Solutions Manual

to accompany

STATISTICS FOR ENGINEERS AND SCIENTISTS

Fourth Edition

by

William Navidi

Table of Contents

Chapter 1 1
Chapter 2 13
Chapter 3 54
Chapter 4 74
Chapter 5 116
Chapter 6 135
Chapter 7 171
Chapter 8 191
Chapter 9 210
Chapter 10 230

Chapter 1

Section 1.1

1. (a) The population consists of all the times the process could be run. It is conceptual.

   (b) The population consist of all the registered voters in the state. It is tangible.

   (c) The population consist of all people with high cholesterol levels. It is tangible.

   (d) The population consist of all concrete specimens that could be made from the new formulation. It is conceptual.

   (e) The population consist of all bolts 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.

9. (a) A controlled experiment

   (b) It is well-justified, because it is based on a controlled experiment rather than an observational study.

Section 1.2

1. False

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

5. The sample size can be any odd number.

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

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

11. The total height of the 20 men is $20 \times 178 = 3560$. The total height of the 30 women is $30 \times 164 = 4920$.
The total height of all 50 people is $3560 + 4920 = 8480$. There are $20 + 30 = 50$ people in total. Therefore the mean height for both groups put together is $8480/50 = 169.6$ cm.

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 $(76+79)/2 = 77.5$.

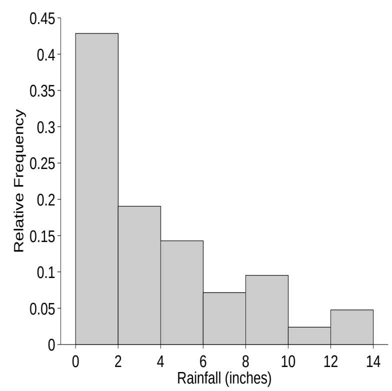
(b) The sample size is $n = 16$. The quintiles have cutpoints $(i/5)(17)$ for $i = 1, 2, 3, 4$. The quintiles are therefore the averages of the sample values in positions 3 and 4, in positions 6 and 7, in positions 10 and 11, and in positions 13 and 14. The quintiles are therefore $(23+41)/2 = 32$, $(46+49)/2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

Section 1.3

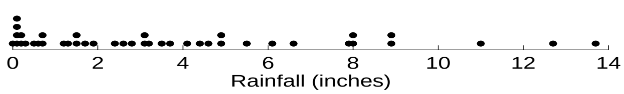
1. (a)

Stem	Leaf
0	011112235677
1	235579
2	468
3	11257
4	14699
5	5
6	16
7	9
8	0099
9	
10	
11	0
12	7
13	7

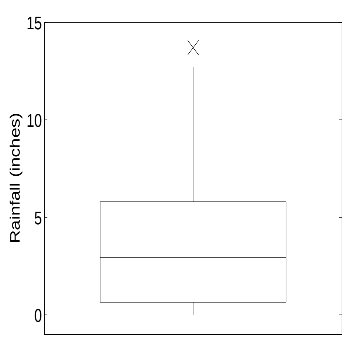
(b) Here is one histogram. Other choices for the endpoints are possible.



(c)



(d)



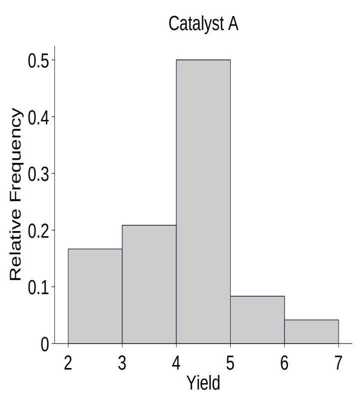
The boxplot shows one outlier.

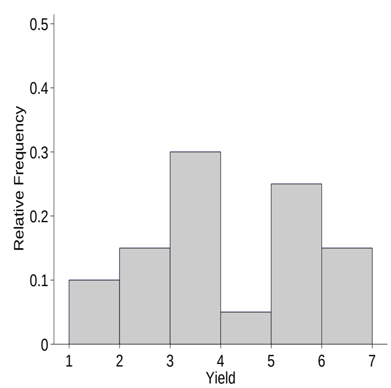
3.

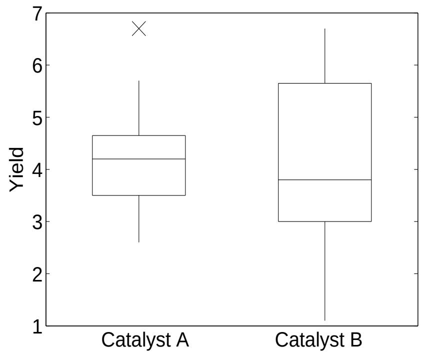
Stem	Leaf
1	1588
2	00003468
3	0234588
4	0346
5	2235666689
6	00233459
7	113558
8	568
9	1225
10	1
11	
12	2
13	06
14	
15	
16	
17	1
18	6
19	9
20	
21	
22	
23	3

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.



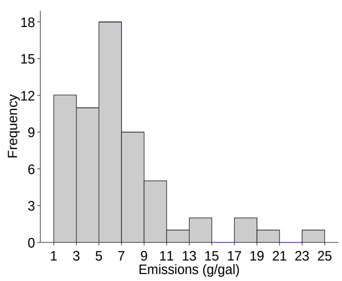


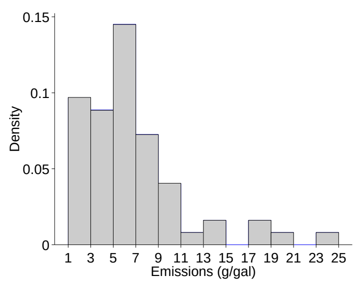
(b) 

(c) The yields for catalyst B are considerably 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. Thus the yields for catalyst B are approximately symmetric. The largest yield 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 240. This sum is approximately $0.14 + 0.10 + 0.05 + 0.01 + 0.02$. This is closest to 30%.

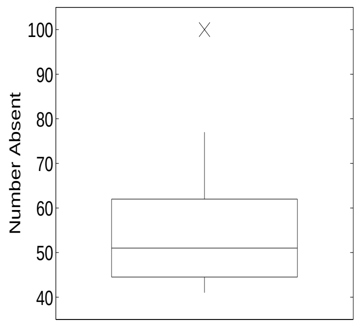
(b) The height of the rectangle over the interval 240–260 is greater than the sum of the heights of the rectangles over the interval 280–340. Therefore there are more men in the interval 240–260 mg/dL.

9. (a) 

(b) 

(c) Yes, the shapes of the histograms are the same.

11. (a)

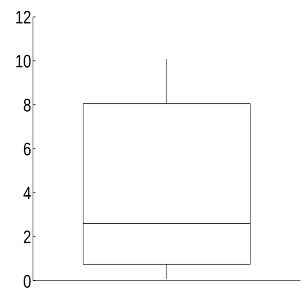


(b) Yes. The value 100 is an outlier.

13. 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 resistances is greater than 5Ω. The answer is (ii).

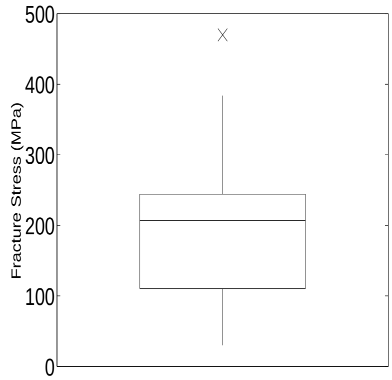
15. (a) IQR = 3rd quartile − 1st quartile. A: IQR = 6.02 − 1.42 = 4.60, B: IQR = 9.13 − 5.27 = 3.86

(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.



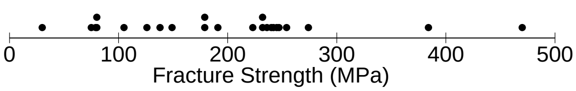
(c) No. The minimum value of -2.235 is an "outlier," since it is more than 1.5 times the interquartile range below the first quartile. The lower whisker should extend to the smallest point that is not an outlier, but the value of this point is not given.

17. (a)



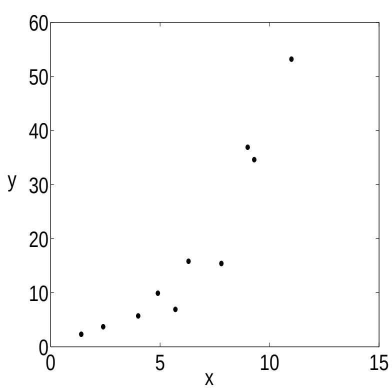
(b) The boxplot indicates that the value 470 is an outlier.

(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.

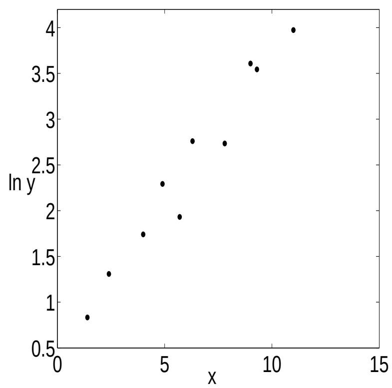
19. (a)



The relationship is non-linear.

(b)

x	1.4	2.4	4.0	4.9	5.7	6.3	7.8	9.0	9.3	11.0
$\ln y$	0.83	1.31	1.74	2.29	1.93	2.76	2.73	3.61	3.54	3.97



The relationship is approximately linear.

(c) It would be easier to work with x and $\ln y$, because the relationship is approximately linear.

Supplementary Exercises for Chapter 1

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

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

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) Incorrect, the total area is greater than 1.

(b) Correct. The total area is equal to 1.

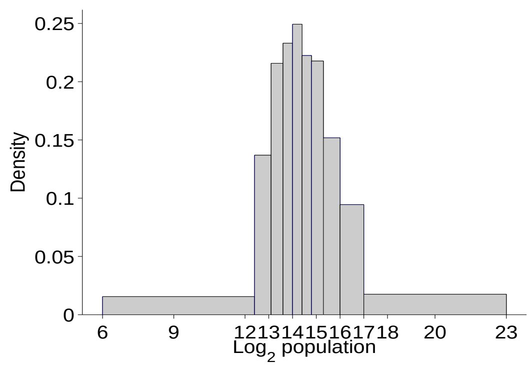
(c) Incorrect. The total area is less than 1.

(d) Correct. The total area is equal to 1.

13. (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.

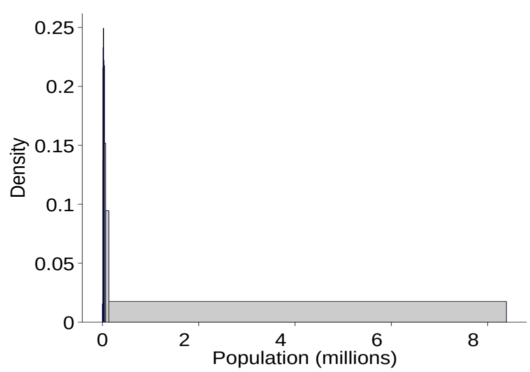
15. (a)



(b) 0.14

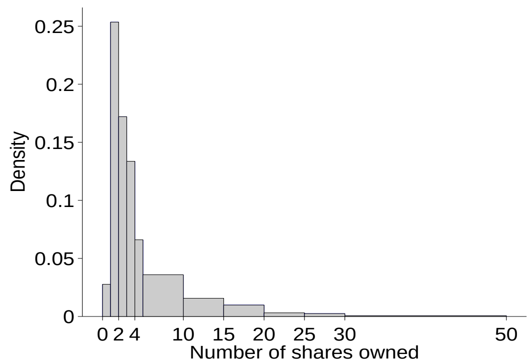
(c) Approximately symmetric

(d)



The data on the raw scale are skewed so much to the right that it is impossible to see the features of the histogram.

17. (a)

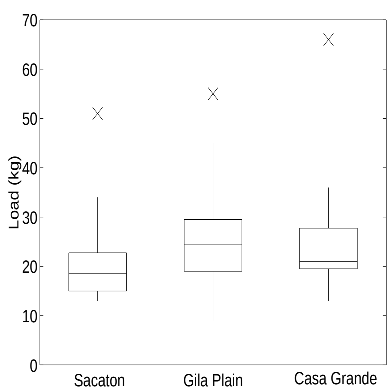


(b) The sample size is 651, so the median is approximated by the point at which the area to the left is 0.5 = 325.5/651. The area to the left of 3 is 295/651, and the area to the left of 4 is 382/651. The point at which the area to the left is 325.5/651 is $3+(325.5-295)/(382-295) = 3.35$.

(c) The sample size is 651, so the first quartile is approximated by the point at which the area to the left is 0.25 = 162.75/651. The area to the left of 1 is 18/651, and the area to the left of 2 is 183/651. The point at which the area to the left is 162.75/651 is $1+(162.75-18)/(183-18) = 1.88$.

(d) The sample size is 651, so the third quartile is approximated by the point at which the area to the left is 0.75 = 488.25/651. The area to the left of 5 is 425/651, and the area to the left of 10 is 542/651. The point at which the area to the left is 488.25/651 is $5+(10-5)(488.25-425)/(542-425) = 7.70$.

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

3. (a) The outcomes are the 16 different strings of 4 true-false answers. These are $\{$TTTT, TTTF, TTFT, TTFF, TFTT, TFTF, TFFT, TFFF, FTTT, FTTF, FTFT, FTFF, FFTT, FFTF, FFFT, FFFF$\}$.

   (b) There are 16 equally likely outcomes. The answers are all the same in two of them, TTTT and FFFF. Therefore the probability is $2/16$ or $1/8$.

   (c) There are 16 equally likely outcomes. There are four of them, TFFF, FTFF, FFTF, and FFFT, for which exactly one answer is "True." Therefore the probability is $4/16$ or $1/4$.

   (d) There are 16 equally likely outcomes. There are five of them, TFFF, FTFF, FFTF, FFFT, and FFFF, for which at most one answer is "True." Therefore the probability is $5/16$.

5. (a) The outcomes are the sequences of candidates that end with either #1 or #2. These are $\{$1, 2, 31, 32, 41, 42, 341, 342, 431, 432$\}$.

   (b) A = $\{$1, 2$\}$

   (c) B = $\{$341, 342, 431, 432$\}$

   (d) C = $\{$31, 32, 341, 342, 431, 432$\}$

   (e) D = $\{$1, 31, 41, 341, 431$\}$

   (f) A and E are mutually exclusive because they have no outcomes in commom.
   B and E are not mutually exclusive because they both contain the outcomes 341, 342, 431, and 432.
   C and E are not mutually exclusive because they both contain the outcomes 341, 342, 431, and 432.
   D and E are not mutually exclusive because they both contain the outcomes 41, 341, and 431.

7. (a)
$$\begin{aligned} P(\text{living room or den}) &= P(\text{living room}) + P(\text{den}) \\ &= 0.26 + 0.22 \\ &= 0.48 \end{aligned}$$

(b) $$\begin{aligned} P(\text{not bedroom}) &= 1 - P(\text{bedroom}) \\ &= 1 - 0.37 \\ &= 0.63 \end{aligned}$$

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

11. (a) False

(b) True

13. (a) $$\begin{aligned} P(S \cup C) &= P(S) + P(C) - P(S \cap C) \\ &= 0.4 + 0.3 - 0.2 \\ &= 0.5 \end{aligned}$$

(b) $P(S^c \cap C^c) = 1 - P(S \cup C) = 1 - 0.5 = 0.5.$

(c) We need to find $P(S \cap C^c)$. Now $P(S) = P(S \cap C) + P(S \cap C^c)$ (this can be seen from a Venn diagram). Now

$$\begin{aligned} P(S \cap C) &= P(S) + P(C) - P(S \cup C) \\ &= 0.4 + 0.3 - 0.5 \\ &= 0.2 \end{aligned}$$

Since $P(S) = 0.4$ and $P(S \cap C) = 0.2$, $P(S \cap C^c) = 0.2$.

15. (a) Let R be the event that a student is proficient in reading, and let M be the event that a student is proficient in mathematics. We need to find $P(R^c \cap M)$. Now $P(M) = P(R \cap M) + P(R^c \cap M)$ (this can be seen from a Venn diagram). We know that $P(M) = 0.78$ and $P(R \cap M) = 0.65$. Therefore $P(R^c \cap M) = 0.13$.

(b) We need to find $P(R \cap M^c)$. Now $P(R) = P(R \cap M) + P(R \cap M^c)$ (this can be seen from a Venn diagram). We know that $P(R) = 0.85$ and $P(R \cap M) = 0.65$. Therefore $P(R \cap M^c) = 0.20$.

(c) First we compute $P(R \cup M)$:
$P(R \cup M) = P(R) + P(M) - P(R \cap M) = 0.85 + 0.78 - 0.65 = 0.98.$
Now $P(R^c \cap M^c) = 1 - P(R \cup M) = 1 - 0.98 = 0.02.$

17.
$$\begin{aligned} P(A \cap B) &= P(A) + P(B) - P(A \cup B) \\ &= 0.98 + 0.95 - 0.99 \\ &= 0.94 \end{aligned}$$

19. (a) False

(b) True

(c) False

(d) True

Section 2.2

1. (a) $(4)(4)(4) = 64$

(b) $(2)(2)(2) = 8$

(c) $(4)(3)(2) = 24$

3. $$\binom{8}{4} = \frac{8!}{4!4!} = 70$$

5. (a) $(8)(7)(6) = 336$

(b) $\binom{8}{3} = \frac{8!}{3!5!} = 56$

7. $(2^{10})(4^5) = 1,048,576$

9. (a) $36^8 = 2.8211 \times 10^{12}$

(b) $36^8 - 26^8 = 2.6123 \times 10^{12}$

(c) $\frac{36^8 - 26^8}{36^8} = 0.9260$

11.
$$\begin{aligned} P(\text{match}) &= P(BB) + P(WW) \\ &= (8/14)(4/6) + (6/14)(2/6) \\ &= 44/84 = 0.5238 \end{aligned}$$

Section 2.3

1. A and B are independent if $P(A \cap B) = P(A)P(B)$. Therefore $P(B) = 0.25$.

3. (a) $5/15$

(b) Given that the first resistor is 50Ω, there are 14 resistors remaining of which 5 are 100Ω. Therefore $P(\text{2nd is } 100\Omega | \text{1st is } 50\Omega) = 5/14$.

(c) Given that the first resistor is 100Ω, there are 14 resistors remaining of which 4 are 100Ω. Therefore $P(\text{2nd is } 100\Omega | \text{1st is } 100\Omega) = 4/14$.

5. Given that a student is an engineering major, it is almost certain that the student took a calculus course. Therefore $P(B|A)$ is close to 1. Given that a student took a calculus course, it is much less certain that the student is an engineering major, since many non-engineering majors take calculus. Therefore $P(A|B)$ is much less than 1, so $P(B|A) > P(A|B)$.

7. 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)
$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(A)P(B) \\ &= 0.2 + 0.15 - (0.2)(0.15) \\ &= 0.32 \end{aligned}$$

9. Let V denote the event that a person buys a hybrid vehicle, and let T denote the event that a person buys a hybrid truck. Then

$$\begin{aligned} P(T\,|V) &= \frac{P(T \cap V)}{P(V)} \\ &= \frac{P(T)}{P(V)} \\ &= \frac{0.05}{0.12} \\ &= 0.417 \end{aligned}$$

11. Let OK denote the event that a valve meets the specification, let R denote the event that a valve is reground, and let S denote the event that a valve is scrapped. Then $P(OK \cap R^c) = 0.7$, $P(R) = 0.2$, $P(S \cap R^c) = 0.1$, $P(OK|R) = 0.9$, $P(S|R) = 0.1$.

(a) $P(R^c) = 1 - P(R) = 1 - 0.2 = 0.8$

(b) $P(S|R^c) = \dfrac{P(S \cap R^c)}{P(R^c)} = \dfrac{0.1}{0.8} = 0.125$

(c)
$$\begin{aligned} P(S) &= P(S \cap R^c) + P(S \cap R) \\ &= P(S \cap R^c) + P(S|R)P(R) \\ &= 0.1 + (0.1)(0.2) \\ &= 0.12 \end{aligned}$$

(d)
$$\begin{aligned} P(R|S) &= \frac{P(S \cap R)}{P(S)} \\ &= \frac{P(S|R)P(R)}{P(S)} \\ &= \frac{(0.1)(0.2)}{0.12} \\ &= 0.167 \end{aligned}$$

(e)
$$\begin{aligned} P(OK) &= P(OK \cap R^c) + P(OK \cap R) \\ &= P(OK \cap R^c) + P(OK|R)P(R) \\ &= 0.7 + (0.9)(0.2) \\ &= 0.88 \end{aligned}$$

(f)
$$\begin{aligned} P(R|OK) &= \frac{P(R \cap OK)}{P(OK)} \\ &= \frac{P(OK|R)P(R)}{P(OK)} \\ &= \frac{(0.9)(0.2)}{0.88} \\ &= 0.205 \end{aligned}$$

(g)
$$\begin{aligned} P(R^c|OK) &= \frac{P(R^c \cap OK)}{P(OK)} \\ &= \frac{0.7}{0.88} \\ &= 0.795 \end{aligned}$$

13. Let $T1$ denote the event that the first device is triggered, and let $T2$ denote the event that the second device is triggered. Then $P(T1) = 0.9$ and $P(T2) = 0.8$.

(a)
$$\begin{aligned} P(T1 \cup T2) &= P(T1) + P(T2) - P(T1 \cap T2) \\ &= P(T1) + P(T2) - P(T1)P(T2) \\ &= 0.9 + 0.8 - (0.9)(0.8) \\ &= 0.98 \end{aligned}$$

(b) $P(T1^c \cap T2^c) = P(T1^c)P(T2^c) = (1-0.9)(1-0.8) = 0.02$

(c) $P(T1 \cap T2) = P(T1)P(T2) = (0.9)(0.8) = 0.72$

(d) $P(T1 \cap T2^c) = P(T1)P(T2^c) = (0.9)(1-0.8) = 0.18$

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

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

(c) $\dfrac{88+165}{88+65+260} = 0.4932$

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

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

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

(c) $P(\text{Gene 2 dominant} \mid \text{Gene 1 dominant}) \quad = \quad \dfrac{P(\text{Gene 1 dominant} \cap \text{Gene 2 dominant})}{P(\text{Gene 1 dominant})}$

$$
\begin{aligned}
&= \frac{56/100}{0.8} \\
&= 0.7
\end{aligned}
$$

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

19. Let R, D, and I denote the events that the senator is a Republian, Democrat, or Independent, respectively, and let M and F denote the events that the senator is male or female, respectively.

(a) $P(R \cap M) = 0.41$

(b)
$$
\begin{aligned}
P(D \cup F) &= P(D) + P(F) - P(D \cap F) \\
&= (0.37 + 0.16) + (0.16 + 0.04) - 0.16 \\
&= 0.57
\end{aligned}
$$

(c)
$$
\begin{aligned}
P(R) &= P(R \cap M) + P(R \cap F) \\
&= 0.41 + 0.04 \\
&= 0.45
\end{aligned}
$$

(d) $P(R^c) = 1 - P(R) = 1 - 0.45 = 0.55$

(e)
$$
\begin{aligned}
P(D) &= P(D \cap M) + P(D \cap F) \\
&= 0.37 + 0.16 \\
&= 0.53
\end{aligned}
$$

(f)
$$
\begin{aligned}
P(I) &= P(I \cap M) + P(I \cap F) \\
&= 0.02 + 0 \\
&= 0.02
\end{aligned}
$$

(g) $P(D \cup I) = P(D) + P(I) = 0.53 + 0.02 = 0.55$

21. (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 $P(\text{second gauge fails}|\text{first gauge fails}) > P(\text{second gauge fails})$.

23. (a) $P(A) = 3/10$

(b) Given that A occurs, there are 9 components remaining, of which 2 are defective.
Therefore $P(B|A) = 2/9$.

(c) $P(A \cap B) = P(A)P(B|A) = (3/10)(2/9) = 1/15$

(d) Given that A^c occurs, there are 9 components remaining, of which 3 are defective.
Therefore $P(B|A) = 3/9$. Now $P(A^c \cap B) = P(A^c)P(B|A^c) = (7/10)(3/9) = 7/30$.

(e) $P(B) = P(A \cap B) + P(A^c \cap B) = 1/15 + 7/30 = 3/10$

(f) No. $P(B) \neq P(B|A)$ [or equivalently, $P(A \cap B) \neq P(A)P(B)$].

25. $n = 10,000$. The two components are a simple random sample from the population. When the population is large, the items in a simple random sample are nearly independent.

27. Let R denote the event of a rainy day, and let C denote the event that the forecast is correct. Then $P(R) = 0.1$, $P(C|R) = 0.8$, and $P(C|R^c) = 0.9$.

(a)
$$\begin{aligned} P(C) &= P(C|R)P(R) + P(C|R^c)P(R^c) \\ &= (0.8)(0.1) + (0.9)(1 - 0.1) \\ &= 0.89 \end{aligned}$$

(b) A forecast of no rain will be correct on every non-rainy day. Therefore the probability is 0.9.

29. Let F denote the event that an item has a flaw. Let A denote the event that a flaw is detected by the first inspector, and let B denote the event that the flaw is detected by the second inspector.

(a)
$$\begin{aligned} P(F|A^c) &= \frac{P(A^c|F)P(F)}{P(A^c|F)P(F)+P(A^c|F^c)P(F^c)} \\ &= \frac{(0.1)(0.1)}{(0.1)(0.1)+(1)(0.9)} \\ &= 0.011 \end{aligned}$$

(b)
$$\begin{aligned} P(F|A^c \cap B^c) &= \frac{P(A^c \cap B^c|F)P(F)}{P(A^c \cap B^c|F)P(F)+P(A^c \cap B^c|F^c)P(F^c)} \\ &= \frac{P(A^c|F)P(B^c|F)P(F)}{P(A^c|F)P(B^c|F)P(F)+P(A^c|F^c)P(B^c|F^c)P(F^c)} \\ &= \frac{(0.1)(0.3)(0.1)}{(0.1)(0.3)(0.1)+(1)(1)(0.9)} \\ &= 0.0033 \end{aligned}$$

31. (a) Each child has probability 0.25 of having the disease. Since the children are independent, the probability that both are disease-free is $0.75^2 = 0.5625$.

(b) Each child has probability 0.5 of being a carrier. Since the children are independent, the probability that both are carriers is $0.5^2 = 0.25$.

(c) Let D denote the event that a child has the disease, and let C denote the event that the child is a carrier. Then $P(D) = 0.25$, $P(C) = 0.5$, and $P(D^c) = 0.75$. We first compute $P(C\,|\,D^c)$, the probability that a child who does not have the disease is a carrier. First, $P(C \cap D^c) = P(C) = 0.5$. Now

$$P(C\,|\,D^c) = \frac{P(C \cap D^c)}{P(D^c)} = \frac{0.5}{0.75} = \frac{2}{3}$$

Since children are independent, the probability that both children are carriers given that neither has the disease is $(2/3)^2 = 4/9$.

(d) Let W_D denote the event that the woman has the disease, Let W_C denote the event that the woman is a carrier, and let W_F denote the event that the woman does not have the disease and is not a carrier. Then $P(W_D) = 0.25$,

$P(W_C) = 0.5$, and $P(W_F) = 0.25$, Let C_D denote the event that the child has the disease.

$$\begin{aligned}
P(C_D) &= P(C_D \cap W_D) + P(C_D \cap W_C) + P(C_D \cap W_F) \\
&= P(C_D \,|\, W_D)P(W_D) + P(C_D \,|\, W_C)P(W_C) + P(C_D \,|\, W_F)P(W_F) \\
&= (0.5)(0.25) + (0.25)(0.5) + (0)(0.25) \\
&= 0.25
\end{aligned}$$

33. Let D represent the event that the man actually has the disease, and let $+$ represent the event that the test gives a positive signal.

We are given that $P(D) = 0.005$, $P(+|D) = 0.99$, and $P(+|D^c) = 0.01$.

It follows that $P(D^c) = 0.995$, $P(-|D) = 0.01$, and $P(-|D^c) = 0.99$.

(a) $$\begin{aligned}
P(D|-) &= \frac{P(-|D)P(D)}{P(-|D)P(D) + P(-|D^c)P(D^c)} \\
&= \frac{(0.01)(0.005)}{(0.01)(0.005) + (0.99)(0.995)} \\
&= 5.08 \times 10^{-5}
\end{aligned}$$

(b) $P(++|D) = 0.99^2 = 0.9801$

(c) $P(++|D^c) = 0.01^2 = 0.0001$

(d) $$\begin{aligned}
P(D|++) &= \frac{P(++|D)P(D)}{P(++|D)P(D) + P(++|D^c)P(D^c)} \\
&= \frac{(0.9801)(0.005)}{(0.9801)(0.005) + (0.0001)(0.995)} \\
&= 0.9801
\end{aligned}$$

35. $P(\text{system functions}) = P[(A \cap B) \cap (C \cup D)]$. 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

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

37. 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, $n = 7$.

Section 2.4

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)

x	1	2	3	4	5
p(x)	0.70	0.15	0.10	0.03	0.02

(b) $P(X \le 2) = P(X = 1) + P(X = 2) = 0.70 + 0.15 = 0.85$

(c) $P(X > 3) = P(X = 4) + P(X = 5) = 0.03 + 0.02 = 0.05$

(d) $\mu_X = 1(0.70) + 2(0.15) + 3(0.10) + 4(0.03) + 5(0.02) = 1.52$

(e) $\sigma_X = \sqrt{1^2(0.70) + 2^2(0.15) + 3^2(0.10) + 4^2(0.03) + 5^2(0.02) - 1.52^2} = 0.9325$

7. (a) $\sum_{x=1}^{5} cx = 1$, so $c(1+2+3+4+5) = 1$, so $c = 1/15$.

(b) $P(X = 2) = c(2) = 2/15 = 0.2$

(c) $\mu_X = \sum_{x=1}^{5} xP(X = x) = \sum_{x=1}^{5} x^2/15 = (1^2 + 2^2 + 3^2 + 4^2 + 5^2)/15 = 11/3$

(d) $\sigma_X^2 = \sum_{x=1}^{5} (x - \mu_X)^2 P(X = x) = \sum_{x=1}^{5} x(x - 11/3)^2/15 = (64/135) + 2(25/135) + 3(4/135) + 4(1/135) + 5(16/135) = 14/9$

Alternatively, $\sigma_X^2 = \sum_{x=1}^{5} x^2 P(X=x) - \mu_X^2 = \sum_{x=1}^{5} x^3/15 - (11/3)^2 = (1^3+2^3+3^3+4^3+5^3)/15 - (11/3)^2 = 14/9$

(e) $\sigma_X = \sqrt{14/9} = 1.2472$

9. (a)

x	$p_1(x)$
0	0.2
1	0.16
2	0.128
3	0.1024
4	0.0819
5	0.0655

(b)

x	$p_2(x)$
0	0.4
1	0.24
2	0.144
3	0.0864
4	0.0518
5	0.0311

(c) $p_2(x)$ appears to be the better model. Its probabilities are all fairly close to the proportions of days observed in the data. In contrast, the probabilities of 0 and 1 for $p_1(x)$ are much smaller than the observed proportions.

(d) No, this is not right. The data are a simple random sample, and the model represents the population. Simple random samples generally do not reflect the population exactly.

11. Let A denote an acceptable chip, and U an unacceptable one.

(a) If the first two chips are both acceptable, then $Y = 2$. This is the smallest possible value.

(b) $P(Y=2) = P(AA) = (0.9)^2 = 0.81$

(c) $P(Y=3|X=1) = \dfrac{P(Y=3 \text{ and } X=1)}{P(X=1)}$.

Now $P(Y=3 \text{ and } X=1) = P(AUA) = (0.9)(0.1)(0.9) = 0.081$, and $P(X=1) = P(A) = 0.9$.

Therefore $P(Y=3|X=1) = 0.081/0.9 = 0.09$.

(d) $P(Y=3|X=2) = \dfrac{P(Y=3 \text{ and } X=2)}{P(X=2)}$.

Now $P(Y=3 \text{ and } X=2) = P(UAA) = (0.1)(0.9)(0.9) = 0.081$, and
$P(X=2) = P(UA) = (0.1)(0.9) = 0.09$.
Therefore $P(Y=3|X=2) = 0.081/0.09 = 0.9$.

(e) If $Y=3$ the only possible values for X are $X=1$ and $X=2$.
Therefore

$$\begin{aligned} P(Y=3) &= P(Y=3|X=1)P(X=1) + P(Y=3|X=2)P(X=2) \\ &= (0.09)(0.9) + (0.9)(0.09) \\ &= 0.162 \end{aligned}$$

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

15. (a) $\displaystyle \mu = \int_0^{\infty} 0.1te^{-0.1t}\,dt$

$$
\begin{aligned}
&= -te^{-0.1t}\Big|_0^{\infty} - \int_0^{\infty} -e^{-0.1t}\,dt \\
&= 0 - 10e^{-0.1t}\Big|_0^{\infty} \\
&= 10
\end{aligned}
$$

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

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

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

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

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

With the process in Exercise 16, the probability is

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

Therefore this process is better than the one in Exercise 16.

19. (a) $P(X > 3) = \int_3^4 (3/64)x^2(4-x)\,dx = \left(\dfrac{x^3}{16} - \dfrac{3x^4}{256}\right)\Bigg|_3^4 = 67/256$

(b) $P(2 < X < 3) = \int_2^3 (3/64)x^2(4-x)\,dx = \left(\dfrac{x^3}{16} - \dfrac{3x^4}{256}\right)\Bigg|_2^3 = 109/256$

(c) $\mu = \int_0^4 (3/64)x^3(4-x)\,dx = \left(\dfrac{3x^4}{64} - \dfrac{3x^5}{320}\right)\Bigg|_0^4 = 2.4$

(d) $\sigma^2 = \int_0^4 (3/64)x^4(4-x)\,dx - \mu^2 = \left(\dfrac{3x^5}{80} - \dfrac{x^6}{128}\right)\Bigg|_0^4 - 2.4^2 = 0.64$

(e) $F(x) = \int_{-\infty}^x f(t)\,dt$
If $x \le 0$, $F(x) = \int_{-\infty}^x 0\,dt = 0$
If $0 < x < 4$, $F(x) = \int_0^x (3/64)t^2(4-t)\,dt = (16x^3 - 3x^4)/256$
If $x \ge 3$, $F(x) = \int_0^4 (3/64)t^2(4-t)\,dt = 1$

21. (a) $P(X < 0.2) = \int_0^{0.2} 12(x^2 - x^3)\,dx = 4x^3 - 3x^4 \Bigg|_0^{0.2} = 0.0272$

(b) $\mu = \int_0^1 12x(x^2 - x^3)\,dx = 3x^4 - \dfrac{12}{5}x^5 \Bigg|_0^1 = 0.6$

(c) The variance is

$$\begin{aligned}
\sigma^2 &= \int_0^1 12x^2(x^2 + x^3)\,dx - \mu^2 \\
&= \left(\frac{12}{5}x^5 + 2x^6\right)\Bigg|_0^1 - 0.6^2 \\
&= 0.04
\end{aligned}$$

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

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

If $0 < x \le 1$, $F(x) = \int_{0}^{x} 12(t^2 - t^3)\,dt = 4t^3 - 3t^4 \Big|_0^x = 4x^3 - 3x^4$

If $x > 1$, $F(x) = \int_{0}^{1} 12(t^2 - t^3)\,dt = 1$

(e) $P(0 < X < 0.8) = \int_{0}^{0.8} 12(x^2 - x^3)\,dx = \left(4x^3 - 3x^4\right) \Big|_0^{0.8} = 0.81922$

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

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

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

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

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

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

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

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

Section 2.5

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) $\mu_{X+4Y} = \mu_X + 4\mu_Y = 9.5 + 4(6.8) = 36.7$

$\sigma_{X+4Y} = \sqrt{\sigma_X^2 + 4^2\sigma_Y^2} = \sqrt{0.4^2 + 16(0.1^2)} = 0.566$

3. Let $X_1, \ldots, X_4$ be the lifetimes of the four transistors. Let $S = X_1 + \cdots + X_4$ be the total lifetime.

$\mu_S = \sum \mu_{X_i} = 4(900) = 3600$

$\sigma_S = \sqrt{\sum \sigma^2_{X_i}} = \sqrt{4(30^2)} = 60$

5. Let $X_1, \ldots, X_5$ be the thicknesses of the five layers. Let $S = X_1 + \cdots + X_5$ be the total thickness.

(a) $\mu_S = \sum \mu_{X_i} = 5(1.2) = 6.0$

(b) $\sigma_S = \sqrt{\sum \sigma^2_{X_i}} = \sqrt{5(0.04^2)} = 0.0894$

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

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

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

11. (a) The number of passenger-miles is 8000(210) = 1,680,000. Let X be the number of gallons of fuel used. Then $\mu_X = 1{,}680{,}000(0.15) = 252{,}000$ gallons.

(b) $\sigma_X = \sqrt{(1{,}680{,}000)(0.01)} = 12.9615$

(c) $\mu_{X/1{,}680{,}000} = (1/1{,}680{,}000)\mu_X = (1/1{,}680{,}000)(252{,}000) = 0.15.$

(d) $\sigma_{X/1{,}680{,}000} = (1/1{,}680{,}000)\sigma_X = (1/1{,}680{,}000)(12.9615) = 7.7152 \times 10^{-6}$

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

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

15. (a) $P(X < 9.98) = \int_{9.95}^{9.98} 10\,dx = 10x \Big|_{9.95}^{9.98} = 0.3$

(b) $P(Y > 5.01) = \int_{5.01}^{5.1} 5\,dy = 5y \Big|_{5.01}^{5.1} = 0.45$

(c) Since X and Y are independent,
$P(X < 9.98 \text{ and } Y > 5.01) = P(X < 9.98)P(Y > 5.01) = (0.3)(0.45) = 0.135$

(d) $\mu_X = \int_{9.95}^{10.05} 10x\,dx = 5x^2 \Big|_{9.95}^{10.05} = 10$

(e) $\mu_Y = \int_{4.9}^{5.1} 5y\,dy = 2.5y^2 \Big|_{4.9}^{5.1} = 5$

17. (a) Let $\mu = 40.25$ be the mean SiO_2 content, and let $\sigma = 0.36$ be the standard deviation of the SiO_2 content, in a randomly chosen rock. Let $\overline{X}$ be the average content in a random sample of 10 rocks.

Then $\mu_{\overline{X}} = \mu = 40.25$, and $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{10}} = \frac{0.36}{\sqrt{10}} = 0.11$.

(b) Let n be the required number of rocks. Then $\frac{\sigma}{\sqrt{n}} = \frac{0.36}{\sqrt{n}} = 0.05$.

Solving for n yields $n = 51.84$. Since n must be an integer, take $n = 52$.

Section 2.6

1. (a) 0.17

   (b) $P(X \geq 1 \text{ and } Y < 2) = P(1,0)+P(1,1)+P(2,0)+P(2,1) = 0.17+0.23+0.06+0.14 = 0.60$

   (c) $P(X < 1) = P(X = 0) = P(0,0)+P(0,1)+P(0,2) = 0.10+0.11+0.05 = 0.26$

   (d) $P(Y \geq 1) = 1-P(Y = 0) = 1-P(0,0)-P(1,0)-P(2,0) = 1-0.10-0.17-0.06 = 0.67$

   (e) $P(X \geq 1) = 1-P(X = 0) = 1-P(0,0)-P(0,1)-P(0,2) = 1-0.10-0.11-0.05 = 0.74$

   (f) $P(Y = 0) = P(0,0)+P(1,0)+P(2,0) = 0.10+0.17+0.06 = 0.33$

   (g) $P(X = 0 \text{ and } Y = 0) = 0.10$

3. (a) $p_{Y|X}(0\,|\,0) = \dfrac{p_{X,Y}(0,0)}{p_X(0)} = \dfrac{0.10}{0.26} = 0.3846$

   $p_{Y|X}(1\,|\,0) = \dfrac{p_{X,Y}(0,1)}{p_X(0)} = \dfrac{0.11}{0.26} = 0.4231$

   $p_{Y|X}(2\,|\,0) = \dfrac{p_{X,Y}(0,2)}{p_X(0)} = \dfrac{0.05}{0.26} = 0.1923$

   (b) $p_{X|Y}(0\,|\,1) = \dfrac{p_{X,Y}(0,1)}{p_Y(1)} = \dfrac{0.11}{0.48} = 0.2292$

   $p_{X|Y}(1\,|\,1) = \dfrac{p_{X,Y}(1,1)}{p_Y(1)} = \dfrac{0.23}{0.48} = 0.4792$

   $p_{X|Y}(2\,|\,1) = \dfrac{p_{X,Y}(2,1)}{p_Y(1)} = \dfrac{0.14}{0.48} = 0.2917$

   (c) $E(Y\,|\,X = 0) = 0p_{Y|X}(0\,|\,0)+1p_{Y|X}(1\,|\,0)+2p_{Y|X}(2\,|\,0) = 0.8077$

   (d) $E(X\,|\,Y = 1) = 0p_{X|Y}(0\,|\,1)+1p_{X|Y}(1\,|\,1)+2p_{X|Y}(2\,|\,1) = 1.0625$

5. (a) $\mu_{X+Y} = \mu_X + \mu_Y = 0.89 + 1.22 = 2.11$

(b) $\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2 + 2\text{Cov}(X,Y)} = \sqrt{1.0379 + 1.1916 + 2(-0.1158)} = 1.4135$

(c) $P(X+Y=3) = P(0,3) + P(1,2) + P(2,1) + P(3,0) = 0.10 + 0.05 + 0.05 + 0.04 = 0.24$

7. (a) $2X + 3Y$

(b) $\mu_{2X+3Y} = 2\mu_X + 3\mu_Y = 2(0.89) + 3(1.22) = 5.44$

(c)
$$\begin{aligned}\sigma_{2X+3Y} &= \sqrt{2^2\sigma_X^2 + 3^2\sigma_Y^2 + 2(2)(3)\text{Cov}(X,Y)} \\ &= \sqrt{2^2(1.0379) + 3^2(1.1916) - 2(2)(3)(-0.1158)} \\ &= 3.6724\end{aligned}$$

9. (a) The marginal probability mass function $p_X(x)$ is found by summing along the rows of the joint probability mass function.

			y			
x	0	1	2	3	4	$p_X(x)$
0	0.06	0.03	0.01	0.00	0.00	0.10
1	0.06	0.08	0.04	0.02	0.00	0.20
2	0.04	0.05	0.12	0.06	0.03	0.30
3	0.00	0.03	0.07	0.09	0.06	0.25
4	0.00	0.00	0.02	0.06	0.07	0.15
$p_Y(y)$	0.16	0.19	0.26	0.23	0.16	

$p_X(0) = 0.10$, $p_X(1) = 0.20$, $p_X(2) = 0.30$, $p_X(3) = 0.25$, $p_X(4) = 0.15$, $p_X(x) = 0$ if $x \neq 0, 1, 2, 3,$ or 4.

(b) The marginal probability mass function $p_Y(y)$ is found by summing down the columns of the joint probability mass function. So $p_Y(0) = 0.16$, $p_Y(1) = 0.19$, $p_Y(2) = 0.26$, $p_Y(3) = 0.23$, $p_Y(4) = 0.16$, $p_Y(y) = 0$ if $y \neq 0, 1, 2, 3,$ or 4.

(c) No. The joint probability mass function is not equal to the product of the marginals. For example, $p_{X,Y}(0,0) = 0.06 \neq p_X(0)p_Y(0)$.

(d) $\mu_X = 0p_X(0)+1p_X(1)+2p_X(2)+3p_X(3)+4p_X(4) = 0(0.10)+1(0.20)+2(0.30)+3(0.25)+4(0.15) = 2.15$

$\mu_Y = 0p_Y(0)+1p_Y(1)+2p_Y(2)+3p_Y(3)+4p_Y(4) = 0(0.16)+1(0.19)+2(0.26)+3(0.23)+4(0.16) = 2.04$

(e)
$$\begin{aligned}\sigma_X^2 &= 0^2p_X(0)+1^2p_X(1)+2^2p_X(2)+3^2p_X(3)+4^2p_X(4)-\mu_X^2\\ &= 0^2(0.10)+1^2(0.20)+2^2(0.30)+3^2(0.25)+4^2(0.15)-2.15^2\\ &= 1.4275\end{aligned}$$

$\sigma_X = \sqrt{1.4275} = 1.1948$

$$\begin{aligned}\sigma_Y^2 &= 0^2p_Y(0)+1^2p_Y(1)+2^2p_Y(2)+3^2p_Y(3)+4^2p_Y(4)-\mu_Y^2\\ &= 0^2(0.16)+1^2(0.19)+2^2(0.26)+3^2(0.23)+4^2(0.16)-2.04^2\\ &= 1.6984\end{aligned}$$

$\sigma_Y = \sqrt{1.6984} = 1.3032$

(f) $\text{Cov}(X,Y) = \mu_{XY} - \mu_X\mu_Y$.

$$\begin{aligned}\mu_{XY} &= (0)(0)p_{X,Y}(0,0)+(0)(1)p_{X,Y}(0,1)+(0)(2)p_{X,Y}(0,2)+(0)(3)p_{X,Y}(0,3)+(0)(4)p_{X,Y}(0,4)\\ &\quad +(1)(0)p_{X,Y}(1,0)+(1)(1)p_{X,Y}(1,1)+(1)(2)p_{X,Y}(1,2)+(1)(3)p_{X,Y}(1,3)+(1)(4)p_{X,Y}(1,4)\\ &\quad +(2)(0)p_{X,Y}(2,0)+(2)(1)p_{X,Y}(2,1)+(2)(2)p_{X,Y}(2,2)+(2)(3)p_{X,Y}(2,3)+(2)(4)p_{X,Y}(2,4)\\ &\quad +(3)(0)p_{X,Y}(3,0)+(3)(1)p_{X,Y}(3,1)+(3)(2)p_{X,Y}(3,2)+(3)(3)p_{X,Y}(3,3)+(3)(4)p_{X,Y}(3,4)\\ &\quad +(4)(0)p_{X,Y}(4,0)+(4)(1)p_{X,Y}(4,1)+(4)(2)p_{X,Y}(4,2)+(4)(3)p_{X,Y}(4,3)+(4)(4)p_{X,Y}(4,4)\\ &= (0)(0)(0.06)+(0)(1)(0.03)+(0)(2)(0.01)+(0)(3)(0.00)+(0)(4)(0.00)\\ &\quad +(1)(0)(0.06)+(1)(1)(0.08)+(1)(2)(0.04)+(1)(3)(0.02)+(1)(4)(0.00)\\ &\quad +(2)(0)(0.04)+(2)(1)(0.05)+(2)(2)(0.12)+(2)(3)(0.06)+(2)(4)(0.03)\\ &\quad +(3)(0)(0.00)+(3)(1)(0.03)+(3)(2)(0.07)+(3)(3)(0.09)+(3)(4)(0.06)\\ &\quad +(4)(0)(0.00)+(4)(1)(0.00)+(4)(2)(0.02)+(4)(3)(0.06)+(4)(4)(0.07)\\ &= 5.44\end{aligned}$$

$\mu_X = 2.15$, $\mu_Y = 2.04$

$\text{Cov}(X,Y) = 5.44 - (2.15)(2.04) = 1.0540$

(g) $\rho_{X,Y} = \dfrac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y} = \dfrac{1.0540}{(1.1948)(1.3032)} = 0.6769$

11. (a) $p_{Y|X}(0\,|\,4) = \dfrac{p_{X,Y}(4,0)}{p_X(4)} = \dfrac{0.00}{0.15} = 0$

$p_{Y|X}(1\,|\,4) = \dfrac{p_{X,Y}(4,1)}{p_X(4)} = \dfrac{0.00}{0.15} = 0$

$$p_{Y|X}(2\,|\,4) = \frac{p_{X,Y}(4,2)}{p_X(4)} = \frac{0.02}{0.15} = 2/15$$

$$p_{Y|X}(3\,|\,4) = \frac{p_{X,Y}(4,3)}{p_X(4)} = \frac{0.06}{0.15} = 2/5$$

$$p_{Y|X}(4\,|\,4) = \frac{p_{X,Y}(4,4)}{p_X(4)} = \frac{0.07}{0.15} = 7/15$$

(b) $p_{X|Y}(0\,|\,3) = \dfrac{p_{X,Y}(0,3)}{p_Y(3)} = \dfrac{0.00}{0.23} = 0$

$$p_{X|Y}(1\,|\,3) = \frac{p_{X,Y}(1,3)}{p_Y(3)} = \frac{0.02}{0.23} = 2/23$$

$$p_{X|Y}(2\,|\,3) = \frac{p_{X,Y}(2,3)}{p_Y(3)} = \frac{0.06}{0.23} = 6/23$$

$$p_{X|Y}(3\,|\,3) = \frac{p_{X,Y}(3,3)}{p_Y(3)} = \frac{0.09}{0.23} = 9/23$$

$$p_{X|Y}(4\,|\,3) = \frac{p_{X,Y}(4,3)}{p_Y(3)} = \frac{0.06}{0.23} = 6/23$$

(c) $E(Y\,|\,X=4) = 0p_{Y|X}(0\,|\,4) + 1p_{Y|X}(1\,|\,4) + 2p_{Y|X}(2\,|\,4) + 3p_{Y|X}(3\,|\,4) + 4p_{Y|X}(4\,|\,4) = 3.33$

(d) $E(X\,|\,Y=3) = 0p_{X|Y}(0\,|\,3) + 1p_{X|Y}(1\,|\,3) + 2p_{X|Y}(2\,|\,3) + 3p_{X|Y}(3\,|\,3) + 4p_{X|Y}(4\,|\,3) = 2.83$

13. (a) $\mu_Z = \mu_{X+Y} = \mu_X + \mu_Y = 1.01 + 1.23 = 2.24$

(b) $\sigma_Z = \sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2 + 2\text{Cov}(X,Y)} = \sqrt{0.8099 + 0.9971 + 2(0.2377)} = 1.511$

(c)
$$\begin{aligned} P(Z=2) &= P(X+Y=2) \\ &= P(X=0 \text{ and } Y=2) + P(X=1 \text{ and } Y=1) + P(X=2 \text{ and } Y=0) \\ &= 0.07 + 0.16 + 0.02 \\ &= 0.25 \end{aligned}$$

15. (a) $p_{Y|X}(0\,|\,3) = \dfrac{p_{X,Y}(3,0)}{p_X(3)} = \dfrac{0.01}{0.07} = 0.1429$

$$p_{Y|X}(1\,|\,3) = \frac{p_{X,Y}(3,1)}{p_X(3)} = \frac{0.02}{0.07} = 0.2858$$

$$p_{Y|X}(2\,|\,3) = \frac{p_{X,Y}(3,2)}{p_X(3)} = \frac{0.02}{0.07} = 0.2858$$

$$p_{Y|X}(3\,|\,3) = \frac{p_{X,Y}(3,3)}{p_X(3)} = \frac{0.02}{0.07} = 0.2858$$

(b) $p_{X|Y}(0\,|\,1) = \dfrac{p_{X,Y}(0,1)}{p_Y(1)} = \dfrac{0.10}{0.34} = 0.2941$

$$p_{X|Y}(1\,|\,1) = \frac{p_{X,Y}(1,1)}{p_Y(1)} = \frac{0.16}{0.34} = 0.4706$$

$$p_{X|Y}(2\,|\,1) = \frac{p_{X,Y}(2,1)}{p_Y(1)} = \frac{0.06}{0.34} = 0.1765$$

$$p_{X|Y}(3\,|\,1) = \frac{p_{X,Y}(3,1)}{p_Y(1)} = \frac{0.02}{0.34} = 0.0588$$

(c) $E(Y\,|\,X=3) = 0p_{Y|X}(0\,|\,3) + 1p_{Y|X}(1\,|\,3) + 2p_{Y|X}(2\,|\,3) + 3p_{Y|X}(3\,|\,3) = 1.71.$

(d) $E(X\,|\,Y=1) = 0p_{X|Y}(0\,|\,1) + 1p_{X|Y}(1\,|\,1) + 2p_{X|Y}(2\,|\,1) + 3p_{X|Y}(3\,|\,1) = 1$

17. (a) $\mathrm{Cov}(X,Y) = \mu_{XY} - \mu_X\mu_Y$

$$\begin{aligned}
\mu_{XY} &= \int_1^2\int_4^5 \frac{1}{6}xy(x+y)\,dy\,dx \\
&= \int_1^2 \frac{1}{6}\left(\frac{x^2y^2}{2} + \frac{xy^3}{3}\right)\Bigg|_4^5 dx \\
&= \int_1^2 \frac{1}{6}\left(\frac{9x^2}{2} + \frac{61x}{3}\right) dx \\
&= \frac{1}{6}\left(\frac{3x^3}{2} + \frac{61x^2}{6}\right)\Bigg|_1^2 \\
&= \frac{41}{6}
\end{aligned}$$

$f_X(x) = \dfrac{1}{6}\left(x + \dfrac{9}{2}\right)$ for $1 \le x \le 2$ (see Example 2.55).

$$\mu_X = \int_1^2 \frac{1}{6}x\left(x+\frac{9}{2}\right)dx = \frac{1}{6}\left(\frac{x^3}{3} + \frac{9x^2}{4}\right)\Bigg|_1^2 = \frac{109}{72}.$$

$f_Y(y) = \dfrac{1}{6}\left(y + \dfrac{3}{2}\right)$ for $4 \le y \le 5$ (see Example 2.55).

$$\mu_Y = \int_4^5 \frac{1}{6}y\left(y+\frac{3}{2}\right)\,dy = \frac{1}{6}\left(\frac{y^3}{3}+\frac{3y^2}{4}\right)\Bigg|_4^5 = \frac{325}{72}.$$

$$\mathrm{Cov}(X,Y) = \mu_{XY} - \mu_X\mu_Y = \frac{41}{6} - \left(\frac{109}{72}\right)\left(\frac{325}{72}\right) = -0.000193.$$

(b) $\rho_{X,Y} = \dfrac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}$.

$$\sigma_X^2 = \int_1^2 \frac{1}{6}x^2\left(x+\frac{9}{2}\right)\,dx - \mu_X^2 = \frac{1}{6}\left(\frac{x^4}{4}+\frac{3x^3}{2}\right)\Bigg|_1^2 - \left(\frac{109}{72}\right)^2 = 0.08314.$$

$$\sigma_Y^2 = \int_4^5 \frac{1}{6}y^2\left(y+\frac{3}{2}\right)\,dx - \mu_Y^2 = \frac{1}{6}\left(\frac{y^4}{4}+\frac{y^3}{2}\right)\Bigg|_4^5 - \left(\frac{325}{72}\right)^2 = 0.08314.$$

$$\rho_{X,Y} = \frac{-0.000193}{\sqrt{(0.08314)(0.08314)}} = -0.00232.$$

19. (a) $\mathrm{Cov}(X,Y) = \mu_{XY} - \mu_X\mu_Y$.

$$\begin{aligned}
\mu_{XY} &= \int_0^1\int_0^1 xy\frac{3(x^2+y^2)}{2}\,dx\,dy \\
&= \int_0^1 \left(\frac{3x^4y}{8}+\frac{3x^2y^3}{4}\right)\Bigg|_0^1 dy \\
&= \int_0^1 \left(\frac{3y}{8}+\frac{3y^3}{4}\right) dy \\
&= \left(\frac{3y^2}{16}+\frac{3y^4}{16}\right)\Bigg|_0^1 dy \\
&= \frac{3}{8}
\end{aligned}$$

$$\mu_X = \int_0^1 x\frac{1+3x^2}{2}\,dx = \left(\frac{x^2}{4}+\frac{3x^4}{8}\right)\Bigg|_0^1 = \frac{5}{8}.$$

$$\mu_Y = \int_0^1 y\frac{1+3y^2}{2}\,dy = \left(\frac{y^2}{4}+\frac{3y^4}{8}\right)\Bigg|_0^1 = \frac{5}{8}.$$

$$\mathrm{Cov}(X,Y) = \frac{3}{8} - \left(\frac{5}{8}\right)^2 = -\frac{1}{64}.$$

(b) $\rho_{X,Y} = \dfrac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}$.

$$\sigma_X^2 = \int_0^1 x^2 \frac{1+3x^2}{2}\,dx - \mu_X^2 = \left(\frac{x^3}{6} + \frac{3x^5}{10}\right)\Bigg|_0^1 - \left(\frac{5}{8}\right)^2 = \frac{73}{960}.$$

$$\sigma_Y^2 = \int_0^1 y^2 \frac{1+3y^2}{2}\,dy - \mu_Y^2 = \left(\frac{y^3}{6} + \frac{3y^5}{10}\right)\Bigg|_0^1 - \left(\frac{5}{8}\right)^2 = \frac{73}{960}.$$

$$\rho_{X,Y} = \frac{-1/64}{\sqrt{73/960}\sqrt{73/960}} = -\frac{15}{73}.$$

(c) $f_{Y\,|\,X}(y\,|\,0.5) = \dfrac{f_{X,Y}(0.5,y)}{f_X(0.5)}$.

For $0 < y < 1$, $f_{X,Y}(0.5,y) = \dfrac{3+12y^2}{8}$, $f_X(0.5) = \dfrac{7}{8}$.

So for $0 < y < 1$, $f_{Y\,|\,X}(y\,|\,0.5) = \dfrac{3+12y^2}{7}$

(d) $E(Y\,|\,X=0.5) = \displaystyle\int_0^1 y f_{Y\,|\,X}(y\,|\,0.5)\,dy = \int_0^1 y\frac{3+12y^2}{7}\,dy = \frac{3y^2+6y^4}{14}\Bigg|_0^1 = \frac{9}{14}$

21. (a) The probability mass function of Y is the same as that of X, so $f_Y(y) = e^{-y}$ if $y > 0$ and $f_Y(y) = 0$ if $y \le 0$. Since X and Y are independent, $f(x,y) = f_X(x)f_Y(y)$.

Therefore $f(x,y) = \begin{cases} e^{-x-y} & x > 0 \text{ and } y > 0 \\ 0 & \text{otherwise} \end{cases}$

(b)
$$\begin{aligned} P(X \le 1 \text{ and } Y > 1) &= P(X \le 1)P(Y > 1) \\ &= \left(\int_0^1 e^{-x}\,dx\right)\left(\int_1^\infty e^{-y}\,dy\right) \\ &= \left(-e^{-x}\Big|_0^1\right)\left(-e^{-y}\Big|_1^\infty\right) \\ &= (1-e^{-1})(e^{-1}) \\ &= e^{-1} - e^{-2} \\ &= 0.2325 \end{aligned}$$

(c) $\mu_X = \displaystyle\int_0^\infty xe^{-x}\,dx = -xe^{-x}\Bigg|_0^\infty - \int_0^\infty e^{-x}\,dx = 0 - (-e^{-x})\Bigg|_0^\infty = 0 + 1 = 1$

(d) Since X and Y have the same probability mass function, $\mu_Y = \mu_X = 1$.
Therefore $\mu_{X+Y} = \mu_X + \mu_Y = 1 + 1 = 2$.

(e)
$$\begin{aligned}
P(X+Y \le 2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{2-x} f(x,y)\,dy\,dx \\
&= \int_0^2 \int_0^{2-x} e^{-x-y}\,dy\,dx \\
&= \int_0^2 e^{-x}\left(-e^{-y}\Big|_0^{2-x}\right) dx \\
&= \int_0^2 e^{-x}(1-e^{x-2})\,dx \\
&= \int_0^2 (e^{-x} - e^{-2})\,dx \\
&= (-e^{-x} - xe^{-2})\Big|_0^2 \\
&= 1 - 3e^{-2} \\
&= 0.5940
\end{aligned}$$

23. (a) $R = 0.3X + 0.7Y$

(b) $\mu_R = \mu_{0.3X+0.7Y} = 0.3\mu_X + 0.7\mu_Y = (0.3)(6) + (0.7)(6) = 6$.
The risk is $\sigma_R = \sigma_{0.3X+0.7Y} = \sqrt{0.3^2\sigma_X^2 + 0.7^2\sigma_Y^2 + 2(0.3)(0.7)\text{Cov}(X,Y)}$.
$\text{Cov}(X,Y) = \rho_{X,Y}\sigma_X\sigma_Y = (0.3)(3)(3) = 2.7$.
Therefore $\sigma_R = \sqrt{0.3^2(3^2) + 0.7^2(3^2) + 2(0.3)(0.7)(2.7)} = 2.52$.

(c) $\mu_R = \mu_{(0.01K)X+(1-0.01K)Y} = (0.01K)\mu_X + (1-0.01K)\mu_Y = (0.01K)(6) + (1-0.01K)(6) = 6$.
$\sigma_R = \sqrt{(0.01K)^2\sigma_X^2 + (1-0.01K)^2\sigma_Y^2 + 2(0.01K)(1-0.01K)\text{Cov}(X,Y)}$.
Therefore $\sigma_R = \sqrt{(0.01K)^2(3^2) + (1-0.01K)^2(3^2) + 2(0.01K)(1-0.01K)(2.7)} = 0.03\sqrt{1.4K^2 - 140K + 10{,}000}$.

(d) σ_R is minimized when $1.4K^2 - 140K + 10000$ is minimized.
Now $\dfrac{d}{dK}(1.4K^2 - 140K + 10000) = 2.8K - 140$, so $\dfrac{d}{dK}(1.4K^2 - 140K + 10000) = 0$ if $K = 50$.
σ_R is minimized when $K = 50$.

(e) For any correlation ρ, the risk is $0.03\sqrt{K^2 + (100-K)^2 + 2\rho K(100-K)}$.
If $\rho \ne 1$ this quantity is minimized when $K = 50$.

25. (a) $\sigma_{M_1} = \sqrt{\sigma^2_{M1}} = \sqrt{\sigma^2_{R+E_1}} = \sqrt{\sigma^2_R + \sigma^2_{E1}} = \sqrt{2^2+1^2} = 2.2361$. Similarly, $\sigma_{M_2} = 2.2361$.

(b) $\mu_{M_1M_2} = \mu_{R^2+E_1R+E_2R+E_1E_2} = \mu_{R^2} + \mu_{E_1}\mu_R + \mu_{E_2}\mu_R + \mu_{E_1}\mu_{E_2} = \mu_{R^2}$

(c) $\mu_{M_1}\mu_{M_2} = \mu_{R+E_1}\mu_{R+E_2} = (\mu_R + \mu_{E_1})(\mu_R + \mu_{E_2}) = \mu_R\mu_R = \mu_R^2$

(d) $\text{Cov}(M_1, M_2) = \mu_{M_1M_2} - \mu_{M_1}\mu_{M_2} = \mu_{R^2} - \mu_R^2 = \sigma_R^2$

(e) $\rho_{M_1,M_2} = \dfrac{\text{Cov}(M_1,M_2)}{\sigma_{M_1}\sigma_{M_2}} = \dfrac{\sigma_R^2}{\sigma_{M_1}\sigma_{M_2}} = \dfrac{4}{(2.2361)(2.2361)} = 0.8$

27. (a) $\text{Cov}(aX, bY) = \mu_{aX\cdot bY} - \mu_{aX}\mu_{bY} = \mu_{abXY} - a\mu_X b\mu_Y = ab\mu_{XY} - ab\mu_X\mu_Y$
$= ab(\mu_{XY} - \mu_X\mu_Y) = ab\text{Cov}(X,Y)$.

(b) $\rho_{aX,bY} = \text{Cov}(aX,bY)/(\sigma_{aX}\sigma_{bY}) = ab\text{Cov}(X,Y)/(ab\sigma_X\sigma_Y) = \text{Cov}(X,Y)/(\sigma_X\sigma_Y) = \rho_{X,Y}$.

29. (a)
$$\begin{aligned} V(X - (\sigma_X/\sigma_Y)Y) &= \sigma_X^2 + (\sigma_X/\sigma_Y)^2\sigma_Y^2 - 2(\sigma_X/\sigma_Y)\text{Cov}(X,Y) \\ &= 2\sigma_X^2 - 2(\sigma_X/\sigma_Y)\text{Cov}(X,Y) \end{aligned}$$

(b)
$$\begin{aligned} V(X - (\sigma_X/\sigma_Y)Y) &\geq 0 \\ 2\sigma_X^2 - 2(\sigma_X/\sigma_Y)\text{Cov}(X,Y) &\geq 0 \\ 2\sigma_X^2 - 2(\sigma_X/\sigma_Y)\rho_{X,Y}\sigma_X\sigma_Y &\geq 0 \\ 2\sigma_X^2 - 2\rho_{X,Y}\sigma_X^2 &\geq 0 \\ 1 - \rho_{X,Y} &\geq 0 \\ \rho_{X,Y} &\leq 1 \end{aligned}$$

(c)
$$\begin{aligned} V(X + (\sigma_X/\sigma_Y)Y) &\geq 0 \\ 2\sigma_X^2 + 2(\sigma_X/\sigma_Y)\text{Cov}(X,Y) &\geq 0 \\ 2\sigma_X^2 + 2(\sigma_X/\sigma_Y)\rho_{X,Y}\sigma_X\sigma_Y &\geq 0 \\ 2\sigma_X^2 + 2\rho_{X,Y}\sigma_X^2 &\geq 0 \\ 1 + \rho_{X,Y} &\geq 0 \\ \rho_{X,Y} &\geq -1 \end{aligned}$$

31.
$$\begin{aligned}
\mu_Y &= \mu_{7.84C+11.44N+O-1.58Fe}\\
&= 7.84\mu_C + 11.44\mu_N + \mu_O - 1.58\mu_{Fe}\\
&= 7.84(0.0247) + 11.44(0.0255) + 0.1668 - 1.58(0.0597)\\
&= 0.5578
\end{aligned}$$

$$\begin{aligned}
\sigma_Y^2 &= \sigma^2_{7.84C+11.44N+O-1.58Fe}\\
&= 7.84^2\sigma_C^2 + 11.44^2\sigma_N^2 + \sigma_O^2 + 1.58^2\sigma_{Fe}^2 + 2(7.84)(11.44)\mathrm{Cov}(C,N) + 2(7.84)\mathrm{Cov}(C,O) - 2(7.84)(1.58)\mathrm{Cov}(C,Fe)\\
&\quad +2(11.44)\mathrm{Cov}(N,O) - 2(11.44)(1.58)\mathrm{Cov}(N,Fe) - 2(1.58)\mathrm{Cov}(O,Fe)\\
&= 7.84^2(0.0131)^2 + 11.44^2(0.0194)^2 + 0.0340^2 + 1.58^2(0.0413)^2 + 2(7.84)(11.44)(-0.0001118)\\
&\quad +2(7.84)(0.0002583) - 2(7.84)(1.58)(0.0002110) + 2(11.44)(-0.0002111) - 2(11.44)(1.58)(0.00007211)\\
&\quad -2(1.58)(0.0004915)\\
&= 0.038100
\end{aligned}$$

$\sigma = \sqrt{0.038100} = 0.1952$

33. (a) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \int_c^d\int_a^b k\,dx\,dy = k\int_c^d\int_a^b dx\,dy = k(d-c)(b-a) = 1.$

Therefore $k = \dfrac{1}{(b-a)(d-c)}$.

(b) $f_X(x) = \int_c^d k\,dy = \dfrac{d-c}{(b-a)(d-c)} = \dfrac{1}{b-a}$

(c) $f_Y(y) = \int_a^b k\,dx = \dfrac{b-a}{(b-a)(d-c)} = \dfrac{1}{d-c}$

(d) $f(x,y) = \dfrac{1}{(b-a)(d-c)} = \left(\dfrac{1}{b-a}\right)\left(\dfrac{1}{d-c}\right) = f_X(x)f_Y(y)$

Supplementary Exercises for Chapter 2

1. 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 $(A\cup B)\cup(C\cup D)$.

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

$P(C \cup D) = P(C) + P(D) - P(C \cap D) = P(C) + P(D) - P(C)P(D) = 0.95 + 0.7 - (0.95)(0.7) = 0.985.$
$P[(A \cup B) \cup (C \cup D)] = P(A \cup B) + P(C \cup D) - P(A \cup B)P(C \cup D) = 0.98 + 0.985 - (0.98)(0.985) = 0.9997.$

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

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

5. Let R be the event that the shipment is returned. Let B_1 be the event that the first brick chosen meets the specification, let B_2 be the event that the second brick chosen meets the specification, let B_3 be the event that the third brick chosen meets the specification, and let B_4 be the event that the fourth brick chosen meets the specification. Since the sample size of 4 is a small proportion of the population, it is reasonable to treat these events as independent, each with probability 0.9.
$P(R) = 1 - P(R^c) = 1 - P(B_1 \cap B_2 \cap B_3 \cap B_4) = 1 - (0.9)^4 = 0.3439.$

7. 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 $P(\text{bit received is the same as the bit sent}) = P(A^c \cap B^c) + P(A \cap B) = P(A^c)P(B^c) + P(A)P(B) = 0.9^2 + 0.1^2 = 0.82.$

9. Let A be the event that two different numbers come up, and let B be the event that one of the dice comes up 6. Then A contains 30 equally likely outcomes (6 ways to choose the number for the first die times 5 ways to choose the number for the second die). Of these 30 outcomes, 10 belong to B, specifically (1,6), (2,6), (3,6), (4,6), (5,6), (6,1), (6,2), (6,3), (6,4), and (6,5). Therefore $P(B \,|\, A) = 10/30 = 1/3$.

11. (a)
$$
\begin{aligned}
P(X \le 2 \text{ and } Y \le 3) &= \int_0^2 \int_0^3 \frac{1}{6} e^{-x/2-y/3}\,dy\,dx \\
&= \int_0^2 \frac{1}{2} e^{-x/2} \left(-e^{-y/3} \Big|_0^3 \right) dx \\
&= \int_0^2 \frac{1}{2} e^{-x/2} (1-e^{-1})\,dx \\
&= (e^{-1}-1) e^{-x/2} \Big|_0^2 \\
&= (1-e^{-1})^2 \\
&= 0.3996
\end{aligned}
$$

(b)
$$
\begin{aligned}
P(X \ge 3 \text{ and } Y \ge 3) &= \int_3^\infty \int_3^\infty \frac{1}{6} e^{-x/2-y/3}\,dy\,dx \\
&= \int_3^\infty \frac{1}{2} e^{-x/2} \left(-e^{-y/3} \Big|_3^\infty \right) dx \\
&= \int_3^\infty \frac{1}{2} e^{-x/2} e^{-1}\,dx \\
&= -e^{-1} e^{-x/2} \Big|_3^\infty \\
&= e^{-5/2} \\
&= 0.0821
\end{aligned}
$$

(c) If $x \le 0$, $f(x,y) = 0$ for all y so $f_X(x) = 0$.

If $x > 0$, $f_X(x) = \int_0^\infty \frac{1}{6} e^{-x/2-y/3}\,dy = \frac{1}{2} e^{-x/2} \left(-e^{-y/3} \Big|_0^\infty \right) = \frac{1}{2} e^{-x/2}$.

Therefore $f_X(x) = \begin{cases} \frac{1}{2} e^{-x/2} & x > 0 \\ 0 & x \le 0 \end{cases}$

(d) If $y \le 0$, $f(x,y) = 0$ for all x so $f_Y(y) = 0$.

If $y > 0$, $f_Y(y) = \int_3^\infty \frac{1}{6} e^{-x/2-y/3}\,dx = \frac{1}{3} e^{-y/3} \left(-e^{-x/2} \Big|_0^\infty \right) = \frac{1}{3} e^{-y/3}$.

Therefore $f_Y(y) = \begin{cases} \frac{1}{3} e^{-y/3} & y > 0 \\ 0 & y \le 0 \end{cases}$

(e) Yes, $f(x,y) = f_X(x) f_Y(y)$.

13. Let D denote the event that a snowboard is defective, let E denote the event that a snowboard is made in the eastern United States, let W denote the event that a snowboard is made in the western United States, and let C denote the event that a snowboard is made in Canada. Then $P(E) = P(W) = 10/28$, $P(C) = 8/28$, $P(D|E) = 3/100$, $P(D|W) = 6/100$, and $P(D|C) = 4/100$.

(a)
$$\begin{aligned} P(D) &= P(D|E)P(E) + P(D|W)P(W) + P(D|C)P(C) \\ &= \left(\frac{10}{28}\right)\left(\frac{3}{100}\right) + \left(\frac{10}{28}\right)\left(\frac{6}{100}\right) + \left(\frac{8}{28}\right)\left(\frac{4}{100}\right) \\ &= \frac{122}{2800} = 0.0436 \end{aligned}$$

(b) $P(D \cap C) = P(D|C)P(C) = \left(\frac{8}{28}\right)\left(\frac{4}{100}\right) = \frac{32}{2800} = 0.0114$

(c) Let U be the event that a snowboard was made in the United States.

Then $P(D \cap U) = P(D) - P(D \cap C) = \frac{122}{2800} - \frac{32}{2800} = \frac{90}{2800}$.

$P(U|D) = \frac{P(D \cap U)}{P(D)} = \frac{90/2800}{122/2800} = \frac{90}{122} = 0.7377$.

15. The total number of pairs of cubicles is $\binom{6}{2} = \frac{6!}{2!4!} = 15$. Each is equally likely to be chosen. Of these pairs, five are adjacent (1 and 2, 2 and 3, 3 and 4, 4 and 5, 5 and 6). Therefore the probability that an adjacent pair of cubicles is selected is $5/15$, or $1/3$.

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

19. The marginal probability mass function $p_X(x)$ is found by summing along the rows of the joint probability mass function.

x	y = 100	150	200	$p_X(x)$
0.02	0.05	0.06	0.11	0.22
0.04	0.01	0.08	0.10	0.19
0.06	0.04	0.08	0.17	0.29
0.08	0.04	0.14	0.12	0.30
$p_Y(y)$	0.14	0.36	0.50	

(a) For additive concentration (X): $p_X(0.02) = 0.22$, $p_X(0.04) = 0.19$, $p_X(0.06) = 0.29$, $p_X(0.08) = 0.30$, and $p_X(x) = 0$ for $x \neq 0.02$, 0.04, 0.06, or 0.08.

For tensile strength (Y): The marginal probability mass function $p_Y(y)$ is found by summing down the columns of the joint probability mass function. Therefore $p_Y(100) = 0.14$, $p_Y(150) = 0.36$, $p_Y(200) = 0.50$, and $p_Y(y) = 0$ for $y \neq 100$, 150, or 200.

(b) No, X and Y are not independent. For example $P(X = 0.02 \cap Y = 100) = 0.05$, but $P(X = 0.02)P(Y = 100) = (0.22)(0.14) = 0.0308$.

(c)
$$\begin{aligned} P(Y \geq 150 \,|\, X = 0.04) &= \frac{P(Y \geq 150 \text{ and } X = 0.04)}{P(X = 0.04)} \\ &= \frac{P(Y = 150 \text{ and } X = 0.04) + P(Y = 200 \text{ and } X = 0.04)}{P(X = 0.04)} \\ &= \frac{0.08 + 0.10}{0.19} \\ &= 0.947 \end{aligned}$$

(d)
$$\begin{aligned} P(Y > 125 \,|\, X = 0.08) &= \frac{P(Y > 125 \text{ and } X = 0.08)}{P(X = 0.08)} \\ &= \frac{P(Y = 150 \text{ and } X = 0.08) + P(Y = 200 \text{ and } X = 0.08)}{P(X = 0.08)} \\ &= \frac{0.14 + 0.12}{0.30} \\ &= 0.867 \end{aligned}$$

(e) The tensile strength is greater than 175 if $Y = 200$. Now

$$P(Y=200\,|\,X=0.02)=\frac{P(Y=200\text{ and }X=0.02)}{P(X=0.02)}=\frac{0.11}{0.22}=0.500,$$
$$P(Y=200\,|\,X=0.04)=\frac{P(Y=200\text{ and }X=0.04)}{P(X=0.04)}=\frac{0.10}{0.19}=0.526,$$
$$P(Y=200\,|\,X=0.06)=\frac{P(Y=200\text{ and }X=0.06)}{P(X=0.06)}=\frac{0.17}{0.29}=0.586,$$
$$P(Y=200\,|\,X=0.08)=\frac{P(Y=200\text{ and }X=0.08)}{P(X=0.08)}=\frac{0.12}{0.30}=0.400.$$

The additive concentration should be 0.06.

21. (a) $$p_{Y|X}(100\,|\,0.06)=\frac{p(0.06,100)}{p_X(0.06)}=\frac{0.04}{0.29}=\frac{4}{29}=0.138$$
$$p_{Y|X}(150\,|\,0.06)=\frac{p(0.06,150)}{p_X(0.06)}=\frac{0.08}{0.29}=\frac{8}{29}=0.276$$
$$p_{Y|X}(200\,|\,0.06)=\frac{p(0.06,200)}{p_X(0.06)}=\frac{0.17}{0.29}=\frac{17}{29}=0.586$$

(b) $$p_{X|Y}(0.02\,|\,100)=\frac{p(0.02,100)}{p_Y(100)}=\frac{0.05}{0.14}=\frac{5}{14}=0.357$$
$$p_{X|Y}(0.04\,|\,100)=\frac{p(0.04,100)}{p_Y(100)}=\frac{0.01}{0.14}=\frac{1}{14}=0.071$$
$$p_{X|Y}(0.06\,|\,100)=\frac{p(0.06,100)}{p_Y(100)}=\frac{0.04}{0.14}=\frac{4}{14}=0.286$$
$$p_{X|Y}(0.08\,|\,100)=\frac{p(0.08,100)}{p_Y(100)}=\frac{0.04}{0.14}=\frac{4}{14}=0.286$$

(c) $$\begin{aligned}E(Y\,|\,X=0.06) &= 100p_{Y|X}(100\,|\,0.06)+150p_{Y|X}(150\,|\,0.06)+200p_{Y|X}(200\,|\,0.06)\\ &= 100(4/29)+150(8/29)+200(17/29)\\ &= 172.4\end{aligned}$$

(d) $$\begin{aligned}E(X\,|\,Y=100) &= 0.02p_{X|Y}(0.02\,|\,100)+0.04p_{X|Y}(0.04\,|\,100)+0.06p_{X|Y}(0.06\,|\,100)+0.08p_{X|Y}(0.08\,|\,100)\\ &= 0.02(5/14)+0.04(1/14)+0.06(4/14)+0.08(4/14)\\ &= 0.0500\end{aligned}$$

23. (a) Under scenario A:

$\mu = 0(0.65) + 5(0.2) + 15(0.1) + 25(0.05) = 3.75$

$\sigma = \sqrt{0^2(0.65) + 5^2(0.2) + 15^2(0.1) + 25^2(0.05) - 3.75^2} = 6.68$

(b) Under scenario B:

$\mu = 0(0.65) + 5(0.24) + 15(0.1) + 20(0.01) = 2.90$

$\sigma = \sqrt{0^2(0.65) + 5^2(0.24) + 15^2(0.1) + 20^2(0.01) - 2.90^2} = 4.91$

(c) Under scenario C:

$\mu = 0(0.65) + 2(0.24) + 5(0.1) + 10(0.01) = 1.08$

$\sigma = \sqrt{0^2(0.65) + 2^2(0.24) + 5^2(0.1) + 10^2(0.01) - 1.08^2} = 1.81$

(d) Let L denote the loss.

Under scenario A, $P(L < 10) = P(L = 0) + P(L = 5) = 0.65 + 0.2 = 0.85$.

Under scenario B, $P(L < 10) = P(L = 0) + P(L = 5) = 0.65 + 0.24 = 0.89$.

Under scenario C, $P(L < 10) = P(L = 0) + P(L = 2) + P(L = 5) = 0.65 + 0.24 + 0.1 = 0.99$.

25. (a) $p(0,0) = P(X = 0 \text{ and } Y = 0) = \left(\frac{3}{10}\right)\left(\frac{2}{9}\right) = \frac{1}{15} = 0.0667$

$p(1,0) = P(X = 1 \text{ and } Y = 0) = \left(\frac{4}{10}\right)\left(\frac{3}{9}\right) + \left(\frac{3}{10}\right)\left(\frac{4}{9}\right) = \frac{4}{15} = 0.2667$

$p(2,0) = P(X = 2 \text{ and } Y = 0) = \left(\frac{4}{10}\right)\left(\frac{3}{9}\right) = \frac{2}{15} = 0.1333$

$p(0,1) = P(X = 0 \text{ and } Y = 1) = \left(\frac{3}{10}\right)\left(\frac{3}{9}\right) + \left(\frac{3}{10}\right)\left(\frac{3}{9}\right) = \frac{3}{15} = 0.2000$

$p(1,1) = P(X = 1 \text{ and } Y = 1) = \left(\frac{4}{10}\right)\left(\frac{3}{9}\right) + \left(\frac{3}{10}\right)\left(\frac{4}{9}\right) = \frac{4}{15} = 0.2667$

$p(0,2) = P(X = 0 \text{ and } Y = 2) = \left(\frac{3}{10}\right)\left(\frac{2}{9}\right) = \frac{1}{15} = 0.0667$

$p(x,y) = 0$ for all other pairs (x,y).

The joint probability mass function is

	y		
x	0	1	2
0	0.0667	0.2000	0.0667
1	0.2667	0.2667	0
2	0.1333	0	0

(b) The marginal probability density function of X is:

$$p_X(0) = p(0,0) + p(0,1) + p(0,2) = \frac{1}{15} + \frac{3}{15} + \frac{1}{15} = \frac{1}{3}$$

$$p_X(1) = p(1,0) + p(1,1) = \frac{4}{15} + \frac{4}{15} = \frac{8}{15}.$$

$$p_X(2) = p(2,0) = \frac{2}{15}$$

$$\mu_X = 0p_X(0) + 1p_X(1) + 2p_X(2) = 0\left(\frac{1}{3}\right) + 1\left(\frac{8}{15}\right) + 2\left(\frac{2}{15}\right) = \frac{12}{15} = 0.8$$

(c) The marginal probability density function of Y is:

$$p_Y(0) = p(0,0) + p(1,0) + p(2,0) = \frac{1}{15} + \frac{4}{15} + \frac{2}{15} = \frac{7}{15}$$

$$p_Y(1) = p(0,1) + p(1,1) = \frac{3}{15} + \frac{4}{15} = \frac{7}{15}.$$

$$p_Y(2) = p(0,2) = \frac{1}{15}$$

$$\mu_Y = 0p_Y(0) + 1p_Y(1) + 2p_Y(2) = 0\left(\frac{7}{15}\right) + 1\left(\frac{7}{15}\right) + 2\left(\frac{1}{15}\right) = \frac{9}{15} = 0.6$$

(d)
$$\begin{aligned}\sigma_X &= \sqrt{0^2 p_X(0) + 1^2 p_X(1) + 2^2 p_X(2) - \mu_X^2} \\ &= \sqrt{0^2(1/3) + 1^2(8/15) + 2^2(2/15) - (12/15)^2} \\ &= \sqrt{96/225} = 0.6532\end{aligned}$$

(e)
$$\begin{aligned}\sigma_Y &= \sqrt{0^2 p_Y(0) + 1^2 p_Y(1) + 2^2 p_Y(2) - \mu_Y^2} \\ &= \sqrt{0^2(7/15) + 1^2(7/15) + 2^2(1/15) - (9/15)^2} \\ &= \sqrt{84/225} = 0.6110\end{aligned}$$

(f) $\text{Cov}(X,Y) = \mu_{XY} - \mu_X \mu_Y$.

$$\begin{aligned}\mu_{XY} &= (0)(0)p(0,0) + (1)(0)p(1,0) + (2)(0)p(2,0) + (0)(1)p(0,1) + (1)(1)p(1,1) + (0)(2)p(0,2) \\ &= (1)(1)\frac{4}{15} = \frac{4}{15}\end{aligned}$$

$$\text{Cov}(X,Y) = \frac{4}{15} - \left(\frac{12}{15}\right)\left(\frac{9}{15}\right) = -\frac{48}{225} = -0.2133$$

(g) $\rho_{X,Y} = \dfrac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} = \dfrac{-48/225}{\sqrt{96/225}\sqrt{84/225}} = -0.5345$

27. (a) $\mu_X = \displaystyle\int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^1 x \frac{6x^2+6x+2}{7}\,dx = \frac{1}{14}(3x^4+4x^3+2x^2)\bigg|_0^1 = \frac{9}{14} = 0.6429$

(b) $\sigma_X^2 = \displaystyle\int_0^1 x^2 \frac{6x^2+6x+2}{7}\,dx - \mu_X^2 = \frac{1}{7}\left(\frac{6x^5}{5} + \frac{3x^4}{2} + \frac{2x^3}{3}\right)\bigg|_0^1 - \left(\frac{9}{14}\right)^2 = \frac{199}{2940} = 0.06769$

(c) $\text{Cov}(X,Y) = \mu_{XY} - \mu_X \mu_Y$.

$$\begin{aligned}
\mu_{XY} &= \int_0^1 \int_0^1 xy\left(\frac{6}{7}\right)(x+y)^2\,dx\,dy \\
&= \int_0^1 \frac{6}{7} y \left(\frac{x^4}{4} + \frac{2x^3 y}{3} + \frac{x^2 y^2}{2}\bigg|_0^1\right) dy \\
&= \int_0^1 \frac{6}{7}\left(\frac{y^3}{2} + \frac{2y^2}{3} + \frac{y}{4}\right) dy \\
&= \frac{6}{7}\left(\frac{y^2}{8} + \frac{2y^3}{9} + \frac{y^4}{8}\right)\bigg|_0^1 \\
&= \frac{17}{42}
\end{aligned}$$

$\mu_X = \dfrac{9}{14}$, computed in part (a). To compute μ_Y, note that the joint density is symmetric in x and y, so the marginal density of Y is the same as that of X. It follows that $\mu_Y = \mu_X = \dfrac{9}{14}$.

$\text{Cov}(X,Y) = \dfrac{17}{42} - \left(\dfrac{9}{14}\right)\left(\dfrac{9}{14}\right) = \dfrac{-5}{588} = -0.008503$.

(d) Since the marginal density of Y is the same as that of X, $\sigma_Y^2 = \sigma_X^2 = \dfrac{199}{2940}$.

Therefore $\rho_{X,Y} = \dfrac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} = \dfrac{-5/588}{\sqrt{199/2940}\sqrt{199/2940}} = \dfrac{-25}{199} = -0.1256$.

29. (a) $p_X(0) = 0.6$, $p_X(1) = 0.4$, $p_X(x) = 0$ if $x \neq 0$ or 1.

(b) $p_Y(0) = 0.4$, $p_Y(1) = 0.6$, $p_Y(y) = 0$ if $y \neq 0$ or 1.

(c) Yes. It is reasonable to assume that knowledge of the outcome of one coin will not help predict the outcome of the other.

(d) $p(0,0) = p_X(0)p_Y(0) = (0.6)(0.4) = 0.24$, $p(0,1) = p_X(0)p_Y(1) = (0.6)(0.6) = 0.36$,
$p(1,0) = p_X(1)p_Y(0) = (0.4)(0.4) = 0.16$, $p(1,1) = p_X(1)p_Y(1) = (0.4)(0.6) = 0.24$,
$p(x,y) = 0$ for other values of (x,y).

31. (a) The possible values of the pair (X,Y) are the ordered pairs (x,y) where each of x and y is equal to 1, 2, or 3. There are nine such ordered pairs, and each is equally likely. Therefore $p_{X,Y}(x,y) = 1/9$ for $x = 1,2,3$ and $y = 1,2,3$, and $p_{X,Y}(x,y) = 0$ for other values of (x,y).

(b) Both X and Y are sampled from the numbers $\{1,2,3\}$, with each number being equally likely.
Therefore $p_X(1) = p_X(2) = p_X(3) = 1/3$, and $p_X(x) = 0$ for other values of x. p_Y is the same.

(c) $\mu_X = \mu_Y = 1(1/3) + 2(1/3) + 3(1/3) = 2$

(d) $\mu_{XY} = \sum_{x=1}^{3}\sum_{y=1}^{3} xyp_{X,Y}(x,y) = \dfrac{1}{9}\sum_{x=1}^{3}\sum_{y=1}^{3} xy = \dfrac{1}{9}(1+2+3)(1+2+3) = 4.$

Another way to compute μ_{XY} is to note that X and Y are independent, so $\mu_{XY} = \mu_X\mu_Y = (2)(2) = 4$.

(e) $\mathrm{Cov}(X,Y) = \mu_{XY} - \mu_X\mu_Y = 4 - (2)(2) = 0$

33. (a) $\mu_X = \int_{-\infty}^{\infty} xf(x)\,dx$. Since $f(x) = 0$ for $x \leq 0$, $\mu_X = \int_0^{\infty} xf(x)\,dx$.

(b) $\mu_X = \int_0^{\infty} xf(x)\,dx \geq \int_k^{\infty} xf(x)\,dx \geq \int_k^{\infty} kf(x)\,dx = kP(X \geq k)$

(c) $\mu_X/k \geq kP(X \geq k)/k = P(X \geq k)$

(d) $\mu_X = \mu_{(Y-\mu_Y)^2} = \sigma_Y^2$

(e) $P(|Y - \mu_Y| \geq k\sigma_Y) = P((Y-\mu_Y)^2 \geq k^2\sigma_Y^2) = P(X \geq k^2\sigma_Y^2)$

(f) $P(|Y - \mu_Y| \geq k\sigma_Y) = P(X \geq k^2\sigma_Y^2) \leq \mu_X/(k^2\sigma_Y^2) = \sigma_Y^2/(k^2\sigma_Y^2) = 1/k^2$

35. (a) If the pooled test is negative, it is the only test performed, so $X = 1$. If the pooled test is positive, then n additional tests are carried out, one for each individual, so $X = n+1$. The possible values of X are therefore 1 and $n+1$.

(b) The possible values of X are 1 and 5. Now $X = 1$ if the pooled test is negative. This occurs if none of the individuals has the disease. The probability that this occurs is $(1-0.1)^4 = 0.6561$. Therefore $P(X = 1) = 0.6561$. It follows that $P(X = 5) = 0.3439$.
So $\mu_X = 1(0.6561)+5(0.3439) = 2.3756$.

(c) The possible values of X are 1 and 7. Now $X = 1$ if the pooled test is negative. This occurs if none of the individuals has the disease. The probability that this occurs is $(1-0.2)^6 = 0.262144$. Therefore $P(X = 1) = 0.262144$. It follows that $P(X = 7) = 0.737856$.
So $\mu_X = 1(0.262144)+7(0.737856) = 5.4271$.

(d) The possible values of X are 1 and $n+1$. Now $X = 1$ if the pooled test is negative. This occurs if none of the individuals has the disease. The probability that this occurs is $(1-p)^n$. Therefore $P(X = 1) = (1-p)^n$. It follows that $P(X = n+1) = 1-(1-p)^n$.
So $\mu_X = 1(1-p)^n + (n+1)(1-(1-p)^n) = n+1-n(1-p)^n$.

(e) The pooled method is more economical if $11-10(1-p)^{10} < 10$. Solving for p yields $p < 0.2057$.

Chapter 3

Section 3.1

1. (ii). The measurements are close together, and thus precise, but they are far from the true value of 100°C, and thus not accurate.

3. (a) True

   (b) False

   (c) False

   (d) True

5. (a) No, we cannot determine the standard deviation of the process from a single measurement.

   (b) Yes, the bias can be estimated to be 2 pounds, because the reading is 2 pounds when the true weight is 0.

7. (a) Yes, the uncertainty can be estimated with the standard deviation of the five measurements, which is 21.3 μg.

   (b) No, the bias cannot be estimated, since we do not know the true value.

9. We can get a more accurate estimate by subtracting the bias of 26.2 μg, obtaining 100.8 μg above 1 kg.

11. (a) No, they are in increasing order, which would be highly unusual for a simple random sample.

    (b) No, since they are not a simple random sample from a population of possible measurements, we cannot estimate the uncertainty.

Section 3.2

1. (a) $\sigma_{4X} = 4\sigma_X = 4(0.3) = 1.2$

   (b) $\sigma_{X-2Y} = \sqrt{\sigma_X^2 + 2^2\sigma_Y^2} = \sqrt{0.3^2 + (2^2)(0.2^2)} = 0.5$

   (c) $\sigma_{2X-3Y} = \sqrt{2^2\sigma_X^2 + 3^2\sigma_Y^2} = \sqrt{(2^2)(0.3^2) + (3^2)(0.2^2)} = 0.8485$

3. Let n be the necessary number of measurements. Then $\sigma_{\overline{X}} = 3/\sqrt{n} = 1$. Solving for n yields $n = 9$.

5. Let X represent the estimated mean annual level of land uplift for the years 1774–1884, and let Y represent the estimated mean annual level of land uplift for the years 1885–1984. Then $X = 4.93$, $Y = 3.92$, $\sigma_X = 0.23$, and $\sigma_Y = 0.19$. The difference is $X - Y = 1.01$ mm.

   The uncertainty is $\sigma_{X-Y} = \sqrt{\sigma_X^2 + \sigma_Y^2} = \sqrt{0.23^2 + 0.19^2} = 0.30$ mm.

7. $d = 3$, $F = 2.2$, $\sigma_F = 0.1$.

   $W = Fd = (2.2)(3) = 6.6\,\text{Nm}$.

   $\sigma_W = \sigma_{3F} = 3\sigma_F = 3(0.1) = 0.3\,\text{Nm}$.

9. $D_W = 1000$, $D_S = 500$, $\sigma_{D_S} = 5$.

   $G = D_S/D_W = 500/1000 = 0.5$

   $\sigma_G = \sigma_{0.001D_S} = 0.001\sigma_{D_S} = 0.001(5) = 0.005$ kg/m^3.

11. (a) $T_a = 36$, $k = 0.025$, $t = 10$, $T_0 = 72$, $\sigma_{T_0} = 0.5$

    $T = T_a + (T_0 - T_a)e^{-kt} = 36 + 0.7788(T_0 - 36) = 36 + 0.7788(72 - 36) = 64.04°\text{F}$

    $\sigma_T = \sigma_{T_a+(T_0-T_a)e^{-kt}} = \sigma_{36+0.7788(T_0-36)} = \sigma_{0.7788(T_0-36)} = 0.7788\sigma_{(T_0-36)} = 0.7788\sigma_{T_0} = 0.7788(0.5) = 0.39°\text{F}$

    (b) $T_0 = 72$, $k = 0.025$, $t = 10$, $T_a = 36$, $\sigma_{T_a} = 0.5$

    $T = T_a + (T_0 - T_a)e^{-kt} = T_a + 0.7788(72 - T_a) = 56.074 + 0.2212T_a = 56.074 + 0.2212(36) = 64.04°\text{F}$

    $\sigma_T = \sigma_{56.074+0.2212T_a} = \sigma_{0.2212T_a} = 0.2212\sigma_{T_a} = 0.2212(0.5) = 0.11°\text{F}$

13. (a) The uncertainty in the average of 9 measurements is approximately equal to $s/\sqrt{9} = 0.081/3 = 0.027$ cm.

(b) The uncertainty in a single measurement is approximately equal to s, which is 0.081 cm.

15. (a) Let $\overline{X} = 87.0$ denote the average of the eight measurements, let $s = 2.0$ denote the sample standard deviation, and let σ denote the unknown population standard deviation. The volume is estimated with $\overline{X}$. The uncertainty is $\sigma_{\overline{X}} = \sigma/\sqrt{8} \approx s/\sqrt{8} = 2.0/\sqrt{8} = 0.7$ mL.

(b) $\sigma_{\overline{X}} \approx s/\sqrt{16} = 2.0/\sqrt{16} = 0.5$ mL

(c) Let n be the required number of measurements. Then $\sigma/\sqrt{n} = 0.4$.
Approximating σ with $s = 2.0$ yields $2.0/\sqrt{n} \approx 0.4$, so $n \approx 25$.

17. Let $\overline{X}$ denote the average of the 10 yields at 65°C, and let $\overline{Y}$ denote the average of the 10 yields at 80°C. Let s_X denote the sample standard deviation of the 10 yields at 65°C, and let s_Y denote the sample standard deviation of the 10 yields at 80°C. Then $\overline{X} = 70.14$, $s_X = 0.897156$, $\overline{Y} = 90.50$, $s_Y = 0.794425$.

(a) At 65°C, the yield is $\overline{X} \pm s_X/\sqrt{10} = 70.14 \pm 0.897156/\sqrt{10} = 70.14 \pm 0.28$.

At 80°C, the yield is $\overline{Y} \pm s_Y/\sqrt{10} = 90.50 \pm 0.794425/\sqrt{10} = 90.50 \pm 0.25$.

(b) The difference is estimated with $\overline{Y} - \overline{X} = 90.50 - 70.14 = 20.36$.
The uncertainty is $\sigma_{\overline{Y}-\overline{X}} = \sqrt{\sigma^2_{\overline{Y}} + \sigma^2_{\overline{X}}} = \sqrt{0.25^2 + 0.28^2} = 0.38$.

19. (a) Let $\sigma_X = 0.05$ denote the uncertainty in the instrument. Then $\sigma_{\overline{X}} = \sigma_X/\sqrt{10} = 0.05/\sqrt{10} = 0.016$.

(b) Let $\sigma_Y = 0.02$ denote the uncertainty in the instrument. Then $\sigma_{\overline{Y}} = \sigma_Y/\sqrt{5} = 0.02/\sqrt{5} = 0.0089$.

(c) The uncertainty in $\frac{1}{2}\overline{X} + \frac{1}{2}\overline{Y}$ is
$\sigma_{0.5\overline{X}+0.5\overline{Y}} = \sqrt{0.5^2\sigma^2_{\overline{X}} + 0.5^2\sigma^2_{\overline{Y}}} = \sqrt{0.5^2(0.05^2/10) + 0.5^2(0.02^2/5)} = 0.0091$.
The uncertainty in $\frac{10}{15}\overline{X} + \frac{5}{15}\overline{Y}$ is
$\sigma_{(2/3)\overline{X}+(1/3)\overline{Y}} = \sqrt{(2/3)^2\sigma^2_{\overline{X}} + (1/3)^2\sigma^2_{\overline{Y}}} = \sqrt{(2/3)^2(0.05^2/10) + (1/3)^2(0.02^2/5)} = 0.011$.
The uncertainty in $\frac{1}{2}\overline{X} + \frac{1}{2}\overline{Y}$ is smaller.

(d) $c_{\text{best}} = \dfrac{\sigma^2_{\overline{Y}}}{\sigma^2_{\overline{X}} + \sigma^2_{\overline{Y}}} = \dfrac{0.02^2/5}{0.05^2/10 + 0.02^2/5} = 0.24.$

The minimum uncertainty is $\sqrt{c^2_{\text{best}}\sigma^2_{\overline{X}} + (1 - c_{\text{best}})^2\sigma^2_{\overline{Y}}} = 0.0078.$

Section 3.3

1. $X = 2.0$, $\sigma_X = 0.3$.

(a) $\dfrac{dY}{dX} = 3X^2 = 12, \quad \sigma_Y = \left|\dfrac{dY}{dX}\right|\sigma_X = 3.6$

(b) $\dfrac{dY}{dX} = \dfrac{1}{\sqrt{2X}} = 0.5, \quad \sigma_Y = \left|\dfrac{dY}{dX}\right|\sigma_X = 0.15$

(c) $\dfrac{dY}{dX} = \dfrac{-3}{X^2} = -0.75, \quad \sigma_Y = \left|\dfrac{dY}{dX}\right|\sigma_X = 0.225$

(d) $\dfrac{dY}{dX} = \dfrac{1}{X} = 0.5, \quad \sigma_Y = \left|\dfrac{dY}{dX}\right|\sigma_X = 0.15$

(e) $\dfrac{dY}{dX} = e^X = e^2 = 7.389, \quad \sigma_Y = \left|\dfrac{dY}{dX}\right|\sigma_X = 2.217$

(f) $\dfrac{dY}{dX} = -\sin X = -\sin 2 = 0.909297, \quad \sigma_Y = \left|\dfrac{dY}{dX}\right|\sigma_X = 0.273$

3. $h = 6$, $r = 5$, $\sigma_r = 0.02$, $V = \pi r^2 h/3 = 157.0796$.

$\dfrac{dV}{dr} = 2\pi rh/3 = 62.8319 \quad \sigma_V = \left|\dfrac{dV}{dr}\right|\sigma_r = 62.8319(0.02) = 1.3 \text{ cm}^3.$

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

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

$T = 1.7289 \pm 0.0058$ s

(b) $L = 0.742$, $T = 1.73$, $\sigma_T = 0.01$, $g = 4\pi^2 LT^{-2} = 29.292986T^{-2} = 9.79$

$$\frac{dg}{dT} = -58.5860T^{-3} = -11.315 \quad \sigma_T = \left|\frac{dg}{dT}\right|\sigma_T = 0.11$$

$g = 9.79 \pm 0.11$ m/s^2

7. (a) $g = 9.80$, $d = 0.15$, $l = 30.0$, $h = 5.33$, $\sigma_h = 0.02$, $F = \sqrt{gdh/4l} = 0.110680\sqrt{h} = 0.2555$

$$\frac{dF}{dh} = 0.055340h^{-1/2} = 0.023970 \quad \sigma_F = \left|\frac{dF}{dh}\right|\sigma_h = 0.0005$$

$F = 0.2555 \pm 0.0005$ m/s

(b) $g = 9.80$, $l = 30.0$, $h = 5.33$, $d = 0.15$, $\sigma_d = 0.03$, $F = \sqrt{gdh/4l} = 0.659760\sqrt{d} = 0.2555$

$$\frac{dF}{dh} = 0.329880d^{-1/2} = 0.851747 \quad \sigma_F = \left|\frac{dF}{dd}\right|\sigma_d = 0.026$$

$F = 0.256 \pm 0.026$ m/s

Note that F is given to only three significant digits in the final answer, in order to keep the uncertainty to no more than two significant digits.

(c) $g = 9.80$, $d = 0.15$, $h = 5.33$, $l = 30.00$, $\sigma_l = 0.04$, $F = \sqrt{gdh/4l} = 1.399562l^{-1/2} = 0.2555$

$$\frac{dF}{dl} = -0.699781l^{-3/2} = -0.00425873 \quad \sigma_F = \left|\frac{dF}{dl}\right|\sigma_l = 0.0002$$

$F = 0.2555 \pm 0.0002$ m/s

9. $m = 750$, $V_0 = 500.0$, $V_1 = 813.2$, $\sigma_{V_0} = \sigma_{V_1} = 0.1$, $D = \dfrac{m}{V_1 - V_0} = \dfrac{750}{V_1 - V_0} = 2.3946.$

Let $V = V_1 - V_0$. Then $V = 813.2 - 500.0 = 313.2$, and

$\sigma_V = \sigma_{V_1 - V_0} = \sqrt{\sigma_{V_1}^2 + \sigma_{V_0}^2} = \sqrt{0.1^2 + 0.1^2} = 0.141421.$

Now $\dfrac{dD}{dV} = -\dfrac{750}{V^2} = -0.007646, \quad \sigma_D = \left|\dfrac{dD}{dV}\right|\sigma_V = 0.0011.$

$D = 2.3946 \pm 0.0011$ g/mL

11. (a) The relative uncertainty is 0.4/20.9 = 1.9%

(b) The relative uncertainty is 0.8/15.1 = 5.3%

(c) The relative uncertainty is 23/388 = 5.9%

(d) The relative uncertainty is 0.009/2.465 = 0.37%

13. $s = 2.2$, $t = 0.67$, $\sigma_t = 0.02$, $g = 2s/t^2 = 4.4/t^2 = 9.802$

$$\ln g = \ln 4.4 - 2\ln t, \quad \frac{d\ln g}{dt} = -2/t = -2.985, \quad \sigma_{\ln g} = \left|\frac{d\ln g}{dt}\right|\sigma_t = 0.060$$

$g = 9.802$ m/s^2 $\pm$ 6.0%

15. (a) $g = 9.80$, $L = 0.855$, $\sigma_L = 0.005$, $T = 2\pi\sqrt{L/g} = 2.00709\sqrt{L} = 1.85588$

$$\ln T = \ln 2.00709 + 0.5\ln L, \quad \frac{d\ln T}{dL} = 0.5/L = 0.584795, \quad \sigma_{\ln T} = \left|\frac{d\ln T}{dL}\right|\sigma_L = 0.0029$$

$T = 1.856$ s $\pm$ 0.29%

(b) $L = 0.855$, $T = 1.856$, $\sigma_T = 0.005$, $g = 4\pi^2 L T^{-2} = 33.754047T^{-2} = 9.799$

$$\ln g = \ln 33.754047 - 2\ln T, \quad \frac{d\ln g}{dT} = -2/T = -1.0776, \quad \sigma_{\ln g} = \left|\frac{d\ln g}{dT}\right|\sigma_T = 0.0054$$

$g = 9.799$ m/s^2 $\pm$ 0.54%

17. (a) $g = 9.80$, $d = 0.20$, $l = 35.0$, $h = 4.51$, $\sigma_h = 0.03$, $F = \sqrt{gdh/4l} = 0.118332\sqrt{h} = 0.2513$

$$\ln F = \ln 0.118332 + 0.5\ln h, \quad \frac{d\ln F}{dh} = 0.5/h = 0.110865, \quad \sigma_{\ln F} = \left|\frac{d\ln F}{dh}\right|\sigma_h = 0.0033$$

$F = 0.2513$ m/s $\pm$ 0.33%

(b) $g = 9.80$, $l = 35.0$, $h = 4.51$, $d = 0.20$, $\sigma_d = 0.008$, $F = \sqrt{gdh/4l} = 0.561872\sqrt{d} = 0.2513$

$$\ln F = \ln 0.561872 + 0.5\ln d, \quad \frac{d\ln F}{dd} = 0.5/d = 2.5, \quad \sigma_{\ln F} = \left|\frac{d\ln F}{dd}\right|\sigma_d = 0.02$$

$F = 0.2513$ m/s $\pm$ 2.0%

(c) $g = 9.80$, $d = 0.20$, $h = 4.51$, $l = 35.00$, $\sigma_l = 0.4$, $F = \sqrt{gdh/4l} = 1.486573 l^{-1/2} = 0.2513$

$\ln F = \ln 1.486573 - 0.5 \ln l, \quad \dfrac{d \ln F}{dl} = -0.5/l = -0.014286, \quad \sigma_{\ln F} = \left|\dfrac{d \ln F}{dl}\right| \sigma_l = 0.0057$

$F = 0.2513$ m/s $\pm$ 0.57%

19. $m = 288.2$, $V_0 = 400.0$, $V_1 = 516.0$, $\sigma_{V_0} = 0.1$, $\sigma_{V_1} = 0.2$, $D = \dfrac{m}{V_1 - V_0} = \dfrac{288.2}{V_1 - V_0} = 2.484$

Let $V = V_1 - V_0$. Then $V = 516.0 - 400.0 = 116.0$, and

$\sigma_V = \sigma_{V_1 - V_0} = \sqrt{\sigma_{V_1}^2 + \sigma_{V_0}^2} = \sqrt{0.1^2 + 0.2^2} = 0.223607$

$\ln D = \ln 288.2 - \ln V$, $\dfrac{d \ln D}{dV} = -1/V = -0.008621, \quad \sigma_{\ln D} = \left|\dfrac{d \ln D}{dV}\right| \sigma_V = 0.0019.$

$D = 2.484$ g/mL $\pm$ 0.19%

Section 3.4

1. (a) $X = 10.0$, $\sigma_X = 0.5$, $Y = 5.0$, $\sigma_Y = 0.1$, $U = XY^2 = 250$

$$\frac{\partial U}{\partial X} = Y^2 = 25.0, \quad \frac{\partial U}{\partial Y} = 2XY = 100.0, \quad \sigma_U = \sqrt{\left(\frac{\partial U}{\partial X}\right)^2 \sigma_X^2 + \left(\frac{\partial U}{\partial Y}\right)^2 \sigma_Y^2} = 16.0$$

$U = 250 \pm 16$

(b) $X = 10.0$, $\sigma_X = 0.5$, $Y = 5.0$, $\sigma_Y = 0.1$, $U = X^2 + Y^2 = 125$

$$\frac{\partial U}{\partial X} = 2X = 20.0, \quad \frac{\partial U}{\partial Y} = 2Y = 10.0, \quad \sigma_U = \sqrt{\left(\frac{\partial U}{\partial X}\right)^2 \sigma_X^2 + \left(\frac{\partial U}{\partial Y}\right)^2 \sigma_Y^2} = 10.0$$

$U = 125 \pm 10$

(c) $X = 10.0$, $\sigma_X = 0.5$, $Y = 5.0$, $\sigma_Y = 0.1$, $U = (X + Y^2)/2 = 17.50$

$$\frac{\partial U}{\partial X} = 1/2, \quad \frac{\partial U}{\partial Y} = Y = 5.0, \quad \sigma_U = \sqrt{\left(\frac{\partial U}{\partial X}\right)^2 \sigma_X^2 + \left(\frac{\partial U}{\partial Y}\right)^2 \sigma_Y^2} = 0.56$$

$U = 17.50 \pm 0.56$

3. (a) $s = 55.2$, $\theta = 0.50$, $\sigma_s = 0.1$, $\sigma_\theta = 0.02$, $h = s\tan\theta = 30.1559$

$$\frac{\partial h}{\partial s} = \tan\theta = 0.5463,\ \frac{\partial h}{\partial \theta} = s\sec^2\theta = 71.6742$$

$$\sigma_h = \sqrt{\left(\frac{\partial h}{\partial s}\right)^2 \sigma_s^2 + \left(\frac{\partial h}{\partial \theta}\right)^2 \sigma_\theta^2} = 1.4345$$

$h = 30.2 \pm 1.4$

(b) $$\sigma_h = \sqrt{\left(\frac{\partial h}{\partial s}\right)^2 \sigma_s^2 + \left(\frac{\partial h}{\partial \theta}\right)^2 \sigma_\theta^2},\quad \frac{\partial h}{\partial s} = 0.5463,\quad \frac{\partial h}{\partial \theta} = 71.6742$$

If $\sigma_s = 0.05$ and $\sigma_\theta = 0.02$, then $\sigma_h = 1.43$.

If $\sigma_s = 0.1$ and $\sigma_\theta = 0.01$, then $\sigma_h = 0.72$.

Reducing the uncertainty in θ to 0.01 radians provides the greater reduction.

5. (a) $P_1 = 10.1$, $\sigma_{P_1} = 0.3$, $P_2 = 20.1$, $\sigma_{P_2} = 0.4$, $P_3 = \sqrt{P_1 P_2} = 14.25$

$$\frac{\partial P_3}{\partial P_1} = 0.5\sqrt{P_2/P_1} = 0.705354$$

$$\frac{\partial P_3}{\partial P_2} = 0.5\sqrt{P_1/P_2} = 0.354432$$

$$\sigma_{P_3} = \sqrt{\left(\frac{\partial P_3}{\partial P_1}\right)^2 \sigma_{P_1}^2 + \left(\frac{\partial P_3}{\partial P_2}\right)^2 \sigma_{P_2}^2} = 0.25$$

$P_3 = 14.25 \pm 0.25$ MPa

(b) $$\sigma_{P_3} = \sqrt{\left(\frac{\partial P_3}{\partial P_1}\right)^2 \sigma_{P_1}^2 + \left(\frac{\partial P_3}{\partial P_2}\right)^2 \sigma_{P_2}^2},\quad \frac{\partial P_3}{\partial P_1} = 0.705354,\quad \frac{\partial P_3}{\partial P_2} = 0.354432$$

If $\sigma_{P_1} = 0.2$ and $\sigma_{P_2} = 0.4$, then $\sigma_{P_3} = 0.20$.

If $\sigma_{P_1} = 0.3$ and $\sigma_{P_2} = 0.2$, then $\sigma_{P_3} = 0.22$.

Reducing the uncertainty in P_1 to 0.2 MPa provides the greater reduction.

7. (a) $p = 2.3$, $\sigma_p = 0.2$, $q = 3.1$, $\sigma_q = 0.2$, $f = pq/(p+q) = 1.320$

$$\frac{\partial f}{\partial p} = q^2/(p+q)^2 = 0.329561$$

$$\frac{\partial f}{\partial q} = p^2/(p+q)^2 = 0.181413$$

$$\sigma_f = \sqrt{\left(\frac{\partial f}{\partial p}\right)^2 \sigma_p^2 + \left(\frac{\partial f}{\partial q}\right)^2 \sigma_q^2} = 0.075$$

$f = 1.320 \pm 0.075$ cm

(b) $\sigma_f = \sqrt{\left(\frac{\partial f}{\partial p}\right)^2 \sigma_p^2 + \left(\frac{\partial f}{\partial q}\right)^2 \sigma_q^2}$, $\quad \frac{\partial f}{\partial p} = 0.329561, \quad \frac{\partial f}{\partial q} = 0.181413$

If $\sigma_p = 0.1$ and $\sigma_q = 0.2$, then $\sigma_f = 0.049$.

If $\sigma_p = 0.2$ and $\sigma_q = 0.1$, then $\sigma_f = 0.068$.

Reducing the uncertainty in p to 0.1 cm provides the greater reduction.

9. (a) $C = 1.25,\ \sigma_C = 0.03, L = 1.2,\ \sigma_L = 0.1, A = 1.30,\ \sigma_A = 0.05,$

$M = A/LC = 0.8667$

$\frac{\partial M}{\partial C} = -A/(LC^2) = -0.6933$

$\frac{\partial M}{\partial L} = -A/(L^2C) = -0.7222$

$\frac{\partial M}{\partial A} = 1/(LC) = 0.6667$

$\sigma_M = \sqrt{\left(\frac{\partial M}{\partial C}\right)^2 \sigma_C^2 + \left(\frac{\partial M}{\partial L}\right)^2 \sigma_L^2 + \left(\frac{\partial M}{\partial A}\right)^2 \sigma_A^2} = 0.082$

$M = 0.867 \pm 0.082$

(b) $\sigma_M = \sqrt{\left(\frac{\partial M}{\partial C}\right)^2 \sigma_C^2 + \left(\frac{\partial M}{\partial L}\right)^2 \sigma_L^2 + \left(\frac{\partial M}{\partial A}\right)^2 \sigma_A^2}$,

$\frac{\partial M}{\partial C} = -A/(LC^2) = -0.6933 \qquad \frac{\partial M}{\partial L} = -A/(L^2C) = -0.7222 \qquad \frac{\partial M}{\partial A} = 1/(LC) = 0.6667$

If $\sigma_C = 0.01$, $\sigma_L = 0.1$, and $\sigma_A = 0.05$, then $\sigma_M = 0.080$.

If $\sigma_C = 0.03$, $\sigma_L = 0.05$, and $\sigma_A = 0.05$, then $\sigma_M = 0.053$.

If $\sigma_C = 0.03$, $\sigma_L = 0.1$, and $\sigma_A = 0.01$, then $\sigma_M = 0.076$.

Reducing the uncertainty in L to 0.05 cm provides the greatest reduction.

11. (a) $\tau_0 = 50,\ \sigma_{\tau_0} = 1, w = 1.2,\ \sigma_w = 0.1, k = 0.29,\ \sigma_k = 0.05,$

$\tau = \tau_0(1 - kw) = 32.6$

$\frac{\partial \tau}{\partial \tau_0} = 1 - kw = 0.652$

$\frac{\partial \tau}{\partial k} = -\tau_0 w = -60$

$\frac{\partial \tau}{\partial w} = -\tau_0 k = -14.5$

$\sigma_\tau = \sqrt{\left(\frac{\partial \tau}{\partial \tau_0}\right)^2 \sigma_{\tau_0}^2 + \left(\frac{\partial \tau}{\partial k}\right)^2 \sigma_k^2 + \left(\frac{\partial \tau}{\partial w}\right)^2 \sigma_w^2} = 3.4$

$\tau = 32.6 \pm 3.4$ MPa

(b) $\sigma_\tau = \sqrt{\left(\dfrac{\partial\tau}{\partial\tau_0}\right)^2 \sigma_{\tau_0}^2 + \left(\dfrac{\partial\tau}{\partial k}\right)^2 \sigma_k^2 + \left(\dfrac{\partial\tau}{\partial w}\right)^2 \sigma_w^2}$,

$\dfrac{\partial\tau}{\partial\tau_0} = 1 - kw = 0.652, \qquad \dfrac{\partial\tau}{\partial k} = -\tau_0 w = -60, \qquad \dfrac{\partial\tau}{\partial w} = -\tau_0 k = -14.5$

If $\sigma_{\tau_0} = 0.1$, $\sigma_k = 0.05$, and $\sigma_w = 0.1$, then $\sigma_\tau = 3.3$.

If $\sigma_{\tau_0} = 1.0$, $\sigma_k = 0.025$, and $\sigma_w = 0.1$, then $\sigma_\tau = 2.2$.

If $\sigma_{\tau_0} = 1.0$, $\sigma_k = 0.05$, and $\sigma_w = 0.01$, then $\sigma_\tau = 3.1$.

Reducing the uncertainty in k to 0.025 mm^{-1} provides the greatest reduction.

(c) If $\sigma_{\tau_0} = 0$, $\sigma_k = 0.05$, and $\sigma_w = 0$, $\sigma_\tau = 3.0$. Thus implementing the procedure would reduce the uncertainty in τ only to 3.0 MPa. It is probably not worthwhile to implement the new procedure for a reduction this small.

13. (a) $g = 3867.4$, $\sigma_g = 0.3$, $b = 1037.0$, $\sigma_b = 0.2$, $m = 2650.4$, $\sigma_m = 0.1$, $y = mb/g = 710.68$.

$\dfrac{\partial y}{\partial g} = -mb/g^2 = -0.18376$

$\dfrac{\partial y}{\partial b} = m/g = 0.685318$

$\dfrac{\partial y}{\partial m} = b/g = 0.268139$

$\sigma_y = \sqrt{\left(\dfrac{\partial y}{\partial g}\right)^2 \sigma_g^2 + \left(\dfrac{\partial y}{\partial m}\right)^2 \sigma_m^2 + \left(\dfrac{\partial y}{\partial b}\right)^2 \sigma_b^2} = 0.15$

$y = 710.68 \pm 0.15$ g

(b) $\sigma_y = \sqrt{\left(\dfrac{\partial y}{\partial g}\right)^2 \sigma_g^2 + \left(\dfrac{\partial y}{\partial m}\right)^2 \sigma_m^2 + \left(\dfrac{\partial y}{\partial b}\right)^2 \sigma_b^2}$,

$\dfrac{\partial y}{\partial g} = -mb/g^2 = -0.18376, \quad \dfrac{\partial y}{\partial b} = m/g = 0.685318, \quad \dfrac{\partial y}{\partial m} = b/g = 0.268139$

If $\sigma_g = 0.1$, $\sigma_b = 0.2$, and $\sigma_m = 0.1$, $\sigma_y = 0.14$.

If $\sigma_g = 0.3$, $\sigma_b = 0.1$, and $\sigma_m = 0.1$, $\sigma_y = 0.09$.

If $\sigma_g = 0.3$, $\sigma_b = 0.2$, and $\sigma_m = 0$, $\sigma_y = 0.15$.

Reducing the uncertainty in b to 0.1 g provides the greatest reduction.

15. (a) $F = 800$, $\sigma_F = 1$, $R = 0.75$, $\sigma_R = 0.1$, $L_0 = 25.0$, $\sigma_{L_0} = 0.1$, $L_1 = 30.0$, $\sigma_{L_1} = 0.1$,

$Y = \dfrac{FL_0}{\pi R^2 (L_1 - L_0)} = 2264$

$$\frac{\partial Y}{\partial F} = \frac{L_0}{\pi R^2(L_1 - L_0)} = 2.82942$$
$$\frac{\partial Y}{\partial R} = \frac{-2FL_0}{\pi R^3(L_1 - L_0)} = -6036.1$$
$$\frac{\partial Y}{\partial L_0} = \frac{FL_1}{\pi R^2(L_1 - L_0)^2} = 543.249$$
$$\frac{\partial Y}{\partial L_1} = \frac{-FL_0}{\pi R^2(L_1 - L_0)^2} = -452.707$$
$$\sigma_Y = \sqrt{\left(\frac{\partial Y}{\partial F}\right)^2 \sigma_F^2 + \left(\frac{\partial Y}{\partial R}\right)^2 \sigma_R^2 + \left(\frac{\partial Y}{\partial L_0}\right)^2 \sigma_{L_0}^2 + \left(\frac{\partial Y}{\partial L_1}\right)^2 \sigma_{L_1}^2} = 608$$
$Y = 2264 \pm 608\ \text{N/mm}^2$

(b) $$\sigma_Y = \sqrt{\left(\frac{\partial Y}{\partial F}\right)^2 \sigma_F^2 + \left(\frac{\partial Y}{\partial R}\right)^2 \sigma_R^2 + \left(\frac{\partial Y}{\partial L_0}\right)^2 \sigma_{L_0}^2 + \left(\frac{\partial Y}{\partial L_1}\right)^2 \sigma_{L_1}^2},$$
$$\frac{\partial Y}{\partial F} = \frac{L_0}{\pi R^2(L_1 - L_0)} = 2.82942, \quad \frac{\partial Y}{\partial R} = \frac{-2FL_0}{\pi R^3(L_1 - L_0)} = -6036.1,$$
$$\frac{\partial Y}{\partial L_0} = \frac{FL_1}{\pi R^2(L_1 - L_0)^2} = 543.249, \quad \frac{\partial Y}{\partial L_1} = \frac{-FL_0}{\pi R^2(L_1 - L_0)^2} = -452.707$$
If $\sigma_F = 0$, $\sigma_R = 0.1$, $\sigma_{L_0} = 0.1$, and $\sigma_{L_1} = 0.1$, then $\sigma_Y = 608$.
If $\sigma_F = 1$, $\sigma_R = 0$, $\sigma_{L_0} = 0.1$, and $\sigma_{L_1} = 0.1$, then $\sigma_Y = 71$.
If $\sigma_F = 1$, $\sigma_R = 0.1$, $\sigma_{L_0} = 0$, and $\sigma_{L_1} = 0.1$, then $\sigma_Y = 605$.
If $\sigma_F = 1$, $\sigma_R = 0.1$, $\sigma_{L_0} = 0.1$, and $\sigma_{L_1} = 0$, then $\sigma_Y = 606$.
R is the only variable that substantially affects the uncertainty in Y.

17. $t = 10$, $T = 54.1$, $\sigma_T = 0.2$, $T_0 = 70.1$, $\sigma_{T_0} = 0.2$, $T_a = 35.7$, $\sigma_{T_a} = 0.1$,
$k = \ln(T_0 - T_a)/t - \ln(T - T_a)/t = 0.1\ln(T_0 - T_a) - 0.1\ln(T - T_a) = 0.0626$
$$\frac{\partial k}{\partial T} = \frac{-1}{10(T - T_a)} = -0.00543478$$
$$\frac{\partial k}{\partial T_0} = \frac{1}{10(T_0 - T_a)} = 0.00290698$$
$$\frac{\partial k}{\partial T_a} = \frac{1}{10(T - T_a)} - \frac{1}{10(T_0 - T_a)} = 0.00252781$$
$$\sigma_k = \sqrt{\left(\frac{\partial k}{\partial T}\right)^2 \sigma_T^2 + \left(\frac{\partial k}{\partial T_0}\right)^2 \sigma_{T_0}^2 + \left(\frac{\partial k}{\partial T_a}\right)^2 \sigma_{T_a}^2} = 0.0013$$
$k = 0.0626 \pm 0.0013\ \text{min}^{-1}$

19. (a) No, they both involve the quantities h and r.

(b) $r = 0.9,\ \sigma_r = 0.1,\ h = 1.7,\ \sigma_h = 0.1,\ R = c\dfrac{2\pi r(h+2r)}{\pi r^2(h+4r/3)} = c\dfrac{2h+4r}{hr+4r^2/3} = 2.68c$

$$\frac{\partial R}{\partial r} = -c\frac{2h^2+16rh/3+16r^2/3}{(4r^2/3+rh)^2} = -2.68052c$$

$$\frac{\partial R}{\partial h} = -c\frac{4r^2/3}{(4r^2/3+rh)^2} = -0.158541c$$

$$\sigma_R = \sqrt{\left(\frac{\partial R}{\partial r}\right)^2\sigma_r^2 + \left(\frac{\partial R}{\partial h}\right)^2\sigma_h^2} = 0.27c$$

$R = 2.68c \pm 0.27c$

21. $\mu = 1.49,\ \tau = 35.2,\ \sigma_\tau = 0.1,\ h = 12.0,\ \sigma_h = 0.3,\ V = \tau h/\mu = \tau h/1.49 = 283.49,$

$\ln V = \ln\tau + \ln h - \ln 1.49$

$$\frac{\partial \ln V}{\partial \tau} = 1/\tau = 0.028409$$

$$\frac{\partial \ln V}{\partial h} = 1/h = 0.083333$$

$$\sigma_{\ln V} = \sqrt{\left(\frac{\partial \ln V}{\partial \tau}\right)^2\sigma_\tau^2 + \left(\frac{\partial \ln V}{\partial h}\right)^2\sigma_h^2} = 0.025$$

$V = 283.49$ mm/s $\pm$ 2.5%

23. $p = 4.3,\ \sigma_p = 0.1,\ q = 2.1,\ \sigma_q = 0.2,\ f = pq/(p+q) = 1.41,\ \ln f = \ln p + \ln q - \ln(p+q)$

$$\frac{\partial \ln f}{\partial p} = 1/p - 1/(p+q) = 0.0763081$$

$$\frac{\partial \ln f}{\partial q} = 1/q - 1/(p+q) = 0.31994$$

$$\sigma_{\ln f} = \sqrt{\left(\frac{\partial \ln f}{\partial p}\right)^2\sigma_p^2 + \left(\frac{\partial \ln f}{\partial q}\right)^2\sigma_q^2} = 0.064$$

$f = 1.41$ cm $\pm$ 6.4%

25. $\theta_1 = 0.216,\ \sigma_{\theta_1} = 0.003,\ \theta_2 = 0.456,\ \sigma_{\theta_2} = 0.005,\ n = \dfrac{\sin\theta_1}{\sin\theta_2} = 0.487,\ \ln n = \ln(\sin\theta_1) - \ln(\sin\theta_2)$

$$\frac{\partial \ln n}{\partial \theta_1} = \cot\theta_1 = 4.5574$$

$$\frac{\partial \ln n}{\partial \theta_2} = -\cot\theta_2 = -2.03883$$

$$\sigma_{\ln n} = \sqrt{\left(\frac{\partial \ln n}{\partial \theta_1}\right)^2 \sigma_{\theta_1}^2 + \left(\frac{\partial \ln n}{\partial \theta_2}\right)^2 \sigma_{\theta_2}^2} = 0.017$$

$n = 0.487 \pm 1.7\%$

27. $F = 750,\ \sigma_F = 1,\ R = 0.65,\ \sigma_R = 0.09,\ L_0 = 23.7,\ \sigma_{L_0} = 0.2,\ L_1 = 27.7,\ \sigma_{L_1} = 0.2,$

$$Y = \frac{FL_0}{\pi R^2(L_1 - L_0)} = 3347.9,\ \ln Y = \ln F + \ln L_0 - \ln\pi - 2\ln R - \ln(L_1 - L_0)$$

$$\frac{\partial \ln Y}{\partial F} = 1/F = 0.001333$$

$$\frac{\partial \ln Y}{\partial L_0} = 1/L_0 + 1/(L_1 - L_0) = 0.292194$$

$$\frac{\partial \ln Y}{\partial R} = -2/R = -3.07692$$

$$\frac{\partial \ln Y}{\partial L_1} = -1/(L_1 - L_0) = -0.25$$

$$\sigma_{\ln Y} = \sqrt{\left(\frac{\partial \ln Y}{\partial F}\right)^2 \sigma_F^2 + \left(\frac{\partial \ln Y}{\partial L_0}\right)^2 \sigma_{L_0}^2 + \left(\frac{\partial \ln Y}{\partial R}\right)^2 \sigma_R^2 + \left(\frac{\partial \ln Y}{\partial L_1}\right)^2 \sigma_{L_1}^2} = 0.29$$

$Y = 3347.9\ \text{N/mm}^2 \pm 29\%$

29. $r = 0.8,\ \sigma_r = 0.1,\ h = 1.9,\ \sigma_h = 0.1$

(a) $S = 2\pi r(h + 2r) = 17.59,\ \ln S = \ln(2\pi) + \ln r + \ln(h + 2r)$

$$\frac{\partial \ln S}{\partial r} = \frac{1}{r} + \frac{2}{h + 2r} = 1.82143$$

$$\frac{\partial \ln S}{\partial h} = \frac{1}{h + 2r} = 0.285714$$

$$\sigma_{\ln S} = \sqrt{\left(\frac{\partial \ln S}{\partial r}\right)^2 \sigma_r^2 + \left(\frac{\partial \ln S}{\partial h}\right)^2 \sigma_h^2} = 0.18$$

$S = 17.59\ \mu\text{m} \pm 18\%$

(b) $V = \pi r^2(h + 4r/3) = 5.965,\ \ln V = \ln(\pi) + 2\ln r + \ln(h + 4r/3)$

$$\frac{\partial \ln V}{\partial r} = \frac{2}{r} + \frac{4}{3h + 4r} = 2.94944$$

$$\frac{\partial \ln V}{\partial h} = \frac{3}{3h + 4r} = 0.337079$$

$$\sigma_{\ln V} = \sqrt{\left(\frac{\partial \ln V}{\partial r}\right)^2 \sigma_r^2 + \left(\frac{\partial \ln V}{\partial h}\right)^2 \sigma_h^2} = 0.30$$

$V = 5.965\ \mu\text{m}^3 \pm 30\%$

(c) $R = c\dfrac{2\pi r(h+2r)}{\pi r^2(h+4r/3)} = c\dfrac{2h+4r}{rh+4r^2/3} = 2.95c$, $\ln R = \ln c + \ln(2h+4r) - \ln(rh+4r^2/3)$

$$\frac{\partial \ln R}{\partial r} = \frac{2}{2r+h} - \frac{8r+3h}{4r^2+3rh} = -1.12801$$

$$\frac{\partial \ln R}{\partial h} = \frac{1}{2r+h} - \frac{3}{4r+3h} = -0.0513644$$

$$\sigma_{\ln R} = \sqrt{\left(\frac{\partial \ln R}{\partial r}\right)^2 \sigma_r^2 + \left(\frac{\partial \ln R}{\partial h}\right)^2 \sigma_h^2} = 0.11$$

$R = 2.95c \pm 11\%$

Note that the uncertainty in R cannot be determined directly from the uncertainties in S and V, because S and V are not independent.

(d) No

31. $R = kld^{-2}$. The relative uncertainties are $\dfrac{\sigma_k}{k} = 0$ (since k is a constant), $\dfrac{\sigma_l}{l} = 0.03$, and $\dfrac{\sigma_d}{d} = 0.02$.

$$\frac{\sigma_R}{R} = \sqrt{\left(\frac{\sigma_k}{k}\right)^2 + \left(\frac{\sigma_l}{l}\right)^2 + \left(-2\frac{\sigma_d}{d}\right)^2} = \sqrt{0^2 + 0.03^2 + (-0.04)^2} = 0.05.$$

The relative uncertainty is 5.0%.

Supplementary Exercises for Chapter 3

1. (a) $X = 25.0$, $\sigma_X = 1$, $Y = 5.0$, $\sigma_Y = 0.3$, $Z = 3.5$, $\sigma_Z = 0.2$. Let $U = XY + Z$.

$$\frac{\partial U}{\partial X} = Y = 5.0, \quad \frac{\partial U}{\partial Y} = X = 25.0, \quad \frac{\partial U}{\partial Z} = 1$$

$$\sigma_U = \sqrt{\left(\frac{\partial U}{\partial X}\right)^2 \sigma_X^2 + \left(\frac{\partial U}{\partial Y}\right)^2 \sigma_Y^2 + \left(\frac{\partial U}{\partial Z}\right)^2 \sigma_Z^2} = 9.0$$

(b) $X = 25.0$, $\sigma_X = 1$, $Y = 5.0$, $\sigma_Y = 0.3$, $Z = 3.5$, $\sigma_Z = 0.2$. Let $U = Z/(X+Y)$.

$$\frac{\partial U}{\partial X} = -Z/(X+Y)^2 = -0.003889, \quad \frac{\partial U}{\partial Y} = -Z/(X+Y)^2 = -0.003889, \quad \frac{\partial U}{\partial Z} = 1/(X+Y) = 0.03333$$

$$\sigma_U = \sqrt{\left(\frac{\partial U}{\partial X}\right)^2 \sigma_X^2 + \left(\frac{\partial U}{\partial Y}\right)^2 \sigma_Y^2 + \left(\frac{\partial U}{\partial Z}\right)^2 \sigma_Z^2} = 0.0078$$

(c) $X = 25.0$, $\sigma_X = 1$, $Y = 5.0$, $\sigma_Y = 0.3$, $Z = 3.5$, $\sigma_Z = 0.2$. Let $U = \sqrt{X(\ln Y + Z)}$.

$$\frac{\partial U}{\partial X} = \frac{1}{2}\sqrt{\frac{\ln Y + Z}{X}} = 0.226041, \quad \frac{\partial U}{\partial Y} = \frac{1}{2Y}\sqrt{\frac{X}{\ln Y + Z}} = 0.221199, \quad \frac{\partial U}{\partial Z} = \frac{1}{2}\sqrt{\frac{X}{\ln Y + Z}} = 1.105996,$$

$$\sigma_U = \sqrt{\left(\frac{\partial U}{\partial X}\right)^2 \sigma_X^2 + \left(\frac{\partial U}{\partial Y}\right)^2 \sigma_Y^2 + \left(\frac{\partial U}{\partial Z}\right)^2 \sigma_Z^2} = 0.32$$

(d) $X = 25.0$, $\sigma_X = 1$, $Y = 5.0$, $\sigma_Y = 0.3$, $Z = 3.5$, $\sigma_Z = 0.2$. Let $U = Xe^{Z^2-2Y}$.

$$\frac{\partial U}{\partial X} = e^{Z^2-2Y} = 9.487736, \quad \frac{\partial U}{\partial Y} = -2Xe^{Z^2-2Y} = -474.38679, \quad \frac{\partial U}{\partial Z} = 2XZe^{Z^2-2Y} = -1660.353771$$

$$\sigma_U = \sqrt{\left(\frac{\partial U}{\partial X}\right)^2 \sigma_X^2 + \left(\frac{\partial U}{\partial Y}\right)^2 \sigma_Y^2 + \left(\frac{\partial U}{\partial Z}\right)^2 \sigma_Z^2} = 361.41$$

3. (a) Let X, Y, and Z represent the measured lengths of the components, and let $T = X + Y + Z$ be the combined length. Then $\sigma_X = \sigma_Y = \sigma_Z = 1.2$.

$$\sigma_T = \sqrt{\sigma_X^2 + \sigma_Y^2 + \sigma_Z^2} = 2.1 \text{ mm}$$

(b) Let σ be the required uncertainty in X and Y. Then $\sigma_T = \sqrt{\sigma^2 + \sigma^2 + \sigma^2} = 0.05$. Solving for σ yields $\sigma = 0.029$ mm.

5. (a) $\gamma = 9800$, $\eta = 0.85$, $\sigma_\eta = 0.02$, $Q = 60$, $\sigma_Q = 1$, $H = 3.71$, $\sigma_H = 0.10$

$$P = \eta\gamma QH = 9800\eta QH = 1.854 \times 10^6$$

$$\frac{\partial P}{\partial \eta} = 9800QH = 2.18148 \times 10^6$$

$$\frac{\partial P}{\partial Q} = 9800\eta H = 30904.3$$

$$\frac{\partial P}{\partial H} = 9800\eta Q = 499800$$

$$\sigma_P = \sqrt{\left(\frac{\partial P}{\partial \eta}\right)^2 \sigma_\eta^2 + \left(\frac{\partial P}{\partial Q}\right)^2 \sigma_Q^2 + \left(\frac{\partial P}{\partial H}\right)^2 \sigma_H^2} = 73188.9$$

$$P = (1.854 \pm 0.073) \times 10^6 \text{ W}$$

(b) The relative uncertainty is $\dfrac{\sigma_P}{P} = \dfrac{0.073 \times 10^6}{1.854 \times 10^6} = 0.039 = 3.9\%$.

(c) $\sigma_P = \sqrt{\left(\frac{\partial P}{\partial \eta}\right)^2 \sigma_\eta^2 + \left(\frac{\partial P}{\partial Q}\right)^2 \sigma_Q^2 + \left(\frac{\partial P}{\partial H}\right)^2 \sigma_H^2}$,

$\frac{\partial P}{\partial \eta} = 9800QH = 2.18148 \times 10^6$, $\quad \frac{\partial P}{\partial Q} = 9800\eta H = 30904.3$, $\quad \frac{\partial P}{\partial H} = 9800\eta Q = 499800$

If $\sigma_\eta = 0.01$, $\sigma_Q = 1$, and $\sigma_H = 0.1$, then $\sigma_P = 6.3 \times 10^4$.

If $\sigma_\eta = 0.02$, $\sigma_Q = 0.5$, and $\sigma_H = 0.1$, then $\sigma_P = 6.8 \times 10^4$.

If $\sigma_\eta = 0.02$, $\sigma_Q = 1$, and $\sigma_H = 0.05$, then $\sigma_P = 5.9 \times 10^4$.

Reducing the uncertainty in H to 0.05 m provides the greatest reduction.

7. (a) $H = 4$, $c = 0.2390$, $\Delta T = 2.75$, $\sigma_{\Delta T} = 0.02$, $m = 0.40$, $\sigma_m = 0.01$

$$C = \frac{cH(\Delta T)}{m} = \frac{0.956(\Delta T)}{m} = 6.57$$

$$\frac{\partial C}{\partial \Delta T} = \frac{0.956}{m} = 2.39$$

$$\frac{\partial C}{\partial m} = \frac{-0.956(\Delta T)}{m^2} = -16.4312$$

$$\sigma_C = \sqrt{\left(\frac{\partial C}{\partial \Delta T}\right)^2 \sigma_{\Delta T}^2 + \left(\frac{\partial C}{\partial m}\right)^2 \sigma_m^2} = 0.17$$

$C = 6.57 \pm 0.17$ kcal.

(b) The relative uncertainty is $\frac{\sigma_C}{C} = \frac{0.17}{6.57} = 0.026 = 2.6\%$.

Alternatively, $\ln C = \ln 0.956 + \ln(\Delta T) - \ln m$, so $\frac{\partial \ln C}{\partial \Delta T} = \frac{1}{\Delta T} = 0.363636$, $\frac{\partial \ln C}{\partial m} = \frac{-1}{m} = -2.5$,

$$\frac{\sigma_C}{C} = \sigma_{\ln C} = \sqrt{\left(\frac{\partial \ln C}{\partial \Delta T}\right)^2 \sigma_{\Delta T}^2 + \left(\frac{\partial \ln C}{\partial m}\right)^2 \sigma_m^2} = 0.026 = 2.6\%$$

(c) $\sigma_C = \sqrt{\left(\frac{\partial C}{\partial \Delta T}\right)^2 \sigma_{\Delta T}^2 + \left(\frac{\partial C}{\partial m}\right)^2 \sigma_m^2}$, $\quad \frac{\partial C}{\partial \Delta T} = \frac{0.956}{m} = 2.39 \quad \frac{\partial C}{\partial m} = \frac{-0.956(\Delta T)}{m^2} = -16.4312$

If $\sigma_{\Delta T} = 0.01$ and $\sigma_m = 0.01$, then $\sigma_C = 0.17$.

If $\sigma_{\Delta T} = 0.02$ and $\sigma_m = 0.005$, then $\sigma_C = 0.10$.

Reducing the uncertainty in the mass to 0.005 g provides the greater reduction.

9. (a) Let X be the measured northerly component of velocity and let Y be the measured easterly component of

velocity. The velocity of the earth's crust is estimated with $V = \sqrt{X^2+Y^2}$.

Now $X = 22.10$, $\sigma_X = 0.34$ and $Y = 14.3$, $\sigma_Y = 0.32$, so $V = \sqrt{22.10^2+14.3^2} = 26.32$, and

$$\frac{\partial V}{\partial X} = X/\sqrt{X^2+Y^2} = 0.83957$$

$$\frac{\partial V}{\partial Y} = Y/\sqrt{X^2+Y^2} = 0.543251$$

$$\sigma_V = \sqrt{\left(\frac{\partial V}{\partial X}\right)^2 \sigma_X^2 + \left(\frac{\partial V}{\partial Y}\right)^2 \sigma_Y^2} = 0.33$$

$V = 26.32 \pm 0.33$ mm/year

(b) Let T be the estimated number of years it will take for the earth's crust to move 100 mm.

By part (a), $V = 26.3230$, $\sigma_V = 0.334222$. We are using extra precision in V and σ_V to get good precision for T and σ_T.

Now $T = 100/V = 3.799$, $\quad \frac{dT}{dV} = -100/V^2 = -0.144321$

$$\sigma_T = \left|\frac{dT}{dV}\right| \sigma_V = 0.144321(0.334222) = 0.048$$

$T = 3.799 \pm 0.048$ years

11. Let X_1 and X_2 represent the estimated thicknesses of the outer layers, and let Y_1, Y_2, Y_3 represent the estimated thicknesses of the inner layers. Then $X_1 = X_2 = 0.50$, $Y_1 = Y_2 = Y_3 = 6.25$, $\sigma_{X_i} = 0.02$ for $i = 1,2$, and $\sigma_{Y_j} = 0.05$ for $j = 1,2,3$.

The estimated thickness of the item is $T = X_1 + X_2 + Y_1 + Y_2 + Y_3 = 19.75$.

The uncertainty in the estimate is $\sigma_T = \sqrt{\sigma_{X_1}^2 + \sigma_{X_2}^2 + \sigma_{Y_1}^2 + \sigma_{Y_2}^2 + \sigma_{Y_3}^2} = 0.09$.

The thickness is 19.75 ± 0.09 mm.

13. (a) The relative uncertainty in λ is $\sigma_\lambda/\lambda = \sigma_{\ln\lambda}$. The given relative uncertainties are $\sigma_{\ln V} = 0.0001$, $\sigma_{\ln I} = 0.0001$, $\sigma_{\ln A} = 0.001$, $\sigma_{\ln l} = 0.01$, $\sigma_{\ln a} = 0.01$. $\ln\lambda = \ln V + \ln I + \ln A - \ln\pi - \ln l - \ln a$.

$$\sigma_{\ln\lambda} = \sqrt{\sigma_{\ln V}^2 + \sigma_{\ln I}^2 + \sigma_{\ln A}^2 + \sigma_{\ln l}^2 + \sigma_{\ln a}^2} = \sqrt{0.0001^2 + 0.0001^2 + 0.001^2 + 0.01^2 + 0.01^2} = 0.014 = 1.4\%$$

(b) If $\sigma_{\ln V} = 0.0001$, $\sigma_{\ln I} = 0.0001$, $\sigma_{\ln A} = 0.001$, $\sigma_{\ln l} = 0.005$, and $\sigma_{\ln a} = 0.01$, then $\sigma_{\ln V} = 0.011$.

If $\sigma_{\ln V} = 0$, $\sigma_{\ln I} = 0$, $\sigma_{\ln A} = 0$, $\sigma_{\ln l} = 0.01$, and $\sigma_{\ln a} = 0.01$, then $\sigma_{\ln V} = 0.014$.

Reducing the uncertainty in l to 0.5% provides the greater reduction.

15. (a) Yes, since the strengths of the wires are all estimated to be the same, the sum of the strengths is the same as the strength of one of them multiplied by the number of wires. Therefore the estimated strength is 80,000 pounds in both cases.

    (b) No, for the ductile wire method the squares of the uncertainties of the 16 wires are added, to obtain $\sigma = \sqrt{16 \times 20^2} = 80$. For the brittle wire method, the uncertainty in the strength of the weakest wire is multiplied by the number of wires, to obtain $\sigma = 16 \times 20 = 320$.

17. (a) $r = 3.00,\ \sigma_r = 0.03,\ v = 4.0,\ \sigma_v = 0.2.\ Q = \pi r^2 v = 113.1.$

$$\frac{\partial Q}{\partial r} = 2\pi r v = 75.3982$$

$$\frac{\partial Q}{\partial v} = \pi r^2 = 28.2743$$

$$\sigma_Q = \sqrt{\left(\frac{\partial Q}{\partial r}\right)^2 \sigma_r^2 + \left(\frac{\partial Q}{\partial v}\right)^2 \sigma_v^2} = 6.1$$

$Q = 113.1 \pm 6.1\ \mathrm{m^3/s}$

    (b) $r = 4.00,\ \sigma_r = 0.04,\ v = 2.0,\ \sigma_v = 0.1.\ Q = \pi r^2 v = 100.5.$

$$\frac{\partial Q}{\partial r} = 2\pi r v = 50.2655$$

$$\frac{\partial Q}{\partial v} = \pi r^2 = 50.2655$$

$$\sigma_Q = \sqrt{\left(\frac{\partial Q}{\partial r}\right)^2 \sigma_r^2 + \left(\frac{\partial Q}{\partial v}\right)^2 \sigma_v^2} = 5.4$$

$Q = 100.5 \pm 5.4\ \mathrm{m^3/s}$

    (c) The relative uncertainty is $\dfrac{\sigma_Q}{Q} = \sigma_{\ln Q}$.

$$\ln Q = \ln \pi + 2 \ln r + \ln v, \quad \sigma_{\ln r} = \frac{\sigma_r}{r} = 0.01, \quad \sigma_{\ln v} = \frac{\sigma_v}{v} = 0.05, \quad \frac{\partial \ln Q}{\partial r} = \frac{2}{r}, \quad \frac{\partial \ln Q}{\partial v} = \frac{1}{v}$$

$$\sigma_{\ln Q} = \sqrt{\left(\frac{\partial \ln Q}{\partial r}\right)^2 \sigma_r^2 + \left(\frac{\partial \ln Q}{\partial v}\right)^2 \sigma_v^2} = \sqrt{2^2 \left(\frac{\sigma_r}{r}\right)^2 + \left(\frac{\sigma_v}{v}\right)^2} = \sqrt{2^2(0.01^2) + 0.05^2} = 0.054$$

Yes, the relative uncertainty in Q can be determined from the relative uncertainties in r and v, and it is equal to 5.4%.

19. (a) $C_0 = 0.03$, $t = 40$, $C = 0.0023$, $\sigma_C = 0.0002$. $k = (1/t)(1/C - 1/C_0) = (1/40)(1/C) - 5/6 = 10.04$

$$\frac{dk}{dC} = \frac{-1}{40C^2} = -4725.90,\ \sigma_k = \left|\frac{dk}{dC}\right|\sigma_C = 4725.90(0.0002) = 0.95$$

$k = 10.04 \pm 0.95\ \text{s}^{-1}$

(b) $C_0 = 0.03$, $t = 50$, $C = 0.0018$, $\sigma_C = 0.0002$. $k = (1/t)(1/C - 1/C_0) = (1/50)(1/C) - 2/3 = 10.4$

$$\frac{dk}{dC} = \frac{-1}{50C^2} = -6172.84,\ \sigma_k = \left|\frac{dk}{dC}\right|\sigma_C = 6172.84(0.0002) = 1.2$$

$k = 10.4 \pm 1.2\ \text{s}^{-1}$

(c) Let $\overline{k} = (\widehat{k}_1 + \widehat{k}_2)/2 = 0.5\widehat{k}_1 + 0.5\widehat{k}_2$. From parts (a) and (b), $\sigma_{\widehat{k}_1} = 0.945180$ and $\sigma_{\widehat{k}_2} = 1.234568$. We are using extra precision in $\sigma_{\widehat{k}_1}$ and $\sigma_{\widehat{k}_2}$ in order to get good precision in $\sigma_{\overline{k}}$.

$$\sigma_{\overline{k}} = \sqrt{0.5^2\sigma^2_{\widehat{k}_1} + 0.5^2\sigma^2_{\widehat{k}_2}} = \sqrt{0.5^2(0.945180^2) + 0.5^2(1.234568^2)} = 0.78$$

(d) The value of c is $c_{\text{best}} = \dfrac{\sigma^2_{\widehat{k}_2}}{\sigma^2_{\widehat{k}_1} + \sigma^2_{\widehat{k}_2}} = \dfrac{1.234568^2}{0.945180^2 + 1.234568^2} = 0.63.$

21. (a) Let s be the measured side of the square. Then $s = 181.2$, $\sigma_s = 0.1$.

The estimated area is $S = s^2 = 32,833$.

$$\frac{dS}{ds} = 2s = 362.4, \quad \sigma_S = \left|\frac{dS}{ds}\right|\sigma_s = (362.4)(0.1) = 36$$

$S = 32,833 \pm 36\ \text{m}^2$

(b) The estimated area of a semicircle is $C = (\pi s^2)/8 = 12,894$.

$$\frac{dC}{ds} = \pi s/4 = 142.314, \quad \sigma_C = \left|\frac{dC}{ds}\right|\sigma_s = (142.314)(0.1) = 14$$

$C = 12,894 \pm 14\ \text{m}^2$

(c) This is not correct. Let s denote the length of a side of the square. Since S and C are both computed in terms of s, they are not independent. In order to compute σ_A correctly, we must express A directly in terms of s: $A = s^2 + 2\pi s^2/8 = s^2(1 + \pi/4)$. So $\sigma_A = \left|\dfrac{dA}{ds}\right|\sigma_s = 2s(1 + \pi/4)\sigma_s = 65\ \text{m}^2$.

23. (a) $P_1 = 8.1,\ \sigma_{P_1} = 0.1,\ P_2 = 15.4,\ \sigma_{P_2} = 0.2,\ P = \sqrt{P_1 P_2} = 11.16871$

$$\frac{\partial P_3}{\partial P_1} = \frac{P_2}{2\sqrt{P_1 P_2}} = 0.689426$$

$$\frac{\partial P_3}{\partial P_2} = \frac{P_1}{2\sqrt{P_1 P_2}} = 0.36262$$

$$\sigma_{P_3} = \sqrt{\left(\frac{\partial P_3}{\partial P_1}\right)^2 \sigma_{P_1}^2 + \left(\frac{\partial P_3}{\partial P_2}\right)^2 \sigma_{P_2}^2} = 0.10$$

$P_3 = 11.16871 \pm 0.10$ MPa

(b) $\dfrac{\partial^2 P_3}{\partial P_1^2} = -(0.25)P_1^{-3/2}P_2^{1/2} = -0.042557,\quad \dfrac{\partial^2 P_3}{\partial P_2^2} = -(0.25)P_2^{-3/2}P_1^{1/2} = -0.011773$

The bias corrected estimate is

$$P_3 - \left(\frac{1}{2}\right)\left[\left(\frac{\partial^2 P_3}{\partial P_1^2}\right)\sigma_{P_1}^2 + \left(\frac{\partial^2 P_3}{\partial P_2^2}\right)\sigma_{P_2}^2\right] = 11.16871 - (0.5)[-0.042557(0.1^2) - 0.011773(0.2^2)] = 11.16916.$$

(c) No. The difference between the two estimates is much less than the uncertainty.

Chapter 4

Section 4.1

1. (a) $X \sim \text{Bernoulli}(p)$, where $p = 0.4$. $\mu_X = p = 0.4$, $\sigma_X^2 = p(1-p) = 0.24$.

   (b) No. A Bernoulli random variable has possible values 0 and 1. The possible values of Y are 0 and 2.

   (c) $Y = 2X$, where $X \sim \text{Bernoulli}(p)$.
   Therefore $\mu_Y = 2\mu_X = 2p = 0.80$, and $\sigma_Y^2 = 2^2\sigma_X^2 = 4p(1-p) = 0.96$.

3. (a) $p_X = 0.05$

   (b) $p_Y = 0.20$

   (c) $p_Z = 0.23$

   (d) Yes, it is possible for there to be both discoloration and a crack.

   (e) No. $p_Z = P(X = 1 \text{ or } Y = 1) \neq P(X = 1) + P(Y = 1)$ because $P(X = 1 \text{ and } Y = 1) \neq 0$.

   (f) No. If the surface has both discoloration and a crack, then $X = 1$, $Y = 1$, and $Z = 1$, but $X + Y = 2$.

5. (a) $p_X = 1/2$

   (b) $p_Y = 1/2$

   (c) $p_Z = 1/4$

   (d) Yes. $P(X = x \text{ and } Y = y) = P(X = x)P(Y = y)$ for all values of x and y.

   (e) Yes. $p_Z = 1/4 = (1/2)(1/2) = p_X p_Y$.

   (f) Yes. If both coins come up heads, then $X = 1$, $Y = 1$, and $Z = 1$, so $Z = XY$. If not, then $Z = 0$, and either X, Y, or both are equal to 0 as well, so again $Z = XY$.

7. (a) Since the possible values of X and Y are 0 and 1, the possible values of the product $Z = XY$ are also 0 and 1. Therefore Z is a Bernoulli random variable.

(b) $p_Z = P(Z = 1) = P(XY = 1) = P(X = 1 \text{ and } Y = 1) = P(X = 1)P(Y = 1) = p_X p_Y$.

Section 4.2

1. (a) $P(X = 1) = \dfrac{7!}{1!(7-1)!}(0.3)^1(1-0.3)^{7-1} = 0.2471$

(b) $P(X = 2) = \dfrac{7!}{2!(7-2)!}(0.3)^2(1-0.3)^{7-2} = 0.3177$

(c) $P(X < 1) = P(X = 0) = \dfrac{7!}{0!(7-0)!}(0.3)^0(1-0.3)^{7-0} = 0.0824$

(d)
$$\begin{aligned} P(X > 4) &= P(X = 5) + P(X = 6) + P(X = 7) \\ &= \frac{7!}{5!(7-5)!}(0.3)^5(1-0.3)^{7-5} + \frac{7!}{6!(7-6)!}(0.3)^6(1-0.3)^{7-6} + \frac{7!}{7!(7-7)!}(0.3)^7(1-0.3)^{7-7} \\ &= 0.002500 + 0.003572 + 0.0002187 \\ &= 0.0288 \end{aligned}$$

(e) $\mu_X = (7)(0.3) = 2.1$

(f) $\sigma_X^2 = (7)(0.3)(1-0.3) = 1.47$

3. (a) $P(X = 2) = \dfrac{4!}{2!(4-2)!}(0.6)^2(1-0.6)^{4-2} = 0.3456$

(b)
$$\begin{aligned} P(X > 2) &= 1 - P(X \le 2) \\ &= 1 - P(X = 0) - P(X = 1) - P(X = 2) \\ &= 1 - \frac{8!}{0!(8-0)!}(0.2)^0(1-0.2)^{8-0} - \frac{8!}{1!(8-1)!}(0.2)^1(1-0.2)^{8-1} - \frac{8!}{2!(8-2)!}(0.2)^2(1-0.2)^{8-2} \\ &= 0.2031 \end{aligned}$$

(c)
$$\begin{aligned}P(X \leq 2) &= P(X=0)+P(X=1)+P(X=2)\\ &= \frac{5!}{0!(5-0)!}(0.4)^0(1-0.4)^{5-0}+\frac{5!}{1!(5-1)!}(0.4)^1(1-0.4)^{5-1}+\frac{5!}{2!(5-2)!}(0.4)^2(1-0.4)^{5-2}\\ &= 0.6826\end{aligned}$$

(d)
$$\begin{aligned}P(3 \leq X \leq 5) &= P(X=3)+P(X=4)+P(X=5)\\ &= \frac{6!}{3!(6-3)!}(0.7)^3(1-0.7)^{6-3}+\frac{6!}{4!(6-4)!}(0.7)^4(1-0.7)^{6-4}+\frac{6!}{5!(6-5)!}(0.7)^5(1-0.7)^{6-5}\\ &= 0.8119\end{aligned}$$

5. Let X be the number of automobiles that violate the standard. Then $X \sim \text{Bin}(12, 0.1)$.

(a) $P(X=3) = \frac{12!}{3!(12-3)!}(0.1)^3(1-0.1)^{12-3} = 0.0852$

(b) $P(X<3) = P(X=0)+P(X=1)+P(X=2) = \frac{10!}{0!(10-0)!}(0.75)^0(1-0.75)^{10-0}+\frac{10!}{1!(10-1)!}(0.75)^1(1-0.75)^{10-1}+\frac{10!}{2!(10-2)!}(0.75)^2(1-0.75)^{10-2} = 0.2824+0.3766+0.2301 = 0.8891$

(c) $P(X=0) = \frac{12!}{0!(12-0)!}(0.1)^0(1-0.1)^{12-0} = 0.2824$

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

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

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

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

9. Let X be the number of women among the five winners. Then $X \sim \text{Bin}(5, 0.6)$.

(a) $P(X \leq 3) = 1 - P(X > 3) = 1 - P(X=4) - P(X=5) = 1 - \dfrac{5!}{4!(5-4)!}(0.6)^4(1-0.6)^{5-4} - \dfrac{5!}{5!(5-5)!}(0.6)^5(1-0.6)^{5-5} = 1 - 0.25920 - 0.07776 = 0.6630$

(b) $P(\text{3 of one gender and 2 of another}) = P(X=2) + P(X=3) = \dfrac{5!}{2!(5-2)!}(0.6)^2(1-0.6)^{5-2} + \dfrac{5!}{3!(5-3)!}(0.6)^3(1-0.6)^{5-3} = 0.2304 + 0.3456 = 0.5760$

11. Let X be the number of defective parts in the sample from vendor A, and let Y be the number of defective parts in the sample from vendor B. Let p_A be the probability that a part from vendor A is defective, and let p_B be the probability that a part from vendor B is defective. Then $X \sim \text{Bin}(100, p_A)$ and $Y \sim \text{Bin}(200, p_B)$. The observed value of X is 12 and the observed value of Y is 10.

(a) $\widehat{p}_A = X/100 = 12/100 = 0.12, \quad \sigma_{\widehat{p}_A} = \sqrt{\dfrac{p_A(1-p_A)}{100}}$

Replacing p_A with $\widehat{p}_A = 0.12$ and n with 100, $\sigma_{\widehat{p}_A} = \sqrt{\dfrac{0.12(1-0.12)}{100}} = 0.032$.

(b) $\widehat{p}_B = Y/200 = 10/200 = 0.05, \quad \sigma_{\widehat{p}_B} = \sqrt{\dfrac{p_B(1-p_B)}{200}}$

Replacing p_B with $\widehat{p}_B = 0.05$ and n with 200, $\sigma_{\widehat{p}_B} = \sqrt{\dfrac{0.05(1-0.05)}{200}} = 0.015$.

(c) The difference is estimated with $\widehat{p}_A - \widehat{p}_B = 0.12 - 0.05 = 0.07$.

The uncertainty is $\sqrt{\sigma^2_{\widehat{p}_A} + \sigma^2_{\widehat{p}_B}} = 0.036$.

13. (a) $P(\text{can be used}) = P(\text{meets spec}) + P(\text{long}) = 0.90 + 0.06 = 0.96$.

(b) Let X be the number of bolts out of 10 that can be used. Then $X \sim \text{Bin}(10, 0.96)$.

$$\begin{aligned}
P(X < 9) &= 1 - P(X \geq 9) \\
&= 1 - P(X = 9) - P(X = 10) \\
&= 1 - \frac{10!}{9!(10-9)!}(0.96)^9(1-0.96)^{10-9} - \frac{10!}{10!(10-10)!}(0.96)^{10}(1-0.96)^{10-10} \\
&= 1 - 0.27701 - 0.66483 \\
&= 0.0582
\end{aligned}$$

15. (a) Let Y be the number of lights that are green. Then $Y \sim \text{Bin}(3, 0.6)$.

$P(Y = 3) = \dfrac{3!}{3!(3-3)!}(0.6)^3(1-0.6)^{3-3} = 0.216$

(b) $X \sim \text{Bin}(5, 0.216)$

(c) $P(X = 3) = \dfrac{5!}{3!(5-3)!}(0.216)^3(1-0.216)^{5-3} = 0.0619$

17. (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 $n = 4$.

19. (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.

21. (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 $X \leq 2$.

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{5!}{0!(5-0)!}(0.3)^0(1-0.3)^{5-0} + \frac{5!}{1!(5-1)!}(0.3)^1(1-0.3)^{5-1}$$
$$+ \frac{5!}{2!(5-2)!}(0.3)^2(1-0.3)^{5-2} = 0.8369$$

(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 $n = 9$.

23. (a) $Y = 10X + 3(100 - X) = 7X + 300$

(b) $X \sim \text{Bin}(100, 0.9)$, so $\mu_X = 100(0.9) = 90$.
$\mu_Y = 7\mu_X + 300 = 7(90) + 300 = \$930$

(c) $\sigma_X = \sqrt{100(0.9)(0.1)} = 3$. $\sigma_Y = 7\sigma_X = \$21$

25. Let p be the probability that a tire has no flaw. The results of Example 4.14 show that the sample proportion is $\widehat{p} = 93/100 = 0.93$ and $\sigma_{\widehat{p}} = 0.0255$. Let X be the number of tires out of four that have no flaw. Then $X \sim \text{Bin}(4, p)$. Let q be the probability that exactly one of four tires has a flaw.

Then $q = P(X = 3) = \dfrac{4!}{3!(4-3)!}(p)^3(1-p)^{4-3} = 4p^3(1-p) = 4p^3 - 4p^4$.

The estimate of q is $\widehat{q} = 4\widehat{p}^3 - 4\widehat{p}^4 = 4(0.93)^3 - 4(0.93)^4 = 0.225$.

$\dfrac{d\widehat{q}}{d\widehat{p}} = 12\widehat{p}^2 - 16\widehat{p}^3 = 12(0.93)^2 - 16(0.93)^3 = -2.4909$

$\sigma_{\widehat{q}} = \left|\dfrac{d\widehat{q}}{d\widehat{p}}\right|\sigma_{\widehat{p}} = (2.4909)(0.0255) = 0.064$

The probability is 0.225 ± 0.064.

Section 4.3

1. (a) $P(X = 1) = e^{-4}\dfrac{4^1}{1!} = 0.0733$

(b) $P(X = 0) = e^{-4}\dfrac{4^0}{0!} = 0.0183$

(c) $P(X < 2) = P(X = 0) + P(X = 1) = e^{-4}\dfrac{4^0}{0!} + e^{-4}\dfrac{4^1}{1!} = 0.0183 + 0.0733 = 0.0916$

(d) $P(X > 1) = 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1) = 1 - 0.0183 - 0.0733 = 0.9084$

(e) Since $X \sim \text{Poisson}(4)$, $\mu_X = 4$.

(f) Since $X \sim \text{Poisson}(4)$, $\sigma_X = \sqrt{4} = 2$.

3. Since the mean number per mile is 3, the mean number per two miles is (2)(3) = 6. Therefore $X \sim \text{Poisson}(6)$,

(a) $P(X = 4) = e^{-6}\dfrac{6^4}{4!} = 0.1339$

(b)
$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) \\ &= e^{-6}\frac{6^0}{0!} + e^{-6}\frac{6^1}{1!} \\ &= 0.0024788 + 0.014873 \\ &= 0.0174 \end{aligned}$$

(c) $$\begin{aligned} P(5 \leq X < 8) &= P(X = 5) + P(X = 6) + P(X = 7) \\ &= e^{-6}\frac{6^5}{5!} + e^{-6}\frac{6^6}{6!} + e^{-6}\frac{6^7}{7!} \\ &= 0.16062 + 0.16062 + 0.13768 \\ &= 0.4589 \end{aligned}$$

(d) Since $X \sim \text{Poisson}(6)$, $\mu_X = 6$.

(e) Since $X \sim \text{Poisson}(6)$, $\sigma_X = \sqrt{6} = 2.45$.

5. Let X be the number of servers that fail. Then X is the number of successes in $n = 1000$ Bernoulli trials, each of which has success probability $p = 0.003$. The mean of X is $np = (1000)(0.003) = 3$. Since n is large and p is small, $X \sim \text{Poisson}(3)$ to a very close approximation.

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

(b) The event that fewer than 998 servers function is the same as the event that more than 2 servers fail, or equivalently, that $X > 2$.

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) = 1 - e^{-3}\frac{3^0}{0!} - e^{-3}\frac{3^1}{1!} - e^{-3}\frac{3^2}{2!} \\ &= 1 - 0.04979 - 0.14936 - 0.22404 \\ &= 0.5768 \end{aligned}$$

(c) $\mu_X = 3$

(d) $\sigma_X = \sqrt{3} = 1.732$

7. (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, $X \sim \text{Poisson}(2)$.

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$\begin{aligned} &= e^{-2}\frac{2^0}{0!}+e^{-2}\frac{2^1}{1!}+e^{-2}\frac{2^2}{2!} \\ &= 0.13534+0.27067+0.27067 \\ &= 0.6767 \end{aligned}$$

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. Let X represent the number of bacteria observed in 0.5 mL. Let λ represent the true concentration in bacteria per mL. Then $X \sim \text{Poisson}(0.5\lambda)$. The observed value of X is 39. The estimated concentration is $\widehat{\lambda} = 39/0.5 = 78$. The uncertainty is $\sigma_{\widehat{\lambda}} = \sqrt{78/0.5} = 12.49$. $\lambda = 78 \pm 12$.

13. (a) Let N be the number of defective components produced. Then $N \sim \text{Poisson}(20)$.

$$P(N=15) = e^{-20}\frac{20^{15}}{15!} = 0.0516$$

(b) Let X be the number that are repairable. Then $X \sim \text{Bin}(15, 0.6)$.

$$P(X=10) = \frac{15!}{10!(15-10)!}(0.6)^{10}(1-0.6)^{15-10} = 0.1859$$

(c) Given N, $X \sim \text{Bin}(N, 0.6)$.

(d) Let N be the number of defective components, and let X be the number that are repairable.

$$\begin{aligned} P(N=15 \cap X=10) &= P(N=15)P(X=10|N=15) \\ &= \left(e^{-20}\frac{20^{15}}{15!}\right)\left(\frac{15!}{10!(15-10)!}(0.6)^{10}(1-0.6)^{15-10}\right) \\ &= 0.00960 \end{aligned}$$

15. (a) Let X be the number of flaws in a 3-meter board. Since the mean number of flaws is 0.45 per meter, $X \sim \text{Poisson}(1.35)$.

$P(X=0) = e^{-1.35}\frac{1.35^0}{0!}$, so $P(X=0) = 0.2592$.

(b) Let n be the length in meters needed so that the probability of no flaw is 0.5. Then $X \sim \text{Poisson}(0.45n)$, and $P(X = 0) = e^{-0.45n}\dfrac{0.45n^0}{0!} = 0.5$. Solving for n yields $n = 1.54$.

17. Let λ_1 be the mean number of chips per cookie in one of Mom's cookies, and let λ_2 be the mean number of chips per cookie in one of Grandma's cookies. Let X_1 and X_2 be the numbers of chips in two of Mom's cookies, and let Y_1 and Y_2 be the numbers of chips in two of Grandma's cookies. Then $X_i \sim \text{Poisson}(\lambda_1)$ and $Y_i \sim \text{Poisson}(\lambda_2)$. The observed values are $X_1 = 14$, $X_2 = 11$, $Y_1 = 6$, and $Y_2 = 8$.

(a) The estimate is $\widehat{\lambda}_1 = \overline{X} = (14+11)/2 = 12.5$.

(b) The estimate is $\widehat{\lambda}_2 = \overline{Y} = (6+8)/2 = 7.0$.

(c) $\sigma_{X_1} = \sigma_{X_2} = \sqrt{\lambda_1}$. The uncertainty is $\sigma_{\overline{X}} = \sqrt{\lambda_1/2}$. Replacing λ_1 with $\widehat{\lambda}_1$, $\sigma_{\overline{X}} = \sqrt{12.5/2} = 2.5$.

(d) $\sigma_{Y_1} = \sigma_{Y_2} = \sqrt{\lambda_2}$. The uncertainty is $\sigma_{\overline{Y}} = \sqrt{\lambda_2/2}$. Replacing λ_2 with $\widehat{\lambda}_2$, $\sigma_{\overline{Y}} = \sqrt{7.0/2} = 1.9$.

(e) The estimate is $\widehat{\lambda}_1 - \widehat{\lambda}_2 = 12.5 - 7.0 = 5.5$.
The uncertainty is $\sigma_{\widehat{\lambda}_1 - \widehat{\lambda}_2} = \sqrt{\sigma^2_{\widehat{\lambda}_1} + \sigma^2_{\widehat{\lambda}_2}} = \sqrt{6.25 + 3.5} = 3.1$.
$\lambda_1 - \lambda_2 = 5.5 \pm 3.1$

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

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

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

(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.

21. Let X be the number of flaws in a one-square-meter sheet of aluminum. Let λ be the mean number of flaws per square meter. Then $X \sim \text{Poisson}(\lambda)$.

Let p be the probability that a one-square-meter sheet of aluminum has exactly one flaw.

Then $p = P(X = 1) = e^{-\lambda}\dfrac{\lambda^1}{1!} = \lambda e^{-\lambda}$. From Example 4.27, $\widehat{\lambda} = 2$, and $\sigma_{\widehat{\lambda}}{}^2 = 0.02$.

Therefore p is estimated with $\widehat{p} = \widehat{\lambda} e^{-\widehat{\lambda}} = 2e^{-2} = 0.271$.

$$\frac{d\widehat{p}}{d\widehat{\lambda}} = e^{-\widehat{\lambda}} - \widehat{\lambda}e^{-\widehat{\lambda}} = e^{-2} - 2e^{-2} = -0.135335$$

$$\sigma_{\widehat{p}} = \left|\frac{d\widehat{p}}{d\widehat{\lambda}}\right| \sigma_{\widehat{\lambda}} = 0.135335\sqrt{0.02} = 0.019$$

$p = 0.271 \pm 0.019$

Section 4.4

1. Let X be the number of units with broken fans among the seven chosen. Then $X \sim \text{H}(20, 8, 7)$.

$$P(X = 3) = \frac{\binom{8}{3}\binom{20-8}{7-3}}{\binom{20}{7}} = 0.3576$$

3. Let X be the number of the day on which the computer crashes. Then $X \sim \text{Geom}(0.1)$.

$P(X = 12) = (0.1)(0.9)^{11} = 0.0314$

5. (a) $Y \sim \text{NB}(3, 0.4)$. $P(Y = 7) = \binom{7-1}{3-1}(0.4)^3(1-0.4)^{7-3} = 0.1244$

(b) $\mu_Y = 3/0.4 = 7.5$

(c) $\sigma_Y^2 = 3(1-0.4)/(0.4^2) = 11.25$

7. (iv). $P(X = x) = p(1-p)^{x-1}$ for $x \geq 1$. Since $1 - p < 1$, this quantity is maximized when $x = 1$.

9. Let X be the number of hours that have elapsed when the fault is detected.
Then $X \sim \text{Geom}(0.8)$.

(a) $P(X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3) = (0.8)(1-0.8)^0 + (0.8)(1-0.8)^1 + (0.8)(1-0.8)^2 = 0.992$

(b) $P(X = 3|X > 2) = \dfrac{P(X = 3 \cap X > 2)}{P(X > 2)} = \dfrac{P(X = 3)}{P(X > 2)}$

Now $P(X = 3) = 0.8(1-0.8)^2 = 0.032$, and
$P(X > 2) = 1 - P(X \leq 2) = 1 - P(X = 1) - P(X = 2) = 1 - 0.8 - 0.8(1-0.8) = 0.04$.

Therefore $P(X = 3|X > 2) = \dfrac{0.032}{0.04} = 0.8$.

(c) $\mu_X = 1/0.8 = 1.25$

11. (a) $X \sim \text{H}(10, 3, 4)$. $P(X = 2) = \dfrac{\binom{3}{2}\binom{10-3}{4-2}}{\binom{10}{4}} = 0.3$

(b) $\mu_X = 4(3/10) = 1.2$

(c) $\sigma_X = \sqrt{4\left(\dfrac{3}{10}\right)\left(\dfrac{7}{10}\right)\left(\dfrac{10-4}{10-1}\right)} = 0.7483$

13. (a) $X \sim \text{H}(60, 5, 10)$, so

$$P(X > 1) = 1 - P(X = 0) - P(X = 1) = 1 - \frac{\binom{5}{0}\binom{60-5}{10-0}}{\binom{60}{10}} - \frac{\binom{5}{1}\binom{60-5}{10-1}}{\binom{60}{10}} = 0.1904$$

(b) $X \sim \text{H}(60, 10, 10)$, so

$$P(X > 1) = 1 - P(X = 0) - P(X = 1) = 1 - \frac{\binom{10}{0}\binom{60-10}{10-0}}{\binom{60}{10}} - \frac{\binom{10}{1}\binom{60-10}{10-1}}{\binom{60}{10}} = 0.5314$$

(c) $X \sim \mathrm{H}(60,20,10)$, so

$$P(X>1) = 1 - P(X=0) - P(X=1) = 1 - \frac{\binom{20}{0}\binom{60-20}{10-0}}{\binom{60}{10}} - \frac{\binom{20}{1}\binom{60-20}{10-1}}{\binom{60}{10}} = 0.9162$$

15. Let X_1 denote the number of orders for a small drink, let X_2 denote the number of orders for a medium drink, let X_3 denote the number of orders for a large drink.

(a) $(X_1, X_2, X_3) \sim \mathrm{MN}(20, 0.25, 0.35, 0.40)$

$P(X_1 = 5, X_2 = 7, X_3 = 8) = \dfrac{20!}{5!7!8!}(0.25)^5(0.35)^7(0.4)^8 = 0.0411$

(b) $X_3 \sim \mathrm{Bin}(20, 0.4)$. $P(X_3 > 10) = \sum_{x=11}^{20} \dfrac{20!}{x!(20-x)!}(0.4)^x(1-0.4)^{20-x} = 0.1275$

Alternatively, use Table A.1 to find $P(X_3 \le 10) = 0.872$. Then $P(X_3 > 10) = 1 - P(X \le 10) = 0.128$

17. $P(X=n) = p(1-p)^{n-1}$. $P(Y=1) = \binom{n}{1} p^1(1-p)^{n-1} = np(1-p)^{n-1}$. So $P(X=n) = (1/n)P(Y=1)$.

Section 4.5

1. (a) Using Table A.2: $1 - 0.1977 = 0.8023$

(b) Using Table A.2: $0.9032 - 0.6554 = 0.2478$

(c) Using Table A.2: $0.8159 - 0.3821 = 0.4338$

(d) Using Table A.2: $0.0668 + (1 - 0.3264) = 0.7404$

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

(b) Using Table A.2: $c = -2.00$

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

(d) Using Table A.2: $c = 0.83$

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

5. (a) $z = (19 - 16)/2 = 1.50$. The area to the right of $z = 1.50$ is 0.0668.

(b) The z-score of the 10th percentile is ≈ -1.28.
The 10th percentile is therefore $\approx 16 - 1.28(2) = 13.44$.

(c) $z = (14.5 - 16)/2 = -0.75$. The area to the right of $z = -0.75$ is 0.2266.
Therefore a lifetime of 14.5 is on the 23rd percentile, approximately.

(d) For 14.5, $z = (14.5 - 16)/2 = -0.75$. For 17, $z = (17 - 16)/2 = 0.5$.
The area between $z = -0.75$ and $z = 0.5$ is $0.6915 - 0.2266 = 0.4649$.

7. (a) $z = (700 - 480)/90 = 2.44$. The area to the right of $z = 2.44$ is 0.0073.

(b) The z-score of the 25th percentile is ≈ -0.67.
The 25th percentile is therefore $\approx 480 - 0.67(90) = 419.7$.

(c) $z = (600 - 480)/90 = 1.33$. The area to the left of $z = 1.33$ is 0.9082.
Therefore a score of 600 is on the 91st percentile, approximately.

(d) For 420, $z = (420 - 480)/90 = -0.67$. For 520, $z = (520 - 480)/90 = 0.44$.
The area between $z = -0.67$ and $z = 0.44$ is $0.6700 - 0.2514 = 0.4186$.

9. (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 $0.7734 - 0.4013 = 0.3721$.

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.

13\. Let X be the diameter of a hole, and let Y be the diameter of a piston. Then $X \sim N(15, 0.025^2)$, and $Y \sim N(14.88, 0.015^2)$. The clearance is given by $C = 0.5X - 0.5Y$.

(a) $\mu_C = \mu_{0.5X-0.5Y} = 0.5\mu_X - 0.5\mu_Y = 0.5(15) - 0.5(14.88) = 0.06$ cm

(b) $\sigma_C = \sqrt{0.5^2\sigma_X^2 + (-0.5)^2\sigma_Y^2} = \sqrt{0.5^2(0.025^2) + 0.5^2(0.015^2)} = 0.01458$ cm

(c) Since C is a linear combination of normal random variables, it is normally distributed, with μ_C and σ_C as given in parts (a) and (b).

The z-score of 0.05 is $(0.05 - 0.06)/0.01458 = -0.69$. The area to the left of $z = -0.69$ is 0.2451.

Therefore $P(C < 0.05) = 0.2451$.

(d) The z-score of the 25th percentile is ≈ -0.67.

The 25th percentile is therefore $\approx 0.06 - 0.67(0.01458) = 0.0502$ cm.

(e) The z-score of 0.05 is $z = (0.05 - 0.06)/0.01458 = -0.69$. The z-score of 0.09 is $z = (0.09 - 0.06)/0.01458 = 2.06$.

The area between $z = -0.69$ and $z = 2.06$ is $0.9803 - 0.2451 = 0.7352$.

Therefore $P(0.05 < C < 0.09) = 0.7352$.

(f) The probability is maximized when $\mu_C = 0.07$ cm, the midpoint between 0.05 and 0.09 cm. Now $\mu_Y = 14.88$, and μ_X must be adjusted so that $\mu_C = 0.5\mu_X - 0.5\mu_Y = 0.07$. Substituting $\mu_C = 0.07$ and $\mu_Y = 14.88$, and solving for μ_X yields $\mu_X = 15.02$ cm.

To find the probability that the clearance will be between 0.05 and 0.09:

The z-score of 0.05 is $z = (0.05 - 0.07)/0.01458 = -1.37$.

The z-score of 0.09 is $z = (0.09 - 0.07)/0.01458 = 1.37$.

The area between $z = -1.37$ and $z = 1.37$ is $0.9147 - 0.0853 = 0.8294$.

Therefore $P(0.05 < C < 0.09) = 0.8294$.

15. (a) The z-score of 12 is $(12 - 12.05)/0.03 = -1.67$. The area to the left of $z = -1.67$ is 0.0475. The proportion is 0.0475.

(b) Let μ be the required value of the mean. This value must be chosen so that the 1st percentile of the distribution is 12. The z-score of the 1st percentile is approximately $z = -2.33$. Therefore $-2.33 = (12 - \mu)/0.03$. Solving for μ yields $\mu = 12.07$ ounces.

(c) Let σ be the required standard deviation. The value of σ must be chosen so that the 1st percentile of the distribution is 12. The z-score of the 1st percentile is approximately $z = -2.33$. Therefore $-2.33 = (12 - 12.05)/\sigma$. Solving for σ yields $\sigma = 0.0215$ ounces.

17. (a) The proportion of strengths that are less than 65 is 0.10. Therefore 65 is the 10th percentile of strength. The z-score of the 10th percentile is approximately $z = -1.28$. Let σ be the required standard deviation. The value of σ must be chosen so that the z-score for 65 is -1.28. Therefore $-1.28 = (65 - 75)/\sigma$. Solving for σ yields $\sigma = 7.8125$ N/m^2.

(b) Let σ be the required standard deviation. The value of σ must be chosen so that the 1st percentile of the distribution is 65. The z-score of the 1st percentile is approximately $z = -2.33$. Therefore $-2.33 = (65 - 75)/\sigma$. Solving for σ yields $\sigma = 4.292$ N/m^2.

(c) Let μ be the required value of the mean. This value must be chosen so that the 1st percentile of the distribution is 65. The z-score of the 1st percentile is approximately $z = -2.33$. Therefore $-2.33 = (65 - \mu)/5$. Solving for μ yields $\mu = 76.65$ N/m^2.

19. Let $a = 1/\sigma$ and let $b = -\mu/\sigma$. Then $Z = aX + b$. Now $a\mu + b = (1/\sigma)\mu - \mu/\sigma = 0$, and $a^2\sigma^2 = (1/\sigma)^2\sigma^2 = 1$. Equation (4.25) now shows that $Z \sim N(0, 1)$.

21. (a) Let $D = R_2 - R_1$. The event that $R_2 > R_1$ is the event that $D > 0$.

$\mu_D = \mu_{R_2} - \mu_{R_1} = 120 - 100 = 20$.

Since R_1 and R_2 are independent, $\sigma_D = \sqrt{\sigma^2_{R_1} + \sigma^2_{R_2}} = \sqrt{5^2 + 10^2} = 11.180$.

Since D is a linear combination of independent normal random variables, D is normally distributed.

The z-score of 0 is $(0 - 20)/11.180 = -1.79$. The area to the right of $z = -1.79$ is $1 - 0.0367 = 0.9633$.

Therefore $P(D > 0) = 0.9633$.

(b) Let $D = R_2 - R_1$. The event that R_2 exceeds R_1 by more than 30 is the event that $D > 30$.

$\mu_D = \mu_{R_2} - \mu_{R_1} = 120 - 100 = 20$.

Since R_1 and R_2 are independent, $\sigma_D = \sqrt{\sigma^2_{R_1} + \sigma^2_{R_2}} = \sqrt{5^2 + 10^2} = 11.180$.

Since D is a linear combination of independent normal random variables, D is normally distributed.

The z-score of 30 is $(30 - 20)/11.180 = 0.89$.

The area to the right of $z = 0.89$ is $1 - 0.8133 = 0.1867$.

Therefore $P(D > 30) = 0.1867$.

23. (a) If $m = 0$, then $X = E$, so $X \sim N(0, 0.25)$.

$P(\text{error}) = P(X > 0.5)$. The z-score of 0.5 is $(0.5 - 0)/\sqrt{0.25} = 1.00$.

The area to the right of $z = 1.00$ is $1 - 0.8413 = 0.1587$.

Therefore $P(\text{error}) = 0.1587$.

(b) If $m = 0$, then $X = E$, so $X \sim N(0, \sigma^2)$.

$P(\text{error}) = P(X > 0.5) = 0.01$. The z-score of 0.5 is $(0.5 - 0)/\sigma$.

Since $P(X > 0.5) = 0.01$, the z-score of 0.5 is 2.33.

Therefore $(0.5 - 0)/\sigma = 2.33$. Solving for σ yields $\sigma = 0.2146$, and $\sigma^2 = 0.04605$.

25. (a) The sample mean is 114.8 J and the sample standard deviation is 5.006 J.

(b) The z-score of 100 is $(100 - 114.8)/5.006 = -2.96$. The area to the left of $z = -2.96$ is 0.0015. Therefore only 0.15% of bolts would have breaking torques less than 100 J, so the shipment would be accepted.

(c) The sample mean is 117.08 J; the sample standard deviation is 8.295 J. The z-score of 100 is $(100-117.08)/8.295 = -2.06$. The area to the left of $z = -2.06$ is 0.0197. Therefore about 2% of the bolts would have breaking torques less than 100 J, so the shipment would not be accepted.

(d) The bolts in part (c) are stronger. In fact, the weakest bolt in part (c) is stronger than the weakest bolt in part (a), the second-weakest bolt in part (c) is stronger than the second-weakest bolt in part (a), and so on.

(e) The method is certainly not valid for the bolts in part (c). This sample contains an outlier (140), so the normal distribution should not be used.

Section 4.6

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 $P(Y \le m) = P(\ln Y \le \ln m) = 0.5$.
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 $P(Y \le y_{90}) = P(\ln Y \le \ln y_{90}) = 0.90$.
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 $P(Y \le m) = P(\ln Y \le \ln m) = 0.5$.

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. (a) $\ln I \sim N(1, 0.2)$, $\ln R \sim N(4, 0.1)$, and I and R are independent. Therefore $\ln V \sim N(5, 0.3)$.
Since $\ln V$ is normal, V is lognormal, with $\mu_V = 5$ and $\sigma_V^2 = 0.3$.

(b) $P(V < 200) = P(\ln V < \ln 200) = P(\ln V < 5.298317)$.

Now $\ln V \sim N(5, 0.3)$, so the z-score of 5.298317 is $(5.298317 - 5)/\sqrt{0.3} = 0.54$.

The area to the left of $z = 0.54$ is 0.7054.

Therefore $P(V < 200) = 0.7054$.

(c) $P(150 < V < 300) = P(\ln 150 < \ln V < \ln 300) = P(5.010635 < \ln V < 5.703782)$.

Now $\ln V \sim N(5, 0.3)$, so the z-score of 5.010635 is $(5.010635 - 5)/\sqrt{0.3} = 0.02$, and the z-score of 5.703782 is $(5.703782 - 5)/\sqrt{0.3} = 1.28$.

The area between $z = 0.02$ and $z = 1.28$ is $0.8997 - 0.5080 = 0.3917$.

Therefore $P(150 < V < 300) = 0.3917$.

(d) The mean of V is $E(V) = e^{\mu+\sigma^2/2} = e^{5+0.3/2} = 172.43$.

(e) Since $\ln V \sim N(5, 0.3)$, the median of $\ln V$ is 5.

Therefore the median of V is $e^5 = 148.41$.

(f) The standard deviation of V is $\sqrt{e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2}} = \sqrt{e^{2(5)+2(0.3)} - e^{2(5)-0.3}} = 101.99$.

(g) Let v_{10} be the 10th percentile of V. Then $P(V \le v_{10}) = P(\ln V \le \ln v_{10}) = 0.10$.

The z-score of the 10th percentile is approximately $z = -1.28$.
Therefore the z-score of $\ln v_{10}$ must be -1.28, so $\ln v_{10}$ satisfies the equation $-1.28 = (\ln v_{10} - 5)/\sqrt{0.3}$.
$\ln v_{10} = 4.2989$, so $v_{10} = e^{4.2989} = 73.619$.

(h) Let v_{90} be the 90th percentile of V. Then $P(V \le v_{90}) = P(\ln V \le \ln v_{90}) = 0.90$.
The z-score of the 90th percentile is approximately $z = 1.28$.
Therefore the z-score of $\ln v_{90}$ must be 1.28, so $\ln v_{90}$ satisfies the equation $1.28 = (\ln v_{90} - 5)/\sqrt{0.3}$.
$\ln v_{90} = 5.7011$, so $v_{90} = e^{5.7011} = 299.19$.

7. Let X represent the withdrawal strength for a randomly chosen annularly threaded nail, and let Y represent the withdrawal strength for a randomly chosen helically threaded nail.

(a) $E(X) = e^{3.82+(0.219)^2/2} = 46.711$ N/mm

(b) $E(Y) = e^{3.47+(0.272)^2/2} = 33.348$ N/mm

(c) First find the probability for annularly threaded nails.
$P(X > 50) = P(\ln X > \ln 50) = P(\ln X > 3.9120)$.
The z-score of 3.9120 is $(3.9120 - 3.82)/0.219 = 0.42$.
The area to the right of $z = 0.42$ is $1 - 0.6628 = 0.3372$.
Therefore the probability for annularly threaded nails is 0.3372.

Now find the probability for helically threaded nails.
$P(Y > 50) = P(\ln Y > \ln 50) = P(\ln Y > 3.9120)$.
The z-score of 3.9120 is $(3.9120 - 3.47)/0.272 = 1.63$.
The area to the right of $z = 1.63$ is $1 - 0.9484 = 0.0516$.
Therefore the probability for helically threaded nails is 0.0516.

Annularly threaded nails have the greater probability. The probability is 0.3372 vs. 0.0516 for helically threaded nails.

(d) First find the median strength for annularly threaded nails.
Let m be the median of X. Then $P(X \le m) = P(\ln X \le \ln m) = 0.5$.
Since $\ln X \sim N(3.82, 0.219^2)$, $P(\ln Y < 3.82) = 0.5$.
Therefore $\ln m = 3.82$, so $m = e^{3.82}$.
Now $P(Y > e^{3.82}) = P(\ln Y > \ln e^{3.82}) = P(\ln Y > 3.82)$.
The z-score of 3.82 is $(3.82 - 3.47)/0.272 = 1.29$.
The area to the right of $z = 1.29$ is $1 - 0.9015 = 0.0985$. Therefore the probability is 0.0985.

(e) The log of the withdrawal strength for this nail is $\ln 20 = 2.996$.

For annularly threaded nails, the z-score of 2.996 is $(2.996 - 3.82)/0.219 = -3.76$.

The area to the left of $z = -3.76$ is less than 0.0001.

Therefore a strength of 20 is extremely small for an annularly threaded nail; less than one in ten thousand such nails have strengths this low.

For helically threaded nails, the z-score of 2.996 is $(2.996 - 3.47)/0.272 = -1.74$.

The area to the left of $z = -1.74$ is 0.0409.

Therefore about 4.09% of helically threaded nails have strengths of 20 or below.

We can be pretty sure that it was a helically threaded nail. Only about 0.01% of annularly threaded nails have strengths as small as 20, while about 4.09% of helically threaded nails do.

9. 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 $P(Y > 1.20) = 0.2090$.

11. $\ln X_1, \ldots, \ln X_n$ are independent normal random variables, so $\ln P = a_1 \ln X_1 + \cdots + a_n \ln X_n$ is a normal random variable. It follows that P is lognormal.

Section 4.7

1. (a) $\mu_T = 1/0.45 = 2.2222$

   (b) $\sigma_T^2 = 1/(0.45^2) = 4.9383$

   (c) $P(T > 3) = 1 - P(T \leq 3) = 1 - (1 - e^{-0.45(3)}) = 0.2592$

   (d) Let m be the median. Then $P(T \leq m) = 0.5$.
   $P(T \leq m) = 1 - e^{-0.45m} = 0.5$, so $e^{-0.45m} = 0.5$.
   Solving for m yields $m = 1.5403$.

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. (a) Let X be the number of pores with diameters less than 3 microns. The probability that a diameter is less than 3 microns is 0.5276.

Therefore $X \sim \text{Bin}(10, 0.5276)$.

$$\begin{aligned} P(X > 7) &= P(X = 8) + P(X = 9) + P(X = 10) \\ &= \frac{10!}{8!(10-8)!}(0.5276)^8(1-0.5276)^{10-8} + \frac{10!}{9!(10-9)!}(0.5276)^9(1-0.5276)^{10-9} \\ &= +\frac{10!}{10!(10-10)!}(0.5276)^{10}(1-0.5276)^{10-10} \\ &= 0.077 \end{aligned}$$

(b) Let X be the number of pores with diameters greater than 11 microns. The probability that a diameter is greater than 11 microns is 0.0639.

Therefore $X \sim \text{Bin}(10, 0.0639)$.

$$P(X = 1) = \frac{10!}{1!(10-1)!}(0.0639)^1(1-0.0639)^{10-1} = 0.353$$

7. 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.

9. Let T be the waiting time between accidents. Then $T \sim \text{Exp}(3)$.

(a) $\mu_T = 1/3$ year

(b) $\sigma_T = 1/3$ year

(c) $P(T > 1) = 1 - P(T \leq 1) = 1 - (1 - e^{-3(1)}) = 0.0498$

(d) $P(T < 1/12) = 1 - e^{-3(1/12)} = 0.2212$

(e) The probability is $P(T \leq 1.5 \,|\, T > 0.5) = 1 - P(T > 1.5 \,|\, T > 0.5)$

By the lack of memory property, $P(T > 1.5 \,|\, T > 0.5) = P(T > 1)$,

so $P(T \leq 1.5 \,|\, T > 0.5) = 1 - P(T > 1) = P(T \leq 1) = 1 - e^{-3(1)} = 0.9502$.

11. $X_1, \ldots, X_5$ are each exponentially distributed with $\lambda = 1/200 = 0.005$.

(a) $P(X_1 > 100) = 1 - P(X_1 \leq 100) = 1 - (1 - e^{-0.005(100)}) = e^{-0.5} = 0.6065$

(b)
$$\begin{aligned} P(X_1 > 100 \text{ and } X_2 > 100 \text{ and } X_3 > 100 \text{ and } X_4 > 100 \text{ and } X_5 > 100) &= \prod_{i=1}^{5} P(X_i > 100) \\ &= (e^{-0.5})^5 \\ &= e^{-2.5} \\ &= 0.0821 \end{aligned}$$

(c) The time of the first replacement will be greater than 100 hours if and only if each of the bulbs lasts longer than 100 hours.

(d) Using parts (b) and (c), $P(T \leq 100) = 1 - P(T > 100) = 1 - 0.0821 = 0.9179$.

(e) $P(T \leq t) = 1 - P(T > t) = 1 - \prod_{i=1}^{5} P(X_i > t) = 1 - (e^{-0.005t})^5 = 1 - e^{-0.025t}$

(f) Yes, $T \sim \text{Exp}(0.025)$

(g) $\mu_T = 1/0.025 = 40$ hours

(h) $T \sim \text{Exp}(n\lambda)$

Section 4.8

1. Let T be the waiting time.

(a) $\mu_T = (0 + 15)/2 = 7.5$ minutes.

(b) $\sigma_T = \sqrt{\dfrac{(15-0)^2}{12}} = 4.3301$ minutes

(c) $P(5 < T < 11) = \dfrac{11-5}{15-0} = 0.4$

(d) The probability that the waiting time is less than 5 minutes on any given morning is $(5-0)/(15-0) = 1/3$. Let X be the number of mornings on which the waiting time is less than 5 minutes. Then $X \sim \text{Bin}(10, 1/3)$.
$P(X = 4) = \dfrac{10!}{4!(10-4)!}(1/3)^4(1-1/3)^{10-4} = 0.2276$.

3. (a) $\mu_T = 4/0.5 = 8$

(b) $\sigma_T = \sqrt{4/0.5^2} = 4$

(c)
$$\begin{aligned}
P(T \le 1) &= 1 - \sum_{j=0}^{4-1} e^{-(0.5)(1)} \frac{[(0.5)(1)]^j}{j!} \\
&= 1 - e^{-(0.5)(1)} \frac{[(0.5)(1)]^0}{0!} - e^{-(0.5)(1)} \frac{[(0.5)(1)]^1}{1!} - e^{-(0.5)(1)} \frac{[(0.5)(1)]^2}{2!} - e^{-(0.5)(1)} \frac{[(0.5)(1)]^3}{3!} \\
&= 1 - 0.60653 - 0.30327 - 0.075816 - 0.012636 \\
&= 0.00175
\end{aligned}$$

(d)
$$\begin{aligned}
P(T \ge 4) &= 1 - P(T \le 3) \\
&= 1 - \left(1 - \sum_{j=0}^{4-1} e^{-(0.5)(3)} \frac{[(0.5)(3)]^j}{j!}\right) \\
&= e^{-(0.5)(3)} \frac{[(0.5)(3)]^0}{0!} + e^{-(0.5)(3)} \frac{[(0.5)(3)]^1}{1!} + e^{-(0.5)(3)} \frac{[(0.5)(3)]^2}{2!} + e^{-(0.5)(3)} \frac{[(0.5)(3)]^3}{3!} \\
&= 0.22313 + 0.33470 + 0.25102 + 0.12551 \\
&= 0.9344
\end{aligned}$$

5. $\mu_T = r/\lambda$ and $\sigma_T^2 = r/\lambda^2$. It follows that $\lambda = r/\mu_T$.
Since $\mu_T = 8$ and $r = 16$, $\lambda = 2$.
$\sigma^2 = 16/(2^2) = 4$.

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

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

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

(c) Let m be the median.
Then $P(T \le m) = 0.5$, 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) $P(T < 5000) = P(T \le 5000) = 1 - e^{-[(0.0001)(5000)]^{1.5}} = 0.2978$

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

11. (a) $P(X_1 > 5) = 1 - P(X_1 \le 5) = 1 - (1 - e^{-[(0.2)(5)]^2}) = e^{-1} = 0.3679$

(b) Since X_2 has the same distribution as X_1, $P(X_2 > 5) = P(X_1 > 5) = e^{-1}$. Since X_1 and X_2 are independent, $P(X_1 > 5 \text{ and } X_2 > 5) = P(X_1 > 5)P(X_2 > 5) = (e^{-1})^2 = e^{-2} = 0.1353$.

(c) The lifetime of the system will be greater than 5 hours if and only if the lifetimes of both components are greater than 5 hours.

(d) $P(T \le 5) = 1 - P(T > 5) = 1 - e^{-2} = 0.8647$

(e) $P(T \le t) = 1 - P(T > t) = 1 - P(X_1 > t)P(X_2 > t) = 1 - (e^{-[(0.2)(t)]^2})^2 = 1 - e^{-0.08t^2} = 1 - e^{-(\sqrt{0.08}t)^2}$

(f) Yes, $T \sim \text{Weibull}(2, \sqrt{0.08}) = \text{Weibull}(2, 0.2828)$.

13. $\mu_{X^2} = \int_a^b x^2 \frac{1}{b-a}\,dx = \frac{a^2+ab+b^2}{3}$

Now $\mu_X = \frac{a+b}{2}$, therefore $\sigma_X^2 = \mu_{X^2} - \mu_X^2 = \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$.

15. (a) The cumulative distribution function of U is

$$F_U(x) = \begin{cases} 0 & x \le 0 \\ x & 0 < x \le 1 \\ 1 & x > 1 \end{cases}$$

(b) $F_X(x) = P(X \le x) = P((b-a)U + a \le x) = P\left(U \le \frac{x-a}{b-a}\right) = F_U\left(\frac{x-a}{b-a}\right) = \begin{cases} 0 & x \le a \\ \frac{x-a}{b-a} & a < x \le b \\ 1 & x > b \end{cases}$

(c) The cdf of X is that of a random variable distributed $U(a,b)$.

Section 4.9

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 $V(\widehat{\mu}_1)$.

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

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

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

The mean squared error of $\widehat{\mu}_1$ is the sum of the variance and the square of the bias, so

$MSE(\widehat{\mu}_1) = \frac{1}{2} + 0^2 = \frac{1}{2}$.

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

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

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

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

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
$MSE(\widehat{\mu}_3) = \dfrac{\sigma^2}{8} + \left(-\dfrac{\mu}{2}\right)^2 = \dfrac{2\mu^2+1}{8}$.

(d) $\widehat{\mu}_3$ has smaller mean squared error than $\widehat{\mu}_1$ whenever $\dfrac{2\mu^2+1}{8} < \dfrac{1}{2}$.
Solving for μ yields $-1.2247 < \mu < 1.2247$.

(e) $\widehat{\mu}_3$ has smaller mean squared error than $\widehat{\mu}_2$ whenever $\dfrac{2\mu^2+1}{8} < \dfrac{5}{9}$.
Solving for μ yields $-1.3123 < \mu < 1.3123$.

5. The probability mass function of X is $f(x;\, p) = p(1-p)^{x-1}$.
The MLE is the value of p that maximizes $f(x;\, p)$, or equivalently, $\ln f(x;\, p)$.
$\dfrac{d}{dp}\ln f(x;\, p) = \dfrac{d}{dp}[\ln p + (x-1)\ln(1-p)] = \dfrac{1}{p} - \dfrac{x-1}{1-p} = 0$.
Solving for p yields $p = \dfrac{1}{x}$. The MLE is $\widehat{p} = \dfrac{1}{X}$.

7. (a) The probability mass function of X is $f(x;\, p) = \dfrac{n!}{x!(n-x)!}(p)^x(1-p)^{n-x}$.
The MLE of p is the value of p that maximizes $f(x;\, p)$, or equivalently, $\ln f(x;\, p)$.
$\dfrac{d}{dp}\ln f(x;\, p) = \dfrac{d}{dp}[\ln n! - \ln x! - \ln(n-x)! + x\ln p + (n-x)\ln(1-p)] = \dfrac{x}{p} - \dfrac{n-x}{1-p} = 0$.
Solving for p yields $p = \dfrac{x}{n}$. The MLE of p is $\widehat{p} = \dfrac{X}{n}$.
The MLE of $\dfrac{p}{1-p}$ is therefore $\dfrac{\widehat{p}}{1-\widehat{p}} = \dfrac{X}{n-X}$.

(b) The MLE of p is $\widehat{p} = \dfrac{1}{\overline{X}}$. The MLE of $\dfrac{p}{1-p}$ is therefore $\dfrac{\widehat{p}}{1-\widehat{p}} = \dfrac{1}{\overline{X}-1}$.

(c) The MLE of λ is $\widehat{\lambda} = \overline{X}$. The MLE of $e^{-\lambda}$ is therefore $e^{-\widehat{\lambda}} = e^{-\overline{X}}$.

9. The joint probability density function of $X_1, \ldots, X_n$ is

$$f(x_1, \ldots, x_n; \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-x_i^2/2\sigma^2} = (2\pi)^{-n/2}\sigma^{-n} e^{-\sum_{i=1}^{n} x_i^2/2\sigma^2}.$$

The MLE is the value of σ that maximizes $f(x_1, \ldots, x_n; \sigma)$, or equivalently, $\ln f(x_1, \ldots, x_n; \sigma)$.

$$\frac{d}{d\sigma} \ln f(x_1, \ldots, x_n; \sigma) = \frac{d}{d\sigma}\left[-(n/2)\ln 2\pi - n\ln\sigma - \sum_{i=1}^{n} \frac{x_i^2}{2\sigma^2}\right] = -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n} x_i^2}{\sigma^3} = 0.$$

Now we solve for σ:

$$-\frac{n}{\sigma} + \frac{\sum_{i=1}^{n} x_i^2}{\sigma^3} = 0$$

$$-n\sigma^2 + \sum_{i=1}^{n} x_i^2 = 0$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2}{n}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

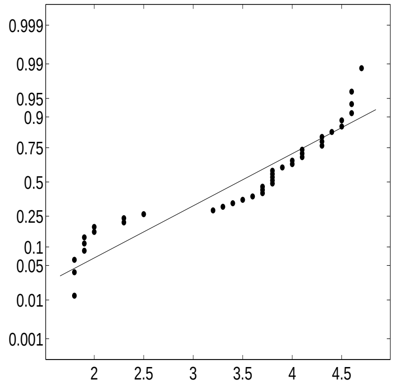
The MLE of σ is $\widehat{\sigma} = \sqrt{\dfrac{\sum_{i=1}^{n} X_i^2}{n}}$.

Section 4.10

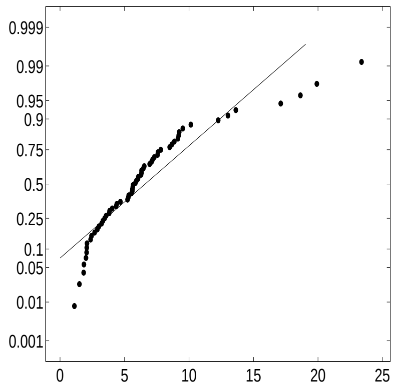
1. (a) No

(b) No

(c) Yes

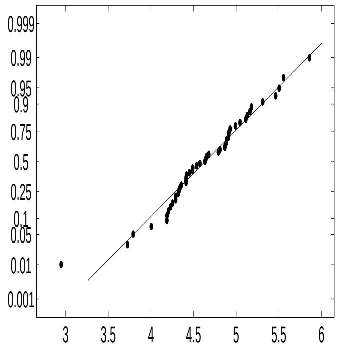
3. 

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.

8. 

The plot is a probability plot of the logs of the data. The points follow a line reasonably well, with the exception of the smallest value. The distribution is approximately normal except perhaps for the left tail, which may be somewhat longer than that of a normal distribution.

Section 4.11

1. (a) Let $X_1, \ldots, X_{144}$ 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. The mean number of red lights encountered per day is

   $\mu_X = 0(0.1) + 1(0.3) + 2(0.3) + 3(0.2) + 4(0.1) = 1.9$.

   The standard deviation is

   $\sigma_X = \sqrt{(0-1.9)^2(0.1) + (1-1.9)^2(0.3) + (2-1.9)^2(0.3) + (3-1.9)^2(0.2) + (4-1.9)^2(0.1)} = 1.1358$.

   Let $X_1, \ldots, X_{100}$ be the numbers of red lights encountered on 100 days. Then $\overline{X}$ is approximately normally distributed with mean $\mu_{\overline{X}} = 1.9$ and $\sigma_{\overline{X}} = 1.1358/\sqrt{100} = 0.11358$.

   The z-score of 2 is $z = (2 - 1.9)/0.11358 = 0.88$.

The area to the right of $z = 0.88$ is $1 - 0.8106 = 0.1894$. $P(\overline{X} > 2) = 0.1894$.

5. (a) Let $X_1, \ldots, X_{60}$ be the weights of the 60 bags.

Then $\overline{X}$ is approximately normally distributed with mean $\mu_{\overline{X}} = 15$ and $\sigma_{\overline{X}} = 5/\sqrt{60} = 0.6455$.

The z-score of 14 is $(14 - 15)/0.6455 = -1.55$.

The area to the left of $z = -1.55$ is 0.0606.

$P(\overline{X} < 14) = 0.0606$.

(b) Let x_{70} denote the 70th percentile

The z-score of the 70th percentile is approximately $z = 0.52$.

Therefore x_{70} satisfies the equation $0.52 = (x_{70} - 15)/0.6455$.

$x_{70} = 15.34$ kg.

(c) Let n be the necessary sample size.

Then $\overline{X}$ is approximately normally distributed with mean $\mu_{\overline{X}} = 15$ and $\sigma_{\overline{X}} = 5/\sqrt{n}$.

Since $P(\overline{X} < 14) = 0.01$, 14 is the 1st percentile of the distribution of $\overline{X}$.

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

Therefore $14 = 15 - 2.33(5/\sqrt{n})$. Solving for n yields $n \approx 136$.

7. (a) Let $X_1, \ldots, X_{50}$ be the times taken by each of the 50 customers. Let $S = X_1 + \cdots + X_{50}$.

Then S is approximately normally distributed with mean $\mu_S = 50(4) = 200$ and standard deviation $\sigma_S = 2\sqrt{50} = 14.142$.

The z-score of 180 is $(180 - 200)/14.142 = -1.41$.

The area to the left of $z = -1.41$ is 0.0793.

$P(S < 180) = 0.0793$.

(b) Let $X_1, \ldots, X_{50}$ be the times taken by each of the 50 customers. Let $S = X_1 + \cdots + X_{50}$.

Then S is approximately normally distributed with mean $\mu_S = 50(4) = 200$ and standard deviation $\sigma_S = 2\sqrt{50} = 14.142$.

Let x_{30} denote the 30th percentile of the total time taken by 50 customers.

The z-score of the 30th percentile is approximately $z = -0.52$.

Therefore $x_{30} = 200 - 0.52(14.142) = 192.65$.

9. Let n be the required number of measurements. Let $\overline{X}$ be the average of the n measurements.

Then the true value is $\mu_{\overline{X}}$, and the standard deviation is $\sigma_{\overline{X}} = \sigma/\sqrt{n} = 0.5/\sqrt{n}$.

Now $P(\mu_{\overline{X}} - 0.10 < \overline{X} < \mu_{\overline{X}} + 0.10) = 0.90$. In any normal population, 90% of the population is within 1.645 standard deviations of the mean.

Therefore $1.645\sigma_{\overline{X}} = 0.1$. Since $\sigma_{\overline{X}} = 0.5/\sqrt{n}$, $n = 67.65$. The smallest value of n is therefore $n = 68$.

11. (a) Let X represent the number of nondefective bearings in a shipment.

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

To find $P(X \geq 440)$, use the continuity correction and find the z-score of 439.5.

The z-score of 439.5 is $(439.5 - 450)/6.7802 = -1.57$.

The area to the right of $z = -1.57$ is $1 - 0.0582 = 0.9418$.

$P(X \geq 440) = 0.9418$.

(b) Let Y represent the number of shipments out of 300 that are acceptable.

From part (a) the probability that a shipment is acceptable is 0.9418, so $Y \sim \text{Bin}(300, 0.9418)$.

It follows that Y is approximately normal with mean $\mu_Y = 300(0.9418) = 282.54$ and standard deviation $\sigma_Y = \sqrt{300(0.9418)(0.0582)} = 4.0551$.

To find $P(Y > 285)$, use the continuity correction and find the z-score of 285.5.

The z-score of 285.5 is $(285.5 - 282.54)/4.0551 = 0.73$.

The area to the right of $z = 0.73$ is $1 - 0.7673 = 0.2327$.

$P(X > 285) = 0.2327$.

(c) Let p be the required proportion of defective bearings, and let X represent the number of defective bearings in a shipment.

Then $X \sim \text{Bin}(500, p)$, so X is approximately normal with mean $\mu_X = 500p$ and standard deviation $\sigma_X = \sqrt{500p(1-p)}$.

The probability that a shipment is acceptable is $P(X \geq 440) = 0.99$.

Using the continuity correction, 439.5 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 $-2.33 = (439.5 - 500p)/\sqrt{500p(1-p)}$.

This equation can be rewritten as $252,714.45p^2 - 442,214.45p + 193160.25 = 0$.

Solving for p yields $p = 0.909$. (0.841 is a spurious root.)

13. (a) Let X be the number of particles emitted by mass A and let Y be the number of particles emitted by mass B in a five-minute time period. Then $X \sim \text{Poisson}(100)$ and $Y \sim \text{Poisson}(125)$.

Now X is approximately normal with mean 100 and variance 100, and Y is approximately normal with mean 125 and variance 125.

The number of particles emitted by both masses together is $S = X + Y$.

Since X and Y are independent, and both approximately normal, S is approximately normal as well, with mean $\mu_S = 100 + 125 = 225$ and standard deviation $\sigma_S = \sqrt{100 + 125} = 15$.

The z-score of 200 is $(200 - 225)/15 = -1.67$. The area to the left of $z = -1.67$ is 0.0475.

$P(S < 200) = 0.0475$.

(b) Let X be the number of particles emitted by mass A and let Y be the number of particles emitted by mass B in a two-minute time period. Then $X \sim \text{Poisson}(40)$ and $Y \sim \text{Poisson}(50)$.

Now X is approximately normal with mean 40 and variance 40, and Y is approximately normal with mean 50 and variance 50.

The difference between the numbers of particles emitted by the two masses is $D = Y - X$.

Mass B emits more particles than mass A if $D > 0$.

Since X and Y are independent, and both approximately normal, D is approximately normal as well, with mean $\mu_D = 50 - 40 = 10$ and standard deviation $\sigma_D = \sqrt{50+40} = 9.486833$.

The z-score of 0 is $(0 - 10)/9.486833 = -1.05$. The area to the right of $z = -1.05$ is $1 - 0.1469 = 0.8531$.

$P(D > 0) = 0.8531$.

15. (a) Let X be the number of particles withdrawn in a 5 mL volume.

Then the mean of X is $50(5) = 250$, so $X \sim \text{Poisson}(250)$, and X is approximately normal with mean $\mu_X = 250$ and standard deviation $\sigma_X = \sqrt{250} = 15.8114$.

The z-score of 235 is $(235 - 250)/15.8114 = -0.95$, and the z-score of 265 is $(265 - 250)/15.8114 = 0.95$.

The area between $z = -0.95$ and $z = 0.95$ is $0.8289 - 0.1711 = 0.6578$.

The probability is 0.6578.

(b) Since the withdrawn sample contains 5 mL, the average number of particles per mL will be between 48 and 52 if the total number of particles is between $5(48) = 240$ and $5(52) = 260$.

From part (a), X is approximately normal with mean $\mu_X = 250$ and standard deviation $\sigma_X = \sqrt{250} = 15.8114$.

The z-score of 240 is $(240 - 250)/15.8114 = -0.63$, and the z-score of 260 is $(260 - 250)/15.8114 = 0.63$.

The area between $z = -0.63$ and $z = 0.63$ is $0.7357 - 0.2643 = 0.4714$.

The probability is 0.4714.

(c) Let X be the number of particles withdrawn in a 10 mL volume.

Then the mean of X is $50(10) = 500$, so $X \sim \text{Poisson}(500)$, and X is approximately normal with mean $\mu_X = 500$ and standard deviation $\sigma_X = \sqrt{500} = 22.3607$.

The average number of particles per mL will be between 48 and 52 if the total number of particles is between $10(48) = 480$ and $10(52) = 520$.

The z-score of 480 is $(480 - 500)/22.3607 = -0.89$, and the z-score of 520 is $(520 - 500)/22.3607 = 0.89$.

The area between $z = -0.89$ and $z = 0.89$ is $0.8133 - 0.1867 = 0.6266$.

The probability is 0.6266.

(d) Let v be the required volume. Let X be the number of particles withdrawn in a volume of v mL.

Then $X \sim \text{Poisson}(50v)$, so X is approximately normal with mean $\mu_X = 50v$ and standard deviation $\sigma_X = \sqrt{50v}$.

The average number of particles per mL will be between 48 and 52 if the total number of particles X is between $48v$ and $52v$.

Since $\mu_X = 50v$, $P(48v < X < 52v) = 0.95$ if the z-score of $48v$ is -1.96 and the z-score of $52v$ is 1.96.

The z-score of 1.96 therefore satisfies the equation $1.96 = (52v - 50v)/\sqrt{50v}$, or equivalently, $\sqrt{v} = 1.96\sqrt{50}/2$.

Solving for v yields $v = 48.02$ ml.

17. (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.

19. Let X be the number of rivets from vendor A that meet the specification, and let Y be the number of rivets from vendor B that meet the specification.

    Then $X \sim \text{Bin}(500, 0.7)$, and $Y \sim \text{Bin}(500, 0.8)$.

    It follows that X is approximately normal with mean $\mu_X = 500(0.7) = 350$ and variance $\sigma_X^2 = 500(0.7)(0.3) = 105$, and Y is approximately normal with mean $\mu_Y = 500(0.8) = 400$ and variance $\sigma_X^2 = 500(0.8)(0.2) = 80$.

    Let $T = X + Y$ be the total number of rivets that meet the specification.

    Then T is approximately normal with mean $\mu_T = \mu_X + \mu_Y = 350 + 400 = 750$, and standard deviation $\sigma_T = \sqrt{\sigma_X^2 + \sigma_Y^2} = \sqrt{105 + 80} = 13.601471$.

    To find $P(T > 775)$, use the continuity correction and find the z-score of 775.5.

    The z-score of 775.5 is $(775.5 - 750)/13.601471 = 1.87$.

    The area to the right of $z = 1.87$ is $1 - 0.9693 = 0.0307$.

    $P(T > 775) = 0.0307$.

Section 4.12

1. (a) $X \sim \text{Bin}(100, 0.03)$, $Y \sim \text{Bin}(100, 0.05)$

   (b) Answers will vary.

   (c) ≈ 0.72

   (d) ≈ 0.18

   (e) The distribution deviates somewhat from the normal.

3. (a) $\mu_A = 6$ exactly (simulation results will be approximate), $\sigma_A^2 \approx 0.25$.

   (b) ≈ 0.16

   (c) The distribution is approximately normal.

5. (a) ≈ 0.25

   (b) ≈ 0.25

   (c) ≈ 0.61

7. (a–c) Answers will vary.

   (d) ≈ 0.025

9. (a) Answers will vary.

   (b) ≈ 2.7

   (c) ≈ 0.34

   (d) ≈ 1.6

   (e) System lifetime is not approximately normally distributed.

(f) Skewed to the right.

11. (a) Answers will vary.

(b) $\approx 10,090$

(c) ≈ 1250

(d) ≈ 0.58

(e) ≈ 0.095

(f) The distribution differs somewhat from normal.

13. (a) $\widehat{\lambda} = 0.25616$

(b–d) Answers will vary.

(e) Bias ≈ 0.037, $\sigma_{\widehat{\lambda}} \approx 0.12$

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 $X \sim \text{Bin}(10, 0.25)$.

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

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

(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 $P(30 \le Y \le 35)$, 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 $0.9922 - 0.8508 = 0.1414$.

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

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

(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 components in a sample of 250 that are defective.

Then $X \sim \text{Bin}(250, 0.07)$, so X is approximately normal with mean $\mu_X = 250(0.07) = 17.5$ and standard deviation $\sigma_X = \sqrt{250(0.07)(0.93)} = 4.0342$.

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

The z-score of 19.5 is $(19.5 - 17.5)/4.0342 = 0.50$.

The area to the left of $z = 0.50$ is 0.6915.

$P(X < 20) = 0.6915$.

(b) Let Y be the number of components in a sample of 10 that are defective. Then $X \sim \text{Bin}(10, 0.07)$.

$P(X \geq 1) = 1 - P(X = 0) = 1 - \dfrac{10!}{0!(10-0)!}(0.07)^0(1-0.07)^{10-0} = 0.5160$.

(c) Let p be the required probability, and let X represent the number of components in a sample of 250 that are defective.

Then $X \sim \text{Bin}(250, p)$, so X is approximately normal with mean $\mu_X = 250p$ and standard deviation $\sigma_X = \sqrt{250p(1-p)}$.

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

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 $-2.33 = (19.5 - 250p)/\sqrt{250p(1-p)}$.

This equation can be rewritten as $62,635.7225p^2 - 11,107.225p + 380.25 = 0$.

Solving for p yields $p = 0.04686$. (0.1271 is a spurious root.)

13. (a) Let λ be the true concentration. Then $\widehat{\lambda} = 56/2 = 28$.

The uncertainty is $\sigma_{\widehat{\lambda}} = \sqrt{\lambda/2}$. Substituting $\widehat{\lambda}$ for λ, $\sigma_{\widehat{\lambda}} = \sqrt{28/2} = 3.7$.

$\lambda = 28.0 \pm 3.7$.

(b) Let v be the required volume, in mL. Then $\sigma_{\widehat{\lambda}} = \sqrt{\lambda/v} = 1$.

Substituting $\widehat{\lambda} = 28$ for λ and solving for v yields $v = 28$ mL.

15. (a) Let X_1, X_2, X_3 be the three thicknesses. Let $S = X_1 + X_2 + X_3$ be the thickness of the stack. Then S is normally distributed with mean $\mu_S = 3(1.5) = 4.5$ and standard deviation $0.2\sqrt{3} = 0.34641$.

The z-score of 5 is $(5 - 4.5)/0.34641 = 1.44$.

The area to the right of $z = 1.44$ is $1 - 0.9251 = 0.0749$.

$P(S > 5) = 0.0749$.

(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 $0.84 = (x_{80} - 4.5)/0.34641$.

Solving for x_{80} yields $x_{80} = 4.7910$ cm.

(c) Let n be the required number. Then $\mu_S = 1.5n$ and $\sigma_S = 0.2\sqrt{n}$.

$P(S > 5) \geq 0.99$, so 5 is less than or equal to the 1st percentile of the distribution of S.

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

Therefore n satisfies the inequality $-2.33 \geq (5 - 1.5n)/(0.2\sqrt{n})$.

The smallest value of n satisfying this inequality is the solution to the equation

$-2.33 = (5 - 1.5n)/(0.2\sqrt{n})$.

This equation can be rewritten $1.5n - 0.466\sqrt{n} - 5 = 0$.

Solving for $\sqrt{n}$ with the quadratic formula yields $\sqrt{n} = 1.9877$, so $n = 3.95$. The smallest integer value of n is therefore $n = 4$.

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

19. (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.

21. (a) $P(X \leq 0) = F(0) = e^{-e^{-0}} = e^{-1}$

(b) $P(X > \ln 2) = 1 - P(X \leq \ln 2) = 1 - F(\ln 2) = 1 - e^{-e^{-\ln 2}} = 1 - e^{-1/2}$

(c) Let x_m be the median of X. Then $F(x_m) = e^{-e^{-x_m}} = 0.5$.

Solving for x_m yields $e^{-x_m} = -\ln 0.5 = \ln 2$, so $x_m = -\ln(\ln 2) = 0.3665$.

23. (a) $f_X(x) = F'(x) = \dfrac{e^{-(x-\alpha)/\beta}}{\beta[1+e^{-(x-\alpha)/\beta}]^2}$

(b) $f_X(\alpha - x) = f_X(\alpha + x) = \dfrac{e^{x/\beta}}{\beta[1+e^{x/\beta}]^2}$

(c) Since $f_X(x)$ is symmetric around α, its center of mass is at $x = \alpha$.

25. (a) $P(X > s) = P(\text{First } s \text{ trials are failures}) = (1-p)^s$

(b)
$$\begin{aligned} P(X > s+t \mid X > s) &= P(X > s+t \text{ and } X > s)/P(X > s) \\ &= P(X > s+t)/P(X > s) \\ &= (1-p)^{s+t}/(1-p)^s \\ &= (1-p)^t \\ &= P(X > t) \end{aligned}$$

Note that if $X > s+t$, it must be the case that $X > s$, which is the reason that $P(X > s+t \text{ and } X > s) = P(X > s+t)$.

(c) Let X be the number of tosses of the penny needed to obtain the first head.
Then $P(X > 5 \mid X > 3) = P(X > 2) = 1/4$.
The probability that the nickel comes up tails twice is also 1/4.

27. (a) $F_Y(y) = P(Y \le y) = P(7X \le y) = P(X \le y/7)$.
Since $X \sim \text{Exp}(\lambda)$, $P(X \le y/7) = 1 - e^{-\lambda y/7}$.

(b) $f_Y(y) = F'_Y(y) = (\lambda/7)e^{-\lambda y/7}$

29. (a) $\dfrac{P(X=x)}{P(X=x-1)} = \dfrac{e^{-\lambda}\lambda^x/x!}{e^{-\lambda}\lambda^{x-1}/(x-1)!} = \dfrac{e^{-\lambda}\lambda^x(x-1)!}{e^{-\lambda}\lambda^{x-1}x!} = \dfrac{\lambda}{x}$

(b) $P(X=x) \ge P(X=x-1)$ if and only if $\dfrac{\lambda}{x} \ge 1$ if and only if $x \le \lambda$.

Chapter 5

Section 5.1

1. (a) 1.96

   (b) 2.33

   (c) 2.57 or 2.58

   (d) 1.28

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) $\overline{X} = 50$, $s = 2$, $n = 100$, $z_{.025} = 1.96$.
   The confidence interval is $50 \pm 1.96(2/\sqrt{100})$, or $(49.608, 50.392)$.

   (b) $\overline{X} = 50$, $s = 2$, $n = 100$, $z_{.005} = 2.58$.
   The confidence interval is $50 \pm 2.58(2/\sqrt{100})$, or $(49.484, 50.516)$.

   (c) $\overline{X} = 50$, $s = 2$, $n = 100$, so the upper confidence bound 50.3 satisfies $50.3 = 50 + z_{\alpha/2}(2/\sqrt{100})$.
   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) $z_{.025} = 1.96$. $1.96(2/\sqrt{n}) = 0.3$, so $n = 171$.

   (e) $z_{.005} = 2.58$. $2.58(2/\sqrt{n}) = 0.3$, so $n = 296$.

7. (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 $n = 327$.

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) $\overline{X} = 11.9$, $s = 1.1$, $n = 140$, $z_{.025} = 1.96$.
The confidence interval is $11.9 \pm 1.96(1.1/\sqrt{140})$, or $(11.718, 12.082)$.

(b) $\overline{X} = 11.9$, $s = 1.1$, $n = 140$, $z_{.005} = 2.58$.
The confidence interval is $11.9 \pm 2.58(1.1/\sqrt{140})$, or $(11.660, 12.140)$.

(c) $\overline{X} = 11.9$, $s = 1.1$, $n = 140$, so the upper confidence bound 11.99 satisfies $11.99 = 11.9 + z_{\alpha/2}(1.1/\sqrt{140})$.
Solving for $z_{\alpha/2}$ yields $z_{\alpha/2} = 0.97$.
The area to the right of $z = 0.97$ is $1 - 0.8340 = 0.1660$, so $\alpha/2 = 0.1660$.
The level is $1 - \alpha = 1 - 2(0.1660) = 0.6680$, or 66.80%.

(d) $z_{.025} = 1.96$. $1.96(1.1/\sqrt{n}) = 0.1$, so $n = 465$.

(e) $z_{.005} = 2.58$. $2.58(1.1/\sqrt{n}) = 0.1$, so $n = 806$.

13. (a) $\overline{X} = 136.9$, $s = 22.6$, $n = 123$, $z_{.02} = 2.05$.
The lower confidence bound is $136.9 - 2.05(22.6/\sqrt{123}) = 132.72$.

(b) The lower confidence bound 134.3 satisfies $134.3 = 136.9 - z_\alpha(22.6/\sqrt{123})$.
Solving for z_α yields $z_\alpha = -1.28$.
The area to the right of $z = -1.28$ is $1 - \alpha = 0.8997$.
The level is 0.8997, or 89.97%.

15. (a) $\overline{X} = 348.2$, $s = 5.1$, $n = 67$, $z_{.01} = 2.33$.
The upper confidence bound is $348.2 + 2.33(5.1/\sqrt{67}) = 349.651$.

(b) The upper confidence bound 349.5 satisfies $349.5 = 348.2 + z_\alpha(5.1/\sqrt{67})$.
Solving for z_α yields $z_\alpha = 2.09$.
The area to the left of $z = 2.09$ is $1 - \alpha = 0.9817$.
The level is 0.9817, or 98.17%.

17. (a) $\overline{X} = 85$, $s = 2$, $n = 60$, $z_{.02} = 2.05$.
The lower confidence bound is $85 - 2.05(2/\sqrt{60}) = 84.471$.

(b) The lower confidence bound 11.7 satisfies $11.7 = 85 - z_\alpha(2/\sqrt{60})$.
Solving for z_α yields $z_\alpha = 1.55$.
The area to the left of $z = 1.55$ is $1 - \alpha = 0.9394$.
The level is 0.9394, or 93.94%.

19. 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.

21. 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).

23. (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.

25. The supervisor is underestimating the confidence. The statement that the mean cost is less than \$160 is a one-sided upper confidence bound with confidence level 97.5%.

Section 5.2

1. (a) $28/70 = 0.4$, or 40%.

(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 $0.1 = 1.96\sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$.
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 $n = 127$.

(f) The upper confidence bound 0.5 satisfies the equation $0.5 = 0.405405 + z_\alpha\sqrt{0.405405(1-0.405405)/(70+4)}$
Solving for z_α yields $z_\alpha = 1.66$. The area to the left of $z = 1.66$ is $1-\alpha = 0.9515$.
The level is 0.9515, or 95.15%.

3. (a) $X = 52$, $n = 70$, $\tilde{p} = (52+2)/(70+4) = 0.72973$, $z_{.025} = 1.96$.
The confidence interval is $0.72973 \pm 1.96\sqrt{0.72973(1-0.72973)/(70+4)}$, or (0.629, 0.831).

(b) $X = 52$, $n = 70$, $\tilde{p} = (52+2)/(70+4) = 0.72973$, $z_{.05} = 1.645$.
The confidence interval is $0.72973 \pm 1.645\sqrt{0.72973(1-0.72973)/(70+4)}$, or (0.645, 0.815).

(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 0.72973 and solving for n yields $n = 300$.

(d) Let n be the required sample size.
Then n satisfies the equation $0.05 = 1.645\sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$.
Replacing $\tilde{p}$ with 0.72973 and solving for n yields $n = 210$.

(e) Let X be the number of 90% confidence intervals that cover the true proportion.
Then $X \sim \text{Bin}(300, 0.90)$, so X is approximately normal with mean $\mu_X = 300(0.90) = 270$ and standard deviation $\sqrt{300(0.90)(0.10)} = 5.196152$.
To find $P(X > 280)$, use the continuity correction and find the z-score of 280.5.
The z-score of 280.5 is $(280.5 - 270)/5.196152 = 2.02$.
The area to the right of $z = 2.02$ is $1 - 0.9783 = 0.0217$.
$P(X > 280) = 0.0217$.

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

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) $X = 89$, $n = 710$, $\tilde{p} = (89+2)/(710+4) = 0.12745$, $z_{.05} = 1.645$.

The confidence interval is $0.12745 \pm 1.645\sqrt{0.12745(1-0.12745)/(710+4)}$, or (0.107, 0.148).

(b) $X = 89$, $n = 710$, $\tilde{p} = (89+2)/(710+4) = 0.12745$, $z_{.025} = 1.96$.

The confidence interval is $0.12745 \pm 1.96\sqrt{0.12745(1-0.12745)/(710+4)}$, or (0.103, 0.152).

(c) $X = 89$, $n = 710$, $\tilde{p} = (89+2)/(710+4) = 0.12745$, $z_{.005} = 2.58$.

The confidence interval is $0.12745 \pm 2.58\sqrt{0.12745(1-0.12745)/(710+4)}$, or (0.095, 0.160).

13. (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 is no preliminary estimate of $\tilde{p}$ available, 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 $n = 253$.

15. (a) $X = 293$, $n = 335$, $\tilde{p} = (293+2)/(335+4) = 0.87021$, $z_{.05} = 1.645$.
The confidence interval is $0.87021 \pm 1.645\sqrt{0.87021(1-0.87021)/(293+4)}$, or (0.840, 0.900).

(b) Let n be the required sample size. Then n satisfies the equation $0.025 = 1.645\sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$.
Replacing $\tilde{p}$ with 0.87021 and solving for n yields $n = 486$.

(c) Let n be the required sample size. Then n satisfies the equation $0.03 = 1.645\sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$.
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 5.3

1. (a) 1.796

(b) 2.447

(c) 63.657

(d) 2.048

3. (a) 95%

(b) 98%

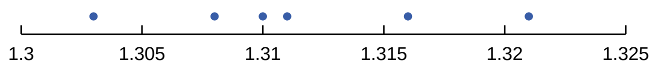
(c) 99%

(d) 80%

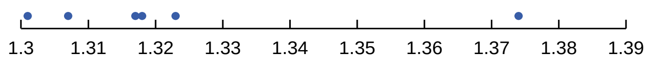
(e) 90%

5. $\overline{X} = 2.938$, $s = 0.47199$, $n = 5$, $t_{5-1,.025} = 2.938$.
The confidence interval is $2.938 \pm 2.938(0.47199/\sqrt{5})$, or (2.352, 3.524).

7. 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).

9. (a) 

(b) Yes, there are no outliers. $\overline{X} = 1.3115$, $s = 0.00628490$, $n = 6$, $t_{6-1,.005} = 4.032$.
The confidence interval is $1.3115 \pm 4.032(0.00628490/\sqrt{6})$, or (1.3012, 1.3218).

(c) 

(d) No, the data set contains an outlier.

11. $\overline{X} = 2.03$, $s = 0.090$, $n = 6$, $t_{6-1,.05} = 2.015$.
The confidence interval is $2.03 \pm 2.015(0.090/\sqrt{6})$, or (1.956, 2.104).

13. $\overline{X} = 0.242$, $s = 0.031$, $n = 10$, $t_{10-1,.025} = 2.262$.
The confidence interval is $0.242 \pm 2.262(0.031/\sqrt{12})$, or (0.2198, 0.2642).

15. (a) SE Mean is StDev/$\sqrt{\text{N}}$, so 0.52640 = 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 5.4

1. $\overline{X} = 620$, $s_X = 20$, $n_X = 80$, $\overline{Y} = 750$, $s_Y = 30$, $n_Y = 95$, $z_{.025} = 1.96$.
The confidence interval is $750 - 620 \pm 1.96\sqrt{20^2/80 + 30^2/95}$, or (122.54, 137.46).

3. $\overline{X} = 30.4$, $s_X = 0.6$, $n_X = 1559$, $\overline{Y} = 31.1$, $s_Y = 0.2$, $n_Y = 1924$, $z_{.005} = 2.58$.
The confidence interval is $31.1 - 30.4 \pm 2.58\sqrt{0.6^2/1559 + 0.2^2/1924}$, or (0.6591, 0.7409).

5. $\overline{X} = 3.64$, $s_X = 0.23$, $n_X = 40$, $\overline{Y} = 3.40$, $s_Y = 0.28$, $n_Y = 43$, $z_{.025} = 1.96$.
The confidence interval is $3.64 - 3.40 \pm 1.96\sqrt{0.23^2/40 + 0.28^2/43}$, or (0.1301, 0.3499).

7. $\overline{X} = 105$, $s_X = 20$, $n_X = 49$, $\overline{Y} = 117$, $s_Y = 15$, $n_Y = 35$, $z_{.01} = 2.33$.
The confidence interval is $117 - 105 \pm 2.33\sqrt{20^2/49 + 15^2/35}$, or (3.100, 20.900).

9. $\overline{X} = 242$, $s_X = 20$, $n_X = 47$, $\overline{Y} = 220$, $s_Y = 31$, $n_Y = 42$, $z_{.025} = 1.96$.
The confidence interval is $242 - 220 \pm 1.96\sqrt{20^2/47 + 31^2/42}$, or (11.018, 32.982).

11. (a) $\overline{X} = 91.1$, $s_X = 6.23$, $n_X = 50$, $\overline{Y} = 90.7$, $s_Y = 4.34$, $n_Y = 40$, $z_{.025} = 1.96$.
The confidence interval is $91.1 - 90.7 \pm 1.96\sqrt{6.23^2/50 + 4.34^2/40}$, or $(-1.789, 2.589)$.

(b) No. Since 0 is in the confidence interval, it may be regarded as being a plausible value for the mean difference in hardness.

13. It is not possible. The amounts of time spent in bed and spent asleep in bed are not independent.

Section 5.5

1. $X = 20$, $n_X = 100$, $\tilde{p}_X = (20+1)/(100+2) = 0.205882$,
$Y = 10$, $n_Y = 150$, $\tilde{p}_Y = (10+1)/(150+2) = 0.072368$, $z_{.05} = 1.645$.
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).

3. (a) $X = 841$, $n_X = 5320$, $\tilde{p}_X = (841+1)/(5320+2) = 0.158211$,
$Y = 134$, $n_Y = 1120$, $\tilde{p}_Y = (134+1)/(1120+2) = 0.120321$, $z_{.01} = 2.33$.
The confidence interval is $0.158211 - 0.120321 \pm 2.33\sqrt{\dfrac{0.158211(1-0.158211)}{5320+2} + \dfrac{0.120321(1-0.120321)}{1120+2}}$,
or $(0.0124, 0.0633)$.

(b) The standard deviation of the difference between the proportions is $\sqrt{p_X(1-p_X)/(n_X+2) + p_Y(1-p_Y)/(n_Y+2)}$.
Estimate $p_X \approx \tilde{p}_X = 0.120321$ and $p_Y \approx \tilde{p}_Y = 0.158211$.

If 1000 additional patients are treated with bare metal stents, the standard deviation of the difference between the proportions is then $\sqrt{0.120321(1-0.120321)/(1120+2) + 0.158211(1-0.158211)/(6320+2)} = 0.01074$.

The width of the 98% confidence interval will be $\pm 2.33(0.01074) = 0.0250$.

If 500 additional patients are treated with drug-coated stents, the standard deviation of the difference between the proportions is then $\sqrt{0.120321(1-0.120321)/(1620+2) + 0.158211(1-0.158211)/(5320+2)} = 0.009502$.

The width of the 98% confidence interval will be $\pm 2.33(0.009502) = 0.0221$.

If 500 additional patients are treated with bare metal stents and 250 additional patients are treated with drug-coated stents, the standard deviation of the difference between the proportions is then
$\sqrt{0.120321(1-0.120321)/(1370+2) + 0.158211(1-0.158211)/(5820+2)} = 0.01000$.

The width of the 98% confidence interval will be $\pm 2.33(0.01000) = 0.0233$.

Therefore the confidence interval would be most precise if 500 new patients are treated with drug-coated stents.

5. $X = 8$, $n_X = 12$, $\tilde{p}_X = (8+1)/(12+2) = 0.642857$,
$Y = 5$, $n_Y = 15$, $\tilde{p}_Y = (5+1)/(15+2) = 0.352941$, $z_{.025} = 1.96$.
The confidence interval is $0.642857 - 0.352941 \pm 1.96\sqrt{\dfrac{0.642857(1-0.642857)}{12+2} + \dfrac{0.352941(1-0.352941)}{15+2}}$,
or $(-0.04862, 0.6285)$.

7. No. The sample proportions come from the same sample rather than from two independent samples.

9. $X = 30$, $n_X = 164$, $\tilde{p}_X = (30+1)/(164+2) = 0.186747$,
$Y = 19$, $n_Y = 185$, $\tilde{p}_Y = (19+1)/(185+2) = 0.106952$, $z_{.005} = 2.58$.
The confidence interval is $0.186747 - 0.106952 \pm 2.58\sqrt{\dfrac{0.186747(1-0.186747)}{164+2} + \dfrac{0.106952(1-0.106952)}{185+2}}$,
or $(-0.0176, 0.1772)$.

11. No, these are not simple random samples.

Section 5.6

1. $\overline{X} = 1.5$, $s_X = 0.25$, $n_X = 7$, $\overline{Y} = 1.0$, $s_Y = 0.15$, $n_Y = 5$.
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 $(0.1234, 0.8766)$.

3. $\overline{X} = 59.6$, $s_X = 5.295701$, $n_X = 10$, $\overline{Y} = 50.9$, $s_Y = 5.321863$, $n_Y = 10$.
The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{5.295701^2}{10} + \dfrac{5.321863^2}{10}\right]^2}{\dfrac{(5.295701^2/10)^2}{10-1} + \dfrac{(5.321863^2/10)^2}{10-1}} = 17, \text{ rounded down to the nearest integer.}$$

$t_{17,.005} = 2.898$, so the confidence interval is $59.6 - 50.9 \pm 2.898\sqrt{\dfrac{5.295701^2}{10} + \dfrac{5.321863^2}{10}}$,
or $(1.8197, 15.5803)$.

5. $\overline{X} = 73.1$, $s_X = 9.1$, $n_X = 10$, $\overline{Y} = 53.9$, $s_Y = 10.7$, $n_Y = 10$.
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).

7. $\overline{X} = 33.8$, $s_X = 0.5$, $n_X = 4$, $\overline{Y} = 10.7$, $s_Y = 3.3$, $n_Y = 8$.

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).

9. $\overline{X} = 4.8$, $s_X = 1.9$, $n_X = 24$, $\overline{Y} = 2.8$, $s_Y = 1.0$, $n_Y = 24$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{1.9^2}{24} + \dfrac{1.0^2}{24}\right]^2}{\dfrac{(1.9^2/24)^2}{24-1} + \dfrac{(1.0^2/24)^2}{24-1}} = 34\text{, rounded down to the nearest integer.}$$

$t_{34,.025} = 2.032$, so the confidence interval is $4.8 - 2.8 \pm 2.032\sqrt{\dfrac{1.9^2}{24} + \dfrac{1.0^2}{24}}$, or (1.1093, 2.8907).

11. $\overline{X} = 36.893077$, $s_X = 3.221054$, $n_X = 13$, $\overline{Y} = 33.000000$, $s_Y = 2.544882$, $n_Y = 9$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{3.221054^2}{13} + \dfrac{2.544882^2}{9}\right]^2}{\dfrac{(3.221054^2/13)^2}{13-1} + \dfrac{(2.544882^2/9)^2}{9-1}} = 19\text{, rounded down to the nearest integer.}$$

$t_{19,.01} = 2.539$, so the confidence interval is $36.893077 - 33.000000 \pm 2.539\sqrt{\dfrac{3.221054^2}{13} + \dfrac{2.544882^2}{9}}$, or (0.7646, 7.0216).

13. $\overline{X} = 229.54286$, $s_X = 14.16885$, $n_X = 7$, $\overline{Y} = 143.95556$, $s_Y = 59.75699$, $n_Y = 9$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{14.16885^2}{7} + \dfrac{59.75699^2}{9}\right]^2}{\dfrac{(14.16885^2/7)^2}{7-1} + \dfrac{(59.75699^2/9)^2}{9-1}} = 9\text{, rounded down to the nearest integer.}$$

$t_{9,.025} = 2.262$, so the confidence interval is $229.54286 - 143.95556 \pm 2.262\sqrt{\dfrac{14.16885^2}{7} + \dfrac{59.75699^2}{9}}$, or (38.931, 132.244).

15. $\overline{X} = 4500.8$, $s_X = 271.6$, $n_X = 5$, $\overline{Y} = 1299.8$, $s_Y = 329.8$, $n_Y = 5$.
The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{271.6^2}{5} + \dfrac{329.8^2}{5}\right]^2}{\dfrac{(271.6^2/5)^2}{5-1} + \dfrac{(329.8^2/5)^2}{5-1}} = 7\text{, rounded down to the nearest integer.}$$

$t_{7,.01} = 2.998$, so the confidence interval is $4500.8 - 1299.8 \pm 2.998\sqrt{\dfrac{271.6^2}{5} + \dfrac{329.8^2}{5}}$, or $(2628.179, 3773.821)$.

Section 5.7

1. $\overline{D} = 6.736667$, $s_D = 6.045556$, $n = 9$, $t_{9-1,.025} = 2.306$.
The confidence interval is $6.736667 \pm 2.306(6.045556/\sqrt{9})$, or (2.090, 11.384).

3. The differences are: 9, 7, 5, 9, 10, 9, 2, 8, 8, 10.
$\overline{D} = 7.7$, $s_D = 2.496664$, $n = 10$, $t_{10-1,.01} = 2.821$.
The confidence interval is $7.7 \pm 2.821(2.496664/\sqrt{10})$, or (5.473, 9.927).

5. The differences are: 25, 32, 39, 23, 38.
$\overline{D} = 31.4$, $s_D = 7.300685$, $n = 5$, $t_{5-1,.05} = 2.132$.
The confidence interval is $31.4 \pm 2.132(7.300685/\sqrt{5})$, or (24.4391, 38.3609).

7. The differences are: $8.4, 8.6, 10.5, 9.6, 10.7, 10.8, 10.7, 11.3, 10.7$.
$\overline{D} = 10.144444$, $s_D = 1.033333$, $n = 9$, $t_{9-1,.025} = 2.306$.
The confidence interval is $10.144444 \pm 2.306(1.033333/\sqrt{9})$, or (9.350, 10.939).

9. (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%.

Section 5.8

1. (a) $\chi^2_{12,.025} = 23.337$

   (b) $\chi^2_{12,.975} = 4.404$

   (c) $\chi^2_{5,.005} = 16.750$

   (d) $\chi^2_{5,.995} = 0.412$

   (e) $\chi^2_{22,.1} = 30.813$

   (f) $\chi^2_{22,.9} = 14.041$

3. $s = 7.5$, $n = 8$, $\alpha = 0.01$. $\chi^2_{7,.005} = 20.278$, $\chi^2_{7,.995} = 0.989$.

   A 99% confidence interval for σ is $\left(\sqrt{\frac{7(7.5^2)}{20.278}}, \sqrt{\frac{7(7.5^2)}{0.989}}\right)$, or (4.41, 19.95).

5. $s = 8$, $n = 18$, $\alpha = 0.05$. $\chi^2_{17,.025} = 30.191$, $\chi^2_{17,.975} = 7.564$.

   A 95% confidence interval for σ^2 is $\left(\frac{17(8^2)}{30.191}, \frac{17(8^2)}{7.564}\right)$, or (36.04, 143.84).

7. (a) $s = 0.0614$

   (b) $s = 0.0614$, $n = 12$, $\alpha = 0.08$. $\chi^2_{11,.01} = 24.725$, $\chi^2_{11,.99} = 3.053$.

   A 95% confidence interval for σ is $\left(\sqrt{\frac{11(0.0641^2)}{24.725}}, \sqrt{\frac{11(0.0641^2)}{3.053}}\right)$, or (0.041, 0.117).

9. $s = 0.08841$, $n = 12$, $\alpha = 0.05$. $\chi^2_{11,.025} = 21.920$, $\chi^2_{11,.975} = 3.816$.

   A 95% confidence interval for σ^2 is $\left(\frac{11(0.08841^2)}{21.920}, \frac{11(0.08841^2)}{3.816}\right)$, or (0.00392, 0.0225).

11. $s = 40$ and the number of degrees of freedom is $k = 100$.
    We therefore approximate the χ^2_{100} distribution with the $N(100, 200)$ distribution.
    The z score for the 2.5 percentile of the normal distribution is $z = -1.96$.
    We therefore estimate $\chi^2_{100,.025} = 100 - 1.96\sqrt{(200)} = 72.281$.
    The z score for the 97.5 percentile of the normal distribution is $z = 1.96$.
    We therefore estimate $\chi^2_{100,.975} = 100 + 1.96\sqrt{(200)} = 127.719$.
    A 95% confidence interval for σ is $\left(\sqrt{\frac{100(40^2)}{127.719}}, \sqrt{\frac{100(40^2)}{72.281}}\right)$, or (35.394, 47.049).

Section 5.9

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 $(79.808, 93.312)$.

Section 5.10

1. (a) $X^* \sim N(8.5, 0.2^2)$, $Y^* \sim N(21.2, 0.3^2)$

   (b) Answers will vary.

   (c) Answers will vary.

   (d) Yes, P is approximately normally distributed.

   (e) Answers will vary.

3. (a) Yes, A is approximately normally distributed.

(b) Answers will vary.

(c) Answers will vary.

5. (a) $N(0.27, 0.40^2/349)$ and $N(1.62, 1.70^2/143)$. Since the values 0.27 and 1.62 are sample means, their variances are equal to the population variances divided by the sample sizes.

(b) No, R is not approximately normally distributed.

(c) Answers will vary.

(d) It is not appropriate, since R is not approximately normally distributed.

7. (a,b,c) Answers will vary.

9. (a) Coverage probability for traditional interval is ≈ 0.89, mean length is ≈ 0.585.
Coverage probability for the Agresti-Coull interval is ≈ 0.98, mean length is ≈ 0.51.

(b) Coverage probability for traditional interval is ≈ 0.95, mean length is ≈ 0.46.
Coverage probability for the Agresti-Coull interval is ≈ 0.95, mean length is ≈ 0.42.

(c) Coverage probability for traditional interval is ≈ 0.92, mean length is ≈ 0.305.
Coverage probability for the Agresti-Coull interval is ≈ 0.96, mean length is ≈ 0.29.

(d) The traditional method has coverage probability close to 0.95 for $n = 17$, but less than 0.95 for both $n = 10$ and $n = 40$.

(e) The Agresti-Coull interval has greater coverage probability for sample sizes 10 and 40, and nearly equal for 17.

(f) The Agresti-Coull method does.

Supplementary Exercises for Chapter 5

1. 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).

3. $X = 1919$, $n_X = 1985$, $\tilde{p}_X = (1919+1)/(1985+2) = 0.966281$,
$Y = 4561$, $n_Y = 4988$, $\tilde{p}_Y = (4561+1)/(4988+2) = 0.914228$, $z_{.005} = 2.58$.
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).

5. $\overline{X} = 6.1$, $s_X = 0.7$, $n_X = 125$, $\overline{Y} = 5.8$, $s_Y = 1.0$, $n_Y = 75$, $z_{.05} = 1.645$.
The confidence interval is $6.1 - 5.8 \pm 1.645\sqrt{0.7^2/125 + 1.0^2/75}$, or (0.084, 0.516).

7. (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) Let n be the required sample size.
Then n satisfies the equation $0.03 = 1.96\sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$.
Replacing $\tilde{p}$ with 0.16484 and solving for n yields $n = 584$.

9. 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.

11. $\overline{X} = 7.909091$, $s_X = 0.359039$, $n_X = 11$, $\overline{Y} = 8.00000$, $s_Y = 0.154919$, $n_Y = 6$.
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 $(-0.420, 0.238)$.

13. Let n be the required sample size.
The 99% confidence interval based on 64 observations has width ± 1.2.

Therefore $1.2 = 2.58\sigma/\sqrt{64}$, so $2.58\sigma = 9.6$.
Now n satisfies the equation $1.0 = 2.58\sigma/\sqrt{n} = 9.6/\sqrt{n}$. Solving for n yields $n = 93$.

15. (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.

17. (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).

19. (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).

21. (a) $\mu = 1.6$, $\sigma_\mu = 0.05$, $h = 15$, $\sigma_h = 1.0$, $\tau = 25$, $\sigma_\tau = 1.0$. $V = \tau h/\mu = 234.375$

$$\frac{\partial V}{\partial \mu} = -\tau h/\mu^2 = -146.484375, \quad \frac{\partial V}{\partial h} = \tau/\mu = 15.625, \quad \frac{\partial V}{\partial \tau} = h/\mu = 9.375$$

$$\sigma_V = \sqrt{\left(\frac{\partial V}{\partial \mu}\right)^2 \sigma_\mu^2 + \left(\frac{\partial V}{\partial h}\right)^2 \sigma_h^2 + \left(\frac{\partial V}{\partial \tau}\right)^2 \sigma_\tau^2} = 19.63862$$

(b) If the estimate of V is normally distributed, then a 95% confidence interval for V is $234.375 \pm 1.96(19.63862)$ or (195.883, 272.867).

(c) The confidence interval is valid. The estimate of V is approximately normal.

23. (a) $\overline{X} = 8.95$, $s = 0.5682$, $n6$, $t_{8-1,.01} = 2.998$.

The prediction interval is $8.95 \pm 2.998(0.5682\sqrt{1+1/8})$, or $(7.143, 10.757)$.

(b) $\overline{X} = 8.95$, $s = 0.5682$, $n6$, $k_{8,.05,.01} = 4.8907$.

The tolerance interval is $8.95 \pm 4.8907(0.5682)$, or $(6.171, 11.729)$.

25. It is not appropriate. The sample contains an outlier, which suggests that the population is not normal.

27. (a) Answers may vary somewhat from the 0.5 and 99.5 percentiles in Exercise 26 parts (c) and (d).

(b) Answers may vary somewhat from the answer to Exercise 26 part (c).

(c) Answers may vary somewhat from the answer to Exercise 26 part (d).

Chapter 6

Section 6.1

1. (a) $\overline{X} = 783$, $s = 120$, $n = 73$. The null and alternate hypotheses are $H_0 : \mu \leq 750$ versus $H_1 : \mu > 750$.

   $z = (783 - 750)/(120/\sqrt{73}) = 2.35$. Since the alternate hypothesis is of the form $\mu > \mu_0$, the P-value is the area to the right of $z = 2.35$.

   Thus $P = 0.0094$.

   (b) The P-value is 0.0094, so if H_0 is true then the sample is in the most extreme 0.94% of its distribution.

3. (a) $\overline{X} = 231.7$, $s = 2.19$, $n = 66$. The null and alternate hypotheses are $H_0 : \mu = 232$ versus $H_1 : \mu \neq 232$.

   $z = (231.7 - 232)/(2.19/\sqrt{66}) = -1.11$. 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 $z = 1.11$ and to the left of $z = -1.11$.

   Thus $P = 0.1335 + 0.1335 = 0.2670$.

   (b) The P-value is 0.2670, so if H_0 is true then the sample is in the most extreme 26.70% of its distribution.

5. (a) $\overline{X} = 4.5$, $s = 2.7$, $n = 80$. The null and alternate hypotheses are $H_0 : \mu \geq 5.4$ versus $H_1 : \mu < 5.4$.

   $z = (4.5 - 5.4)/(2.7/\sqrt{80}) = -2.98$. 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) If the mean number of sick days were 5.4, the probability of observing a sample mean less than or equal to the observed value of 4.5 would be 0.0014. Since 0.0014 is a small probability, we are convinced that the mean number of sick days is less than 5.4.

7. (a) $\overline{X} = 715$, $s = 24$, $n = 60$. The null and alternate hypotheses are $H_0 : \mu \geq 740$ versus $H_1 : \mu < 740$.

   $z = (715 - 740)/(24/\sqrt{60}) = -8.07$. 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) $\overline{X} = 8.24$, $s = 16.33$, $n = 126$. The null and alternate hypotheses are $H_0: \mu \geq 10$ versus $H_1: \mu < 10$.

   $z = (8.24 - 10)/(16.33/\sqrt{126}) = -1.21$. Since the alternate hypothesis is of the form $\mu < \mu_0$, the P-value is the area to the left of $z = -1.21$.

   Thus $P = 0.1131$.

   (b) If the mean profit margin were 10%, the probability of observing a sample mean less than or equal to the observed value of 8.24 would be 0.1131. Since 0.1131 is not a small probability, it is plausible that the mean is 10% or more.

11. (ii) 5. 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. $\overline{X} = 5.2$ and $\sigma_{\overline{X}} = \sigma/\sqrt{n} = 0.1$. The null and alternate hypotheses are $H_0: \mu = 5.0$ versus $H_1: \mu \neq 5.0$.

    $z = (5.2 - 5.0)/0.1 = 2.00$. 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 $z = 2.00$ and to the left of $z = -2.00$.

    Thus $P = 0.0228 + 0.0228 = 0.0456$.

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 $H_0: \mu \leq 3.5$ versus $H_1: \mu > 3.5$.

    $z = (4.07114 - 3.5)/0.2153 = 2.65$.

    (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. $P = 0.5$. 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 2%.

   (b) False. $P > 0.01$, so the result is not statistically significant at the 1% level.

   (c) True. The null hypothesis is rejected at any level greater than or equal to 2%.

   (d) False. $P > 0.01$, so the null hypothesis is not rejected at the 1% level.

7. (iii). Since the decrease is statistically significant, it is reasonable to conclude that the homicide rate did actually decrease. However, we cannot determine what caused the decrease.

9. (a) $H_0 : \mu \leq 8$. If H_0 is rejected, we can conclude that $\mu > 8$, and that the new battery should be used.

   (b) $H_0 : \mu \leq 60,000$. If H_0 is rejected, we can conclude that $\mu > 60,000$, and that the new material should be used.

   (c) $H_0 : \mu = 10$. If H_0 is rejected, we can conclude that $\mu \neq 10$, and that the flow meter should be recalibrated.

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) $H_0 : \mu = 1.2$. 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. $X = 35$, $n = 1000$, $\widehat{p} = 35/1000 = 0.035$.
   The null and alternate hypotheses are $H_0: p \geq 0.05$ versus $H_1: p < 0.05$.
   $z = (0.035 - 0.05)/\sqrt{0.05(1-0.05)/1000} = -2.18$.
   Since the alternate hypothesis is of the form $p < p_0$, the P-value is the area to the left of $z = -2.18$, so $P = 0.0146$. We can conclude that the goal has been met.

3. $X = 29$, $n = 50$, $\widehat{p} = 29/50 = 0.58$.
   The null and alternate hypotheses are $H_0: p \leq 0.50$ versus $H_1: p > 0.50$.
   $z = (0.58 - 0.50)/\sqrt{0.50(1-0.50)/50} = 1.13$.
   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 $H_0: p \leq 0.50$ versus $H_1: p > 0.50$.
   $z = (0.548 - 0.50)/\sqrt{0.50(1-0.50)/500} = 2.15$.
   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 $H_0: p \leq 0.70$ versus $H_1: p > 0.70$.
   $z = (0.733 - 0.70)/\sqrt{0.70(1-0.70)/150} = 0.89$.
   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. $X = 42$, $n = 300$, $\widehat{p} = 42/300 = 0.14$.
   The null and alternate hypotheses are $H_0: p = 0.12$ versus $H_1: p \neq 0.12$.
   $z = (0.14 - 0.12)/\sqrt{0.12(1-0.12)/300} = 1.07$.
   Since the alternate hypothesis is of the form $p = p_0$, the P-value is the sum of the areas to the right of $z = 1.07$ and to the left of $z = -1.07$, so $P = 0.2846$. We cannot conclude that the locus is not in Hardy-Weinberg equilibrium.

11. $X = 73$, $n = 100$, $\widehat{p} = 73/100 = 0.73$.
    The null and alternate hypotheses are $H_0: p \leq 0.60$ versus $H_1: p > 0.60$.
    $z = (0.73 - 0.60)/\sqrt{0.60(1-0.60)/100} = 2.65$.

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 $H_0: p \geq 0.7$ versus $H_1: \mu < 0.7$. $n = 500$. 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) $\overline{X} = 100.01$, $s = 0.0264575$, $n = 3$. There are $3 - 1 = 2$ degrees of freedom.

The null and alternate hypotheses are $H_0: \mu = 100$ versus $H_1: \mu \neq 100$.

$t = (100.01 - 100)/(0.0264575/\sqrt{3}) = 0.6547$.

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 $t = 0.6547$ and to the left of $t = -0.6547$.

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

We conclude that the scale may well be calibrated correctly.

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

3. (a) $H_0: \mu \leq 5$ versus $H_1: \mu > 5$

(b) $\overline{X} = 6.5$, $s = 1.9$, $n = 8$. There are $8 - 1 = 7$ degrees of freedom.

From part (a), the null and alternate hypotheses are $H_0: \mu \leq 5$ versus $H_1: \mu > 5$.

$t = (6.5 - 5)/(1.9/\sqrt{8}) = 2.233$.

Since the alternate hypothesis is of the form $\mu > \mu_0$, the P-value is the area to the right of $t = 2.233$.

From the t table, $0.025 < P < 0.050$. A computer package gives $P = 0.030$.

(c) Yes, the P-value is small, so we conclude that $\mu > 5$.

5. (a) $\overline{X} = 6.7$, $s = 3.9$, $n = 20$. There are $20 - 1 = 19$ degrees of freedom.

The null and alternate hypotheses are $H_0: \mu \geq 10$ versus $H_1: \mu < 10$.

$t = (6.7 - 10)/(3.9/\sqrt{20}) = -3.784$.

Since the alternate hypothesis is of the form $\mu < \mu_0$, the P-value is the area to the left of $t = -3.784$.

From the t table, $0.0005 < P < 0.001$. A computer package gives $P = 0.0006$.

We can conclude that the system meet this guideline.

(b) $\overline{X} = 6.7$, $s = 3.9$, $n = 20$. There are $20 - 1 = 19$ degrees of freedom.

The null and alternate hypotheses are $H_0: \mu \geq 7.5$ versus $H_1: \mu < 7.5$.

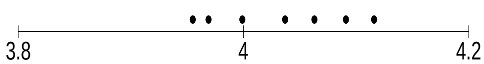
$t = (6.7 - 7.5)/(3.9/\sqrt{19}) = -0.917$.

Since the alternate hypothesis is of the form $\mu < \mu_0$, the P-value is the area to the left of $t = -0.917$.

From the t table, $0.10 < P < 0.25$. A computer package gives $P = 0.1852$.

We cannot conclude that the system meet this guideline.

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 $H_0: \mu = 4$ versus $H_1: \mu \neq 4$.

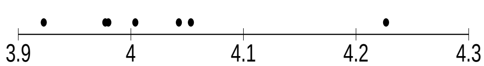
$t = (4.032857 - 4)/(0.061244/\sqrt{7}) = 1.419$.

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. $\overline{X} = 457.8$, $s = 317.7$, $n = 18$. There are $18 - 1 = 17$ degrees of freedom.

The null and alternate hypotheses are $H_0: \mu = 10$ versus $H_1: \mu \neq 0$.

$t = (457.8 - 0)/(317.7/\sqrt{18}) = 6.114$.

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 $t = 6.114$ and to the left of $t = -6.114$. From the t table, $P < 0.001$. A computer package gives $P = 1.1474 \times 10^{-4}$.

We cannot conclude that the mean reduction is greater than 10%.

11. $\overline{X} = 0.242$, $s = 0.031$, $n = 10$. There are $10 - 1 = 9$ degrees of freedom.

The null and alternate hypotheses are $H_0: \mu \geq 0.3$ versus $H_1: \mu < 0.3$.

$t = (0.242 - 0.3)/(0.031/\sqrt{10}) = -5.9165$.

Since the alternate hypothesis is of the form $\mu < \mu_0$, the P-value is the area to the left of $t = -5.9165$.

From the t table, $P < 0.0005$. A computer package gives $P = 0.000112$.

We can conclude that the concentration is less than 0.3 ppm.

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

(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. $\overline{X} = 8.5$, $s_X = 1.9$, $n_X = 58$, $\overline{Y} = 11.9$, $s_Y = 3.6$, $n_Y = 58$.

   The null and alternate hypotheses are $H_0: \mu_X - \mu_Y \geq 0$ versus $H_1: \mu_X - \mu_Y < 0$.

   $z = (8.5 - 11.9 - 0)/\sqrt{1.9^2/58 + 3.6^2/58} = -6.36$. Since the alternate hypothesis is of the form $\mu_X - \mu_Y < \Delta$, the P-value is the area to the left of $z = -6.36$.

   Thus $P \approx 0$.

   We can conclude that the mean hospital stay is shorter for patients receiving C4A-rich plasma.

3. $\overline{X} = 5.92$, $s_X = 0.15$, $n_X = 42$, $\overline{Y} = 6.05$, $s_Y = 0.16$, $n_Y = 37$.

   The null and alternate hypotheses are $H_0: \mu_X - \mu_Y = 0$ versus $H_1: \mu_X - \mu_Y \neq 0$.

   $z = (5.92 - 6.05 - 0)/\sqrt{0.15^2/42 + 0.16^2/37} = -3.71$. 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 $z = 3.71$ and to the left of $z = -3.71$.

   Thus $P = 0.0001 + 0.0001 = 0.0002$.

   We can conclude that the mean dielectric constant differs between the two types of asphalt.

5. $\overline{X} = 40$, $s_X = 12$, $n_X = 75$, $\overline{Y} = 42$, $s_Y = 15$, $n_Y = 100$.

   The null and alternate hypotheses are $H_0: \mu_X - \mu_Y > 0$ versus $H_1: \mu_X - \mu_Y \leq 0$.

   $z = (40 - 42 - 0)/\sqrt{12^2/75 + 15^2/100} = -0.98$. 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.

7. (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 $H_0: \mu_X - \mu_Y \leq 0$ versus $H_1: \mu_X - \mu_Y > 0$.

$z = (7.79 - 7.64 - 0)/\sqrt{1.06^2/80 + 1.31^2/80} = 0.80$. 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 $H_0: \mu_X - \mu_Y = 0$ versus $H_1: \mu_X - \mu_Y \neq 0$.

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.

9. (a) $\overline{X} = 4.7$, $s_X = 7.2$, $n_X = 77$, $\overline{Y} = 2.6$, $s_Y = 5.9$, $n_Y = 79$.

The null and alternate hypotheses are $H_0: \mu_X - \mu_Y \leq 0$ versus $H_1: \mu_X - \mu_Y > 0$.

$z = (4.7 - 2.6 - 0)/\sqrt{7.2^2/77 + 2.6^2/79} = 1.99$. Since the alternate hypothesis is of the form $\mu_X - \mu_Y > \Delta$, the P-value is the area to the right of $z = 1.99$.

Thus $P = 0.0233$.

We can conclude that the mean weight loss is greater for those on the low-carbohydrate diet.

(b) $\overline{X} = 4.7$, $s_X = 7.2$, $n_X = 77$, $\overline{Y} = 2.6$, $s_Y = 5.9$, $n_Y = 79$.

The null and alternate hypotheses are $H_0: \mu_X - \mu_Y \leq 1$ versus $H_1: \mu_X - \mu_Y > 1$.

$z = (4.7 - 2.6 - 1)/\sqrt{7.2^2/77 + 2.6^2/79} = 1.04$. Since the alternate hypothesis is of the form $\mu_X - \mu_Y > \Delta$, the P-value is the area to the right of $z = 1.04$.

Thus $P = 0.1492$.

We cannot conclude that the mean weight loss on the low-carbohydrate diet is more than 1 kg greater than that of the low-fat diet.

11. $\overline{X} = 7.38$, $s_X = 12.83$, $n_X = 92$, $\overline{Y} = 8.20$, $s_Y = 9.84$, $n_Y = 123$.

The null and alternate hypotheses are $H_0: \mu_X - \mu_Y \geq 0$ versus $H_1: \mu_X - \mu_Y < 0$.

$z = (7.38 - 8.20 - 0)/\sqrt{12.83^2/92 + 9.84^2/123} = -0.51$. Since the alternate hypothesis is of the form $\mu_X - \mu_Y < \Delta$, the P-value is the area to the left of $z = -0.51$.

Thus $P = 0.3050$.

We cannot conclude that the mean number of hours per week increased between 2002 and 2004.

13. (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 6.6

1. (a) $H_0: p_X - p_Y \geq 0$ versus $H_1: p_X - p_Y < 0$

(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 $H_0: p_X - p_Y \geq 0$ versus $H_1: p_X - p_Y < 0$.

$$z = \frac{0.960 - 0.970}{\sqrt{0.96375(1 - 0.96375)(1/1000 + 1/600)}} = -1.04.$$

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.

3. $X = 245$, $n_X = 307$, $\widehat{p}_X = 245/307 = 0.7980$, $Y = 304$, $n_Y = 347$, $\widehat{p}_Y = 304/347 = 0.8761$,

$\widehat{p} = (245 + 304)/(307 + 347) = 0.8394$.

The null and alternate hypotheses are $H_0: p_X - p_Y \geq 0$ versus $H_1: p_X - p_Y < 0$.

$$z = \frac{0.7980 - 0.8761}{\sqrt{0.8394(1 - 0.8394)(1/307 + 1/347)}} = -2.71.$$

Since the alternate hypothesis is of the form $p_X - p_Y < 0$, the P-value is the area to the left of $z = -2.71$.

Thus $P = 0.0034$.

We can conclude that the method is more accurate for new age songs. firms.

5. $X = 102$, $n_X = 230$, $\widehat{p}_X = 102/230 = 0.443478$, $Y = 20$, $n_Y = 72$, $\widehat{p}_Y = 20/72 = 0.277778$,

$\widehat{p} = (102 + 20)/(230 + 72) = 0.403974$.

The null and alternate hypotheses are $H_0: p_X - p_Y \leq 0$ versus $H_1: p_X - p_Y > 0$.

$z = \dfrac{0.443478 - 0.277778}{\sqrt{0.403974(1 - 0.403974)(1/230 + 1/72)}} = 2.50$. Since the alternate hypothesis is of the form $p_X - p_Y > 0$, the P-value is the area to the right of $z = 2.50$.

Thus $P = 0.0062$.

We can conclude that the proportion of patch users is greater among European-Americans.

7. $X = 20$, $n_X = 1200$, $\widehat{p}_X = 20/1200 = 0.016667$, $Y = 15$, $n_Y = 1500$, $\widehat{p}_Y = 15/1500 = 0.01$,
$\widehat{p} = (20+15)/(1200+1500) = 0.012963$.

The null and alternate hypotheses are $H_0: p_X - p_Y \le 0$ versus $H_1: p_X - p_Y > 0$.

$$z = \frac{0.016667 - 0.01}{\sqrt{0.012963(1-0.012963)(1/1200+1/1500)}} = 1.52.$$

Since the alternate hypothesis is of the form $p_X - p_Y > 0$, the P-value is the area to the right of $z = 1.52$.

Thus $P = 0.0643$.

The evidence suggests that heavy packaging reduces the proportion of damaged shipments, but may not be conclusive.

9. $X = 76$, $n_X = 164$, $\widehat{p}_X = 76/164 = 0.46341$, $Y = 48$, $n_Y = 96$, $\widehat{p}_Y = 48/96 = 0.5$,
$\widehat{p} = (76+48)/(164+96) = 0.47692$.

The null and alternate hypotheses are $H_0: p_X - p_Y \ge 0$ versus $H_1: p_X - p_Y < 0$.

$$z = \frac{0.46341 - 0.5}{\sqrt{0.47692(1-0.47692)(1/164+1/96)}} = -0.57.$$

Since the alternate hypothesis is of the form $p_X - p_Y < 0$, the P-value is the area to the right of $z = -0.57$.

Thus $P = 0.2843$.

We cannot conclude that the proportion of patients who fall is less for those who exercise.

11. $X = 195$, $n_X = 5320$, $\widehat{p}_X = 195/5320 = 0.036654$, $Y = 33$, $n_Y = 1120$, $\widehat{p}_Y = 33/1120 = 0.029463$,
$\widehat{p} = (195+33)/(5320+1120) = 0.035403$.

The null and alternate hypotheses are $H_0: p_X - p_Y = 0$ versus $H_1: p_X - p_Y \ne 0$.

$$z = \frac{0.036654 - 0.029463}{\sqrt{0.035403(1-0.035403)(1/5320+1/1120)}} = 1.18.$$

Since the alternate hypothesis is of the form $p_X - p_Y \ne 0$, the P-value is the sum of the areas to the right of $z = 1.18$ and to the left of $z = -1.18$.

Thus $P = 0.1190 + 0.1190 = 0.2380$.

We cannot conclude that the proportions differ between the two groups.

13. No, because the two samples are not independent.

15. (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 6.7

1. (a) $\overline{X} = 3.05$, $s_X = 0.34157$, $n_X = 4$, $\overline{Y} = 1.8$, $s_Y = 0.90921$, $n_Y = 4$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{0.34157^2}{4} + \dfrac{0.90921^2}{4}\right]^2}{\dfrac{(0.34157^2/4)^2}{4-1} + \dfrac{(0.90921^2/4)^2}{4-1}} = 3, \text{ rounded down to the nearest integer.}$$

$t_3 = (3.05 - 1.8 - 0)/\sqrt{0.34157^2/4 + 0.90921^2/4} = 2.574$.

The null and alternate hypotheses are $H_0: \mu_X - \mu_Y \leq 0$ versus $H_1: \mu_X - \mu_Y > 0$.

Since the alternate hypothesis is of the form $\mu_X - \mu_Y > \Delta$, the P-value is the area to the right of $t = 2.574$.

From the t table, $0.025 < P < 0.050$. A computer package gives $P = 0.041$.

We can conclude that the mean strength of crayons made with dye B is greater than that made with dye A.

(b) $\overline{X} = 3.05$, $s_X = 0.34157$, $n_X = 4$, $\overline{Y} = 1.8$, $s_Y = 0.90921$, $n_Y = 4$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{0.34157^2}{4} + \dfrac{0.90921^2}{4}\right]^2}{\dfrac{(0.34157^2/4)^2}{4-1} + \dfrac{(0.90921^2/4)^2}{4-1}} = 3, \text{ rounded down to the nearest integer.}$$

$t_3 = (3.05 - 1.8 - 1)/\sqrt{0.34157^2/4 + 0.90921^2/4} = 0.5148$.

The null and alternate hypotheses are $H_0: \mu_X - \mu_Y \leq 1$ versus $H_1: \mu_X - \mu_Y > 1$.

Since the alternate hypothesis is of the form $\mu_X - \mu_Y > \Delta$, the P-value is the area to the right of $t = 0.5148$.

From the t table, $0.25 < P < 0.40$. A computer package gives $P = 0.321$.

We cannot conclude that the mean strength of crayons made with dye B exceeds that of those made with dye A by more than 1 J.

3. $\overline{X} = 2.31$, $s_X = 0.89$, $n_X = 15$, $\overline{Y} = 2.8$, $s_Y = 1.1$, $n_Y = 15$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{0.89^2}{15} + \dfrac{1.1^2}{15}\right]^2}{\dfrac{(0.89^2/15)^2}{15-1} + \dfrac{(1.1^2/15)^2}{15-1}} = 26\text{, rounded down to the nearest integer.}$$

$t_{26} = (2.31 - 2.8 - 0)/\sqrt{0.89^2/15 + 1.1^2/15} = -1.341$.

The null and alternate hypotheses are $H_0: \mu_X - \mu_Y = 0$ versus $H_1: \mu_X - \mu_Y \neq 0$.

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 $t = 1.341$ and to the left of $t = -1.341$.

From the t table, $0.10 < P < 0.20$. A computer package gives $P = 0.191$.

We cannot conclude that the penetration resistance differs between the two depths.

5. $\overline{X} = 24.8$, $s_X = 0.91043$, $n_X = 10$, $\overline{Y} = 19.5$, $s_Y = 1.3106$, $n_Y = 10$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{0.91043^2}{10} + \dfrac{1.3106^2}{10}\right]^2}{\dfrac{(0.91043^2/10)^2}{10-1} + \dfrac{(1.3106^2/10)^2}{10-1}} = 16\text{, rounded down to the nearest integer.}$$

$t_{16} = (24.8 - 19.5 - 0)/\sqrt{0.91043^2/10 + 1.3106^2/10} = 10.502$.

The null and alternate hypotheses are $H_0: \mu_X - \mu_Y \leq 0$ versus $H_1: \mu_X - \mu_Y > 0$.

Since the alternate hypothesis is of the form $\mu_X - \mu_Y > \Delta$, the P-value is the area to the right of $t = 10.502$.

From the t table, $P < 0.0005$. A computer package gives $P = 6.914 \times 10^{-9}$.

We can conclude that the mean height of plants grown with slow-release fertilizer is greater.

7. $\overline{X} = 53.0$, $s_X = 1.41421$, $n_X = 6$, $\overline{Y} = 54.5$, $s_Y = 3.88587$, $n_Y = 6$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{1.41421^2}{6} + \dfrac{3.88587^2}{6}\right]^2}{\dfrac{(1.41421^2/6)^2}{6-1} + \dfrac{(3.88587^2/6)^2}{6-1}} = 6\text{, rounded down to the nearest integer.}$$

$t_6 = (53.0 - 54.5 - 0)/\sqrt{1.41421^2/6 + 3.88587^2/6} = -0.889$.

The null and alternate hypotheses are $H_0: \mu_X - \mu_Y \geq 0$ versus $H_1: \mu_X - \mu_Y < 0$.

Since the alternate hypothesis is of the form $\mu_X - \mu_Y < \Delta$, the P-value is the area to the left of $t = -0.889$.

From the t table, $0.10 < P < 0.25$. A computer package gives $P = 0.204$.

We cannot conclude that the mean cost of the new method is less than that of the old method.

9. $\overline{X} = 8.38$, $s_X = 0.96$, $n_X = 15$, $\overline{Y} = 9.83$, $s_Y = 1.02$, $n_Y = 15$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{0.96^2}{15} + \dfrac{1.02^2}{15}\right]^2}{\dfrac{(0.96^2/15)^2}{15-1} + \dfrac{(1.02^2/15)^2}{15-1}} = 27\text{, rounded down to the nearest integer.}$$

$t_{27} = (8.38 - 9.83 - 0)/\sqrt{0.96^2/15 + 1.02^2/15} = -4.009$.

The null and alternate hypotheses are $H_0: \mu_X - \mu_Y \geq 0$ versus $H_1: \mu_X - \mu_Y < 0$.

Since the alternate hypothesis is of the form $\mu_X - \mu_Y < \Delta$, the P-value is the area to the left of $t = -4.009$.

From the t table, $P < 0.0005$. A computer package gives $P = 0.000216$..

We can conclude that 6” spacing provides a higher mean failure pressure.

11. $\overline{X} = 36.893$, $s_X = 3.2211$, $n_X = 13$, $\overline{Y} = 33.000$, $s_Y = 2.5449$, $n_Y = 9$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{3.2211^2}{13} + \dfrac{2.5449^2}{9}\right]^2}{\dfrac{(3.2211^2/13)^2}{13-1} + \dfrac{(2.5449^2/9)^2}{9-1}} = 19\text{, rounded down to the nearest integer.}$$

$t_{19} = (36.893 - 33.000 - 0)/\sqrt{3.2211^2/13 + 2.5449^2/9} = 3.1601$.

The null and alternate hypotheses are $H_0: \mu_X - \mu_Y \leq 0$ versus $H_1: \mu_X - \mu_Y > 0$.

Since the alternate hypothesis is of the form $\mu_X - \mu_Y > \Delta$, the P-value is the area to the right of $t = 3.1601$.

From the t table, $0.001 < P < 0.005$. A computer package gives $P = 0.0025774$.

We can conclude that the mean sodium content is higher for brand B.

13. (a) $\overline{X} = 27.3$, $s_X = 5.2$, $n_X = 12$, $\overline{Y} = 32.7$, $s_Y = 4.1$, $n_Y = 14$.

The number of degrees of freedom is

$$\nu = \frac{\left[\dfrac{5.2^2}{12} + \dfrac{4.1^2}{14}\right]^2}{\dfrac{(5.2^2/12)^2}{12-1} + \dfrac{(4.1^2/14)^2}{14-1}} = 20\text{, rounded down to the nearest integer.}$$

$t_{20} = (27.3 - 32.7 - 0)/\sqrt{5.2^2/12 + 4.1^2/14} = -2.9056$.

The null and alternate hypotheses are $H_0: \mu_X - \mu_Y \geq 0$ versus $H_1: \mu_X - \mu_Y < 0$.

Since the alternate hypothesis is of the form $\mu_X - \mu_Y < \Delta$, the P-value is the area to the left of $t = -2.9056$.

From the t table, $0.001 < P < 0.005$. A computer package gives $P = 0.0043724$.

We can conclude that the mean time to sleep is less for the new drug.

15. $\overline{X} = 62.714$, $s_X = 3.8607$, $n_X = 7$, $\overline{Y} = 60.4$, $s_Y = 5.461$, $n_Y = 10$.

The number of degrees of freedom is

$$\nu = \frac{\left[\frac{3.8607^2}{7} + \frac{5.461^2}{10}\right]^2}{\frac{(3.8607^2/7)^2}{7-1} + \frac{(5.461^2/10)^2}{10-1}} = 14, \text{ rounded down to the nearest integer.}$$

$t_{14} = (62.714 - 60.4 - 0)/\sqrt{3.8607^2/7 + 5.461^2/10} = 1.024$.

The null and alternate hypotheses are $H_0: \mu_X - \mu_Y = 0$ versus $H_1: \mu_X - \mu_Y \neq 0$.

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 $t = 1.024$ and to the left of $t = -1.024$.

From the t table, $0.20 < P < 0.50$. A computer package gives $P = 0.323$.

We cannot conclude that the mean permeability coefficient at 60°C differs from that at 61°C.

17. (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 6.8

1. $\overline{D} = 0.17214$, $s_D = 0.44316$, $n = 14$. There are $14 - 1 = 13$ degrees of freedom.

The null and alternate hypotheses are $H_0: \mu_D = 0$ versus $H_1: \mu_D \neq 0$.

$t = (.17214 - 0)/(0.44316/\sqrt{14}) = 1.459$.

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 $t = 1.459$ and to the left of $t = -1.459$.

From the t table, $0.10 < P < 0.20$. A computer package gives $P = 0.16821$.

We cannot conclude that the mean amount absorbed differs between the brand name and the generic drug.

3. The differences are $4, -1, 0, 7, 3, 4, -1, 5, -1, 5$.

$\overline{D} = 2.5$, $s_D = 2.9907$, $n = 10$. There are $10 - 1 = 9$ degrees of freedom.

The null and alternate hypotheses are $H_0: \mu_D = 0$ versus $H_1: \mu_D \neq 0$.

$t = (2.5 - 0)/(2.9907/\sqrt{10}) = 2.643$.

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 $t = 2.643$

and to the left of $t = -2.643$.

From the t table, $0.02 < P < 0.05$. A computer package gives $P = 0.027$.

We can conclude that the etch rates differ between center and edge.

5. 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 $H_0 : \mu_D = 0$ versus $H_1 : \mu_D \neq 0$.

$t = (-10 - 0)/(9.3808/\sqrt{8}) = -3.015$.

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.

7. The differences are $35, 57, 14, 29, 31$.

$\overline{D} = 33.2$, $s_D = 15.498$, $n = 5$.

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

The null and alternate hypotheses are $H_0 : \mu_D \leq 0$ versus $H_1 : \mu_D > 0$.

$t = (33.2 - 0)/(15.498/\sqrt{5}) = 4.790$.

Since the alternate hypothesis is of the form $\mu_D > \Delta$, the P-value is the area to the right of $t = 4.790$.

From the t table, $0.001 < P < 0.005$. A computer package gives $P = 0.0044$.

We can conclude that the mean strength after 6 days is greater than the mean strength after 3 days.

9. The differences are $1, 2, -1, 0, 2, -1, 2$.

$\overline{D} = 0.714286$, $s_D = 1.38013$, $n = 7$. There are $7 - 1 = 6$ degrees of freedom.

The null and alternate hypotheses are $H_0 : \mu_D = 0$ versus $H_1 : \mu_D \neq 0$.

$t = (0.714286 - 0)/(1.38013/\sqrt{7}) = 1.3693$.

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 $t = 1.3693$ and to the left of $t = -1.3693$.

From the t table, $0.20 < P < 0.50$. A computer package gives $P = 0.220$.

We cannot conclude that the mean response differs between the two drugs.

11. (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 $H_0 : \mu_R - \mu_B \leq 0$ versus $H_1 : \mu_R - \mu_B > 0$.

$\overline{D} = 3.21$ $s_D = 1.1338$, $n = 10$. 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 $H_0: \mu_R - \mu_B \leq 2$ vs. $H_1: \mu_R - \mu_B > 2$.

$\overline{D} = 3.21$ $s_D = 1.1338$, $n = 10$. 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.

13. (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 6.9

1. (a) The signed ranks are

x	$x - 14$	Signed Rank
20	6	1
21	7	2
5	-9	-3
24	10	4
25	11	5
32	18	6
43	29	7

The null and alternate hypotheses are $H_0: \mu \leq 14$ versus $H_1: \mu > 14$.

The sum of the positive signed ranks is $S_+ = 25$. $n = 7$.

Since the alternate hypothesis is of the form $\mu > \mu_0$, the P-value is the area under the signed rank probability mass function corresponding to $S_+ \geq 25$.

From the signed-rank table, $P = 0.0391$.

We can conclude that the mean concentration is greater than 14 g/L.

(b) The signed ranks are

x	$x-30$	Signed Rank
32	2	1
25	-5	-2
24	-6	-3
21	-9	-4
20	-10	-5
43	13	6
5	-25	-7

The null and alternate hypotheses are $H_0: \mu \geq 14$ versus $H_1: \mu < 30$.

The sum of the positive signed ranks is $S_+ = 7$. $n = 7$.

Since the alternate hypothesis is of the form $\mu < \mu_0$, the P-value is the area under the signed rank probability mass function corresponding to $S_+ \leq 7$.

From the signed-rank table, $P > 0.1094$.

We cannot conclude that the mean concentration is less than 30 g/L.

(c) The signed ranks are

x	$x-18$	Signed Rank
20	2	1
21	3	2
24	6	3
25	7	4
5	-13	-5
32	14	6
43	25	7

The null and alternate hypotheses are $H_0: \mu = 18$ versus $H_1: \mu \neq 18$.

The sum of the positive signed ranks is $S_+ = 23$. $n = 7$.

Since the alternate hypothesis is of the form $\mu \neq \mu_0$, the P-value is twice the area under the signed rank probability mass function corresponding to $S_+ \geq 23$.

From the signed-rank table, $P = 2(0.0781) = 0.1562$.

We cannot conclude that the mean concentration differs from 18 g/L.

3. (a) The signed ranks are

x	$x-45$	Signed Rank
41.1	−3.9	−1
41.0	−4.0	−2
40.0	−5.0	−3
38.1	−6.9	−4
52.3	7.3	5
37.3	−7.7	−6
36.2	−8.8	−7
34.6	−10.4	−8
33.6	−11.4	−9
33.4	−11.6	−10
59.4	14.4	11
30.3	−14.7	−12
30.1	−14.9	−13
28.8	−16.2	−14
63.0	18.0	15
26.8	−18.2	−16
67.8	22.8	17
19.0	−26.0	−18
72.3	27.3	19
14.3	−30.7	−20
78.6	33.6	21
80.8	35.8	22
9.1	−35.9	−23
81.2	36.2	24

The null and alternate hypotheses are $H_0: \mu \geq 45$ versus $H_1: \mu < 45$.

The sum of the positive signed ranks is $S_+ = 134$. $n = 24$.

Since $n > 20$, compute the z-score of S_+ and use the z table.

$$z = \frac{S_+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} = -0.46.$$

Since the alternate hypothesis is of the form $\mu < \mu_0$, the P-value is the area to the left of $z = -0.46$. From the z table, $P = 0.3228$.

We cannot conclude that the mean conversion is less than 45.

(b) The signed ranks are

x	$x-30$	Signed Rank
30.1	0.1	1
30.3	0.3	2
28.8	−1.2	−3
26.8	−3.2	−4
33.4	3.4	5
33.6	3.6	6
34.6	4.6	7
36.2	6.2	8
37.3	7.3	9
38.1	8.1	10
40.0	10.0	11
19.0	−11.0	−12.5
41.0	11.0	12.5
41.1	11.1	14
14.3	−15.7	−15
9.1	−20.9	−16
52.3	22.3	17
59.4	29.4	18
63.0	33.0	19
67.8	37.8	20
72.3	42.3	21
78.6	48.6	22
80.8	50.8	23
81.2	51.2	24

The null and alternate hypotheses are $H_0: \mu \leq 30$ versus $H_1: \mu > 30$.

The sum of the positive signed ranks is $S_+ = 249.5$. $n = 24$.

Since $n > 20$, compute the z-score of S_+ and use the z table.

$$z = \frac{S_+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} = 2.84.$$

Since the alternate hypothesis is of the form $\mu > \mu_0$, the P-value is the area to the right of $z = 2.84$. From the z table, $P = 0.0023$.

We can conclude that the mean conversion is greater than 30.

(c) The signed ranks are

x	$x-55$	Signed Rank
52.3	-2.7	-1
59.4	4.4	2
63.0	8.0	3
67.8	12.8	4
41.1	-13.9	-5
41.0	-14.0	-6
40.0	-15.0	-7
38.1	-16.9	-8
72.3	17.3	9
37.3	-17.7	-10
36.2	-18.8	-11
34.6	-20.4	-12
33.6	-21.4	-13
33.4	-21.6	-14
78.6	23.6	15
30.3	-24.7	-16
30.1	-24.9	-17
80.8	25.8	18
28.8	-26.2	-19.5
81.2	26.2	19.5
26.8	-28.2	-21
19.0	-36.0	-22
14.3	-40.7	-23
9.1	-45.9	-24

The null and alternate hypotheses are $H_0: \mu = 55$ versus $H_1: \mu \neq 55$.

The sum of the positive signed ranks is $S_+ = 70.5$. $n = 24$.

Since $n > 20$, compute the z-score of S_+ and use the z table.

$$z = \frac{S_+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} = -2.27.$$

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 $z = 2.27$ and to the left of $z = -2.27$.

From the z table, $P = 0.0116 + 0.0116 = 0.0232$.

We can conclude that the mean conversion differs from 55.

5. The signed ranks are

x	$x-0$	Signed Rank
0.01	0.01	2
0.01	0.01	2
-0.01	-0.01	-2
0.03	0.03	4
0.05	0.05	5.5
-0.05	-0.05	-5.5
-0.07	-0.07	-7
-0.11	-0.11	-8
-0.13	-0.13	-9
0.15	0.15	10

The null and alternate hypotheses are $H_0: \mu = 0$ versus $H_1: \mu \neq 0$.

The sum of the positive signed ranks is $S_+ = 23.5$. $n = 10$.

Since the alternate hypothesis is of the form $\mu \neq \mu_0$, the P-value is twice the area under the signed rank probability mass function corresponding to $S_+ \leq 23.5$.

From the signed-rank table, $P > 2(0.1162) = 0.2324$.

We cannot conclude that the gauges differ.

7. The ranks of the combined samples are

Value	Rank	Sample
12	1	X
13	2	X
15	3	X
18	4	Y
19	5	X
20	6	X
21	7	X
23	8	Y
24	9	Y
27	10	X
30	11	Y
32	12	Y
35	13	Y
40	14	Y

The null and alternate hypotheses are $H_0: \mu_X - \mu_Y = 0$ versus $H_1: \mu_X - \mu_Y \neq 0$.

The test statistic W is the sum of the ranks corresponding to the X sample.

$W = 34$. The sample sizes are $m = 7$ and $n = 7$.

Since the alternate hypothesis is of the form $\mu_X - \mu_Y \neq \Delta$, the P-value is twice the area under the rank-sum probability mass function corresponding to $W \leq 34$.

From the rank-sum table, $P = 2(0.0087) = 0.0174$.

We can conclude that the mean recovery times differ.

9. The ranks of the combined samples are

Value	Rank	Sample
51	1	Y
59	2	Y
64	4	X
64	4	Y
64	4	Y
66	6	X
69	7	X
72	8	X
74	9.5	X
74	9.5	Y
75	11.5	X
75	11.5	Y
77	14	X
77	14	X
77	14	Y
80	16.5	Y
80	16.5	Y
81	18.5	Y
81	18.5	Y
83	20.5	X
83	20.5	Y
85	22.5	X
85	22.5	Y
86	24	Y
88	25	X
91	26	X

The null and alternate hypotheses are $H_0: \mu_X - \mu_Y = 0$ versus $H_1: \mu_X - \mu_Y \neq 0$.

The test statistic W is the sum of the ranks corresponding to the X sample.

$W = 168$. The sample sizes are $m = 12$ and $n = 14$.

Since n and m are both greater than 8, compute the z-score of W and use the z table.

$$z = \frac{W - m(m+n+1)/2}{\sqrt{mn(m+n+1)/12}} = 0.31.$$

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 $z = 0.31$ and to the left of $z = -0.31$. From the z table, $P = 0.3783 + 0.3783 = 0.7566$.

We cannot conclude that the mean scores differ.

Section 6.10

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 $H_0: p_1 = 0.85,\ p_2 = 0.10,\ p_3 = 0.05$

(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 $O_{.1} = 1603$ and $O_{.2} = 2036$. The column totals are $O_{1.} = 504$, $O_{2.} = 229$, $O_{3.} = 686$, $O_{4.} = 1105$, $O_{5.} = 1115$.

The grand total is $O_{..} = 3639$.

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

	Poor	Near Poor	Low Income	Middle Income	High Income
Men	222.01	100.88	302.19	486.76	491.16
Women	281.99	128.12	383.81	618.24	623.84

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

$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{5}(O_{ij} - E_{ij})^2/E_{ij} = 108.35$.

From the χ^2 table, $P < 0.005$. A computer package gives $P \approx 0$.

We can conclude that the proportions in the various income categories differ between men and women.

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 $H_0: p_1 = 0.10,\ p_2 = 0.70,\ p_3 = 0.15,\ p_4 = 0.05$

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 6.11

1. $s^2 = 96$, $n = 11$. The null and alternate hypotheses are $H_0 : \sigma^2 \leq 50$, $H_1 : \sigma^2 > 50$.
The test statistic is $(11-1)(96)/50 = 19.2$. There are $11-1 = 10$ degrees of freedom.
Since the alternate hypothesis is of the form $\sigma^2 > \sigma_0^2$, the P-value is the area to the right of $\chi^2_{10} = 19.2$.
From the χ^2 table, $0.025 < P < 0.05$. A computer package gives $P = 0.0378$.
There is reasonably strong evidence that $\sigma^2 > 50$.

3. $s^2 = 64$, $n = 25$. The null and alternate hypotheses are $H_0 : \sigma^2 = 225$, $H_1 : \sigma^2 \neq 225$.
The test statistic is $(25-1)(64)/225 = 6.8267$. There are $25-1 = 24$ degrees of freedom.
Since the alternate hypothesis is of the form $\sigma^2 \neq \sigma_0^2$, the P-value is twice the area to the left of $\chi^2_{24} = 6.8267$.
From the χ^2 table, $P < 0.01$. A computer package gives $P = 0.000463$.
There is sufficient evidence to contradict the claim.

5. $s^2 = 18.49$, $n = 25$. The null and alternate hypotheses are $H_0 : \sigma^2 \geq 25$, $H_1 : \sigma^2 < 25$.
The test statistic is $(25-1)(18.49)/25 = 17.750$. There are $25-1 = 24$ degrees of freedom.
Since the alternate hypothesis is of the form $\sigma^2 < \sigma_0^2$, the P-value is the area to the left of $\chi^2_{24} = 17.750$.
From the χ^2 table, $P > 0.10$. A computer package gives $P = 0.185$.
We cannot conclude that the pediatrician's claim is true.

7. $s^2 = 7569$, $n = 20$. The null and alternate hypotheses are $H_0 : \sigma^2 = 10{,}000$, $H_1 : \sigma^2 \neq 10,000$.
The test statistic is $(20-1)(7569)/10{,}000 = 14.381$. There are $20-1 = 19$ degrees of freedom.
Since the alternate hypothesis is of the form $\sigma^2 \neq \sigma_0^2$, the P-value is twice the area to the left of $\chi^2_{19} = 14.381$.
From the χ^2 table, $P > 0.20$. A computer package gives $P = 0.478$.
There is not sufficient evidence to contradict the claim.

9. $\nu_1 = 7$, $\nu_2 = 20$. From the F table, the upper 5% point is 2.51.

11. (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 $P = 0.02$.

13. The sample variance of the sodium contents for the brand A is $s_1^2 = 6.4764$. The sample size is $n_1 = 9$.
The sample variance of the sodium contents for brand B is $s_2^2 = 10.3572$. The sample size is $n_2 = 13$.
The null and alternate hypotheses are $H_0 : \sigma_1^2/\sigma_2^2 = 1$ versus $H_1 : \sigma_1^2/\sigma_2^2 \neq 1$.
The test statistic is $F = s_2^2/s_1^2 = 1.6020$. The numbers of degrees of freedom are 12 and 8.
Since this is a two-tailed test, the P-value is twice the area to the right of 1.6020 under the $F_{12,8}$ probability

density function.

From the F table, $P > 0.2$. A computer package gives $P = 0.513$.

We cannot conclude that the variance of the sodium content differs between the brands.

Section 6.12

1. (a) True. H_0 is rejected at any level greater than or equal to 0.03.

   (b) False. 2% is less than 0.03.

   (c) False. 10% is greater than 0.03.

3. (a) $H_0 : \mu \geq 90$ versus $H_1 : \mu < 90$

   (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) H_0 says that the claim is true. Therefore H_0 is true, so rejecting it is a type I error.

   (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. The 1% level. The error in question is rejecting H_0 when true, which is a type I error. When the level is smaller, it is less likely to reject H_0, and thus less likely to make a type I error.

Section 6.13

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 0.90, not 0.10.

   (d) False. The power is not the probability that H_0 is 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. (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 $P(\overline{X} \le 49,400)$.

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 $\overline{X} \le x_5$, 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 $\overline{X} \le 49,177.5$.

The power is therefore $P(\overline{X} \le 49,177.5)$ 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 $\overline{X} \le x_0$ where $P(\overline{X} \le x_0) = 0.80$ 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 $\overline{X} \le 49,920$.

The level is $P(\overline{X} \le 49,920)$ 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 $n = 618$.

7. (ii). Since 7 is farther from the null mean of 10 than 8 is, the power against the alternative $\mu = 7$ will be greater than the power against the alternative $\mu = 8$.

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 6.14

1. Several tests have been performed, so we cannot interpret the P-values in the way that we do when only one test is performed.

3. (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.

5. The original P-value must be $0.05/20 = 0.0025$.

7. (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.

Section 6.15

1. (a) Using Method 1, the limits of the 95% confidence interval are the 2.5 and 97.5 percentiles of the bootstrap means, $\overline{X}_{.025}$ and $\overline{X}_{.975}$.

The 2.5 percentile is $(Y_{25} + Y_{26})/2 = 38.3818$, and the 97.5 percentile is $(Y_{975} + Y_{976})/2 = 38.53865$.

The confidence interval is (38.3818, 38.53865). Values outside this interval can be rejected at the 5% level.

Therefore null hypotheses (ii) and (iv) can be rejected.

(b) Using Method 1, the limits of the 90% confidence interval are the 5th and 95th percentiles of the bootstrap means, $\overline{X}_{.05}$ and $\overline{X}_{.95}$.

The 5th percentile is $(Y_{50} + Y_{51})/2 = 38.39135$, and the 95th percentile is $(Y_{950} + Y_{951})/2 = 38.5227$.

The confidence interval is (38.39135, 38.5227). Values outside this interval can be rejected at the 10% level.

Therefore null hypotheses (i), (ii), and (iv) can be rejected.

3. No, the value 103 is an outlier.

5. (a) $s_A^2 = 200.28$, $s_B^2 = 39.833$, $s_A^2/s_B^2 = 5.02$.

   (b) No, the F-test requires the assumption that the data are normally distributed. These data contain an outlier (103), so the F-test should not be used.

   (c) $P \approx 0.37$

7. (a) The test statistic is $t = \dfrac{\overline{X} - 7}{s/\sqrt{7}}$. H_0 will be rejected if $|t| > 2.447$.

   (b) The power is approximately 0.60.

9. (a) $X = 22.10$, $\sigma_X = 0.34$, $Y = 14.30$, $\sigma_Y = 0.32$, $V = \sqrt{X^2 + Y^2} = 26.323$.

   $\dfrac{\partial V}{\partial X} = X/\sqrt{X^2+Y^2} = 0.83957$, $\quad \dfrac{\partial V}{\partial Y} = Y/\sqrt{X^2+Y^2} = 0.54325$,

   $$\sigma_V = \sqrt{\left(\frac{\partial V}{\partial X}\right)^2 \sigma_X^2 + \left(\frac{\partial V}{\partial Y}\right)^2 \sigma_Y^2} = 0.3342.$$

   $V = 26.323$, $\sigma_V = 0.3342$

   (b) $V = 26.323$, $\sigma_V = 0.3342$. The null and alternate hypotheses are $H_0 : \mu_V \leq 25$ versus $H_1 : \mu_V > 25$.

   $z = (26.323 - 25)/0.3342 = 3.96$.

   Since the alternate hypothesis is of the form $\mu > \mu_0$, the P-value is the area to the right of $z = 3.96$.

   Thus $P \approx 0$.

   (c) Yes, V is approximately normally distributed.

Supplementary Exercises for Chapter 6

1. 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 $H_0 : \mu_1 - \mu_2 \geq 0$ versus $H_1 : \mu_1 - \mu_2 < 0$. 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.

3. This requires a test for a population proportion. Let p represent the population proportion of defective parts under the new program. The appropriate null and alternate hypotheses are $H_0: p \geq 0.10$ versus $H_1: p < 0.10$. The test statistic is the sample proportion of defective parts. The z table should be used to find the P-value.

5. (a) $H_0: \mu \geq 16$ versus $H_1: \mu < 16$

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

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

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

(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.

7. (a) $H_0: \mu_1 - \mu_2 = 0$ versus $H_1: \mu_1 - \mu_2 \neq 0$

(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 $H_0: \mu_D = 0$ versus $H_1: \mu_D \neq 0$.

$t = (1608.143 - 0)/(2008.147/\sqrt{7}) = 2.119$.

(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, $0.05 < P < 0.010$. A computer package gives $P = 0.078$.

The null hypothesis is suspect, but one would most likely not firmly conclude that it is false.

9. $\overline{X} = 5.6$, $s = 1.2$, $n = 85$. The null and alternate hypotheses are $H_0: \mu \leq 5$ versus $H_1: \mu > 5$.

$z = (5.6 - 5)/(1.2/\sqrt{85}) = 4.61$.

Since the alternate hypothesis is of the form $\mu > \mu_0$, the P-value is the area to the right of $z = 4.61$.

Thus $P \approx 0$.

11. (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%.

13. (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.

15. $X = 37$, $n = 37 + 458 = 495$, $\widehat{p} = 37/495 = 0.074747$.

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.

17. (a) Both samples have a median of 20.

(b) The ranks of the combined samples are

Value	Rank	Sample
−10	1	Y
−9	2	Y
−8	3	Y
−7	4	Y
−6	5	Y
−5	6	Y
−4	7	Y
1	8	X
2	9	X
3	10	X
4	11	X
5	12	X
6	13	X
7	14	X
20	15.5	X
20	15.5	Y
21	17	Y
22	18	Y
23	19	Y
24	20	Y
25	21	Y
26	22	Y
27	23	Y
40	24	X
50	25	X
60	26	X
70	27	X
80	28	X
90	29	X
100	30	X

The null and alternate hypotheses are H_0 : median of X − median of $Y = 0$ versus H_1 : median of X − median of $Y \neq 0$.

The test statistic W is the sum of the ranks corresponding to the X sample.

$W = 281.5$. The sample sizes are $m = 15$ and $n = 15$.

Since n and m are both greater than 8, compute the z-score of W and use the z table.

$$z = \frac{W - m(m+n+1)/2}{\sqrt{mn(m+n+1)/12}} = 2.03.$$

Since the alternate hypothesis is of the form median of X − median of $Y \neq \Delta$, the P-value is the sum of the areas to the right of $z = 2.03$ and to the left of $z = -2.03$.

From the z table, $P = 0.0212 + 0.0212 = 0.0424$.

The P-value is fairly small, and so it provides reasonably strong evidence that the population medians are different.

(c) No, the X sample is heavily skewed to the right, while the Y sample is strongly bimodal. It does not seem reasonable to assume that these samples came from populations of the same shape.

19. (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 appropriate null and alternate hypotheses are $H_0 : \mu_A - \mu_B \leq 0$ versus $H_1 : \mu_A - \mu_B > 0$.

(b) $\overline{A} = 54.919$, $s_A = 2.5522$, $n_A = 16$, $\overline{B} = 53.019$, $s_B = 2.7294$, $n_B = 16$.

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 $H_0 : \mu_A - \mu_B \leq 0$ versus $H_1 : \mu_A - \mu_B > 0$.

Since the alternate hypothesis is of the form $\mu_A - \mu_B > \Delta$, 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 $P = 0.026$.

Since $P \leq 0.05$, we can conclude at the 5% level that the means are different.

21. (a) Yes.

(b) The conclusion is not justified. The engineer is concluding that H_0 is true because the test failed to reject.

23. 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. $\overline{x} = \overline{y} = 4$, $\sum_{i=1}^n (x_i - \overline{x})^2 = 28$, $\sum_{i=1}^n (y_i - \overline{y})^2 = 28$, $\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = 23$.

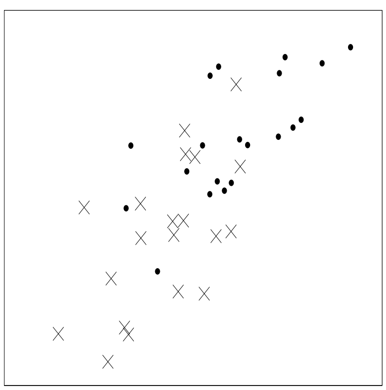
$$r = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}\sqrt{\sum_{i=1}^n (y_i - \overline{y})^2}} = 0.8214.$$

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

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

The correlation coefficient between temperature and stirring rate is

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

(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.

9. (a) $\overline{x} = 201.3$, $\overline{y} = 202.2$, $\sum_{i=1}^{n}(x_i - \overline{x})^2 = 2164.1$, $\sum_{i=1}^{n}(y_i - \overline{y})^2 = 1571.6$, $\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = 1716.4$, $n = 10$.

$r = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} = 0.930698.$

$W = \dfrac{1}{2}\ln\dfrac{1+r}{1-r} = 1.663582$, $\sigma_W = \sqrt{1/(10-3)} = 0.377964$.

A 95% confidence interval for μ_W is $W \pm 1.96\sigma_W$, or $(0.92279, 2.40438)$.

A 95% confidence interval for ρ is $\left(\dfrac{e^{2(0.92279)} - 1}{e^{2(0.92279)} + 1}, \dfrac{e^{2(2.40438)} - 1}{e^{2(2.40438)} + 1}\right)$, or $(0.7272, 0.9838)$.

(b) The null and alternate hypotheses are $H_0 : \rho \geq 0.9$ versus $H_1 : \rho < 0.9$.

$r = 0.930698$, $W = \dfrac{1}{2}\ln\dfrac{1+r}{1-r} = 1.663582$, $\sigma_W = \sqrt{1/(10-3)} = 0.377964$.

Under H_0, take $\rho = 0.9$, so $\mu_W = \dfrac{1}{2}\ln\dfrac{1+0.9}{1-0.9} = 1.47222$.

The null distribution of W is therefore normal with mean 1.47222 and standard deviation 0.377964.

The z-score of 1.663582 is $(1.663582 - 1.47222)/0.377964 = 0.51$. Since the alternate hypothesis is of the form $\rho > \rho_0$, the P-value is the area to the right of $z = 0.51$.

Thus $P = 0.3050$.

We cannot conclude that $\rho > 0.9$.

(c) $r = 0.930698$, $n = 10$, $U = r\sqrt{n-2}/\sqrt{1-r^2} = 7.1965$.

Under H_0, U has a Student's t distribution with $10 - 2 = 8$ degrees of freedom.

Since the alternate hypothesis is of the form $\rho > \rho_0$, the P-value is the area to the right of $t = 7.1965$.

From the t table, $P < 0.001$. A computer package gives $P = 4.64 \times 10^{-5}$.

We conclude that $\rho > 0$.

11. $r = -0.9515$, $W = \dfrac{1}{2}\ln\dfrac{1+r}{1-r} = -1.847395$, $\sigma_W = \sqrt{1/(30000-3)} = 0.00577379$.

A 95% confidence interval for μ_W is $W \pm 1.96\sigma_W$, or $(-1.85871, -1.83608)$.

A 95% confidence interval for ρ is $\left(\dfrac{e^{2(-1.85871)}-1}{e^{2(-1.85871)}+1}, \dfrac{e^{2(-1.83608)}-1}{e^{2(-1.83608)}+1}\right)$, or $(-0.9526, -0.9504)$.

13. $r = -0.509$, $n = 23$, $U = r\sqrt{n-2}/\sqrt{1-r^2} = -2.7098$.

Under H_0, U has a Student's t distribution with $23-2=21$ degrees of freedom.

Since the alternate hypothesis is $H_1 : \rho \neq 0$, 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 7.2

1. (a) $245.82 + 1.13(65) = 319.27$ pounds

(b) The difference in y predicted from a one-unit change in x is the slope $\widehat{\beta}_1 = 1.13$. Therefore the change in the number of lbs of steam predicted from a change of 5°C is $1.13(5) = 5.65$ pounds.

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

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

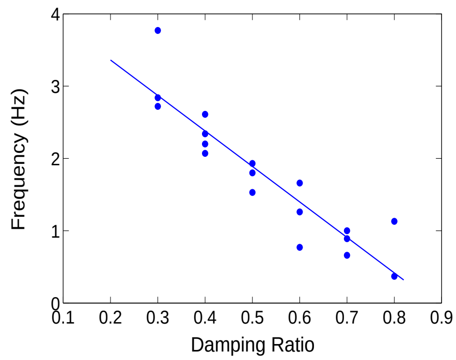
(b) Let x be the required height. Then $19 = -0.2967 + 0.2738x$, so $x = 70.477$ in.

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

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. (a)



(b) $\overline{x} = 0.527778$, $\overline{y} = 1.752778$, $\sum_{i=1}^n (x_i - \overline{x})^2 = 0.476111$, $\sum_{i=1}^n (y_i - \overline{y})^2 = 13.672761$, $\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = -2.335389$.

$\widehat{\beta}_1 = \dfrac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = -4.905134$ and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} = 4.341599$.

The equation of the least-squares line is $y = 4.341599 - 4.905134x$.

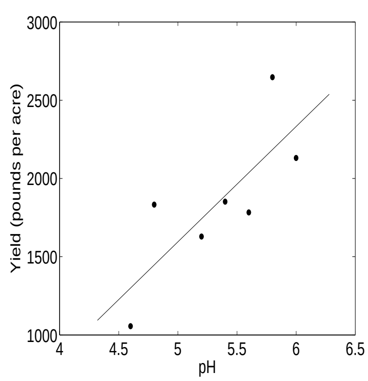
(c) By $4.905134(0.2) = 0.9810$ Hz.

(d) $4.341599 - 4.905134(0.75) = 0.6627$

(e) No, values greater than 1 are outside the range of the data.

(f) Let x be the required damping ratio. Then $2.0 = 4.341599 - 4.905134x$, so $x = 0.47738$.

11. (a)



The linear model is appropriate.

(b) $\overline{x} = 5.342857$, $\overline{y} = 1847.285714$, $\sum_{i=1}^n (x_i - \overline{x})^2 = 1.577143$, $\sum_{i=1}^n (y_i - \overline{y})^2 = 1398177.429$, $\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = 1162.514286$.

$\widehat{\beta}_1 = \dfrac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = 737.101449$ and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} = -2090.942029$.

The equation of the least-squares line is $y = -2090.942029 + 737.101449x$.

(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.

		Fitted Value	Residual
x	y	$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x$	$e = y - \widehat{y}$
4.6	1056	1299.7	−243.72
4.8	1833	1447.1	385.86
5.2	1629	1742.0	−112.99
5.4	1852	1889.4	−37.41
5.6	1783	2036.8	−253.83
5.8	2647	2184.2	−462.75
6.0	2131	2331.7	−200.67

(d) $737.101449(0.1) = 73.71$. Increase by 73.71 pounds per acre.

(e) $-2090.942029 + 737.10144(5.5) = 1963.1$ hours.

(f) No, because 7 is outside the range of the pH values present in the data.

(g) Let x be the required pH. Then $1500 = -2090.942029 + 737.10144x$, so $x = 4.872$.

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

15. (iii) equal to \$34,900. Since 70 inches is equal to $\overline{x}$, the predicted y value, $\widehat{y}$ will be equal to $\overline{y} = 34,900$.
To see this, note that $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1\overline{x} = (\overline{y} - \widehat{\beta}_1\overline{x}) + \widehat{\beta}_1\overline{x} = \overline{y}$.

Section 7.3

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) $\widehat{\beta}_1 = 0.329642$, $\quad s_{\widehat{\beta}_1} = 0.0546207$, $\quad n = 12$. There are $12 - 2 = 10$ degrees of freedom.

The null and alternate hypotheses are $H_0 : \beta_1 \geq 0.5$ versus $H_1 : \beta_1 < 0.5$.

$t = (0.329642 - 0.5)/0.0546207 = -3.119.$

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) $x = 40$, $\quad \widehat{y} = 7.623276 + 0.329642(40) = 20.808952.$

$s_{\widehat{y}} = s\sqrt{\dfrac{1}{n} + \dfrac{(x - \overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 1.834204$. 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) $x = 40$, $\quad \widehat{y} = 7.623276 + 0.329642(40) = 20.808952.$

$s_{\text{pred}} = s\sqrt{1 + \dfrac{1}{n} + \dfrac{(x - \overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 4.621721$. 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) $\overline{x} = 1.547286$, $\overline{y} = 0.728571$, $\sum_{i=1}^n (x_i - \overline{x})^2 = 0.141471$, $\sum_{i=1}^n (y_i - \overline{y})^2 = 0.0246857$, $\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = 0.0561429$, $n = 7$.

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = 0.396849 \text{ and } \widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} = 0.114532.$$

The least-squares line is $y = 0.114532 + 0.396849x$

(b) $$r^2 = \frac{[\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})]^2}{\sum_{i=1}^n (x_i - \overline{x})^2 \sum_{i=1}^n (y_i - \overline{y})^2} = 0.902557, \quad s = \sqrt{\frac{(1 - r^2)\sum_{i=1}^n (y_i - \overline{y})^2}{n-2}} = 0.0219338,$$

(c) $$s_{\widehat{\beta}_1} = \frac{s}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.058315.$$

There are $7 - 2 = 5$ degrees of freedom. $t_{5,.025} = 2.571$.

Therefore a 95% confidence interval for β_1 is $0.396849 \pm 2.571(0.058315)$, or $(0.247, 0.547)$.

(d) $$\widehat{y} = 0.709806, \quad s_{\widehat{y}} = s\sqrt{\frac{1}{n} + \frac{(x - \overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.00873675.$$

$t_{5,.025} = 2.571$. A 95% confidence interval is $0.709806 \pm 2.571(0.00873675)$ or (0.687, 0.732).

(e) Let $y_{1.5} = \beta_0 + \beta_1(1.5)$ denote the mean COA of bricks whose pore volume is 1.5.

The null and alternate hypotheses are $H_0: y_{1.5} \geq 0.75$ versus $H_1: y_{1.5} < 0.75$.

$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1(1.5) = 0.709806$, $\quad s_{\widehat{y}} = s\sqrt{\dfrac{1}{n} + \dfrac{(x-\overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.00873675$.

$t = (0.709806 - 0.75)/0.00873675 = -4.601$. There are $7 - 2 = 5$ degrees of freedom.

Since the alternate hypothesis is $y_{1.5} < 0.75$, the P-value is the area to the left of $t = -4.601$.

From the t table, $0.001 < P < 0.005$. A computer package gives $P = 0.00292$.

We may conclude that the mean COA of bricks whose pore volume is 1.5 is less than 0.75 volts.

(f) $\widehat{y} = 0.709806$, $\quad s_{\text{pred}} = s\sqrt{1 + \dfrac{1}{n} + \dfrac{(x-\overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.0236098$.

$t_{5,.025} = 2.571$. A 95% prediction interval is $0.709806 \pm 2.571(0.0236098)$ or (0.649, 0.770).

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 $x - \overline{x}$.

$\overline{x} = 21.7548$. The value 20 is closest to $\overline{x}$ and the value 15 is the farthest from $\overline{x}$.

Therefore the confidence interval at 20 would be the shortest, and the confidence interval at 15 would be the longest.

13. $s^2 = (1-r^2)\sum_{i=1}^n (y_i - \overline{y})^2/(n-2) = (1-0.9111)(234.19)/(17-2) = 1.388$.

15. (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 $P = 0.0005 + 0.0005 = 0.001$.

17. (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.

Section 7.4

1. (a) $\ln y = -0.4442 + 0.79833 \ln x$

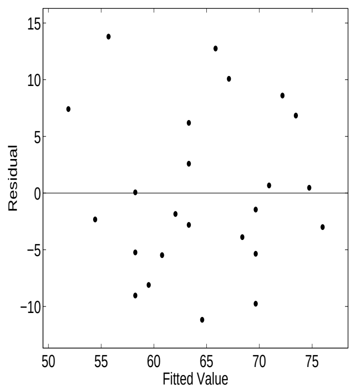
(b) $\widehat{y} = e^{\ln \widehat{y}} = e^{-0.4442+0.79833(\ln 2500)} = 330.95$.

(c) $\widehat{y} = e^{\ln \widehat{y}} = e^{-0.4442+0.79833(\ln 1600)} = 231.76$.

(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).

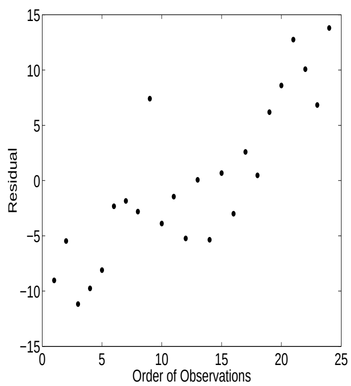
3. (a) $y = 20.162 + 1.269x$

(b)



There is no apparent pattern to the residual plot. The linear model looks fine.

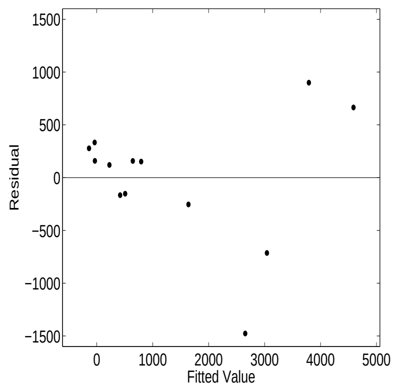
(c)



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.

5. (a) $y = -235.32 + 0.695x$.

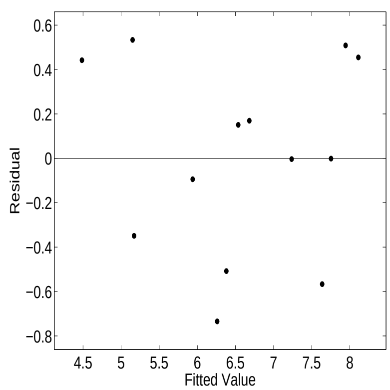
(b)



The residual plot shows a pattern, with positive residuals at the higher and lower fitted values, and negative residuals in the middle. The model is not appropriate.

(c) $\ln y = -0.0745 + 0.925 \ln x$.

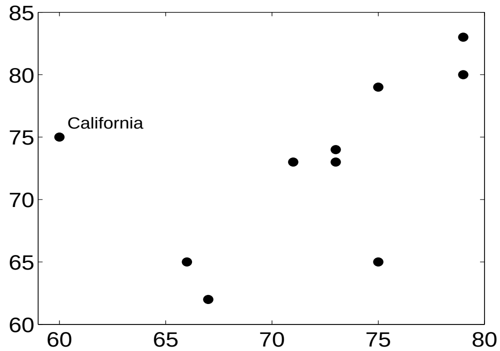
(d)



The residual plot shows no obvious pattern. The model is appropriate.

(e) The log model is more appropriate. The 95% prediction interval is $(197.26, 1559.76)$.

7. (a)



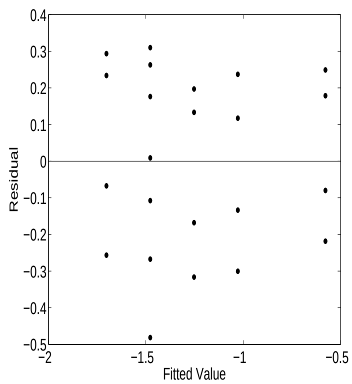
(b) $y = 30.122 + 0.596x$.

(c) $y = -22.714 + 1.305$.

(d) Yes, the outlier is influential. The equation of the least-squares line changes substantially when the outlier is removed.

9. (a) The model is $\log_{10} y = \beta_0 + \beta_1 \log_{10} x + \varepsilon$. Note that the natural log (ln) could be used in place of $\log_{10}$, but common logs are more convenient since partial pressures are expressed as powers of 10.

(b)



The least-squares line is $\log_{10} y = -3.277 - 0.225 \log_{10} x$. The linear model appears to fit quite well.

(c) The theory says that the coefficient β_1 of $\log_{10} x$ in the linear model is equal to -0.25. The estimated value is $\widehat{\beta} = -0.225$. We determine whether the data are consistent with the theory by testing the hypotheses $H_0 : \beta_1 = -0.25$ vs. $H_1 : \beta_1 \neq -0.25$. The value of the test statistic is $t = 0.821$. There are 21 degrees of freedom, so $0.20 < P < 0.50$ (a computer package gives $P = 0.210$). We do not reject H_0, so the data are consistent with the theory.

11. (a) $y = 2049.87 - 4.270x$

(b) (12, 2046) and (13, 1954) are outliers. The least-squares line with (12, 2046) deleted is $y = 2021.85 - 2.861x$. The least-squares line with (13, 1954) deleted is $y = 2069.30 - 5.236x$. The least-squares line with both outliers deleted is $y = 2040.88 - 3.809x$.

(c) The slopes of the least-squares lines are noticeably affected by the outliers. They ranged from -2.861 to -5.236.

13. 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 $\beta_1 = b$.

15. (a) A physical law.

(b) It would be better to redo the experiment. If the results of an experiment violate a physical law, then something was wrong with the experiment, and you can't fix it by transforming variables.

Supplementary Exercises for Chapter 7

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

$\widehat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = 1.016879$ and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1\overline{x} = -0.038981$.

(b) 0

(c) The molar absorption coefficient M.

(d) $r^2 = \dfrac{[\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})]^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2} = 0.989726$, $s = \sqrt{\dfrac{(1 - r^2)\sum_{i=1}^{n}(y_i - \overline{y})^2}{n - 2}} = 0.0474028$,

$s_{\widehat{\beta}_0} = s\sqrt{\dfrac{1}{n} + \dfrac{\overline{x}^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}} = 0.0910319$.

The null and alternate hypotheses are $H_0 : \beta_0 = 0$ versus $H_1 : \beta_0 \neq 0$.

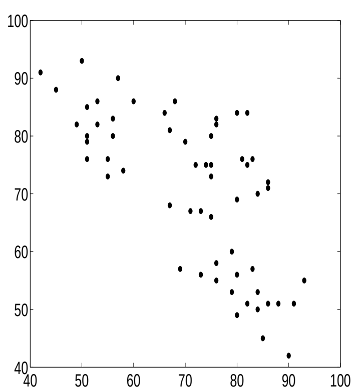
There are $n - 2 = 3$ degrees of freedom. $t = (-0.038981 - 0)/0.0910319 = -0.428$.

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 $t = 0.428$ and to the left of $t = -0.428$.

From the t table, $0.50 < P < 0.80$. A computer package gives $P = 0.698$.

The data are consistent with the Beer-Lambert law.

3. (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 $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} = 120.175090$.

$T_{i+1} = 120.175090 - 0.695669 T_i$.

(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).

5. (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 $H_0 : \beta_0 = 0$ versus $H_1 : \beta_0 \neq 0$.

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 $H_0 : \beta_1 = 1$ versus $H_1 : \beta_1 \neq 1$.

$\widehat{\beta}_1 = 0.941818$, $s_{\widehat{\beta}_1} = \dfrac{s}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.010858$.

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.

7. (a) $\overline{x} = 92.0$, $\overline{y} = 0.907407$, $\sum_{i=1}^n (x_i - \overline{x})^2 = 514.0$, $\sum_{i=1}^n (y_i - \overline{y})^2 = 4.738519$, $\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = 20.1$, $n = 27$.

$\widehat{\beta}_1 = \dfrac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = 0.0391051$ and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} = -2.690258$

$y = -2.690258 + 0.0391051x$

(b) $r^2 = \dfrac{[\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})]^2}{\sum_{i=1}^n (x_i - \overline{x})^2 \sum_{i=1}^n (y_i - \overline{y})^2} = 0.165877$, $\quad s = \sqrt{\dfrac{(1-r^2)\sum_{i=1}^n (y_i - \overline{y})^2}{n-2}} = 0.397618.$

$\widehat{\beta}_0 = -2.690258$, $\quad s_{\widehat{\beta}_0} = s\sqrt{\dfrac{1}{n} + \dfrac{\overline{x}^2}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 1.615327.$

$\widehat{\beta}_1 = 0.0391051$, $\quad s_{\widehat{\beta}_1} = \dfrac{s}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.0175382.$

There are $27 - 2 = 25$ degrees of freedom. $t_{25,.025} = 2.060$.

Therefore a 95% confidence interval for β_0 is $-2.690258 \pm 2.060(1.615327)$, or $(-6.0171, 0.6366)$. A 95% confidence interval for β_1 is $0.0391051 \pm 2.060(0.0175382)$, or $(0.00298, 0.0752)$.

(c) $x = 93$, $\quad \widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1(93) = 0.94651.$

(d) $x = 93$, $\quad \widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1(93) = 0.94651.$

$s_{\widehat{y}} = s\sqrt{\dfrac{1}{n} + \dfrac{(x - \overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.078506.$

There are 25 degrees of freedom. $t_{25,.025} = 2.060$.

Therefore a 95% confidence interval for the mean response is $0.94651 \pm 2.060(0.078506)$, or (0.785, 1.108).

(e) $x = 93$, $\quad \widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1(93) = 0.94651.$

$s_{\text{pred}} = s\sqrt{1 + \dfrac{1}{n} + \dfrac{(x - \overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.405294.$

There are 25 degrees of freedom. $t_{25,.025} = 2.060$.

Therefore a 95% prediction interval is $0.94651 \pm 2.060(0.405294)$, or (0.112, 1.781).

9. (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 $v_i = \ln y_i$.

$\overline{u} = 1.755803$, $\overline{v} = -0.563989$, $\sum_{i=1}^n (u_i - \overline{u})^2 = 0.376685$, $\quad \sum_{i=1}^n (v_i - \overline{v})^2 = 0.160487$,

$\sum_{i=1}^n (u_i - \overline{u})(v_i - \overline{v}) = 0.244969$, $\quad n = 5$.

$\widehat{\beta}_1 = \dfrac{\sum_{i=1}^n (u_i - \overline{u})(v_i - \overline{v})}{\sum_{i=1}^n (u_i - \overline{u})^2} = 0.650328$ 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 $H_0: r = 0.5$ versus $H_1: r \neq 0.5$.

$\widehat{r} = 0.650328, \; s = 0.0198005, \; s_{\widehat{r}} = \dfrac{s}{\sqrt{\sum_{i=1}^n (u_i - \overline{u})^2}} = 0.0322616.$

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 $r \neq 0.5$.

11. (a) $\overline{x} = 41.428157, \overline{y} = 33.428571, \sum_{i=1}^n (x_i - \overline{x})^2 = 6135.428571, \;\; \sum_{i=1}^n (y_i - \overline{y})^2 = 4899.428157,$
$\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = 2451.428571.$

$\widehat{\beta}_1 = \dfrac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = 0.399553$ and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} = 16.875664.$

The least-squares line is $y = 16.875664 + 0.399553x$.

(b) $r^2 = \dfrac{[\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})]^2}{\sum_{i=1}^n (x_i - \overline{x})^2 \sum_{i=1}^n (y_i - \overline{y})^2} = 0.199916, \;\; s = \sqrt{\dfrac{(1 - r^2)\sum_{i=1}^n (y_i - \overline{y})^2}{n - 2}} = 18.073814,$

$\widehat{\beta}_1 = 0.399553, \;\; s_{\widehat{\beta}_1} = \dfrac{s}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.230742.$

There are $14 - 2 = 12$ degrees of freedom. $t = (0.399553 - 0)/0.230742 = 1.732$

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 $t = 1.732$ and to the left of $t = -1.732$.

From the t table, $0.10 < P < 0.20$. A computer package gives $P = 0.1089$.

We cannot conclude that $\beta_1 \neq 0$.

(c) $r^2 = \dfrac{[\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})]^2}{\sum_{i=1}^n (x_i - \overline{x})^2 \sum_{i=1}^n (y_i - \overline{y})^2} = 0.199916, \;\; s = \sqrt{\dfrac{(1 - r^2)\sum_{i=1}^n (y_i - \overline{y})^2}{n - 2}} = 18.073814,$

$\widehat{\beta}_1 = 0.399553, \;\; s_{\widehat{\beta}_1} = \dfrac{s}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.230742.$

There are $14 - 2 = 12$ degrees of freedom. $t = (0.399553 - 1)/0.230742 = -2.602$

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 $t = 2.602$ and to the left of $t = -2.602$.

From the t table, $0.02 < P < 0.05$. A computer package gives $P = 0.0231$.

We can conclude that $\beta_1 \neq 1$.

(d) The data do not provide sufficient evidence to conclude that the linear model is useful for predicting chlorine concentration from pressure, because we cannot conclude that $\beta_1 \neq 0$.

13. (a) $\bar{x} = 225$, $\bar{y} = 86.48333$, $\sum_{i=1}^n (x_i - \bar{x})^2 = 37500$, $\sum_{i=1}^n (y_i - \bar{y})^2 = 513.116667$,
$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 4370$, $n = 12$.

$\widehat{\beta}_1 = \dfrac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = 0.116533$ and $\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x} = 60.263333$.

The estimate of σ^2 is $s^2 = \dfrac{(1 - r^2)\sum_{i=1}^n (y_i - \bar{y})^2}{n - 2}$.

$r^2 = \dfrac{[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})]^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2} = 0.992466$, $s^2 = 0.386600$.

(b) The null and alternate hypotheses are $H_0 : \beta_0 = 0$ versus $H_1 : \beta_0 \neq 0$.

$s = \sqrt{s^2} = 0.621772$. $\widehat{\beta}_0 = 60.263333$, $s_{\widehat{\beta}_0} = s\sqrt{\dfrac{1}{n} + \dfrac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} = 0.744397$.

There are $12 - 2 = 10$ degrees of freedom. $t = (60.263333 - 0)/0.744397 = 80.956$.

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 $t = 80.956$ and to the left of $t = -80.956$.

From the t table, $P < 0.001$. A computer package gives $P = 2.0 \times 10^{-15}$.

We can conclude that $\beta_0 \neq 0$.

(c) The null and alternate hypotheses are $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$.

$s = \sqrt{s^2} = 0.621772$. $\widehat{\beta}_1 = 0.116533$, $s_{\widehat{\beta}_1} = \dfrac{s}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} = 0.0032108$.

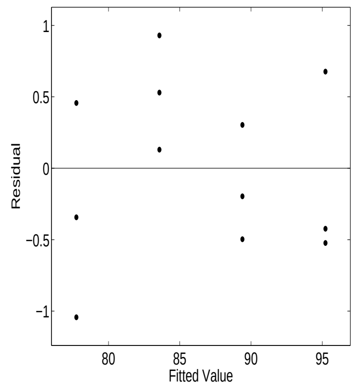
There are $12 - 2 = 10$ degrees of freedom. $t = (0.116533 - 0)/0.0032108 = 36.294$.

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 $t = 36.294$ and to the left of $t = -36.294$.

From the t table, $P < 0.001$. A computer package gives $P = 6.0 \times 10^{-12}$.

We can conclude that $\beta_1 \neq 0$.

(d)



The linear model appears to be appropriate.

(e) $\widehat{\beta}_1 = 0.116533$, $s_{\widehat{\beta}_1} = 0.0032108$.

There are $n-2$ = 10 degrees of freedom. $t_{10,.025} = 2.228$.

Therefore a 95% confidence interval for the slope is $0.116533 \pm 2.228(0.0032108)$, or (0.10938, 0.12369).

(f) $x = 225$, $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1(225) = 86.483333$.

$$s_{\widehat{y}} = s\sqrt{\frac{1}{n} + \frac{(x-\overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.179490.$$

There are 10 degrees of freedom. $t_{10,.025} = 2.228$.

Therefore a 95% confidence interval for the mean response is $86.483333 \pm 2.228(0.179490)$, or (86.083, 86.883).

(g) $x = 225$, $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1(225) = 86.483333$.

$$s_{\text{pred}} = s\sqrt{1 + \frac{1}{n} + \frac{(x-\overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 0.647160.$$

There are 10 degrees of freedom. $t_{10,.025} = 2.228$.

Therefore a 95% prediction interval is $86.483333 \pm 2.228(0.647160)$, or (85.041, 87.925).

15. (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.

17. (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 $\widehat{t} = 145.736 - 0.05180(1) = 145.68$.

19. (a) We need to minimize the sum of squares $S = \sum(y_i - \widehat{\beta}x_i)^2$.

We take the derivative with respect to $\widehat{\beta}$ and set it equal to 0, obtaining $-2\sum x_i(y_i - \widehat{\beta}x_i) = 0$.

Then $\sum x_i y_i - \widehat{\beta}\sum x_i^2 = 0$, so $\widehat{\beta} = \sum x_i y_i / \sum x_i^2$.

(b) Let $c_i = x_i/\sum x_i^2$. Then $\widehat{\beta} = \sum c_i y_i$, so $\sigma^2_{\widehat{\beta}} = \sum c_i^2 \sigma^2 = \sigma^2 \sum x_i^2/(\sum x_i^2)^2 = \sigma^2/\sum x_i^2$.

21. From the answer to Exercise 20, we know that $\sum_{i=1}^n (x_i - \overline{x}) = 0$, $\sum_{i=1}^n \overline{x}(x_i - \overline{x}) = 0$, and $\sum_{i=1}^n x_i(x_i - \overline{x}) = \sum_{i=1}^n (x_i - \overline{x})^2$. Now

$$
\begin{aligned}
\mu_{\widehat{\beta}_0} &= \sum_{i=1}^n \left[\frac{1}{n} - \frac{\overline{x}(x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2}\right]\mu_{y_i} \\
&= \sum_{i=1}^n \left[\frac{1}{n} - \frac{\overline{x}(x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2}\right](\beta_0 + \beta_1 x_i) \\
&= \beta_0 \sum_{i=1}^n \frac{1}{n} + \beta_1 \sum_{i=1}^n \frac{x_i}{n} - \beta_0 \frac{\sum_{i=1}^n \overline{x}(x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2} - \beta_1 \frac{\sum_{i=1}^n x_i \overline{x}(x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2} \\
&= \beta_0 + \beta_1\overline{x} - 0 - \beta_1\overline{x}\frac{\sum_{i=1}^n x_i(x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2} \\
&= \beta_0 + \beta_1\overline{x} - 0 - \beta_1\overline{x} \\
&= \beta_0
\end{aligned}
$$

23.

$$
\begin{aligned}
\sigma^2_{\widehat{\beta}_0} &= \sum_{i=1}^n \left[\frac{1}{n} - \frac{\overline{x}(x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2}\right]^2 \sigma^2 \\
&= \sum_{i=1}^n \left[\frac{1}{n^2} - \frac{2\overline{x}}{n}\frac{(x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2} + \overline{x}^2 \frac{\sum_{i=1}^n (x_i - \overline{x})^2}{[\sum_{i=1}^n (x_i - \overline{x})^2]^2}\right]\sigma^2 \\
&= \left[\sum_{i=1}^n \frac{1}{n^2} - 2\frac{\overline{x}}{n}\frac{\sum_{i=1}^n (x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2} + \overline{x}^2 \frac{\sum_{i=1}^n (x_i - \overline{x})^2}{[\sum_{i=1}^n (x_i - \overline{x})^2]^2}\right]\sigma^2 \\
&= \left[\frac{1}{n} - 2\frac{\overline{x}}{n}(0) + \frac{\overline{x}^2}{\sum_{i=1}^n (x_i - \overline{x})^2}\right]\sigma^2 \\
&= \left[\frac{1}{n} + \frac{\overline{x}^2}{\sum_{i=1}^n (x_i - \overline{x})^2}\right]\sigma^2
\end{aligned}
$$

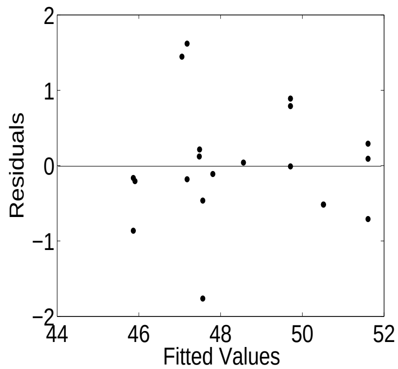
Chapter 8

Section 8.1

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 kg/mm^2.

   (c) By 0.4249(5) = 2.1245 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.

   $F_{9,17} = \dfrac{\text{SSR}/9}{\text{SSE}/17} = \dfrac{(\text{SST} - \text{SSE})/9}{\text{SSE}/17} = 59.204$.

   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) $\text{SSE} = \sum_{i=1}^{6} e_i^2 = 0.52914$, $\text{SST} = \sum_{i=1}^{n} (y_i - \overline{y})^2 = 16.70833$.

(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 $H_0 : \beta_1 = \beta_2 = 0$ 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) $\widehat{\beta}_2 = 2.9205$, $s_{\widehat{\beta}_2} = 0.038261$. 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 8.2

1. (a)

Predictor	Coef	StDev	T	P
Constant	6.3347	2.1740	2.9138	0.009
x_1	1.2915	0.1392	9.2776	0.000

β_0 differs from 0 ($P = 0.009$), β_1 differs from 0 ($P = 0.000$).

(b)

Predictor	Coef	StDev	T	P
Constant	53.964	8.7737	6.1506	0.000
x_2	−0.9192	0.2821	−3.2580	0.004

β_0 differs from 0 ($P = 0.000$), β_1 differs from 0 ($P = 0.004$).

(c)

Predictor	Coef	StDev	T	P
Constant	12.844	7.5139	1.7094	0.104
x_1	1.2029	0.1707	7.0479	0.000
x_2	−0.1682	0.1858	−0.90537	0.377

β_0 may not differ from 0 ($P = 0.104$), β_1 differs from 0 ($P = 0.000$), β_2 may not differ from 0 ($P = 0.377$).

(d) The model in part (a) is the best. When both x_1 and x_2 are in the model, only the coefficient of x_1 is significantly different from 0. In addition, the value of R^2 is only slightly greater (0.819 vs. 0.811) for the model containing both x_1 and x_2 than for the model containing x_1 alone.

3. (a) Plot (i) came from Engineer B, and plot (ii) came from Engineer A. We know this because the variables x_1 and x_2 are both significantly different from 0 for Engineer A but not for Engineer B. Therefore Engineer B is the one who designed the experiment to have the dependent variables nearly collinear.

(b) Engineer A's experiment produced the more reliable results. In Engineer B's experiment, the two dependent variables are nearly collinear.

5. (a) For $R_1 < 4$, the least squares line is $R_2 = 1.233 + 0.264R_1$. For $R_1 \geq 4$, the least squares line is $R_2 = -0.190 + 0.710R_1$.

(b) The relationship is clearly non-linear when $R_1 < 4$.

Quadratic Model

Predictor	Coef	StDev	T	P
Constant	1.2840	0.26454	4.8536	0.000
R_1	0.21661	0.23558	0.91947	0.368
R_1^2	0.0090189	0.044984	0.20049	0.843

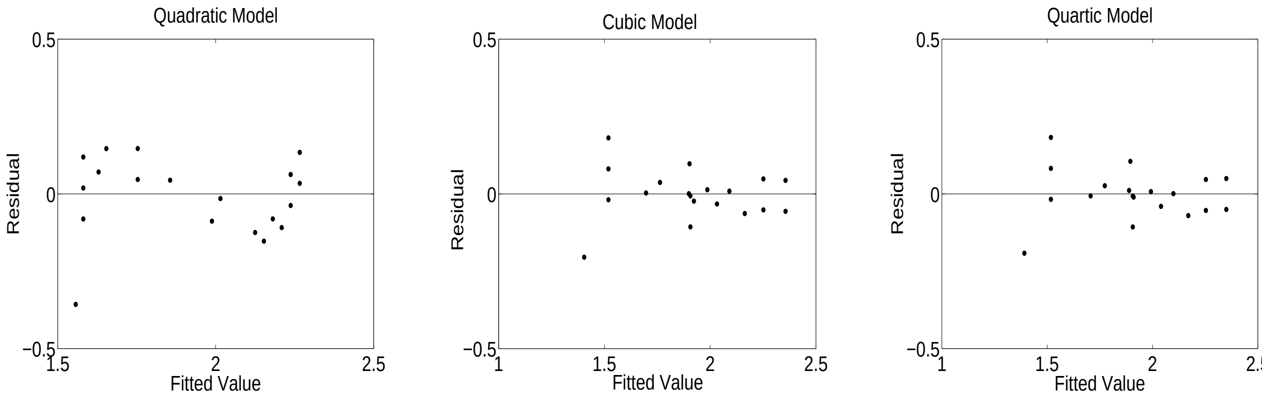
Cubic Model

Predictor	Coef	StDev	T	P
Constant	−1.8396	0.56292	−3.2680	0.004
R_1	4.4987	0.75218	5.9809	0.000
R_1^2	−1.7709	0.30789	−5.7518	0.000
R_1^3	0.22904	0.039454	5.8053	0.000

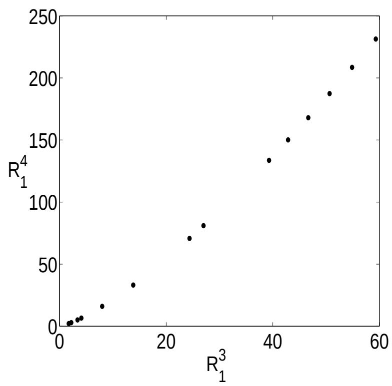
Quartic Model

Predictor	Coef	StDev	T	P
Constant	−2.6714	2.0117	−1.3279	0.200
R_1	6.0208	3.6106	1.6675	0.112
R_1^2	−2.7520	2.2957	−1.1988	0.245
R_1^3	0.49423	0.61599	0.80234	0.432
R_1^4	−0.02558	0.05930	−0.43143	0.671

(c)



(d)



The correlation coefficient between R_1^3 and R_1^4 is 0.997.

(e) R_1^3 and R_1^4 are nearly collinear.

(f) The cubic model is best. The quadratic is inappropriate because the residual plot exhibits a pattern. The residual plots for both the cubic and quartic models look good, however, there is no reason to include R_1^4 in the model since it merely confounds the effect of R_1^3.

Section 8.3

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. (v). x_2^2, x_1x_2, and x_1x_3 all have large P-values and thus may not contribute significantly to the fit.

5. (iv). Carbon and Silicon both have large P-values and thus may not contribute significantly to the fit.

7. 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.

9. (a) $\text{SSE}_{\text{full}} = 7.7302$, $\text{SSE}_{\text{reduced}} = 7.7716$, $n = 165$, $p = 7$, $k = 4$.

$$F = \frac{(\text{SSE}_{\text{reduced}} - \text{SSE}_{\text{full}})/(p-k)}{\text{SSE}_{\text{full}}/(n-p-1)} = 0.2803.$$

(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.

11. $\text{SSE}_{\text{full}} = 9.37$, $\text{SSE}_{\text{reduced}} = 27.49$, $n = 24$, $p = 6$, $k = 3$. The value of the F statistic for testing the plausibility of the reduced model is

$$F = \frac{(\text{SSE}_{\text{reduced}} - \text{SSE}_{\text{full}})/(p-k)}{\text{SSE}_{\text{full}}/(n-p-1)} = 10.958.$$

There are $p - k = 3$ and $n - p - 1 = 17$ degrees of freedom.

From the F table, $P < 0.001$. A computer package gives $P = 0.000304$.

The reduced model is not plausible.

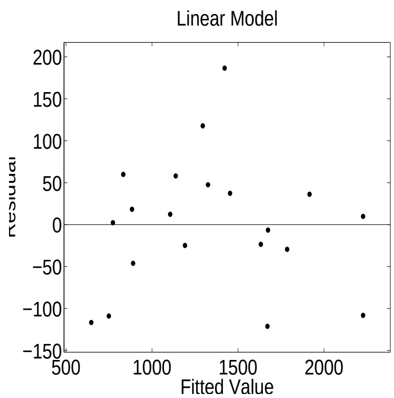
13. (a)

Predictor	Coef	StDev	T	P
Constant	37.989	53.502	0.71004	0.487
x	1.0774	0.041608	25.894	0.000

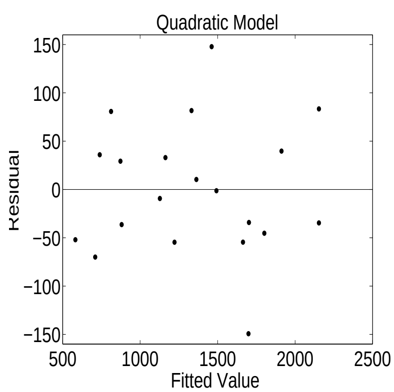
(b)

Predictor	Coef	StDev	T	P
Constant	−253.45	132.93	−1.9067	0.074
x	1.592	0.22215	7.1665	0.000
x^2	−0.00020052	0.000085328	−2.3499	0.031

(c)



(d)



(e) The quadratic model seems more appropriate. The P-value for the quadratic term is fairly small (0.031), and the residual plot for the quadratic model exhibits less pattern. (There are a couple of points somewhat detached from the rest of the plot, however.)

(f) 1683.5

(g) (1634.7, 1732.2)

15. (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.

17. 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) $\hat{y} = 46.802 - 130.11(0.15) - 807.10(0.01) + 3580.5(0.15)(0.01) = 24.6\%$.

(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.

3. (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) $\text{SSR} = \text{SST} - \text{SSE} = 104.09 - 17.28 = 86.81$.

   (g) $\text{MSR} = 86.81/2 = 43.405$.

   (h) $F = \text{MSR}/\text{MSE} = 43.405/1.44 = 30.14$.

   (i) $2 + 12 = 14$.

5. (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

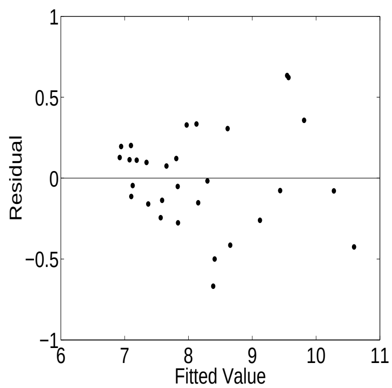
S = 0.33050　　R-sq = 92.0%　　R-sq(adj) = 90.7%

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			

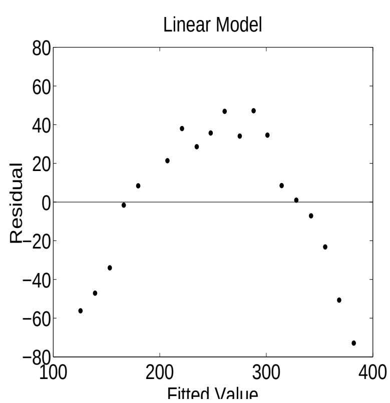
Comparing this model with the one in part (a), $F_{3,24} = \dfrac{(7.8405 - 2.6462)/(5-2)}{2.6462/24} = 15.70$, $P < 0.001$. A computer package gives $P = 7.3 \times 10^{-6}$.

(e)

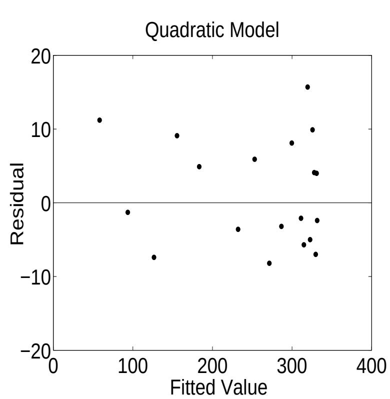
Vars	R-Sq	R-Sq(adj)	C-p	S	Speed	Pause	Speed2	Pause2	Speed*Pause
1	61.5	60.1	92.5	0.68318		X			
1	60.0	58.6	97.0	0.69600					X
2	76.9	75.2	47.1	0.53888		X		X	
2	74.9	73.0	53.3	0.56198	X	X			
3	90.3	89.2	7.9	0.35621	X	X		X	
3	87.8	86.4	15.5	0.39903		X		X	X
4	92.0	90.7	4.8	0.33050	X	X	X	X	
4	90.5	89.0	9.2	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.

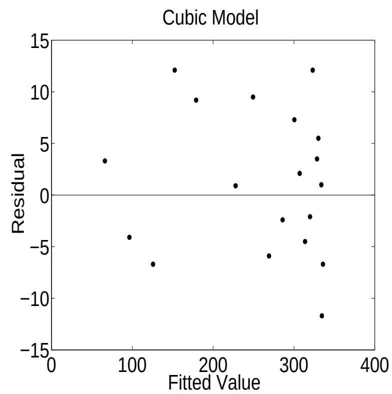
7.



The residual plot shows an obvious curved pattern, so the linear model is not appropriate.



There is no obvious pattern to the residual plot, so the quadratic model appears to fit well.



There is no obvious pattern to the residual plot, so the cubic model appears to fit well.

9. (a) Under model 1, the prediction is $-320.59+0.37820(1500)-0.16047(400)=182.52$.
Under model 2, the prediction is $-380.1+0.41641(1500)-0.5198(150)=166.55$.
Under model 3, the prediction is $-301.8+0.3660(1500)-0.2106(400)+0.164(150)=187.56$.

(b) Under model 1, the prediction is $-320.59+0.37820(1600)-0.16047(300)=236.39$.
Under model 2, the prediction is $-380.1+0.41641(1600)-0.5198(100)=234.18$.
Under model 3, the prediction is $-301.8+0.3660(1600)-0.2106(300)+0.164(100)=237.02$.

(c) Under model 1, the prediction is $-320.59+0.37820(1400)-0.16047(200)=176.80$.
Under model 2, the prediction is $-380.1+0.41641(1400)-0.5198(75)=163.89$.
Under model 3, the prediction is $-301.8+0.3660(1400)-0.2106(200)+0.164(75)=180.78$.

(d) (iv). The output does not provide much to choose from between the two-variable models. In the three-variable model, none of the coefficients are significantly different from 0, even though they were significant in the two-variable models. This suggest collinearity.

11. (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).

13. (a) Let y_1 represent the lifetime of the sponsor's paint, y_2 represent the lifetime of the competitor's paint, x_1 represent January temperature, x_2 represent July temperature, and x_3 represent Precipitation.

Then one good model for y_1 is $y_1 = -4.2342 + 0.79037x_1 + 0.20554x_2 - 0.082363x_3 - 0.0079983x_1x_2 - 0.0018349x_1^2$.

A good model for y_2 is $y_2 = 6.8544 + 0.58898x_1 + 0.054759x_2 - 0.15058x_3 - 0.0046519x_1x_2 + 0.0019029x_1x_3 - 0.0035069x_1^2$.

(b) Substitute $x_1 = 26.1$, $x_2 = 68.9$, and $x_3 = 13.3$ to obtain $\widehat{y}_1 = 13.83$ and $\widehat{y}_2 = 13.90$.

15. (a) Linear Model

Predictor	Coef	StDev	T	P
Constant	0.25317	0.0065217	38.819	0.000
x	-0.041561	0.040281	-1.0318	0.320

(b) Quadratic Model

Predictor	Coef	StDev	T	P
Constant	0.21995	0.0038434	57.23	0.000
x	0.58931	0.06146	9.5886	0.000
x^2	-2.2679	0.2155	-10.524	0.000

(c) Cubic Model

Predictor	Coef	StDev	T	P
Constant	0.22514	0.0068959	32.648	0.000
x	0.41105	0.20576	1.9977	0.069
x^2	-0.74651	1.6887	-0.44206	0.666
x^3	-3.6728	4.043	-0.90843	0.382

(d) Quartic Model

Predictor	Coef	StDev	T	P
Constant	0.23152	0.013498	17.152	0.000
x	0.10911	0.58342	0.18702	0.855
x^2	3.4544	7.7602	0.44515	0.665
x^3	-26.022	40.45	-0.64333	0.533
x^4	40.157	72.293	0.55548	0.590

(e) The quadratic model. The coefficient of x^3 in the cubic model is not significantly different from 0. Neither is the coefficient of x^4 in the quartic model.

(f) $\widehat{y} = 0.21995 + 0.58931(0.12) - 2.2679(0.12^2) = 0.258$

17. (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.

19. (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) $\widehat{\beta}_1 = 0.059718$, $s_{\widehat{\beta}_1} = 0.0088901$.

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 $H_0 : \beta_2 \geq 0$ versus $H_1 : \beta_2 < 0$.

From the output, the test statistic for testing $H_0 : \beta_2 = 0$ versus $H_1 : \beta_2 \neq 0$ 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.

21. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$.

23. (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.)

$$\begin{aligned} z \quad = \quad & 8663.2 - 313.31x_3 - 14.46x_6 + 0.358x_7 - 0.078746x_8 \\ & + 13.998x_9 + 230.24x_{10} - 188.16x_{13} + 5.4133x_{14} + 1928.2x_{15} \\ & - 8.2533x_{16} + 294.94x_{19} + 129.79x_{22} - 3020.7x_{23} \end{aligned}$$

(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.

$$\begin{aligned} w \;=\; & 700.56 - 21.701x_2 - 20.000x_3 + 21.813x_4 + 62.599x_5 + 0.016156x_7 - 0.012689x_8 \\ & + 1.1315x_9 + 15.245x_{10} + 1.1103x_{11} - 20.523x_{13} - 90.189x_{15} - 0.77442x_{16} + 7.5559x_{19} \\ & + 5.9163x_{20} - 7.5497x_{21} + 12.994x_{22} - 271.32x_{23} \end{aligned}$$

(h) The 13-variable model below has Mallows' C_p equal to 8.0.

$$\begin{aligned} w \;=\; & 567.06 - 23.582x_2 - 16.766x_3 + 90.482x_5 + 0.0082274x_7 - 0.011004x_8 + 0.89554x_9 \\ & + 12.131x_{10} - 11.984x_{13} - 0.67302x_{16} + 11.097x_{19} + 4.6448x_{20} + 11.108x_{22} - 217.82x_{23} \end{aligned}$$

(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
Age	5	3.8081	0.76161	7.9115	0.000
Error	73	7.0274	0.096266		
Total	78	10.835			

(b) Yes, $F_{5,73} = 7.9115$, $P < 0.01$ ($P = 5.35 \times 10^{-6}$).

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) From Exercise 13, MSE $= 2.3023$, so $s = \sqrt{2.3023} = 1.517$.

(b) The MINITAB output for the power calculation is

```
Power and Sample Size

One-way ANOVA

Alpha = 0.05  Assumed standard deviation = 1.517  Number of Levels = 4

   SS  Sample  Target                   Maximum
Means    Size   Power  Actual Power  Difference
    2      18     0.9      0.912468           2

The sample size is for each level.
```

(c) The MINITAB output for the power calculation is

```
Power and Sample Size

One-way ANOVA

Alpha = 0.05  Assumed standard deviation = 2.2755  Number of Levels = 4

   SS  Sample  Target                   Maximum
Means    Size   Power  Actual Power  Difference
    2      38     0.9      0.902703           2

The sample size is for each level.
```

17. (a)

Source	DF	SS	MS	F	P
Machine	4	6862	1715.5	7.8825	0.000
Error	30	6529.1	217.64		
Total	34	13391			

(b) (b) Yes, $F_{4,30} = 7.8825$, $P \approx 0$ ($P = 0.00018171$).

19. (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.

(c) $t_{6,.025/15} = 4.698$ (the value obtained by interpolating is 4.958). The value of MSE is 0.00508. The 5% critical value is therefore $4.698\sqrt{2(0.00508)/2} = 0.335$. Any pair that differs by more than 0.335 may 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.

(d) The Tukey-Kramer method is more powerful, since its critical value is smaller (0.284 vs. 0.335).

(e) Either the Bonferroni or the Tukey-Kramer method can be used.

3. (a) MSE $= 2.9659$, $J_i = 12$ for all i. There are 7 comparisons to be made. Now $t_{88,.025/7} = 2.754$, so the 5% critical value is $2.754\sqrt{2.9659(1/12+1/12)} = 1.936$. All the sample means of the non-control formulations differ from the sample mean of the control formulation by more than this amount. Therefore we conclude at the 5% level that all the non-control formulations differ from the control formulation.

(b) There are 7 comparisons to be made. We should use the Studentized range value $q_{7,88,.05}$. This value is not in the table, so we will use $q_{7,60,.05} = 4.31$, which is only slightly larger. The 5% critical value is $4.31\sqrt{2.9659/12} = 2.14$. All of the non-control formulations differ from the sample mean of the control formulation by more than this amount. Therefore we conclude at the 5% level that all the non-control formulations differ from the control formulation.

(c) The Bonferroni method is more powerful, because it is based on the actual number of comparisons being made, which is 7. The Tukey-Kramer method is based on the largest number of comparisons that could be made, which is $(7)(8)/2 = 28$.

5. (a) $t_{16,.025/6} = 3.0083$ (the value obtained by interpolating is 3.080). The value of MSE is 2.3023. The 5% critical value is therefore $3.0083\sqrt{2(2.3023)/5} = 2.8869$. We may conclude that the mean for 750°C differs from the means for 850°C and 900°C, and that the mean for 800° differs from the mean for 900°C.

(b) $q_{4,16,.05} = 4.05$. The value of MSE is 2.3023. The 5% critical value is therefore $4.05\sqrt{2.3023/5} = 2.75$. We may conclude that the mean for 750°C differs from the means for 850°C and 900°C, and that the mean for 800° differs from the mean for 900°C.

(c) The Tukey-Kramer method is more powerful, because its critical value is smaller.

7. (a) $t_{16,.025/3} = 2.6730$ (the value obtained by interpolating is 2.696). The value of MSE is 2.3023. The 5% critical value is therefore $2.6730\sqrt{2(2.3023)/5} = 2.5651$. We may conclude that the mean for 900°C differs from the means for 750°C and 800°C.

(b) $q_{4,16,.05} = 4.05$. The value of MSE is 2.3023. The 5% critical value is therefore $4.05\sqrt{2.3023/5} = 2.75$. We may conclude that the mean for 900°C differs from the means for 750°C and 800°C.

(c) The Bonferroni method is more powerful, because its critical value is smaller.

9. (a) $t_{73,.025} = 1.993$, $MSE = 0.096266$, the s ample sizes are 12 and 15. The sample means are $\overline{X}_1 = 1.6825$, $\overline{X}_6 = 2.0353$. The 95% confidence interval is $0.3528 \pm 1.993\sqrt{0.096266(1/12 + 1/15)}$, or $(0.1133, 0.5923)$.

(b) The sample sizes are $J_1 = 12$, $J_2 = 12$, $J_3 = 13$, $J_4 = 12$, $J_5 = 15$, $J_6 = 15$. $MSE = 0.096266$. We should use the Studentized range value $q_{6,73,.05}$. This value is not in the table, so we will use $q_{6,60,.05} = 4.16$, which is only slightly larger. The values of $q_{6,60,.05}\sqrt{(MSE/2)(1/J_i + 1/J_j)}$ and the values of the differences $|\overline{X}_{i.} - \overline{X}_{j.}|$ are presented in the following two tables.

$q_{6,60,.05}\sqrt{(MSE/2)(1/J_i + 1/J_j)}$

	1	2	3	4	5	6
1	–	0.37260	0.36536	0.37260	0.35348	0.35348
2	0.37260	–	0.36536	0.37260	0.35348	0.35348
3	0.36536	0.36536	–	0.36536	0.34584	0.34584
4	0.37260	0.37260	0.36536	–	0.35348	0.35348
5	0.35348	0.35348	0.34584	0.35348	–	0.33326
6	0.35348	0.35348	0.34584	0.35348	0.33326	–

$|\overline{X}_{i.} - \overline{X}_{j.}|$

	1	2	3	4	5	6
1	0	0.0075	0.49904	0.15083	0.5475	0.35283
2	0.0075	0	0.49154	0.14333	0.54	0.34533
3	0.49904	0.49154	0	0.34821	0.048462	0.14621
4	0.15083	0.14333	0.34821	0	0.39667	0.202
5	0.5475	0.54	0.048462	0.39667	0	0.19467
6	0.35283	0.34533	0.14621	0.202	0.19467	0

The differences that are significant at the 5% level are: mean 1 differs from means 3 and 5; mean 2 differs from means 3 and 5; and mean 4 differs from mean 5.

11. (a) $t_{8,.025} = 2.306$, $\text{MSE} = 1.3718$. The sample means are $\overline{X}_1 = 1.998$ and $\overline{X}_3 = 5.300$. The sample sizes are $J_1 = 5$ and $J_3 = 3$. The 95% confidence interval is therefore $3.302 \pm 2.306\sqrt{1.3718(1/5+1/3)}$, or $(1.330, 5.274)$.

(b) 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.

13. (a) $\overline{X}_{..} = 88.04$, $I = 4$, $J = 5$, $\text{MSTr} = \sum_{i=1}^{I} J(\overline{X}_{i.} - \overline{X}_{..})^2/(I-1) = 19.554$.

$F = \text{MSTr}/\text{MSE} = 19.554/3.85 = 5.08$. 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.

15. 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)

Main Effects of Solution

Solution	
NaCl	−9.1148
Na_2HPO_4	9.1148

Main Effects of Temperature

Temperature	
25°C	1.8101
37°C	−1.8101

Interactions

	Temperature	
Solution	25°C	37°C
NaCl	−0.49983	0.49983
Na_2HPO_4	0.49983	−0.49983

(b)

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			

(c) Yes, $F_{1,20} = 0.015632$, $P > 0.10$ ($P = 0.902$).

(d) Yes, since the additive model is plausible. The mean yield stress differs between Na_2HPO_4 and NaCl: $F_{1,20} = 5.1983$, $0.01 < P < 0.05$ ($P = 0.034$).

(e) 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	10.121	0.008
Pressure	2	35.663	17.832	10.607	0.002
Interaction	2	4.3544	2.1772	1.2951	0.310
Error	12	20.173	1.6811		
Total	17	77.205			

(b) Yes. $F_{2,12} = 1.2951$, $P > 0.10$ ($P = 0.310$).

(c) Yes, since the additive model is plausible. The mean strength differs between the two adhesives: $F_{1,12} = 10.121$, $P < 0.01$ ($P = 0.008$).

(d) Yes, since the additive model is plausible. The mean strength differs among the pressures: $F_{2,12} = 10.607$, $P < 0.01$ ($P = 0.002$).

9. (a)

Main Effects of Speed

80	−13.074
120	−5.7593
150	19.463

Main Effects of Time

5	−8.4259
10	−0.2037
15	8.6296

Interactions

	Time		
Speed	5	10	15
80	5.6481	0.75926	−6.4074
120	2.3704	−0.018519	−2.3519
150	−8.0185	−0.74074	8.7593

(b)

Source	DF	SS	MS	F	P
Speed	2	10796	5397.9	63.649	0.000
Time	2	2619.1	1309.6	15.442	0.000
Interaction	4	1357.5	339.38	4.0018	0.007
Error	45	3816.3	84.807		
Total	53	18589			

(c) No, $F_{4,45} = 4.0018$, $P < 0.01$ ($P = 0.007$)

(d) No, because the additive model is rejected.

(e) No, because the additive model is rejected.

11. (a)

Main Effects of Material

CPTi-ZrO_2	0.044367
TiAlloy-ZrO_2	−0.044367

Main Effects of Neck Length

Short	−0.018533
Medium	−0.024833
Long	0.043367

Interactions

	Neck Length		
	Short	Medium	Long
CPTi-ZrO_2	0.0063333	−0.023767	0.017433
TiAlloy-ZrO_2	0.0063333	0.023767	−0.017433

(b)

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			

(c) 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 $P > 0.10$ ($P = 0.184$).

(d) Yes, since the additive model is plausible. The mean coefficient of friction differs between CPTi-ZrO_2 and TiAlloy-ZrO_2: $F_{1,24} = 23.630$, $P < 0.001$.

(e) Yes, since the additive model is plausible. The mean coefficient of friction is not the same for all neck lengths: $F_{2,24} = 5.6840$, $P \approx 0.01$. To determine which pairs of effects differ, we use $q_{3,24,.05} = 3.53$. We compute $3.53\sqrt{0.002499/10} = 0.056$. We conclude that the effect of long neck length differs from both short and medium lengths, but we cannot conclude that the effects of short and medium lengths differ from each other.

13. (a)

Main Effects of Concentration

15	0.16667
40	−0.067778
100	−0.098889

Main Effects of Delivery Ratio

1:1	0.73333
1:5	−0.30000
1:10	−0.43333

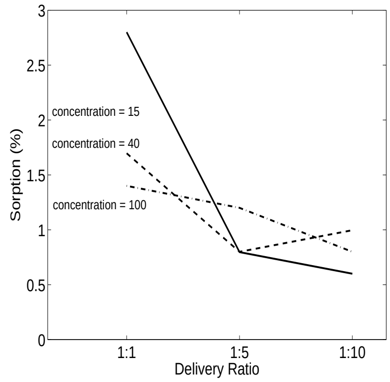
Interactions

	Delivery Ratio		
Concentration	1:1	1:5	1:10
15	0.66778	−0.30222	−0.36556
40	−0.20111	−0.064444	0.26556
100	−0.46667	0.36667	0.10000

(b)

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			

(c) No. The The value of the test statistic is 17.587, its null distribution is $F_{4,18}$, and $P \approx 0$.

(d) 

The slopes of the line segments are quite different from one another, indicating a high degree of interaction.

15. (a)

Main Effects of Attachment	
Nail	−1.3832
Adhesive	1.3832

Main Effects of Length	
Quarter	−7.1165
Half	−2.5665
Full	9.683

Interactions

	Length: Quarter	Half	Full
Nail	0.48317	0.33167	−0.51633
Adhesive	−0.48317	−0.33167	0.51633

(b)

Source	DF	SS	MS	F	P
Attachment	1	114.79	114.79	57.773	0.000
Length	2	3019.8	1509.9	759.94	0.000
Interaction	2	10.023	5.0115	2.5223	0.090
Error	54	107.29	1.9869		
Total	59	3251.9			

(c) The additive model is barely plausible: $F_{2,54} = 2.5223$, $0.05 < P < 0.10$ ($P = 0.090$).

(d) Yes, the attachment method does affect the critical buckling load: $F_{1,54} = 57.773$, $P \approx 0$.

(e) Yes, the side member length does affect the critical buckling load: $F_{2,54} = 759.94$, $P \approx 0$. To determine which effects differ at the 5% level, we should use $q_{3,54,.05}$. This value is not found in Table A.8, so we approximate it with $q_{3,40,.05} = 3.44$. We compute $3.44\sqrt{1.9869/20} = 1.08$. We conclude that the effects of quarter, half and full all differ from each other.

17. (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. $F_{2,9} = 13.53$, $0.01 < P < 0.001$ ($P = 0.002$).

19. (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) NaOH concentration is the blocking factor, age is the treatment factor.

(b)

Source	DF	SS	MS	F	P
Treatment	3	386.33	128.78	211.14	0.000
Blocks	4	13.953	3.4882	5.7192	0.008
Error	12	7.3190	0.6099		
Total	19	407.60			

(c) Yes, $F_{3,12} = 211.14$, $P \approx 0$ ($P = 1.202 \times 10^{-10}$)

(d) $q_{4,12,.05} = 4.20$, MSAB = 0.6099, and $J = 5$. The 5% critical value is therefore $4.20\sqrt{0.6099/5} = 1.4669$. The sample means are $\overline{X}_0 = 55.46$, $\overline{X}_4 = 45.22$, $\overline{X}_8 = 46.26$, and $\overline{X}_{12} = 44.47$. We therefore conclude that age 0 differs from ages 4, 8, and 12, and that age 8 differs from age 12.

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)

Source	DF	SS	MS	F	P
Waterworks	3	1253.5	417.84	4.8953	0.005
Block	14	1006.1	71.864	0.84193	0.622
Error	42	3585.0	85.356		
Total	59	5844.6			

(b) Yes, $F_{3,42} = 4.8953$, $P = 0.005$.

(c) To determine which effects differ at the 5% level, we should use $q_{4,42,.05}$. This value is not found in Table A.8, so we approximate it with $q_{4,40,.05} = 3.79$. The 5% critical value is $3.79\sqrt{85.356/15} = 9.04$. The sample means are $\overline{X}_A = 34.000$, $\overline{X}_B = 22.933$, $\overline{X}_C = 24.800$, $\overline{X}_D = 31.467$. We can conclude that A differs from both B and C.

(d) The *P*-value for the blocking factor is large (0.622), suggesting that the blocking factor (time) has only a small effect on the outcome. It might therefore be reasonable to ignore the blocking factor and perform a one-way ANOVA.

9. (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)

Term	Effect
A	−119.25
B	259.25
C	−82.75
AB	101.75
AC	−6.25
BC	−52.75
ABC	−2.25

(b) No, since the design is unreplicated, there is no error sum of squares.

(c) The AB interaction is among the larger effects.

(d) No, because the AB interaction is large.

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) From Exercise 11, MSE $= 29.026$, so $s = \sqrt{29.026} = 5.388$.

(b) The MINITAB output for the power calculation is

```
Power and Sample Size

One-way ANOVA

Alpha = 0.05  Assumed standard deviation = 5.388  Number of Levels = 5

   SS  Sample  Target                    Maximum
Means    Size   Power  Actual Power   Difference
   50      10     0.9      0.901970           10

The sample size is for each level.
```

(c) The MINITAB output for the power calculation is

```
Power and Sample Size

One-way ANOVA

Alpha = 0.05  Assumed standard deviation = 8.082  Number of Levels = 5

   SS  Sample  Target                    Maximum
Means    Size   Power  Actual Power   Difference
   50      22     0.9      0.913650           10

The sample size is for each level.
```

17. (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.

19.

Source	DF	SS	MS	F	P
Location	2	3.0287	1.5144	9.4427	0.000
Error	107	17.16	0.16037		
Total	109	20.189			

Yes, $F_{2,107} = 9.4427$, $P \approx 0$.

21. (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, $F_{2,9} = 749.53$, $P \approx 0.000$.

23.

Source	DF	SS	MS	F	P
Angle	6	936.4	156.07	20.302	0.000
Error	39	299.81	7.6874		
Total	45	1236.2			

Yes, $F_{6,39} = 20.302$, $P \approx 0$.

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 $n = 200$. $\overline{p} = 1.64/30 = 0.054667$.
   The centerline is $\overline{p} = 0.054667$.
   The LCL is $\overline{p} - 3\sqrt{\overline{p}(1-\overline{p})/200} = 0.00644$.
   The UCL is $\overline{p} + 3\sqrt{\overline{p}(1-\overline{p})/200} = 0.1029$.

3. Yes, the only information needed to compute the control limits is $\overline{p}$ and the sample size n. In this case, $n = 100$, and $\overline{p} = (622/50)/100 = 0.1244$.
   The control limits are $\overline{p} \pm 3\sqrt{\overline{p}(1-\overline{p})/n}$, so
   LCL = 0.0254 and UCL = 0.2234.

5. (iv). The sample size must be large enough so the mean number of defectives per sample is at least 10.

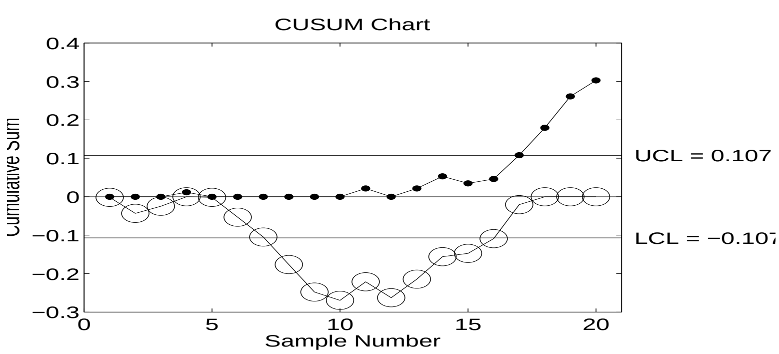
7. It was out of control. The UCL is 23.13.

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)



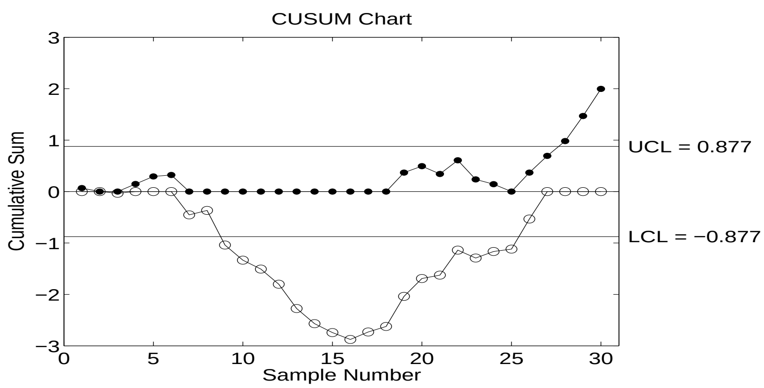
(d) The process is out of control on sample 8.

(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 $\sigma_{\overline{X}} = (0.577)(1.14)/3 = 0.219$.

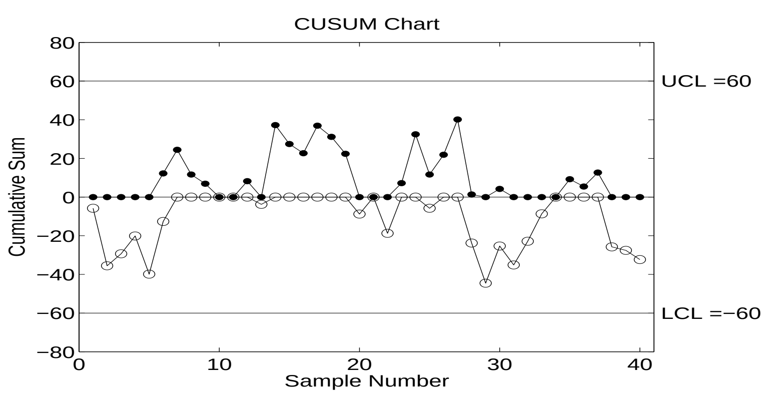
(c)



(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) $\widehat{\mu} = \overline{\overline{X}} = 0.248$, $\overline{s} = 0.002$, $LSL = 0.246$, $USL = 0.254$. The sample size is $n = 6$.

   $\widehat{\sigma} = \overline{s}/c_4$. From the control chart table, $c_4 = 0.9515$.

   Therefore $\widehat{\sigma} = 0.002102$.

   Since $\widehat{\mu}$ is closer to LSL than to USL, $C_{pk} = (\widehat{\mu} - LSL)/(3\widehat{\sigma}) = 0.3172$.

   (b) No. Since $C_{pk} < 1$, the process capability is not 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 15.50.

   (b) $LSL = 15.40$, $USL = 15.60$, $\widehat{\sigma} = 0.017562$.

   If $\mu = 15.50$, then $C_{pk} = (15.60 - 15.50)/[3(0.017562)] = 1.898$.

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 $n = 300$. $\overline{p} = 5.83/100 = 0.0583$.

   The centerline is $\overline{p} = 0.0583$

   The LCL is $\overline{p} - 3\sqrt{\overline{p}(1 - \overline{p})/300} = 0.0177$.

   The UCL is $\overline{p} + 3\sqrt{\overline{p}(1 - \overline{p})/300} = 0.0989$.

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.

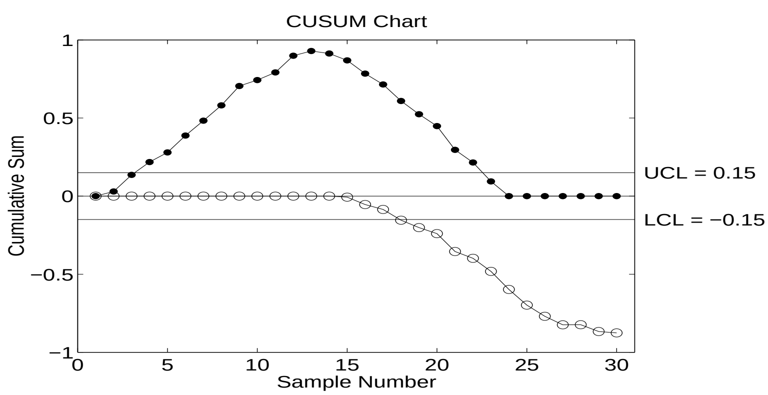
(b) The upper and lower limits for the $\overline{X}$-chart are $\overline{\overline{X}} - A_2\overline{R}$ and $\overline{\overline{X}} + A_2\overline{R}$, respectively.
From the control chart table, $A_2 = 1.023$. $\overline{R} = 0.110$ and $\overline{\overline{X}} = 5.095$.
Therefore LCL = 4.982 and 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 $n = 300$.
The mean number of defectives over the last 20 days is 12.0.

Therefore $\overline{p} = 12.0/300 = 0.04$.
The control limits are $\overline{p} \pm 3\sqrt{\overline{p}(1-\overline{p})/n}$.
Therefore LCL = 0.00606 and UCL = 0.0739.

(b) Sample 7. The proportion of defective chips is then $1/300 = 0.0033$, which is less than the lower control limit.

(c) No, this special cause improves the process. It should be preserved rather than eliminated.

Appendix B

1. $\dfrac{\partial v}{\partial x} = 3 + 2y^4, \qquad \dfrac{\partial v}{\partial y} = 8xy^3$

2. $\dfrac{\partial w}{\partial x} = \dfrac{3x^2}{x^2+y^2} - \dfrac{2x(x^3+y^3)}{(x^2+y^2)^2}, \qquad \dfrac{\partial w}{\partial y} = \dfrac{3y^2}{x^2+y^2} - \dfrac{2y(x^3+y^3)}{(x^2+y^2)^2}$

3. $\dfrac{\partial z}{\partial x} = -\sin x \sin y^2, \qquad \dfrac{\partial z}{\partial y} = 2y \cos x \cos y^2$

4. $\dfrac{\partial v}{\partial x} = ye^{xy}, \qquad \dfrac{\partial v}{\partial y} = xe^{xy}$

5. $\dfrac{\partial v}{\partial x} = e^x(\cos y + \sin z), \qquad \dfrac{\partial v}{\partial y} = -e^x \sin y, \qquad \dfrac{\partial v}{\partial z} = e^x \cos z$

6. $\dfrac{\partial w}{\partial x} = \dfrac{x}{\sqrt{x^2+4y^2+3z^2}}, \qquad \dfrac{\partial w}{\partial y} = \dfrac{4y}{\sqrt{x^2+4y^2+3z^2}}, \qquad \dfrac{\partial w}{\partial z} = \dfrac{3z}{\sqrt{x^2+4y^2+3z^2}}$

7. $\dfrac{\partial z}{\partial x} = \dfrac{2x}{x^2+y^2}, \qquad \dfrac{\partial z}{\partial y} = \dfrac{2y}{x^2+y^2}$

8. $\dfrac{\partial v}{\partial x} = \dfrac{2xy}{x^2y+z} - ze^{y^2}\sin(xz), \qquad \dfrac{\partial v}{\partial y} = \dfrac{x^2}{x^2y+z} + 2yze^{y^2}\cos(xz), \qquad \dfrac{\partial v}{\partial z} = \dfrac{1}{x^2y+z} - xe^{y^2}\sin(xz)$

9. $\dfrac{\partial v}{\partial x} = \sqrt{\dfrac{y^5}{x}} - \dfrac{3}{2}\sqrt{\dfrac{y^3}{x}}, \qquad \dfrac{\partial v}{\partial y} = 5\sqrt{xy^3} - \dfrac{9}{2}\sqrt{xy}$

10. $\dfrac{\partial z}{\partial x} = \dfrac{xy\cos(x^2y)}{\sqrt{\sin(x^2y)}}, \qquad \dfrac{\partial z}{\partial y} = \dfrac{x^2\cos(x^2y)}{2\sqrt{\sin(x^2y)}}$