_1 < 0$, $r < 0$. So $r = -\sqrt{r^2} = -\sqrt{0.220} = -0.469$.

   (e) Since $n = 120$ is large, use the z table to construct a confidence interval.
   $z_{.05} = 1.645$, so a 90% confidence interval is $43.62 \pm 1.645(1.20)$, or $(41.65, 45.59)$.

   (f) No. A reasonable range of predicted values is given by the 95% prediction interval, which is $(20.86, 66.37)$.

5. (a) $H_0 : \beta_A - \beta_B = 0$

   (b) $\widehat{\beta}_A$ and $\widehat{\beta}_B$ are independent and normally distributed with means β_A and β_B, respectively, and estimated standard deviations $s_{\widehat{\beta}_A} = 0.13024$ and $s_{\widehat{\beta}_B} = 0.03798$.
   Since $n = 120$ is large, the estimated standard deviations can be treated as good approximations to the true standard deviations.
   $\widehat{\beta}_A = -0.7524$ and $\widehat{\beta}_B = -0.13468$.
   The test statistic is $z = (\widehat{\beta}_A - \widehat{\beta}_B)/\sqrt{s^2_{\widehat{\beta}_A} + s^2_{\widehat{\beta}_B}} = -4.55$.
   Since the alternate hypothesis is of the form $\beta_A - \beta_B \neq 0$, the P-value is the sum of the areas to the right of $z = 4.55$ and to the left of $z = -4.55$.
   Thus $P \approx 0 + 0 = 0$.
   We can conclude that the effect of humidity differs between the two cities.

7. (a) $\overline{x} = 20.888889$, $\overline{y} = 62.888889$, $\sum_{i=1}^n (x_i - \overline{x})^2 = 1036.888889$, $\sum_{i=1}^n (y_i - \overline{y})^2 = 524.888889$, $\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = -515.111111$, $n = 9$.
   $\widehat{\beta}_1 = \dfrac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = -0.496785$ and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} = 73.266181$.
   The least-squares line is $y = 73.266181 - 0.496785x$

   (b) $r^2 = \dfrac{[\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})]^2}{\sum_{i=1}^n (x_i - \overline{x})^2 \sum_{i=1}^n (y_i - \overline{y})^2} = 0.487531$, $\quad s = \sqrt{\dfrac{(1 - r^2)\sum_{i=1}^n (y_i - \overline{y})^2}{n - 2}} = 6.198955$,
   $s_{\widehat{\beta}_0} = s\sqrt{\dfrac{1}{n} + \dfrac{\overline{x}^2}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 4.521130$, $\quad s_{\widehat{\beta}_1} = \dfrac{s}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.192510$.
   There are $9 - 2 = 7$ degrees of freedom. $t_{7,.025} = 2.365$.

Therefore a 95% confidence interval for β_0 is $73.266181 \pm 2.365(4.521130)$, or $(62.57, 83.96)$.

The 95% confidence interval for β_1 is $-0.496785 \pm 2.365(0.192510)$, or $(-0.952, -0.0415)$.

(c) $\widehat{y} = 60.846550$, $s_{\text{pred}} = s\sqrt{1 + \dfrac{1}{n} + \dfrac{(x-\overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 6.582026$.

$t_{7,.025} = 2.365$. A 95% prediction interval is $60.846550 \pm 2.365(6.582026)$ or $(45.28, 76.41)$.

(d) The standard error of prediction is $s_{\text{pred}} = s\sqrt{1 + \dfrac{1}{n} + \dfrac{(x-\overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}}$.

Given two values for x, the one that is farther from $\overline{x}$ will have the greater value of s_{pred}, and thus the wider prediction interval. Since $\overline{x} = 20.888889$, the prediction interval for $x = 30$ will be wider than the one for $x = 15$.

9. (a) $\overline{x} = 21.5075$ $\quad \overline{y} = 4.48$, $\quad \sum_{i=1}^n (x_i - \overline{x})^2 = 1072.52775$, $\quad \sum_{i=1}^n (y_i - \overline{y})^2 = 112.624$, $\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = 239.656$, $\quad n = 40$.

$\widehat{\beta}_1 = \dfrac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = 0.223450$ and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} = -0.325844$.

The least-squares line is $y = -0.325844 + 0.223450x$

(b) $r^2 = \dfrac{[\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})]^2}{\sum_{i=1}^n (x_i - \overline{x})^2 \sum_{i=1}^n (y_i - \overline{y})^2} = 0.475485$, $\quad s = \sqrt{\dfrac{(1-r^2)\sum_{i=1}^n (y_i - \overline{y})^2}{n-2}} = 1.246816$,

$s_{\widehat{\beta}_0} = s\sqrt{\dfrac{1}{n} + \dfrac{\overline{x}^2}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.842217$, $\quad s_{\widehat{\beta}_1} = \dfrac{s}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.0380713$.

There are $40 - 2 = 38$ degrees of freedom. $t_{38,.025} \approx 2.024$.

Therefore a 95% confidence interval for β_0 is $-0.325844 \pm 2.024(0.842217)$, or $(-2.031, 1.379)$.

The 95% confidence interval for β_1 is $0.223450 \pm 2.024(0.0380713)$, or $(0.146, 0.301)$.

(c) The prediction is $\widehat{\beta}_0 + \widehat{\beta}_1(20) = -0.325844 + 0.223450(20) = 4.143150$.

(d) $\widehat{y} = 4.14350$, $s_{\widehat{y}} = s\sqrt{\dfrac{1}{n} + \dfrac{(x-\overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.205323$.

$t_{38,.025} \approx 2.024$. A 95% confidence interval is $4.14350 \pm 2.024(0.205323)$ or $(3.727, 4.559)$.

(e) $\widehat{y} = 4.14350$, $s_{\text{pred}} = s\sqrt{1 + \dfrac{1}{n} + \dfrac{(x-\overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 1.263609$.

$t_{38,.025} \approx 2.024$. A 95% prediction interval is $4.14350 \pm 2.024(1.263609)$ or $(1.585, 6.701)$.

11. The width of a confidence interval is proportional to $s_{\widehat{y}} = s\sqrt{\dfrac{1}{n} + \dfrac{(x-\overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}}$.

Since s, n, $\overline{x}$, and $\sum_{i=1}^n (x_i - \overline{x})^2$ are the same for each confidence interval, the width of the confidence interval is an increasing function of the difference $x - \overline{x}$.

$\overline{x} = 1.51966$. The value 1.5 is closest to $\overline{x}$ and the value 1.8 is the farthest from $\overline{x}$.

Therefore the confidence interval at 1.5 would be the shortest, and the confidence interval at 1.8 would be the longest.

13. (a) $t = 1.71348/6.69327 = 0.256$.

(b) $n = 25$, so there are $n - 2 = 23$ degrees of freedom. The P-value is for a two-tailed test, so it is equal to the sum of the areas to the right of $t = 0.256$ and to the left of $t = -0.256$.

Thus $P = 0.40 + 0.40 = 0.80$.

(c) $s_{\widehat{\beta}_1}$ satisfies the equation $3.768 = 4.27473/s_{\widehat{\beta}_1}$, so $s_{\widehat{\beta}_1} = 1.13448$.

(d) $n = 25$, so there are $n - 2 = 23$ degrees of freedom. The P-value is for a two-tailed test, so it is equal to the sum of the areas to the right of $t = 3.768$ and to the left of $t = -3.768$.

Thus $P = 0.0005 + 0.0005 = 0.001$.

15. (a) $\widehat{y} = 106.11 + 0.1119(4000) = 553.71$.

(b) $\widehat{y} = 106.11 + 0.1119(500) = 162.06$.

(c) Below. For values of x near 500, there are more points below the least squares estimate than above it.

(d) There is a greater amount of vertical spread on the right side of the plot than on the left.

17. $r = -0.509$, $n = 23$, $U = r\sqrt{n-2}/\sqrt{1-r^2} = -2.7098$.

Under H_0, U has a Student's t distribution with $23 - 2 = 21$ degrees of freedom.

Since the alternate hypothesis is $H_1\colon \rho \neq 0$, the P-value is the sum of the areas to the right of $t = 2.7098$ and to the left of $t = -2.7098$.

From the t table, $0.01 < P < 0.02$. A computer package gives $P = 0.0131$.

We conclude that $\rho \neq 0$.

Section 8.2

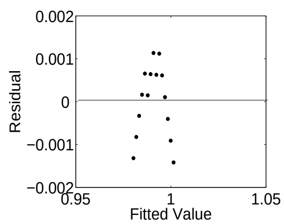
1. (a) $\ln y = -0.4442 + 0.79833 \ln x$

   (b) $\widehat{y} = e^{\widehat{\ln y}} = e^{-0.4442+0.79833(\ln 2500)} = 330.95.$

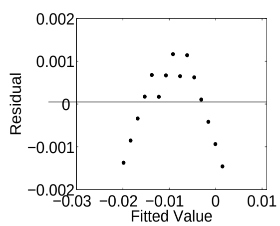
   (c) $\widehat{y} = e^{\widehat{\ln y}} = e^{-0.4442+0.79833(\ln 1600)} = 231.76.$

   (d) The 95% prediction interval for $\ln y$ is given as (3.9738, 6.9176).
   The 95% prediction interval for y is therefore $(e^{3.9738}, e^{6.9176})$, or (53.19, 1009.89).

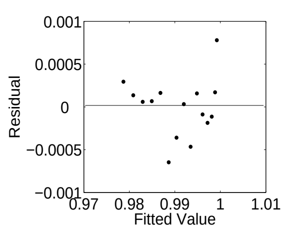
3. (a) $y = 1.0014 - 0.015071t$



   (b) $\ln y = 0.0014576 - 0.015221t$



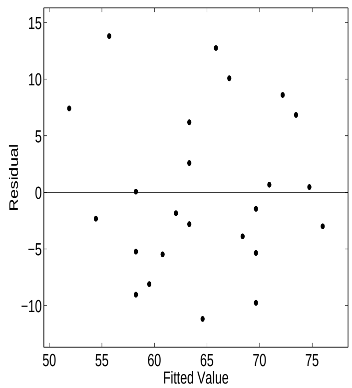
   (c) $y = 0.99922 - 0.012385t^{1.5}$



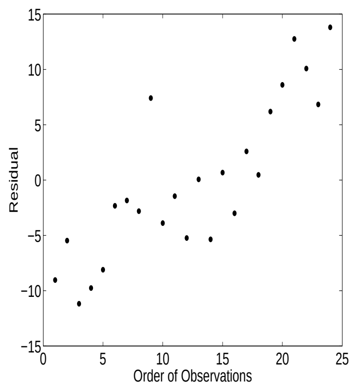
   (d) The model $y = 0.99922 - 0.012385t^{1.5}$ fits best. Its residual plot shows the least pattern.

   (e) The estimate is $y = 0.99922 - 0.012385(0.75^{1.5}) = 0.991$.

5. (a) $y = 20.162 + 1.269x$

(b) 

There is no apparent pattern to the residual plot. The linear model looks fine.

(c) 

The residuals increase over time. The linear model is not appropriate as is. Time, or other variables related to time, must be included in the model.

7. The equation becomes linear upon taking the log of both sides: $\ln W = \beta_0 + \beta_1 \ln L$, where $\beta_0 = \ln a$ and $\beta_1 = b$.

9. (a) A physical law.

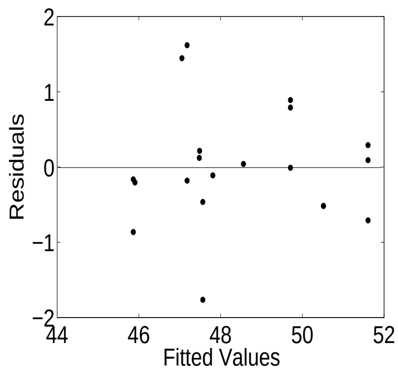
   (b) It would be better to redo the experiment. If the results of anexperiment violate a physical law, then something was wrong with the experiment, and you can't fix it by transforming variables.

Section 8.3

1. (a) The predicted strength is $26.641 + 3.3201(8.2) - 0.4249(10) = 49.62$ kg/mm^2.

(b) By $3.3201(10) = 33.201\ \text{kg/mm}^2$.

(c) By $0.4249(5) = 2.1245\ \text{kg/mm}^2$.

3. 

The linear model is reasonable. There is no obvious pattern to the plot.

5. (a) $\widehat{y} = 56.145 - 9.046(3) - 33.421(1.5) + 0.243(20) - 0.5963(3)(1.5) - 0.0394(3)(20) + 0.6022(1.5)(20) + 0.6901(3^2) + 11.7244(1.5^2) - 0.0097(20^2) = 25.465$.

(b) No, the predicted change depends on the values of the other independent variables, because of the interaction terms.

(c) $R^2 = \text{SSR}/\text{SST} = (\text{SST} - \text{SSE})/\text{SST} = (6777.5 - 209.55)/6777.5 = 0.9691$.

(d) There are 9 degrees of freedom for regression and $27 - 9 - 1 = 17$ degrees of freedom for error.

$$F_{9,17} = \frac{\text{SSR}/9}{\text{SSE}/17} = \frac{(\text{SST} - \text{SSE})/9}{\text{SSE}/17} = 59.204.$$

From the F table, $P < 0.001$. A computer package gives $P = 4.6 \times 10^{-11}$.

The null hypothesis can be rejected.

7. (a) $\widehat{y} = -0.21947 + 0.779(2.113) - 0.10827(0) + 1.3536(1.4) - 0.0013431(730) = 2.3411$ liters

(b) By $1.3536(0.05) = 0.06768$ liters

(c) Nothing is wrong. In theory, the constant estimates FEV_1 for an individual whose values for the other variables are all equal to zero. Since these values are outside the range of the data (e.g., no one has zero height), the constant need not represent a realistic value for an actual person.

9. (a) $\widehat{y} = -1.7914 + 0.00026626(1500) + 9.8184(1.04) - 0.29982(17.5) = 3.572$.

   (b) By $9.8184(0.01) = 0.098184$.

   (c) Nothing is wrong. The constant estimates the pH for a pulp whose values for the other variables are all equal to zero. Since these values are outside the range of the data (e.g., no pulp has zero density), the constant need not represent a realistic value for an actual pulp.

   (d) From the output, the confidence interval is (3.4207, 4.0496).

   (e) From the output, the prediction interval is (2.2333, 3.9416).

   (f) Pulp B. The standard deviation of its predicted pH (SE Fit) is smaller than that of Pulp A (0.1351 vs. 0.2510).

11. (a) $t = -0.58762/0.2873 = -2.05$.

   (b) $s_{\widehat{\beta}_1}$ satisfies the equation $4.30 = 1.5102/s_{\widehat{\beta}_1}$, so $s_{\widehat{\beta}_1} = 0.3512$.

   (c) $\widehat{\beta}_2$ satisfies the equation $-0.62 = \widehat{\beta}_2/0.3944$, so $\widehat{\beta}_2 = -0.2445$.

   (d) $t = 1.8233/0.3867 = 4.72$.

   (e) $\text{MSR} = 41.76/3 = 13.92$.

   (f) $F = \text{MSR}/\text{MSE} = 13.92/0.76 = 18.316$.

   (g) $\text{SSE} = 46.30 - 41.76 = 4.54$.

   (h) $3 + 6 = 9$.

13. (a) $\widehat{y} = 267.53 - 1.5926(30) - 1.3897(35) - 1.0934(30) - 0.002658(30)(30) = 135.92°\text{F}$

   (b) No. The change in the predicted flash point due to a change in acetic acid concentration depends on the butyric acid concentration as well, because of the interaction between these two variables.

   (c) Yes. The predicted flash point will change by $-1.3897(10) = -13.897°\text{F}$.

15. (a) The residuals are the values $e_i = y_i - \widehat{y}_i$ for each i. They are shown in the following table.

x	y	Fitted Value $\widehat{y}$	Residual $e = y - \widehat{y}$
150	10.4	10.17143	0.22857
175	12.4	12.97429	−0.57429
200	14.9	14.54858	0.35142
225	15	14.89429	0.10571
250	13.9	14.01143	−0.11144
275	11.9	11.90001	−0.00001

(b) $\text{SSE} = \sum_{i=1}^{6} e_i^2 = 0.52914$, $\quad \text{SST} = \sum_{i=1}^{n} (y_i - \overline{y})^2 = 16.70833$.

(c) $s^2 = \text{SSE}/(n-3) = 0.17638$

(d) $R^2 = 1 - \text{SSE}/\text{SST} = 0.96833$

(e) $F = \dfrac{\text{SSR}/2}{\text{SSE}/3} = \dfrac{(\text{SST} - \text{SSE})/2}{\text{SSE}/3} = 45.864$. There are 2 and 3 degrees of freedom.

(f) Yes. From the F table, $0.001 < P < 0.01$. A computer package gives $P = 0.0056$. Since $P \leq 0.05$, the hypothesis $H_0\colon \beta_1 = \beta_2 = 0$ can be rejected at the 5% level.

17. (a) $\widehat{y} = 1.18957 + 0.17326(0.5) + 0.17918(5.7) + 0.17591(3.0) - 0.18393(4.1) = 2.0711$.

(b) 0.17918

(c) PP is more useful, because its P-value is small, while the P-value of CP is fairly large.

(d) The percent change in GDP would be expected to be larger in Sweden, because the coefficient of PP is negative.

19. (a)

Predictor	Coef	StDev	T	P
Constant	−0.012167	0.01034	−1.1766	0.278
Time	0.043258	0.043186	1.0017	0.350
Time^2	2.9205	0.038261	76.33	0.000

$y = -0.012167 + 0.043258t + 2.9205t^2$

(b) $\widehat{\beta}_2 = 2.9205$, $\quad s_{\widehat{\beta}_2} = 0.038261$. There are $n - 3 = 7$ degrees of freedom.

$t_{7,.025} = 2.365$. A 95% confidence interval is therefore $2.9205 \pm 2.365(0.038261)$, or (2.830, 3.011).

(c) Since $a = 2\beta_2$, the confidence limits for a 95% confidence interval for a are twice the limits of the confidence interval for β_2. Therefore a 95% confidence interval for a is (5.660, 6.022).

(d) $\widehat{\beta}_0$: $t_7 = -1.1766$, $P = 0.278$, $\widehat{\beta}_1$: $t_7 = 1.0017$, $P = 0.350$, $\widehat{\beta}_2$: $t_7 = 76.33$, $P = 0.000$.

(e) No, the P-value of 0.278 is not small enough to reject the null hypothesis that $\beta_0 = 0$.

(f) No, the P-value of 0.350 is not small enough to reject the null hypothesis that $\beta_1 = 0$.

Section 8.4

1. (a) False. There are usually several models that are about equally good.

   (b) True.

   (c) False. Model selection methods can suggest models that fit the data well.

   (d) True.

3. (iv). Carbon and Silicon both have large P-values and thus may not contribute significantly to the fit.

5. The four-variable model with the highest value of R^2 has a lower R^2 than the three-variable model with the highest value of R^2. This is impossible.

7. (a) $\text{SSE}_{\text{full}} = 7.7302, \quad \text{SSE}_{\text{reduced}} = 7.7716, \quad n = 165, \quad p = 7, \quad k = 4.$

$$F = \frac{(\text{SSE}_{\text{reduced}} - \text{SSE}_{\text{full}})/(p-k)}{\text{SSE}_{\text{full}}/(n-p-1)} = 0.2803.$$

   (b) 3 degrees of freedom in the numerator and 157 in the denominator.

   (c) $P > 0.10$ (a computer package gives $P = 0.840$). The reduced model is plausible.

(d) This is not correct. It is possible for a group of variables to be fairly strongly related to an independent variable, even though none of the variables individually is strongly related.

(e) No mistake. If y is the dependent variable, then the total sum of squares is $\sum(y_i - \overline{y})^2$. This quantity does not involve the independent variables.

9. (a)

Predictor	Coef	StDev	T	P
Constant	25.613	10.424	2.4572	0.044
x_1	0.18387	0.12353	1.4885	0.180
x_2	−0.015878	0.0040542	−3.9164	0.006

(b)

Predictor	Coef	StDev	T	P
Constant	14.444	16.754	0.86215	0.414
x_1	0.17334	0.20637	0.83993	0.425

(c)

Predictor	Coef	StDev	T	P
Constant	40.370	3.4545	11.686	0.000
x_2	−0.015747	0.0043503	−3.6197	0.007

(d) The model containing x_2 as the only independent variable is best. There is no evidence that the coefficient of x_1 differs from 0.

11. The model $y = \beta_0 + \beta_1 x_2 + \varepsilon$ is a good one. One way to see this is to compare the fit of this model to the full quadratic model. The ANOVA table for the full model is

Source	DF	SS	MS	F	P
Regression	5	4.1007	0.82013	1.881	0.193
Residual Error	9	3.9241	0.43601		
Total	14	8.0248			

The ANOVA table for the model $y = \beta_0 + \beta_1 x_2 + \varepsilon$ is

Source	DF	SS	MS	F	P
Regression	1	2.7636	2.7636	6.8285	0.021
Residual Error	13	5.2612	0.40471		
Total	14	8.0248			

From these two tables, the F statistic for testing the plausibility of the reduced model is

$$\frac{(5.2612 - 3.9241)/(5-1)}{3.9241/9} = 0.7667.$$

The null distribution is $F_{4,9}$, so $P > 0.10$ (a computer package gives $P = 0.573$). The large P-value indicates that the reduced model is plausible.

Supplementary Exercises for Chapter 8

1. (a) $\overline{x} = 18.142857$, $\overline{y} = 0.175143$, $\sum_{i=1}^n (x_i - \overline{x})^2 = 418.214286$, $\sum_{i=1}^n (y_i - \overline{y})^2 = 0.0829362$, $\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = 3.080714$, $n = 14$.

$\widehat{\beta}_1 = \dfrac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = 0.00736635$ and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} = 0.0414962$.

$y = 0.0414962 + 0.00736635x$

(b) $r^2 = \dfrac{[\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})]^2}{\sum_{i=1}^n (x_i - \overline{x})^2 \sum_{i=1}^n (y_i - \overline{y})^2} = 0.273628$, $s = \sqrt{\dfrac{(1 - r^2)\sum_{i=1}^n (y_i - \overline{y})^2}{n-2}} = 0.0708535$.

$\widehat{\beta}_1 = 0.00736635$, $s_{\widehat{\beta}_1} = \dfrac{s}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.00346467$.

There are $14 - 2 = 12$ degrees of freedom. $t_{12,.025} = 2.179$.

Therefore a 95% confidence interval for the slope is $0.00736635 \pm 2.179(0.00346467)$, or $(-0.00018, 0.01492)$.

(c) $x = 20$, $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1(20) = 0.188823$.

$s_{\widehat{y}} = s\sqrt{\dfrac{1}{n} + \dfrac{(x - \overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.0199997$.

There are 12 degrees of freedom. $t_{12,.025} = 2.179$.

Therefore a 95% confidence interval for the mean response is $0.188823 \pm 2.179(0.0199997)$, or $(0.145, 0.232)$.

(d) $x = 20$, $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1(20) = 0.188823$.

$s_{\text{pred}} = s\sqrt{1 + \dfrac{1}{n} + \dfrac{(x - \overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.073622$.

There are 12 degrees of freedom. $t_{12,.05} = 1.782$.

Therefore a 90% prediction interval is $0.188823 \pm 1.782(0.073622)$, or $(0.0576, 0.320)$.

3. (a) $\ln y = \beta_0 + \beta_1 \ln x$, where $\beta_0 = \ln k$ and $\beta_1 = r$.

(b) Let $u_i = \ln x_i$ and let $v_i = \ln y_i$.

$\overline{u} = 1.755803$, $\overline{v} = -0.563989$, $\sum_{i=1}^n (u_i - \overline{u})^2 = 0.376685$, $\sum_{i=1}^n (v_i - \overline{v})^2 = 0.160487$, $\sum_{i=1}^n (u_i - \overline{u})(v_i - \overline{v}) = 0.244969$, $n = 5$.

$\widehat{\beta}_1 = \dfrac{\sum_{i=1}^n (u_i - \overline{u})(v_i - \overline{v})}{\sum_{i=1}^n (u_i - \overline{u})^2} = 0.650328$ and $\widehat{\beta}_0 = \overline{v} - \widehat{\beta}_1 \overline{u} = -1.705838$.

The least-squares line is $\ln y = -1.705838 + 0.650328 \ln x$.

Therefore $\widehat{r} = 0.650328$ and $\widehat{k} = e^{-1.705838} = 0.18162$.

(c) The null and alternate hypotheses are $H_0\colon r = 0.5$ versus $H_1\colon r \neq 0.5$.

$\widehat{r} = 0.650328$, $s = 0.0198005$, $s_{\widehat{r}} = \dfrac{s}{\sqrt{\sum_{i=1}^{n}(u_i - \overline{u})^2}} = 0.0322616$.

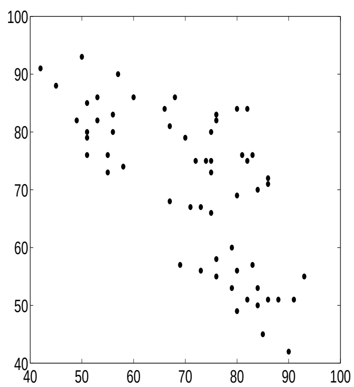
There are $5 - 2 = 3$ degrees of freedom. $t = (0.650328 - 0.5)/0.0322616 = 4.660$.

Since the alternate hypothesis is of the form $r \neq r_0$, the P-value is the sum of the areas to the right of $t = 4.660$ and to the left of $t = -4.660$.

From the t table, $0.01 < P < 0.02$. A computer package gives $P = 0.019$.

We can conclude that $r \neq 0.5$.

5. (a)



(b) $\overline{x} = 71.101695$, $\overline{y} = 70.711864$, $\sum_{i=1}^{n}(x_i - \overline{x})^2 = 10505.389831$, $\sum_{i=1}^{n}(y_i - \overline{y})^2 = 10616.101695$, $\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = -7308.271186$, $n = 59$.

$\widehat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = -0.695669$ and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1\overline{x} = 120.175090$.

$T_{i+1} = 120.175090 - 0.695669T_i$.

(c) $r^2 = \dfrac{[\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})]^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2} = 0.478908$, $s = \sqrt{\dfrac{(1 - r^2)\sum_{i=1}^{n}(y_i - \overline{y})^2}{n - 2}} = 9.851499$,

$s_{\widehat{\beta}_1} = \dfrac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} = 0.096116$.

There are $n - 2 = 57$ degrees of freedom. $t_{57,.025} \approx 2.002$.

Therefore a 95% confidence interval for β_1 is $-0.695669 \pm 2.002(0.096116)$, or $(-0.888, -0.503)$.

(d) $\widehat{y} = 120.175090 - 0.695669(70) = 71.4783$ minutes.

(e) $\widehat{y} = 71.4783$, $s_{\widehat{y}} = s\sqrt{\dfrac{1}{n} + \dfrac{(x-\overline{x})^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}} = 1.286920$.

There are $59 - 2 = 57$ degrees of freedom. $t_{57,.01} \approx 2.394$.

Therefore a 98% confidence interval is $71.4783 \pm 2.394(1.286920)$, or (68.40, 74.56).

(f) $\widehat{y} = 71.4783$, $s_{\text{pred}} = s\sqrt{1 + \dfrac{1}{n} + \dfrac{(x-\overline{x})^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}} = 9.935200$.

$t_{57,.005} \approx 2.6649$. A 95% prediction interval is $71.4783 \pm 2.6649(9.935200)$ or (45.00, 97.95).

7. (a) $\overline{x} = 50$, $\overline{y} = 47.909091$, $\sum_{i=1}^{n}(x_i-\overline{x})^2 = 11000$, $\sum_{i=1}^{n}(y_i-\overline{y})^2 = 9768.909091$,
$\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y}) = 10360$, $n = 11$.

$\widehat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2} = 0.941818$ and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1\overline{x} = 0.818182$.

(b) $r^2 = \dfrac{[\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})]^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2\sum_{i=1}^{n}(y_i-\overline{y})^2} = 0.998805$, $s = \sqrt{\dfrac{(1-r^2)\sum_{i=1}^{n}(y_i-\overline{y})^2}{n-2}} = 1.138846$.

$\widehat{\beta}_0 = 0.818182$, $s_{\widehat{\beta}_0} = s\sqrt{\dfrac{1}{n} + \dfrac{\overline{x}^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}} = 0.642396$.

The null and alternate hypotheses are $H_0\colon \beta_0 = 0$ versus $H_1\colon \beta_0 \neq 0$.

There are $11 - 2 = 9$ degrees of freedom. $t = (0.818182 - 0)/0.642396 = 1.274$.

Since the alternate hypothesis is of the form $\beta_0 \neq b$, the P-value is the sum of the areas to the right of $t = 1.274$ and to the left of $t = -1.274$.

From the t table, $0.20 < P < 0.50$. A computer package gives $P = 0.235$.

It is plausible that $\beta_0 = 0$.

(c) The null and alternate hypotheses are $H_0\colon \beta_1 = 1$ versus $H_1\colon \beta_1 \neq 1$.

$\widehat{\beta}_1 = 0.941818$, $s_{\widehat{\beta}_1} = \dfrac{s}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}} = 0.010858$.

There are $11 - 2 = 9$ degrees of freedom. $t = (0.941818 - 1)/0.010858 = -5.358$.

Since the alternate hypothesis is of the form $\beta_1 \neq b$, the P-value is the sum of the areas to the right of $t = 5.358$ and to the left of $t = -5.358$.

From the t table, $P < 0.001$. A computer package gives $P = 0.00046$.

We can conclude that $\beta_1 \neq 1$.

(d) Yes, since we can conclude that $\beta_1 \neq 1$, we can conclude that the machine is out of calibration.

Since two coefficients were tested, some may wish to apply the Bonferroni correction, and multiply the P-value for β_1 by 2. The evidence that $\beta_1 \neq 1$ is still conclusive.

(e) $x = 20$, $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1(20) = 19.65455$.

$s_{\widehat{y}} = s\sqrt{\dfrac{1}{n} + \dfrac{(x-\overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.47331$. There are 9 degrees of freedom. $t_{9,.025} = 2.262$.

Therefore a 95% confidence interval for the mean response is $19.65455 \pm 2.262(0.47331)$, or (18.58, 20.73).

(f) $x = 80$, $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1(80) = 76.163636$.

$s_{\widehat{y}} = s\sqrt{\dfrac{1}{n} + \dfrac{(x-\overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.47331$. There are 9 degrees of freedom. $t_{9,.025} = 2.262$.

Therefore a 95% confidence interval for the mean response is $76.163636 \pm 2.262(0.47331)$, or (75.09, 77.23).

(g) No, when the true value is 20, the result of part (e) shows that a 95% confidence interval for the mean of the measured values is (18.58, 20.73). Therefore it is plausible that the mean measurement will be 20, so that the machine is in calibration.

9. (ii). The standard deviation $s_{\widehat{y}}$ is not given in the output. To compute $s_{\widehat{y}}$, the quantity $\sum_{i=1}^n (x_i - \overline{x})^2$ must be known.

11. (a) If $f = 1/2$ then $1/f = 2$. The estimate is $\widehat{t} = 145.736 - 0.05180(2) = 145.63$.

(b) Yes. $r = -\sqrt{\text{R-Sq}} = -0.988$. Note that r is negative because the slope of the least-squares line is negative.

(c) If $f = 1$ then $1/f = 1$. The estimate is $\widehat{t} = 145.736 - 0.05180(1) = 145.68$.

13. (a) $\widehat{y} = 46.802 - 130.11(0.15) - 807.10(0.01) + 3580.5(0.15)(0.01) = 24.6\%$.

(b) By $130.11(0.05) - 3580.5(0.006)(0.05) = 5.43\%$.

(c) No, we need to know the oxygen content, because of the interaction term.

15. (a) $\widehat{\beta}_0$ satisfies the equation $0.59 = \widehat{\beta}_0/0.3501$, so $\widehat{\beta}_0 = 0.207$.

(b) $s_{\widehat{\beta}_1}$ satisfies the equation $2.31 = 1.8515/s_{\widehat{\beta}_1}$, so $s_{\widehat{\beta}_1} = 0.8015$.

(c) $t = 2.7241/0.7124 = 3.82$.

(d) $s = \sqrt{\text{MSE}} = \sqrt{1.44} = 1.200$.

(e) There are 2 independent variables in the model, so there are 2 degrees of freedom for regression.

(f) SSR = SST − SSE = 104.09 − 17.28 = 86.81.

(g) MSR = 86.81/2 = 43.405.

(h) F = MSR/MSE = 43.405/1.44 = 30.14.

(i) $2 + 12 = 14$.

17. (a)

Predictor	Coef	StDev	T	P
Constant	10.84	0.2749	39.432	0.000
Speed	−0.073851	0.023379	−3.1589	0.004
Pause	−0.12743	0.013934	−9.1456	0.000
Speed^2	0.0011098	0.00048887	2.2702	0.032
Pause^2	0.0016736	0.00024304	6.8861	0.000
Speed·Pause	−0.00024272	0.00027719	−0.87563	0.390

S = 0.33205 R-sq = 92.2% R-sq(adj) = 90.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	31.304	6.2608	56.783	0.000
Residual Error	24	2.6462	0.11026		
Total	29	33.95			

(b) We drop the interaction term Speed·Pause.

Predictor	Coef	StDev	T	P
Constant	10.967	0.23213	47.246	0.000
Speed	−0.079919	0.022223	−3.5961	0.001
Pause	−0.13253	0.01260	−10.518	0.000
Speed^2	0.0011098	0.00048658	2.2809	0.031
Pause^2	0.0016736	0.0002419	6.9185	0.000

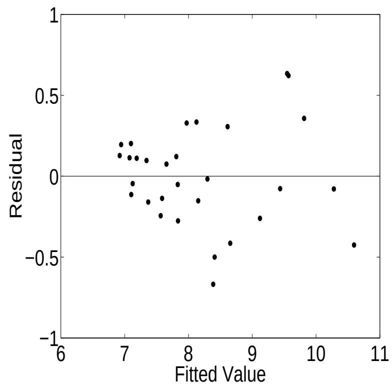
S = 0.33050 R-sq = 92.0% R-sq(adj) = 90.7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	31.22	7.8049	71.454	0.000
Residual Error	25	2.7307	0.10923		
Total	29	33.95			

Comparing this model with the one in part (a), $F_{1,24} = \dfrac{(2.7307 - 2.6462)/(5-4)}{2.6462/24} = 0.77$, $P > 0.10$. A computer package gives $P = 0.390$ (the same as the P-value for the dropped variable).

(c)



There is a some suggestion of heteroscedasticity, but it is hard to be sure without more data.

(d)

Predictor	Coef	StDev	T	P
Constant	9.9601	0.21842	45.601	0.000
Pause	-0.13253	0.020545	-6.4507	0.000
Pause^2	0.0016736	0.00039442	4.2431	0.000

S = 0.53888 R-sq = 76.9% R-sq(adj) = 75.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	26.11	13.055	44.957	0.000
Residual Error	27	7.8405	0.29039		
Total	29	33.95			

Comparing this model with the one in part (a), $F_{3,24} = \dfrac{(7.8405 - 2.6462)/(5-2)}{2.6462/24} = 15.70$, $P < 0.001$. A computer package gives $P = 7.3 \times 10^{-6}$.

(e)

Vars	R-Sq	R-Sq(adj)	C-p	S	Speed	Pause	Speed2	Pause2	Speed*Pause
1	61.5	60.1	92.5	0.68318		X			
1	60.0	58.6	97.0	0.69600					X
2	76.9	75.2	47.1	0.53888		X		X	
2	74.9	73.0	53.3	0.56198	X	X			
3	90.3	89.2	7.9	0.35621	X	X		X	
3	87.8	86.4	15.5	0.39903		X		X	X
4	92.0	90.7	4.8	0.33050	X	X	X	X	
4	90.5	89.0	9.2	0.35858	X	X		X	X
5	92.2	90.6	6.0	0.33205	X	X	X	X	X

(f) The model containing the dependent variables Speed, Pause, Speed^2 and Pause^2 has both the lowest value of C_p and the largest value of adjusted R^2.

19. (a) Linear Model

Predictor	Coef	StDev	T	P
Constant	40.751	5.4533	7.4728	0.000
Hardwood	0.54013	0.61141	0.88341	0.389

S = 12.308 R-sq = 4.2% R-sq(adj) = −1.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	118.21	118.21	0.78041	0.38866
Residual Error	18	2726.5	151.47		
Total	19	2844.8			

Quadratic Model

Predictor	Coef	StDev	T	P
Constant	12.683	3.9388	3.2199	0.005
Hardwood	10.067	1.1028	9.1287	0.000
Hardwood^2	−0.56928	0.063974	−8.8986	0.000

S = 5.3242 R-sq = 83.1% R-sq(adj) = 81.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	2362.9	1181.4	41.678	0.000
Residual Error	17	481.90	28.347		
Total	19	2844.8			

Cubic Model

Predictor	Coef	StDev	T	P
Constant	27.937	2.9175	9.5755	0.000
Hardwood	0.48749	1.453	0.3355	0.742
Hardwood^2	0.85104	0.20165	4.2204	0.001
Hardwood^3	−0.057254	0.0080239	−7.1354	0.000

S = 2.6836 R-sq = 95.9% R-sq(adj) = 95.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	2729.5	909.84	126.34	0.000
Residual Error	16	115.23	7.2018		
Total	19	2844.8			

Quartic Model

Predictor	Coef	StDev	T	P
Constant	30.368	4.6469	6.5351	0.000
Hardwood	−1.7962	3.6697	−0.48946	0.632
Hardwood^2	1.4211	0.8632	1.6463	0.120
Hardwood^3	−0.10878	0.076229	−1.4271	0.174
Hardwood^4	0.0015256	0.0022438	0.67989	0.507

S = 2.7299 R-sq = 96.1% R-sq(adj) = 95.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	2733	683.24	91.683	0.00
Residual Error	15	111.78	7.4522		
Total	19	2844.8			

The values of SSE and their degrees of freedom for models of degrees 1, 2, 3, and 4 are:

Linear	18	2726.55
Quadratic	17	481.90
Cubic	16	115.23
Quartic	15	111.78

To compare quadratic vs. linear, $F_{1,17} = \dfrac{(2726.55 - 481.90)/(18 - 17)}{481.90/17} = 79.185$.

$P \approx 0$. A computer package gives $P = 8.3 \times 10^{-8}$.

To compare cubic vs. quadratic, $F_{1,16} = \dfrac{(481.90 - 115.23)/(17 - 16)}{115.23/16} = 50.913$.

$P \approx 0$. A computer package gives $P = 2.4 \times 10^{-6}$.

To compare quartic vs. cubic, $F_{1,15} = \dfrac{(115.23 - 111.78)/(16 - 15)}{111.78/15} = 0.463$.

$P > 0.10$. A computer package gives $P = 0.507$.

The cubic model is selected by this procedure.

(b) The cubic model is $y = 27.937 + 0.48749x + 0.85104x^2 - 0.057254x^3$. The estimate y is maximized when $dy/dx = 0$. $dy/dx = 0.48749 + 1.70208x - 0.171762x^2$. Therefore $x = 10.188$ ($x = -0.2786$ is a spurious root).

21. (a)

Predictor	Coef	StDev	T	P
Constant	-0.093765	0.092621	-1.0123	0.335
x_1	0.63318	2.2088	0.28666	0.780
x_2	2.5095	0.30151	8.3233	0.000
x_1^2	5.318	8.2231	0.64672	0.532
x_2^2	-0.3214	0.17396	-1.8475	0.094
x_1x_2	0.15209	1.5778	0.09639	0.925

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	20.349	4.0698	894.19	0.000
Residual Error	10	0.045513	0.0045513		
Total	15	20.394			

(b) The model containing the variables x_1, x_2, and x_2^2 is a good one. Here are the coefficients along with their standard deviations, followed by the analysis of variance table.

Predictor	Coef	StDev	T	P
Constant	-0.088618	0.068181	-1.2997	0.218
x_1	2.1282	0.30057	7.0805	0.000
x_2	2.4079	0.13985	17.218	0.000
x_2^2	-0.27994	0.059211	-4.7279	0.000

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	20.346	6.782	1683.9	0.000
Residual Error	12	0.048329	0.0040275		
Total	15	20.394			

The F statistic for comparing this model to the full quadratic model is

$$F_{2,10} = \frac{(0.048329 - 0.045513)/(12 - 10)}{0.045513/10} = 0.309,\ P > 0.10,$$

so it is reasonable to drop x_1^2 and x_1x_2 from the full quadratic model. All the remaining coefficients are significantly different from 0, so it would not be reasonable to reduce the model further.

(c) The output from the MINITAB best subsets procedure is

```
Response is y

                                                x x x
                                                1 2 1
                           Mallows          x x ^ ^ x
Vars  R-Sq  R-Sq(adj)        C-p         S  1 2 2 2 2
   1  98.4       98.2       61.6   0.15470    X
```

1	91.8	91.2	354.1	0.34497				X	
2	99.3	99.2	20.4	0.10316	X	X			
2	99.2	99.1	25.7	0.11182		X	X		
3	99.8	99.7	2.2	0.062169		X	X	X	
3	99.8	99.7	2.6	0.063462	X	X		X	
4	99.8	99.7	4.0	0.064354	X	X	X	X	
4	99.8	99.7	4.1	0.064588		X	X	X	X
5	99.8	99.7	6.0	0.067463	X	X	X	X	X

The model with the best adjusted R^2 (0.99716) contains the variables x_2, x_1^2, and x_2^2. This model is also the model with the smallest value of Mallows' C$_p$ (2.2). This is not the best model, since it contains x_1^2 but not x_1. The model containing x_1, x_2, and x_2^2, suggested in the answer to part (b), is better. Note that the adjusted R^2 for the model in part (b) is 0.99704, which differs negligibly from that of the model with the largest adjusted R^2 value.

23. (a)

Predictor	Coef	StDev	T	P
Constant	1.1623	0.17042	6.8201	0.006
t	0.059718	0.0088901	6.7174	0.007
t^2	−0.00027482	0.000069662	−3.9450	0.029

(b) Let x be the time at which the reaction rate will be equal to 0.05. Then $0.059718 - 2(0.00027482)x = 0.05$, so $x = 17.68$ minutes.

(c) $\widehat{\beta}_1 = 0.059718$, $\quad s_{\widehat{\beta}_1} = 0.0088901$.

There are 6 observations and 2 dependent variables, so there are $6 - 2 - 1 = 3$ degrees of freedom for error.

$t_{3,.025} = 3.182$.

A 95% confidence interval is $0.059718 \pm 3.182(0.0088901)$, or $(0.0314, 0.0880)$.

(d) The reaction rate is decreasing with time if $\beta_2 < 0$. We therefore test $H_0\colon \beta_2 \geq 0$ versus $H_1\colon \beta_2 < 0$.

From the output, the test statistic for testing $H_0\colon \beta_2 = 0$ versus $H_1\colon \beta_2 \neq 0$ is is $t = -3.945$.

The output gives $P = 0.029$, but this is the value for a two-tailed test.

For the one-tailed test, $P = 0.029/2 = 0.0145$.

It is reasonable to conclude that the reaction rate decreases with time.

25. (a) The 17-variable model containing the independent variables x_1, x_2, x_3, x_6, x_7, x_8, x_9, x_{11}, x_{13}, x_{14}, x_{16}, x_{18}, x_{19}, x_{20}, x_{21}, x_{22}, and x_{23} has adjusted R^2 equal to 0.98446. The fitted model is

$$\begin{aligned} y \quad = \quad & -1569.8 - 24.909x_1 + 196.95x_2 + 8.8669x_3 - 2.2359x_6 \\ & - 0.077581x_7 + 0.057329x_8 - 1.3057x_9 - 12.227x_{11} + 44.143x_{13} \\ & + 4.1883x_{14} + 0.97071x_{16} + 74.775x_{18} + 21.656x_{19} - 18.253x_{20} \\ & + 82.591x_{21} - 37.553x_{22} + 329.8x_{23} \end{aligned}$$

(b) The 8-variable model containing the independent variables x_1, x_2, x_5, x_8, x_{10}, x_{11}, x_{14}, and x_{21} has Mallows' C_p equal to 1.7. The fitted model is

$$y = -665.98 - 24.782x_1 + 76.499x_2 + 121.96x_5 + 0.024247x_8 + 20.4x_{10} - 7.1313x_{11} + 2.4466x_{14} + 47.85x_{21}$$

(c) Using a value of 0.15 for both α-to-enter and α-to-remove, the equation chosen by stepwise regression is $y = -927.72 + 142.40x_5 + 0.081701x_7 + 21.698x_{10} + 0.41270x_{16} + 45.672x_{21}$.

(d) The 13-variable model below has adjusted R^2 equal to 0.95402. (There are also two 12-variable models whose adjusted R^2 is only very slightly lower.)

$$\begin{aligned} z &= 8663.2 - 313.31x_3 - 14.46x_6 + 0.358x_7 - 0.078746x_8 \\ &\quad +13.998x_9 + 230.24x_{10} - 188.16x_{13} + 5.4133x_{14} + 1928.2x_{15} \\ &\quad -8.2533x_{16} + 294.94x_{19} + 129.79x_{22} - 3020.7x_{23} \end{aligned}$$

(e) The 2-variable model $z = -1660.9 + 0.67152x_7 + 134.28x_{10}$ has Mallows' C_p equal to -4.0.

(f) Using a value of 0.15 for both α-to-enter and α-to-remove, the equation chosen by stepwise regression is $z = -1660.9 + 0.67152x_7 + 134.28x_{10}$

(g) The 17-variable model below has adjusted R^2 equal to 0.97783.

$$\begin{aligned} w &= 700.56 - 21.701x_2 - 20.000x_3 + 21.813x_4 + 62.599x_5 + 0.016156x_7 - 0.012689x_8 \\ &\quad + 1.1315x_9 + 15.245x_{10} + 1.1103x_{11} - 20.523x_{13} - 90.189x_{15} - 0.77442x_{16} + 7.5559x_{19} \\ &\quad + 5.9163x_{20} - 7.5497x_{21} + 12.994x_{22} - 271.32x_{23} \end{aligned}$$

(h) The 13-variable model below has Mallows' C_p equal to 8.0.

$$\begin{aligned} w &= 567.06 - 23.582x_2 - 16.766x_3 + 90.482x_5 + 0.0082274x_7 - 0.011004x_8 + 0.89554x_9 \\ &\quad + 12.131x_{10} - 11.984x_{13} - 0.67302x_{16} + 11.097x_{19} + 4.6448x_{20} + 11.108x_{22} - 217.82x_{23} \end{aligned}$$

(i) Using a value of 0.15 for both α-to-enter and α-to-remove, the equation chosen by stepwise regression is $w = 130.92 - 28.085x_2 + 113.49x_5 + 0.16802x_9 - 0.20216x_{16} + 11.417x_{19} + 12.068x_{21} - 78.371x_{23}$.

Chapter 9

Section 9.1

1. (a)

Source	DF	SS	MS	F	P
Temperature	3	202.44	67.481	59.731	0.000
Error	16	18.076	1.1297		
Total	19	220.52			

(b) Yes. $F_{3,\,16} = 59.731$, $P < 0.001$ ($P \approx 0$).

3. (a)

Source	DF	SS	MS	F	P
Treatment	4	19.009	4.7522	2.3604	0.117
Error	11	22.147	2.0133		
Total	15	41.155			

(b) No. $F_{4,\,11} = 2.3604$, $P > 0.10$ ($P = 0.117$).

5. (a)

Source	DF	SS	MS	F	P
Site	3	1.4498	0.48327	2.1183	0.111
Error	47	10.723	0.22815		
Total	50	12.173			

(b) No. $F_{3,\,47} = 2.1183$, $P > 0.10$ ($P = 0.111$).

7. (a)

Source	DF	SS	MS	F	P
Group	3	0.19218	0.064062	1.8795	0.142
Error	62	2.1133	0.034085		
Total	65	2.3055			

(b) No. $F_{3,\,62} = 1.8795$, $P > 0.10$ ($P = 0.142$).

9. (a)

Source	DF	SS	MS	F	P
Temperature	2	148.56	74.281	10.53	0.011
Error	6	42.327	7.0544		
Total	8	190.89			

(b) Yes. $F_{2,\,6} = 10.53$, $0.01 < P < 0.05$ ($P = 0.011$).

11. No, $F_{3,16} = 15.83$, $P < 0.001$ ($P \approx 4.8 \times 10^{-5}$).

13. (a)

Source	DF	SS	MS	F	P
Temperature	3	58.650	19.550	8.4914	0.001
Error	16	36.837	2.3023		
Total	19	95.487			

(b) Yes, $F_{3,\,16} = 8.4914$, $0.001 < P < 0.01$ ($P = 0.0013$).

15. (a)

Source	DF	SS	MS	F	P
Grade	3	1721.4	573.81	9.4431	0.000
Error	96	5833.4	60.765		
Total	99	7554.9			

(b) Yes, $F_{3,96} = 9.4431$, $P < 0.001$ ($P \approx 0$).

17. (a)

Source	DF	SS	MS	F	P
Soil	2	2.1615	1.0808	5.6099	0.0104
Error	23	4.4309	0.19265		
Total	25	6.5924			

(b) Yes. $F_{2,\,23} = 5.6099$, $0.01 < P < 0.05$ ($P = 0.0104$).

Section 9.2

1. (a) Yes, $F_{5,6} = 46.64$, $P \approx 0$.

(b) $q_{6,6,.05} = 5.63$. The value of MSE is 0.00508. The 5% critical value is therefore $5.63\sqrt{0.00508/2} = 0.284$. Any pair that differs by more than 0.284 can be concluded to be different. The following pairs meet this criterion: A and B, A and C, A and D, A and E, B and C, B and D, B and E, B and F, D and F.

3. The sample sizes are $J_1 = 16$, $J_2 = 9$, $J_3 = 14$, $J_4 = 12$. MSE = 0.22815. We should use the Studentized range value $q_{4,47,.05}$. This value is not in the table, so we will use $q_{4,40,.05} = 3.79$, which is only slightly larger. The values of $q_{4,40,.05}\sqrt{(\text{MSE}/2)(1/J_i + 1/J_j)}$ are presented in the table on the left, and the values of the differences $|\overline{X}_{i.} - \overline{X}_{j.}|$ are presented in the table on the right.

	1	2	3	4
1	–	0.53336	0.46846	0.48884
2	0.53336	–	0.54691	0.56446
3	0.46846	0.54691	–	0.50358
4	0.48884	0.56446	0.50358	–

	1	2	3	4
1	0	0.50104	0.16223	0.18354
2	0.50104	0	0.33881	0.3175
3	0.16223	0.33881	0	0.02131
4	0.18354	0.31750	0.02131	0

None of the differences exceeds its critical value, so we cannot conclude at the 5% level that any of the treatment means differ.

5. The sample means are $\overline{X}_1 = 1.998$, $\overline{X}_2 = 3.0000$, $\overline{X}_3 = 5.300$. The sample sizes are $J_1 = 5$, $J_2 = J_3 = 3$. The upper 5% point of the Studentized range is $q_{3,8,.05} = 4.04$. The 5% critical value for $|\overline{X}_1 - \overline{X}_2|$ and for $|\overline{X}_1 - \overline{X}_3|$ is $4.04\sqrt{(1.3718/2)(1/5+1/3)} = 2.44$, and the 5% critical value for $|\overline{X}_2 - \overline{X}_3|$ is $4.04\sqrt{(1.3718/2)(1/3+1/3)} = 2.73$. Therefore means 1 and 3 differ at the 5% level.

7. (a) $\overline{X}_{..} = 88.04$, $\quad I = 4$, $\quad J = 5$, $\quad \text{MSTr} = \sum_{i=1}^{I} J(\overline{X}_{i.} - \overline{X}_{..})^2/(I-1) = 19.554$.

$F = \text{MSTr}/\text{MSE} = 19.554/3.85 = 5.08$. There are 3 and 16 degrees of freedom, so $0.01 < P < 0.05$ (a computer package gives $P = 0.012$). The null hypothesis of no difference is rejected at the 5% level.

(b) $q_{4,\,16\,.05} = 4.05$, so catalysts whose means differ by more than $4.05\sqrt{3.85/5} = 3.55$ are significantly different at the 5% level. Catalyst 1 and Catalyst 2 both differ significantly from Catalyst 4.

9. The value of the F statistic is $F = \text{MSTr}/\text{MSE} = 19.554/\text{MSE}$. The upper 5% point of the $F_{3,16}$ distribution is 3.24. Therefore the F test will reject at the 5% level if $19.554/\text{MSE} \geq 3.24$, or, equivalently, if $\text{MSE} \leq 6.035$.

The largest difference between the sample means is $89.88 - 85.79 = 4.09$. The upper 5% point of the Studentized range distribution is $q_{4,16,.05} = 4.05$. Therefore the Tukey-Kramer test will fail to find any differences significant at the 5% level if $4.09 < 4.05\sqrt{\text{MSE}/5}$, or equivalently, if $\text{MSE} > 5.099$.

Therefore the F test will reject the null hypothesis that all the means are equal, but the Tukey-Kramer test will not find any pair of means to differ at the 5% level, for any value of MSE satisfying $5.099 < \text{MSE} < 6.035$.

Section 9.3

1. Let I be the number of levels of oil type, let J be the number of levels of piston ring type, and let K be the number of replications. Then $I = 4$, $J = 3$, and $K = 3$.

(a) The number of degrees of freedom for oil type is $I - 1 = 3$.

(b) The number of degrees of freedom for piston ring type is $J - 1 = 2$.

(c) The number of degrees of freedom for interaction is $(I - 1)(J - 1) = 6$.

(d) The number of degrees of freedom for error is $IJ(K - 1) = 24$.

(e) The mean squares are found by dividing the sums of squares by their respective degrees of freedom.

The F statistics are found by dividing each mean square by the mean square for error. The number of degrees of freedom for the numerator of an F statistic is the number of degrees of freedom for its effect, and the number of degrees of freedom for the denominator is the number of degrees of freedom for error.

P-values may be obtained from the F table, or from a computer software package.

Source	DF	SS	MS	F	P
Oil	3	1.0926	0.36420	5.1314	0.007
Ring	2	0.9340	0.46700	6.5798	0.005
Interaction	6	0.2485	0.041417	0.58354	0.740
Error	24	1.7034	0.070975		
Total	35	3.9785			

(f) Yes. $F_{6,\,24} = 0.58354$, $P > 0.10$ $(P = 0.740)$.

(g) No, some of the main effects of oil type are non-zero. $F_{3,\,24} = 5.1314$, $0.001 < P < 0.01$ $(P = 0.007)$.

(h) No, some of the main effects of piston ring type are non-zero. $F_{2,\,24} = 6.5798$, $0.001 < P < 0.01$ $(P = 0.005)$.

3. (a) Let I be the number of levels of mold temperature, let J be the number of levels of alloy, and let K be the number of replications. Then $I = 5$, $J = 3$, and $K = 4$.

The number of degrees of freedom for mold temperature is $I - 1 = 4$.

The number of degrees of freedom for alloy is $J - 1 = 2$.

The number of degrees of freedom for interaction is $(I - 1)(J - 1) = 8$.

The number of degrees of freedom for error is $IJ(K - 1) = 45$.

The mean squares are found by dividing the sums of squares by their respective degrees of freedom.

The F statistics are found by dividing each mean square by the mean square for error. The number of degrees of freedom for the numerator of an F statistic is the number of degrees of freedom for its effect, and the number of degrees of freedom for the denominator is the number of degrees of freedom for error.

P-values may be obtained from the F table, or from a computer software package.

Source	DF	SS	MS	F	P
Mold Temp.	4	69738	17434.5	6.7724	0.000
Alloy	2	8958	4479.0	1.7399	0.187
Interaction	8	7275	909.38	0.35325	0.939
Error	45	115845	2574.3		
Total	59	201816			

(b) Yes. $F_{8,45} = 0.35325$, $P > 0.10$ ($P = 0.939$).

(c) No, some of the main effects of mold temperature are non-zero. $F_{4,\,45} = 6.7724$, $P < 0.001$ ($P \approx 0$).

(d) Yes. $F_{3,\,45} = 1.7399$, $P > 0.10$, ($P = 0.187$).

5. (a)

Source	DF	SS	MS	F	P
Solution	1	1993.9	1993.9	5.1983	0.034
Temperature	1	78.634	78.634	0.20500	0.656
Interaction	1	5.9960	5.9960	0.015632	0.902
Error	20	7671.4	383.57		
Total	23	9750.0			

(b) Yes, $F_{1,20} = 0.015632$, $P > 0.10$ ($P = 0.902$).

(c) Yes, since the additive model is plausible. The mean yield stress differs between Na_2HPO_4 and NaCl: $F_{1,20} = 5.1983$, $0.01 < P < 0.05$ ($P = 0.034$).

(d) There is no evidence that the temperature affects yield stress: $F_{1,20} = 0.20500$, $P > 0.10$ ($P = 0.656$).

7. (a)

Source	DF	SS	MS	F	P
Adhesive	1	17.014	17.014	6.7219	0.024
Curing Pressure	2	35.663	17.832	7.045	0.009
Interaction	2	39.674	19.837	7.8374	0.007
Error	12	30.373	2.5311		
Total	17	122.73			

(b) No. $F_{2,\,12} = 7.8374$, $0.001 < P < 0.01$ ($P = 0.007$).

(c) No, because the additive model is not plausible.

(d) No, because the additive model is not plausible.

9. (a)

Source	DF	SS	MS	F	P
Taper Material	1	0.059052	0.059052	23.630	0.000
Neck Length	2	0.028408	0.014204	5.6840	0.010
Interaction	2	0.0090089	0.0045444	1.8185	0.184
Error	24	0.059976	0.002499		
Total	29	0.15652			

(b) Yes, the interactions may plausibly be equal to 0. The value of the test statistic is 1.8185, its null distribution is $F_{2,24}$, and $P > 0.10$ ($P = 0.184$).

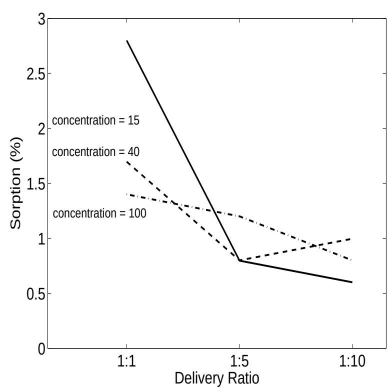
(c) Yes, since the additive model is plausible. The mean coefficient of friction differs between CPTi-ZrO_2 and TiAlloy-ZrO_2: $F_{1,24} = 23.630$, $P < 0.001$.

11. (a)

Source	DF	SS	MS	F	P
Concentration	2	0.37936	0.18968	3.8736	0.040
Delivery Ratio	2	7.34	3.67	74.949	0.000
Interaction	4	3.4447	0.86118	17.587	0.000
Error	18	0.8814	0.048967		
Total	26	12.045			

(b) No. The The value of the test statistic is 17.587, its null distribution is $F_{4,18}$, and $P \approx 0$.

(c)



The slopes of the line segments are quite different from one another, indicating a high degree of interaction.

13. (a)

Source	DF	SS	MS	F	P
Wafer	2	114661.4	57330.7	11340.1	0.000
Operator	2	136.78	68.389	13.53	0.002
Interaction	4	6.5556	1.6389	0.32	0.855
Error	9	45.500	5.0556		
Total	17	114850.3			

(b) There are differences among the operators. $F_{2,\,9} = 13.53$, $0.01 < P < 0.001$ ($P = 0.002$).

15. (a)

Source	DF	SS	MS	F	P
PVAL	2	125.41	62.704	8.2424	0.003
DCM	2	1647.9	823.94	108.31	0.000
Interaction	4	159.96	39.990	5.2567	0.006
Error	18	136.94	7.6075		
Total	26	2070.2			

(b) Since the interaction terms are not equal to 0, ($F_{4,18} = 5.2567$, $P = 0.006$), we cannot interpret the main effects. Therefore we compute the cell means. These are

	DCM (ml)		
PVAL	50	40	30
0.5	97.8	92.7	74.2
1.0	93.5	80.8	75.4
2.0	94.2	88.6	78.8

We conclude that a DCM level of 50 ml produces greater encapsulation efficiency than either of the other levels. If DCM = 50, the PVAL concentration does not have much effect. Note that for DCM = 50, encapsulation efficiency is maximized at the lowest PVAL concentration, but for DCM = 30 it is maximized at the highest PVAL concentration. This is the source of the significant interaction.

Section 9.4

1. (a) Liming is the blocking factor, soil is the treatment factor.

(b)

Source	DF	SS	MS	F	P
Soil	3	1.178	0.39267	18.335	0.000
Block	4	5.047	1.2617	58.914	0.000
Error	12	0.257	0.021417		
Total	19	6.482			

(c) Yes, $F_{3,12} = 18.335$, $P \approx 0$.

3. (a) Let I be the number of levels for lighting method, let J be the number of levels for blocks, and let K be the number of replications. Then $I = 4$, $J = 3$, and $K = 3$.

The number of degrees of freedom for treatments is $I - 1 = 3$.

The number of degrees of freedom for blocks is $J - 1 = 2$.

The number of degrees of freedom for interaction is $(I - 1)(J - 1) = 6$.

The number of degrees of freedom for error is $IJ(K - 1) = 24$.

The mean squares are found by dividing the sums of squares by their respective degrees of freedom.

The F statistics are found by dividing each mean square by the mean square for error. The number of degrees of freedom for the numerator of an F statistic is the number of degrees of freedom for its effect, and the number of degrees of freedom for the denominator is the number of degrees of freedom for error.

P-values may be obtained from the F table, or from a computer software package.

Source	DF	SS	MS	F	P
Lighting	3	9943	3314.33	3.3329	0.036
Block	2	11432	5716.00	5.7481	0.009
Interaction	6	6135	1022.50	1.0282	0.431
Error	24	23866	994.417		
Total	35	51376			

(b) Yes. The P-value for interactions is large (0.431).

(c) Yes. The P-value for lighting is small (0.036).

5. (a)

Source	DF	SS	MS	F	P
Variety	9	339032	37670	2.5677	0.018
Block	5	1860838	372168	25.367	0.000
Error	45	660198	14671		
Total	59	2860069			

(b) Yes, $F_{9,45} = 2.5677$, $P = 0.018$.

7. (a) One motor of each type should be tested on each day. The order in which the motors are tested on any given day should be chosen at random. This is a randomized block design, in which the days are the blocks. It is not a completely randomized design, since randomization occurs only within blocks.

(b) The test statistic is $\dfrac{\sum_{i=1}^{5}(\overline{X}_{i.} - \overline{X}_{..})^2}{\sum_{j=1}^{4}\sum_{i=1}^{5}(X_{ij} - \overline{X}_{i.} - \overline{X}_{.j} - \overline{X}_{..})^2/12}$.

Section 9.5

1.

	A	B	C	D
1	$-$	$-$	$-$	$-$
ad	$+$	$-$	$-$	$+$
bd	$-$	$+$	$-$	$+$
ab	$+$	$+$	$-$	$-$
cd	$-$	$-$	$+$	$+$
ac	$+$	$-$	$+$	$-$
bc	$-$	$+$	$+$	$-$
$abcd$	$+$	$+$	$+$	$+$

The alias pairs are $\{A, BCD\}$, $\{B, ACD\}$, $\{C, ABD\}$, $\{D, ABC\}$, $\{AB, CD\}$, $\{AC, BD\}$, and $\{AD, BC\}$

3. (a)

Variable	Effect	DF	Sum of Squares	Mean Square	F	P
A	6.75	1	182.25	182.25	11.9508	0.009
B	9.50	1	361.00	361.00	23.6721	0.001
C	1.00	1	4.00	4.00	0.2623	0.622
AB	2.50	1	25.00	25.00	1.6393	0.236
AC	0.50	1	1.00	1.00	0.0656	0.804
BC	0.75	1	2.25	2.25	0.1475	0.711
ABC	-2.75	1	30.25	30.25	1.9836	0.197
Error		8	122.00	15.25		
Total		15	727.75			

(b) Factors A and B (temperature and concentration) seem to have an effect on yield. There is no evidence that pH has an effect. None of the interactions appear to be significant. Their P-values are all greater than 0.19.

(c) Since the effect of temperature is positive and statistically significant, we can conclude that the mean yield is higher when temperature is high.

5. (a)

Variable	Effect
A	3.3750
B	23.625
C	1.1250
AB	−2.8750
AC	−1.3750
BC	−1.6250
ABC	1.8750

(b) No, since the design is unreplicated, there is no error sum of squares.

(c) No, none of the interaction terms are nearly as large as the main effect of factor B.

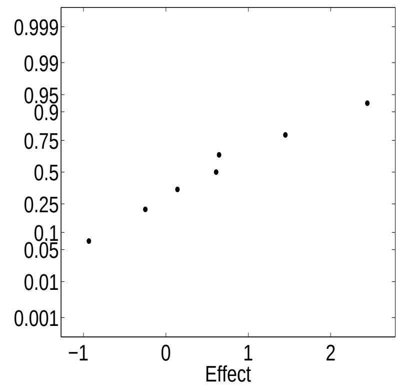
(d) If the additive model is known to hold, then the ANOVA table below shows that the main effect of B is not equal to 0, while the main effects of A and C may be equal to 0.

Variable	Effect	DF	Sum of Squares	Mean Square	F	P
A	3.3750	1	22.781	22.781	2.7931	0.170
B	23.625	1	1116.3	1116.3	136.86	0.000
C	1.1250	1	2.5312	2.5312	0.31034	0.607
Error		4	32.625	8.1562		
Total		7	1174.2			

7. (a)

Variable	Effect
A	2.445
B	0.140
C	−0.250
AB	1.450
AC	0.610
BC	0.645
ABC	−0.935

(b) No, since the design is unreplicated, there is no error sum of squares.



(c) The estimates lie nearly on a straight line, so none of the factors can clearly be said to influence the resistance.

9. (a)

Variable	Effect
A	1.2
B	3.25
C	−16.05
D	−2.55
AB	2
AC	2.9
AD	−1.2
BC	1.05
BD	−1.45
CD	−1.6
ABC	−0.8
ABD	−1.9
ACD	−0.15
BCD	0.8
$ABCD$	0.65

(b) Factor C is the only one that really stands out.

11. (a)

Variable	Effect	DF	Sum of Squares	Mean Square	F	P
A	14.245	1	811.68	811.68	691.2	0.000
B	8.0275	1	257.76	257.76	219.5	0.000
C	−6.385	1	163.07	163.07	138.87	0.000
AB	−1.68	1	11.29	11.29	9.6139	0.015
AC	−1.1175	1	4.9952	4.9952	4.2538	0.073
BC	−0.535	1	1.1449	1.1449	0.97496	0.352
ABC	−1.2175	1	5.9292	5.9292	5.0492	0.055
Error		8	9.3944	1.1743		
Total		15	1265.3			

(b) All main effects are significant, as is the AB interaction. Only the BC interaction has a P value that is reasonably large. All three factors appear to be important, and they seem to interact considerably with each other.

13. (ii) The sum of the main effect of A and the $BCDE$ interaction.

Supplementary Exercises for Chapter 9

1.

Source	DF	SS	MS	F	P
Gypsum	3	0.013092	0.0043639	0.28916	0.832
Error	8	0.12073	0.015092		
Total	11	0.13383			

The value of the test statistic is $F_{3,8} = 0.28916$; $P > 0.10$ ($P = 0.832$). There is no evidence that the pH differs with the amount of gypsum added.

3.

Source	DF	SS	MS	F	P
Day	2	1.0908	0.54538	22.35	0.000
Error	36	0.87846	0.024402		
Total	38	1.9692			

We conclude that the mean sugar content differs among the three days ($F_{2,36} = 22.35$, $P \approx 0$).

5. (a) No. The variances are not constant across groups. In particular, there is an outlier in group 1.

(b) No, for the same reasons as in part (a).

(c)

Source	DF	SS	MS	F	P
Group	4	5.2029	1.3007	8.9126	0.000
Error	35	5.1080	0.14594		
Total	39	10.311			

We conclude that the mean dissolve time differs among the groups ($F_{4,35} = 8.9126$, $P \approx 0$).

7. The recommendation is not a good one. The engineer is trying to interpret the main effects without looking at the interactions. The small P-value for the interactions indicates that they must be taken into account. Looking at the cell means, it is clear that if design 2 is used, then the less expensive material performs just as well as the more expensive material. The best recommendation, therefore, is to use design 2 with the less expensive material.

9. (a)

Source	DF	SS	MS	F	P
Base	3	13495	4498.3	7.5307	0.000
Instrument	2	90990	45495	76.164	0.000
Interaction	6	12050	2008.3	3.3622	0.003
Error	708	422912	597.33		
Total	719	539447			

(b) No, it is not appropriate because there are interactions between the row and column effects ($F_{6,708} = 3.3622$, $P = 0.003$).

11. (a)

Source	DF	SS	MS	F	P
Channel Type	4	1011.7	252.93	8.7139	0.001
Error	15	435.39	29.026		
Total	19	1447.1			

Yes. $F_{4,15} = 8.7139$, $P = 0.001$.

(b) $q_{5,20,.05} = 4.23$, MSE $= 29.026$, $J = 4$. The 5% critical value is therefore $4.23\sqrt{29.026/4} = 11.39$. The sample means for the five channels are $\overline{X}_1 = 44.000$, $\overline{X}_2 = 44.100$, $\overline{X}_3 = 30.900$, $\overline{X}_4 = 28.575$, $\overline{X}_5 = 44.425$. We can therefore conclude that channels 3 and 4 differ from channels 1, 2, and 5.

13.

Source	DF	SS	MS	F	P
Well Type	4	5.7523	1.4381	1.5974	0.175
Error	289	260.18	0.90028		
Total	293	265.93			

No. $F_{4,289} = 1.5974$, $P > 0.10$ ($P = 0.175$).

15. (a)

Variable	Effect	Variable	Effect	Variable	Effect	Variable	Effect
A	3.9875	AB	−0.1125	BD	−0.0875	ACD	0.4875
B	2.0375	AC	0.0125	CD	0.6375	BCD	−0.3125
C	1.7125	AD	−0.9375	ABC	−0.2375	$ABCD$	−0.7125
D	3.7125	BC	0.7125	ABD	0.5125		

(b) The main effects are noticeably larger than the interactions, and the main effects for A and D are noticeably larger than those for B and C.

(c)

Variable	Effect	DF	Sum of Squares	Mean Square	F	P
A	3.9875	1	63.601	63.601	68.415	0.000
B	2.0375	1	16.606	16.606	17.863	0.008
C	1.7125	1	11.731	11.731	12.619	0.016
D	3.7125	1	55.131	55.131	59.304	0.001
AB	−0.1125	1	0.050625	0.050625	0.054457	0.825
AC	0.0125	1	0.000625	0.000625	0.00067231	0.980
AD	−0.9375	1	3.5156	3.5156	3.7818	0.109
BC	0.7125	1	2.0306	2.0306	2.1843	0.199
BD	−0.0875	1	0.030625	0.030625	0.032943	0.863
CD	0.6375	1	1.6256	1.6256	1.7487	0.243
Interaction		5	4.6481	0.92963		
Total		15	158.97			

We can conclude that each of the factors A, B, C, and D has an effect on the outcome.

(d) The F statistics are computed by dividing the mean square for each effect (equal to its sum of squares) by the error mean square 1.04. The degrees of freedom for each F statistic are 1 and 4. The results are summarized in the following table.

Variable	Effect	DF	Sum of Squares	Mean Square	F	P
A	3.9875	1	63.601	63.601	61.154	0.001
B	2.0375	1	16.606	16.606	15.967	0.016
C	1.7125	1	11.731	11.731	11.279	0.028
D	3.7125	1	55.131	55.131	53.01	0.002
AB	−0.1125	1	0.050625	0.050625	0.048678	0.836
AC	0.0125	1	0.000625	0.000625	0.00060096	0.982
AD	−0.9375	1	3.5156	3.5156	3.3804	0.140
BC	0.7125	1	2.0306	2.0306	1.9525	0.235
BD	−0.0875	1	0.030625	0.030625	0.029447	0.872
CD	0.6375	1	1.6256	1.6256	1.5631	0.279
ABC	−0.2375	1	0.22563	0.22563	0.21695	0.666
ABD	0.5125	1	1.0506	1.0506	1.0102	0.372
ACD	0.4875	1	0.95063	0.95063	0.91406	0.393
BCD	−0.3125	1	0.39062	0.39062	0.3756	0.573
$ABCD$	−0.7125	1	2.0306	2.0306	1.9525	0.235

(e) Yes. None of the P-values for the third- or higher-order interactions are small.

(f) We can conclude that each of the factors A, B, C, and D has an effect on the outcome.

17. (a)

Source	DF	SS	MS	F	P
H_2SO_4	2	457.65	228.83	8.8447	0.008
$CaCl_2$	2	38783	19391	749.53	0.000
Interaction	4	279.78	69.946	2.7036	0.099
Error	9	232.85	25.872		
Total	17	39753			

(b) The P-value for interactions is 0.099. One cannot rule out the additive model.

(c) Yes, $F_{2,9} = 8.8447$, $0.001 < P < 0.01$ ($P = 0.008$).

(d) Yes, $F_{2,9} = 749.53$, $P \approx 0.000$.

Chapter 10

Section 10.1

1. (a) Count

   (b) Continuous

   (c) Binary

   (d) Continuous

3. (a) is in control

   (b) has high capability

5. (a) False. Being in a state of statistical control means only that no special causes are operating. It is still possible for the process to be calibrated incorrectly, or for the variation due to common causes to be so great that much of the output fails to conform to specifications.

   (b) False. Being out of control means that some special causes are operating. It is still possible for much of the output to meet specifications.

   (c) True. This is the definition of statistical control.

   (d) True. This is the definition of statistical control.

Section 10.2

1. (a) The sample size is $n = 4$. The upper and lower limits for the R-chart are $D_3\overline{R}$ and $D_4\overline{R}$, respectively.
   From the control chart table, $D_3 = 0$ and $D_4 = 2.282$.
   $\overline{R} = 143.7/30 = 4.79$. Therefore LCL $= 0$, and UCL $= 10.931$.

   (b) The sample size is $n = 4$. The upper and lower limits for the S-chart are $B_3\overline{s}$ and $B_4\overline{s}$, respectively.
   From the control chart table, $B_3 = 0$ and $B_4 = 2.266$.
   $\overline{s} = 62.5/30 = 2.08333$. Therefore LCL $= 0$ and UCL $= 4.721$.

(c) The upper and lower limits for the $\overline{X}$-chart are $\overline{\overline{X}} - A_2\overline{R}$ and $\overline{\overline{X}} + A_2\overline{R}$, respectively.

From the control chart table, $A_2 = 0.729$. $\overline{R} = 143.7/30 = 4.79$ and $\overline{\overline{X}} = 712.5/30 = 23.75$.

Therefore LCL $= 20.258$ and UCL $= 27.242$.

(d) The upper and lower limits for the $\overline{X}$-chart are $\overline{\overline{X}} - A_3\overline{s}$ and $\overline{\overline{X}} + A_3\overline{s}$, respectively.

From the control chart table, $A_3 = 1.628$. $\overline{s} = 62.5/30 = 2.08333$ and $\overline{\overline{X}} = 712.5/30 = 23.75$.

Therefore LCL $= 20.358$ and UCL $= 27.142$.

3. (a) The sample size is $n = 5$. The upper and lower limits for the R-chart are $D_3\overline{R}$ and $D_4\overline{R}$, respectively.

From the control chart table, $D_3 = 0$ and $D_4 = 2.114$.

$\overline{R} = 0.1395$. Therefore LCL $= 0$ and UCL $= 0.2949$. The variance is in control.

(b) The upper and lower limits for the $\overline{X}$-chart are $\overline{\overline{X}} - A_2\overline{R}$ and $\overline{\overline{X}} + A_2\overline{R}$, respectively.

From the control chart table, $A_2 = 0.577$. $\overline{R} = 0.1395$ and $\overline{\overline{X}} = 2.505$.

Therefore LCL $= 2.4245$ and UCL $= 2.5855$. The process is out of control for the first time on sample 8.

(c) The 1σ limits are $\overline{\overline{X}} - A_2\overline{R}/3 = 2.478$ and $\overline{\overline{X}} + A_2\overline{R}/3 = 2.5318$, respectively.

The 2σ limits are $\overline{\overline{X}} - 2A_2\overline{R}/3 = 2.4513$ and $\overline{\overline{X}} + 2A_2\overline{R}/3 = 2.5587$, respectively.

The process is out of control for the first time on sample 7, where 2 out of the last three samples are below the lower 2σ control limit.

5. (a) $\overline{X}$ has a normal distribution with $\mu = 14$ and $\sigma_{\overline{X}} = 3/\sqrt{5} = 1.341641$.

The 3σ limits are $12 \pm 3(1.341641)$, or 7.97508 and 16.02492.

The probability that a point plots outside the 3σ limits is $p = P(\overline{X} < 7.97508) + P(\overline{X} > 16.02492)$.

The z-score for 7.97508 is $(7.97508 - 14)/1.341641 = -4.49$.

The z-score for 16.02492 is $(16.02492 - 14)/1.341641 = 1.51$.

The probability that a point plots outside the 3σ limits is the sum of the area to the left of $z = -4.49$ and the area to the right of $z = 1.51$.

Therefore $p = 0.0000 + 0.0655 = 0.0655$.

The ARL is $1/p = 1/0.0655 = 15.27$.

(b) Let m be the required value. Since the shift is upward, $m > 12$.

The probability that a point plots outside the 3σ limits is $p = P(\overline{X} < 7.97508) + P(\overline{X} > 16.02492)$.

Since ARL $= 4$, $p = 1/4$. Since $m > 12$, $P(\overline{X} > 16.02492) > P(\overline{X} < 7.97508)$.

Find m so that $P(\overline{X} > 16.02492) = 1/4$, and check that $P(\overline{X} < 7.97508) \approx 0$.

The z-score for 16.02492 is $(16.02492 - m)/1.341641$. The z-score with an area of $1/4 = 0.25$ to the right is approximately $z = 0.67$.

Therefore $0.67 = (16.02492 - m)/1.341641$, so $m = 15.126$.

Now check that $P(\overline{X} < 7.97508) \approx 0$.

The z-score for 7.97508 is $(7.97508 - 15.126)/1.341641 = -5.33$. So $P(\overline{X} < 7.97508) \approx 0$.

Therefore $m = 15.126$.

(c) We will find the required value for $\sigma_{\overline{X}}$.

The probability that a point plots outside the 3σ limits is $p = P(\overline{X} < 12 - 3\sigma_{\overline{X}}) + P(\overline{X} > 12 + 3\sigma_{\overline{X}})$.

Since ARL $= 4$, $p = 1/4$. Since the process mean is 14, $P(\overline{X} > 12 + 3\sigma_{\overline{X}}) > P(\overline{X} > 12 - 3\sigma_{\overline{X}})$.

Find σ so that $P(\overline{X} > 12 + 3\sigma_{\overline{X}}) = 1/4$, and check that $P(\overline{X} < 12 - 3\sigma_{\overline{X}}) \approx 0$.

The z-score for $12 + 3\sigma_{\overline{X}}$ is $(12 + 3\sigma_{\overline{X}} - 14)/\sigma_{\overline{X}}$. The z-score with an area of $1/4 = 0.25$ to the right is approximately $z = 0.67$.

Therefore $(12 + 3\sigma_{\overline{X}} - 14)/\sigma_{\overline{X}} = 0.67$, so $\sigma_{\overline{X}} = 0.8584$.

Now check that $P(\overline{X} < 12 - 3\sigma_{\overline{X}}) \approx 0$.

$12 - 3\sigma_{\overline{X}} = 9.425$. The z-score for 9.425 is $(9.425 - 14)/0.8584 = -5.33$, so $P(\overline{X} < 12 - 3\sigma_{\overline{X}}) \approx 0$.

Therefore $\sigma_{\overline{X}} = 0.8584$.

Since $n = 5$, $\sigma_{\overline{X}} = \sigma/\sqrt{5}$.

Therefore $\sigma = 1.92$.

(d) Let n be the required sample size. Then $\sigma_{\overline{X}} = 3/\sqrt{n}$.

From part (c), $\sigma_{\overline{X}} = 0.8584$. Therefore $3/\sqrt{n} = 0.8584$, so $n = 12.214$.

Round up to obtain $n = 13$.

7. The probability of a false alarm on any given sample is 0.0027, and the probability that there will not be a false alarm on any given sample is 0.9973.

(a) The probability that there will be no false alarm in the next 50 samples is $0.9973^{50} = 0.874$. Therefore the probability that there will be a false alarm within the next 50 samples is $1 - 0.874 = 0.126$.

(b) The probability that there will be no false alarm in the next 100 samples is $0.9973^{100} = 0.763$. Therefore the probability that there will be a false alarm within the next 50 samples is $1 - 0.763 = 0.237$.

(c) The probability that there will be no false alarm in the next 200 samples is $0.9973^{200} = 0.582$.

(d) Let n be the required number. Then $0.9973^n = 0.5$, so $n \ln 0.9973 = \ln 0.5$. Solving for n yields $n = 256.37 \approx 257$.

9. (a) The sample size is $n = 8$. The upper and lower limits for the S-chart are $B_3\overline{s}$ and $B_4\overline{s}$, respectively.

From the control chart table, $B_3 = 0.185$ and $B_4 = 1.815$.

$\overline{s} = 0.0880$. Therefore LCL $= 0.01628$ and UCL $= 0.1597$. The variance is in control.

(b) The upper and lower limits for the $\overline{X}$-chart are $\overline{\overline{X}} - A_3\overline{s}$ and $\overline{\overline{X}} + A_3\overline{s}$, respectively.

From the control chart table, $A_3 = 1.099$. $\overline{s} = 0.0880$ and $\overline{\overline{X}} = 9.9892$.

Therefore LCL $= 9.8925$ and UCL $= 10.0859$. The process is out of control for the first time on sample 3.

(c) The 1σ limits are $\overline{\overline{X}} - A_3\overline{s}/3 = 9.9570$ and $\overline{\overline{X}} + A_3\overline{s}/3 = 10.0214$, respectively.

The 2σ limits are $\overline{\overline{X}} - 2A_3\overline{s}/3 = 9.9247$ and $\overline{\overline{X}} + 2A_3\overline{s}/3 = 10.0537$, respectively.

The process is out of control for the first time on sample 3, where one sample exceeds the upper 3σ control limit.

11. (a) The sample size is $n = 5$. The upper and lower limits for the S-chart are $B_3\overline{s}$ and $B_4\overline{s}$, respectively.

From the control chart table, $B_3 = 0$ and $B_4 = 2.089$.

$\overline{s} = 0.4647$. Therefore LCL $= 0$ and UCL $= 0.971$. The variance is in control.

(b) The upper and lower limits for the $\overline{X}$-chart are $\overline{\overline{X}} - A_3\overline{s}$ and $\overline{\overline{X}} + A_3\overline{s}$, respectively.

From the control chart table, $A_3 = 1.427$. $\overline{s} = 0.4647$ and $\overline{\overline{X}} = 9.81$.

Therefore LCL $= 9.147$ and UCL $= 10.473$. The process is in control.

(c) The 1σ limits are $\overline{\overline{X}} - A_3\overline{s}/3 = 9.589$ and $\overline{\overline{X}} + A_3\overline{s}/3 = 10.031$, respectively.

The 2σ limits are $\overline{\overline{X}} - 2A_3\overline{s}/3 = 9.368$ and $\overline{\overline{X}} + 2A_3\overline{s}/3 = 10.252$, respectively.

The process is out of control for the first time on sample 9, where 2 of the last three sample means are below the lower 2σ control limit.

13. (a) The sample size is $n = 4$. The upper and lower limits for the S-chart are $B_3\overline{s}$ and $B_4\overline{s}$, respectively.

From the control chart table, $B_3 = 0$ and $B_4 = 2.266$.

$\overline{s} = 3.082$. Therefore LCL $= 0$ and UCL $= 6.984$.

The variance is out of control on sample 8. After deleting this sample, $\overline{\overline{X}} = 150.166$ and $\overline{s} = 2.911$. The new limits for the S-chart are 0 and 6.596. The variance is now in control.

(b) The upper and lower limits for the $\overline{X}$-chart are $\overline{\overline{X}} - A_3\overline{s}$ and $\overline{\overline{X}} + A_3\overline{s}$, respectively.

From the control chart table, $A_3 = 1.628$. $\overline{s} = 2.911$ and $\overline{\overline{X}} = 150.166$.

Therefore LCL $= 145.427$ and UCL $= 154.905$. The process is in control.

(c) The 1σ limits are $\overline{\overline{X}} - A_3\overline{s}/3 = 148.586$ and $\overline{\overline{X}} + A_3\overline{s}/3 = 151.746$, respectively.

The 2σ limits are $\overline{\overline{X}} - 2A_3\overline{s}/3 = 147.007$ and $\overline{\overline{X}} + 2A_3\overline{s}/3 = 153.325$, respectively.

The process is in control (recall that sample 8 has been deleted).

Section 10.3

1. The sample size is $n = 300$. $\overline{p} = 1.42/40 = 0.0355$.
   The centerline is $\overline{p} = 0.0355$.
   The LCL is $\overline{p} - 3\sqrt{\overline{p}(1-\overline{p})/300} = 0.00345$.
   The UCL is $\overline{p} + 3\sqrt{\overline{p}(1-\overline{p})/300} = 0.06755$.

3. Yes, the only information needed to compute the control limits is $\overline{p}$ and the sample size n. In this case, $n = 200$, and $\overline{p} = (748/40)/200 = 0.0935$.
   The control limits are $\overline{p} \pm 3\sqrt{\overline{p}(1-\overline{p})/n}$, so
   LCL $= 0.0317$ and UCL $= 0.1553$.

5. (iv). The sample size must be large enough so the mean number of defectives per sample is at least 10.

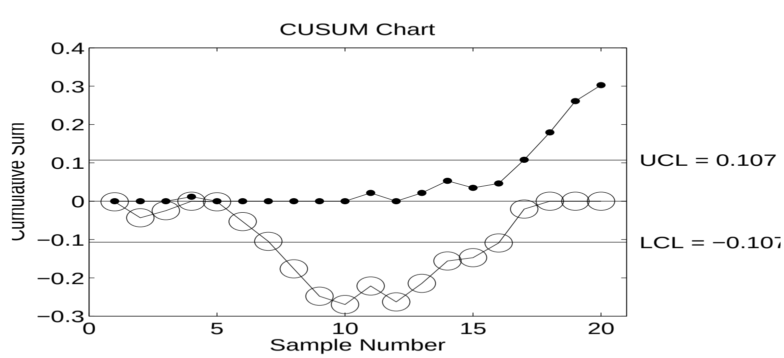
7. It was out of control. The UCL is 23.13.

Section 10.4

1. (a) No samples need be deleted.

   (b) The estimate of $\sigma_{\overline{X}}$ is $A_2\overline{R}/3$. The sample size is $n = 5$.
   $\overline{R} = 0.1395$. From the control chart table, $A_2 = 0.577$.
   Therefore $\sigma_{\overline{X}} = (0.577)(0.1395)/3 = 0.0268$.

   (c)



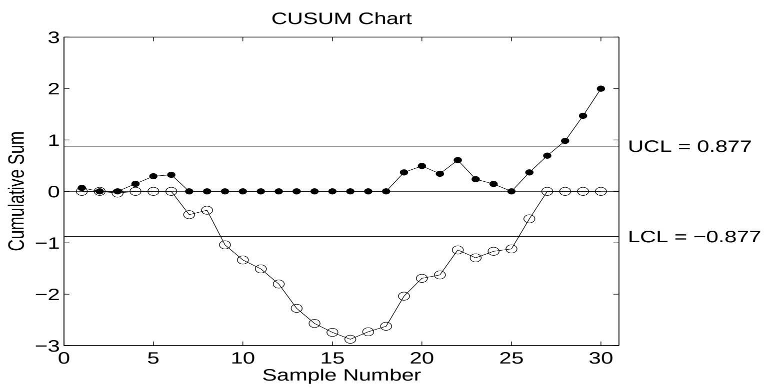
(d) The process is out of control on sample 8.

(e) The Western Electric rules specify that the process is out of control on sample 7.

3. (a) No samples need be deleted.

(b) The estimate of $\sigma_{\overline{X}}$ is $A_2\overline{R}/3$. The sample size is $n = 5$.
$\overline{R} = 1.14$. From the control chart table, $A_2 = 0.577$.
Therefore $\sigma_{\overline{X}} = (0.577)(1.14)/3 = 0.219$.

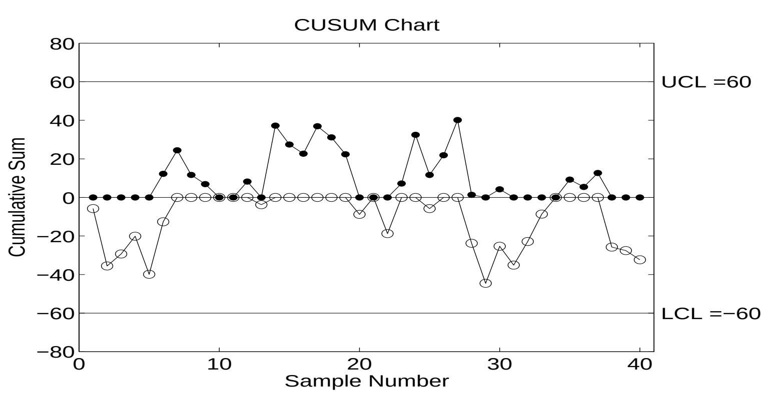
(c)



(d) The process is out of control on sample 9.

(e) The Western Electric rules specify that the process is out of control on sample 9.

5. (a)



(b) The process is in control.

Section 10.5

1. (a) $\widehat{\mu} = \overline{\overline{X}} = 0.205$, $\overline{s} = 0.002$, $LSL = 0.18$, $USL = 0.22$. The sample size is $n = 4$.

   $\widehat{\sigma} = \overline{s}/c_4$. From the control chart table, $c_4 = 0.9213$.

   Therefore $\widehat{\sigma} = 0.002171$.

   Since $\widehat{\mu}$ is closer to USL than to LSL, $C_{pk} = (USL - \widehat{\mu})/(3\widehat{\sigma}) = 2.303$.

   (b) Yes. Since $C_{pk} > 1$, the process capability is acceptable.

3. (a) The capability is maximized when the process mean is equidistant from the specification limits. Therefore the process mean should be set to 0.20.

   (b) $LSL = 0.18$, $USL = 0.22$, $\widehat{\sigma} = 0.002171$.

   If $\mu = 0.20$, then $C_{pk} = (0.22 - 0.20)/[3(0.002171)] = 3.071$.

5. (a) Let μ be the optimal setting for the process mean.

   Then $C_p = (USL - \mu)/(3\sigma) = (\mu - LSL)/(3\sigma)$, so $1.2 = (USL - \mu)/(3\sigma) = (\mu - LSL)/(3\sigma)$.

   Solving for LSL and USL yields $LSL = \mu - 3.6\sigma$ and $USL = \mu + 3.6\sigma$.

   (b) The z-scores for the upper and lower specification limits are $z = \pm 3.60$.

   Therefore, using the normal curve, the proportion of units that are non-conforming is the sum of the areas under the normal curve to the right of $z = 3.60$ and to the left of $z = -3.60$.

   The proportion is $0.0002 + 0.0002 = 0.0004$.

   (c) Likely. The normal approximation is likely to be inaccurate in the tails.

Supplementary Exercises for Chapter 10

1. The sample size is $n = 250$. $\overline{p} = 2.98/50 = 0.0596$.

   The centerline is $\overline{p} = 0.0596$

   The LCL is $\overline{p} - 3\sqrt{\overline{p}(1-\overline{p})/250} = 0.0147$.

   The UCL is $\overline{p} + 3\sqrt{\overline{p}(1-\overline{p})/250} = 0.1045$.

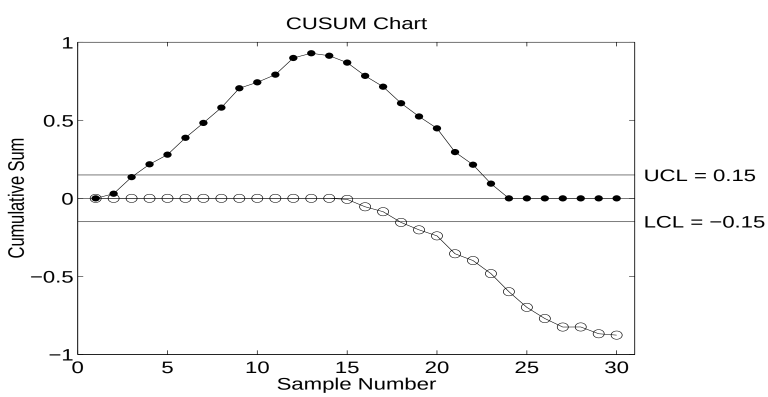
3. (a) The sample size is $n = 3$. The upper and lower limits for the R-chart are $D_3\overline{R}$ and $D_4\overline{R}$, respectively. From the control chart table, $D_3 = 0$ and $D_4 = 2.575$.
$\overline{R} = 0.110$. Therefore LCL = 0 and UCL = 0.283. The variance is in control.

(b) The upper and lower limits for the $\overline{X}$-chart are $\overline{\overline{X}} - A_2\overline{R}$ and $\overline{\overline{X}} + A_2\overline{R}$, respectively.
From the control chart table, $A_2 = 1.023$. $\overline{R} = 0.110$ and $\overline{\overline{X}} = 5.095$.
Therefore LCL = 4.982 and UCL = 5.208. The process is out of control on sample 3.

(c) The 1σ limits are $\overline{\overline{X}} - A_2\overline{R}/3 = 5.057$ and $\overline{\overline{X}} + A_2\overline{R}/3 = 5.133$, respectively.
The 2σ limits are $\overline{\overline{X}} - 2A_2\overline{R}/3 = 5.020$ and $\overline{\overline{X}} + 2A_2\overline{R}/3 = 5.170$, respectively.
The process is out of control for the first time on sample 3, where a sample mean is above the upper 3σ control limit.

5. (a) No samples need be deleted.

(b) The estimate of $\sigma_{\overline{X}}$ is $A_2\overline{R}/3$. The sample size is $n = 3$.
$\overline{R} = 0.110$. From the control chart table, $A_2 = 1.023$.
Therefore $\sigma_{\overline{X}} = (1.023)(0.110)/3 = 0.0375$.

(c) 

(d) The process is out of control on sample 4.

(e) The Western Electric rules specify that the process is out of control on sample 3. The CUSUM chart first signaled an out-of-control condition on sample 4.

7. (a) The sample size is $n = 500$.
The mean number of defectives over the last 25 days is 22.4.

Therefore $\overline{p} = 22.4/500 = 0.0448$.
The control limits are $\overline{p} \pm 3\sqrt{\overline{p}(1-\overline{p})/n}$.
Therefore LCL = 0.0170 and UCL = 0.0726

(b) Sample 12. The proportion of defective chips is then $7/500 = 0.014$, which is below the lower control limit.

(c) No, this special cause improves the process. It should be preserved rather than eliminated.