Chapter 1

Basics of Probability

1.1 Introduction

1.2 Basic Concepts and Definitions

1.2.1 Sample space = $\{TTT, TTH, THT, HTT, THH, HTH, HHT, HHH\}$,

$A = \{TTT, TTH, THT, HTT, THH, HTH, HHT\}$,

$P(A) = 7/8$

1.2.2 The sample space is given in the table below:

RRRR	RRRL	RRLR	RLRR	LRRR
RRLL	RLRL	RLLR	LLRR	LRRL
LRLR	RLLL	LRLL	LLRL	LLLR
LLLL				

From the sample space, we see there are 6 outcomes where exactly 2 cars turn left. Assuming each outcome is equally likely, the probability that exactly 2 cars turn left is $6/16 = 3/8$. This assumption is probably not very reasonable since at a given intersection, cars typically turn one direction more often than the other.

1.2.3

a. Since there are 365 days in a year and exactly one day is the person's birthday, the probability of guessing correctly is 1/365.
b. By similar reasoning as part a., the probability of guessing the correct month is 1/12.

1.2.4

a. The sample space is given in the table below:

	Die					
Coin	1	2	3	4	5	6
T	1T	2T	3T	4T	5T	6T
H	1H	2H	3H	4H	5H	6H

b. The event of getting a tail and at least a 5 is $A = \{5T, 6T\}$ and $P(A) = 2/12 = 1/6$

1.2.5

a. $P(X \leq 3) = \frac{3-(-2)}{5-(-2)} = \frac{5}{7}$
b. $P(X \geq -1.5) = \frac{3-(-1.5)}{5-(-2)} = \frac{6.5}{7}$
c. $P(-1 \leq X \leq 4) = \frac{4-(-1)}{5-(-2)} = \frac{5}{7}$
d. $P(0 \leq X \leq 6) = \frac{5-0}{5-(-2)} = \frac{5}{7}$
e. $P(X \geq 4) = \frac{5-4}{5-(-2)} = \frac{1}{7}$
f. $P(X = -0.75) = 0$

1.2.6

a. The sample space is given in the table below:

	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36

b. The distribution of the product is given in the table below:

x	1	2	3	4	5	6	8	9	10	12	15	16	18	20	24	25	30	36
# of Outcomes	1	2	2	3	2	4	2	1	2	4	2	1	2	2	2	1	2	1
$P(x)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{9}$	$\frac{1}{18}$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{9}$	$\frac{1}{18}$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{36}$

c. $P(5 \leq X \leq 25) = P(X = 5) + \cdots + P(X = 25) = \frac{1}{18} + \cdots + \frac{1}{36} = \frac{25}{36}$

1.2.7

a. $P(B) = \frac{15}{39} \approx 0.385$, theoretical approach
b. $P(B) = \frac{8}{20} \approx 0.40$, relative frequency approach
c. In part a., the sample space consists of the 39 students in this class. Out of these students, exactly 15 got a B. Thus the probability in part a. is an exact value. In part b., the sample space consists of the 115 students. Selecting 20 students from this group is 20 trials of the random experiment of selecting a student from this group. So the probability in part b. is an estimate.
d. Neither group of students in parts a. and b. is a good representation of all statistics students in the U.S., so neither estimate would apply to all students from the U.S.

1.2.8 a. unusual, b. not unusual, c. not unusual, d. unusual, e. not unusual

1.2.9 Since there are many wrong answers and only one right answer, the probability of guessing the right answer is much smaller than that of guessing a wrong one.

1.2.10 The probability of 0.10 is a theoretical probability and therefore describes an "average in the long run." It does not tell us exactly what will happen on any one trial, or any set of 10 trials.

1.2.11 The probability of hitting a ring is the area of the ring divided by the total area of the dartboard.

$$
\begin{array}{lll}
\text{Total area} = \pi(9^2) = 81\pi & & \\
\text{Area of bullseye} = \pi(1^2) = \pi & \Rightarrow & P(\text{bullseye}) = \frac{\pi}{81\pi} \approx 0.0123 \\
\text{Area of ring 1} = \pi(3^2 - 1^2) = 8\pi & \Rightarrow & P(\text{Hitting ring 1}) = \frac{8\pi}{81\pi} \approx 0.0988 \\
\text{Area of ring 2} = \pi(5^2 - 3^2) = 16\pi & \Rightarrow & P(\text{Hitting ring 2}) = \frac{16\pi}{81\pi} \approx 0.1975 \\
\text{Area of ring 3} = \pi(7^2 - 5^2) = 24\pi & \Rightarrow & P(\text{Hitting ring 3}) = \frac{24\pi}{81\pi} \approx .2963 \\
\text{Area of ring 4} = \pi(9^2 - 7^2) = 32\pi & \Rightarrow & P(\text{Hitting ring 4}) = \frac{32\pi}{81\pi} \approx 0.395
\end{array}
$$

1.3 Counting Problems

1.3.1

a. $26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 = 17,576,000$
b. $26 \cdot 25 \cdot 24 \cdot 10 \cdot 9 \cdot 8 = 11,232,000$
c. $26 \cdot 25 \cdot 26 \cdot 9 \cdot 10 \cdot 10 = 15,210,000$

1.3.2

a. $62 \cdot 62 \cdot 62 \cdot 62 = 14,776,336$

b. $\dfrac{1}{14,776,336} = 6.77 \times 10^{-8}$

c. $\dfrac{1}{62} = 0.0161$

1.3.3

a. The customer has a yes/no choice for each of the mayo and mustard. Since there are 4 types and bread and 5 types of meat, there are $4 \cdot 5 \cdot 2 \cdot 2 = 80$ different possibilities.

b. $\binom{6}{3} = 20$

c. Since the customer has a choice of yes or no for each of the 6 vegetables, there $2^6 = 64$ different choices.

d. $4 \cdot 5 \cdot 2^6 \cdot 2 \cdot 2 = 5120$

e. Let n denote the number of sauces. Then the total number of choices would be $5120 \cdot 2^n$. For this to be at least 40,000, we need $40,000/5120 \approx 7.8 \leq 2^n$. But $2^3 = 8$, so the owner need at least 3 sauces.

1.3.4 There are 2 sub-choices: 1. the arrangement of the 3 boys, and 2. the arrangement of the 2 girls. There are 3! possibilities for the first and 2! for the second. Thus the total number of possibilities is $3! \cdot 2! = 12$

1.3.5 If the awards are all different, then the order in which they are given matters, so the total number ways they can be awarded is

$$_{15}P_3 = \frac{15!}{12!} = 2730.$$

If the awards are all the same, then the order does not matter, so the total number ways they can be awarded is

$$\binom{15}{3} = 455.$$

1.3.6 $3! \cdot 4! \cdot 5! \cdot 3! \cdot = 103,680$

1.3.7

a. As an arrangement problem where some items are identical, we have a total of 8 flags, 5 are identical, and 3 are identical. So the total number of different arrangements is $\frac{8!}{5! \cdot 3!} = 56$.

b. Another way to think of this problem is that we need to choose the positions of the 3 green flags from the total of 8 positions on the pole. The number of ways to do this is $\binom{8}{3} = 56$.

1.3.8 We can think of the arrivals as an arrangement of 5 customers, 2 of them are identical, and another 2 are identical. The total number of arrangements is then

$$\frac{5!}{2! \cdot 1! \cdot 2!} = 30.$$

1.3.9 There are 3 sub-choices: 1. The selection of 2 of 3 math faculty, 2. the selection of 3 of 5 theology faculty, and 3. The selection of 1 of 3 chemistry faculty. The number of ways this can be done is

$$\binom{3}{2}\cdot\binom{5}{3}\cdot\binom{3}{1} = 90.$$

1.3.10

a. $\binom{9}{7} = 36$
b. $\binom{6}{4} = 15$
c. If he answers all 3 of the first 3, then he must choose 4 of the last 6 questions. The number of ways to do this is $\binom{6}{4} = 15$. If he chooses 2 of the first 3, he then has to choose 5 of the last 6. The total number of ways this can be donw i

$$\binom{3}{2}\cdot\binom{6}{5} = 18.$$

Overall he has $15 + 18 = 33$ choices.

1.3.11

$$\begin{aligned}(x+3)^9 &= \sum_{k=0}^{9}\binom{9}{k}x^{9-k}3^k \\ &= \binom{9}{0}x^{9-0}3^0 + \binom{9}{1}x^{9-1}3^1 + \binom{9}{2}x^{9-2}3^2 + \cdots + \binom{9}{9}x^{9-9}3^9 \\ &= x^9 + 27x^8 + 324x^7 + 2,268x^6 + 10,206x^5 + 30,618x^4 \\ &\quad + 61,236x^3 + 78,732x^2 + 59,049x + 19,683\end{aligned}$$

1.3.12 $\binom{21}{7}x^{21-7}y^7 = 116,280x^{14}y^7$

1.3.13 There are n sub-choices, each of which has 2 choices, so the total number of choices is

$$\underbrace{2\cdot 2\cdots 2}_{n \text{ times}} = 2^n.$$

1.3.14 $\binom{10}{5} + \binom{10}{6} = 462$

1.3.15

a. $n(\text{Four of a kind}) = \binom{13}{1}\cdot 48 = 624$

$P(\text{Four of a kind}) = \dfrac{624}{\binom{52}{5}} = 2.4\times 10^{-4}$

b. $n(\text{Full house}) = \binom{13}{1}\cdot\binom{4}{3}\cdot\binom{12}{1}\cdot\binom{4}{2} = 3,744$

$P(\text{Full house}) = \dfrac{3,744}{\binom{52}{5}} = 0.00144$

c. $n(\text{Three of a kind}) = \binom{13}{1} \cdot \binom{4}{3} \cdot \binom{12}{2} \cdot \binom{4}{1} \cdot \binom{4}{1} = 54,912$

$P(\text{Three of a kind}) = \dfrac{54,912}{\binom{52}{5}} = 0.0211$

1.3.16 $n(3 \text{ red}) = \binom{15}{3} \cdot \binom{10}{5} = 114,660 \quad P(3 \text{ red}) = \dfrac{114,660}{\binom{25}{8}} = 0.106$

1.3.17

a. $\dfrac{5 \cdot 6 \cdot 3 \cdot 2}{\binom{16}{4}} \approx 0.0989$

b. $\dfrac{\binom{6}{4}}{\binom{16}{4}} \approx 0.00824$

c. $\dfrac{\binom{5}{4}}{\binom{16}{4}} \approx 0.00275$

d. $\dfrac{\binom{5}{3} \cdot \binom{3}{1}}{\binom{16}{4}} \approx 0.0165$

1.3.18

a. $5 \cdot 7 = 35$

b. $5 \cdot 7 \cdot 7 \cdot 5 = 1,125$

c. $5 \cdot 7 \cdot 6 \cdot 4 = 840$

1.3.19 In each of these parts, we can think of making a number by simply rearranging the digits.

a. There are 4 distinct digits, so the number of arrangements is $4! = 24$.

b. There are 4 digits, 2 of which are identical, so the number of arrangements is $\dfrac{4!}{2!} = 12$.

c. There are 4 digits, 3 of which are identical, so the number of arrangements is $\dfrac{4!}{3!} = 4$.

1.3.20 $\binom{21}{4} \cdot \binom{5}{2} \cdot 6! = 43,092,000$

1.3.21

a. $n(\text{Total \# of ways}) = 10 \cdot 10 \cdot 10 = 1000$

b. The number of ways to choose the numbers 8, 4, and 6 in any order is $3! = 6$. Therefore, $P(\text{winning}) = \dfrac{6}{1000} = 0.006$.

c. The number of ways to choose the numbers 1, 1, and 5 in any order is $\dfrac{3!}{2!1!} = 3$. Therefore, $P(\text{winning}) = \dfrac{3}{1000} = 0.003$.

d. There is only 1 way to choose the numbers 9, 9, and 9. Therefore, $P(\text{winning}) = \dfrac{1}{1000} = 0.001$.

e. To maximize the probability of winning, three different numbers should be chosen. The actual values of the numbers do not matter because the probability of choosing three different numbers will be the same for any three numbers.

1.3.22 Since there are 4 types of hamburgers, there are 4 different ways in which all 3 customers can order the same type. In general, there are $4 \cdot 4 \cdot 4 = 64$ ways of the 3 customers choosing their hamburgers. Assuming the choices are made independently and that each customer is equally likely to choose each type,

$$P(\text{All 3 order the same}) = \frac{4}{64} = \frac{1}{16} = 0.0625.$$

1.3.23 Note:

1. There are 2 choices for the person seated on the left.
2. There are $(n-1)$ choices for the chair of the person seated on the left.
3. There are $(n-2)!$ ways of arranging the other $(n-2)$ students.
4. There are in general $n!$ ways of arranging all n students.

Since the students are randomly arranged,

$$P(\text{Katie and Cory sit next to each other}) = \frac{2(n-1)(n-2)!}{n!} = \frac{2}{n}.$$

1.3.24

a. We can think of this as though we are placing the r red cubes in the positions between the blue cubes and that not more than 1 red cube can go in any one of these positions. Since there are $n-1$ positions between the blue cubes, there are $\binom{n-1}{r}$ ways this can be done.
b. When the r red cubes are placed between the blue cubes as in part a., the blue cubes are divided into $r+1$ subgroups, each containing at least 1 blue cube. We can think of each of these subgroups as the cubes being placed into one of the $r+1$ empty bins. So the number of ways this can be done is the same number as in part a, $\binom{n-1}{r}$.

1.3.25 Since there are 10 choices for each of the 5 decimal places, plus the possibility of choosing the number 1, the size of the sample space is

$$n(S) = (10 \cdot 10 \cdot 10 \cdot 10 \cdot 10) + 1 = 100,001.$$

Therefore, the probability of selecting any one of these numbers is

$$P(X = x) = \frac{1}{100,001} = 9.9999 \times 10^{-6}.$$

1.3.26 By the definition of combinations,

$$\binom{n_1+n_2}{n_1} = \frac{(n_1+n_2)!}{n_1!\,(n_1+n_2-n_1)!} = \frac{(n_1+n_2)!}{n_1!\,n_2!} \quad \text{and}$$
$$\binom{n_1+n_2}{n_2} = \frac{(n_1+n_2)!}{n_2!\,(n_1+n_2-n_2)!} = \frac{(n_1+n_2)!}{n_2!\,n_1!}.$$

Thus these two quantities are equal.

1.3.27 By the definition of combinations,

$$\begin{aligned}\binom{n-1}{r}+\binom{n-1}{r-1} &= \frac{(n-1)!}{r!(n-1-r)!}+\frac{(n-1)!}{(r-1)!(n-1-(r-1))!}\\ &= \frac{(n-1)!}{r!(n-1-r)!}+\frac{(n-1)!}{(r-1)!(n-r)!}\\ &= \frac{(n-1)!}{r!(n-r-1)!}\cdot\frac{n-r}{n-r}+\frac{(n-1)!}{(r-1)!(n-r)!}\cdot\frac{r}{r}\\ &= \frac{(n-r)(n-1)!}{r!(n-r)!}+\frac{r(n-1)!}{r!(n-r)!}\\ &= \frac{((n-r)+r)(n-1)!}{r!(n-r)!}\\ &= \frac{n(n-1)!}{r!(n-r)!}=\frac{n!}{r!(n-r)!}=\binom{n}{r}.\end{aligned}$$

1.4 Axioms of Probability and the Addition Rule

1.4.1

a. Not disjoint, both alarms could fail in the same trial.
b. Not disjoint, both women could say no.
c. Disjoint, the student's grade cannot be both a C or better *and* a D or F.
d. Not disjoint, the son could play on both the swings *and* the slide during one trip to the park.
e. Disjoint, assuming the flower is a solid color, the flower cannot be both red *and* blue. If we assume the flower could have more than one color, then the events are not disjoint.
f. Disjoint, if the truck has a damaged tail light, it will not be free of defects.

1.4.2 To determine $P(R)$, $P(K)$, $P(R\cap K)$, and $P(R\cup K)$, we first must know the number of cards that are red, $n(R)$, and the number of cards that are kings $n(K)$. Once we calculate $P(R)$, $P(K)$, and $P(R\cap K)$, we can use them to calculate $P(R\cup K)$. Note

$$\begin{aligned} n(R) &= 26, \quad n(K)=4,\\ P(R) &= \frac{n(R)}{\#\text{ of cards in a deck}}=\frac{26}{52}=\frac{1}{2},\\ P(K) &= \frac{n(K)}{\#\text{ of cards in a deck}}=\frac{4}{52}=\frac{1}{13},\\ P(R\cap K) &= \frac{\#\text{ of cards that are both red } \textit{and} \text{ a king}}{\#\text{ of cards in a deck}}=\frac{2}{52}=\frac{1}{26},\\ P(R\cup K) &= P(R)+P(K)-P(R\cap K)=\frac{1}{2}+\frac{1}{13}-\frac{1}{26}=\frac{7}{13}\end{aligned}$$

1.4.3 The probability of event A, $P(A)$, occurring is one minus the probability of the opposite event, $P(\bar{A})$, occurring.

a. $P(\bar{A}) = 1 - P(A) = 1 - \dfrac{1}{1987} = \dfrac{1986}{1987}$
b. $P(\bar{A}) = 1 - P(A) = 1 - 0.25 = 0.75$
c. $P(\bar{A}) = 1 - P(A) = 1 - 0.8 = 0.2$

1.4.4 By the axioms of probability and elementary set theory,

$$P(A) = P(1) + P(2) + P(3) = \left(\frac{1}{2}\right)^1 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 = 0.875,$$
$$P(B) = P(5) + P(6) + P(7) + P(8) + P(9) + P(10) = \left(\frac{1}{2}\right)^5 + \cdots + \left(\frac{1}{2}\right)^{10} = 0.0615,$$
$$P(A \cap B) = P(\emptyset) = 0,$$
$$P(A \cup B) = P(A) + P(B) = 0.875 + 0.0615 = 0.9365,$$

and

$$\begin{aligned} P(C) &= 1 - P(\bar{C}) = 1 - [P(1) + P(2) + P(3) + P(4)] \\ &= 1 - \left[\left(\frac{1}{2}\right)^1 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^4\right] = 1 - 0.9375 = 0.0625 \end{aligned}$$

1.4.5 This cannot happen because these probabilities do not satisfy probability Axiom 2. Let $S = \{n \colon 1 \leq n \leq 10\}$. By the Axiom 2, $P(S)$ should equal 1, but we see that

$$P(S) = P(1) + P(2) + \cdots + P(10) = \left(\frac{1}{3}\right)^1 + \left(\frac{1}{3}\right)^2 + \cdots + \left(\frac{1}{3}\right)^{10} \approx 0.500.$$

1.4.6 Let A be the event that the horse finishes fifth or better and let B be the event that the horse finishes in the top three The bookie claims that $P(A) = 0.35$ and $P(B) = 0.45$, but $B \subset A$ so by Theorem 1.4.3 $P(B) \leq P(A)$. But since $0.45 > 0.35$, these probabilities are not consistent with the axioms of probability.

1.4.7 Letting A be the event it rains today and B be the event it rains tomorrow, the forecaster claims that $P(A) = 0.25$, $P(B) = 0.50$, $P(A \cap B) = 0.10$, and $P(A \cup B) = 0.70$. But

$$P(A) + P(B) - P(A \cap B) = 0.25 + 0.50 - 0.10 = 0.65 \neq 0.70 = P(A \cup B),$$

so these probabilities do not satisfy the Addition Rule and thus they are not consistent with the axioms of probability.

1.4.8 To simplify notation, let $|I|$ denote the length of interval I.

a. Note that by the definition of the length of an interval and since B is a subinterval of S, $0 \leq |B| \leq |S|$ so that $0 \leq P(B) \leq 1$. Thus this assignment is consistent with axioms 1 and 2.

b. To be consistent with axiom 3, we need

$$P(\text{in } A \text{ or } B) = P(\text{in } A) + P(\text{in } B) = \frac{|A|}{|S|} + \frac{|B|}{|S|} = \frac{|A| + |B|}{|S|}.$$

Thus the only assignment that is consistent with axiom 3 is

$$P(\text{in } A \text{ or } B) = \frac{|A| + |B|}{|S|}.$$

1.4.9 From the axioms of Probability we know that $P(S) = 1$. Using the Addition Rule,

$$1 = P(S) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$1 = 0.6 + 0.8 - P(A \cap B) \quad \Rightarrow \quad P(A \cap B) = 0.4$$

1.4.10 Let A be the event that the first student is late and B be the event that the second student is late. We are told that $P(A) = 0.35$, $P(B) = 0.65$, and $P(A \cap B) = 0.25$. By the Addition Rule, the probability that at least of the students is late is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.35 + 0.65 - 0.25 = 0.75.$$

1.4.11 Let A and B be the events that die 1 and die 2 are greater than three, respectively. The probability of winning is $P(A \cup B)$. By the Addition Rule, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, so we must also take into account the probability, $P(A \cap B)$, that both dice will be greater than 3. Since this probability is not zero, $P(A \cup B) < 1$ and the gambler is not guaranteed to win.

1.4.12 Using the frequencies from the table,

$$P(A) = \frac{360}{581}, \quad P(B) = \frac{221}{581}, \quad P(A \cap B) = 0,$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{360}{581} + \frac{221}{581} - 0 = \frac{581}{581} = 1,$$
$$P(A \cup C) = P(A) + P(C) - P(A \cap C) = \frac{360}{581} + \frac{62}{581} - \frac{4}{581} = \frac{418}{581}, \quad \text{and}$$
$$P(B \cap D) = \frac{15}{581}.$$

A and B are disjoint because a person cannot test both positive and negative for marijuana in one test. B and D are not disjoint because from the table we see that there are 15 people who said they never used marijuana and still tested positive.

1.4.13 Let A be the event that at least two students share a birth month. It is easier to calculate the probability, $P(\bar{A})$, that all five students have a different birth month and then use Theorem 1.4.2 to calculate $P(A)$.

$$P(\bar{A}) = \frac{12 \cdot 11 \cdot 10 \cdot 9 \cdot 8}{12^5} = 0.3819 \quad \Rightarrow \quad P(A) = 1 - P(\bar{A}) = 1 - 0.3819 = 0.6181.$$

Since in a class of 15 students, every student cannot have a different birth month, the probability that at least two students share a birth month is $P(A) = 1$.

1.4.14

a. There are $36/2 = 18$ odd numbers, so the probability of winning is $P(A) = \frac{18}{28}$ and $P(\bar{A}) = 1 - \frac{18}{38} = \frac{20}{38}$. The odds against winning are then $\frac{20/18}{18/38} = \frac{20}{18} = \frac{10}{9}$, or $10 : 9$.

b. Using the Theoretical Approach from Section 1.2, we have $P(A) = \frac{n(A)}{n(S)}$ and $P(\bar{A}) = \frac{n(\bar{A})}{n(S)}$. So the odds against A are

$$\frac{P(\bar{A})}{P(A)} = \frac{n(\bar{A})/n(S)}{n(A)/n(S)} = \frac{n(\bar{A})}{n(A)}.$$

c. From part b., we know that $\frac{P(\bar{A})}{P(A)} = \frac{n(\bar{A})}{n(A)} = \frac{a}{b}$, so $n(\bar{A}) = ka$ and $n(A) = kb$ for some integer $k \geq 1$. But $S = A \cup \bar{A}$ and $A \cap \bar{A} = \emptyset$, so $n(S) = n(\bar{A}) + n(A) = ka + kb$. Therefore,

$$P(A) = \frac{n(A)}{n(S)} = \frac{kb}{ka + kb} = \frac{b}{a + b}.$$

1.4.15 Suppose $P(\{n\}) = p$ for all $n = 1, 2, \cdots$ and some $p > 0$. By Axiom 3, we need

$$P(S) = P(\{1\} \cup \{2\} \cup \cdots) = P(\{1\}) + P(\{2\}) + \cdots = p + p + \cdots$$

But this is a divergent series, so it does not have a sum equal to 1. This contradictions Axiom 2. Thus such a number p does not exist.

1.4.16 Let A and B be events of a sample space S. Prove the following relationships:

a. By elementary set theory, $A \cap B \subset B$, so by Theorem 1.4.3, $P(A \cap B) \leq P(A)$.

b. By elementary algebra and Theorem 1.4.2,

$$\begin{aligned} & P(B) \leq P(A) \\ \Rightarrow\ & -P(B) \geq -P(A) \\ \Rightarrow\ & 1 - P(B) \geq 1 - P(A) \\ \Rightarrow\ & P\left(\bar{B}\right) \geq P\left(\bar{A}\right) \end{aligned}$$

c. By axioms 2 and 3, the hint, and Theorem 1.4.2,

$$\begin{aligned} & S = \left(A \cap \bar{B}\right) \cup \bar{A} \cup (A \cap B) \\ \Rightarrow\ & 1 = P\left[\left(A \cap \bar{B}\right) \cup \bar{A} \cup (A \cap B)\right] \\ \Rightarrow\ & 1 = P\left(A \cap \bar{B}\right) + P\left(\bar{A}\right) + P(A \cap B) \\ \Rightarrow\ & P\left(A \cap \bar{B}\right) = 1 - P\left(\bar{A}\right) - P(A \cap B) \\ \Rightarrow\ & P\left(A \cap \bar{B}\right) = P(A) - P(A \cap B). \end{aligned}$$

d. Note that $A \cap \bar{B} \subset \bar{B}$ so that $P\left(A \cap \bar{B}\right) \leq P\left(\bar{B}\right) \quad \Rightarrow \quad -P\left(A \cap \bar{B}\right) \geq -P\left(\bar{B}\right)$. Combining this with part c., we get

$$P(A \cap B) = P(A) - P\left(A \cap \bar{B}\right) \geq P(A) - P\left(\bar{B}\right) = P(A) - [1 - P(B)] = P(A) + P(B) - 1.$$

1.4.17 Let A and B denote the events that person A and B are on-time, respectively. We are told that $P(A) = 0.35$, $P(B) = 0.65$, and $P(A \cap B) = 0.10$. The probability that person A is on-time but person B is not, is, by exercise 1.4.16 c.,

$$P(A \cap \bar{B}) = P(A) - P(A \cap B) = 0.35 - 0.10 = 0.25.$$

1.4.18 By the addition rule,

$$\begin{aligned}
&P(A \cup B \cup C) \\
&= P(A \cup (B \cup C)) \\
&= P(A) + P(B \cup C) - P(A \cap (B \cup C)) \\
&= P(A) + P(B) + P(C) - P(B \cap C) - P((A \cap B) \cup (A \cap C)) \\
&= P(A) + P(B) + P(C) - P(B \cap C) - [P(A \cap B) + P(A \cap C) - P((A \cap B) \cap (A \cap C))] \\
&= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).
\end{aligned}$$

1.5 Conditional Probability and the Multiplication Rule

1.5.1

a. Let B be the event the flower is blue, R be the event it is red, Y be the event it is yellow, S be the event that it is short, and T be the event that it is tall. Using Table **??** and the definition of conditional probability, we have the following

1. The probability that a flower is blue given that it is tall is

$$P(B \,|\, T) = \frac{P(B \cap T)}{P(T)} = \frac{53/390}{158/390} \approx 0.335.$$

2. The probability that it is short given that it is blue or yellow is

$$P(S \,|\, (B \cup Y)) = \frac{P(S \cap (B \cup Y))}{P(B \cup Y)} = \frac{(125 + 22)/390}{(178 + 102)/390} = \frac{147}{289} \approx 0.525.$$

3. The probability that it is red or blue given that it is tall is

$$P((R \cup B) \,|\, T) = \frac{P((R \cup B) \cap T)}{P(T)} = \frac{(53 + 25)/390}{158/390} = \frac{78}{159} \approx 0.494.$$

b. 1. Let B_1 and B_2 denote the events that the first and second flowers chosen are blue, respectively. Since we know that the flowers are chosen without replacement, we can use the multiplication rule to calculate the probability that both flowers are blue. So

$$P(B_1 \cap B_2) = P(B_1) \cdot P(B_2 \mid B_1) = \left(\frac{178}{390}\right) \cdot \left(\frac{177}{389}\right) \approx 0.208.$$

2. Let R be the event that the first flower is red and $\bar{Y}$ be the event that the second flower is not yellow (in other words, it is blue or red). As before, we can use the multiplication rule to calculate the probability that the first flower is red and the second is not yellow. So

$$P(R \cap \bar{Y}) = P(R) \cdot P(\bar{Y} \mid R) = \left(\frac{110}{390}\right) \cdot \left(\frac{109 + 178}{389}\right) \approx 0.208.$$

c. To find the probability that the first flower is short and the second is blue, we must first find the probabilities for each of the two cases described in the hint and then add them together. We can use the multiplication rule to calculate the probabilities of both cases. Let S be the event the first flower is short, B_1 be the event the first is blue, and B_2 be the event the second is blue. Then the probability that the first is short and blue and the second is blue is

$$P((S \cap B_1) \cap B_2) = P(S \cap B_1) \cdot P(B_2 \mid (S \cap B_1)) = \frac{125}{390} \cdot \frac{177}{389} \approx 0.1458.$$

and the probability that the first is short and not blue and the second is blue is

$$P((S \cap \bar{B}_1) \cap B_2) = P(S \cap \bar{B}_1) \cdot P(B_2 \mid (S \cap \bar{B}_1)) = \frac{85 + 22}{390} \cdot \frac{178}{389} \approx 0.1285.$$

We now add these probabilities together, so the probability that the first is short and the second is blue is

$$P = P((S \cap B_1) \cap B_2) + P((S \cap \bar{B}_1) \cap B_2) = 0.1458 + 0.1285 \approx 0.271.$$

1.5.2 Let T be the event that a car has a broken tail light and H be the event that a car has a broken head light. We are given that $P(T) = 0.3$, $P(H) = 0.15$, and $P(T \cap H) = 0.1$. We can use the definition of conditional probability to find the following

a. The probability that a car has a broken head light given that it has a broken tail light is

$$P(H \mid T) = \frac{P(H \cap T)}{P(T)} = \frac{0.1}{0.3} = \frac{1}{3}.$$

b. The probability that a car has a broken tail light given that it has a broken head light is

$$P(T \mid H) = \frac{P(H \cap T)}{P(H)} = \frac{0.1}{0.15} = \frac{2}{3}.$$

c. Here we use the addition rule from Section 1.5. The probability that a car has either a broken head light or a broken tail light is

$$P(H \cup T) = P(H) + P(T) - P(H \cap T) = 0.15 + 0.3 - 0.1 = 0.35.$$

1.5.3 Let G be the event that a package of cheese is good and let B be the event that it is bad. We apply the multiplication rule in a manner similar to Example 1.5.5 to calculate the probabilities.

a. The probability that the first three packages are all good is

$$P(G_1 \cap G_2 \cap G_3) = \frac{12}{15} \cdot \frac{11}{14} \cdot \frac{10}{13} = \frac{44}{91}.$$

b. The probability that the first three packages are all bad is

$$P(B_1 \cap B_2 \cap B_3) = \frac{3}{15} \cdot \frac{2}{14} \cdot \frac{1}{13} = \frac{1}{455}$$

c. The probability that the first two packages are good and the third package is bad is

$$P(G_1 \cap G_2 \cap B) = \frac{12}{15} \cdot \frac{11}{14} \cdot \frac{3}{13} = \frac{66}{455}.$$

d. To calculate the probability that exactly one of the first three packages is bad, we must calculate probabilities for three different cases and then add them together. In the first case, we calculate that the probability that the first package is bad is

$$P(B \cap G_1 \cap G_2) = \frac{3}{15} \cdot \frac{12}{14} \cdot \frac{11}{13} = \frac{66}{455}.$$

Likewise, the probabilities that the second and third packages are bad are

$$P(G_1 \cap B \cap G_2) = \frac{12}{15} \cdot \frac{3}{14} \cdot \frac{11}{13} = \frac{66}{455} \text{ and } P(G_1 \cap G_2 \cap B) = \frac{12}{15} \cdot \frac{11}{14} \cdot \frac{3}{13} = \frac{66}{455}.$$

We can now add these together to find that the probability that exactly one of the first three packages is bad is

$$P = \frac{66}{455} + \frac{66}{455} + \frac{66}{455} = \frac{198}{455}$$

1.5.4 We are given that for a randomly selected motorcycle owner, A is the event that the person is male and B is the event that the person is named Pat. Since we can assume that there are many male names, the conditional probability that a person is named Pat given that that person is a male, $P(B \mid A)$, is relatively small. Since we can also assume that many of the people named Pat are also male, the probability that a motorcycle owner is male given that the person is named Pat, $P(A \mid B)$, is relatively large. Thus, $P(A \mid B)$ is the higher conditional probability.

1.5.5 After two kings have been dealt, there are 50 cards left, two of which are kings. Let K denote the event of getting a king and O denote the event of getting a card other than a king. There are three ways that two of the next three cards are also kings. Using notation similar to that in Example 1.5.6, their probabilities are:

$$P(KKO) = \frac{2}{50} \cdot \frac{1}{49} \cdot \frac{48}{48} = \frac{1}{1225},$$
$$P(KOK) = \frac{2}{50} \cdot \frac{48}{49} \cdot \frac{1}{48} = \frac{1}{1225}, \quad \text{and}$$
$$P(OKK) = \frac{48}{50} \cdot \frac{2}{49} \cdot \frac{1}{48} = \frac{1}{1225}.$$

Since these three ways are disjoint, the probability that two of the next three cards are also kings is the sum of these three probabilities, 3/1225.

1.5.6 His total will be more than 21 if the next card is worth 5 or more. After the first three cards, there are 49 cards left, 35 of which are worth 5 or more (note that an Ace is worth 1 in this situation). Thus the probability his total will be more than 21 is 35/49.

1.5.7 Let L denote the event you like the selected candy and $\bar{L}$ denote the event you don't like the candy.

a. $P(L \cap L) = \dfrac{5}{12} \cdot \dfrac{4}{11} = \dfrac{5}{33}$
b. $P\left[\left(L \cap \bar{L}\right) \cup \left(\bar{L} \cap L\right)\right] = \dfrac{5}{12} \cdot \dfrac{7}{11} + \dfrac{7}{12} \cdot \dfrac{5}{11} = \dfrac{35}{66}$
c. $P\left(\bar{L} \cap \bar{L}\right) = \dfrac{7}{12} \cdot \dfrac{6}{11} = \dfrac{7}{22}$

1.5.8 Let R, B, and G denote the events the selected cube is red, blue, and green, respectively.

a. The two cubes can be the same color in three different ways. Their probabilities are:

$$P(RR) = \frac{5}{14} \cdot \frac{4}{13} = \frac{10}{91},$$
$$P(BB) = \frac{6}{14} \cdot \frac{5}{13} = \frac{15}{91}, \quad \text{and}$$
$$P(GG) = \frac{3}{14} \cdot \frac{2}{13} = \frac{3}{91}.$$

Since these outcomes are disjoint, the probability that both cubes are the same color is the sum of these probabilities, 4/13.

b. Let S denote the event both cubes are the same color. By the definition of conditional probability, the probability that both cubes are blue given that they are the same color is

$$P(BB \mid S) = \frac{P(BB \cap S)}{P(S)}.$$

But the events BB and S can occur only if both cubes are blue. So by the calculations in part a.,

$$P(BB \mid S) = \frac{P(BB \cap S)}{P(S)} = \frac{P(BB)}{P(S)} = \frac{15/91}{4/13} = \frac{15}{28}.$$

1.5.9

a. On the first selection, there are 5 cubes, one of which is labeled 1, so the probability that the first cube selected is labeled 1 is 1/5.

b. The only way that the third cube selected can be labeled 3 is if the first two cubes are not labeled 3 and the third is labeled 3. The probability of this occurring is

$$\frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} = \frac{1}{5}.$$

c. The only way that the second and third cubes selected can be labeled 2 and 3, is if the first cube is not labeled 2 or 3, the second cube is labeled 2, and the third cube is labeled 3. The probability of this occurring is

$$\frac{3}{5} \cdot \frac{1}{4} \cdot \frac{1}{3} = \frac{1}{20}.$$

1.5.10 Let O denote the event of rolling something other than a 3. Using complements, we get

$$P(\text{at least one 3}) = 1 - P(\text{no 3's}) = 1 - P(OOOOO) = 1 - \left(\frac{5}{6}\right)^5 = \frac{4651}{7779} \approx 0.5981.$$

1.5.11

a. Note that since the selections are done with replacement, the probability that someone other than Jose is selected at each selection is $(N-1)/N$. Then using complements,

$$\begin{aligned} P(\text{Jose is selected}) &= 1 - P(\text{Jose is not selected}) \\ &= 1 - P(n \text{ of the other } N-1 \text{ people are selected}) \\ &= 1 - \left(\frac{N-1}{N}\right)^n \\ &= 1 - \left(1 - \frac{1}{N}\right)^n. \end{aligned}$$

b. If the selections are done without replacement, then the probability that Jose is *not* selected in each selection changes from one selection to the next. Again using complements,

$$\begin{aligned} P(\text{Jose is selected}) &= 1 - P(\text{Jose is not selected}) \\ &= 1 - \frac{N-1}{N} \cdot \frac{N-2}{N-1} \cdots \frac{N-n}{N-(n-1)} \\ &= 1 - \frac{N-n}{N} \\ &= \frac{n}{N}. \end{aligned}$$

1.5.12

a. By the definition of conditional probability,

$$P(A \mid E) = \frac{P(A \cap E)}{P(E)}.$$

But since probabilities must be non-negative, this quotient must be non-negative. Also, since $A \cap E \subset E$, by Theorem 1.4.3, $P(A \cap E) \leq P(E)$ so that

$$P(A \mid E) = \frac{P(A \cap E)}{P(E)} \leq \frac{P(E)}{P(E)} = 1.$$

Thus $0 \leq P(A \mid E) \leq 1$.

b. Note that since $E \subset S$, $E \cap S = E$ so that

$$P(S \mid E) = \frac{P(S \cap E)}{P(E)} = \frac{P(E)}{P(E)} = 1.$$

c. Note that since $A \cap B = \emptyset$, $(A \cup B) \cap E = (A \cap E) \cup (B \cap E)$ and that $(A \cap E) \cap (B \cap E) = \emptyset$ so that by the definition of conditional probability,

$$\begin{aligned} P(A \cup B \mid E) &= \frac{P((A \cup B) \cap E)}{P(E)} \\ &= \frac{P((A \cap E) \cup (B \cap E))}{P(E)} \\ &= \frac{P(A \cap E) + P(B \cap E)}{P(E)} \\ &= \frac{P(A \cap E)}{P(E)} + \frac{P(B \cap E)}{P(E)} \\ &= P(A \mid E) + P(B \mid E). \end{aligned}$$

1.6 Bayes' Theorem

1.6.1

a. Since every outcome is either even or odd, these events do form a partition of the sample space.

b. The injury could occur somewhere other than the leg or head, so the union of these events is not the entire sample space. Also, the boy could be injured on both the leg and head, so these events are not disjoint. Thus these events do not form a partition of the sample space.

c. The package could be, for instance, a 24 oz package of green beans, so the union of these events is not the entire sample space. Also, the package could be a 12 oz package of carrots, so these events are not disjoint. Thus these events do not form a partition of the sample space.
d. Since the satellite cannot land on both the ground and in the water at the same time, these events are disjoint. Also, their union is the entire sample space. So these events do form a partition of the sample space.
e. The computer could be the correct price, so the union of these events is not the entire sample space. Thus these events do not form a partition of the sample space.
f. The voter could be independent or belong to another party, so the union of these events is not the entire sample space. Thus these events do not form a partition of the sample space.
g. The chips could be, for instance, sour cream and onion flavored, so the union of these events is not the entire sample space. Thus these events do not form a partition of the sample space.

1.6.2 Let A_1 and A_2 be the events the airplane is produced by machine A and B, respectively, and B be the event the airplane is defective. We are told that $P(A_1) = 0.6$, $P(A_2) = 0.4$, $P(B \mid A_1) = 0.02$, and $P(B \mid A_2) = 0.03$. Thus

$$P(A_2 \mid B) = \frac{0.4(0.03)}{0.6(0.02) + 0.4(0.03)} = 0.5.$$

1.6.3 Let A_1 and A_2 be the events the person is female and male, respectively, and B be the event the person supports increased enforcement. We are told that $P(A_1) = 0.6$, $P(A_2) = 0.4$, $P(B \mid A_1) = 0.7$, and $P(B \mid A_2) = 0.2$. Thus

$$P(A_2 \mid B) = \frac{0.4(0.2)}{0.6(0.7) + 0.4(0.2)} = 0.16.$$

1.6.4 Let A_1, A_2, and A_3 be the events the customer comes to shop, eat, and socialize, respectively, and B be the event the customer is age 20 or less. We are told that $P(A_1) = 0.8$, $P(A_2) = 0.15$, $P(A_3) = 0.05$, $P(B \mid A_1) = 0.05$, $P(B \mid A_2) = 0.35$, and $P(B \mid A_3) = 0.9$. Thus

$$P(A_1 \mid B) = \frac{0.8(0.05)}{0.8(0.05) + 0.15(0.35) + 0.05(0.9)} = \frac{16}{55} \approx 0.291.$$

1.6.5 Let A_1, A_2, and A_3 be the events the cube was originally in bag 1, 2, and 3, respectively, and B be the event the cube is red. We are told that $P(A_1) = 1/3$, $P(A_2) = 1/3$, $P(A_3) = 1/3$, $P(B \mid A_1) = 3/10$, $P(B \mid A_2) = 4/10$, and $P(B \mid A_3) = 5/10$. Thus

$$P(A_1 \mid B) = \frac{(1/3)(3/10)}{(1/3)(3/10) + (1/3)(4/10) + (1/3)(5/10)} = \frac{1}{4}.$$

1.6.6

a. If $P(B \mid A_2) = 1 - p$, we replace 0.1 with $1 - p$ in the calculation of $P(A_3 \mid B)$ in Example 1.6.2 yielding

$$P(A_3 \mid B) = \frac{1/3}{1/3 + 1/3\,(1-p) + 1/3}.$$

b. By elementary calculus, $\lim_{p \to 0} P(A_3 \mid B) = 1/3$. This limit means that if p is "close" to 1, and we don't find the plane in region 2 (meaning event B has occurred), the plane could be in any one of the three regions with equal probability.
c. By elementary calculus, $\lim_{p \to 1} P(A_3 \mid B) = 1/2$. This limit means that if p is "close" to 1, and we don't find the plane in region 2 (meaning event B has occurred), the plane is very likely *not* in region 2 and is in region 1 or 3 with equal probability.

1.6.7 Let A_1 and A_2 be the events the woman has cancer and does not have cancer, respectively, and B be the event she tests positive. We are told that $P(A_1) = 0.009$, $P(A_2) = 0.991$, $P(B \mid A_1) = 0.9$, and $P(B \mid A_2) = 0.06$. Thus

$$P(A_1 \mid B) = \frac{0.009(0.9)}{0.009(0.9) + 0.991(0.06)} \approx 0.1199.$$

This result does agree with that found using natural frequencies.

1.6.8 Consider a randomly selected group of 1000 people. In this group we would expect

- $1000(0.005) = 5$ to actually use marijuana and 995 to not use,
- $5(0.975) \approx 5$ of those who use marijuana will test positive, and
- $999(0.135) \approx 134$ of those who do not use will test positive.

Thus we would expect a total of $5 + 134 = 139$ to test positive. Of these, only 5 actually use marijuana. Therefore, the probability of actually using marijuana given a positive result is $5/139 \approx 0.0360$. This result is nearly identical to that found in the example.

1.6.9 Consider a randomly selected group of 1000 50-year-old musicians. In this group we would expect

- $1000(0.02) = 20$ to smoke and 980 to not smoke,
- $20(0.05) = 1$ of those who smoke will die in the next year, and
- $980(0.005) \approx 5$ of those who do not smoke will die in the next year.

Thus we would expect a total of $1 + 5 = 6$ to die in the next year. Out of these, only 1 is a smoker. Therefore, the probability that the person was a smoker given that the person dies in the next year is $1/6 \approx 0.167$.

1.6.10 Let A_1 and A_2 be the events the person is wearing a Bulldogs and Blue Jays t-shirt, respectively, and B_1 be the event the selected letter is a vowel. We are told that $P(A_1) = 0.3$, $P(A_2) = 0.7$, $P(B_1 \mid A_1) = 2/8$, and $P(B_1 \mid A_2) = 3/8$. Thus

$$P(A_1 \mid B_1) = \frac{0.3(2/8)}{0.3(2/8) + 0.7(3/8)} = \frac{2}{9}.$$

Now let B_2 be the event the selected letter is a consonant. We are told that $P(B_2 \mid A_1) = 6/8$, and $P(B_2 \mid A_2) = 5/8$. Thus

$$P(A_1 \mid B_2) = \frac{0.3(6/8)}{0.3(6/8) + 0.7(5/8)} = \frac{18}{53}.$$

1.6.11 Let A_1 and A_2 be the events as defined in Example 1.6.4 and B be the event we select a red cube. We are told that $P(A_1) = 0.9$, $P(A_2) = 0.1$, $P(B \mid A_1) = 9/10$, and $P(B \mid A_2) = 1/10$. Thus

$$P(A_1 \mid B) = \frac{0.9(9/10)}{0.9(9/10) + 0.1(1/10)} = \frac{81}{82} \approx 0.988.$$

Since this probability is greater than 0.9, selecting a red cube would make us more confident that there are nine red cubes and one blue cube in the bag.

1.6.12 Let A_1 and A_2 be the events the first student did and did not do the problems, respectively, and B be the event the first student fails the test. The professor believes that $P(A_1) = 0.95$, $P(A_2) = 0.05$, $P(B \mid A_1) = 0.15$, and $P(B \mid A_2) = 0.9$. Thus

$$P(A_1 \mid B) = \frac{0.95(0.15)}{0.95(0.15) + 0.05(0.9)} = 0.76.$$

1.6.13

a. Approximately $950(0.135) \approx 128$ will get a false positive result, and $50(0.025) \approx 1$ will get a false negative result.
b. Approximately $50(0.975) \approx 49$ will get "caught," 1 will "get away with it," and 128 will be falsely accused of using marijuana.
c. If we gave the test to every student, we will falsely accuse about 3 times as many students of using marijuana than we actually catch. This is not good, so we woul not recommend giving the test to every student.

1.7 Independent Events

1.7.1

a. The two sisters probably have similar tastes in men, so it is very likely they will both give the same answer. Thus it is not reasonable to assume the events are independent.
b. Since they are in different cities, is it reasonable to assume the events are independent.
c. Since they are eating together, they will likely order food from the same vendor, so what one eats will likely depend on what the other eats. Thus it is not reasonable to assume the events are independent.
d. If the boy sits on the bench the entire time, he is less likely to skin his knee than if he didn't sit on the bench. Thus it is not reasonable to assume the events are independent.
e. Since the cities are far apart, it is reasonable to assume that the weather in one city is independent of the other. Thus it is reasonable to assume the events are independent.
f. If the truck has a dented rear fender, it probably got hit by something from the back. This increases the probability it has a damanaged tail light. Thus it is not reasonable to assume the events are independent.

g. If the car runs out of gas, it may be a sign the man is absent-minded which increases the probability he forgot the gas for the airplane. So it is not reasonable to assume the events are independent.

1.7.2

a. Assuming independence, the probability that all three engines fail on the the same flight is $0.0001^3 = 1 \times 10^{-12}$.
b. If a single mechanic changes the oil in all three engines before a flight, there is a higher probability that a common mistake may be made on the engines. This makes the events of engines failures not independent.

1.7.3

a. The probability of a miss is $1-0.15 = 0.85$ so that the probability of five misses is $0.85^5 = 0.444$, assuming independence.
b. Out of n darts,

$$P(\text{at least one hit}) = 1 - P(n \text{ misses}) = 1 - 0.85^n.$$

For this probability to be at least 0.75, we need $1 - 0.85^n > 0.75 \Rightarrow n > \ln(0.25)/\ln(0.85) = 8.53$. Thus he must throw at least 9 darts.

1.7.4 Since there are 4 equally likely outcomes and each event contains 2 outcomes, $P(A) = P(B) = P(C) = 1/2$. Now,

$$A \cap B = A \cap C = B \cap C = \{1\} \quad \Rightarrow \quad P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2}$$

which shows that the events are pairwise independent. But

$$A \cap B \cap C = \{1\} \quad \Rightarrow \quad P(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$$

which shows that the events are not mutually independent.

1.7.5 Note that there are a total of $21 + n$ students in the class, 15 of which are girls, $6 + n$ of which are freshmen, and 6 of which a freshmen girls. Then

$$P(A) = \frac{15}{21+n}, \quad P(B) = \frac{6+n}{21+n}, \quad \text{and} \quad P(A \cap B) = \frac{6}{21+n}.$$

For A and B to be independent, we need

$$\frac{15}{21+n} \cdot \frac{6+n}{21+n} = \frac{6}{21+n}.$$

Solving this equation using elementary algebra yields $n = 4$.

1.7.6 The event that all of the cubes selected are blue can happen in six disjoint ways: by getting n on the die and then selecting n blue cubes where $n = 1, \ldots, 6$. Since the result of the die and the cubes selected are independent,

$$P[(n \text{ on the die}) \cap (n \text{ blue cubes})] = P(n \text{ on the die}) \cdot P(n \text{ blue cubes}).$$

We use this result to calculate the probability for each value of n below:

$$\begin{aligned}
n = 1 &: \frac{1}{6} \cdot \frac{10}{15} = \frac{1}{9} \\
n = 2 &: \frac{1}{6} \cdot \frac{10}{15} \cdot \frac{9}{14} = \frac{1}{14} \\
n = 3 &: \frac{1}{6} \cdot \frac{10}{15} \cdot \frac{9}{14} \cdot \frac{8}{13} = \frac{4}{91} \\
n = 4 &: \frac{1}{6} \cdot \frac{10}{15} \cdot \frac{9}{14} \cdot \frac{8}{13} \cdot \frac{7}{12} = \frac{1}{39} \\
n = 5 &: \frac{1}{6} \cdot \frac{10}{15} \cdot \frac{9}{14} \cdot \frac{8}{13} \cdot \frac{7}{12} \cdot \frac{6}{11} = \frac{2}{143} \\
n = 6 &: \frac{1}{6} \cdot \frac{10}{15} \cdot \frac{9}{14} \cdot \frac{8}{13} \cdot \frac{7}{12} \cdot \frac{6}{11} \cdot \frac{5}{10} = \frac{1}{143}
\end{aligned}$$

Then,

$$P(\text{all blue cubes}) = \frac{1}{9} + \frac{1}{14} + \frac{4}{91} + \frac{1}{39} + \frac{2}{143} + \frac{1}{143} = \frac{703}{2574} \approx 0.273.$$

1.7.7 Note that $P(S) = 0.25$ and $P(N) = 0.75$.

a. $P(SSSNN) = (0.25)^3(0.75)^2 = 0.00879$
b. $P(SNNSS) = (0.25)(0.75)^2(0.25)^2 = 0.00879$ By similar calculations, $P(NSSNS) = 0.00879$. These probabilities are the exact same as in part a.
c. We need to choose the three positions for the voters who support it. Since the order does not matter, there are $\binom{5}{3} = 10$ ways of doing this.
d. There are 10 different outcomes in which three voters support the issue and two do not support it. Each one of these outcomes has probability 0.00879, so the overall probability is $10(0.00879) = 0.0879$.

1.7.8 Let p denote the probability that a single box contains a "W." Then assuming independence,

$$\begin{aligned}
0.36 &= 1 - P(\text{losing}) \\
&= 1 - P(\text{all 4 boxes do not contain a W}) \\
&= 1 - (1 - p)^4.
\end{aligned}$$

Solving this equation for p yields $p = 1 - 0.64^{1/4} \approx 0.1056$.

1.7.9

a. Since the tires are selected without replacement, the selections are dependent.

b. Note that

$$\begin{aligned} P(\text{shipment is accepted}) &= P(\text{all 5 tires pass inspection}) \\ &= P(T_1 \cap \cdots \cap T_5) \\ &= \frac{490}{500} \cdot \frac{489}{499} \cdots \frac{486}{496} \approx 0.9035. \end{aligned}$$

c. If the selections are done with replacement, then

$$\begin{aligned} P(\text{shipment is accepted}) &= P(T_1 \cap \cdots \cap T_5) \\ &= P(T_1) \cdots P(T_5) = \left(\frac{490}{500}\right)^5 \approx 0.9039. \end{aligned}$$

d. No, there is not much difference.

1.7.10 Assuming independence of the samples,

$$\begin{aligned} P(\text{combined sample tests positive}) &= 1 - P(\text{combined sample tests negative}) \\ &= 1 - P(\text{all 10 individual samples test negative}) \\ &= 1 - 0.99^{10} \\ &\approx 0.0956. \end{aligned}$$

1.7.11 Note that the probability of *not* getting a 3 is 5/6. Since the rolls are independent,

$$P(\text{at least one 3 in } n \text{ rolls}) = 1 - P(\text{no 3's}) = 1 - \left(\frac{5}{6}\right)^n.$$

For this probability to be at least 0.5, we need $n > \ln(0.5)/\ln(5/6) \approx 3.8$. Thus 4 is the smallest value of n.

1.7.12

a. $P(\text{all four work properly}) = 0.95^4 \approx 0.8145$

b. $P(\text{all four fail to work properly}) = 0.05^4 = 6.25 \times 10^{-6}$

c. There are four outcomes in which exactly one system works properly: the first works, and the other three do not; the second works, and the other three do not; and so on. The probability of each of these outcomes is $0.95(0.05)^3 = 0.0001875$. Thus

$$P(\text{exactly one works properly}) = 4(0.0001875) = 0.000475.$$

d. Using independent and disjoint events

$$\begin{aligned} P(\text{at least two work properly}) &= 1 - P(\text{fewer than 2 work properly}) \\ &= 1 - P(\text{0 or 1 work properly}) \\ &= 1 - [P(\text{0 works properly}) + P(\text{1 works properly})] \\ &= 1 - [6.25 \times 10^{-6} + 0.000475] \\ &\approx 0.9995 \end{aligned}$$

1.7.13 Let $HHHH$ denote the outcome that the child in happy in minute 1, 2, 3, and 4. There are four outcomes in which the child is happy in minute 1 and happy in minute 4, the probabilities of which are

$$P(HHHH) = p^3, \;\; P(HHSH) = p\,(1-p)^2,$$
$$P(HSHH) = (1-p)^2 p, \;\text{ and }\; P(HSSH) = (1-p)p\,(1-p).$$

Then since these outcomes are disjoint,

$$\begin{aligned} &P(\text{happy in minute 1 and happy in minute 4}) \\ &= p^3 + p\,(1-p)^2 + (1-p)^2 p + (1-p)p\,(1-p) \\ &= p\left(4p^2 - 6p + 3\right). \end{aligned}$$

1.7.14

a. Since A and B are disjoint, $A \cap B = \emptyset$ so that $P(A \cap B) = 0$. But since $P(A) \neq 0 \neq P(B)$, $P(A) \cdot P(B) \neq 0$. Thus $P(A \cap B) \neq P(A) \cdot P(B)$ proving that A and B are dependent by definition.
b. Note that $\emptyset \cap A = \emptyset$ for any event A. Thus $P(\emptyset \cap A) = 0 = P(\emptyset) \cdot P(A)$, proving that $\emptyset$ and A are independent.

1.7.15

a. Note that $B = (B \cap A) \cup \left(B \cap \bar{A}\right)$ and that $\emptyset = (B \cap A) \cap \left(B \cap \bar{A}\right)$ so that

$$\begin{aligned} P(B) &= P\left[(B \cap A) \cup \left(B \cap \bar{A}\right)\right] \\ &= P(B \cap A) + P\left(B \cap \bar{A}\right) \\ &= P(B)P(A) + P\left(B \cap \bar{A}\right) \end{aligned}$$

$$\begin{aligned} \Rightarrow P\left(B \cap \bar{A}\right) &= P(B) - P(B)P(A) \\ &= P(B)[1 - P(A)] \\ &= P(B)P\left(\bar{A}\right). \end{aligned}$$

Thus $\bar{A}$ and B are independent by definition.

b. Replacing B with $\bar{B}$ in the calculations in part a. and using the fact that A and $\bar{B}$ are independent shows that $P\left(\bar{B} \cap \bar{A}\right) = P\left(\bar{B}\right) P\left(\bar{A}\right)$. Thus $\bar{A}$ and $\bar{B}$ are independent by definition.

1.7.16 If not switching, we win only if the door with the prize is initially selected. The probability of this is $1/N$. If switching, consider the following events:

1. The real prize is behind door 1. In this event, we cannot win, so the probability of winning is 0.
2. The real prize is behind door N. This event occurs with probability $1/N$. We win only if we switch to door N. Since there are $N-2$ doors to switch to, we select door N with probability $1/(N-2)$. Thus the probability of winning in this event is $1/[N(N-2)]$.
3. The real prize is behind one of doors 2 through $N-1$. This event occurs with probability $(N-2)/N$. We win only if we switch to the door with the real prize. Since there are $N-2$ doors to switch to, we select the winning door with probability $1/(N-2)$. Thus the probability of winning in this event is $(N-2)/N \cdot 1/(N-2) = 1/N$.

We win if one of these three disjoint events occur. Thus the probability of winning if we switch is

$$0 + \frac{1}{N(N-2)} + \frac{1}{N} = \frac{N-1}{N(N-2)}.$$

Chapter 2

Discrete Random Variables

2.1 Introduction

2.2 Probability Mass Functions

2.2.1 By the definition of the pmf,

$$
\begin{aligned}
P(X \in \{2, 7, 11\}) &= 1/36 + 6/36 + 2/36 = 1/4 \\
P(3 \leq X < 10) &= 2/36 + 3/36 + 4/36 + 5/36 + 6/36 + 5/36 + 4/36 = 29/36 \\
P(12 \leq X \leq 15) &= 1/36
\end{aligned}
$$

2.2.2

a. The pmf is described in the table below. We see that the largest value of f is $f(5) = 5/15$, so the mode is 5. The median is 4 sice $P(X \leq 4) = 10/15$ and $P(X \geq 4) = 9/15$ and $x = 4$ is the only number for which both of these probabilities are greater than 1/2.

$\boldsymbol{x}$	1	2	3	4	5
$\boldsymbol{f(x)}$	1/15	2/15	3/15	4/15	5/15

b. The probability histogram is shown in the left half of Figure 2.1.

c. The graph of the cdf is shown in the right half of Figure 2.1.

2.2.3

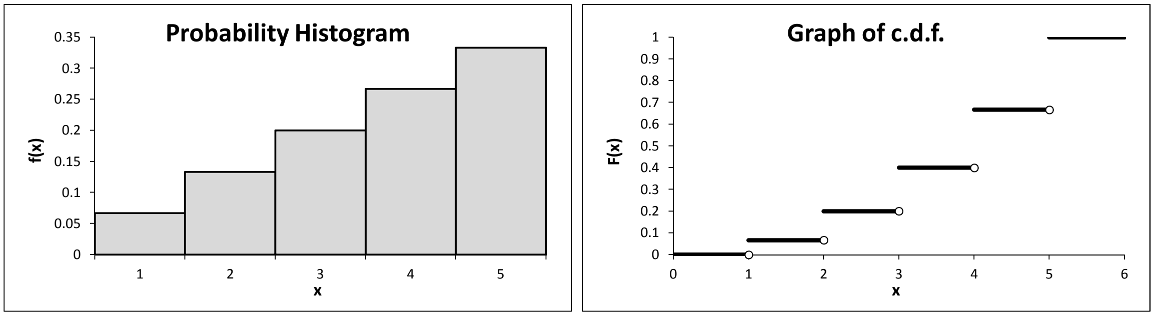


Figure 2.1

a. We need $1/a + 2/a + 3/a = 6/a$ to equal 1. Thus we need $a = 6$.
b. We need $a(1/3)^0 + a(1/3)^1 + a(1/3)^2 + \cdots$ to equal 1. But this series is a geometric series and its sum is $a/(1 - 1/3) = 3a/2$. Thus we need $a = 2/3$
c. We need $0.2 + 0.36 + a = 0.56 + a$ to equal 1. Thus we need $a = 0.44$.

2.2.4 By property 2 $\sum_{x \in R} f(x) = 1$. But by property 1, all the terms in this series are positive. The only way for this to happen is if each term is less than or equal to 1.

2.2.5 Note that $(1/2)^x > 0$ for all x, so this function satisfies the first property of a pmf Now, by properties of geometric series,

$$\sum_{x \in R} f(x) = \sum_{x=1}^{\infty} \left(\frac{1}{2}\right)^x = \sum_{x=1}^{\infty} \frac{1}{2} \left(\frac{1}{2}\right)^x = \frac{1/2}{1 - 1/2} = 1$$

so that the function satisfies the second property of a pmf

2.2.6 Note that if $a > 0$, then $a/x^2 > 0$ for all x, so this function satisfies the first property of a pmf To satisfy the second property, we need

$$\sum_{x \in R} f(x) = \sum_{x=1}^{\infty} \frac{a}{x^2} = a \sum_{x=1}^{\infty} \frac{1}{x^2}.$$

But this series is a p-series with $p = 2$ thus it converges to some number, call it c. Since $1/x^2 > 0$ for all x, $c > 0$. Then for $a = 1/c$, $\sum_{x=1}^{\infty} 1/x^2 = 1$ so that the second property is satisfied.

2.2.7 The pmf is $f(x) = 1/k$ for all $x \in R$. Then by definition of the cdf, for $x \geq 1$,

$$F(x) = P(X \leq x) = f(1) + \cdot + f(\lfloor x \rfloor) = \frac{\lfloor x \rfloor}{k}$$

where $\lfloor x \rfloor$ is the greatest integer less than or equal to x. For $x < 1$, $F(x) = 0$.

2.2.8 The possible values of X and its distribution are shown in the tables below.

	Second Die					
First Die	**1**	**2**	**3**	**4**	**5**	**6**
1	1	2	3	4	5	6
2	2	2	3	4	5	6
3	3	3	3	4	5	6
4	4	4	4	4	5	6
5	5	5	5	5	5	6
6	6	6	6	6	6	6

x	1	2	3	4	5	6
$f(x)$	1/36	3/36	5/36	7/36	9/36	11/36

2.2.9 The possible values of X and its distribution are shown in the tables below.

	Die					
Coin	**1**	**2**	**3**	**4**	**5**	**6**
Heads	2	3	4	5	6	7
Tails	1	2	3	4	5	6

x	1	2	3	4	5	6	7
$f(x)$	1/12	2/12	2/12	2/12	2/12	2/12	1/12

2.2.10 By the definition of conditional probability,

$$P(X \geq 3 \,|\, X \geq 1) = \frac{P(X \geq 3 \cap X \geq 1)}{P(X \geq 1)} = \frac{P(X \geq 3)}{P(X \geq 1)} = \frac{0.4 + 0.1}{0.1 + 0.3 + 0.4 + 0.1} = \frac{5}{9}$$

2.2.11 The smallest value of X occurs when all the flips are heads resulting in $X = -n$. The largest value occurs when all the flips are tails resulting in $X = n$. Thus the range of X is $\{-n, -n+1, \ldots, n-1, n\}$ and $P(X = n) = (1/2)^n$.

2.2.12

a. $P(X > 0) = 0.2 + 0.1 = 0.3$
b. $P(X > 3) = 0.1$
c. $P(X < -3) = 0.4$
d. $P(X < -8) = 0$

2.2.13 The pmf is described in the table below.

$\boldsymbol{x}$	0	1	2	3	4
$\boldsymbol{y}$	-1	2	5	8	11
$\boldsymbol{f(y)}$	0.1	0.1	0.3	0.4	0.1

2.2.14

a. The probability histogram is shown in the left half of Figure 2.2. We see the largest value of f occurs at $x = 0$, so the mode is 0.

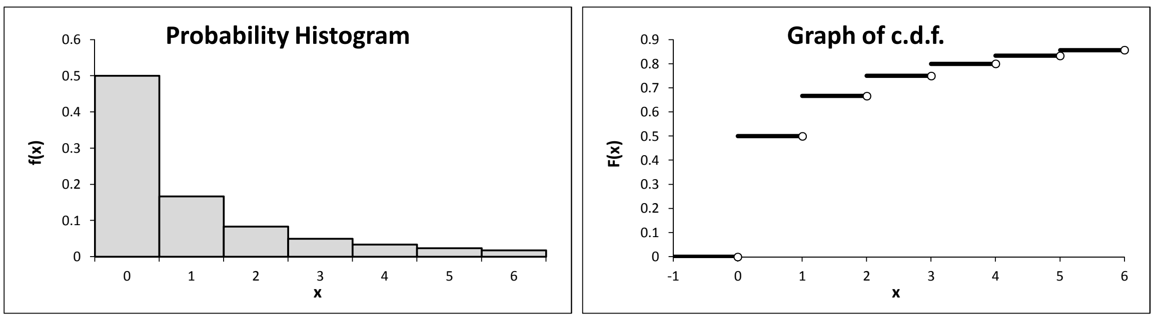


Figure 2.2

b. The graph of the cdf is shown in the right half of Figure 2.2.

c. By the definition of conditional probability,

$$P(X \geq 5 \,|\, X \geq 2) = \frac{P(X \geq 5 \cap X \geq 2)}{P(X \geq 2)} = \frac{P(X \geq 5)}{P(X \geq 2)}$$
$$= \frac{1 - P(X \leq 4)}{1 - P(X \leq 1)} = \frac{1 - (1/2 + 1/6 + 1/12 + 1/20 + 1/30)}{1 - (1/2 + 1/6)} = \frac{1}{2}$$

2.2.15 Let x be a positive integer. Then $X = x$ only if the first $x - 1$ flips are heads and the last flip is tails. The probability of this occurring is $(1/2)^x$. This the pmf of X is $f(x) = (1/2)^x$ for $x = 1, 2, \ldots$.

2.2.16 The pmf is given in the table below.

$\boldsymbol{x}$	-1	0	2	3
$\boldsymbol{f(x)}$	3	4	2	1

2.2.17 The pmf is given in the table below.

$\boldsymbol{x}$	-2	0	5	15
$\boldsymbol{f(x)}$	0.2	0.4	0.39	0.01

2.2.18 By definition, $F(a) = P(X \leq a)$ and $F(b) = P(X \leq b)$. Let A be the set of outcomes for which $X \leq a$ and B be the set of outcomes for which $X \leq b$. Then $P(X \leq a) = P(A)$ and $P(X \leq b) = P(B)$. However, if $k \in A$, then $X(k) \leq a \leq b$ so that $k \in B$. Thus $A \subset B$ and by Theorem 1.4.3, $P(A) \leq P(B)$ proving that $F(a) \leq F(b)$.

2.2.19 The two mistakes are

1. She has a higher probability of getting 80% of less than getting a 90% or less. Since there are more ways to get a 90% or less, this probability should be higher.
2. Since the highest grade is 100%, the probability of getting 100% or less is 1. But she claims this probability is only 0.98.

2.2.20 Note that $X = 0$ only if all three guesses are wrong. The probability of this happening is $(3/4)^3 = 27/64$. $X = 1$ if exactly one is correct and two are wrong which can happen in three ways: CWW, WCW, and WWC. Each one of these outcomes has probability $(1/4)(3/4)^2 = 9/64$, so $P(X = 1) = 3(9/64) = 27/64$. By similar arguments $P(X = 2) = 9/64$ and $P(X = 3) = 1/64$. The distribution is summarized in the table below

x	0	1	2	3
$f(x)$	27/64	27/64	9/64	1/64

2.2.21 Write 1 - 6 on one die and 0, 0, 0, 6, 6, 6 on the other.

2.2.22 A random variable cannot have more than one median because the median is defined to be the smallest value of m satisfying the conditions in the definition. There can only be one smallest value.

2.2.23 X is neither discrete nor continuous. X can take any value between 100 and 1000, but since the repair costs can be as high as \$5,000, many payouts will be exactly 1,000 so $P(X = 1000) \neq 0$.

2.3 The Hypergeometric and Binomial Distributions

2.3.1 Let X denote the number of sour gallons purchased. Then X has a hypergeometric distribution with $N_1 = 6$, $N_2 = 19$, $N = 25$, and $n = 5$ so that the pmf is $f(x) = \frac{\binom{6}{x}\binom{19}{5-x}}{\binom{25}{5}}$.

a. $P(\text{exactly two are sour}) = P(X = 2) = f(2) = 0.2736$
b. $P(\text{less than three are sour}) = P(X \leq 2) = f(0) + f(1) + f(2) = 0.9302$
c. $P(\text{more than three are sour}) = P(X \geq 4) = f(4) + f(5) = 0.0055$

2.3.2 Let X denote the number of captured birds that have been tagged. Then X has a hypergeometric distribution with $N_1 = 50$, $N_2 = 450$, $N = 500$, and $n = 10$ so that the pmf is $f(x) = \frac{\binom{50}{x}\binom{450}{10-x}}{\binom{500}{10}}$. Then $P(X = 4) = f(5) = 0.0104$.

2.3.3 Let X denote the number of chosen cubes that are not red. Then X has a hypergeometric distribution with $N_1 = 10$, $N_2 = 3$, $N = 13$, and $n = 3$ so that the pmf is $f(x) = \frac{\binom{10}{x}\binom{3}{3-x}}{\binom{13}{3}}$. Then $P(X = 2) = f(2) = 0.4721$.

2.3.4 Note that X has a hypergeometric distribution with $N_1 = 9$, $N_2 = 6$, $N = 15$, and $n = 5$ so that the pmf is $f(x) = \frac{\binom{9}{x}\binom{6}{5-x}}{\binom{15}{5}}$. Then by the definition of conditional probability,

$$P(X = 5 \mid X > 3) = \frac{P[(X = 5) \cap (X > 3)]}{P(X > 3)} = \frac{P(X = 5)}{P(X > 3)} = \frac{f(5)}{f(4) + f(5)} = 0.1429$$

2.3.5 Let X denote the number of games won out of the next 10. Then X is $b(10,\, 0.65)$ and its pmf is $f(x) = \binom{10}{x}0.65^x 0.35^{10-x}$. So the probability that the team wins eight of its next ten games is $P(X = 8) = f(8) = 0.1757$.

2.3.6 Let X denote the number of boys in a family with 7 children. Then X is $b(7,\, 0.5)$ and its pmf is $f(x) = \binom{7}{x}0.5^x 0.5^{7-x}$.

a. $P(\text{exactly four boys}) = f(4) = 0.2734$
b. $P(\text{more boys that girls}) = P(X \geq 4) = f(4) + f(5) + f(6) + f(7) = 0.5$
c. $P(\text{less than five boys}) = P(X \leq 4) = 1 - P(X \geq 5) = 1 - f(5) - f(6) - f(7) = 0.7734$

2.3.7 Let X denote the number of packages that weigh less than one pound out of 8. Then X is $b(8,\, 0.02)$ and its pmf is $f(x) = \binom{8}{x}0.02^x 0.098^{8-x}$. So

$$P(\text{at least } 1) = 1 - P(\text{none}) = 1 - P(X = 0) = 1 - f(0) = 0.1492.$$

2.3.8 Let X denote the number of passengers that arrive for the flight out of 25 ticket-holders. Then X is $b(25,\, 0.8)$ and its pmf is $f(x) = \binom{25}{x}0.8^x 0.2^{25-x}$. So

$$P(\text{not enough seats}) = P(X \geq 24) = f(24) + f(25) = 0.0274.$$

2.3.9 Let X denote the number of patients suffering a headache out of 50. Then X is $b(50,\, 0.14)$ and its pmf is $f(x) = \binom{50}{x}0.14^x 0.86^{50-x}$. So

$$P(10 \text{ or fewer}) = P(X \leq 10) = f(0) + \cdots + f(10) = 0.9176.$$

2.3.10 Let X denote the number of points selected such that $x \leq y$ out of 10. Figure 2.3 shows the rectangle $0 \leq x \leq 1$, $0 \leq y \leq 2$ and the portion in which $x \leq y$. We see that the total area of the region is 2 and the area of the portion in which $x \leq y$ is $2 - (1/2)(1)(1) = 3/2$. Thus the probability of selecting a point such that $x \leq y$ is $(3/2)/2 = 3/4$. This means that X is $b(10,\, 0.75)$ and its pmf is $f(x) = \binom{10}{x}0.75^x 0.25^{10-x}$. Thus

$$P(8 \text{ or more}) = P(X \geq 8) = f(8) + f(9) + f(10) = 0.5256.$$

2.3.11

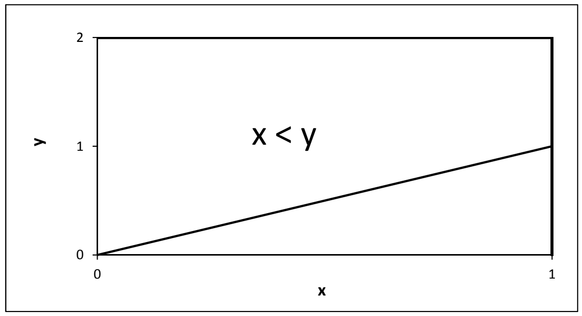


Figure 2.3

a. Let X denote the number of hits in 4 at-bats. Then X is $b(4,\, 0.33)$ and its pmf is $f(x) = \binom{4}{x} 0.33^x 0.67^{4-x}$. So

$$P(\text{at least one hit in a game}) = P(X \geq 1) = 1 - P(X = 0) = 1 - f(0) = 0.7985.$$

b. Let X denote the number of games with at least one hit in 56 games. Then X is $b(56,\, 0.7985)$ and its pmf is $f(x) = \binom{56}{x} 0.7985^x 0.2015^{56-x}$. So

$$P(\text{at least one hit in each of 56 games}) = P(X = 56) = f(56) = 0.000003.$$

2.3.12

a. Let X denote the number of chocolate candies out of 10 candies chosen. If the choices are independent, then X is $b(10,\, 0.2)$ and its pmf is $f(x) = \binom{10}{x} 0.2^x 0.8^{10-x}$. So

$$P(\text{choosing 6 or more chocolate pieces}) = P(X \geq 6) = f(6) + \cdots + f(10) = 0.0064.$$

b. Since the probability in part a. is less than 0.05, we conclude that the assumption is probably not correct. This means the father's claim is probably not valid. He appears to prefer chocolate candies.

c. We use the same random variable as in part a. Then

$$P(\text{choosing 1 or fewer chocolate pieces}) = P(X \leq 1) = f(0) + f(1) = 0.3758.$$

Since this probability is greater than 0.05, the assumption does not appear to be incorrect, so the father's claim appears to be valid.

2.3.13

a. If the claim were true, then we would expect to get about $0.15(50) = 7.5$ peanuts. Thus the two peanuts observed is less than expected.

b. Let X denote the number of peanuts out of 10 chosen nuts. If the claim were true, then X is $b(50,\, 0.15)$ and its pmf is $f(x) = \binom{50}{x} 0.15^x 0.85^{50-x}$. So

$$P(2 \text{ peanuts or fewer}) = P(X \leq 2) = f(0) + f(1) + f(2) = 0.0142.$$

c. Using complementary events,

$$P(2 \text{ peanuts or more}) = P(X \geq 2) = 1 - P(X \leq 1) = 1 - (f(0) + f(1)) = 0.9971.$$

d. Since the probability in part b. is less than 0.05, we conclude that the claim is probably not valid.

e. If x is less than expected, we need to calculate $P(X \leq x)$. If x is more than expected, we need to calculate $P(X \geq x)$.

2.3.14 We have 2 customers that buy shoes, 1 buys other equipment, and 2 that browse. By Definition 1.3.4, the number of ways these customers can be arranged is

$$\frac{5!}{2! \cdot 1! \cdot 2!} = 30$$

Now consider the outcome that the first 2 buy shoes, the third buys other equipment, and the last 2 browse. The probability of this outcome is $0.35^2 \cdot 0.25 \cdot 0.4^2 = 0.0049$. Thus the probability that exactly 2 buy shoes and 1 buys other sporting equipment is $30(0.0049) = 0.147$.

2.3.15 The probability that the first three are too light, the next two are too heavy, and the last 45 are within an acceptable range is $0.06^3 \cdot 0.08^2 \cdot 0.86^{45}$. The number of ways these weights can be arranged is $50!/(3! \cdot 2! \cdot 45!)$ so the probability that exactly three are too light, five are too heavy, and 45 are within an acceptable range is

$$\frac{50!}{3! \cdot 2! \cdot 45!} 0.06^3 \cdot 0.08^2 \cdot 0.86^{45} = 0.0330.$$

2.3.16 Note that $X = n$ means that the r^{th} success occurs on the n^{th} trial. This means there were $r-1$ successes and $n-r$ "failures" on the first $n-1$ trials and a success on the n^{th} trial. The probability of any one of such outcomes is

$$p^{r-1}(1-p)^{n-r}\, p^1 = p^r(1-p)^{n-r}.$$

To determine the number of such outcomes, we need to count the number of ways the $r-1$ successes can be arranged among the $n-1$ trials. The number of ways this can be done is $\binom{n-1}{r-1}$. Thus

$$f(n) = P(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r} \quad \text{for } n = r,\, r+1,\, \ldots$$

2.3.17

a. The only outcome in which the first red cube is obtained on the third selection is BBR. The probability of this occurring is $0.7^2 \cdot 0.3 = 0.147$.
b. The only way the first red cube can obtained on the 15^{th} selection is if the first 14 cubes are blue and the last is red. The probability of this occurring is $0.7^{14} \cdot 0.3 = 0.002$.
c. The only way the first success can occur on the x^{th} trial is if the first $x-1$ trials are failures and the last trial is a success. The probability of this occurring is $(1-p)^{x-1}p$. Thus the pmf is

$$f(x) = P(X = x) = (1-p)^{x-1}p \text{ for } x = 1, 2, \ldots$$

2.4 The Poisson Distribution

2.4.1

a. Most people arrive shortly before the movie starts. Few people arrive after the movie starts. So these occurrences are not randomly scattered over the interval f 6:30 P.M. to 7:15 P.M.
b. People generally consume calories in large batches at meal time. Over the interval of 6:00 A.M. to 10:00 P.M., a person typically consumes a large number of calories during breakfast, which may last only a few minutes, and then no other calories during the interval. So the calories are not randomly scattered over the interval.
c. It seems reasonable to assume the fertilizer granules are randomly scattered over the entire lawn.
d. It seems reasonable to assume that typos are randomly scattered throughout a line.
e. Some areas such as cities are densely populated while rural areas are not. Thus people are not randomly scattered across the entire United States.

2.4.2 Let X denote the number of customers who arrive during a one-hour interval of time. Then X has a Poisson distribution with parameter $\lambda = 25$ and

$$P(X = 20) = \frac{25^{20}}{20!} e^{-25} = 0.0520.$$

2.4.3 Let X denote the number of flaws in a 100 ft section of cloth. Then X has a Poisson distribution with parameter $\lambda = 5$ and

$$P(X = 0) = \frac{5^0}{0!} e^{-5} = 0.0067.$$

This this probability is very small, it would be "unusual" to receive a 100 ft section of cloth with no flaws from this factory.

2.4.4

a. The average number of births per day is $248/365 = 0.679$.

b. Let X denote the number of births in a day. Then X has a Poisson distribution with parameter $\lambda = 0.679$ and

$$P(X = 1) = \frac{0.679^1}{1!}\, e^{-0.679} = 0.344.$$

c. Using the same variable as in part a., we have

$$\begin{aligned} P(X < 3) &= P[(X = 0) \cup (X = 1) \cup (X = 2)] \\ &= \frac{0.679^0}{0!}\, e^{-0.679} + \frac{0.679^1}{1!}\, e^{-0.679} + \frac{0.679^2}{2!}\, e^{-0.679} \\ &= 0.968. \end{aligned}$$

d. Using complementary events,

$$P(X \geq 3) = 1 - P(X < 3) = 1 - 0.968 = 0.032.$$

2.4.5 Customers arrive at a supermarket checkout stand at an average of 3 per hour. Assuming the Poisson distribution applies, calculate the probabilities of the following events:

a. Let X denote the number of arrivals in an hour. Then X has a Poisson distribution with parameter $\lambda = 3$ and

$$P(X = 2) = \frac{3^2}{2!}\, e^{-3} = 0.224.$$

b. Let Y denote the number arrivals in an two-hour interval of time. Then Y has a Poisson distribution with parameter $\lambda = 6$ and

$$P(Y = 2) = \frac{6^2}{2!}\, e^{-6} = 0.0446.$$

c. If a total of exactly 2 people arrive between the hours of 9:00 A.M. and 10:00 A.M. or between 1:00 P.M. and 2:00 P.M., then we could have 2 people arrive during one of the hour intervals and 0 during the other, or 1 arrives during each interval. Using the variable from part a., and since the arrivals are independent, we have

$$\begin{aligned} P(\text{2 arrive during first interval and 0 during second}) &= P(X = 2) \cdot P(X = 0) \\ &= 0.224 \cdot 0.0498 = 0.01116, \\ P(\text{0 arrive during first interval and 2 during second}) &= P(X = 0) \cdot P(X = 2) \\ &= 0.01116, \quad \text{and} \\ P(\text{1 arrive during first interval and 1 during second}) &= P(X = 1) \cdot P(X = 1) \\ &= 0.1494 \cdot 0.1494 = 0.02232 \end{aligned}$$

Since the three events considered above are disjoint,

$$P(\text{total of exactly 2 people arrive}) = 0.01116 + 0.01116 + 0.02232 = 0.04464.$$

d. The answer to part c. is exactly the same as part b. This make sense because arrivals are independent and do not depend on the time of day. So the probability of an event happening over 2 disjoint one-hour intervals should be the same the event happening over a 2-hour interval of time.

2.4.6 The pmf of X is $f(x) = \frac{\lambda^x}{x!} e^{-\lambda}$ so

$$\begin{aligned} P(X = 7) &= P(X = 8) \\ \Rightarrow \frac{\lambda^7}{7!} e^{-\lambda} &= \frac{\lambda^8}{8!} e^{-\lambda} \\ \Rightarrow \frac{8!}{7!} &= \frac{\lambda^8}{\lambda^7} \\ \Rightarrow 8 &= \lambda \end{aligned}$$

and

$$P(X = 6) = \frac{8^6}{6!} e^{-8} = 0.122.$$

2.4.7 Let X denote the number of calls received during a 5-minute interval of time. Then X has a Poisson distribution and since there are 12 5-minute intervals in an hours, the parameter is $\lambda = 6/12 = 0.5$ and

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{0.5^0}{0!} e^{-0.5} = 0.393.$$

2.4.8 The average number of chips in brand A is 22.2 and the average in brand B is 24.15. The frequencies, relative frequencies, and theoretical probabilities calculated using the Poisson distribution and these averages are shown in the table below. We see that for neither brand are the relative frequencies very close to the theoretical probabilities. Thus the number of chocolate chips is neither brand is reasonably described by a Poisson distribution.

	x	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
	Freq	1	1	1	2	2	0	0	5	5	1	0	1	1	0	0
A	**Rel Freq**	0.05	0.05	0.05	0.10	0.10	0.00	0.00	0.25	0.25	0.05	0.00	0.05	0.05	0.00	0.00
	$f(x)$	0.04	0.05	0.06	0.07	0.08	0.08	0.09	0.08	0.08	0.07	0.06	0.05	0.04	0.03	0.02
	Freq	0	1	2	0	1	1	2	2	2	2	0	2	1	1	3
B	**Rel Freq**	0.00	0.05	0.10	0.00	0.05	0.05	0.10	0.10	0.10	0.10	0.00	0.10	0.05	0.05	0.15
	$f(x)$	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.08	0.08	0.08	0.07	0.07	0.06	0.05	0.04

2.4.9 The total number of knots observed is $0(28)+1(56)+\cdots+7(1) = 399$. Since there were a total of 200 boards, the average number of knots on a board is $\lambda = 399/200 \approx 2$. The frequencies, relative frequencies, and theoretical probabilities calculated using the Poisson distribution with $\lambda = 2$ are shown in the table below. We see that the relative frequencies are generally close to the theoretical probabilities. This indicates that the number of knots on a board is reasonably described by a Poisson distribution.

x	0	1	2	3	4	5	6	7
Freq	28	56	51	35	19	7	3	1
Rel Freq	0.14	0.28	0.26	0.18	0.10	0.04	0.02	0.01
$f(x)$	0.135	0.271	0.271	0.180	0.090	0.036	0.012	0.003

2.4.10 Since there are 100 small squares and a total of 30 dots, the average number of dots in a square is $\lambda = 30/100 = 0.3$. The frequencies, relative frequencies, and theoretical probabilities calculated using the Poisson distribution with $\lambda = 0.3$ are shown in the table below. We see that the relative frequencies are generally close to the theoretical probabilities. This indicates that the number of dots in a randomly chosen small square is reasonably described by a Poisson distribution.

x	0	1	2	3
Freq	72	26	2	3
Rel Freq	0.72	0.26	0.02	0.03
$f(x)$	0.741	0.222	0.033	0.003

2.4.11 Let X denote the number of winners out of 2000 tickets. Then X is $b(2000, 0.001)$ and approximately Poisson with parameter $\lambda = 2000(0.001) = 2$. Notice that the rule-of-thumb conditions are met, so the Poisson distribution offers a reasonable approximation.

a. $P(X = 1) \approx \frac{2^1}{1!}\, e^{-2} = 0.271$

b. $P(4 \le X \le 6) \approx \frac{2^4}{4!}\, e^{-2} + \frac{2^5}{5!}\, e^{-2} + \frac{2^6}{6!}\, e^{-2} = 0.138$

c. $P(X < 3) \approx \frac{2^0}{0!}\, e^{-2} + \frac{2^1}{1!}\, e^{-2} + \frac{2^2}{2!}\, e^{-2} = 0.677$

2.4.12 Let X denote the number of students out of 200 that get the flu. Then assuming the 4% rate has not changed, X is $b(200, 0.04)$ and approximately Poisson with parameter $\lambda = 200(0.04) = 8$. Notice that the rule-of-thumb conditions are met, so the Poisson distribution offers a reasonable approximation. Thus

$$P(X \le 2) \approx \frac{8^0}{0!}\, e^{-8} + \frac{8^1}{1!}\, e^{-8} + \frac{8^2}{2!}\, e^{-8} = 0.0138.$$

Since this probability is small, it appears that the assumption of a 4% rate is incorrect. Thus it appears the vaccine program has been effective.

2.4.13 Let X denote the number of cars that arrive at the toll booth in a one-minute interval of time. Then X is Poisson with parameter $\lambda = 4$ so the probability that exactly 10 cars arrive in a one-minute interval of time is

$$P(X = 10) = \frac{4^{10}}{10!}\, e^{-4} = 0.00529.$$

Now let Y denote the number of one-minute intervals of time out of 1000 in which exactly 10 cars arrive. Then Y is $b(1000,\, 0.00529)$ and approximately Poisson with parameter $\lambda = 1000(0.00529) = 5.292$. Notice that the rule-of-thumb conditions are met, so the Poisson distribution offers a reasonable approximation of probabilities of Y. Then the probability that exactly 10 cars arrive in 8 out of 1000 different one-minute intervals of time is

$$P(Y = 8) \approx \frac{5.292^8}{8!}\, e^{-5.292} = 0.0767.$$

2.4.14 The total area occupied by the cow paties is $20,000(0.5) = 10,000$ ft^2. Thus the probability that a single step hits a patty is $10,000/(1000 \cdot 1000) = 0.01$. Let X denote the number of steps out of 400 that hit a patty. Then X is $b(400,\, 0.01)$ and approximately Poisson with parameter $\lambda = 400(0.01) = 4$. Notice that the rule-of-thumb conditions are met, so the Poisson distribution offers a reasonable approximation. Thus

$$P(X \geq 1) = 1 - P(X = 0) \approx 1 - \frac{4^0}{0!}\, e^{-4} = 0.982.$$

2.4.15 Let X have a Poisson distribution with parameter λ.

a. Note that for $x = 1,\, 2,\, \ldots,$

$$\begin{aligned} \frac{P(X = x)}{P(X = x - 1)} &= \frac{\frac{\lambda^x}{x!}\, e^{-\lambda}}{\frac{\lambda^{x-1}}{(x-1)!}\, e^{-\lambda}} \\ &= \frac{\lambda^x}{x!} \cdot \frac{(x-1)!}{\lambda^{x-1}} \\ &= \frac{\lambda}{x}. \end{aligned}$$

b. If $P(X = x) \geq P(X = x - 1)$, then by part a.,

$$1 \leq \frac{P(X = x)}{P(X = x - 1)} = \frac{\lambda}{x} \quad \Rightarrow \quad x \leq \lambda.$$

c. By part b., $P(X = x) \geq P(X = x - 1)$ for all $x \leq \lambda$. By a similar argument, we can show that $P(X = x) \leq P(X = x - 1)$ for all $x > \lambda$. Let $f(x)$ denote the pmf of X and $\lfloor \lambda \rfloor$ denote the greatest integer less than or equal to λ. The above arguments show that

$$f(0) \leq \cdots f\left(\lfloor \lambda \rfloor - 1\right) \leq f\left(\lfloor \lambda \rfloor\right) > f\left(\lfloor \lambda \rfloor + 1\right) \geq \cdots$$

Thus f attains its maximum value at $X = \lfloor \lambda \rfloor$, and by definition, the mode of X is $\lfloor \lambda \rfloor$.

2.4.16 Let X denote the number of calls received in a one-hour interval of time. Then X has a Poisson distribution with parameter $\lambda = 8.4$.

a. By part c. of exercise 2.4.15, the mode of this distribution is $\lfloor 8.4 \rfloor = 8$. This is the most likely number of calls received in an hour.
b. By part a. of exercise 2.4.15,

$$\frac{P(X=8)}{P(X=7)} = \frac{8}{8} = 1 \quad \Rightarrow \quad P(X=8) = P(X=7).$$

showing that receiving 8 calls in an hour is just as likely as receiving 7 calls.

2.5 Mean and Variance

2.5.1

a. $\mu = 0(0.1) + \cdots + 4(0.1) = 2.3$,
$\sigma^2 = (0-2.3)^2(0.1) + \cdots + (4-2.3)^2(0.1) = 1.21$,
$\sigma = \sqrt{1.21} = 1.1$
b. $\mu = -2(1/16) + \cdots + 15(1/16) = 6$,
$\sigma^2 = (-2-6)^2(1/16) + \cdots + (15-6)^2(1/16) = 112/3 = 37.33$,
$\sigma = \sqrt{112/3} = 6.11$
c. $\mu = 3(1) = 3$,
$\sigma^2 = (3-3)^2(1) = 0$,
$\sigma = \sqrt{0} = 0$
d. The values of the pmf are $f(0) = 1/4$, $f(1) = 1/2$, and $f(2) = 1/4$ so that
$\mu = 1(1/2) + 2(1/4) = 1$,
$\sigma^2 = (0-1)^2(1/4) + (1-1)^2(1/2) + (2-1)^2(1/4) = 1/2$,
$\sigma = \sqrt{1/2} = 0.707$

2.5.2

a. The pmf is shown in the table below.

x	0	1	2	3	4
y	-1	2	5	8	11
$f(y)$	0.1	0.1	0.3	0.4	0.1

b. $\mu = -1(0.1) + \cdots + 11(0.1) = 5.9$
c. We see that $E(Y) = 3E(X) - 1$.

2.5.3 The volumes of the boxes are 3 ft^3, 21 ft^3, and 60 ft^3, so the mean volume is $3(0.35) + 21(0.45) + 60(0.2) = 22.5$ ft^3.

2.5.4 Note that X has a uniform distribution with range $R = \{1, \ldots, 9\}$, so $E(X) = (9+1)/2 = 5$ and $Var(X) = (9^2 - 1)/12 = 20/3$.

2.5.5 To find the mean, we need to find the parameter λ. Let X denote the number of yellow candies in a bag and $f(x)$ denote its pmf We are told that $P(X = 0) = 0.0183$. Thus

$$f(0) = \frac{\lambda^0}{0!} e^{-\lambda} = 0.0183 \quad \Rightarrow \quad \lambda = -\ln 0.0183 \approx 4$$

Thus the mean number of yellow candies in a bag is 4.

2.5.6 Suppose X is $b(5,\, 0.20)$.

a. The pmf is given in the table below.

$\boldsymbol{x}$	0	1	2	3	4	5
$\boldsymbol{f(x)}$	0.3277	0.4096	0.2048	0.0512	0.0064	0.0003

b. $\mu = 0(0.3277) + \cdots + 5(0.0003) = 1$, $\sigma^2 = (0-1)^2(0.3277) + \cdots + (5-1)^2(0.0003) = 0.8$

c. Note that $\mu = 5(0.2) = np$ and $\sigma^2 = 5(0.2)(0.8) = np(1-p)$.

2.5.7 The probability of getting a heart on either selection is 1/4. The probability of getting 0 hearts is $(3/4)(3/4) = 9/16$, the probability of getting 1 heart is $2(1/4)(3/4) = 3/8$, and the probability of getting 2 hearts is $(1/4)(1/4) = 1/16$. Let X denote the number of hearts selected. The pmf of X is given in the table below.

$\boldsymbol{x}$	0	1	2
$\boldsymbol{f(x)}$	9/16	3/8	1/16

Thus the mean and variance of the number of hearts selected are

$$\mu = 0(9/16) + 1(3/8) + 2(1/16) = 1/2 \text{ and}$$

$$\sigma^2 = (0 - 1/2)^2(9/16) + (1 - 1/2)^2(3/8) + (2 - 1/2)^2(1/16) = 3/8.$$

2.5.8 The pmf of X is given in the table below.

$\boldsymbol{x}$	0	1
$\boldsymbol{f(x)}$	$1-p$	p

Then $E(X) = 0(1-p) + 1(p) = p$ and $Var(X) = (0-p)^2(1-p) + (1-p)^2(p) = p\,(1-p)$.

2.5.9 The probability that X takes values "close" to 5 is large, so it has a relatively small standard deviation. The probability that Z takes values "far" from 5 is large, so it has a relatively large standard deviation. Z can take values "close" to 5 and "far" from 5 with relatively equal probability, so its standard deviation is somewhere between that of X and Z.

2.5.10 The male either "loses" 155 or "gains" $75,000 - 155 = 74,845$. The pmf of X is given in the table below.

$\boldsymbol{x}$	-155	74845
$\boldsymbol{f(x)}$	0.9985	0.0015

Then $E(X) = -155(0.9985) + 74845(0.0015) = -42.5$. This negative expected value means policy holders expect to "lose" money while the insurance company profits with the sale of many such policies.

2.5.11 Let X denote the profit from a single play of the game. Then $X =$ (value of spin) $- 2.75$. The pmf of X is given in the table below and

$$E(X) = -1.75(1/4) + \cdots 1.25(1/4) = -1/4.$$

Spin	1	2	3	4
$\boldsymbol{x}$	-1.75	-0.75	0.25	1.25
$\boldsymbol{f(x)}$	1/4	1/4	1/4	1/4

2.5.12 The variable X has a uniform distribution with parameter k. Then

$$E(X) = \frac{k+1}{2} = 9.5 \quad \Rightarrow \quad k = 18.$$

2.5.13 Suppose an office building contains 15 different offices. Ten have 20 employees each, four have 100 employees each, and one has 400 employees for a total of 1000 employees.

a. There are a total of 1000 employees and 15 offices, so the average number of employees per office is $1000/15 = 200/3 = 66.7$.
b. Since there are a total of $10(20) = 200$ employees that work in an office with 20 employees, the probability of selecting an employee from an office with 20 employees is 200/1000. Thus $f(20) = 200/1000$. The other values of f are calculated in a similar fashion. The pmf of X is given in the table below.

$\boldsymbol{x}$	20	100	400
$\boldsymbol{f(x)}$	200/1000	400/1000	400/1000

c. $E(X) = 20(200/1000) + 100(400/1000) + 400(400/1000) = 204.$

2.5.14 The expected value of X is

$$E(X) = \sum_{x=1}^{\infty} x \cdot \frac{6}{\pi^2 x^2} = \frac{6}{\pi^2} \sum_{x=1}^{\infty} \frac{1}{x}.$$

But this is a p-series with $p = 1$ which diverges. Thus this expected value does not exist.

2.5.15 Using the identity $\sum_{n=1}^{\infty} nx^{n-1} = \frac{x}{(1-x)^2}$ for $|x| < 1$, the expected value of T is

$$E(T) = \sum_{t=1}^{\infty} t \cdot \frac{1}{10} \left(\frac{9}{10}\right)^{t-1} = \frac{1}{10} \sum_{t=1}^{\infty} t \left(\frac{9}{10}\right)^{t-1} = \frac{1}{10} \cdot \frac{9/10}{(1-9/10)^2} = 9.$$

2.5.16 Suppose it costs \$$n$ to play the game, and let X denote the "profit" from one play of the game. Then $X =$ (result from die) $- n$. The pmf of X is given in the table below.

Die	1	2	3
$\boldsymbol{x}$	$1-n$	$2-n$	$3-n$
$\boldsymbol{f(x)}$	$3/6$	$2/6$	$1/6$

Then

$$E(X) = (1-n)(3/6) + (2-n)(2/6) + (3-n)(1/6) = (10-6n)/6.$$

For this to equal 0, we need $n = 10/6 \approx 1.67$. Thus a "fair" price for the game is \$1.67.

2.5.17

a. The game lasts n flips only if the first $n-1$ flips are heads and the n^{th} flip is tails. The probability of this is $(1/2)^{n-1}(1/2) = 1/2^n$. The pmf of X is given in the table below.

$\boldsymbol{n}$	1	2	3	4
$\boldsymbol{x}$	2^0	2^1	2^2	2^3
$\boldsymbol{f(x)}$	$1/2$	$1/2^2$	$1/2^3$	$1/2^4$

b. You win less than \$25 if the game lasts 4 flips or fewer. Thus

$$P(X \leq 25) = \frac{1}{2} + \cdots + \frac{1}{2^4} = 0.9375,$$

meaning we are very likely to lose money on this game so we would *not* want to play the game.

c. Note

$$E(X) = 2^0 \frac{1}{2} + 2^1 \frac{1}{2^2} + 2^2 \frac{1}{2^3} + \cdots = \sum_{n=1}^{\infty} 2^{n-1} \frac{1}{2^n} = \sum_{n=1}^{\infty} \frac{1}{2}.$$

But this series diverges to infinity, so $E(X)$ does not exist. This means the game has an infinite expected value.

d. Since the expected value of the game is so large, we might be willing to play the game regardless of the price. This does not agree with the answer in part b.

2.6 Functions of a Random Variable

2.6.1 $E(X) = 0(1/4) + \cdots + 3(1/4) = 13/8$

$E\left[X^2\right] = 0^2(1/4) + \cdots + 3^2(1/4) = 31/8$

$Var(X) = 31/8 - (13/8)^2 = 79/64$

$E\left[X^2 + 3X + 1\right] = E\left[X^2\right] + 3E(X) + 1 = 31/8 + 3(13/8) + 1 = 39/4$

$E\left[\sqrt{X+5}\right] = \sqrt{0+5}(1/4) + \cdots + \sqrt{3+5}(1/4) = 2.564$

2.6.2 The pmf of X in table form is shown below

$\boldsymbol{x}$	0	1	2
$\boldsymbol{f(x)}$	1/4	1/2	1/4

$E(X) = 0(1/4) + 1(1/2) + 2(1/4) = 1$

$E\left[X^2\right] = 0^2(1/4) + 1^2(1/2) + 2^2(1/4) = 3/2$

$Var(X) = 3/2 - 1^1 = 1/2$

$E\left[4X^2 + 6X - 9\right] = 4E\left[X^2\right] + 6E(X) - 9 = 4(3/2) + 6(1) - 9 = 3$

2.6.3

a. Note that X has a uniform distribution with parameter $k = 4$ so that $E(X) = (4+1)/2 = 5/2$.

b. Note that $E(Y) = E[X - n] = E(X) - n = 5/2 - n$.
c. For $E(Y)$ to equal 0, we need $n = 5/2$.

2.6.4

a. The new mean daily production will be twice the original mean daily production.
b. By Theorem 2.6.1
$$E(Y) = E[2X] = 2E(X) = 2\mu_X.$$
This does agree with the answer in part a.
c. Since the new daily production is higher, there are more possible values of the daily production so that its value will be more “spread out.” This means the variance of the new daily production will be higher than original.

d. By Theorem 2.6.2
$$Var(Y) = Var[2X] = 2^2 Var(X) = 4\sigma_X^2.$$
This shows that the new variance is four times the old variance which agrees with the answer to part c.

2.6.5 Let $\mu_X = E(X)$. By Theorem 2.6.1, $E[aX] = aE(X) = a\mu_X$, and by the definition of variance,

$$\begin{aligned} Var[aX] &= E\left[(aX - a\mu_X)^2\right] = E\left[a^2 (X - \mu_X)^2\right] \\ &= a^2 E\left[(X - \mu_X)^2\right] = a^2 Var(X). \end{aligned}$$

2.6.6 Since X takes on only one value, its pmf is $f(x) = 1$ for $x = b$. Thus $E(X) = b$ and

$$Var(X) = E\left[(X - b)^2\right] = (b - b)^2(1) = 0.$$

2.6.7

a. The range of Y is $R = \{a, a + 1, \ldots, k + a - 1\}$. Since each value of X is equally likely, each value of Y is also equally likely and the pmf of Y is $f(y) = 1/k$ for all $y \in R$.
b. Note that

$$\begin{aligned} E(Y) &= E[X + a - 1] = E(X) + a - 1 = \frac{k+1}{2} + a - 1 \\ &= \frac{k + 1 + 2a - 2}{2} = \frac{a + (k + a - 1)}{2} = \frac{a+b}{2}. \end{aligned}$$

2.6.8 Note that by exercise 2.6.7, $E(X) = (5 + 11)/2 = 8$ so that

$$E(Y) = E[120X - 200] = 120E(X) - 200 = 120(8) - 200 = 760.$$

2.6.9 Note that μ_X and σ_X^2 are constants. So by linearity properties of expected value,

$$E(Y) = E\left[\frac{X - \mu_X}{\sigma_X}\right] = \frac{1}{\sigma_X}\left[E(X) - E(\mu_X)\right] = \frac{1}{\sigma_X}(\mu_X - \mu_X) = 0$$

and by the definition of σ_X,

$$Var(Y) = E\left[(Y - 0)^2\right] = E\left[\left(\frac{X - \mu_X}{\sigma_X}\right)^2\right] = \frac{1}{\sigma_X^2} E\left[(X - \mu_X)^2\right] = \frac{1}{\sigma_X^2}\sigma_X^2 = 1.$$

2.6.10

a. $5 = E[X - 5] = E(X) - 5 \;\Rightarrow\; E(X) = 10$
b. $30 = E\left[(X - 5)^2\right] = E\left[X^2\right] - 10E(X) + 25 = E\left[X^2\right] - 75 \;\Rightarrow\; E\left[X^2\right] = 105$
c. $Var(X) = E\left[X^2\right] - [E(X)]^2 = 105 - 10^2 = 5$

2.6.11 Note that $Y = X - 6$ so that

$$E(Y) = E[X - 6] = E(X) - 6 = \frac{11+1}{2} - 6 = 0.$$

2.6.12 By Example 2.5.4, $E(X) = 3$ and

$$E\left[X^2\right] = \sum_{x \in R} x^2 f(x) = \lambda^2 + \lambda = 3^2 + 3 = 12.$$

Then

$$E(Y) = E\left[75 - 3X - X^2\right] = 75 - 3E(X) - E\left[X^2\right] = 75 - 3(3) - 12 = 54.$$

2.6.13 For example, take the variable in exercise 2.6.1 where we have $E(X) = 13/8$ but $E\left[X^2\right] = 31/8 \neq (13/8)^2$.

2.6.14 Consider the function $f(a) = E\left[(X - a)^2\right]$. The by expanding and using the linearity properties of expected value,

$$f(a) = E\left[(X - a)^2\right] = E\left[X^2\right] - 2aE(X) + a^2.$$

Taking the derivative with respect to a (using the fact that $E\left[(X - a)^2\right]$ and $E(X)$ are constants) and setting equal to 0 yields

$$\frac{df}{da} = -2E(X) + 2a = 0 \quad \Rightarrow \quad a = E(X).$$

Thus $a = E(X)$ is the only critical value of f. From the expression for df/da, we see that $df/da < 0$ for $a < E(X)$ and $df/da > 0$ for $a > E(X)$. Thus $a = E(X)$ is the location of the global minimum value of f.

2.6.15

a. The values of P for the different values of X are shown in the third column of the table below. Note that $P(P = -6) = P(X = 0) = 0.15$ so that the probabilities of the values of P are as given in the second column. The value of $E(P)$ is given in the last row of the table.
b. See the table below.

		Values of P			
x	$f(x)$	$s = 1$	$s = 2$	$s = 3$	$s = 4$
0	0.15	-6	-12	-18	-24
1	0.35	20	14	8	2
2	0.25	20	40	34	28
3	0.15	20	40	60	54
4	0.10	20	40	60	80
$E(P) =$		16.1	23.1	23.6	20.2

c. From the table, we see that the maximum expected profit occurs when $s = 3$. Thus the manager should order 3 steroes.

2.7 The Moment-Generating Function

2.7.1 In each part below, the mgf for a discrete random variable X is given. For each, describe the pmf of X in table form, find $M'(t)$ and $M''(t)$, and use these derivatives to calculate μ and σ^2.

a. The pmf is given in the table below.

$\boldsymbol{x}$	-1	3	4
$\boldsymbol{f(x)}$	1/5	2/5	2/5

Now,

$$M'(t) = -\frac{1}{5}\,e^{-t} + \frac{6}{5}\,e^{3t} + \frac{8}{5}\,e^{4t} \quad \text{and} \quad M''(t) = \frac{1}{5}\,e^{-t} + \frac{18}{5}\,e^{3t} + \frac{32}{5}\,e^{4t}$$

so that

$$\mu = M'(0) = 13/5 \quad \text{and} \quad \sigma^2 = M''(0) - \left[M'(0)\right]^2 = 51/5 - (13/5)^2 = 86/25.$$

b. The pmf is given in the table below.

$\boldsymbol{x}$	1	2	3	4	5
$\boldsymbol{f(x)}$	0.2	0.2	0.2	0.2	0.2

Now,

$$M'(t) = \sum_{x=1}^{5} x(0.2)e^{tx} \quad \text{and} \quad M''(t) = \sum_{x=1}^{5} x^2(0.2)e^{tx}$$

so that

$$\mu = M'(0) = 3 \quad \text{and} \quad \sigma^2 = M''(0) - \left[M'(0)\right]^2 = 11 - (3)^2 = 2.$$

2.7.2 Comparing $M(t) = \left(0.25 + 0.75e^t\right)^{25}$ to the mgf in Example 2.7.3, we see that this is the mgf of a random variable that is $b(25,\, 0.75)$.

2.7.3 By the definition of a mgf, $M(0) = E\left[e^{0\cdot X}\right] = E[1] = 1$.

2.7.4 By the definition of the mgf,

$$M(t) = E\left[e^{tX}\right] = e^{t(0)}p^0(1-p)^{1-0} + e^{t(1)}p^1(1-p)^{1-1} = (1-p) + p\,e^t.$$

Then

$$M'(t) = p\,e^t \quad \text{and} \quad M''(t) = p\,e^t$$

so that

$$E(X) = M'(0) = p \quad \text{and} \quad Var(X) = M''(0) - \left[M'(0)\right]^2 = p - p^2 = p(1-p).$$

2.7.5 By the rule for the derivative of the natural logarithm and the quotient rule,

$$R'(t) = \frac{M'(t)}{M(t)} \quad \text{and} \quad R''(t) = \frac{M(t)M''(t) - M'(t)M'(t)}{[M(t)]^2}$$

so that

$$\begin{aligned} R'(0) = \frac{M'(0)}{M(0)} = \frac{\mu}{1} = \mu \quad \text{and} \quad R''(0) &= \frac{M(0)M''(0) - M'(0)M'(0)}{[M(0)]^2} \\ &= \frac{M''(0) - [M'(0)]^2}{1^2} = \sigma^2. \end{aligned}$$

2.7.6 By the definition of $E(X)$,

$$E(X) = \sum_{x=1}^{\infty} x \frac{a}{x^2} = a \sum_{x=1}^{\infty} \frac{1}{x}.$$

But this is a divergent p-series, so $E(X)$ does not exist. If such a variable without an expected value had a mgf, then by the second part of Theorem 2.7.1, the expected value would exist and equal $M'(0)$. But this is a contradiction. So a variable without an expected value cannot have a mgf

2.7.7 By definition of the mgf and linearity properties of expected value,

$$M_Y(t) = E\left[e^{Yt}\right] = E\left[e^{(aX+b)t}\right] = E\left[e^{(at)X} e^{bt}\right] = e^{bt} E\left[e^{(at)X}\right].$$

But $E\left[e^{(at)X}\right] = M_X(at)$ so that $M_Y(t) = e^{bt} M_X(at)$ as desired.

2.7.8 By exercise 2.7.7, the mgf of $Y = 2X + 4$ is

$$M_Y(t) = e^{4t} M_X(2t) = e^{4t}\left(0.5 + 0.3e^{2t} + 0.2e^{6t}\right) = 0.5e^{4t} + 0.3e^{6t} + 0.2e^{10t}.$$

2.7.9 By exercise 2.7.7, the mgf of $Y = aX + b$ is $M_Y(t) = e^{bt} M_X(at)$ so that

$$\begin{aligned} M_Y'(t) &= b\,e^{bt} M_X(at) + a\,e^{bt} M_X'(at), \\ M_Y''(t) &= b^2\,e^{bt} M_X(at) + 2ab\,e^{bt} M_X'(at) + a^2\,e^{bt} M_X''(at), \\ E(Y) &= M_Y'(0) = b\,e^0 M_X(0) + a\,e^0 M_X'(0) = b + aE(X), \quad \text{and} \\ Var(Y) &= M_Y''(0) - \left[M_Y'(0)\right]^2 = b^2\,e^0 M_X(0) + 2ab\,e^0 M_X'(0) + a^2\,e^0 M_X''(0) - \left[b + aE(X)\right]^2 \\ &= b^2 + 2abE(X) + a^2 M_X''(0) - b^2 - 2abE(X) - a^2 E(X) \\ &= a^2\left[M_X''(0) - E(X)\right] \\ &= a^2 Var(X). \end{aligned}$$

2.7.10 Let $M_X(t) = [\,q + p\,e^t]^n$ denote the mgf of X. By exercise 2.7.7, the mgf of $F = n - X$ is $M_F(t) = e^{nt} M_X(-t)$ so that

$$M_F(t) = e^{nt} M_X(-t) = e^{nt}\left[\,q + p\,e^{-t}\right]^n = \left[e^t\left(q + p\,e^{-t}\right)\right]^n = \left[\,p + q\,e^t\right]^n.$$

This is the mgf of a random variable that is $b(n,\, q)$.

2.7.11

a. X is $b(50,\ 0.0956)$ and $E(X) = 50(0.0956) = 4.78$.
b. Note that $T = 50 + 10X$ so that $E(T) = 50 + 10E(X) = 97.8$.
c. By combining samples, the lab expects to run about 100 tests. If they tested each individual sample, they would have to run exactly 500 tests. So they are eliminating about 400 tests by combining samples.
d. In general, $E(T) = 50 + 10E(X)$. For $E(T)$ to be less than 500, we need $E(X) < 45$. Now let p_1 denote the probability that a combined sample tests positive. Then $E(X) = 50p_1$. For $E(X)$ to be less than 45, we need $p_1 < 45/50 = 0.9$. Generalizing the results of Exercise 1.7.10, the probability that a combined sample tests positive is $p_1 = 1 - (1-p)^{10}$. Thus we need

$$1 - (1-p)^{10} < 0.9 \quad \Rightarrow \quad p < 0.206.$$

Thus the method of testing combined samples saves work for $p < 0.206$.

2.7.12 By the definition of the mgf,

$$M(t) = \left[e^{Xt}\right] = \sum_{x=1}^{\infty} e^{xt} q^{x-1} p = \sum_{x=1}^{\infty} e^{t} p \left(q\, e^{t}\right)^{x-1}.$$

Now, this is a geometric series which converges if $|e^t q| < 1 \Rightarrow t < \ln(1/q)$ so that

$$M(t) = \frac{p\, e^t}{1 - qe^t} \quad \text{for} \quad t < \ln\left(\frac{1}{q}\right).$$

Then, using the quotient rule and after algebraic simplification,

$$M'(t) = \frac{p\, e^t}{\left(1 - qe^t\right)^2} \quad \text{and} \quad M''(t) = \frac{p\, e^t\left(1 + qe^t\right)}{\left(1 - qe^t\right)^3}.$$

Then using the fact that $1 - q = p$, we get

$$E(X) = M'(0) = \frac{p\, e^0}{\left(1 - qe^0\right)^2} = \frac{p}{(1-q)^2} = \frac{1}{p}, \quad \text{and}$$

$$Var(Y) = M''(0) - \left[M'(0)\right]^2 = \frac{p\, e^0\left(1 + qe^0\right)}{\left(1 - qe^0\right)^3} - \left(\frac{1}{p}\right)^2$$

$$= \frac{p\,(1+q)}{p^3} - \frac{1}{p^2} = \frac{q}{p^2}.$$

2.7.13 Note that

$$M'(t) = r\left(\frac{p\, e^t}{1 - qe^t}\right)^{r-1}\left(\frac{\left(1 - qe^t\right) p\, e^t - p\, e^t\left(-q\, e^t\right)}{\left(1 - qe^t\right)^2}\right)$$

$$= r\left(\frac{p\, e^t}{1 - qe^t}\right)^{r-1}\left(\frac{p\, e^t}{\left(1 - qe^t\right)^2}\right)$$

$$= r\left(\frac{p\, e^t}{1 - qe^t}\right)^{r}\left(\frac{1}{1 - qe^t}\right).$$

Then using the fact that $1-q=p$, we get

$$E(X)=M'(0)=r\left(\frac{p}{1-q}\right)^r\left(\frac{1}{1-q}\right)=\frac{r}{p}.$$

2.7.14 The pmf of X is $f(x)=\frac{\lambda^x}{x!}e^{-\lambda}$. Then by the definition of the mgf and the Taylor series $e^{\lambda e^t}=\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!}$,

$$M(t)=E\left[e^{Xt}\right]=\sum_{x=0}^{\infty}e^{xt}\frac{\lambda^x}{x!}e^{-\lambda}=e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!}=e^{-\lambda}e^{\lambda e^t}=e^{\lambda(e^t-1)}.$$

2.7.15 Note that

$$M'(t)=e^{\lambda(e^t-1)}=e^{\lambda(e^t-1)}\left(\lambda e^t\right)=\lambda e^{t+\lambda(e^t-1)}\quad\text{and}$$
$$M''(t)=\lambda e^{t+\lambda(e^t-1)}\left(1+\lambda e^t\right)$$

so that

$$E(X)=M'(0)=\lambda e^{0+\lambda(e^0-1)}=\lambda,\quad\text{and}$$
$$Var(Y)=M_Y''(0)-\left[M_Y'(0)\right]^2=\lambda e^{0+\lambda(e^0-1)}\left(1+\lambda e^0\right)-\lambda^2=\lambda(1+\lambda)-\lambda^2=\lambda.$$

These derivatives were much easier to calculate than manipulating the infinite series in Example 2.5.4.

2.7.16

a. Assuming the claim is true, X is $b(100,0.2)$ so that $\mu=100(0.2)=20$ and $\sigma=\sqrt{100(0.2)(0.8)}=4$.
b. The "usual" minimum value is $20-2(4)=12$ and the "usual" maximum value is $20+2(4)=28$.
c. Since $X=10$ is less than the "usual" minimum value, it is considered "unusual." This indicates the claim may be incorrect.

2.7.17 The pmf of a random variable X with a Poisson distribution with parameter λ is $f(x)=\frac{\lambda^x}{x!}e^{-\lambda}$. Using the Taylor series $e^{\lambda t}=\sum_{x=0}^{\infty}\frac{(\lambda t)^x}{x!}$, the probability-generating function is

$$P(t)=\sum_{x=0}^{\infty}\frac{\lambda^x}{x!}e^{-\lambda}t^x=e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda t)^x}{x!}=e^{-\lambda}e^{\lambda t}=e^{\lambda(t-1)}.$$

Note that this Taylor series converges for all t, so $P(t)$ is defined for all t. Then

$$P'(t)=\lambda\,e^{\lambda(t-1)}\quad\Rightarrow\quad E(X)=P'(1)=\lambda\,e^{\lambda(1-1)}=\lambda.$$

2.7.18 The pmf of a random variable X that is $b(n, p)$ is $f(x) = \binom{n}{x} p^x q^{n-x}$ for $x = 0, 1, \ldots, n$. Using the binomial theorem $(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^{n-x} b^x$, the probability-generating function is

$$P(t) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} t^x = \sum_{x=0}^{n} \binom{n}{x} (p\,t)^x q^{n-x} = (q + p\,t)^n.$$

Note that this series is finite, so it converges for all t. Then using the fact that $q + p = 1$,

$$P'(t) = n(q + p\,t)^{n-1} p \;\Rightarrow\; E(X) = P'(1) = n(q+p)^{n-1} p = np.$$

2.7.19 The pmf of a random variable X with a geometric distribution is $f(x) = q^{x-1} p$ for $x = 1, 2, \ldots$ where $0 \le p \le 1$ and $q = 1 - p$. Then

$$P(t) = \sum_{x=1}^{\infty} q^{x-1} p\, t^x = \frac{p}{q} \sum_{x=1}^{\infty} (q\,t)(q\,t)^{x-1}.$$

But this is a geometric series which converges if $|q\,t| < 1 \Rightarrow |t| < 1/q$. For $|t| < 1/q$,

$$P(t) = \frac{p}{q}\left(\frac{q\,t}{1 - t}\right) = \frac{p\,t}{1 - q\,t}.$$

Then using the quotient rule,

$$P'(t) = \frac{p}{(1 - q\,t)^2} \;\Rightarrow\; E(X) = P'(1) = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

Chapter 3

Continuous Random Variables

3.1 Introduction

3.2 Definitions

3.2.1

a. Note that $(1-x)^4 \geq 0$ for all x, so $f(x) \geq 0$ for all x in its domain. Also,

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 5(1-x)^4 dx = -(1-x)^5\big|_0^1 = 0 + 1 = 1.$$

Thus the first two requirements for a pdf are met.

b. The graph of the pdf is shown in the left half of Figure 3.1.

c. The cdf is $F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_0^x 5(1-t)^4 dt = -(1-t)^5 + 1$ for $0 < x \leq 1$. The graph of the cdf is shown in the right half of Figure 3.1.

d. $P(0 < X \leq 0.5) = \int_0^{0.5} 5(1-x)^4 dx = 0.96875$, $P(X \leq 0) = 0$, $P(X > 0.25) = \int_{0.25}^1 5(1-x)^4 dx \approx 0.2373$

3.2.2

a. Note that all values of $f(x)$ are non-negative, so $f(x) \geq 0$ for all x in its domain. Also,

$$\int_{-\infty}^{\infty} f(x)\,dx = 0.25(0.5) + 0.15(2) + 0.35(1) + 0.25(0.9) = 1.$$

Thus the first two requirements for a pdf are met.

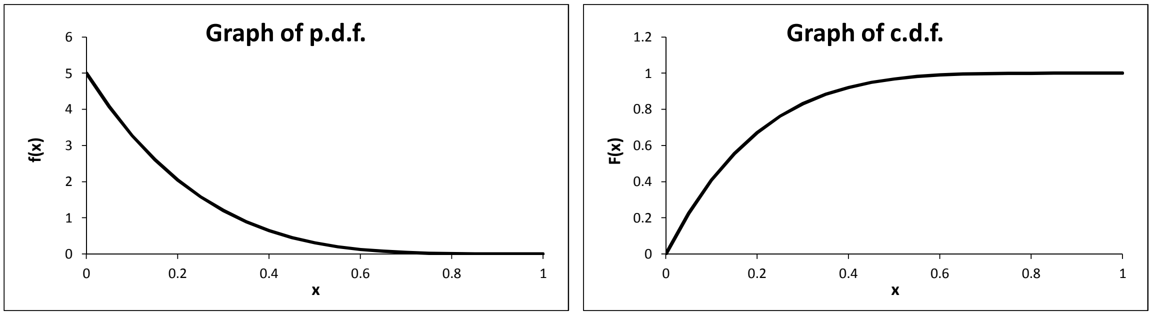


Figure 3.1

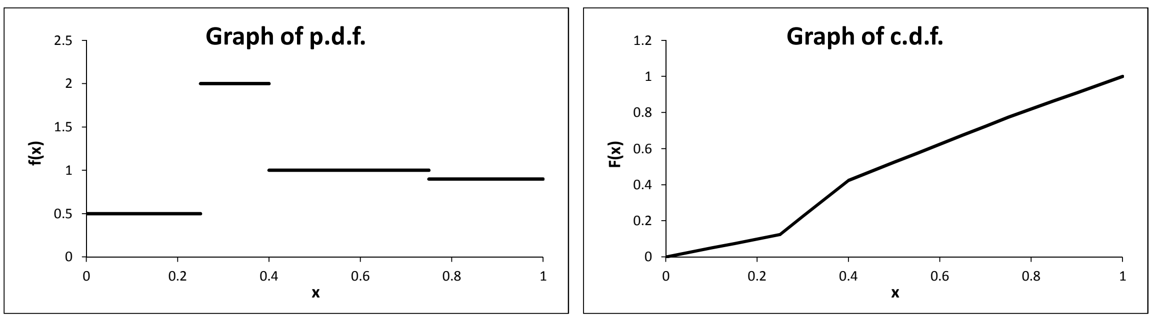


Figure 3.2

b. The graph of the pdf is shown in the left half of Figure 3.2.

c. From the graph of the pdf, we see that for $0 < x \leq 0.25$, $F(x) = 0.5x$. For $0.25 < x \leq 0.4$, $F(x) = 0.5(0.25) + 2(x - 0.25) = -0.375 + 2x$. The calculations for other intervals are similar. The cdf as a piecewise function is shown below. The graph of the cdf is shown in the right half of Figure 3.2.

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 0.5x & 0 < x \leq 0.25 \\ -0.375 + 2x & 0.25 < x \leq 0.4 \\ 0.025 + x & 0.4 < x \leq 0.75 \\ 0.1 + 0.9x & 0.75 < x \leq 1 \\ 1 & x > 1 \end{cases}$$

d. $P(0 < X \leq 0.5) = 0.5(0.25) + 2(0.15) + 1(0.1) = 0.525$, $P(X \leq 0) = 0$, $P(X > 0.25) = 1 - 0.5(0.25) = 0.875$

3.2.3

a. Note that $0.25 > 0$ and on the interval $0.5 \leq x1$, $-x + 2.5 > 0$, so $f(x) \geq 0$ for all x in its domain. Also,

$$\int_{-\infty}^{\infty} f(x)\, dx = 0.25(0.5) + \int_{0.5}^{1} -x + 0.25\, dx = 0.125 + 0.875 = 1.$$

Thus the first two requirements for a pdf are met.

b. The graph of the pdf is shown in the left half of Figure 3.3.

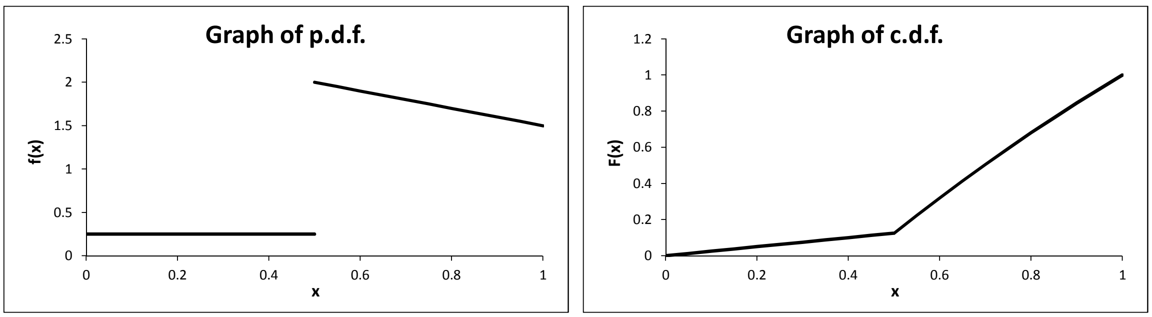


Figure 3.3

c. From the graph of the pdf, we see that for $0 < x \leq 0.5$, $F(x) = 0.25x$. For $0.5 < x \leq 1$, $F(x) = 0.25(0.5) + \int_{0.5}^{x} -t + 2.5\, dt = -0.5x^2 + 2.5x - 1$. The cdf as a piecewise function is shown below. The graph of the cdf is shown in the right half of Figure 3.3.

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 0.25x & 0 < x \leq 0.5 \\ -0.5x^2 + 2.5x - 1 & 0.5 < x \leq 1 \\ 1 & x > 1 \end{cases}$$

d. $P(0 < X \leq 0.5) = 0.25(0.5) = 0.125$, $P(X \leq 0) = 0$, $P(X > 0.25) = 1 - F(0.25) = 1 - 0.25(0.25) = 0.9375$

3.2.4

a. Note that $\int_0^a (5/6)\, x^3 dx = 5/24\, a^4$. For this to equal 1, we must have $a^4 = 24/5$ so that $a = (24/5)^{1/4} \approx 1.48$.

b. Note that $\int_{-a}^{a} 0.16x^2 dx = (8/75)\, a^3$. For this to equal 1, we must have $a^3 = 75/8$ so that $a = (75/8)^{1/3} \approx 2.109$.

c. Note that $\int_{-1}^{1} a\left(1 - x^4\right)\, dx = (8/5)\, a$. For this to equal 1, we must have $a = 5/8$.

d. Note that $\int_0^2 a\sqrt{3x}\, dx = (4\sqrt{6}/3)\, a$. For this to equal 1, we need $a = 3/(4\sqrt{6}) \approx 0.306$.

e. Note that $\int_1^3 a\, dx = 2a$. For this to equal 1, we need $a = 1/2$.

f. Note that $\int_0^{\infty} ae^{-2x} dx = a/2$. For this to equal 1, we need $a = 2$.

g. Note that $\int_{10}^{\infty} a/x^2 dx = a/10$. For this to equal 1, we need $a = 10$.

3.2.5

a. $\mu = \int_0^{10} x \cdot 1/10\, dx = 5$, $\sigma^2 = \int_0^{10} (x-5)^2 \cdot 1/10\, dx = 25/3$
b. $\mu = \int_0^2 x \cdot 3/8 \left(4x - 2x^2\right)\, dx = 2$, $\sigma^2 = \int_0^{10} (x-1)^2 \cdot 3/8 \left(4x - 2x^2\right)\, dx = 1/5$
c. $\mu = \int_0^{\infty} x \cdot 1/5 e^{-x/5}\, dx = 5$, $\sigma^2 = \int_0^{\infty} (x-5)^2 \cdot 1/5 e^{-x/5}\, dx = 25$
d. $\mu = \int_1^{\infty} x \cdot 3/x^4\, dx = 3/2$, $\sigma^2 = \int_1^{\infty} (x - 3/2)^2 \cdot 3/x^4\, dx = 3/4$

3.2.6

a. $F(x) = \int_{-1}^{x} 3/4 \left(1 - t^2\right)\, dt = -1/4 \left(x^3 - 3x - 2\right)$
b. To find p_1 we set $-1/4 \left(x^3 - 3x - 2\right) = 0.25$. Using software to solve this cubic polynomial for x yields three real solutions, 1.879, -0.347, and -1.532. The only one of the solutions that falls in the domain of the random variable is $x = -0.347$. To find p_2 and p_3, we set $F(x)$ equal to 0.5 and 0.75, respectively, yielding $p_2 = 0$ and $p_3 = 0.347$.

3.2.7

a. $F(x) = \int_5^x 5/t^2\, dt = (x-5)/x$
b. To find $\pi_{0.10}$, we set $(x-5)/x = 0.10$ and solve yielding $x = 50/9$. To find $\pi_{0.99}$, we set $(x-5)/x = 0.99$ and solve yielding $x = 500$.

3.2.8

a. $\mu = \int_{1000}^{1200} x \cdot 1/200\, dx = 1100$
b. $P(X > 1150) = \int_{1150}^{1200} 1/200\, dx = 1/4$
c. By the definition of conditional probability,

$$P(X > 1150 \mid X > 1100) = \frac{P[(X > 1150) \cap (X > 1100)]}{P(X > 1100)}.$$

But the only way $(X > 1150) \cap (X > 1100)$ can occur is if $X > 1150$, so $P[(X > 1150) \cap (X > 1100)] = P(X > 1150) = 1/4$. Also, $P(X > 1100) = \int_{1100}^{1200} 1/200\, dx = 1/2$. Thus

$$P(X > 1150 \mid X > 1100) = \frac{1/4}{1/2} = \frac{1}{2}.$$

3.2.9

a. The graph of the pdf is shown in the left half of Figure 3.4. We see that $f(x)$ is much larger near $x = 2$ than $x = 1$ or 3. Thus the daily demand is much more likely to be near 2,000 than 1,000 or 3,000.
b. $\mu = \int_1^3 x \cdot 3/4 \left(1 - (x-2)^2\right)\, dx = 2$. To find the mode, we set $f'(x) = -3(x-2)/2 = 0$ and solve, yielding $x = 2$.

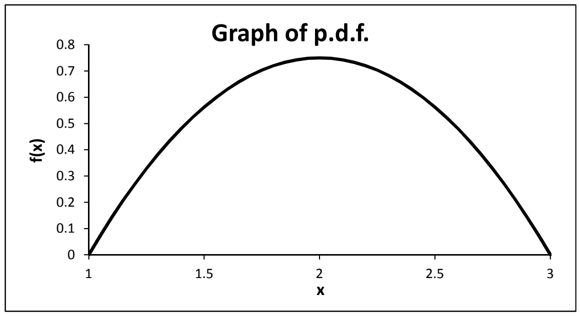


Figure 3.4

c. If the station has 2700 gallons of gasoline in stock at the beginning of a week, then there is not enough to meet demand only if $X > 2.7$. The probably of this is $P(X > 2.7) = \int_{2.7}^{3} 3/4\left(1-(x-2)^2\right)\, dx = 0.06075$.

d. We need $0.9 = P(X < a) = \int_1^a 3/4\left(1-(x-2)^2\right)\, dx = -1/4\left(a^3 - 6a^2 + 9a - 4\right)$. Solving this cubi equation for a using software yields three real solutions: 3.346, 2.608, and 0.0458. The only one of these solutions that lies in the domain of the random variable is $a = 2.608$. Thus the station must have 2608 gallons in stock.

3.2.10 $f(x) = F'(x) = 2x\,e^{-x^2}$ for $x \geq 0$

3.2.11

a. $P(X \leq 1.5) = F(1.5) = (1.5)^2/4 = 9/16$
b. $P(X \geq 0.5) = 1 - P(X < 0.5) = 1 - F(0.5) = 1 - 0.5/4 = 7/8$
c. $P(0.5 < X \leq 1.75) = F(1.75) - F(0.5) = (1.75)^2/4 - 0.5/4 = 41/64$
d. By the definition of conditional probability,

$$\begin{aligned} P(X \geq 0.5 \,|\, X \leq 1.5) &= \frac{P[(X \geq 0.5) \cap (X \leq 1.5)]}{P(X \leq 1.5)} \\ &= \frac{F(1.5) - F(0.5)}{F(1.5)} = \frac{(1.5)^2/4 - 0.5/4}{(1.5)^2/4} = \frac{7}{9}. \end{aligned}$$

e. Taking the derivative of each piece yields

$$f(x) = \begin{cases} 0 & x \leq 0 \\ 1/4 & 0 < x \leq 1 \\ x/2 & 1 < x \leq 2 \\ 0 & x > 2 \end{cases}$$

3.2.12 This means the boy weighs more than about 90% of other one-year olds.

3.2.13 For $t \neq 0$, $M(t) = \int_1^4 e^{xt} \cdot 1/3\, dx = \left(e^{4t} - e^t\right)/(3t)$. For $t = 0$, $M(0) = \int_1^4 e^{x0} \cdot 1/3\, dx = 1$.

3.2.14 Note that $M(t) = \int_0^\infty e^{xt} e^{-x}\, dx = \int_0^\infty e^{(t-1)x}\, dx$. But the exponent is non-negative for $t \geq 1$ which would mean that the integral does not converge. Thus the integral converges, and $M(t)$ exists only for $t < 1$. For $t < 1$, this integral converges to $1/(1-t)$.

3.2.15 By defintion and properties of expected value, $M(0) = E\left[e^{0x}\right] = E(1) = 1$.

3.2.16 By defintion, $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$. But $(x-\mu)^2 \geq 0$ and by definition of the pdf, $f(x) \geq 0$ for all x. Thus the integrand is non-negative for all x. By properties of the definite integral, the value of this integral, assuming it exists, will be non-negative.

3.2.17 By defintion,

$$P\left(x_0 - \frac{\epsilon}{2} < X < x_0 + \frac{\epsilon}{2}\right) = \int_{x_0-\epsilon/2}^{x_0+\epsilon/2} f(x)\, dx$$

But since f is continuous on $[x_0 - \epsilon/2,\, x_0 + \epsilon/2]$, by the mean value theorem for integrals, there exists an $x_1 \in [x_0 - \epsilon/2,\, x_0 + \epsilon/2]$ such that the integral equals $\epsilon \cdot f(x_1)$. But ϵ is "small" so that $[x_0 - \epsilon/2,\, x_0 + \epsilon/2]$ is "narrow." Thus x_1 is "close" to x_0 and $f(x_1) \approx f(x_0)$. Therefore,

$$P\left(x_0 - \frac{\epsilon}{2} < X < x_0 + \frac{\epsilon}{2}\right) = \epsilon \cdot f(x_1) \approx \epsilon \cdot f(x_0).$$

3.2.18

a. Tyson has a positive probability of parking any distance from the door. Since he looks for a spot close to the door, X_T may have a relatively large probability of being small. But since he sometimes has to park far away, X_T may have a relatively large probability of being large. Kylie only parks far away, so X_K is small with probability 0. Based on these arguments, Possible graphs of the pdf's are shown in Figure 3.5.

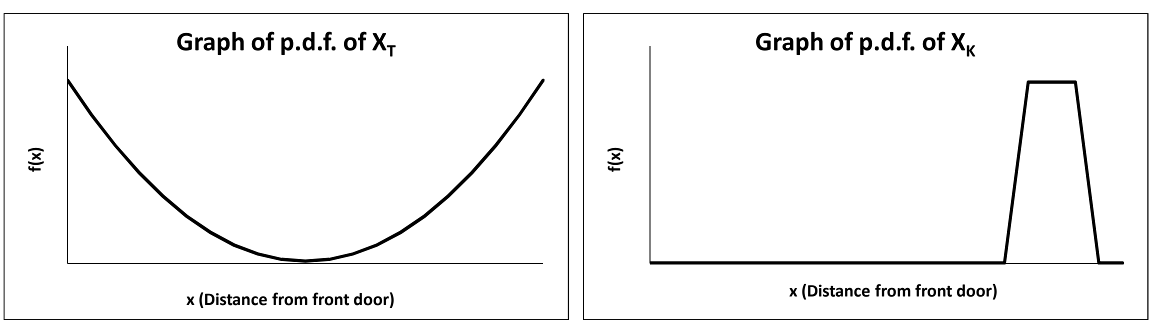


Figure 3.5

b. The mean of X_K is probably larger than the mean of X_T since Tyson sometimes parks close to the door while Kylie never does. The variance of X_T is probably larger than the variance of X_K since X_T can take on a wider range of values thatn X_K.

c. Since Kylie always parks in roughly the same spot everyday, it takes her about the same amount of time to walk from her car to the front door everyday so that the range of Y_K is narrow. This is not the case for Tyson. Some days he may have a short distance to walk and some days he has a long distance to walk. So the variance of Y_T is probably larger than the variance of Y_K.
d. Since it takes Kylie roughly the same amount of time to reach the front door everyday, she will have about the same amount of free-time before school everyday. Tyson will have a lot of time on some days, and not as much on other days.

3.3 The Uniform and Exponential Distributions

3.3.1

a. This does not have a uniform distribution because the dart is more likely to lad "far" from the center than "near" the center.
b. Since the students are instructed to randomly cut the string, it is reasonable to assume that X is uniform with $a = 0$ and $b = 6$.
c. The student is more likely to arrive "close" to the class start time than a long time before or after the start time. So X is more likely to take values close to 0 than other values. Thus is it not reasonable to assume that X has a uniform distribution.
d. Since they agree to meet *sometime* between 6:00 and 6:15 PM with no "target" time, it is reasonable to assume that X is uniform with $a = 0$ and $b = 15$.
e. Since they agree to meet at 6:00 PM, it is more likely she will arrive "close" to 6:00 than a long time before or after 6:00. So X is more likely to take values close to 0 than any other values. Thus is it not reasonable to assume that X has a uniform distribution.

3.3.2 We know from the text that $E(X) = (a + b)/2$ so that

$$
\begin{aligned}
Var(X) &= \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a}\,dx \\
&= \frac{1}{3(b-a)} \left(x - \frac{a+b}{2}\right)^3 \Bigg|_{x=a}^{x=b} \\
&= \frac{1}{3(b-a)} \left[\left(\frac{-a+b}{2}\right)^3 - \left(\frac{a-b}{2}\right)^3\right] \\
&= \frac{1}{3(b-a)} \left[\left(\frac{b-a}{2}\right)^3 + \left(\frac{b-a}{2}\right)^3\right] \\
&= \frac{1}{3(b-a)} \left(\frac{2(b-a)^3}{2^3}\right) \\
&= \frac{(b-a)^2}{12}.
\end{aligned}
$$

Also, for $t \neq 0$,

$$M(t) = \int_a^b e^{tx} \frac{1}{b-a}\, dx = \frac{1}{t(b-a)} e^{tx} \Big|_{x=a}^{x=b} = \frac{e^{tb} - e^{ta}}{t(b-a)}.$$

3.3.3

a. $E(X) = (25 + 15)/2 = 20$
b. $Var(X) = (25 - 15)^2/12 = 25/3$
c. $P(18 < X < 23) = (23 - 18)/(25 - 15) = 1/2$
d. $P(12 < X < 18) = (18 - 15)/(25 - 15) = 3/10$

3.3.4

a. $P(4X + 2 < 4) = P(X < 1/2) = 1/2$,
b. $P(3 < 4X + 2 < 5) = P(1/4 < X < 3/4) = 1/2$,

3.3.5 By subtracting c and dividing by $d - c$, we get

$$P(c + 0.5 < X \cdot (d - c) + c < c + 1.5) = P\left(\frac{0.5}{d-c} < X < \frac{1.5}{d-c}\right) = \frac{1}{d-c}$$

(note that $d - c > 1.5$ so that $1.5/(d - c) < 1$). Using the same algebra, we get

$$2P(c < X \cdot (d - c) + c < c + 0.5) = 2P\left(0 < X < \frac{0.5}{d-c}\right) = 2\left(\frac{0.5}{d-c}\right) = \frac{1}{d-c}.$$

Thus these two probabilities are indeed equal.

3.3.6 Since $-2 \leq X \leq 2$, $P(|X| > 1) = P[(-2 < X < -1) \cup (1 < X < 2)] = P(-2 < X < -1) + P(1 < X < 2) = 1/4 + 1/4 = 1/2$.

3.3.7

a. Note that $P(X < d) = \frac{d-a}{b-a}$. Setting this equal to 0.75 and solving for d yields $d = 0.75b + 0.25a$.

b. Note that $P(c < X) = \frac{b-c}{b-a}$. Setting this equal to 0.90 and solving for c yields $c = 0.9a + 0.1b$.

c. Note that since $0 < c < \frac{b-a}{2}$, $\frac{a+b}{2} - c > a$ and $\frac{a+b}{2} + c < b$ so that

$$P\left(\frac{a+b}{2} - c < X < \frac{a+b}{2} + c\right) = \frac{2c}{b-a}.$$

Setting this equal to 0.6 and solving for c yields $c = 0.3(b - a)$.

d. Note that $P(c < X < d) = \frac{d-c}{b-a}$. Setting this equal to 0.3 yields $d - c = 0.3(b - a)$. But there are an infinite number of values of d and c in the interval (a, b) that solve this equation.

3.3.8 The pdf of X is $f(x) = 1/4$ for $15 < x < 19$ and 0 otherwise. The engineer mistakely subtracted 16 and 14 and then multiplied by 1/4 obtaining 1/2. To correctly calculate the probability, we need to note that $P(14 < X < 16) = P(14 < X < 15) = 1/4$.

3.3.9

a. Let X denote the time (in minutes) past 6:00 AM the passenger arrives. Then X is $U(0, 40)$. Since the train arrives at 6:00, 6:20, and 6:40, the passenger will have to wait less than 5 minutes if $15 < X \leq 20$ or $35 < X \leq 40$ so that

$$\begin{aligned} P(\text{waiting longer than 5 min}) &= P[(15 < X \leq 20) \cup (35 < X \leq 40)] \\ &= P(15 < X \leq 20) + P(35 < X \leq 40) \\ &= 5/40 + 5/40 = 1/4. \end{aligned}$$

b. The passenger will have to wait more than 10 minutes if $0 < X < 10$ or $20 < X < 30$. This has probability of $10/40 + 10/40 = 1/2$.

3.3.10 Let X denote the time (in minutes) past 9:00 AM the call arrives. Then X is $U(0, 60)$ so that $P(20 < X < 45) = 5/60 = 5/12$.

3.3.11 Let X denote the time (in hours) between the arrivals of the calls. Since the number of calls received in an hour is described by a Poisson distribution with parameter $\lambda = 4$, X has an exponential distribution with $\lambda = 4$ so that $P(X > 20/60) = e^{-4(1/3)} = 0.264$.

3.3.12 Note that since $E(X) = 0.5$, the X has parameter $\lambda = 1/0.5 = 2$. The dart will hit the bullseye if X is less than the radius of the bullseye so that $P(\text{hitting the bullseye}) = P(X < 5/16) = 1 - e^{-2(5/16)} = 0.465$.

3.3.13

a. By the definition of conditional probability,

$$\begin{aligned} P(X > 2+1 \,|\, X > 1) &= \frac{P[(X > 2+1) \cap (X > 1)]}{P(X > 1)} = \frac{P(X > 3)}{P(X > 1)} \\ &= \frac{e^{-2(3)}}{e^{-2(2)}} = e^{-2(2)} = P(X > 2). \end{aligned}$$

b. By the definition of conditional probability,

$$\begin{aligned} P(X > a+b \,|\, X > b) &= \frac{P[(X > a+b) \cap (X > b)]}{P(X > b)} = \frac{P(X > a+b)}{P(X > b)} \\ &= \frac{e^{-\lambda(a+b)}}{e^{-\lambda(b)}} = e^{-\lambda(a))} = P(X > a). \end{aligned}$$

3.3.14

a. Since X is exponentially distributed with parameter $\lambda = 0.001$, $P(X > 200) = e^{-0.001(200)} = 0.819$.

b. Using the results from exercise 3.3.13, we get

$$P(X > 1100 \mid X > 900) = P(X > 200 + 900 \mid X > 900) = P(X > 200) = 0.819.$$

c. The results from part b. mean that the probability the bulb lasts at least another 200 hours after it has lasted 900 hours (call this probability p_1) is equal to the probably that a brand new bulb lasts at least 200 hours (call this probability p_2). This does not seem reasonable. It is more reasonable to believe that p_1 should be less than p_2. Thus the assumption that X has an exponential distribution does not seem reasonable.

3.3.15

a. To find $\pi_{0.25}$, we set $F(x) = 1 - e^{-1/4x} = 0.25$ and solve for x, yielding $\pi_{0.25} = 1.15$.
b. To find $\pi_{0.95}$, we set $F(x) = 1 - e^{-0.1x} = 0.95$ and solve for x, yielding $\pi_{0.95} = 29.96$.
c. To find π_p, we set $F(x) = 1 - e^{-\lambda x} = p$ and solve for x, yielding $\pi_p = \ln(1-p)/(-\lambda)$.

3.3.16 Note that the cdf of X is $F(x) = (x-a)/(b-a)$ for $a < x < b$.

a. To find $\pi_{0.12}$, we set $F(x) = (x-0)/(1-0) = 0.12$ and solve for x, yielding $\pi_{0.12} = 0.12$.
b. To find $\pi_{0.99}$, we set $F(x) = (x+8)/(15.5+8) = 0.99$ and solve for x, yielding $\pi_{0.99} = 15.265$.

c. To find π_p, we set $F(x) = (x-a)/(b-a) = p$ and solve for x, yielding $\pi_p = p(b-a) + a$.

3.3.17 To find the median of an exponential distribution, we set $F(x) = 1 - e^{-\lambda x} = 0.5$ and solve for x, yielding $m = \ln(0.5)/(-\lambda)$.

To find the median of a uniform distribution, we set $F(x) = (x-a)/(b-a) = 0.5$ and solve for x, yielding $m = 0.5(b-a) + a$.

3.3.18 The calculations are summarized in the table below and the relative frequency histogram is shown in Figure 3.6. We see that the relative frequencies are close to the theoretical probabilities and the histogram is "flat." This means that it is reasonable to assume that X is $U(5,\, 20)$.

Interval	(5, 8]	(8, 11]	(11, 14]	(14, 17]	(17, 20]
Freq	6	7	6	7	4
Rel Freq	0.20	0.23	0.20	0.23	0.13
Theo Prob	0.2	0.2	0.2	0.2	0.2

3.3.19 The mean of the data is $\bar{x} = 5.15$ so if X has an exponential distribution, then its parameter $\lambda \approx 1/5.15 = 0.194$. The calculations are summarized in the table below and the relative frequency histogram is shown in Figure 3.7. We see that the relative frequencies are not close to the theoretical probabilities and the histogram does not have the shape of an exponential density curve. This means that it is not reasonable to assume that X has an exponential distribution.

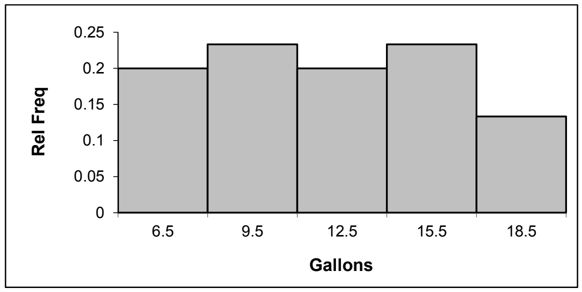


Figure 3.6

Interval	(0, 2]	(2, 4]	(4, 6]	(6, 8]	(8, 10]	(10, 12]
Freq	1	8	13	5	2	1
Rel Freq	0.03	0.27	0.43	0.17	0.07	0.03
Theo Prob	0.32	0.22	0.15	0.10	0.07	0.05

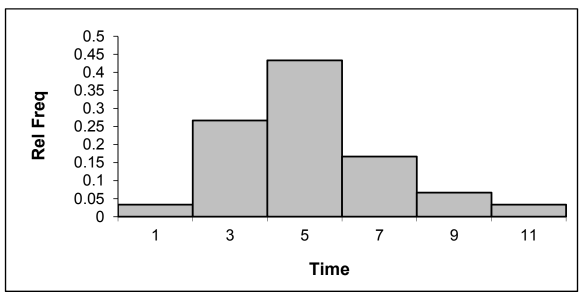


Figure 3.7

3.3.20

a. To find To find the skewness of the exponential distribution, first note that $\mu = \sigma = 1/\lambda$. Next note that

$$M'''(t) = \frac{d}{dt} M''(t) = \frac{d}{dt}\left(\frac{2\lambda}{(\lambda - t)^3}\right) = \frac{6\lambda}{(\lambda - t)^4}$$

so that $E\left[X^3\right] = M'''(0) = 6/\lambda^3$ and

$$\gamma_1 = \frac{6/\lambda^3 - 3(1/\lambda)(1/\lambda)^2 - (1/\lambda)^3}{(1/\lambda)^3} = \frac{2/\lambda^3}{(1/\lambda)^3} = 2.$$

b. To find the skewness of the uniform distribution, we have $\mu = (a+b)/2$ and $\sigma^2 = (b-a)^2/12$ so that

$$\gamma_1 = E\left[\left(\frac{X-(a+b)/2}{\sqrt{(b-a)^2/12}}\right)^3\right] = \left(\frac{12}{(b-a)^2}\right)^{3/2} E\left[\left(X - \frac{a+b}{2}\right)^3\right].$$

Now by definition of the expected value of a function of a random variable and elementary calculus,

$$E\left[\left(X - \frac{a+b}{2}\right)^3\right] = \int_a^b \left(X - \frac{a+b}{2}\right)^3 \frac{1}{b-a}\,dx = 0.$$

Thus the skewness is $\gamma_1 = 0$.

3.3.21 We have $\mu = \sigma = 1/\lambda$. By the results of the calculations immediately following Example 3.3.5, $E\left[X^2\right] = M''(0) = 2/\lambda^2$. By the solution to exercise 3.3.20, $E\left[X^3\right] = M'''(0) = 6/\lambda^3$. Also, using the results of this solution,

$$M''''(t) = \frac{d}{dt} M'''(t) = \frac{d}{dt}\left(\frac{6\lambda}{(\lambda-t)^4}\right) = \frac{24\lambda}{(\lambda-t)^5}$$

so that $E\left[X^4\right] = M''''(0) = 24/\lambda^4$. Thus

$$\gamma_2 = \frac{24/\lambda^4 - 4(1/\lambda)\left(6/\lambda^3\right) + 2(1/\lambda)^2\left(2/\lambda^2\right) + 5(1/\lambda)^4}{(1/\lambda)^4} - 3 = \frac{9/\lambda^4}{(1/\lambda)^4} - 3 = 6.$$

3.4 The Normal Distribution

Note: All probabilities were found using Table C.1.

3.4.1

a. $P(Z > 1.78) = 1 - 0.9625 = 0.0375$
b. $P(-1.54 < Z \le 0.98) = 0.8365 - 0.0618 = 0.7747$
c. $P(|Z| > 2.5) = P[(Z < -2.5) \cup (Z > 2.5)] = 2(0.0062) = 0.0124$
d. $P(Z^2 + 1 < 2.6) = P(Z^2 < 1.6) = P(|Z| < \sqrt{1.6}) = P(-1.26 < Z < 1.26) = 0.8962 - 0.1038 = 0.7924$

3.4.2

a. $P(X > 13.4) = P[Z > (13.4-12)/1.4] = P(Z > 1) = 1 - 0.8413 = 0.1587$
b. $P(10.5 < X \le 14.5) = P[(10.5-12)/1.4 < Z \le (14.5-12)/1.4] = P(-1.07 < Z \le 1.79) = 0.9633 - 0.1423 = 0.8210$
c. $P(|X-12| > 0.9) = P[(X - 12 < -0.9) \cup (X - 12 > 0.9)] = P[(X < 11.1) \cup (X > 12.9)] = P((Z < -0.64) \cup (Z > 0.64)] = 0.2611 + (1 - 0.7389) = 0.5222$

d. $P(X^2 + 10 < 110) = P(|X| < 10) = P(-10 < X < 10) = P(-15.71 < Z < -1.43) = 0.0764 - 0.0001 = 0.0763$

e. $P(X > 14 \,|\, X > 13.4) = \frac{P[(X>14)\cap(X>13.4)]}{P(X>13.4)} = \frac{P(X>14)}{P(X>13.4)} = \frac{P(Z>1.43)}{P(Z>1)} = \frac{1-0.9236}{1-0.8413} = 0.4814$

3.4.3

a. Note that $P(150 < X < 190) = P(-0.76 < Z < 0.62) = 0.7324 - 0.2236 = 0.5088$. Thus about 50.88% of people weigh between 150 and 190 pounds.

b. From Table C.1, we see that $P(Z < 1.645) \approx 0.95$. To find $\pi_{0.95}$, we then solve $1.645 = (x - 172)/29$ for x yielding $\pi_{0.45} = 168.4$.

c. From Table C.1, we see that $P(Z < -0.125) \approx 0.45$. To find $\pi_{0.45}$, we then solve $-0.125 = (x - 172)/29$ for x yielding $\pi_{0.95} = 219.7$.

3.4.4 Note that if $x = \mu + a\,\sigma$ where a is any number, then the z-score of x is $z = (\mu + a\,\sigma - \mu)/\sigma = a$. We use this formula to calculate the following probabilities:

a. $P(\mu - \sigma < X \le \mu + \sigma) = P(-1 < Z \le 1) = 0.8413 - 0.1587 = 0.6826 \approx 0.68$

b. $P(\mu - 2\sigma < X \le \mu + 2\sigma) = P(-2 < Z \le 2) = 0.9772 - 0.0228 = 0.9544 \approx 0.95$

c. $P(\mu - 3\sigma < X \le \mu + 3\sigma) = P(-3 < Z \le 3) = 0.9987 - 0.0013 = 0.9974 \approx 0.997$

3.4.5

a. Packages between 31.4 and 33.6 ounces are within one standard deviation of the mean, so about 68% of packages fall within this range.

b. Packages between 30.3 and 34.7 ounces are within two standard deviations of the mean, so about 95% of packages fall within this range.

c. Packages between 29.2 and 35.8 ounces are within three standard deviations of the mean, so about 99.7% of packages fall within this range.

3.4.6 The emperical rule says that $P(\mu - \sigma < X \le \mu + \sigma) \approx 0.68$. This means that $P(X \le \mu + \sigma) \approx 0.68 + 0.5(1 - 0.68) = 0.84$. But by definition, $\pi_{0.84}$ is a number such that $P(X \le \pi_{0.84}) = 0.84$. This shows that $\pi_{0.84} \approx \mu + \sigma$. By similar arguments, $\pi_{0.975} \approx \mu + 2\sigma$ and $\pi_{0.999} \approx \mu + 3\sigma$.

3.4.7 In general, to find a number $c > 0$ such that $P(-c < Z \le c) = p$, we look through Table C.1 for a number c such that $P(-c < Z) = 0.5(1 - p)$. This yields the solutions: a. $c = 1.645$, b. $c = 1.96$, c. $c = 2.575$.

3.4.8 Let X denote the actual amount spent on gasoline. Then X is $N\left(185,\, 10^2\right)$ so that

$$P(X > 200) = P(Z > (200 - 185)/10) = P(Z > 1.5) = 1 - 0.9332 = 0.0668.$$

Since this probability is small, the budgeted amount seems reasonable.

3.4.9 Let X denote the actual diameter of a randomly selected bolt. Then X is $N\left(3.002,\, 0.002^2\right)$. A bolt is accepted if $2.995 < X < 3.005$ so that

$$\begin{aligned} P(\text{A bolt is scrapped}) &= 1 - P(\text{A bolt is accepted}) \\ &= 1 - P(2.995 < X < 3.005) = 1 - P(-3.5 < Z < 1.5) \\ &= 1 - (0.9332 - 0.0001) = 0.0669. \end{aligned}$$

This means tha about 6.69% of all bolts will be scrapped.

3.4.10 We need to find μ such that $P(X < 12) = 0.015$. Examining Table C.1, we see that the corresponding z-score is -2.17. Thus we need $-2.17 = (12 - \mu)/0.1$. Solving this for μ yields $\mu = 12.217$.

3.4.11 Note that

$$\begin{aligned} P(125 < Y \leq 175) &= P\left(125 < e^X \leq 175\right) \\ &= P\left(\leq 125 < X \leq \ln 175\right) \\ &= P(-0.17 < Z < 0.16) = 0.5636 - 0.4325 = 0.1311 \end{aligned}$$

3.4.12 The calculations are summarized in the table below and the relative frequency histogram is shown in Figure 3.8. We see that the relative frequencies are close to the theoretical probabilities and the histogram resembles the shape of a bell-curve. This means that it is reasonable to assume that the height has a normal distribution.

Interval	$(-\infty, 61]$	(61, 63]	(63, 65]	(65, 67]	$(67, \infty)$
Freq	3	6	11	8	2
Rel Freq	0.100	0.200	0.367	0.267	0.067
Theo Prob	0.149	0.256	0.307	0.201	0.087

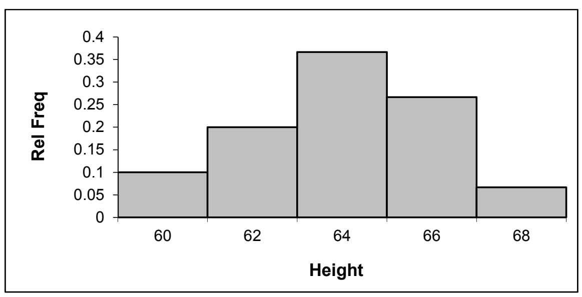


Figure 3.8

3.4.13 The mean of a normally distributed random variable equals the horizontal location of the peak of the bell-curve. Based on this, we see that Y has the largest mean and X has the smallest. The variance is determined by the shape of the bell-curve. The shorter and "fatter" the curve is, the higher the variance. Based on this, we see that Z has the largest variance and X has the smallest.

3.4.14 First note that from the text, we have

$$E\left(X^2\right) = M''(0) = \sigma^2 + \mu^2 = 1^1 + 6^2 = 37.$$

Now, the length of one piece of spaghetti is X and the length of the other piece is $10 - X$ so that the area of the rectangle is $A = X(10 - X) = 10X - X^2$. Then by properties of expected value,

$$E(A) = E\left[10X - X^2\right] = 10E(X) - E\left(X^2\right) = 10(6) - 37 = 23.$$

3.4.15

a. Note that $f(-z) = \frac{1}{\sqrt{2\pi}}\, e^{-(-z)^2/2} = \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2} = f(z)$.

b. Note that

$$f(\mu - h) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(\mu-h-\mu)^2/\left(2\sigma^2\right)} = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(-h)^2/\left(2\sigma^2\right)} = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-h^2/\left(2\sigma^2\right)} \text{ and}$$
$$f(\mu + h) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(\mu+h-\mu)^2/\left(2\sigma^2\right)} = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-h^2/\left(2\sigma^2\right)}.$$

Thus $f(\mu - h) = f(\mu + h)$ as desired.

c. By definition, the median of a random variable X that is $N\left(\mu,\, \sigma^2\right)$ is a number m such that $P(X < m) = 0.5$. From Table C.1, we see that the corresponding z-score is 0 so that $0 = (m - \mu)/\sigma$. Solving this for m, we get $m = \mu$. Thus the median of the normal distribution is μ.

d. The mode is a value of x for which $f(x)$ is maximized. We have

$$\frac{d}{dx}\, f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/\left(2\sigma^2\right)} \left(\frac{-2(x-\mu)}{2\sigma^2}\right).$$

Setting this equal to 0 and solving for x yields $x = \mu$. Since the sign of $f'(x)$ is determined by the sign of the fraction on the right-end, we see that $f'(x) > 0$ for $x < \mu$ and $f'(x) < 0$ for $x > \mu$. Thus by the first-derivative test, μ is the location of a local maximum. This proves that μ is the mode of the normal distribution.

e. As shown in part d., the maximum value of $f(x)$ occurs at $x = \mu$. But $f(\mu) = 1/\left(\sigma\sqrt{2\pi}\right)$. This is the maximum value of $f(x)$.

3.4.16 The mgf of X is $M_X(t) = \exp\left\{\mu t + \frac{\sigma^2}{2}\, t^2\right\}$. By exercise 2.7.7, the mgf of $Z = \frac{X-\mu}{\sigma} = \frac{1}{\sigma}\, X + \frac{-\mu}{\sigma}$ is

$$\begin{aligned}
M_Z(t) &= \exp\left\{\frac{-\mu}{\sigma}\, t\right\} M_X\left(\frac{1}{\sigma}\, t\right) \\
&= \exp\left\{\frac{-\mu}{\sigma}\, t\right\} \exp\left\{\mu\left(\frac{1}{\sigma}\, t\right) + \frac{\sigma^2}{2}\left(\frac{1}{\sigma}\, t\right)^2\right\} \\
&= \exp\left\{\frac{-\mu}{\sigma}\, t + \mu\left(\frac{1}{\sigma}\, t\right) + \frac{\sigma^2}{2}\left(\frac{1}{\sigma}\, t\right)^2\right\} \\
&= \exp\left\{\frac{t^2}{2}\right\}.
\end{aligned}$$

But this is the mgf of a random variable with a standard normal distribution. Thus Z has a standard normal distribution which proves Theorem 3.4.1.

3.4.17 Note that

$$\frac{df}{dx} = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/\left(2\sigma^2\right)} \left(\frac{-2(x-\mu)}{2\sigma^2}\right) \quad \text{and}$$
$$\frac{d^2f}{dx^2} = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/\left(2\sigma^2\right)} \left[\left(\frac{-2(x-\mu)}{2\sigma^2}\right)^2 + \left(\frac{-2}{2\sigma^2}\right)\right]$$

The sign of $\frac{d^2f}{dx^2}$ is determined by the expression in the square brackets. Simplifying this we get

$$\left[\left(\frac{-2(x-\mu)}{2\sigma^2}\right)^2 + \left(\frac{-2}{2\sigma^2}\right)\right] = \frac{(x-\mu)^2-\sigma^2}{\sigma^4}.$$

Setting this equal to 0 yields $(x-\mu)^2 = \sigma^2$ which has solutions $x = \mu \pm \sigma$. A quick check of the signs shows that $\frac{d^2f}{dx^2} > 0$ for $x < \mu - \sigma$ and $x > \mu + \sigma$ and that $\frac{d^2f}{dx^2} < 0$ for $\mu - \sigma < x < \mu + \sigma$. Thus $\frac{d^2f}{dx^2}$ changes signs at $x = \mu \pm \sigma$.

3.4.18 First note that

$$M''(t) = e^{t^2/2}\left(1+t^2\right),$$
$$M'''(t) = e^{t^2/2}\left(t^3+3t\right), \quad \text{and}$$
$$M''''(t) = e^{t^2/2}\left(t^4+6t^2+3\right)$$

so that $E\left[X^2\right] = M''(0) = 1$, $E\left[X^3\right] = M'''(0) = 0$, and $E\left[X^4\right] = M''''(0) = 3$. Thus the skewness is

$$\gamma_1 = \frac{E\left[X^3\right] - 3\mu\sigma^2 - \mu^3}{\sigma^3} = \frac{0 - 3(0)(1) - 0^3}{1^3} = 0$$

and the kurtosis is

$$\gamma_2 = \frac{E\left[X^4\right] - 4\mu E\left[X^3\right] + 2\mu^2 E\left[X^2\right] + 5\mu^4}{\sigma^4} - 3 = \frac{3 - 4(0)(0) + 2(0)^2(1) + 5(0)^4}{1^4} - 3 = 0.$$

3.4.19 Plugging $z = 0.1236$ into this formula yields $P(Z \le 0.1236) \approx 0.54917$.

3.4.20 By basic algebra and properties of limits, we have

$$\begin{aligned}
\lim_{z\to\infty} \frac{P\left(Z > z + \frac{a}{z}\right)}{P(Z > z)} &= \lim_{z\to\infty} \frac{1 - \Phi\left(z + \frac{a}{z}\right)}{1 - \Phi(z)} \\
&= \lim_{z\to\infty} \frac{\frac{1}{(z - a/z)\sqrt{2\pi}}\, e^{-(z-a/z)^2/2}}{\frac{1}{z\sqrt{2\pi}}\, e^{-z^2/2}} \\
&= \lim_{z\to\infty} \frac{z}{z - a/z}\, e^{-a}\, e^{-a^2/\left(2z^2\right)} \\
&= e^{-a}.
\end{aligned}$$

3.5 Functions of Continuous Random Variables

3.5.1 First note that $Y \geq (b-a)\cdot 0 + a = a$ and $Y \leq (b-a)\cdot 1 + a = b$, so $a \leq Y \leq b$. Now,

$$F(y) = P(Y \leq y) = P\left[(b-a)X + a \leq y\right] = P\left(X \leq \frac{y-a}{b-a}\right)$$

but $0 \leq \dfrac{y-a}{b-a} \leq 1$, so $P\left(X \leq \dfrac{y-a}{b-a}\right) = \dfrac{y-a}{b-a}$. Thus

$$f(y) = \frac{d}{dy}F(y) = \frac{d}{dy}\left(\frac{y-a}{b-a}\right) = \frac{1}{b-a}$$

which shows that Y is $U(a,\, b)$.

3.5.2 First note that since $0 < X < 1$, $0 < Y < 1$. Then for $0 < y < 1$,

$$F(y) = P(Y \leq y) = P\left(X^2 \leq y\right) = P\left(X \leq \sqrt{y}\right) = \int_0^{\sqrt{y}} 2x\,dx = y.$$

Thus $f(y) = F'(y) = 1$. Therefore, Y is $U(0,\, 1)$.

3.5.3 First note that $X > 0$ and $Y > 0$. Then for $y > 0$,

$$F(y) = P(Y \leq y) = P\left(X^2 \leq y\right) = P\left(X \leq \sqrt{y}\right) = \int_0^{\sqrt{y}} \frac{x}{4}\,e^{-x/2}\,dx$$

and

$$f(y) = \frac{dF}{dy} = \frac{\sqrt{y}}{4}\,e^{-\sqrt{y}/2}\left(\frac{1}{2}\,y^{-1/2}\right) = \frac{1}{8}\,e^{-\sqrt{y}/2}.$$

3.5.4 First note that $a < X < b$ and $a^2 \leq Y \leq b^2$. Then for $a^2 \leq y \leq b^2$

$$F(y) = P(Y \leq y) = P\left(X^2 \leq y\right) = P\left(X \leq \sqrt{y}\right) = \int_0^{\sqrt{y}} \frac{1}{b-a}\,dx$$

and

$$f(y) = \frac{dF}{dy} = \frac{1}{b-a}\left(\frac{1}{2}\,y^{-1/2}\right) = \frac{1}{2\sqrt{y}(b-a)}.$$

3.5.5 The area is $A = X^2$ by exercise 3.5.4, the pdf of A is $\frac{1}{2\sqrt{a}(3-2)}$ so that

$$P(5 < A < 6) = \int_5^6 \frac{1}{2\sqrt{a}}\,da = \sqrt{6} - \sqrt{5} \approx 0.213.$$

3.5.6 First note that since $0 < X < 1$, $-1 < Y < 1$. Then for $-1 < y < 1$,

$$F(y) = P(Y \le y) = P\left((1-2X)^3 \le y\right) = P\left(X \ge \frac{1-y^{1/3}}{2}\right) = \int_{\left(1-y^{1/3}\right)/2}^{1} 3(1-2x)^2\,dx$$

and

$$f(y) = \frac{dF}{dy} = -3\left[1 - 2\left(\frac{1-y^{1/3}}{2}\right)\right]^2 \cdot \left[\frac{1}{2}\left(-\frac{1}{3}y^{-2/3}\right)\right] = \frac{1}{2}.$$

Thus Y is $U(-1,\, 1)$.

3.5.7 First note that since $-4 \le X \le 5$, $0 \le Y \le 25$. Then

$$F(y) = P(Y \le y) = P\left(X^2 \le y\right) = P\left(|X| \le \sqrt{y}\right) = P\left(-\sqrt{y} \le X \le \sqrt{y}\right).$$

Now, if $0 \le y \le 16$, then $-\sqrt{y} \ge -4$ so that

$$P\left(-\sqrt{y} \le X \le \sqrt{y}\right) = \frac{2\sqrt{y}}{9}.$$

If $16 < y \le 25$, then $-\sqrt{y} < -4$ so that

$$P\left(-\sqrt{y} \le X \le \sqrt{y}\right) = P\left(-4 \le X \le \sqrt{y}\right) = \frac{\sqrt{y}+4}{9}.$$

Thus we have

$$F(y) = \begin{cases} 2\sqrt{y}/9 & 0 \le y \le 16 \\ (\sqrt{y}+4)/9 & 16 < y \le 25 \end{cases}$$

Since $f(y) = \frac{dF}{dy}$, we have

$$f(y) = \begin{cases} 1/(9\sqrt{y}) & 0 \le y \le 16 \\ 1/(18\sqrt{y}) & 16 < y \le 25 \end{cases}$$

3.5.8 First note that since $-1 \le X \le 5$, $0 \le Y \le 25$. Then

$$F(y) = P(Y \le y) = P\left(X^2 \le y\right) = P\left(|X| \le \sqrt{y}\right) = P\left(-\sqrt{y} \le X \le \sqrt{y}\right).$$

Now, if $0 \le y \le 1$, then $-\sqrt{y} \ge -1$ so that

$$P\left(-\sqrt{y} \le X \le \sqrt{y}\right) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{x^2}{42}\,dx = \frac{y^{3/2}}{63}.$$

If $1 < y \le 25$, then $-\sqrt{y} < -1$ so that

$$P\left(-\sqrt{y} \le X \le \sqrt{y}\right) = \int_{-1}^{\sqrt{y}} \frac{x^2}{42}\,dx = \frac{y^{3/2}+1}{126}.$$

Thus we have

$$F(y) = \begin{cases} y^{3/2}/63 & 0 \le y \le 11 \\ \left(y^{3/2}+1\right)/126 & 1 < y \le 25 \end{cases}$$

Since $f(y) = \frac{dF}{dy}$, we have

$$f(y) = \begin{cases} \sqrt{y}/42 & 0 \le y \le 1 \\ \sqrt{y}/84 & 1 < y \le 25 \end{cases}$$

3.5.9 The pdf of X is $f(x) = \dfrac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$ and the cdf of Y is

$$F(y) = P(Y \le y) = P\left(X^2 \le y\right) = P\left(-\sqrt{y} \le X \le \sqrt{y}\right) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2} dx.$$

So by the fundamental theorem of calculus,

$$\begin{aligned} f(y) &= \frac{d}{dy} \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \left[e^{-y/2} \left(\frac{1}{2}\, y^{-1/2} \right) - e^{-y/2} \left(-\frac{1}{2}\, y^{-1/2} \right) \right] \\ &= \frac{1}{\sqrt{2\pi y}}\, e^{-y/2}. \end{aligned}$$

3.5.10

a. Let $Z = 1 - Y$. Then since $0 \le Y \le 1$, $0 \le Z \le 1$. Then the cdf of Z is

$$F(z) = P(0 \le Z \le z) = P(0 \le 1 - Y \le z) = P(1 - z \le Y \le 1) = z.$$

Thus the pdf of Z is $\frac{d}{dz}(z) = 1$ which shows that $Z = 1 - Y$ is $U(0,\, 1)$.

b. The arguments of ln in both formulas are values of variables that are $U(0,\, 1)$. Thus both formulas will generate values of a random variable with the same distribution, exponential in this case.

3.5.11 Note that since $Y = e^X$, $Y > 0$. Then for any $y > 0$, the cdf of Y is

$$F(y) = P(Y \le y) = P\left(e^X \le y\right) = P(X \le \ln y) = \int_{-\infty}^{\ln y} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/\left(2\sigma^2\right)}\, dx$$

and

$$f(y) = \frac{dF}{dy} = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(\ln y - \mu)^2/\left(2\sigma^2\right)} \left(\frac{1}{y}\right).$$

3.5.12 First note that Y does have the range (0, 1) since if we choose any number p between 0 and 1, then Y can equal p since the equation $p = 1/((1+e^{-x})$ has the solution $x = -ln(1/p - 1)$. Now, the cdf of Y is

$$F(y) = P(Y \leq y) = P\left(\frac{1}{1+e^{-X}} \leq y\right) = P\left[X \geq \ln\left(\frac{1}{y} - 1\right)\right] = \int_{\ln(1/y-1)}^{\infty} \frac{e^{-x}}{(1+e^{-x})^2}\,dx$$

and the pdf is

$$\begin{aligned} f(y) = \frac{dF}{dy} &= -\frac{e^{-\ln(1/y-1)}}{\left(1+e^{-\ln(1/y-1)}\right)^2}\left(\frac{-1/y^2}{1/y-1}\right) \\ &= -\frac{(1/y-1)^{-1}}{(1+(1/y-1)^{-1})^2}\left(\frac{-1/y^2}{1/y-1}\right) \\ &= -\frac{(1/y-1)^{-1}}{\left(\frac{(1/y)^2}{(1/y-1)^2}\right)}\left(\frac{-1/y^2}{1/y-1}\right) \\ &= -\frac{(1/y-1)}{(1/y)^2}\left(\frac{-1/y^2}{1/y-1}\right) \\ &= 1. \end{aligned}$$

Thus Y is $U(0, 1)$.

3.5.13 First consider the case where $a > 0$. The cdf of Y is

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) = \int_{-\infty}^{(y-b)/a} f_X(x)\,dx$$

and since $|a| = a$, the pdf is

$$f_Y(y) = \frac{dF}{dy} = f_X\left(\frac{y-b}{a}\right)\left(\frac{1}{a}\right) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

If $a < 0$, when we derive the cdf, we reverse the inequality when dividing by the negative number a and get the cdf

$$F_Y(y) = P\left(X \geq \frac{y-b}{a}\right) = \int_{(y-b)/a}^{\infty} f_X(x)\,dx$$

and since $|a| = -a$, the pdf is

$$f_Y(y) = \frac{dF}{dy} = -f_X\left(\frac{y-b}{a}\right)\left(\frac{1}{a}\right) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

This establishes the result.

3.5.14 The pdf of X is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}.$$

By exercise 3.5.13, the pdf of $Y = aX + b$ is

$$\begin{aligned} f_Y(y) &= \frac{1}{|a|}\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-\left((y-b)/a-\mu\right)^2}{2\sigma^2}\right\} \\ &= \frac{1}{(|a|\sigma)\sqrt{2\pi}} \exp\left\{\frac{-\left(y-(a\mu+b)\right)^2}{2(a\sigma)^2}\right\}. \end{aligned}$$

But since $\sqrt{(a\sigma)^2} = |a|\sigma$, this pdf is that of a variable that is $N\left(a\mu+b,\, (a\sigma)^2\right)$.

3.5.15 First note that by the definition of a cdf, $0 < F(x) < 1$ so that $0 < Y < 1$. The cdf of Y is

$$F_Y(y) = P(Y \le y) = P(F_X(X) \le y).$$

Now let c be the unique number between a and b such that $F_X(c) = y$. Since F_X is continuous and strictly increasing, such a unique c exists. Then

$$P(F_X(X) \le y) = P(X \le c) = F_X(c) = y.$$

Thus $F_Y(y) = y$ and the pdf of Y is $f_Y(y) = 1$ which proves that Y is $U(0,\, 1)$.

3.5.16 Results will vary, but the histogram should indicate the numbers have a uniform distribution.

3.5.17

a. $F(x) = \int_0^x \frac{1}{2\sqrt{t}}\, dt = \sqrt{x}$ for $0 < x < 1$
b. To find the inverse cdf, we set $y = F(x)$ and solve for x yielding $x = y^2$. Thus $F^{-1}(y) = y^2$.

3.5.18

a. $F(x) = \int_0^x 2(1-t)\, dt = -x^2 + 2x$ for $0 < x < 1$
b. To find the inverse cdf, we set $y = F(x)$ obtaining $0 = -x^2 + 2x - y$. Solving for x using the quadratic formula yields $x = 1 \pm \sqrt{1-y}$. Since x must be between 0 and 1, the only one of these solutions that makes sense is $x = 1 - \sqrt{1-y}$. Thus $F^{-1}(y) = 1 - \sqrt{1-y}$.

3.6 Joint Distributions

3.6.1

a. The outcomes are shown in the table below.

	First Cube				
Second Cube	R_1	R_2	R_3	B_1	B_2
R_1	-	R_2R_1	R_3R_1	B_1R_1	B_2R_1
R_2	R_1R_2	-	R_3R_2	B_1R_2	B_2R_2
R_3	R_1R_3	R_2R_3	-	B_1R_3	B_2R_3
B_1	R_1B_1	R_2B_1	R_3B_1	-	B_2B_1
B_2	R_1B_2	R_2B_2	R_3B_2	B_1B_2	-

b. The joint and marginal pmf's are shown below.

			$\boldsymbol{x}$		
	$\boldsymbol{f(x, y)}$	**0**	**1**	**2**	$\boldsymbol{P(Y = y)}$
	0	0	0	6/20	6/20
$\boldsymbol{y}$	**1**	0	12/20	0	12/20
	2	2/20	0	0	2/20
	$\boldsymbol{P(X = x)}$	2/20	12/20	6/20	

c. Note that $P(X + Y \geq 2) = 1$ since $X + Y$ must equal 2. Values of X and Y that satisfy the inequality $2X - Y \leq 3$ are (1, 1) and (0, 2) so that $P(2X - Y \leq 3) = 12/20 + 2/20 = 7/10$

3.6.2

a. The joint and marginal pmf's are shown below.

			$\boldsymbol{x}$			
	$\boldsymbol{f(x, y)}$	**1**	**2**	**3**	**4**	$\boldsymbol{P(Y = y)}$
	1	1/16	0	0	0	1/16
$\boldsymbol{y}$	**2**	1/16	2/16	0	0	3/16
	3	1/16	1/16	3/16	0	5/16
	4	1/16	1/16	1/16	4/16	7/16
	$\boldsymbol{P(X = x)}$	4/16	4/16	4/16	4/16	

b. $P(X + Y = 4) = 0 + 2/16 + 1/16 = 3/16$ and $P(X \geq Y) = 1/16 + 2/16 + 3/16 + 4/16 = 5/8$.

3.6.3

a. Note that everything in the joint pmf is positive, so $f(x, y) \geq 0$ for all x and y in the range. Also,

$$\sum_{(x,\, y) \in R}\sum f(x,\, y) = 4/39 + 5/39 + 6/39 + 7/39 + 8/39 + 9/39 = 1.$$

Thus both properties are satisfied.

b. The marginal pmf's are

$$f_X(x) = \sum_{y=1}^{3} \frac{3x+y}{39} = \frac{3x+1}{39} + \frac{3x+2}{39} + \frac{3x+3}{39} = \frac{3x+2}{13} \quad \text{and}$$

$$f_Y(y) = \sum_{x=1}^{2} \frac{3x+y}{39} = \frac{3(1)+y}{39} + \frac{3(2)+y}{39} = \frac{9+2y}{39}$$

X and Y are not independent since $\frac{3x+2}{13} \cdot \frac{9+2y}{39} \neq \frac{3x+y}{39}$

c. Note

$$\begin{aligned} P(Y \geq X) &= \sum_{x=1}^{2} \sum_{y=x}^{3} \frac{3x+y}{39} \\ &= \frac{3(1)+1}{39} + \frac{3(1)+2}{39} + \frac{3(1)+3}{39} + \frac{3(2)+2}{39} + \frac{3(2)+3}{39} = \frac{32}{39}, \\ P(X+1 \leq Y) &= \sum_{x=1}^{2} \sum_{y=x+1}^{3} \frac{3x+y}{39} = \frac{3(1)+2}{39} + \frac{3(1)+3}{39} + \frac{3(2)+3}{39} = \frac{20}{39}, \quad \text{and} \\ P\left(X^2 \leq Y^2\right) &= P(X \leq Y) = \frac{32}{39} \end{aligned}$$

3.6.4

a. The marginal pdf of both variables is 1/2, so the joint pdf is $f(x, y) = 1/4$.

b. Note

$$\begin{aligned} P(Y \leq X) &= \int_{-1}^{1} \int_{-1}^{x} \frac{1}{4}\, dy\, dx = \frac{1}{2}, \\ P(Y \leq |X|) &= \int_{-1}^{0} \int_{-1}^{-x} \frac{1}{4}\, dydx + \int_{0}^{1} \int_{-1}^{x} \frac{1}{4}\, dy\, dx = \frac{3}{8} + \frac{3}{8} = \frac{3}{4} \quad \text{and} \\ P\left(X^2 + Y^2 \leq 1\right) &= \iint_R \frac{1}{4}\, dy\, dx = \frac{\pi}{4} \quad \text{where } R \text{ is a circle with radius 1} \end{aligned}$$

3.6.5

a. In the example, we calculated the probability that the man has to wait longer than five minutes is 2/9. By identical calculations, the probability that the woman has to wait longer than five minutes is also 2/9. Since the arrivals are idependent, the probability that the first to arrive has to wait longer than five minutes is $2/9 + 2/9 = 4/9$.

b. For the man to wait longer than ten minutes, he cannot arrive any later than 6:05, so

$$P(Y > X + 10) = \int_{0}^{5} \int_{x+10}^{15} \left(\frac{1}{15}\right)^2 dy\, dx = \frac{1}{18}.$$

c. If the woman arrives before the man, then $Y < X$ and

$$P(Y < X) = \int_0^{15} \int_0^{x} \left(\frac{1}{15}\right)^2 \, dy \, dx = \frac{1}{2}.$$

d. This scenario is almost identical to that in the example, except the marginal pdf's are 1/30 and the joint pdf is $(1/30)^2$. Also, if the man waits longer than 5 minutes, he cannot arrive any later than 6:25 so that

$$P(Y > X + 5) = \int_0^{25} \int_{x+5}^{30} \left(\frac{1}{30}\right)^2 \, dy \, dx = \frac{25}{72}.$$

3.6.6

a. The graphs are shown in Figure 3.9. We see that the man is more likely to arrive closer to 6:15 than 6:00 while the opposite is true for the woman.

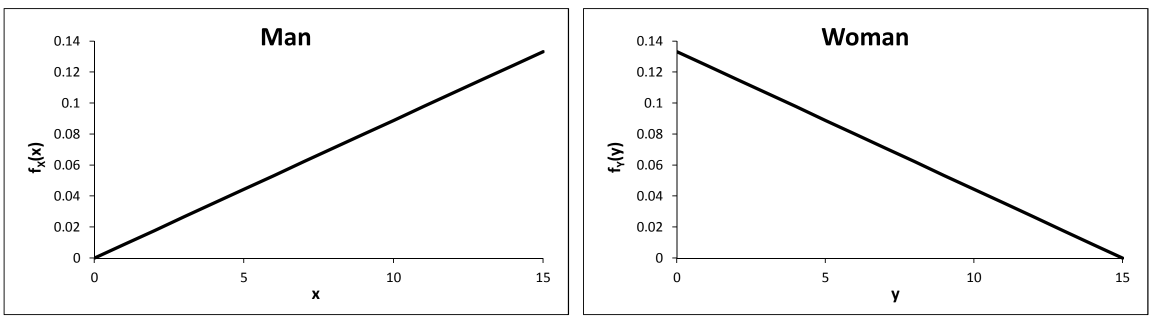


Figure 3.9

b. The joint pdf is the product of the marginal pdf's, so the probability that the man has to wait more than five minutes for the woman to arrive is

$$P(Y > X + 5) = \int_0^{10} \int_{x+5}^{15} \left(\frac{2}{225}\,x\right) \cdot \left(-\frac{2}{225}\,y + \frac{2}{15}\right) dy \, dx = \frac{8}{243}.$$

3.6.7 The marginal pdf's are

$$f_X(x) = \int_0^1 (x + 0.5)(y + 0.5) \, dy = x + 0.5 \text{ and}$$
$$f_Y(y) = \int_0^1 (x + 0.5)(y + 0.5) \, dx = y + 0.5$$

Since the product of the marginal pdf's equals the joint pdf, X and Y are independent.

3.6.8

a. Notice that everything in the joint pdf is positive, so $f(x, y) \geq 0$ for all x and y. Also, using elementary calculus, we can verify that

$$\int_0^1 \int_0^1 \frac{24}{11}\left(x^2 + \frac{xy}{2}\right) dy\, dx = 1.$$

b. The marginal pdf's are

$$f_X(x) = \int_0^1 \frac{24}{11}\left(x^2 + \frac{xy}{2}\right) dy = \frac{6x(4x+1)}{11} \quad \text{and}$$
$$f_Y(y) = \int_0^1 \frac{24}{11}\left(x^2 + \frac{xy}{2}\right) dx = \frac{2x(4x+1)}{11}$$

c. Note

$$P(Y > X) = \int_0^1 \int_x^1 \frac{24}{11}\left(x^2 + \frac{xy}{2}\right) dy\, dx = \frac{7}{22}.$$

3.6.9

a. The marginal pdf's are

$$f_X(x) = \int_0^1 6xy^2\, dy = 2x \quad \text{and}$$
$$f_Y(y) = \int_0^1 6xy^2\, dx = 3y^2$$

b. Note that $f_X(x)f_Y(y) = 2x\left(3y^2\right) = 6xy^2 = f(x, y)$ which proves that X and Y are independent.

c. Note

$$P(X + Y < 1) = P(Y < 1 - X) = \int_0^1 \int_0^{1-x} 6xy^2\, dy\, dx = \frac{1}{10}.$$

3.6.10

a. The marginal pdf's are

$$f_X(x) = \int_0^x \frac{1}{2}\, dy = \frac{x}{2} \quad \text{and}$$
$$f_Y(y) = \int_y^1 \frac{1}{2}\, dx = \frac{2-y}{2}$$

b. Note that $f_X(x)f_Y(y) = \frac{x}{2}\left(\frac{2-y}{2}\right) \neq \frac{1}{2} = f(x, y)$ which proves that X and Y are dependent.

c. Note

$$P(Y < X/2) = \int_0^1 \int_0^{x/2} \frac{1}{2}\, dy\, dx = \frac{1}{2}.$$

3.6.11

a. Note that both X and Y take values of 1, 2, and 3 with equal probability. The marginal distributions are shown in the table below.

$\boldsymbol{x, y}$	1	2	3
$\boldsymbol{f_X(x)}$	1/3	1/3	1/3
$\boldsymbol{f_Y(y)}$	1/3	1/3	1/3

b. Note that X and Y must have the same value. The joint distribution is shown in the table below.

			$\boldsymbol{x}$	
	$\boldsymbol{f(x, y)}$	**1**	**2**	**3**
	1	1/3	0	0
$\boldsymbol{y}$	**2**	0	1/3	0
	3	0	0	1/3

c. Note that $f(1, 1) = 1/3 \neq f_X(1)f_Y(1) = (1/3)(1/3) = 1/9$. Thus X and Y are not independent.

d. For any outcome of a random experiment, a random variable takes exactly one value. Thus, if we know the value of X, then we already know the value of X, so that X and X are not independent.

3.6.12 The marginal pdf's are all 1 so that the joint pdf is 1 and

$$P(X \geq YZ) = \int_0^1 \int_0^1 \int_{yz}^1 1\, dx\, dy\, dz = \frac{3}{4}.$$

3.6.13 Note that

$$P\left(x_0 - \frac{\epsilon}{2} < X < x_0 + \frac{\epsilon}{2},\, y_0 - \frac{\epsilon}{2} < Y < y_0 + \frac{\epsilon}{2}\right) = \iint_R f(x, y)\, dy\, dx$$

where R is the square $x_0 - \frac{\epsilon}{2} < y < x_0 + \frac{\epsilon}{2}$, $y_0 - \frac{\epsilon}{2} < y < y_0 + \frac{\epsilon}{2}$ on the $x - y$ plane. Geometrically, this double integral can be thought of as the volume of a 3-dimensional region bounded above by the curve $z = f(x, y)$ and whose base is the square R. The area of this square is ϵ^2. Since ϵ is a small number, the height of the region is approximately $f(x_0, y_0)$, the height at the center point of region R. Thus the volume of this region, and the value of the probability, is approximately $\epsilon^2 \cdot f(x_0, y_0)$.

3.6.14

a. Since both X and Y are $U(0, 1)$ and are independent, their joint pdf if $f(x, y) = 1$, so any probability involving X and Y is simply the area of the region of interest. Drawing the region $y \leq 0.5x$ inside the rectangle $0 \leq y \leq 1$, $0 \leq x \leq 1$ yields a right-triangle with vertices of $(0, 0)$, $(1, 0)$, and $(1, 0.5)$. The area of this triangle is $1/2 \cdot (1) \cdot (1/2) = 1/4$.

b. Drawing the region $1.5x \leq y < 2.5x$ yields a triangle with vertices of (0, 0), (2/5, 1), and (2/3, 1). Taking the leg of this triangle on the top of the square as the base, gives the base a length of $2/3 - 2/5 = 4/15$. The height of the triangle is the perpendicular distance from the base to the tip of the triangle, which is 1. Thus the area of the triangle is $1/2 \cdot (4/15) \cdot (1) = 2/15$.

3.6.15

a. Note that both marginal pdf's are 1 so that the joint pdfis also 1. Then

$$P(1.618 - 0.05 < Y/X < 1.618 + 0.05) = \iint_R 1\, dy\, dx$$

where R is the region illustrated in Figure 3.10. Since the integrand is a constant 1, the value of the integral equals the area of region R. This region is a triangle. Taking the "base" to be the leg on the line $y = 1$, the length of the base is

$$\frac{1}{1.568} - \frac{1}{1.668} \approx 0.03823.$$

The "height" is then the perpendiculat distance from the base to the tip of the triangle, which is 1 so that the area of the triangle is $0.5(0.03823)(1) = 0.01912$. This is the probability that the ratio Y/X is within 0.05 of 1.618. By similar calculations, the probability that the ratio X/Y is within 0.05 of 1.618 has the same value. Since these two events are disjoint,

$$P(\text{the ratio } Y/X \text{ or } X/Y \text{ is within 0.05 of 1.618}) = 2(0.01912) = 0.03823.$$

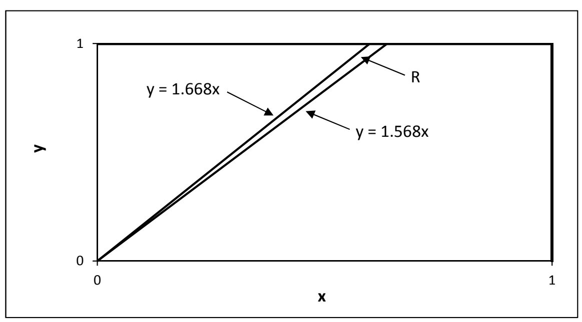


Figure 3.10

b. Using independence and complementary events, we get

$$\begin{aligned} P(\text{at least one pair is within 0.05}) &= 1 - P(\text{none is within 0.05}) \\ &= 1 - (1 - 0.03823)^{100} \\ &= 0.9797 \end{aligned}$$

c. Since the probability in part b. is so large, finding at least one pair of measurements that is "close" to the Golden Ratio is not unusual, even if the measurements are independent. This makes the claim somewhat questionable.

3.7 Functions of Independent Random Variables

3.7.1 By the definition of the expected value of a function of random variables, the expected values are

$$a.\ E[X+Y] = \int_0^1\int_0^1 (x+y)6xy^2\,dy\,dx = 17/12,$$
$$b.\ E[X] = \int_0^1\int_0^1 x6xy^2\,dy\,dx = 2/3,$$
$$c.\ E[Y] = \int_0^1\int_0^1 y6xy^2\,dy\,dx = 3/4,\ \text{ and}$$
$$d.\ E\left[X^2+Y\right] = \int_0^1\int_0^1 \left(x^2+y\right)6xy^2\,dy\,dx = 5/4.$$

3.7.2 By the definition of the expected value of a function of random variables, the expected values are

$$a.\ E[2X+Y] = \int_0^2\int_0^x (2x+y)(1/2)\,dy\,dx = 10/3,$$
$$b.\ E\left[3X^2+Y^3\right] = \int_0^2\int_0^x \left(3x^2+y^3\right)(1/2)\,dy\,dx = 34/5,\ \text{ and}$$
$$c.\ E[|Y-X|] = E[X-Y] = \int_0^2\int_0^x (x-y)(1/2)\,dy\,dx = 2/3\ \text{ since } X > Y.$$

3.7.3 We the region $0 \le x \le 15$, $0 \le y \le 15$ into two sub-regions as shown in Figure 3.11. So that

$$E[|X-Y|] = \int_0^{15}\int_0^x (x-y)(1/15)^2\,dy\,dx + \int_0^{15}\int_x^{15} (y-x)(1/15)^2\,dy\,dx = 75/2 + 5/2 = 5$$

3.7.4 The marginal pdf's are $f_X(x) = 1/8$ and $f_Y(y) = 1/10$ so that the joint pdf is 1/80 and

$$E\left[X^2Y+3\right] = \int_2^{10}\int_{10}^{20} \left(x^2y+3\right)(1/80)\,dy\,dx = 623.$$

3.7.5 The joint pdf of X and Θ is

$$f(x,\theta) = f_X(x)f_\Theta(\theta) = \frac{4}{\pi},\ \ 0 < x < 1/2,\ 0 < \theta < \pi/2.$$

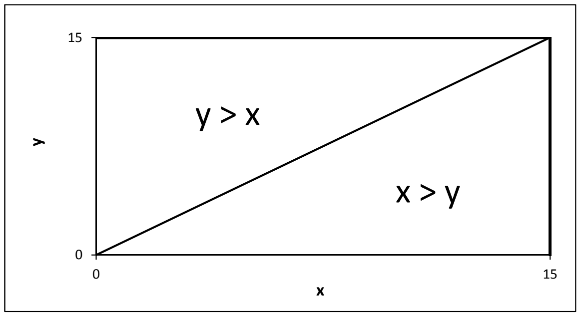


Figure 3.11

so that

$$\begin{aligned} E(h) &= \int_0^{1/2}\int_0^{\pi/2} \frac{x}{\cos\theta}\cdot\frac{4}{\pi}\,d\theta\,dx \\ &= \frac{4}{\pi}\int_0^{1/2} x\ln|\sec\theta+\tan\theta|\Big|_{\theta=0}^{\theta=\pi/2}\,dx \\ &= \frac{4}{\pi}\int_0^{1/2} x\left[\ln|\sec\pi/2+\tan\pi/2| - \ln|\sec 0+\tan 0|\right]\,dx \end{aligned}$$

But $\sec\pi/2$ and $\tan\pi/2$ do not exist, so this integral does not converge and $E(h)$ does not exist.

3.7.6

a. Let X denote the amount he gets from his maternal grandparents and Y the amount he gets from his paternal grandparents. Then the expected total amount is $E[X+Y] = E(X)+E(Y) = 12+15 = 27$.
b. Let X, Y, and Z denote the number of small, medium, and large baskets sold, respectively. Then the expected total profit is $E[5X+8Y+10Z] = 5E(X)+8E(Y)+10E(Z) = 5(25)+8(18)+10(9) = 359$.
c. Let X and Y denote the amount of ginger ale and fruit punch, respectively, poured into a bowl. Then since X and Y are independent, the variance of the volume of punch in the bowls is $Var(X+Y) = Var(X)+Var(Y) = 0.25^2+0.12^2 = 0.0769$.
d. Let X and Y denote the lengths of the long and short sides, respectively. Then the perimeter is $2X+2Y$, and since X and Y are independent, $2X$ and $2Y$ are also independent so that the variance of the perimeter is

$$Var(2X+2Y) = Var(2X)+Var(2Y) = 2^2Var(X)+2^2Var(Y) = 4(2^2)+4(0.5^2) = 17.$$

3.7.7 The variable X is not independent with itself, so we cannot apply the second part of Theorem 3.7.2 to $Var[X+X]$. The correct value of $Var[2X]$ is $2^2(5) = 20$. The first part of Theorem 3.7.1 does not require independence, so it does apply to $E[X+X]$.

3.7.8

a. The pmf is shown in the table below.

$\boldsymbol{x}$	0	1
$\boldsymbol{f(x)}$	$1-p$	p

From the distribution, we get

$$E(X_i) = 0(1-p) + 1(p) = p \text{ and}$$
$$Var(X_i) = (0-p)^2(1-p) + (1-p)^2(p) = p(1-p).$$

b. Note that since the trials are independent, then so are the X_i's. Thus by Theorem 3.7.1,

$$E(X) = E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = p + \cdots + p = np \text{ and}$$
$$Var(X) = Var(X_1) + \cdots + E(X_n) = p(1-p) + \cdots + p(1-p) = np(1-p).$$

3.7.9

a. The outcomes are shown in the table below.

	First Cube				
Second Cube	R_1	R_2	R_3	B_1	B_2
R_1	-	R_2R_1	R_3R_1	B_1R_1	B_2R_1
R_2	R_1R_2	-	R_3R_2	B_1R_2	B_2R_2
R_3	R_1R_3	R_2R_3	-	B_1R_3	B_2R_3
B_1	R_1B_1	R_2B_1	R_3B_1	-	B_2B_1
B_2	R_1B_2	R_2B_2	R_3B_2	B_1B_2	-

b. From the list of outcomes, we see that the first cube is red 12 times, the first cube is blue 8 times, the second cube is red 12 times, and the second cube is blue 8 times. The distributions of X_1 and X_2 are shown in the table below.

$\boldsymbol{x}$	0	1
$\boldsymbol{f_{X_1}(x)}$	8/20	12/20
$\boldsymbol{f_{X_2}(x)}$	8/20	12/20

From the distributions, we get

$$E(X_1) = 0(8/20) + 1(12/20) = 3/5 \text{ and}$$
$$E(X_2) = 0(8/20) + 1(12/20) = 3/5.$$

c. Note that

$$E(X) = E(X_1 + X_2) = E(X_1) + E(X_2) = 3/5 + 3/5 = 6/5.$$

d. The correct formula is $E(X) = nN_1/(N_1 + N_2)$.

3.7.10 Note that $P(X > Y) = P(X - Y > 0)$. But the variable $X - Y$ is

$$N\left(965 - 995,\ 19^2 + 31^2\right) = N(-30,\ 1322)$$

so that

$$P(X - Y > 0) = P\left(Z > \frac{0 - (-30)}{\sqrt{1322}}\right) = P(Z > 0.83) = 0.2033.$$

3.7.11

a. Let Y_i denote the weight of the i^{th} "3-pound" package selected, $i = 1, 2$. Then each Y_i is $N\left(3.11,\ 0.09^2\right)$. Then with X_i as defined in the example, we want to find

$$P(X_1 + \cdots + X_6 > Y_1 + Y_2) = P(X_1 + \cdots + X_6 - Y_1 - Y_2 > 0).$$

But this random variable is

$$N\left(6(1.05) - 2(3.11),\ 6(0.05^2) + 2(-1)^2(0.09^2)\right) = N(0.08,\ 0.0312)$$

so that

$$P(X_1 + \cdots + X_6 - Y_1 - Y_2 > 0) = P\left(Z > \frac{0 - 0.08}{\sqrt{0.0312}}\right) = P(Z > -0.45) = 0.6736.$$

b. Generalizing the calculations in part a., the variable $T_1 - T_2$ is

$$N\left(3n(1.05) - n(3.11),\ 3n(0.05^2) + n(-1)^2(0.09^2)\right) = N(0.04n,\ 0.0156n)$$

so that

$$P(T_1 - T_2 > 0) = P\left(Z > \frac{0 - 0.04n}{\sqrt{0.0156n}}\right) \approx P(Z > -0.32\sqrt{n}).$$

c. As $n \to \infty$, $-0.32\sqrt{n} \to -\infty$ so that $P(Z > -0.32\sqrt{n}) \to 1$. They would want to buy the 1-pound packages.

d. Generalizing the calculations in part b., the variable $T_1 - T_2$ is

$$N\left(3n\mu_X - n\mu_Y,\ 3n\sigma_X^2 + n\sigma_Y^2\right)$$

so that

$$P(T_1 - T_2 > 0) = P\left(Z > \frac{0 - (3n\mu_X - n\mu_Y)}{\sqrt{3n\sigma_X^2 + n\sigma_Y^2}}\right)$$

The value of this probability is broken down into three cases:

(a) If $3\mu_X = \mu_Y$, then this z-score is 0 so the probability is 0.5.

(b) If $3\mu_X > \mu_Y$, then this z-score is negative, so the probability $\to 1$ as $n \to \infty$.

(c) If $3\mu_X < \mu_Y$, then this z-score is positive, so the probability $\to 0$ as $n \to \infty$.

3.7.12 Let Y denote the weight of the "3-pound" package. Then the weight of one of the bags made from this package is $Y/3$. If X denotes the weight of a randomly chosen "1-pound" package, we want to find $P(Y/3 > X) = P(Y - 3X > 0)$. The distributions of X and Y are as given in Example 3.7.6 so that the distribution of $Y - 3X$ is

$$N\left(1(3.11) - 3(1.05),\ 1^2(0.09^2) + (-3)^2(0.05^2)+\right) = N(-0.04,\ 0.0306)$$

and $P(Y - 3X > 0) = P(Z > 0.23) = 0.4090$.

3.7.13

a. The three packages do not have to weigh the same. Denoting the total weight as $X_1+X_2+X_3$ allows the packages to have different weights, but the same distribution. Denoting the total weight as $3X$ means that the total weight is 3 times the weight of one package, implying the packages all weigh the same.

b. The variance of $3X$ is $3^2(0.05^2) = 0.0225$ while the variance of $X_1 + X_2 + X_3$ is $1^2 \cdot 0.05^2 + 1^2(0.05^2)+1^2(0.05^2) = 0.0075$. Thus the variance of $3X$ is three times as large as the variance of $X_1 + X_2 + X_3$.

c. Since $3X$ and $X_1 + X_2 + X_3$ do not have the same variance, they do not have the same distribution.

3.7.14 Note that $P(W > 90) = P[(X + Y)/2 > 90] = P(X + Y > 180)$. But the variable $X + Y$ is

$$N\left(85 + 80,\ 4^2 + 5^2\right) = N(165,\ 41)$$

and $P(X + Y > 180) = P(Z > 2.34) = 0.0096$

3.7.15 Let X_i denote the length of the i^{th} movie. Then the total length of the n movies is $T = X_i + \cdots X_n$. But each X_i is $N\left(110,\ 15^2\right)$ so that the distribution of T is $N\left(110n,\ 15^2 n\right)$ and

$$P(T > 8 \text{ hours}) = P(T > 480 \text{ min}) = P\left(Z > \frac{480 - 110n}{15\sqrt{n}}\right).$$

From Table C.1, we see that $P(Z > -1.645) \approx 0.95$. Thus we need

$$\frac{480 - 110n}{15\sqrt{n}} < -1.645 \quad \Rightarrow \quad 110n - 24.675\sqrt{n} - 480 > 0.$$

Solving this inequality with the quadratic formula yields $\sqrt{n} > 2.204$ which means $n > 4.85$. Thus he must rent at least 5 movies.

3.7.16 Since X and Y are independent, Theorem 3.7.5 tells us that X^2 and Y^2 are also independent random variables. Then by Theorem 3.7.2,

$$Var\left[X^2 + Y^2\right] = Var\left[X^2\right] + Var\left[Y^2\right] = 2 + 3 = 5.$$

3.8 The Central Limit Theorem

3.8.1 Note that the sample mean $\bar{X}_{12}$ is $N(0.920,\ 0.003/12) = N(0.920,\ 0.00025)$

a. $P\left(\bar{X}_{12} < 0.915\right) = P(Z < -0.32) = 0.3745$
b. $P\left(0.915 < \bar{X}_{12} < 0.922\right) = P(-0.32 < Z < 0.13) = 0.5517 - 0.3745 = 0.1772$
c. The sample mean being less than 0.920 simply means that when you add all the masses and divide by 12, the result is less than 0.920. It does not mean that all the masses are less than 0.920

3.8.2 Let X denote the IQ score of a randomly chosen person. Then X is $N\left(100,\ 15^2\right)$.

a. $P\left(X > 110\right) = P(Z > 0.67) = 0.2514$
b. The sample mean $\bar{X}_{25}$ is $N\left(100,\ 15^2/25\right) = N(100,\ 9)$ so that

$$P\left(\bar{X}_{25} > 110\right) = P(Z > 3.33) = 0.0004.$$

3.8.3 The sample mean $\bar{X}_45$ is $N\left(16,\ 0.08^2/48\right) = N(16,\ 0.000133)$ so that

$$P\left(\bar{X}_48 > 16.04\right) = P(Z > 3.46) = 0.0003.$$

Since this probability is so small, the observed sample mean of $\bar{x} = 16.04$ is very unlike to have occurred if $\mu = 16$. This indicates that μ is probably greater than 16 meaning the bottles are filled with an amount greater than 16.00 oz.

3.8.4 Since X is $b(500,\ 1/5)$, the population mean and variance are $\mu = 500(1/5) = 100$ and $\sigma^2 = 500(1/5)(4/5) = 80$. Thus the sample mean $\bar{X}_{50}$ is approximately $N\left(100,\ 80/50\right) = N(100,\ 1.6)$ so that

$$P\left(\bar{X}_{50} < 99\right) \approx P(Z < -0.79) = 0.2148.$$

3.8.5 Scores on a certain standardized test have a normal distribution with mean of $\mu = 1515$ and variance of $\sigma^2 = 300^2$. A random sample of n scores is selected.

a. Note that $\bar{X}_n$ is $N(1515,\ 300^2/n)$. The calculations of the probabilities are summarized in the table below.

n	$Var\left(\bar{X}_n\right)$	z_1	z_2	$P\left(1510 < \bar{X}_n < 1520\right)$
5	18000	-0.04	0.04	0.0320
50	1800	-0.12	0.12	0.0956
250	360	-0.26	0.26	0.2052
500	180	-0.37	0.37	0.2886
5000	18	-1.18	1.18	0.7620

b. The probability gets larger.
c. Based on these calculations, it appears that $\lim\limits_{n\to\infty} P\left(\mu - \varepsilon < \bar{X}_n < \mu + \varepsilon\right) = 1$.

3.8.6 Note that $\bar{X}_n$ is $N(172,\, 29^2/n)$. Then

$$P\left(\bar{X}_n > 3500/n\right) = P\left(Z > \frac{3500/n - 172}{29/\sqrt{n}}\right).$$

From Table C.1, we see that $P(Z > 2.325) \approx 0.01$. Thus we need

$$\frac{3500/n - 172}{29/\sqrt{n}} > 2.325 \quad \Rightarrow \quad 0 > 172n + 67.425\sqrt{n} - 3500$$

This inequality is quadratic in $\sqrt{n}$. Solving this using the quadratic formula yields $\sqrt{n} < 4.319$ so that $n < 18.65$. This means the largest value of n is 18. (**Note**: Another way to solve this problem is to calculate $P\left(\bar{X}_n > 3500/n\right)$ for different values of n and find the largest such that the probability is less than 0.01.)

3.8.7

a. Based on the given assumptions, $\bar{X}_{109}$ is $N(98.6,\, 0.06^2/109) = N(98.6,\, 0.000033)$ so that

$$P\left(\bar{X}_{109} < 98.3\right) = P(Z < -9090) = 0.0000.$$

b. Since this probability is so small, the assumption that the mean body temperature of the population of healthy adults is 98.6 °F is probably not correct.

3.8.8

a. Since $\bar{X}_n$ is approximately $N\left(\mu,\, \sigma^2/n\right)$, its mgf is approximately

$$M_X(t) = \exp\left\{\mu t + \frac{\left(\sigma^2/n\right)t^2}{2}\right\}$$

By exercise 2.7.7, the mgf of $Y = \bar{X}_n - \mu$ is approximately

$$M_X(t) = \exp\left\{-\mu t\right\}\exp\left\{\mu t + \frac{\left(\sigma^2/n\right)t^2}{2}\right\} = \exp\left\{\frac{\left(\sigma^2/n\right)t^2}{2}\right\}$$

But this is the mgf of a random variable that is $N\left(0,\, \sigma^2/n\right)$. Thus Y approximately has this distribution.

b. Note that

$$P\left(\mu - \varepsilon < \bar{X}_n < \mu + \varepsilon\right) = P\left(-\varepsilon < \bar{X}_n - \mu < \varepsilon\right).$$

But by part a., the variable $\bar{X}_n - \mu$ is approximately $N\left(0,\, \sigma^2/n\right)$ so that

$$P\left(-\varepsilon < \bar{X}_n - \mu < \varepsilon\right) \approx P\left(\frac{-\varepsilon}{\sigma/\sqrt{n}} < Z < \frac{\varepsilon}{\sigma/\sqrt{n}}\right) = P\left(\frac{-\sqrt{n}\varepsilon}{\sigma} < Z < \frac{\sqrt{n}\varepsilon}{\sigma}\right).$$

As $n \to \infty$, $-\sqrt{n}\varepsilon/\sigma \to -\infty$ and $\sqrt{n}\varepsilon/\sigma \to \infty$ so that

$$P\left(\frac{-\sqrt{n}\varepsilon}{\sigma} < Z < \frac{\sqrt{n}\varepsilon}{\sigma}\right) \to 1.$$

3.8.9 We want to find $P\left(-10 < \bar{X}_{100} - \mu < 10\right)$. By exercise 3.8.8, the variable $\bar{X}_{100} - \mu$ is approximately $N(0,\, 325^2/100) = N(0,\, 1056.25)$ so that

$$P\left(-10 < \bar{X}_{100} - \mu < 10\right) = P(-0.31 < Z < 0.31) = 0.2434.$$

3.8.10 We want to find n such that $P\left(-50 < \bar{X}_n - \mu < 5\right) \geq 0.95$. By exercise 3.8.8, the variable $\bar{X}_n - \mu$ is approximately $N(0,\, 325^2/n)$ so that

$$P\left(-50 < \bar{X}_n - \mu < 50\right) = \left(\frac{-50}{325/\sqrt{n}} < Z < \frac{50}{325/\sqrt{n}}\right).$$

From Table C.1, we see that $P(-1.96 < Z < 1.96) = 0.95$. Thus we need

$$\frac{50}{325/\sqrt{n}} > 1.96 \quad \Rightarrow \quad n > 162.3.$$

Thus the researcher needs a sample size of at least 163.

3.8.11

a. By section 3.3, the mean and variance are $\mu = \sigma^2 = 1$. A typical graph of the exponential pdf is given in that section. This distribution is not anywhere "close" to normal.
b. The marginal pdf's are $f_{X_1}((x_1) = e^{-x_1}$ and $f_{X_2}((x_2) = e^{-x_2}$. Since X_1 and X_2 are independent, the joint pdf is $f\,(x_1,\, x_2) = e^{-x_1 - x_2}$.
c. Note that since $\bar{X}_2 < 1/2$, and X_1 and X_2 cannot be negative, neither variable can be larger than 1. Thus

$$\begin{aligned} P\left(\bar{X}_2 < 1/2\right) &= P\left[\left(X_1 + X_2\right)/2 < 1/2\right] \\ &= P\left[X_2 < 1 - X_1\right] \\ &= \int_0^1 \int_0^{1-x_1} e^{-x_1 - x_2}\, dx_2\, dx_1 \approx 0.2642 \end{aligned}$$

d. According to the central limit theorem, $\bar{X}_2$ is approximately $N(1,\, 1/2)$ so that

$$P\left(\bar{X}_2 < 1/2\right) \approx P(Z < -0.71) = 0.2389.$$

This estimate is rather poor with a relative error of about 9.6%. This estimate is poor because the population distribution is not "close" to normal and the sample size is very small.

3.8.12 By the central limit theorem, $\bar{X}_n$ is approximately $N\left(\mu_X,\, \sigma_X^2/n\right)$ and $\bar{Y}_n$ is approximately $N\left(\mu_Y, \sigma_Y^2/n\right)$. By Theorem 3.7.4, the variable $D_n = \bar{X}_n - \bar{Y}_n$ is then approximately

$$N\left(\mu_X - \mu_Y,\, 1^2\sigma_X^2/n + (-1)^2\sigma_Y^2/n\right) = N\left(\mu_X - \mu_Y,\, \left(\sigma_X^2 + \sigma_Y^2\right)/n\right).$$

3.8.13 Let $\bar{X}_{30}$ and $\bar{Y}_{30}$ denote the sample means of the men and women, respectively. By exercise 3.8.12, the variable $\bar{X}_n - \bar{Y}_n$ is approximately

$$N\left(69 - 63.6,\ \left(2.8^2/n + 2.5^2\right)/30\right) = N(5.4,\ 0.4697)$$

so that

$$P\left(5.25 < \bar{X}_n - \bar{Y}_n < 5.75\right) \approx P(-0.22 < Z < 0.51) = 0.2821$$

3.8.14

a. $\lim_{N\to\infty}\left(1 - \dfrac{n-1}{N-1}\right) = 1$

b. This limit tells us that the finite population correction is not necessary when n is very small in comparison to the population size.

3.9 The Gamma and Related Distributions

3.9.1

a. W has a gamma distribution with parameters $r = 2$ and $\lambda = 10/60 = 1/6$ so that its pdf is

$$f(w) = \frac{((1/6)^2}{(2-1)!}\, w^{2-1}\, e^{-(1/6)w} = \frac{w\, e^{-w/6}}{36} \quad \text{for } w > 0,$$

its mean is $\mu = 2/(1/6) = 12$, and its variance is $\sigma^2 = 2/(1/6)^2 = 72$.

b. If the second call arrives before 8:15 AM, then $W < 15$, so

$$P(W < 15) = \int_0^{15} \frac{w\, e^{-w/6}}{36}\, dw = 0.7127.$$

c. If $r = 8$, then the pdf is

$$f(w) = \frac{((1/6)^8}{(8-1)!}\, w^{8-1}\, e^{-(1/6)w} = \frac{w^7\, e^{-w/6}}{6^8 \cdot 7!} \quad \text{for } w > 0,$$

so that

$$P(W < 45) = \int_0^{45} \frac{w^7\, e^{-w/6}}{6^8 \cdot 7!}\, dw = 0.4754.$$

3.9.2 Let W denote the wait time (in hours) until the second replacement of the batteries. Then W has a gamma distribution with parameters $r = 2$ and $\lambda = 1/10$ so that its pdf is

$$f(w) = \frac{((1/10)^2}{(2-1)!}\, w^{2-1}\, e^{-(1/10)w} = \frac{w\, e^{-w/10}}{100} \quad \text{for } w > 0.$$

Now, the two sets of batteries last the entire trip if the time until the second replacement is greater than eight hours. Thus we need to know $P(W > 8)$, so

$$P(W > 8) = 1 - P(W < 8) = 1 - \int_0^8 \frac{w\, e^{-w/10}}{100}\, dw = 0.8088.$$

3.9.3 By definition,

$$\Gamma(0.75) = \int_0^\infty y^{0.75-1}\, e^{-y}\, dy \approx 1.225.$$

Then, $\Gamma(1.75) \approx 0.75(1.225) = 0.919$ and $\Gamma(2.75) \approx 1.75(0.919) = 1.608$

3.9.4 a. $P(X \le 25) = 0.95$ b. $P(X > 7.261) = 0.05$ c. $P(5.229 < X \le 8.547) = 0.09$

3.9.5 Let W denote the wait time until the 10^{th} customer arrives. Then W has a gamma distribution with parameters $r = 10$ and $\lambda = 1/2$. But this is the same as a chi-square distribution with $n = 20$ degrees of freedom. Thus using Table C.4, we find that

$$P(W < 34.17) = 0.975$$

.

3.9.6

a. $P(T \le 2.447) = 0.975$ b. $P(T > 3.143) = 0.01$
c. $P(-1.943 < T \le 1.943) = 0.9$ d. $P(-3.707 < T \le 1.943) = 0.945$

3.9.7 Notice that in $f(t)$, the Student-t pdf, the only place t appears is in the denominator as t^2. Replacing t with $-t$ will not change the value of this term. So $f(t) = f(-t)$ for any number of degrees of freedom n.

3.9.8 Note that by equation (3.7)

$$T^2 = \frac{Z^2}{C/n} \tag{3.1}$$

But Z^2 is $\chi^2(1)$ and C is $\chi^2(n)$ so that T^2 fits the definition of a variable that has an F distribution with 1 and n degrees of freedom as given in equation (3.8).

3.9.9

a. The graph is shown in Figure 3.12. We see that this is the same as a $U(0,\, 1)$ distribution.
b. The graphs are shown in Figure 3.12.
c. If X has a beta distribution with $a = 5$ and $b = 1$, then its pdf is

$$f(x) = \frac{\Gamma(5+1)}{\Gamma(5)\,\Gamma(1)}\, x^{5-1}(1-x)^{1-1} = \frac{5!}{4! \cdot 0!}\, x^4(1-x)^0 = 5x^4$$

so that

$$P(X > 0.5) = \int_{0.5}^1 5x^4\, dx = 31/32 = 0.96875$$

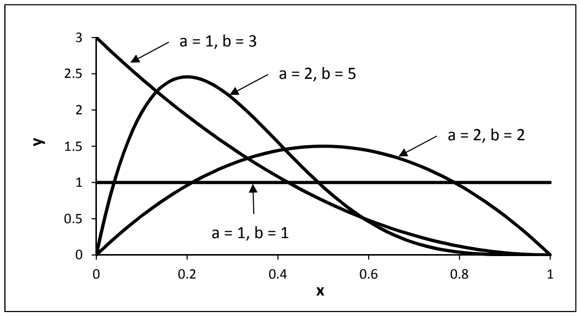


Figure 3.12

3.9.10 By the definition of expected value,

$$E(X) = \int_0^\infty x\,\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}\,dx = \int_0^\infty \frac{x}{\lambda}\,e^{-(x/\lambda)^k}\,k\left(\frac{x}{\lambda}\right)^{k-1}\,dx.$$

Now let $y = (x/\lambda)^k$. Then when $x = 0$, $y = 0$; when $x \to \infty$, $y \to \infty$;

$$dy = k\left(\frac{x}{\lambda}\right)^{k-1}\frac{1}{\lambda}\,dx \quad \text{and} \quad \frac{x}{\lambda} = y^{1/k}.$$

By substitution and the definition of $\Gamma(t)$,

$$E(X) = \lambda\int_0^\infty y^{1/k}\,e^{-y}\,dy = \lambda\int_0^\infty y^{(1+1/k)-1}\,e^{-y}\,dy = \lambda\,\Gamma\left(1 + 1/k\right).$$

3.9.11 The χ^2 distribution is a special case of the gamma distribution where $\lambda = 1/2$ and $r = n/2$. Thus

$$E(X) = \frac{r}{\lambda} = \frac{n/2}{1/2} = n.$$

3.9.12 The mgf of X is $M_X(t) = (1-2t)^{-n/2}$ and the mgf of Y is $M_Y(t) = (1-2t)^{-m/2}$. By Theorem 3.7.3, the mgf of W is

$$M_W(t) = M_X(t)\cdot M_Y(t) = (1-2t)^{-n/2}\cdot(1-2t)^{-m/2} = (1-2t)^{(-n-m)/2}.$$

But this is the mgf of a random variable that is $\chi^2(n+m)$. Thus W has this distribution.

3.9.13 We let $W = W_1 + W_2 + \cdots + W_n$ as in Example 3.9.2 where each W_i has an exponential distribution with parameter λ. By the calculations immediately following Example 3.3.5, the mgf of each W_i is $\lambda/(\lambda - t)$ for $t < \lambda$. By Theorem 3.7.3, since these n variables are mutually independent, the mgf of W is the product of their mgf's so that

$$M_W(t) = \left(\frac{\lambda}{\lambda - t}\right)^n \quad \text{for } t < \lambda.$$

3.9.14 By the definition of the m.g.f, the mgf of a random variable W with a gamma distribution and parameters λ and r is

$$M_W(t) = E\left[e^{Wt}\right] = \int_0^\infty e^{wt}\,\frac{\lambda^r}{\Gamma(r)}\,w^{r-1}\,e^{-\lambda w}\,dw = \frac{\lambda^r}{\Gamma(r)}\int_0^\infty w^{r-1}\,e^{(t-\lambda)w}\,dw.$$

Notice that for this integral to converge, we need $t - \lambda < 0 \quad\Rightarrow\quad t < \lambda$. Now let $y = (\lambda - t)\,w$. Then when $w = 0$, $y = 0$; when $w \to \infty$, $y \to \infty$ (since $\lambda - t > 0$);

$$w = \frac{y}{\lambda - t} \quad \text{and} \quad dw = \frac{dy}{\lambda - t}.$$

By substitution and the definition of $\Gamma(t)$,

$$\begin{aligned} M_W(t) &= \frac{\lambda^r}{\Gamma(r)}\int_0^\infty w^{r-1}\,e^{(t-\lambda)w}\,dw \\ &= \frac{\lambda^r}{\Gamma(r)}\int_0^\infty \frac{y^{r-1}}{(\lambda-t)^{r-1}}\,e^{-y}\,\frac{dy}{\lambda-t} \\ &= \frac{\lambda^r}{\Gamma(r)(\lambda-t)^r}\int_0^\infty y^{r-1}\,e^{-y}\,dy \\ &= \frac{\lambda^r}{\Gamma(r)(\lambda-t)^r}\,\Gamma(r) \\ &= \frac{\lambda^r}{(\lambda-t)^r} \quad \text{for } t < \lambda. \end{aligned}$$

3.9.15 a. $P(F \le 1.88) = 0.975$ b. $P(F > 2.12) = 0.01$ c. $P(1.70 \le F \le 2.12) = 0.04$

3.9.16 a. $P(F > 4.95) = 0.05$ b. $P(F \le 10.67) = 0.99$ c. $P(F > 10.67 \cup F \le 4.95) = 0.01 + 0.95 = 0.96$

3.9.17 In Figure 3.22, we see that the density curve is very tall over the interval (0.5, 1.5] but very short over (2, 4]. Geometrically, we also see that there is more area under the curve over the first interval than the second. Thus $P(0.5 < F \le 1.5) > P(2 < F \le 4)$.

3.9.18 By definition of the F-distribution with n and d degrees of freedom, U is $\chi^2(n)$ and V is $\chi^2(d)$ and they are independent. Then

$$\frac{1}{F} = \frac{V/d}{U/n}$$

so that $1/F$ has an F-distribution with d and n degrees of freedom by definition.

3.9.19

a. Let $u = \sqrt{2\pi}$ so that $\sqrt{2}du = y^{-1/2}dy$ and

$$\begin{aligned} \Gamma\left(\frac{1}{2}\right) &= \int_0^\infty y^{-1/2}e^{-y}dy = \int_0^\infty \sqrt{2}e^{-u^2/2}du = 2\sqrt{\pi}\int_0^\infty \frac{1}{\sqrt{2\pi}}e^{-u^2/2}du \\ &= 2\sqrt{\pi}\cdot\frac{1}{2} = \sqrt{\pi}. \end{aligned}$$

b. From exercise 3.5.9, the pdf of X is

$$f(x) = \frac{1}{\sqrt{2\pi x}}\, e^{-x/2}.$$

But the pdf of a gamma distribution with $\lambda = r = 1/2$ is

$$f(x) = \frac{(1/2)^{1/2}}{\Gamma(1/2)}\, x^{-1/2}\, e^{-x/2} = \frac{(1/2)^{1/2}}{\sqrt{\pi}}\, x^{-1/2}\, e^{-x/2} = \frac{1}{\sqrt{2\pi x}}\, e^{-x/2}$$

which is the same as the pdf of X. Thus X has a gamma distribution with $\lambda = r = 1/2$. Since this is the case, the mgf of X is

$$M(t) = (1 - t/(1/2))^{-1/2} = (1 - 2t)^{-1/2}.$$

c. By Theorem 3.7.5, the variables $Y_1^2, Y_2^2, \ldots, Y_n^2$ are independent. Thus X is the sum of independent random variables, each with mgf $M(t) = (1 - 2t)^{-1/2}$. Hence by Theorem 3.7.3, the mgf of X is

$$M_X(t) = \underbrace{(1 - 2t)^{-1/2} \cdots (1 - 2t)^{-1/2}}_{n \text{ times}} = (1 - 2t)^{-n/2}.$$

d. The mgf in part c. is exactly that of a random variable with a gamma distribution with $\lambda = 1/2$ and $r = n/2$. Thus X has the pdf given in Theorem 3.9.1.

3.10 Approximating the Binomial Distribution

3.10.1 People often travel in groups, such as families, and thus arrive together. So the arrivals are not mutually independent, but independence is a reasonable simplifying assumption.

3.10.2 Let Y be $N(950(0.35), 950(0.35)(1 - 0.35)) = N(332.5, 216.125)$.

a. $P(310 \le X \le 350) \approx P(310 - 0.5 < Y \le 350 + 0.5) = P(-1.56 < Z \le 1.22) = 0.8888 - 0.0594 = 0.8294$
b. $P(X \le 315) \approx P(Y \le 315 + 0.5) = P(Z \le -1.16) = 0.1230$
c. $P(340 \le X) \approx P(340 - 0.5 < Y) = P(0.48 < Z) = 0.3156$

3.10.3 Suppose X is $b(75, 0.02)$.

a. Note that $np = 1.5$ and $n(1 - p) = 73.5$. Since $np < 5$, the rule-of-thumb indicates that we cannot use equation (3.9) to approximate values of the probability of X.
b. Using software, $P(X \le 1) = 0.5561$.

c. Let Y be $N(75(0.02),\ 75(0.02)(1-0.02)) = N(1.5,\ 1.47)$. Then according to equation (3.9),

$$P(X \leq 1) \approx P(Y \leq 1 + 0.5) = P(Z \leq 0) = 0.5.$$

This approximation has a relative error of $0.0561/0.5561 \approx 10\%$, so this is not a good approximation.

3.10.4 Let X denote the number of gamblers out of 5,000 that win. Then X is $b(5,000,\ 1/38)$. Let Y be $N(5000(1/38),\ 5000(1/38)(1-1/38)) = N(131.58,\ 128.12)$. Then the probability that more than 140 of them win is

$$P(140 < X) = P(141 \leq X) \approx P(141 - 0.5 < Y) = P(0.79 < Z) = 0.2148.$$

3.10.5

a. Let X denote the number of voters out of 825 that actually did vote for the proposition. Then X is $b(825,\ 0.4)$. Let Y be $N(825(0.4),\ 825(0.4)(1-0.4)) = N(330,\ 198)$. Then the probability that more than 429 of them actually did vote for the proposition is

$$P(429 < X) = P(430 \leq X) \approx P(430 - 0.5 < Y) = P(7.07 < Z) = 0.0000.$$

b. The fact that this probability is so small indicates that observing 429 or more that actually did vote for the proposition is extremely unusual, assuming the people are being honest. This suggests that people are not honestly responding to the survey.

3.10.6 Let X denote the number of points selected such that $x \leq y$ out of 500. Figure 3.13 shows the rectangle $0 \leq x \leq 1$, $0 \leq y \leq 2$ and the portion in which $x \leq y$. We see that the total area of the region is 2 and the area of the portion in which $x \leq y$ is $2 - (1/2)(1)(1) = 3/2$. Thus the probability of selecting a point such that $x \leq y$ is $(3/2)/2 = 3/4$. This means that X is $b(500,\ 0.75)$. Let Y be $N(500(0.75),\ 500(0.75)(1-0.75)) = N(375,\ 93.75)$. Then the probability that for 365 or more of these points, $x \leq y$, is

$$P(365 \leq X) \approx P(365 - 0.5 < Y) = P(-1.08 < Z) = 0.8599.$$

3.10.7 Let X denote the number of students out of n that regularly use credit cards. Then X is $b(n,\ 0.25)$. Let Y be $N(n(0.25),\ n(0.25)(1-0.25)) = N(0.25n,\ 0.1875n)$. We want to find n such that

$$0.95 < P(50 \leq X) \approx P(50 - 0.5 < Y) = P\left(\frac{49.5 - 0.25n}{\sqrt{0.1875n}} < Z\right).$$

From Table C.1, we see that $P(-1.645 < Z) \approx 0.95$. Thus we want

$$\frac{49.5 - 0.25n}{\sqrt{0.1875n}} < -1.645 \quad \Rightarrow \quad 0 < 0.25n - 0.7123\sqrt{n} - 49.5.$$

This inequality is quadratic in $\sqrt{n}$. Solving it with the quadratic formula yields $\sqrt{n} > 15.57$ so that $n > 242.4$. Thus the researcher needs a sample size of at least 243.

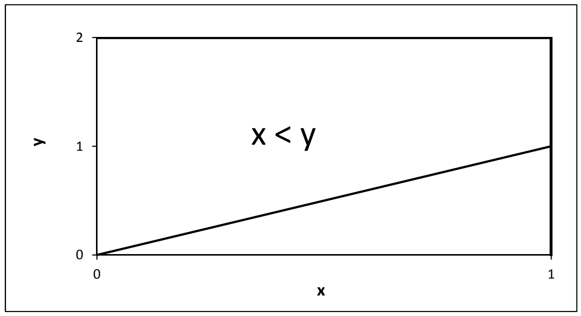


Figure 3.13

3.10.8

a. Note that X is $b(n,\, 1/2)$ so that $\mu = n/2$ and $\sigma^2 = n/4$.
b. Note that

$$\begin{aligned} P\left(\left|\frac{X}{n} - \frac{1}{2}\right| < \epsilon\right) &= P\left(\left|X - \frac{n}{2}\right| < n\epsilon\right) \\ &> 1 - \frac{n/4}{(n\epsilon)^2} \quad \text{by Chevyshev's inequality} \\ &= 1 - \frac{1}{4n\epsilon^2}. \end{aligned}$$

c. By elementary calculus, $\lim_{n\to\infty}\left(1 - \frac{1}{4n\epsilon^2}\right) = 1$. Note that X/n is the relative frequency of tails and $|X/n - 1/2|$ is the difference between the relative frequency and the theoretical probability and ϵ is an upper bound on this difference. The fact that this limit equals 1 means that we can make this difference as small as we want, with probability as close to 1 as we want, by choosing a large enough value of n. In simpler terms, this means that as n gets larger, the relative frequency gets closer to the theoretical probability.

Chapter 4

Statistics

4.1 What is Statistics?

4.1.1

a. Population - All voters in the city, Sample - 50 voters. This is biased because not everyone in the city lives in a house with a 3-car garage, so the sample is not representative of the entire population.
b. Population - All students at the university, Sample - Students in the 4 classes. This is biased because the sample includes only math students, so the sample is not representative of the entire population.
c. Population - All students at the university, Sample - 50 students selected from the student directory. This is not biased because the students were randomly selected from the population.
d. Populations - All potatoes of the 3 varieties, Samples - 3 bags of potatoes. These samples are biased because the samples are not randomly chosen from the respective populations. All potatoes in a single bag probably came from the same crop, so they are not representative of their population.
e. Population - All fellow students, Sample - 15 friends. This is biased because he asked only his friends, so the sample is not representative of the entire population.
f. Population - All fiber optic cable manufactured in a day, Sample - Selected pieces of cable. This is not biased because the pieces are selected throughout the day.

4.1.2

a. Statistic - The biologist surely did not observe all the ghost crabs on the beach.
b. Parameter - All the names in the book were surveyed.

c. Parameter - All of the records were surveyed.
d. Statistic - The mother surely did not observe all of the teenagers in the mall.
e. Statistic - The officer surely did not survey every mile of highway in the region. The percentage is based on what was actually observed.

4.1.3 a. Quantitative, continuous; b. Qualitative; c. Quantitative, discrete; d. Quantitative, continuous; e. Quantitative, discrete

4.1.4 a. Discrete, b. Continuous, c. Continuous, d. Discrete, e. Continuous

4.1.5 a. Nominal, b. Interval, c. Ratio, d. Ratio, e. Ordinal

4.1.6 a. Ratio, b. Nominal, c. Interval, d. Nominal, e. Interval

4.1.7 a. Observational, b. Experiment, c. Experiment, d. Observational, e. Experiment

4.1.8 a. Convenience, b. Systematic, c. Voluntary response, d. Cluster, e. Stratified

4.1.9

a. A drug may work better for some types of cancer than others. The factor of type of cancer needs to be accounted for.
b. The individual professors could influence homework scores, so there are nuisance variables not accounted for.
c. One factor not accounted for is the surface area of the parachute.
d. The researcher needs to take into account lifestyle, income, diet, and other habits when trying to determine “causes” of good health.
e. Different gardners grow tomatoes in different ways. Factors such as amount of irrigation, amount and type of fertilizer, soil type, and others are different from gardner to gardner. These factors need to be accounted for.

4.1.10

a. Since all the candies are mixed, this is a simple random sample.

b. Since the row is randomly chosen, the sample is random. But only groups of students who sit in the same row can be selected, so the sample is not a simple random sample.
c. Since the people are randomly chosen and the population is half male and half female, every member has an equal probability of being selected so the sample is random. However, we could not, for instance, get a sample of 25 males and 75 females, so the sample is not a simple random sample.
d. There is nothing random about this sample at all, so it is neither a random sample nor a simple random sample.

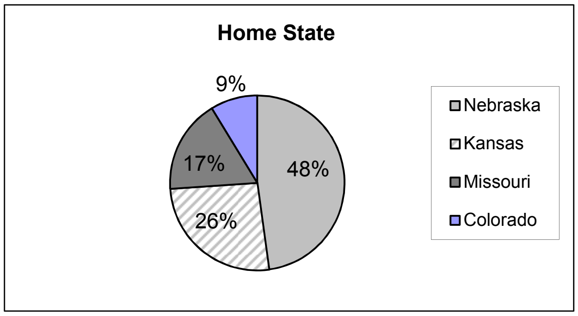
4.1.11 No. Temperatures are at the interval level of measurement. Although there is a zero starting point, 0°F does not mean no temperature, so ratios such as “twice as hot” are meaningless.

4.1.12 Label the dice as first, second, and third. Roll the 3 dice. The number on the first die gives the room number, the second die gives the row number, and the third gives the student number. Repeat this 20 times.

4.2 Summarizing Data

4.2.1

a. Since the data collected are names of states, these data are qualitative.
b. The charts are shown in Figure 4.1.



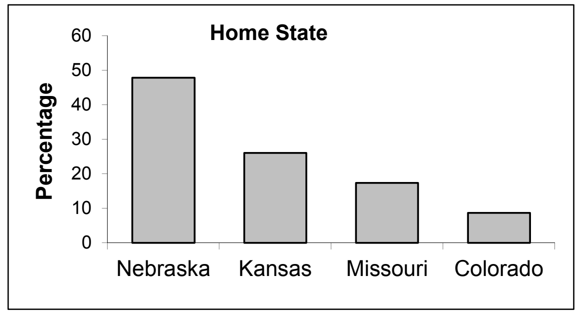
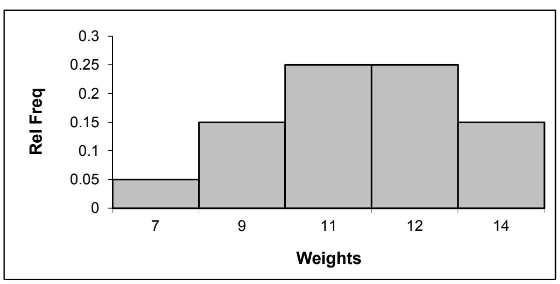


Figure 4.1

c. Since these data are qualitiative, it is not appropriate to calculate the mean, standard deviation, or the 32^{th} percentile.
d. The mode is Nebraska.

4.2.2

a. The histograms are shown in Figure 4.2. The weights appear to have a normal distribution.



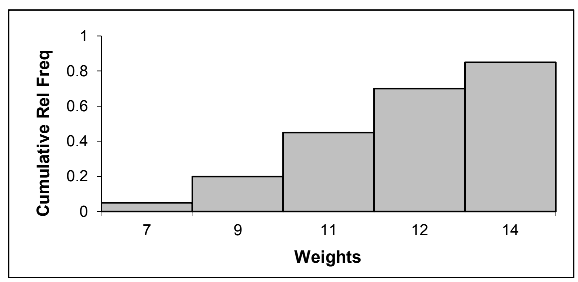


Figure 4.2

b. The mean is $\bar{x} = (6.1 + \cdots 17)/20 = 11.925$, the variance is $s^2 = [(6.1 - 11.925)^2 + \cdots + (17 - 11.952)^2]/(20-1) = 6.985$, and the standard deviation is $s = \sqrt{6.985} = 2.943$.
c. To find the first quartile, $L = 20 \cdot 0.25 = 5$ so that $\pi_{0.25} = 0.5\,(x_5 + x_6) = 0.5(10.1 + 10.4) = 10.25$. By similar calculations, $\pi_{0.5} = 11.75$ and $\pi_{0.75} = 13.5$
d. Note that there are 13 out of 20 data values less than 13.2. So the percentile rank of 13.2 is $13/20 \times 100\% = 65\%$.

4.2.3

a. The mean is $\bar{x} = (0.783 + \cdots 0.976)/30 = 0.890$, the variance is $s^2 = [(0.783 - 0.89)^2 + \cdots (0.976 - 0.89)^2]/(30 - 1) = 0.00208$, and the standard deviation is $s = \sqrt{0.00208} = 0.0456$

b. The histograms are shown in Figure 4.3. The weights appear to have a normal distribution.

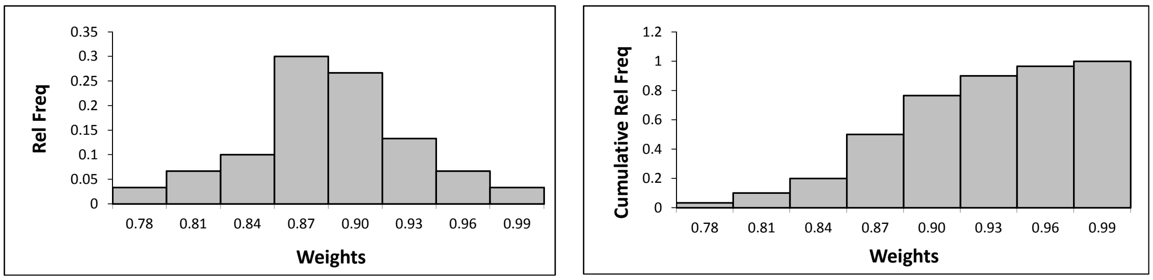


Figure 4.3

4.2.4

a. The mean is $\bar{x} = (2.09 + \cdots 3.26)/30 = 2.621$, the variance is $s^2 = [(2.08 - 2.621)^2 + \cdots + (3.26 - 2.621)^2]/(30 - 1) = 0.0858$, and the standard deviation is $s = \sqrt{0.0858} = 0.2930$.

b. The histograms are shown in Figure 4.4. The percent shrink appears to have a normal distribution.

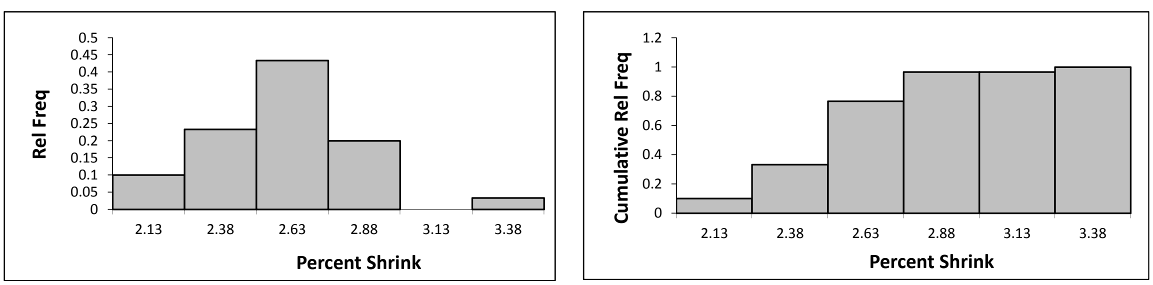


Figure 4.4

c. By part a., $\mu \approx \bar{x} = 2.621$ and $\sigma \approx s = 0.293$ so that the range is approximately

$$(2.621 - 3(0.293),\, 2.621 + 3(0.293)) \quad \Rightarrow \quad (1.742,\, 3.5)\,.$$

d. Yes, the range in part c. falls between 1% and 4%.

4.2.5

a. The bar for trucks is more than twice as long as that for cars. This is not an accurate conclusion because the numbers show the actual distances are about 150 and 200 feet. The distance for trucks is more than that for cars, but the distance is not nearly twice as long.
b. The problem with the graph is that the horizontal axis does not start at 0. To make the graph more accurate, we need the axis to start at 0 as in Figure 4.5.

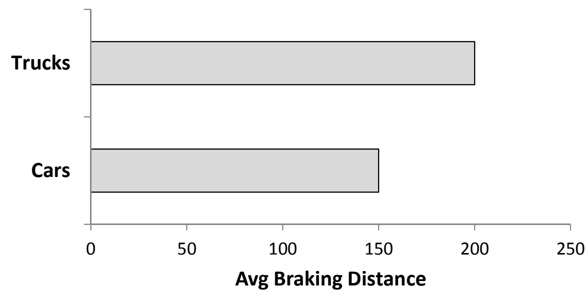


Figure 4.5

4.2.6

a. The 5-number summaries are shown in the table below.

Type	Min	Q_1	Median	Q_3	Max
G	25	74	82	97	117
PG	87	96	100	114	152
PG-13	95	105	117	130	178
R	95	105	117	130	178

b. The boxplots (as drawn by Minitab) are shown in Figure 4.6. We see that the PG and PG-13 data have outliers. The PG and R data have a very narrow range between the min and median and their medians are about the same. The PG-13 movies tend to be the longest of all the types. The G movies tend to be the shortest.

4.2.7 The histograms are shown in Figure 4.7. We see that a higher proportion of the European players are at the higher end of the interval [160, 190] whereas the heights of the players from the Americas are spread more uniformly throughout the interval. This suggests that European players tend to be a bit taller than those from the Americas.

4.2.8 Since the data values 73 and 99 appear twice and no other data value appears more often, the modes at 73 and 99. The histogram is shown in Figure 4.8. We see that there are two peaks, so the data is bimodal. This suggests that there are two types of students in this class: A students and C and below students with not many B students.

4.2.9

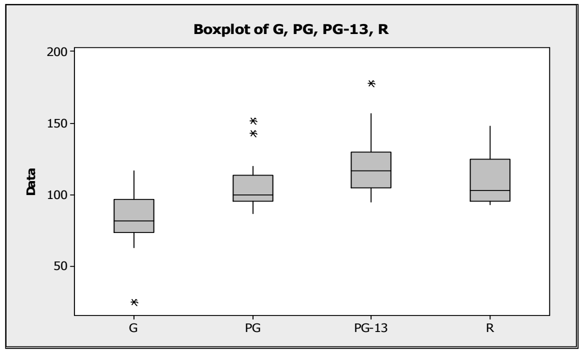


Figure 4.6

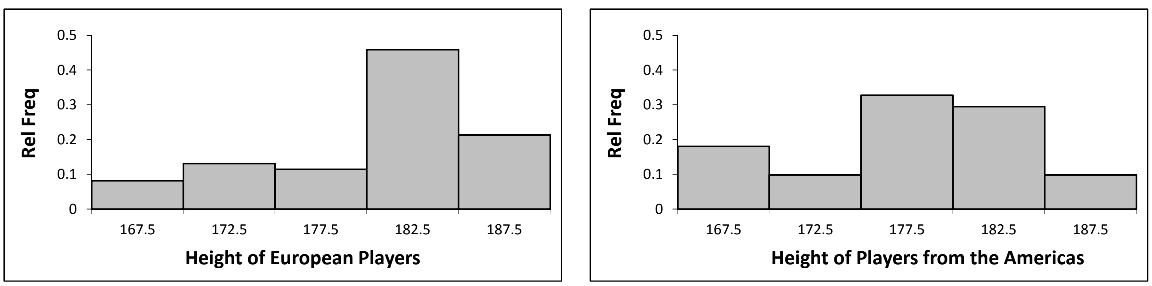


Figure 4.7

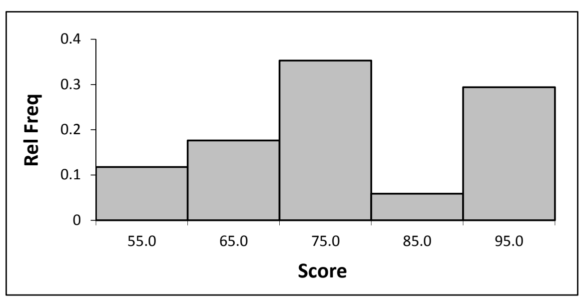


Figure 4.8

a. Yes. If a set of data contains negative values, then it is possible for the mean to be negative.

b. No. Since the variance is the sum of squares, it can never be negative.

c. Yes. Consider the data set $\{-1, 0, 1\}$. Its mean is 0 and its variance is 1, so it is possible that the variance can be larger than the mean.

d. No. By definition $\pi_{0.25}$ is a value that is greater than about 25% of the data values while $\pi_{0.75}$

is greater than about 50%. So $\pi_{0.25}$ cannot be larger than $\pi_{0.75}$. They could be eqaul if the data set contains some repeated data values.

e. No. Consider the data set {1, 2, 3, 10}. The mean is 4 which is greater than more than half the data values.

f. No. Consider the data set {1, 1, 1, 1}. The mean is 1 which is not greater than any of the data values.

4.2.10

a. The min and max of this data are 0 and 7.3, so midrange = $(7.3 - 0)/2 = 3.65$.

b. The median is 1.8. This is much less than the midrange.

c. The phrase "in the middle, halfway between the min and max" is describing the midrange.

4.2.11 If x is a value of a normally distributed random variable with mean μ and standard deviation σ, in section 3.4 we defined the z-score of x to be $z = (x - \mu)/\sigma$. Now, if x is a data value from a set of data with mean $\bar{x}$ and standard deviation s, we define the z-score of x to be

$$z = \frac{x - \bar{x}}{s}.$$

The z-score is a measure of how far x is from the mean.

a. If $x > \bar{x}$, then $x - \bar{x} > 0$ so that $z > 0$ since $s > 0$. A similar argument shows that if $x < \bar{x}$, then $z < 0$

b. If $x = (\bar{x} + n \cdot s)$ where n is any number, then

$$z = \frac{x - \bar{x}}{s} = \frac{(\bar{x} + n \cdot s) - \bar{x}}{s} = \frac{n \cdot s}{s} = n.$$

If x_1 is "further" from the mean than x_2, then we could say $x_1 = \bar{x} + n_1 \cdot s$ and $x_1 = \bar{x} + n_2 \cdot s$ for some numbers n_1 and n_2 such that $|n_1| > |n_2|$ so that how does the the magnitude of the z-score of x_1 is greater than that of x_2?

c. Bob's z-score is $z = (80 - 90)/5 = -2$ and Joe's is $z = (50 - 55)/10 = -0.5$. Since Joe's z-score is smaller in magnitude, Joe's test score is relatively closer to the mean than Bob's test score. So Joe did relatively better.

4.2.12 We can freely choose 4 of the data values. Let S denote the sum of these values. The last data value, call it x must be chosen such that $\bar{x} = (S + x)/5 = 3.2$. There is only one solution to this equation, so there is only one possible value for x. In general, if we want a sample of n data values with a specified mean, we can freely choose $n - 1$ of them.

4.2.13

a. The histograms are shown in Figure 4.9.

b. We see that Class 1 has many data values equal to the mean and not many very far from the mean, so it's standard deviation, or "average distance from the mean," is relatively small. Class 2 exhibits the exact opposite, so its standard deviation is relatively large. Class 3 has some data values close to the mean and some far from the mean, so its standard deviation is between that of classes 1 and 3.

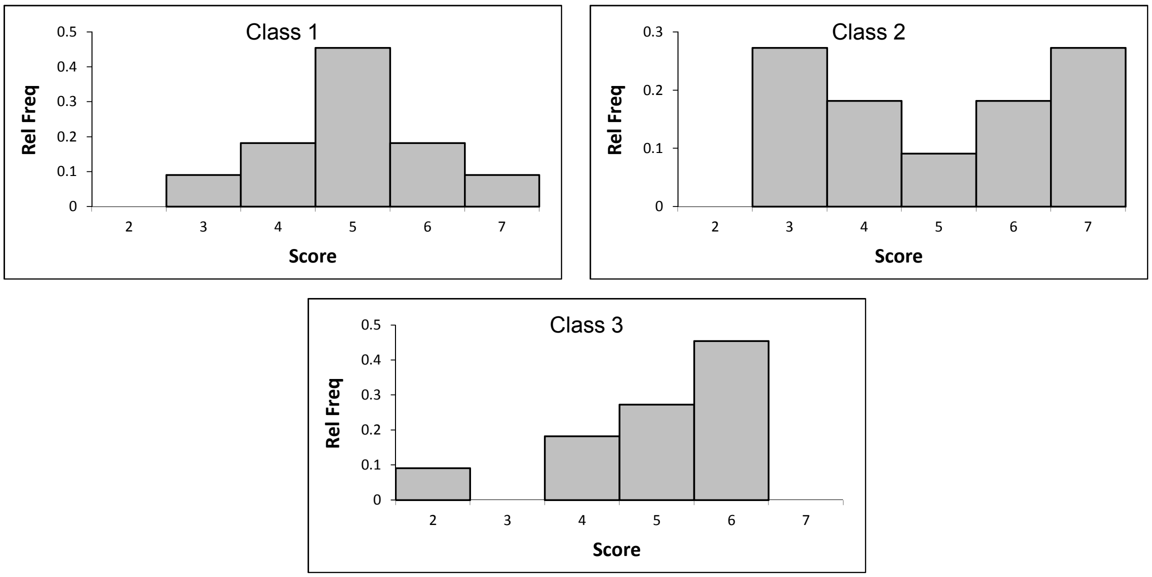


Figure 4.9

c. The sample standard deviations are Class 1: 1.095, Class 2: 1.67, Class 3: 1.26. This agrees with the conclusions in part b.

4.2.14 The mean, median, mode and skewness of the data sets are summarized in the table below, and the histograms are shown in Figure 4.10. We see that set 1 is skewed right, set 2 is symmetric, and set 3 is skewed left. Skewed right corresponds to a positive skewness, skewed left to a negative skewness, and symmetric to a skewness of 0.

	Mean	**Median**	**Mode**	**Skewness**
Set 1	2.91	3	2	0.86
Set 2	5	5	5	0
Set 3	5	5	6	-1.45

4.2.15

a. The calculations of the sample kurtoses are summarized in the table below. Note that $n = 11$ for each data set. The histograms are shown in Figure 4.11.

	$\bar{x}$	$\sum (x_i - \bar{x})^4$	$\sum (x_i - \bar{x})^2$	**Kurtosis**
Set 1	5	2	2	5.00
Set 2	4	36	12	0.42
Set 3	3	198	30	-0.13

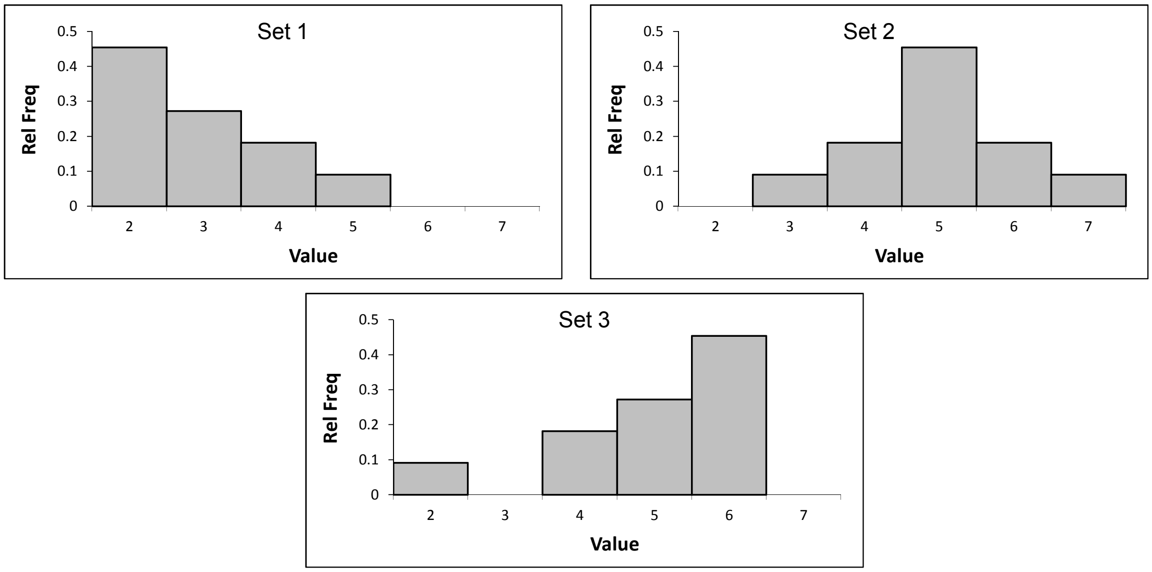


Figure 4.10

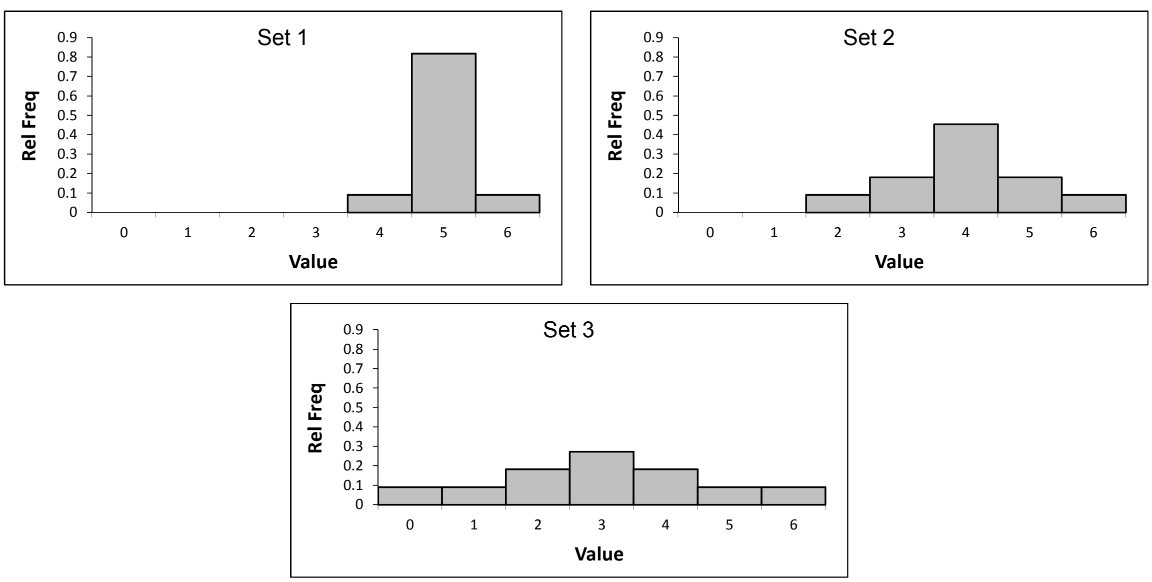


Figure 4.11

b. Data set 1 has a "sharp peak" and a high kurtosis while data set 2 has a very "flat peak" and a small kurtosis. Thus the sharper the peak, the higher the kurtosis.

c. Note that in the formula for the kurtosis, the first term is non-negative. Therefore, the kurtosis must be greater than the second term, $-\frac{3(n-1)^2}{(n-2)(n-3)}$. This lower bound is a rational expression and the degree of the numerator equals the degree of the denominator, so the limit as $n \to \infty$ equals the ratio of leading coefficients, $-3/1 = -3$.

4.2.16 Call Class 1 the class with 10 students and Class 2 the class with 20 students. The calculations of the means and standard deviations are summarized in the table below. We see that adding the outlier to class 1 increased the mean by over 2 and the standard deviation increased by 133%. Adding the outlier to class 2 increased the mean by only 1.2 and the standard deviation increased by 78%. Thus the effect of an outlier is more significant on a small set of data.

	Class 1	**Class 2**	**Class 1 with outlier**	**Class 2 with outlier**
Mean	62.67	63.00	64.80	64.00
StDev	3.16	3.08	7.38	5.48

4.2.17

a. The means, medians, and outliers are summarized in the table below.

	Mean	**Median**	**Outliers**
Set 1	2.44	2.5	none
Set 2	2.18	2.5	1
Set 3	2.48	2.5	1, 4

b. The first two statements are true, and the last two are false.

4.2.18

a. This set of data has 10 values. To calculate a 20% trimmed mean, we would discard the smallest and the largest data value and calculate the mean of the remaining 8 values. These 8 values have a mean of 12.55.

b. Replacing 14.2 with the outlier 25.8 does not affect the value of the 20% trimmed mean.

4.2.19

a. Height: $CV = 3.19/67.54 \times 100\% = 4.72\%$, Money: $CV = 2.56/9.35 \times 100\% = 27.38\%$

b. Based on the coefficients of variation, Money has more variation. This is not surprising because the amount of money students carry with them can range from 0 to very large amounts while heights are relatively similar.

4.2.20 The calculations are summarized in the table below.

x	f	$f \cdot x$	$f \cdot x^2$
11	1	11	121
12	1	12	144
13	5	65	845
14	7	98	1372
15	1	15	225
Sum =	15	201	2707

$\bar{x} = \frac{1}{15}(201) = 13.4$

$s = \sqrt{\frac{15(2707)-201^2}{15(15-1)}} = 0.986$

4.2.21 The calculations are summarized in the table below. The estimate of $\bar{x}$ is 0.1 units high and the estimate of s is 0.01 units high, so these are very good estimates.

x	f	$f \cdot x$	$f \cdot x^2$
0.5	10	5	2.5
1.5	7	10.5	15.75
2.5	6	15	37.5
3.5	3	10.5	36.75
4.5	2	9	40.5
5.5	1	5.5	30.25
6.5	0	0	0
7.5	1	7.5	56.25
Sum =	30	63	219.5

$\bar{x} = \frac{1}{30}(63) = 2.1$

$s = \sqrt{\frac{30(219.5)-63^2}{30(30-1)}} = 1.73$

4.2.22 If $x_1 = x_2 = \ldots = x_n$, let x denote this common value. Then

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{1}{n}\sum x = \frac{1}{n}(n\,x) = x \text{ and}$$
$$s = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1}\sum (x - x)^2} = 0.$$

4.2.23 Note that

$$\begin{aligned}
s^2 &= \frac{1}{n-1}\sum (x_i - \bar{x})^2 \\
&= \frac{1}{n-1}\sum \left(x_i - \frac{1}{n}\sum x_i\right)^2 \\
&= \frac{1}{n-1}\sum \left[{x_i}^2 - \frac{2}{n}x_i\left(\sum x_i\right) + \frac{1}{n^2}\left(\sum x_i\right)^2\right] \\
&= \frac{1}{n-1}\left[\sum \left({x_i}^2\right) - \frac{2}{n}\left(\sum x_i\right)\left(\sum x_i\right) + \frac{n}{n^2}\left(\sum x_i\right)^2\right] \\
&= \frac{1}{n-1}\left[\sum \left({x_i}^2\right) - \frac{1}{n}\left(\sum x_i\right)^2\right] \\
&= \frac{n\sum \left({x_i}^2\right) - \left(\sum x_i\right)^2}{n(n-1)}.
\end{aligned}$$

4.3 Maximum Likelihood Estimates

4.3.1 The MLE of λ is $\bar{x} = (3+6+5)/3 = 14/3$.

4.3.2 The likelihood function is

$$\begin{aligned}
L(p) &= f(2)\cdot f(5)\cdot f(6) \\
&= p\,(1-p)^{2-1}\cdot p\,(1-p)^{5-1}\cdot p\,(1-p)^{6-1} \\
&= p^3(1-p)^{11}.
\end{aligned}$$

Then

$$\begin{aligned}
L'(p) &= 3p^2(1-p)^{11} + p^3(11)(1-p)^{10}(-1) \\
&= p^2(1-p)^{10}(3-14p).
\end{aligned}$$

Setting this derivative equal to 0 and solving yields $p = 0,\ 3/14,\ 1$. Since $L(0) = L(1) = 0$, the MLE is $\hat{p} = 3/14$.

4.3.3 The likelihood function is

$$\begin{aligned}
L(\theta) &= f(0.8)\cdot f(1.5)\cdot f(2.3) \\
&= \frac{3(0.8)^2}{\theta^3}\,e^{-(0.8/\theta)^3}\cdot\frac{3(1.5)^2}{\theta^3}\,e^{-(1.5/\theta)^3}\cdot\frac{3(2.3)^2}{\theta^3}\,e^{-(2.3/\theta)^3} \\
&= \frac{205.6752}{\theta^9}\,e^{-16.054/\theta^3}
\end{aligned}$$

Then

$$L'(\theta) = 205.6752\left(-9\theta^{-10}\right)e^{-16.054/\theta^3} + \frac{205.6752}{\theta^9}e^{-16.054/\theta^3}\left[-16.054\left(-3\theta^{-4}\right)\right]$$
$$= \frac{e^{-16.054/\theta^3}}{\theta^{10}}\left(-1851.0768 + 9905.729\theta^{-3}\right).$$

Setting this derivative equal to 0 and solving yields $\theta \approx 1.749$. Thus the MLE is $\hat{\theta} = 1.749$.

4.3.4 The likelihood function is

$$L(\theta) = f(1.8) \cdot f(5.2) \cdot f(9.6) = \frac{1}{\theta} \cdot \frac{1}{\theta} \cdot \frac{1}{\theta} = \frac{1}{\theta^3}.$$

Since $L(\theta)$ is a decreasing function, L is maximized when θ is as small as possible. Since every observed value of X must be less than θ, the smallest possible value of θ, based on the sample, is 9.6. Thus the MLE of θ is $\hat{\theta} = 9.6$.

4.3.5 The likelihood function is

$$L(p) = f(y_1)\cdots f(y_1)$$
$$= p^{y_1}(1-p)^{1-y_1}\cdots p^{y_2}(1-p)^{1-y_2}$$
$$= P^{(y_1+\cdots+y_n)}(1-p)^{n-(y_1+\cdots+y_n)}$$
$$= p^k(1-p)^{n-k}.$$

Then

$$L'(p) = kp^{k-1}(1-p)^{n-k} + p^k(n-k)(1-p)^{n-k-1}(-1)$$
$$= p^{k-1}(1-p)^{n-k-1}(k-pn).$$

Setting this derivative equal to 0 and solving yields $p = k/n$. Thus the MLE is $\hat{p} = k/n$.

4.3.6 Note that since X is $b(n, p)$, $E(X) = np$. Then since n is a constant,

$$E\left(\hat{Y}\right) = E\left[\frac{X}{n}\right] = \frac{1}{n}E(X) = \frac{1}{n}(np) = p.$$

Thus $\hat{P}$ is an unbiased estimator of p by definition.

4.3.7 From Example 4.3.2, we see that the likelihood function is

$$L(\lambda) = \lambda^n e^{-\lambda \sum x_i}$$

where the sum is take from $i = 1$ to n. Taking the natural logarithm of this yields

$$\ln\left[L(\lambda)\right] = \ln\left(\lambda^n\right) + \ln\left(e^{-\lambda \sum x_i}\right)$$
$$= n\ln(\lambda) - \lambda \sum x_i.$$

Taking the derivative with respect to λ yields

$$\frac{d}{d\lambda} \ln\left[L(\lambda)\right] = \frac{n}{\lambda} - \sum x_i.$$

Setting this equal to 0 and solving for λ yields $\lambda = n/\sum x_i = 1/\bar{x}$. Thus the MLE of λ is $\hat{\lambda} = 1/\bar{x}$.

4.3.8 From Example 4.3.4, the natural logarithm of the likelihood function is

$$\ln\left[L(\mu,\, \theta)\right] = -\frac{n}{2}\ln(2\pi\theta) + \frac{k}{\theta}.$$

where $\theta = \sigma^2$ and $k = -\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2$. Taking the partial derivative with respect to μ yields

$$\frac{\partial}{\partial\mu}\left[\ln\left(L(\mu,\, \theta)\right)\right] = 0 + \frac{1}{\theta}\frac{\partial k}{\partial\mu}.$$

Now,

$$\frac{\partial k}{\partial\mu} = -\frac{1}{2}\sum_{i=1}^{n} 2\left(x_i - \mu\right)(-1) = \left(\sum_{i=1}^{n} x_i\right) - n\mu.$$

For $\frac{\partial}{\partial\mu}\left[\ln\left(L(\mu,\, \theta)\right)\right]$ to equal 0, we must have $\left(\sum_{i=1}^{n} x_i\right) - n\mu = 0$. Solving this for μ yields

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}.$$

Thus the MLE of μ is the sample mean $\bar{x}$.

4.3.9 Note that since each X_i has the same distribution as X, $\text{Var}(X) = \sigma^2 = E\left[(X_i - \mu)^2\right]$ for each i. Thus by linearity properties of the expected value, we have

$$\begin{aligned}
E\left(\hat{\Sigma}^2\right) &= E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} E\left(X_i - \mu\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\sigma^2 \\
&= \frac{1}{n}\left(n\sigma^2\right) \\
&= \sigma^2.
\end{aligned}$$

Thus $\hat{\Sigma}^2$ is an unbiased estimator of σ^2 by definition.

4.4 Sampling Distributions

4.4.1

a. $\hat{p} = 296/500 = 0.592$
b. Assuming $p = 0.5$, then in $n = 500$ trials, $\hat{P}$ is approximately $N(0.5,\, 0.5(1-0.5)/500) = N(0.5,\, 0.0005)$ so that

$$P\left(\hat{P} > 0.592\right) \approx P\left(Z > \frac{0.592 - 0.5}{\sqrt{0.0005}}\right) = P(Z > 4.11) = 0.0000.$$

c. Since the probability in part b. is so small, it suggests the assumption that $p = 0.5$ is not valid.

4.4.2

a. Assuming $\mu = 98.6$ and $\sigma = 0.58$, the central limit theorem says that $\bar{X}_{110}$ is approximately $N\left(98.6,\, 0.58^2/110\right) = N(98.6,\, 0.00306)$ so that

$$P\left(\bar{X}_{110} \le 98.1\right) \approx P\left(Z \le \frac{98.1 - 98.6}{\sqrt{0.00306}}\right) = P(Z \le -9.04) = 0.0000.$$

b. Since the probability in part b. is so small, it suggests that the average body temperature is *not* 98.6 °F?

4.4.3

a. Since X is $b(n,\, p)$, $Var(X) = np\,(1-p)$ so that

$$Var\left(\hat{P}\right) = Var\left(\frac{X}{n}\right) = \frac{1}{n^2}\,Var(X) = \frac{1}{n^2}\,np\,(1-p) = \frac{p\,(1-p)}{n}.$$

b. As $n \to \infty$, $Var\left(\hat{P}\right) \to 0$.
c. Interpretating variance as a measure of how "spread out" the values of a random variable are, we see that as $n \to \infty$, the value of $\hat{P}$ is typically closer to p.
d. A larger sample size would typically yield a better estimate of p since a larger sample size results in a smaller value of $Var\left(\hat{P}\right)$.

4.4.4

a. Let X_i, $i = 1, \ldots, n$ denote n independently selected values from this population. Then each X_i has mean μ and variance σ^2 so that

$$Var\left(\bar{X}_n\right) = Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2}\left(n\,\sigma^2\right) = \frac{\sigma^2}{n}.$$

b. As $n \to \infty$, $Var\left(\bar{X}_n\right) \to 0$.
c. Interpretating variance as a measure of how "spread out" the values of a random variable are, we see that as $n \to \infty$, the value of $\bar{X}_n$ is typically closer to μ.
d. A larger sample size would typically yield a better estimate of μ since a larger sample size results in a smaller value of $Var\left(\bar{X}_n\right)$.
e. If σ^2 were "large," then $Var\left(\bar{X}_n\right)$ would be large. This means it is less likely that the value of $\bar{X}_n$ is "close" to μ as if σ^2 were "small." This means that a sample mean is a better estimate of μ when σ^2 is "small" than when σ^2 is "large."
f. If $\sigma^2 = 0$, then $Var\left(\bar{X}_n\right) = 0$. This means that every value of $\bar{X}_n$ would equal μ so that a sample mean would be a perfect estimate of μ.

4.4.5

a. If $n = 25$ and $p = 0.85$, then $\hat{P}$ is approximately $N(0.85, 0.85(1-0.85)/25) = N(0.85, 0.0051)$ so that

$$P\left(\hat{P} > 0.95\right) \approx P\left(Z > \frac{0.95 - 0.85}{\sqrt{0.0051}}\right) = P(Z > 1.40) = 0.0808.$$

b. When $n = 25$, 50, 100, and 200, $Var\left(\hat{P}\right) = 0.0051$, 0.00255, 0.001275, and 0.0006375, respectively. Calculating $P\left(\hat{P} < 0.75\right)$ as in part a., we get probabilities of 0.0808, 0.0239, 0.0026, and 0.0000, respectively.
c. Note that 0.75 is "far" from the population proportion. In part b., we see that as n gets larger, the probability that the sample proportion is less than 0.75 (i.e. that the sample proportion is "far" from the population proportion) gets smaller. This is true in general.

4.4.6

a. The values of s^2 calculated using this alternative formula are shown in the table below.

Sample		$\bar{x}$	s^2	s
2	2	2	0	0
2	3	2.5	0.25	0.5
2	9	5.5	12.25	3.5
3	2	2.5	0.25	0.5
3	3	3	0	0
3	9	6	9	3
9	2	5.5	12.25	3.5
9	3	6	9	3
9	9	9	0	0
		Mean =	4.778	1.556

b. The mean of the nine "sample variances" is 4.778. This does not equal $\sigma^2 = 86/9 = 9.556$.

c. This formula for sample variance result in a biased estimator of σ^2. It appears under-estimate σ^2.

d. The "sample standard deviations" are shown in the table above.

e. The mean of the nine "sample standard deviations" is 1.556. This results in an even more biased estimate of $\sigma = 3.0912$.

4.4.7

a. The range of values of X is $9 - 2 = 7$ and the median of X is 3.

b. The sample ranges and medians are shown in the table below.

Sample		Range	Median
2	2	0	2
2	3	1	2.5
2	9	7	5.5
3	2	1	2.5
3	3	0	3
3	9	6	6
9	2	7	5.5
9	3	6	6
9	9	0	9
	Mean =	3.111	4.667

c. The mean of the sample ranges is 3.11 and the mean of the medians is 4.667. This shows that the sample range is a biased estimator of the population range and the sample median is a biased estimator of the population median.

d. Based on these results, a sample range is not a good estimate of a population range and a sample median is not a good estimate of the population median.

4.4.8

a. The population MAD is

$$\text{Population MAD} = \frac{1}{3}\left[|2 - 14/3| + |3 - 14/3| + |9 - 14/3|\right] = 26/38.667.$$

b. The sample MAD's are shown in the table below.

Sample		$\bar{x}$	MAD
2	2	2	0
2	3	2.5	0.5
2	9	5.5	3.5
3	2	2.5	0.5
3	3	3	0
3	9	6	3
9	2	5.5	3.5
9	3	6	3
9	9	9	0
Mean =		4.667	1.556

c. The mean of the 9 sample MAD's os 1.556 which is much smaller than the population MAD.

d. Based on this result, a sample MAD is not a good estimate of the population MAD.

4.5 Confidence Intervals for a Proportion

4.5.1 The approximate values of z_α and $z_{\alpha/2}$ are summarized in the table below.

$\boldsymbol{\alpha}$	0.4	0.3	0.005
$\boldsymbol{z_\alpha}$	0.255	0.525	2.575
$\boldsymbol{z_{\alpha/2}}$	0.84	1.035	2.81

4.5.2 $0.65 - 0.02 \le p \le 0.65 + 0.02 \quad \Rightarrow \quad 0.63 \le p \le 0.67$

4.5.3 This statement means that the true proportion could be somewhere between 0.03 and 0.88 which does not narrow down the range of possible values of the proportion very much.

4.5.4 The sample proportion is in the middle of the confidence interval, so $\hat{p} = (0.444 + 0.222)/2 = 0.333$ and the margin of error is half the width of the interval, so $E = (0.444 - 0.222)/2 = 0.111$. Thus an equivalent form of the confidence interval is 0.333 ± 0.111.

4.5.5 This is not a random sample of all college students who use credit cards so this data cannot be used to construct a *valid* confidence interval.

4.5.6 The parameter is

$$p = \text{The proportion of all college students who are left-handed.}$$

We have $n = 87$, and the sample proportion is $\hat{p} = 12/87 \approx 0.1379$. At the 90% confidence level, $\alpha = 0.10$, so the critical value is $z_{0.10/2} = 1.645$, the margin of error is

$$E = 1.645\sqrt{\frac{0.1379(1 - 0.1379)}{87}} \approx 0.0608,$$

and the confidence interval is

$$0.1379 - 0.0608 \leq p \leq 0.521 + 0.0608 \;\Rightarrow\; 0.0771 \leq p \leq 0.1987.$$

Since this confidence interval contains 0.10, the belief that about 10% of the population are left-handed appears to be reasonable.

4.5.7 The parameter is

$$p = \text{The proportion of all students who think they look more like their father.}$$

We have $n = 100$, and the sample proportion is $\hat{p} = 71/100 = 0.71$. At the 99% confidence level, $\alpha = 0.01$, so the critical value is $z_{0.01/2} = 2.576$, the margin of error is

$$E = 2.576\sqrt{\frac{0.71(1 - 0.71)}{100}} \approx 0.1168,$$

and the confidence interval is

$$0.71 - 0.1168 \leq p \leq 0.71 + 0.1168 \;\Rightarrow\; 0.5932 \leq p \leq 0.8268.$$

Since 0.5 is less than the lower limit of this confidence interval, the claim appears to be valid.

4.5.8 For all the confidence intervals, $n = 735$ and $\hat{p} = 383/735 = 0.521$. The calculations are summarized in the table below. We see that as the confidence level increases, the margin of error and the width of the confidence interval also increase.

	C.V.	***E***	**Confidence Interval**
80% C.L.	1.282	0.024	$0.497 \leq p \leq 0.545$
90% C.L.	1.645	0.030	$0.491 \leq p \leq 0.551$
99% C.L.	2.576	0.047	$0.474 \leq p \leq 0.568$

4.5.9 For all the confidence intervals, $\hat{p} = 0.64$ and the critical value is $z_{0.05/2} = 1.96$. The calculations are summarized in the table below. We see that as the sample size increases, the margin of error and the width of the confidence interval decrease. This means that as n increases, the quality of $\hat{p}$ as as estimate of p improves.

n	***E***	**Confidence Interval**
100	0.094	$0.546 \leq p \leq 0.734$
200	0.067	$0.573 \leq p \leq 0.707$
500	0.042	$0.598 \leq p \leq 0.682$

4.5.10

a. The parameter is

$$p = \text{The proportion of all adults that are "pro-choice."}$$

We have $n = 1002$, and the sample proportion is $\hat{p} = 531/1000 = 0.531$. At the 90% confidence level, $\alpha = 0.10$, so the critical value is $z_{0.10/2} = 1.645$, the margin of error is

$$E = 1.645\sqrt{\frac{0.531(1-0.531)}{1000}} \approx 0.0260,$$

and the confidence interval is

$$0.531 - 0.0260 \leq p \leq 0.531 + 0.0260 \Rightarrow 0.505 \leq p \leq 0.557.$$

Since 0.5 is less than the lower limit of this confidence interval, it appear that a majority are "pro-choice."

b. At the 99% confidence level, the critical value is $z_{0.01/2} = 2.576$. Redoing the calculations as in part a. yields $E = 0.0406$ and the confidence interval $0.4904 \leq p \leq 0.5716$. Since this interval contains 0.5, we conclude that it is possible that $p < 0.5$ which is a different conclusion than in part a.

4.5.11

a. The parameter is

$$p = \text{The proportion of all readers who like the new design.}$$

We have $n = 540$, and the sample proportion is $\hat{p} = 309/500 = 0.618$. At the 90% confidence level, $\alpha = 0.10$, so the critical value is $z_{0.10/2} = 1.645$, the margin of error is

$$E = 1.645\sqrt{\frac{0.618(1-0.618)}{500}} \approx 0.0253,$$

and the confidence interval is

$$0.618 - 0.0253 \leq p \leq 0.618 + 0.0253 \Rightarrow 0.5927 \leq p \leq 0.6433.$$

b. Since 0.5 is less than the lower limit of this confidence interval, it appear that most readers like the new design.

c. This is a "voluntary response" sample. Most times samples like this consist mostly of people with very strong opinions, so this is not a random sample. Since it is a bad sample, we must question the conclusion.

4.5.12

a. The parameter is

$$p = \text{The proportion of all flips of a coin that the student correctly predicts.}$$

We have $n = 75$, and the sample proportion is $\hat{p} = 43/75 = 0.5733$. At the 95% confidence level, $\alpha = 0.05$, so the critical value is $z_{0.05/2} = 1.96$, the margin of error is

$$E = 1.96\sqrt{\frac{0.5733(1 - 0.5733)}{75}} \approx 0.1119,$$

and the confidence interval is

$$0.5733 - 0.1119 \leq p \leq 0.5733 + 0.1119 \Rightarrow 0.4614 \leq p \leq 0.6852.$$

b. Since this interval contains 0.5, it does not appear that p is significantly different from 0.5. Thus the results could have happened by chance, so the student's claim does not appear to be valid.

4.5.13

a. The parameter is

$$p = \text{The proportion of all voters who support the candidate.}$$

We have $n = 500$, and the sample proportion is $\hat{p} = 0.54$. At the 95% confidence level, $\alpha = 0.05$, so the critical value is $z_{0.05/2} = 1.96$, the margin of error is

$$E = 1.96\sqrt{\frac{0.54(1 - 0.54)}{500}} \approx 0.0437,$$

and the confidence interval is

$$0.54 - 0.0437 \leq p \leq 0.54 + 0.0437 \Rightarrow 0.4963 \leq p \leq 0.5837.$$

b. Using similar calculations as in part a., with $n = 1000$ and $\hat{p} = 0.54$, we get $E = 0.03089$ and a confidence interval of $0.509 \leq p \leq 0.571$. Notice that the margin of error decreased by about 0.013 and that now this lower limit is greater than 0.5. Thus we are more confident that the candidate has a majority support.

c. With $n = 2000$ and $\hat{p} = 0.54$, we get $E = 0.02184$ and a confidence interval of $0.5186 \leq p \leq 0.56184$. Notice that the margin of error decreased by about 0.011, less than the decrease from part a. to part b. So the additional cost of surveying another 1000 voters probably was not worth it.

4.5.14

a. The parameter is

$$p = \text{The proportion of all voters who say they voted for the proposition.}$$

We have $n = 825$, and the sample proportion is $\hat{p} = 429/825 = 0.52$. At the 99% confidence level, $\alpha = 0.01$, so the critical value is $z_{0.01/2} = 2.576$, the margin of error is

$$E = 2.576\sqrt{\frac{0.52(1-0.52)}{825}} \approx 0.0448,$$

and the confidence interval is

$$0.52 - 0.0448 \le p \le 0.52 + 0.0448 \;\Rightarrow\; 0.4752 \le p \le 0.5648.$$

b. The lower limit of the interval is much larger than 0.4 which suggests that much more than 40% of voters *say* they voted for the proposition. This suggests the voters are not being honest and saying they voted for it when they actually did not.

4.5.15

a. The parameter is

$$p = \text{The proportion of all essays that are graded within two days.}$$

We have $n = 200$, and the sample proportion is $\hat{p} = 168/200 = 0.84$. At the 99% confidence level, $\alpha = 0.01$, so the critical value is $z_{0.01} = 2.326$, the margin of error is

$$E = 2.326\sqrt{\frac{0.84(1-0.84)}{200}} \approx 0.0603,$$

and the confidence interval is

$$0.84 - 0.0603 \le p \le 1 \;\Rightarrow\; 0.7797 \le p \le 1.$$

b. Since this lower limit is greater than 0.75, the confidence interval does support the professor's claim.

4.5.16 The parameter is

$$p = \text{The proportion of all new-born piglets that are female.}$$

We have $n = 201$, and the sample proportion is $\hat{p} = 90/201 = 0.4478$. At the 90% confidence level, $\alpha = 0.10$, so the critical value is $z_{0.10} = 1.282$, the margin of error is

$$E = 1.282\sqrt{\frac{0.4478(1-0.4478)}{201}} \approx 0.0450,$$

and the confidence interval is

$$0 \le p \le 0.4478 + 0.0450 \;\Rightarrow\; 0 \le p \le 0.4928.$$

Since this upper limit is less than 0.5, it does appear that less than half of all new-born piglets are female.

4.5.17 Since we are using a 90% confidence level, the probability that a randomly selected sample will yield a confidence interval that contains the true value of p is 0.9. Let X denote the number of intervals out of the 50 that contain the true value of p. Then X is $b(50,\,0.9)$ so that using software we get

$$P(X \ge 46) = P(X = 46) + \cdots + P(X = 50) = 0.4311.$$

4.5.18 Note that by multiplying by $\sqrt{p\,(1-p)/n}$ we get

$$-z_{\alpha/2} \le \frac{\hat{P} - p}{\sqrt{p\,(1-p)/n}} \le z_{\alpha/2} \;\Rightarrow\; -z_{\alpha/2}\sqrt{\frac{p\,(1-p)}{n}} \le \hat{P} - p \le z_{\alpha/2}\sqrt{\frac{p\,(1-p)}{n}}$$

then subtracting $\hat{P}$ we get

$$-\hat{P} - z_{\alpha/2}\sqrt{\frac{p\,(1-p)}{n}} \le -p \le -\hat{P} + z_{\alpha/2}\sqrt{\frac{p\,(1-p)}{n}}$$

and finally multiplying by -1 and reversing the inequalities we get

$$\hat{P} - z_{\alpha/2}\sqrt{\frac{p\,(1-p)}{n}} \le p \le \hat{P} + z_{\alpha/2}\sqrt{\frac{p\,(1-p)}{n}}.$$

4.5.19

a. At the 95% confidence level, the critical value is $z_{0.05/2} = 1.96$. For the adjusted Wald method with $x = 20$ and $n = 50$,

$$\hat{p} = \frac{20+2}{40+4} = 0.4074, \quad \text{and} \quad E = 1.96\sqrt{\frac{0.4074\,(1-0.4074)}{50+4}} = 0.1311$$

and the confidence interval is

$$0.4074 - 0.1311 \le p \le 0.4074 + 0.1311 \;\Rightarrow\; 0.2763 \le p \le 0.5385.$$

b. For the Wilson Score method, $\hat{p} = 20/50 = 0.4$ and the confidence interval is

$$\frac{0.4 + \frac{1.96^2}{2(50)} \pm 1.96\sqrt{\frac{0.4(1-0.4)+\frac{1.96^2}{4(50}}{50}}}{1 + \frac{1.96^2}{50}} \;\Rightarrow\; 0.2761 \le p \le 0.5388$$

c. Using methods as in previous exercises, the Wald interval is $0.3696 \leq p \leq 0.4304$ and using calculations similar to part a., the Adjusted Wald interval is $0.3701 \leq p \leq 0.4307$. There is very little difference between these intervals. For "large" values of n, $n+4$ is not that much different than n, and $x+2$ is not that much different from x (assuming x is also "large"). Thus for large n the formulas for $\hat{p}$ and E in the adjusted Wald method are almost identical to those for the Wald method. So the resulting confidence intervals are almost identical.

4.6 Confidence Intervals for a Mean

4.6.1 The parameter is

$$\mu = \text{The mean runtime of all G-rated movies.}$$

We have $n = 15$. At the 95% confidence level, the critical value is $t_{0.05/2}(15-1) = 2.145$ so that the margin of error is $E = 2.145\,\frac{21.2}{\sqrt{15}} \approx 11.741$ and the confidence interval is

$$80.6 - 11.741 \leq \mu \leq 80.6 + 11.741 \;\Rightarrow\; 68.859 \leq \mu \leq 92.341.$$

4.6.2 The parameter is

$$\mu = \text{The mean run time of all top-ten finishers in the GPAC.}$$

We have $n = 10$. At the 90% confidence level, the critical value is $t_{0.10/2}(10-1) = 1.833$ so that the margin of error is $E = 1.833\,\frac{19.8}{\sqrt{10}} \approx 11.5$ and the confidence interval is

$$1606 - 11.5 \leq \mu \leq 1606 + 11.5 \;\Rightarrow\; 1594.5 \leq \mu \leq 1617.5.$$

4.6.3

a. We have $n = 50$. At the 90% confidence level, the critical value is $t_{0.05/2}(50-1) \approx 2.009$. The confidence intervals are summarized in the table below

s	E	C.I.
75	21.31	$213.69 \leq \mu \leq 256.31$
55	15.63	$219.37 \leq \mu \leq 250.63$
35	9.94	$225.06 \leq \mu \leq 244.94$
15	4.26	$230.74 \leq \mu \leq 239.26$

b. As the sample standard deviation decreases, the margin of error also decreases.

c. If we want a good estimate of a population mean, we would want a sample with a small standard deviation so that we have a small margin of error and narrow confidence interval.

4.6.4

a. Note that since σ and $z_{\alpha/2}$ are fixed values,

$$\lim_{n\to\infty} E = \lim_{n\to\infty} z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 0.$$

b. A larger sample size provides a better estimate of μ since the margin of error is smaller.

4.6.5

a. The parameter is

$$\mu = \text{The mean ratio (overall height)/(navel height of all students at the university.}$$

We have $n = 39$. At the 99% confidence level, the critical value is $t_{0.01/2}(39-1) \approx 2.704$ so that the margin of error is $E = 2.704\,\frac{0.0474}{\sqrt{39}} \approx 0.0205$ and the confidence interval is

$$1.6494 - 0.0205 \le \mu \le 1.6494 + 0.0205 \;\Rightarrow\; 1.6289 \le \mu \le 1.6699.$$

b. Since this interval contains 1.6494, the results do support the claim.

c. These results cannot be used to make any conclusion about the mean ratio of all people in the world because the sample is not a random sample of people from all over the world.

4.6.6

a. The calculations of the confidence intervals are summarized in the table below.

	$\bar{x}$	s	n	**C.V.**	E	**C.I.**
BN Sodium	137.79	60.69	74	1.665	11.75	$126.04 \le \mu \le 149.54$
Gen Sodium	158.22	73.3	62	1.671	15.56	$142.66 \le \mu \le 173.78$
BN Sugar	7.77	3.36	74	1.665	0.65	$7.12 \le \mu \le 8.42$
Gen Sugar	7.48	4.49	62	1.671	0.95	$6.53 \le \mu \le 8.43$

b. The intervals for sodium overlap, and so do the intervals for sugar. So there does not appear to be a significant difference between the brands.

4.6.7

a. The calculations of the confidence intervals are summarized in the table below.

	$\bar{x}$	s	n	**C.V.**	E	**C.I.**
Buttered	81.75	32.8976	4	3.182	52.34	$29.41 \le \mu \le 134.09$
Kettle	29.5	19.0875	4	3.182	30.37	$-0.87 \le \mu \le 59.87$

b. Since these intervals overlap, there does not appear to be a significant difference between these two population means.

4.6.8

a. The calculations of the confidence intervals are summarized in the table below.

	$\bar{x}$	s	n	C.V.	E	Confidence Interval
Tan	2.511	0.268	30	2.042	0.1	$2.41 \leq \mu \leq 2.61$
Black	2.621	0.293	30	2.042	0.11	$2.51 \leq \mu \leq 2.73$

b. Since these intervals overlap, there does not appear to be a significant difference between these two population means.

4.6.9 Note that by multiplying by $\sigma/\sqrt{n}$ we get

$$-z_{\alpha/2} \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2} \quad \Rightarrow \quad -z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \bar{X}_n - \mu \leq z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

then subtracting $\bar{X}_n$ we get

$$-\bar{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

and multiplying by -1 and reversing the inequalities we get

$$\bar{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

4.6.10

a. The parameter is

$$\mu = \text{The mean height of all supermodels.}$$

We have $n = 9$. At the 99% confidence level, the critical value is $t_{0.01}(9 - 1) = 2.896$ so that the margin of error is $E = 2.896\,\frac{1.4}{\sqrt{9}} \approx 1.35$ and the confidence interval is

$$70.3 - 1.35 \leq \mu < \infty \;\Rightarrow\; 68.95 \leq \mu < \infty.$$

b. The lower confidence interval limit is much greater than 63.5. This suggests that supermodels are taller than typical women.

4.6.11

a. The parameter is

$$\mu = \text{The mean number of times students call home each week.}$$

We have $n = 150$. At the 90% confidence level, the critical value is $t_{0.10/2}(150 - 1) \approx 2.351$ so that the margin of error is $E = 1.655\,\frac{1.85}{\sqrt{150}} \approx 0.250$. The finite population correction factor is

$$C = \sqrt{\frac{650 - 150}{650 - 1}} \approx 0.878,$$

and the confidence interval is

$$2.3 - 0.878(0.25) \le \mu \le 2.3 + 0.878(0.25) \Rightarrow 2.081 \le \mu \le 2.520.$$

b. The calculations of the confidence intervals are summarized in the table below. The confidence interval for μ_1 is narrower than the interval for μ_2 this suggests that 10 is probably a slightly better estimate of μ_1 than μ_2.

	$\bar{x}$	s	n	N	C.V.	E	C	C.I.
Pop 1	10	1	100	200	1.66	0.17	0.70888	$9.88 \le \mu_1 \le 10.12$
Pop 2	10	1	100	5000000	1.66	0.17	0.99999	$9.83 \le \mu_2 \le 10.17$

c. If n is fixed, then $\lim_{N\to\infty} C = \lim_{N\to\infty} \sqrt{\dfrac{N-n}{N-1}} = 1$. This mean that when the population size is very large, the correction factor will not significantly change the confidence interval, so the factor is not necessary.

d. When $n = N$, $C = 0$. This means that $\bar{x}$ is a perfect estimate of μ.

4.6.12 a. We have UCL: $y = 62.1 + 3(0.813)/\sqrt{5} = 63.19$ and LCL: $y = 62.1 - 3(0.813)/\sqrt{5} = 61.01$ and the center line is $y = 62.1$. The control chart is shown in Figure 4.12.

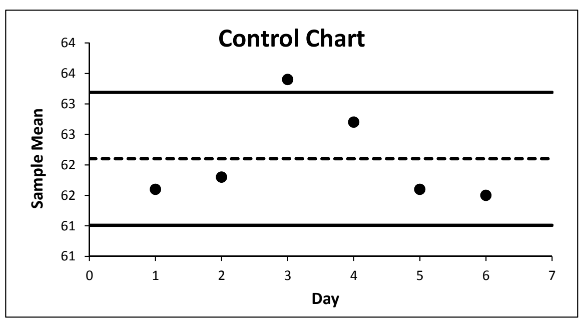


Figure 4.12

b. From the control chart, we see that one sample mean is above the UCL, so the process is not in statistical control.

4.7 Confidence Intervals for a Variance

4.7.1 The parameter is

$$\sigma^2 = \text{The variance of the tensil strength of this type of steel sheet.}$$

We have $n = 26$. At the 90% confidence level the critical values are

$$a = \chi^2_{1-0.1/2}(26-1) = 14.61 \quad \text{and} \quad b = \chi^2_{0.1/2}(26-1) = 37.65.$$

so that the confidence interval estimate of σ^2 is

$$\frac{(26-1)2.41}{37.65} \le \sigma^2 \le \frac{(26-1)2.41}{14.61} \Rightarrow 1.600 \le \sigma^2 \le 4.124.$$

4.7.2 The parameter is

$$\sigma = \text{The standard deviation of the average velocity of all baseballs hit with a metal bat.}$$

We have $n = 26$. At the 95% confidence level the critical values are

$$a = \chi^2_{1-0.05/2}(26-1) = 13.12 \quad \text{and} \quad b = \chi^2_{0.05/2}(26-1) = 40.65.$$

so that the confidence interval estimate of σ^2 is

$$\frac{(26-1)15.58^2}{40.65} \le \sigma^2 \le \frac{(26-1)15.58^2}{13.12} \Rightarrow 149.3 \le \sigma^2 \le 462.5$$

and the confidence interval of σ is $12.22 \le \sigma \le 21.51$.

4.7.3 The parameter is

$$\sigma = \text{The standard deviation of the mass of all such candies.}$$

From the data we have $n = 17$ and $s = 0.0509$. At the 99% confidence level the critical values are

$$a = \chi^2_{1-0.01/2}(17-1) = 5.142 \quad \text{and} \quad b = \chi^2_{0.01/2}(17-1) = 34.27.$$

so that the confidence interval estimate of σ^2 is

$$\frac{(17-1)0.0509^2}{34.27} \le \sigma^2 \le \frac{(17-1)0.0509^2}{5.142} \Rightarrow 0.0012 \le \sigma^2 \le 0.0081$$

and the confidence interval of σ is $0.0348 \le \sigma \le 0.0898$.

4.7.4 The parameter is

$$\sigma^2 = \text{The variance of the lengths of all such bolts.}$$

We have $n = 50$ and $s^2 = 0.0023$. At the 99% confidence level the critical value for an interval giving an upper bound on σ^2 is

$$a = \chi^2_{1-0.01}(50-1) \approx 29.71$$

so that a one-sided confidence interval estimate of σ^2 is

$$0 \le \sigma^2 \le \frac{(50-1)0.0023}{29.71} \Rightarrow 0 \le \sigma^2 \le 0.0038$$

Since this upper limis is less than 0.004, it appears that the manufacturing process meets the specifications.

4.7.5 The calculations of the confidence intervals are summarized in the table below. Since these intervals overlap, there does not appear to be a significant difference between the variances of these two populations.

	s^2	n	a	b	C.I.
Male	35.05	18	8.672	27.59	$21.60 \le \sigma_M^2 \le 68.71$
Female	18.40	8	2.167	14.07	$9.15 \le \sigma_F^2 \le 59.44$

4.7.6 Suppose a random sample of size $n = 10$ is taken from a normally distributed population with variance $\sigma^2 = 1.5$.

a. As shown in the hint, $P\left(S^2 \le 2.446\right) \approx P\left(C \le 14.68\right)$ where C is $\chi^2(9)$. Examining Table C.4, we see that $\chi^2_{0.1}(9) = 14.68$. This means that $P\left(C \le 14.68\right) = 1 - 0.1 = 0.9$. Thus $P\left(S^2 \le 2.446\right) \approx 0.9$.
b. By similar arguments as in part a., the event $S^2 \le 0.83955$ corresponds to the event $C \le (31-1)(0.83955)/1.5 = 16.791$ where C is $\chi^2(31-1)$. Examining Table C.4, we see that $\chi^2_{0.975}(30) = 16.79$. This means that $P\left(C \le 16.97\right) = 1 - 0.975 = 0.025$. Thus $P\left(S^2 \le 0.83955\right) \approx 0.025$.

4.7.7 A confidence interval estimate of σ^2 is not symmetric around s^2, meaning s^2 is not halfway between the upper and lower limits. So it is not appropriate to express such a confidence interval in the form $s^2 \pm E$ where E is half the width of the confidence interval.

4.7.8 By taking the recriprocal and reversing the inequalities, we get

$$a \le \frac{(n-1)S^2}{\sigma^2} \le b \Rightarrow \frac{1}{b} \le \frac{\sigma^2}{(n-1)S^2} \le \frac{1}{a}.$$

Then multiplying by $(n-1)S^2$ yields

$$\frac{(n-1)S^2}{b} \le \sigma^2 \le \frac{(n-1)S^2}{a}.$$

4.7.9

a. For $\alpha = 0.05$ and $r = 100$, $z_{0.5/2} = 1.96$,

$$a \approx \frac{1}{2}\left[-1.96 + \sqrt{2(100)-1}\right]^2 = 73.77, \quad b \approx \frac{1}{2}\left[1.96 + \sqrt{2(100)-1}\right]^2 = 129.07$$

From Table C.4, we find that the exact values of a and b are $a = \chi^2_{1-0.5/2}(100) = 74.22$ and $b = \chi^2_{0.05/2}(100) = 129.6$. The relative errors are $|73.77 - 74.22|/74.22 = 0.006$ and $|129.07 - 129.6|/129.6 = 0.004$. These relative errors are extremely small, so the estimates are very good.

b. For $\alpha = 0.01$ and $r = 15$, $z_{0.01/2} = 2.576$,

$$a \approx \frac{1}{2}\left[-2.576 + \sqrt{2(15)-1}\right]^2 = 3.949, \quad b \approx \frac{1}{2}\left[2.576 + \sqrt{2(15)-1}\right]^2 = 31.682$$

From Table C.4, we find that the exact values of a and b are $a = \chi^2_{1-0.01/2}(15) = 4.601$ and $b = \chi^2_{0.01/2}(15) = 32.80$. The relative errors are $|3.949 - 4.601|/4.601 = 0.1417$ and $|31.682 - 32.80|/32.80 = 0.0341$. These relative errors are much larger than in part a. This suggests that the approximation is better for larger values of r.

c. For $\alpha = 0.1$ and $r = 150$, $z_{0.1/2} = 1.645$,

$$a \approx \frac{1}{2}\left[-1.645 + \sqrt{2(150)-1}\right]^2 = 122.4, \quad b \approx \frac{1}{2}\left[1.645 + \sqrt{2(150)-1}\right]^2 = 179.3$$

so that the confidence interval estimate of σ^2 is

$$\frac{(150-1)2.36}{179.3} \le \sigma^2 \le \frac{(150-1)2.36}{122.4} \;\Rightarrow\; 1.961 \le \sigma^2 \le 2.873$$

4.7.10 The calculations are summarized in the table below. We see that $s_p^2 = 0.1773/14 = 0.01266$ and $s_p = \sqrt{0.01266} = 0.11253$.

Day	$\boldsymbol{s_i^2}$	$\boldsymbol{n_i}$	$\boldsymbol{(n_i-1)\,s_i^2}$	$\boldsymbol{n_i - 1}$
1	0.0085	5	0.0339	4
2	0.0065	3	0.0131	2
3	0.0235	4	0.0704	3
4	0.0265	2	0.0265	1
5	0.0084	5	0.0335	4
		Sum =	0.1773	14

4.8 Confidence Intervals for Differences

4.8.1 The parameters are

p_1 = the proportion of all plain M&M's that are blue, and
p_2 = the proportion of all peanut M&M's that are blue.

We have $\hat{p}_1 = 62/419 \approx 0.148$ and $\hat{p}_2 = 31/142 \approx 0.218$ so that $\hat{p}_1 - \hat{p}_2 = -0.07$ and

$$\hat{p} = \sqrt{\frac{0.148\,(1-0.148)}{419} + \frac{0.218\,(1-0.218)}{142}} \approx 0.0387.$$

At the 95% confidence level, $z_{0.05/2} = 1.96$ so that the margin of error is $E = (0.0387)1.96 \approx 0.0759$ and the confidence interval is

$$-0.07 - 0.0759 \leq (p_1 - p_2) \leq -0.07 + 0.0759 \quad \Rightarrow \quad -0.1459 \leq (p_1 - p_2) \leq 0.0059.$$

Since this interval contains 0, it appears that the two types of M&M's contain approximately the same proportion of blue candies.

4.8.2 The parameters are

$p_1 =$ the proportion of all patients given the medication that experience drowsiness, and
$p_2 =$ the proportion of all patients given a placebo that experience drowsiness.

We have $\hat{p}_1 = 39/1500 = 0.026$ and $\hat{p}_2 = 19/1550 \approx 0.0123$ so that $\hat{p}_1 - \hat{p}_2 = 0.0137$ and

$$\hat{p} = \sqrt{\frac{0.026\,(1 - 0.026)}{1500} + \frac{0.0123\,(1 - 0.0123)}{1550}} \approx 0.00497.$$

At the 95% confidence level, $z_{0.05/2} = 1.96$ so that the margin of error is $E = (0.00497)1.96 \approx 0.00974$ and the confidence interval is

$$0.0137 - 0.00974 \leq (p_1 - p_2) \leq 0.0137 + 0.00974 \quad \Rightarrow \quad 0.00396 \leq (p_1 - p_2) \leq 0.0234.$$

Since the lower limit of this interval is positive, it suggests that the difference is positive so that those who receive the medicine appear to have a higher proportion of drowsiness. However, the upper limit is 0.0234 suggesting that the difference is very small. Because these proportions are so low, drowsiness is not a major concern.

4.8.3 The parameters are defined in the exercise. We have $\hat{p}_1 = 228/(228 + 255) = 0.472$ and $\hat{p}_2 = 503/(503 + 495) \approx 0.504$ so that $\hat{p}_1 - \hat{p}_2 = -0.032$ and

$$\hat{p} = \sqrt{\frac{0.472\,(1 - 0.472)}{483} + \frac{0.504\,(1 - 0.504)}{998}} \approx 0.0277.$$

At the 95% confidence level, $z_{0.05/2} = 1.96$ so that the margin of error is $E = (0.0277)1.96 \approx 0.0543$ and the confidence interval is

$$-0.032 - 0.0543 \leq (p_1 - p_2) \leq -0.032 + 0.0543 \quad \Rightarrow \quad -0.0863 \leq (p_1 - p_2) \leq 0.0223.$$

Because this interval contains 0, there does not appear to be much difference between morning and evening in terms of the proportion of male users.

4.8.4 The parameters are

$p_1 =$ the proportion of all residents living near the location that are in favor, and
$p_2 =$ the proportion of all residents living far from the location that are in favor.

a. We have $\hat{p}_1 = 94/125 = 0.752$ and $\hat{p}_2 = 108/135 = 0.8$ so that $\hat{p}_1 - \hat{p}_2 = -0.048$. This difference is small, so there does not appear to be much difference.

b. We have

$$\hat{p} = \sqrt{\frac{0.752\,(1-0.752)}{125} + \frac{0.8\,(1-0.8)}{135}} \approx 0.0517.$$

At the 95% confidence level, $z_{0.05/2} = 1.96$ so that the margin of error is $E = (0.0517)1.96 \approx 0.101$ and the confidence interval is

$$-0.048 - 0.101 \le (p_1 - p_2) \le -0.048 + 0.101 \quad \Rightarrow \quad -0.149 \le (p_1 - p_2) \le 0.053.$$

Because this interval contains 0, there does not appear to be much difference between the two groups of citizens?

c. Let $p =$ the proportion of all residents that are in favor. Combining the samples into one large sample, we get $\hat{p} = (94 + 108)/(125 + 135) = 0.777$ and the margin of error is

$$E = 1.96\sqrt{\frac{0.777(1-0.777)}{125+135}} \approx 0.0506,$$

and the confidence interval is

$$0.777 - 0.0506 \le p \le 0.777 + 0.0506 \;\Rightarrow\; 0.726 \le p \le 0.828.$$

Since this lower limit is greater than 0.7, it appears that there is at least 70% support.

4.8.5 Since proportions are between 0 and 1, the largest difference between two proportions occurs when one is 1 and the other is 0 which results in a difference of absolute value 1. The smallest difference occurs when the two are equal, resulting in a difference of 0.

4.8.6 The parameters are

$$\mu_1 = \text{the mean volume of all bottles of the first brand, and}$$
$$\mu_2 = \text{the mean volume of all bottles of the second brand.}$$

We have $n_1 = 4$, $\bar{x}_1 = 510.75$, $s_1 = 1.25$, $n_2 = 4$, $\bar{x}_2 = 505$, $s_2 = 0.816$, $\bar{x}_1 - \bar{x}_2 = 5.75$. The pooled standard deviation is

$$s_p = \sqrt{\frac{(4-1)1.25^2 + (4-1)0.816^2}{4+4-2}} = 1.0555.$$

At the 95% confidence level, the critical value is $t_{0.05/2}(4+4-2) = 2.447$ so that the margin of error is $E = 1.0555(2.447)\sqrt{1/4 + 1/4} = 1.826$ and the confidence interval is

$$5.75 - 1.826 \le (\mu_1 - \mu_2) \le 5.75 + 1.826 \quad \Rightarrow \quad 3.924 \le (\mu_1 - \mu_2) \le 7.576.$$

Since the lower limit is positive, it suggests that there is a significant difference between the population means.

4.8.7 The parameters are

$$\mu_1 = \text{the mean average velocity of all balls hit with a metal bat, and}$$
$$\mu_2 = \text{the mean average velocity of all balls hit with a wood bat.}$$

We have $n_1 = 30$, $\bar{x}_1 = 95.83$, $s_1 = 15.678$, $n_2 = 30$, $\bar{x}_2 = 81.16$, $s_2 = 10.279$, $\bar{x}_1 - \bar{x}_2 = 14.67$. The degrees of freedom are

$$r = \frac{\left(\frac{15.678^2}{30} + \frac{10.279^2}{30}\right)^2}{\frac{1}{30-1}\left(\frac{15.678^2}{30}\right)^2 + \frac{1}{30-1}\left(\frac{10.279^2}{30}\right)^2} = 50.04$$

so that at the 95% confidence level, the critical value is $t_{0.05/2}(50) = 2.009$ so that the margin of error is $E = 2.009\sqrt{15.678^2/30 + 10.279^2/30} = 6.877$ and the confidence interval is

$$14.67 - 6.877 \leq (\mu_1 - \mu_2) \leq 14.67 + 6.877 \quad \Rightarrow \quad 7.793 \leq (\mu_1 - \mu_2) \leq 21.547.$$

Since this interval contains 10, the claim that baseballs hit with metal bats travel on average about 10 mph fasterappears to be reasonable.

4.8.8 A biologist wants to compare the mean weight of bluegill and the mean weight of catfish in a lake. She captures 23 bluegill and records a mean weight of 1.25 lb with a standard deviation of 0.15 lb. She also captures 18 catfish and records a mean of 4.36 lb with a standard deviation of 1.02 lb. Assume that the weights of both populations are normally distributed.

a. The sample standard deviations are quite a bit different, so it is not reasonable to assume the population variances are equal.
b. The parameters are

$$\mu_1 = \text{the mean weight of all bluegill in the lake, and}$$
$$\mu_2 = \text{the mean weight of all catfish in the lake.}$$

We have $n_1 = 23$, $\bar{x}_1 = 1.25$, $s_1 = 0.15$, $n_2 = 18$, $\bar{x}_2 = 4.36$, $s_2 = 1.02$, $\bar{x}_1 - \bar{x}_2 = -3.11$. The degrees of freedom are

$$r = \frac{\left(\frac{0.15^2}{23} + \frac{1.02^2}{18}\right)^2}{\frac{1}{23-1}\left(\frac{0.15^2}{23}\right)^2 + \frac{1}{18-1}\left(\frac{1.02^2}{18}\right)^2} = 17.58$$

so that at the 95% confidence level, the critical value is $t_{0.05/2}(17) = 2.110$ so that the margin of error is $E = 2.110\sqrt{0.15^2/23 + 1.02^2/18} = 0.511$ and the confidence interval is

$$-3.11 - 0.511 \leq (\mu_1 - \mu_2) \leq -3.11 + 0.511 \quad \Rightarrow \quad -3.621 \leq (\mu_1 - \mu_2) \leq -2.599.$$

4.8.9 he parameters are

$$\mu_1 = \text{the mean GPA of all education majors at the university, and}$$
$$\mu_2 = \text{the mean GPA of all non-education majors at the university.}$$

We have $n_1 = 47$, $\bar{x}_1 = 3.46$, $s_1 = 0.336$, $n_2 = 53$, $\bar{x}_2 = 3.38$, $s_2 = 0.449$, $\bar{x}_1 - \bar{x}_2 = 0.08$. The calculations of the confidence intervals are summarized in the table below. We see that these confidence intervals are all very similar. Since they all contain 0, there does not appear to be much difference between the mean GPA's of education and non-education majors at this university. However, 0 is near the left end of each interval, so it might be that the education majors have a slightly higher mean GPA.

	df	**C.V.**	s_p	E	**C.I.**
Z-interval	-	1.960	-	0.1544	$-0.0744 \leq (\mu_1 - \mu_2) \leq 0.2344$
Equal Var	98	1.984	0.400	0.159	$-0.0790 \leq (\mu_1 - \mu_2) \leq 0.2390$
Non Equal Var	95.4	1.985	-	0.1564	$-0.0764 \leq (\mu_1 - \mu_2) \leq 0.2364$

4.8.10 The parameters are

$$\mu_1 = \text{the mean run time of all AA batteries of brand A, and}$$
$$\mu_2 = \text{the mean run time of all AA batteries of brand B.}$$

a. From the data, we have $n_1 = 5$, $\bar{x}_1 = 667$, $s_1 = 38.09$, $n_2 = 5$, $\bar{x}_2 = 719.4$, $s_2 = 120.8$, $\bar{x}_1 - \bar{x}_2 = -52.4$. The degrees of freedom are

$$r = \frac{\left(\frac{38.09^2}{5} + \frac{120.8^2}{5}\right)^2}{\frac{1}{5-1}\left(\frac{38.09^2}{5}\right)^2 + \frac{1}{5-1}\left(\frac{120.8^2}{5}\right)^2} = 4.79$$

so that at the 95% confidence level, the critical value is $t_{0.05/2}(4) = 2.776$ so that the margin of error is $E = 2.776\sqrt{38.09^2/5 + 120.8^2/5} = 157.2$ and the confidence interval is

$$-52.4 - 157.2 \leq (\mu_1 - \mu_2) \leq -52.4 + 157.2 \quad \Rightarrow \quad -209.6 \leq (\mu_1 - \mu_2) \leq 104.8.$$

b. Using software, the critical value is $t_{0.05/2}(4.79) = 2.605$ yielding $E = 147.6$ and a confidence interval of $[-200, 95.2]$.

c. The difference between the intervals is due to the fact that when calculating by hand, we rounded r down from 4.79 to 4 to find the critical value whereas the software did not round it down. Since r is so small, this creates a relatively large difference between the critical values used in parts a. and b.

4.8.11 The parameters are

$$\mu_1 = \text{the mean number of chocolate chips in all cookies of brand A, and}$$
$$\mu_2 = \text{the mean number of chocolate chips in all cookies of brand B.}$$

From the data, we have $n_1 = 20$, $\bar{x}_1 = 22.2$, $s_1 = 3.222$, $n_2 = 20$, $\bar{x}_2 = 24.15$, $s_2 = 4.133$, $\bar{x}_1 - \bar{x}_2 = -1.95$. The pooled standard deviation is

$$s_p = \sqrt{\frac{(20-1)3.222^2 + (20-1)4.133^2}{20+20-2}} = 3.706.$$

At the 90% confidence level, the critical value is $t_{0.1/2}(20+20-2) \approx 1.684$ so that the margin of error is $E = 3.706(1.684)\sqrt{1/20+1/20} = 1.974$ and the confidence interval is

$$-1.95 - 1.974 \leq (\mu_1 - \mu_2) \leq -1.95 + 1.974 \quad \Rightarrow \quad -3.924 \leq (\mu_1 - \mu_2) \leq 0.024.$$

Since this interval contains 0, there does not appear to be a significant difference between the brands. However, 0 is very near the upper limit, so brand B could contain more.

4.8.12 The parameters are

$$\mu_1 = \text{the mean time for all male students, and}$$
$$\mu_2 = \text{the mean time for all female students.}$$

From the data, we have $n_1 = 15$, $\bar{x}_1 = 28.373$, $s_1 = 11.228$, $n_2 = 15$, $\bar{x}_2 = 29.953$, $s_2 = 7.448$, $\bar{x}_1 - \bar{x}_2 = -1.58$. The pooled standard deviation is

$$s_p = \sqrt{\frac{(15-1)11.228^2 + (15-1)7.448^2}{15+15-2}} = 9.527.$$

At the 95% confidence level, the critical value is $t_{0.05/2}(15+15-2) = 2.048$ so that the margin of error is $E = 9.527(2.048)\sqrt{1/15+1/15} = 7.125$ and the confidence interval is

$$-1.58 - 7.125 \leq (\mu_1 - \mu_2) \leq -1.58 + 7.125 \quad \Rightarrow \quad -8.705 \leq (\mu_1 - \mu_2) \leq 5.545.$$

Since this interval contains 0, there does not appear to be a significant difference between the abilities of males and females to estimate when 30 seconds has passed.

4.8.13 The parameter is

$$\mu_d = \text{The mean difference (first error - second error) of all students.}$$

The values of the differences in the sample are shown in the table below. We have $n = 18$, $\bar{x} = 0.3611$, and $s = 4.92$. At the 95% confidence level, the critical value is $t_{0.05/2}(18-1) = 2.110$ so that the margin of error is $E = 2.110\,\frac{4.92}{\sqrt{18}} = 2.447$ and the confidence interval is

$$0.3611 - 2.447 \leq \mu_d \leq 0.3611 + 2.447 \;\Rightarrow\; -2.086 \leq \mu_d \leq 2.808.$$

Since the confidence interval contains 0, the second error does not appear to be smaller than the first, on average.

First Error	2	2	9	4	6	3	15	2	1.5	2	0	5	1	2	2	5	10	4
Second Error	7	1	2	2	2	9	5	3	10	7	5	2	3	2	0	1	5	3
Difference	-5	1	7	2	4	-6	10	-1	-8.5	-5	-5	3	-2	0	2	4	5	1

4.8.14 The parameter is

$$\mu_d = \text{The mean difference (before - after) of all athletes who took this training program.}$$

The values of the differences in the sample are shown in the table below. We have $n = 8$, $\bar{x} = 1.613$, and $s = 1.131$. At the 95% confidence level, the critical value is $t_{0.05/2}(8-1) = 2.365$ so that the margin of error is $E = 2.365\,\frac{1.131}{\sqrt{8}} = 0.946$ and the confidence interval is

$$1.613 - 0.946 \leq \mu_d \leq 1.613 + 0.946 \;\Rightarrow\; 0.667 \leq \mu_d \leq 2.559.$$

Since the lower limit is positive, it appears that the training program does have a significant effect on track times.

Before	53.8	54.2	52.9	53	55.1	58.2	56.7	52
After	52.7	53.8	51.1	52.7	54.2	55.4	53.3	49.8
Difference	1.1	0.4	1.8	0.3	0.9	2.8	3.4	2.2

4.8.15 The parameter is

$$\mu_d = \text{The mean difference (with - withou) of all athletes.}$$

The values of the differences in the sample are shown in the table below. We have $n = 10$, $\bar{x} = 0.6$, and $s = 1.075$. At the 95% confidence level, the critical value is $t_{0.05/2}(10-1) = 2.262$ so that the margin of error is $E = 2.262\,\frac{1.075}{\sqrt{10}} = 0.769$ and the confidence interval is

$$0.6 - 0.769 \leq \mu_d \leq 0.6 + 0.769 \;\Rightarrow\; -0.169 \leq \mu_d \leq 1.369.$$

Since this interval contains 0, it does not appear that shoes significantly affect vertical jump height.

With	15.5	18.5	22.5	17.5	16.5	17.5	20.5	15.5	19.5	18.5
Without	15.5	17.5	22.5	15.5	15.5	18.5	21.5	14.5	17.5	17.5
Difference	0	1	0	2	1	-1	-1	1	2	1

4.8.16

a. We havd $n_1 = n_2 = 10$ so that by this alternative formula, $r = 9$ so that the critical value is $t_{0.05/2}(9) = 2.262$ and the margin of error is $E = 2.262\sqrt{6.5^2/10 + 4.5^2/10} = 5.655$. This yields the confidence interval $-4.355 \leq \mu_1 - \mu_2 \leq 6.955$.

b. In part a., we have a much smaller degrees of freedom and a wider confidence interval than in the example.

c. Without loss of generality, we assume that $n_1 \geq n_2$. To simplify notation, let $a = s_1^2/n_1^2$ and $b = s_2^2/n_2^2$ so that

$$\begin{aligned} r &= \frac{(a+b)^2}{\dfrac{a^2}{n_1 - 1} + \dfrac{b^2}{n_2 - 1}} \\ &= \frac{(n_2 - 1)\,(a+b)^2}{\left(\dfrac{n_2 - 1}{n_1 - 1}\right) a^2 + b^2} \quad \text{by multiplying by } \frac{n_2 - 1}{n_2 - 1} \\ &\geq \frac{(n_2 - 1)\,(a+b)^2}{a^2 + b^2} \quad \text{since } \frac{n_2 - 1}{n_1 - 1} \leq 1 \\ &\geq \frac{(n_2 - 1)\left(a^2 + b^2\right)}{a^2 + b^2} \quad \text{since } (a+b)^2 > a^2 + b^2 \\ &= n_2 - 1 \end{aligned}$$

Thus $r \geq n_2 - 1$. Since $n_2 - 1 \leq n_1 - 1$, r is greater than the smaller of $n_1 - 1$ and $n_2 - 1$.

d. Using a smaller value of r yields a larger critical value and a wider confidence interval. This means that more than $100(1 - \alpha)\%$ of confidence intervals would contain the true value of $(\mu_1 - \mu_2)$.

4.8.17

a. If r increases, the critical value gets smaller.
b. The confidence interval would be narrower.
c. Rounding down gives a wider confidence interval which mean we are more confident that the interval contains the true value of the difference. It is better to be more confident than less confident.

4.8.18

a. The parameters are defined in the exercise. We have $\hat{p}_1 = 19/175 = 0.109$ and $\hat{p}_2 = 6/154 = 0.0390$ so that $\hat{p}_1 - \hat{p}_2 = 0.07$. We have

$$\hat{p} = \sqrt{\frac{0.109\,(1 - 0.109)}{175} + \frac{0.0390\,(1 - 0.0390)}{154}} \approx 0.0283.$$

At the 95% confidence level, $z_{0.05} = 1.645$ so that the margin of error is $E = 0.0283(1.645) = 0.0465$ and the confidence interval is

$$0.07 - 0.0465 \leq (p_1 - p_2) \leq 1 \quad \Rightarrow \quad 0.0235 \leq (p_1 - p_2) \leq 1.$$

b. Since this lower limit is greater than 0.02, it appears that the percentage of defects has been reduced by at least two percentage points.

4.8.19 The parameters are defined in the exercise. We have $n_1 = 27$ $\bar{x}_1 = 600$, $s_1 = 80$, $n_2 = 25$, $\bar{x}_2 = 533$, $s_2 = 118$, $\bar{x}_1 - \bar{x}_2 = 67$. The pooled standard deviation is

$$s_p = \sqrt{\frac{(27-1)80^2 + (25-1)118^2}{27+25-2}} = 100.06.$$

At the 95% confidence level, the critical value is $t_{0.05}(27+25-2) = 1.676$ so that the margin of error is $E = 100.06(1.676)\sqrt{1/27 + 1/25} = 46.55$ and the confidence interval is

$$67 - 46.55 \leq (\mu_1 - \mu_2) < \infty \quad \Rightarrow \quad 20.45 \leq (\mu_1 - \mu_2) < \infty.$$

Since this lower limit is greater than 20, it appears that the score increase by at least 20 points, so the data do support the claim.

4.9 Sample Size

4.9.1 Starting with $E = z_{\alpha/2}\sqrt{\hat{p}\left(1-\hat{p}\right)/n}$, we square both sides, multiply by n, and divide by E to get $n = z_{\alpha/2}^2 \hat{p}\left(1-\hat{p}\right)/E^2$.

4.9.2 We have $\hat{p} = 0.75$, $E = 0.015$, and $z_{0.10/2} = 1.645$ so that

$$n = \frac{1.645^2(0.75)\left(1-0.75\right)}{0.015^2} = 2255.02.$$

Thus they must survey at least 2256 students.

4.9.3 We have $\hat{p} = 0.5$, $E = 0.04$, and $z_{0.05/2} = 1.96$ so that

$$n = \frac{1.96^2(0.5)\left(1-0.5\right)}{0.04^2} = 600.25.$$

Thus they must survey at least 601 voters.

4.9.4 Consider the function $f(x) = x(1-x) = x - x^2$ over the interval $0 \leq x \leq 1$. Then $f'(x) = 1 - 2x$. Setting f' equal to 0 yields $x = 1/2$. Then $f(0) = f(1) = 0$ and $f(1/2) = 1/4$. Thus the maximum value of f is 1/4 so that $x(1-x) \leq 1/4$ for all x between 0 and 1. Changing x to $\hat{p}$, we get $\hat{p}\left(1-\hat{p}\right) \leq 1/4$ for all values of $\hat{p}$ between 0 and 1.

4.9.5 We have $E = 0.25$ and $z0.1/2 = 1.645$. Taking 1.3 as an estimate of σ we get a necessary sample size of

$$n = \frac{1.645^2(1.3^2)}{0.25^2} = 73.2$$

so that we need a total of 74. Since we have already sampled 15, we need an additional 59 fish.

4.9.6

a. We have $E = 0.02$ and $z0.05/2 = 1.96$. Taking 0.2 as an estimate of σ we get a necessary sample size of
$$n = \frac{1.96^2(0.2^2)}{0.02^2} = 384.2$$
so that we need to sample at least 385 batteries.

b. The parameter is
$$\mu = \text{The mean life-span of the new batteries.}$$
We have $n = 400$. At the 95% confidence level, the critical value is $t_{0.05/2}(400-1) \approx 1.966$ so that the margin of error is $E = 1.966\,\frac{0.18}{\sqrt{400}} \approx 0.01769$ and the confidence interval is
$$9.05 - 0.01769 \le \mu \le 9.05 + 0.01769 \;\Rightarrow\; 9.032 \le \mu \le 9.068.$$
Since this lower limit is greater than 8.9, it appears that the mean life span has increased.

4.9.7 A student would like to estimate the mean GPA of all education majors at a large university. She believes that most of the GPA's will lie between 2.5 and 3.8.

a. According to equation (4.10), $\sigma \approx (3.8 - 2.5)/4 = 0.325$.

b. We have $E = 0.1$ and $z0.01/2 = 2.576$ so that
$$n = \frac{2.576^2(0.325^2)}{0.1^2} = 70.04.$$
Thus she needs to survey at least 71 students.

4.9.8 We have $E = 0.1\sigma$ and $z0.1/2 = 1.645$ so that
$$n = \frac{1.645^2\sigma^2}{(0.1\sigma)^2} = 270.6. \tag{4.1}$$
Thus the sample size needs to be at least 271.

4.9.9 The population of professors is probably more homogeneous than the general population. This means that the variance of the population of professors is probably less than 15^2. This means that the sample size calculated in Example 4.9.3 is probably larger than needed.

4.9.10

a. Using $E = 0.02$ and $\hat{p} = 0.5$, we get necessary sample sizes of 80%: 1027, 90%: 1692, 95%: 2401, and 99%: 4148.

b. As the confidence level increases, the necessary sample size also increases.

c. Using $z_{0.05/2} = 1.96$ and $\hat{p} = 0.5$, we get necessary sample sizes of $E = 0.015$: 4269, $E = 0.01$: 9604, $E = 0.005$: 38416

d. As the margin of error decreases, the necessary sample size increases.

4.9.11

a. We have $\hat{p} = 0.5$, $E = 0.01$, and $z_{0.01/2} = 2.576$ so that

$$n = \frac{2.576^2(0.5)(1-0.5)}{0.01^2} = 16590.$$

b. They can lower their confidence level or increase the desired margin of error.

4.9.12

a. With $E = 0.1$ at the 95% confidence level, the sample size is $n = 1.96^2\hat{p}(1-\hat{p})/0.1^2 = 384.16\hat{p}(1-\hat{p})$. Now consider the function $f(x) = 38.016x(1-x) = 384.16\left(x - x^2\right)$ over the interval $0 \le x \le 1$. Then $f'(x) = 384.16(1-2x)$. Setting f' equal to 0 yields $x = 1/2$. Then $f(0) = f(1) = 0$ and $f(1/2) = 96.04$. Thus f is maximized at $x = 1/2$. This means that the largest sample size occurs when $\hat{p} = 0.5$.
b. Using the same arguments as in part a., we get that the largest sample size occurs when $\hat{p} = 0.5$.
c. Since a larger random sample will always give a smaller margin of error, when in doubt, we would prefer to use a larger sample size than a smaller one. Using $\hat{p} = 0.5$ gives us the sample size at least as large as we need to achieve the desired margin of error.

4.9.13 We need to find the smallest value of n such that $(n-1)\left(\frac{1}{a} - \frac{1}{b}\right) < 1.52$ where $a = \chi^2_{0.975}(n-1)$ and $b = \chi^2_{0.025}(n-1)$. The calculations in the following table show that the necessary sample size is $n = 21$.

n	$a = \chi^2_{0.975}(n-1)$	$b = \chi^2_{0.025}(n-1)$	$(n-1)\left(\frac{1}{a} - \frac{1}{b}\right)$
18	7.5642	30.1910	1.6844
19	8.2307	31.5264	1.6160
20	8.9065	32.8523	1.5549
21	9.5908	34.1696	1.5000

4.9.14

a. We have $N = 1000$, $n = 100$, $\hat{p} = 65/100 = 0.65$ and $z_{0.1/2} = 1.645$ so that the margin of error is

$$E = 1.645\sqrt{\frac{0.65(1-0.65)}{100}\left(\frac{1000-100}{1000-1}\right)} = 0.0745 \qquad (4.2)$$

and the confidence interval is

$$0.65 - 0.0745 \le p \le 0.65 + 0.0745 \;\Rightarrow\; 0.5755 \le p \le 0.7245.$$

b. Using $N = 800$, $E = 0.05$, and $\hat{p} = 0.5$, we get

$$k = \frac{1.645^2(0.5)(1-0.5)}{0.05^2} = 270.6 \;\Rightarrow\; n = \frac{270.6}{1 + \frac{270.6-1}{800}} = 203.$$

c. Note that k does not depend on N, so $k = 270.6$ for every value of N. The values of n are summarized in the table below. We see that as N gets larger, n gets closer to k

N	2,000	5,000	10,000	100,000	1,000,000
n	239	257	264	270	271

d. The sample size given by formula (4.7) is the same as k. The sample size for a small population given by formula (4.12) is less than k because the denominator is greater than 1. Thus using formula (4.7) for a small population gives us an over-estimate of the necessary sample size.

e. This "finite population correction" is a rational expression in N where the degree of the numerator equals the degree of the denominator. Thus the limit as $N \to \infty$ equals the ratio of the leading coefficients, $1/1 = 1$. This means that as the population gets larger, the correction term is not needed.

4.9.15

a. We begin by dividing both sides by $z_{\alpha/2}$, squaring both sides, and adding the fractions on the right-hand side to get

$$\left(\frac{E}{z_{\alpha/2}}\right)^2 = \frac{\hat{p}_1(1-\hat{p}_1) + \hat{p}_2(1-\hat{p}_2)}{n}.$$

Taking the reciprocal of both sides and multiplying by $[\hat{p}_1(1-\hat{p}_1) + \hat{p}_2(1-\hat{p}_2)]$ yields

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 [\hat{p}_1(1-\hat{p}_1) + \hat{p}_2(1-\hat{p}_2)].$$

b. Using $E = 0.04$, $\hat{p}_1 = \hat{p}_2 = 0.5$, and $z_{0.05/2} = 1.96$, we get

$$n = \left(\frac{1.96}{0.04}\right)^2 [0.5(1-0.5) + 0.5(1-0.5)] = 1200.5$$

so that the necessary sample size is 1201.

4.9.16

a. We begin by dividing both sides by $z_{\alpha/2}$, squaring both sides, and adding the fractions on the right-hand side to get

$$\left(\frac{E}{z_{\alpha/2}}\right)^2 = \frac{\sigma_1^2 + \sigma_2^2}{n}$$

Taking the reciprocal of both sides and multiplying by $\left(\sigma_1^2 + \sigma_2^2\right)$ yields

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 \left(\sigma_1^2 + \sigma_2^2\right).$$

b. Using $E = 0.75$, $\sigma_1^2 = 1.5$, $\sigma_2^2 = 5.6$, and $z_{0.01/2} = 2.576$, we get

$$n = \left(\frac{2.576}{0.75}\right)^2 (1.5 + 5.6) = 83.8$$

so that the necessary sample size is 84.

4.10 Assessing Normality

4.10.1 Replacing y_k with y and x_k with x, we see that these points satisfy the relation $y = (x - \bar{x})/s = (1/s)x - \bar{x}/s$. But s and $\bar{x}/s$ are constant for a given set of data, so the relation is of the form $y = mx + b$ where m and b are constants. Graphically, this relationship forms a straight line.

4.10.2 The calculations and quantile plot are shown in Figure 4.13. The dots do not lie very close to the straight line, so it is not reasonable to assume the population is normally distributed.

k	x	Percentile	z-score
1	52.3	0.0833	-1.3830
2	56.8	0.1667	-0.9674
3	57.8	0.2500	-0.6745
4	60.7	0.3333	-0.4307
5	70.9	0.4167	-0.2104
6	89.9	0.5000	0.0000
7	91.7	0.5833	0.2104
8	92.6	0.6667	0.4307
9	92.7	0.7500	0.6745
10	92.8	0.8333	0.9674
11	93.9	0.9167	1.3830
$\bar{x} =$	77.46		
$s =$	17.59		

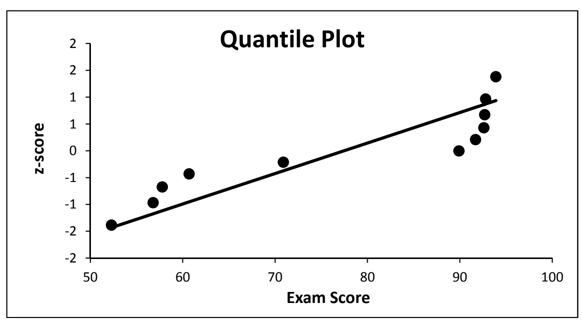


Figure 4.13

4.10.3 The calculations and quantile plot are shown in Figure 4.14. The dots exhibit a pattern, so it is not reasonable to assume the population is normally distributed.

k	x	Percentile	z-score
1	0	0.0526	-1.6199
2	1	0.1053	-1.2521
3	1.5	0.1579	-1.0031
4	2	0.2105	-0.8046
5	2	0.2632	-0.6336
6	2	0.3158	-0.4795
7	2	0.3684	-0.3360
8	2	0.4211	-0.1992
9	2	0.4737	-0.0660
10	3	0.5263	0.0660
11	4	0.5789	0.1992
12	4	0.6316	0.3360
13	5	0.6842	0.4795
14	5	0.7368	0.6336
15	6	0.7895	0.8046
16	9	0.8421	1.0031
17	10	0.8947	1.2521
18	15	0.9474	1.6199

$\bar{x} =$ 4.194

$s =$ 3.785

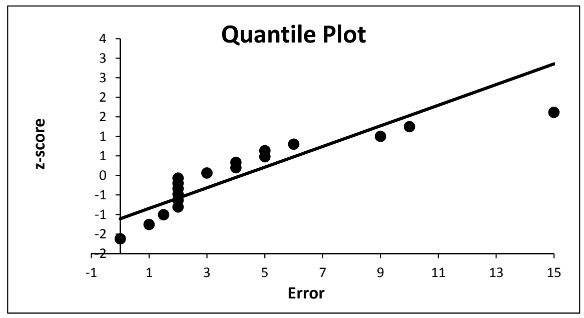


Figure 4.14

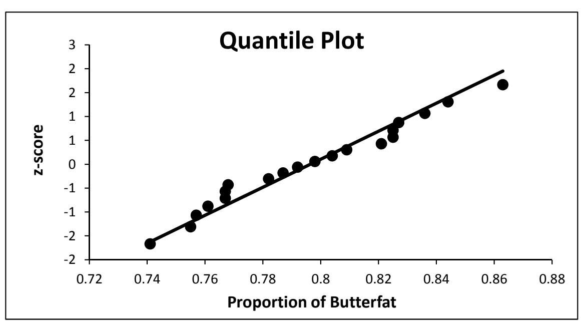


Figure 4.15

4.10.4 The quantile plot is shown in Figure 4.15. The dots lie close to the straight line and do not exhibit any pattern, so it is reasonable to assume the population is normally distributed.

4.10.5 The quantile plots are shown in Figure 4.16. The normality of the name brand and generic B times are somewhat questionable. However, the dots do not lie that far from the line to reject normality. So it appears to be reasonable to assume that all three populations are normal.

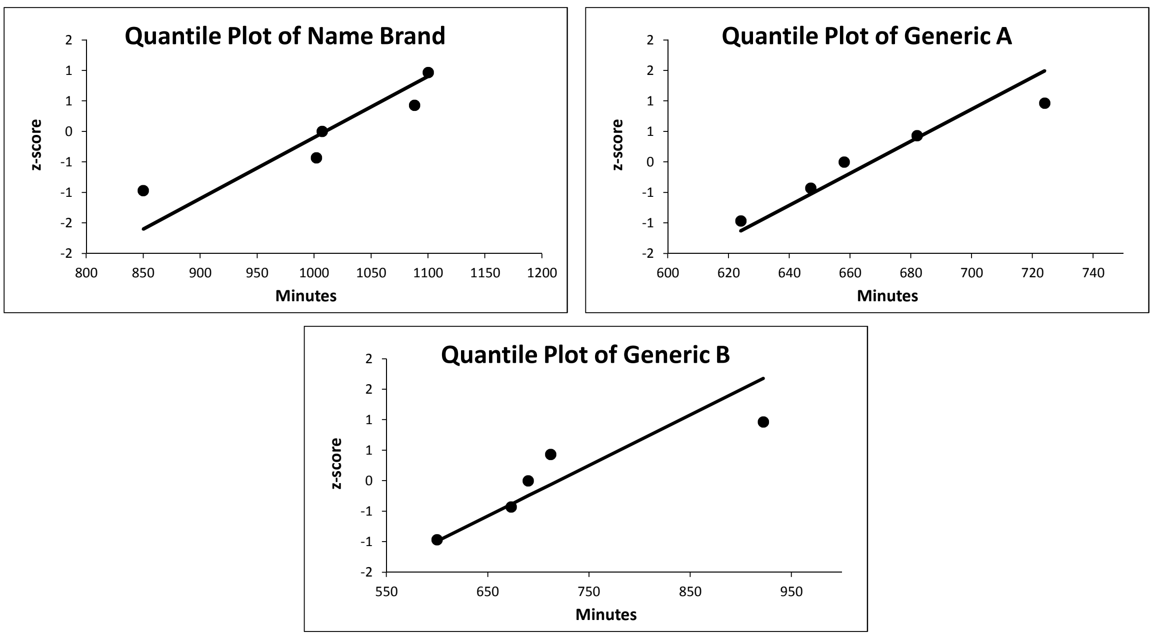


Figure 4.16

4.10.6 The quantile plots are shown in Figure 4.17. It appears reasonable to assume that brand B is normal. For brand A, the numerous data values of 23 and 24 cause us to question the assumption of normality, but the assumption does not appear to be too unreasonable.

4.10.7 The quantile plots are shown in Figure 4.18. The assumptions of normality appear reasonable for both populations.

4.10.8 The quantile plot is shown in Figure 4.19. The assumption of normality appears to be reasonable.

4.10.9 The calculations and probability plot are shown in Figure 4.20. We see that the dots do lie near the straight line. Thus the population appears to have an exponential distribution.

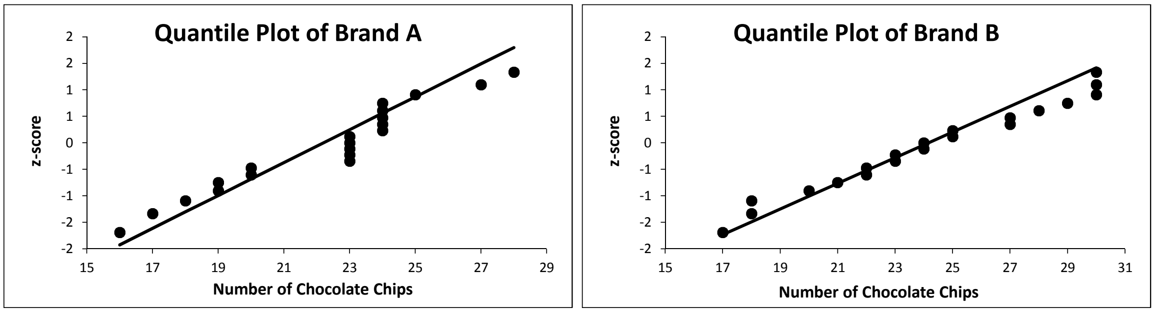


Figure 4.17

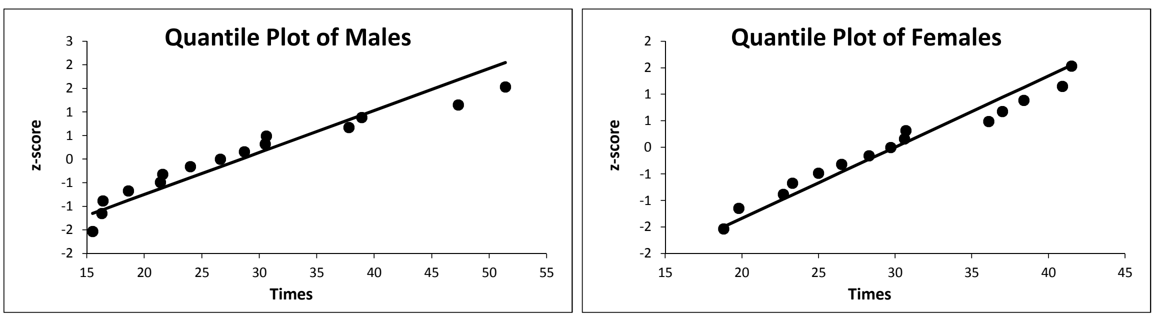


Figure 4.18

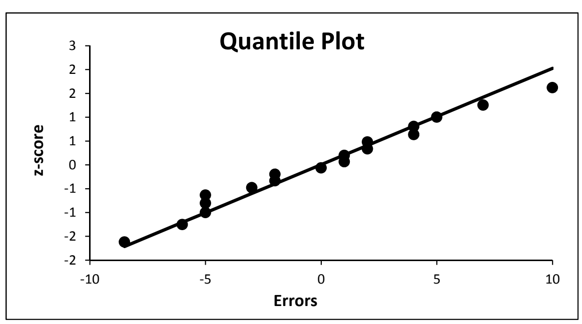


Figure 4.19

k	x	p_k	π_k
1	0.1	0.0625	0.1291
2	0.2	0.1250	0.2671
3	0.3	0.1875	0.4153
4	0.5	0.2500	0.5754
5	0.7	0.3125	0.7494
6	1.2	0.3750	0.9400
7	1.7	0.4375	1.1507
8	1.7	0.5000	1.3863
9	2.1	0.5625	1.6534
10	2.2	0.6250	1.9617
11	2.5	0.6875	2.3263
12	2.9	0.7500	2.7726
13	3.4	0.8125	3.3480
14	4.5	0.8750	4.1589
15	5.5	0.9375	5.5452

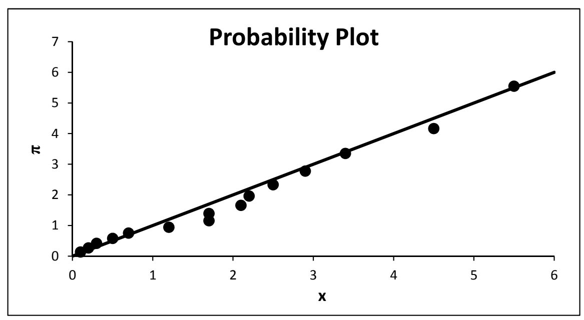


Figure 4.20

Chapter 5

Hypothesis Testing

5.1 Introduction

5.1.1 The claim is made about the population of all kindergarten students. In order to prove her claim, she would need to weigh *every* kindergarten student. She does not do this. She weighs only a sample of students. The sample mean of $\bar{x} = 48.6$ is only an estimate of the population mean she is making a claim about.

5.1.2 This is the correct conclusion because she is making a claim about only those candies in the particular bag. She weighs *every* candy so the mean of 0.048 is the exact value of what she is making a claim about.

5.1.3 A sample proportion less than 0.5 cannot support a claim that a population proportion is greater than 0.5. So no hypothesis test is needed.

5.1.4

a. $p =$ proportion of all citizens of the United States that own a home
b. $p =$ proportion of all introduction to statistics students that spend more than one hour studying for their tests.
c. $\mu =$ mean IQ score of all professors
d. $\sigma =$ standard deviation of all scores on the SAT
e. $\mu =$ mean weight of all 12 oz packages of shredded cheddar cheese

5.1.5

a. $z_{0.10} = 1.282$, $(-\infty, -1.282]$, P-value $= P(Z \leq -1.24) = 0.1075$, do not reject H_0
b. $z_{0.05/2} = 1.960$, $(-\infty, -1.960] \cup [1.960, \infty)$, P-value $= 2 \cdot P(Z > 2.67) = 0.0076$, reject H_0

c. $z_{0.01} = 2.326$, $[2.326, \infty)$, P-value $= P(Z > 0.87) = 0.1922$, do not reject H_0

5.1.6

a. There is sufficient evidence to reject the claim.
b. The data do not support the claim.
c. The data support the claim.

5.1.7 The claim is that $p < 0.75$ so that the hypotheses are

$$\boldsymbol{H_0} : p = 0.75, \qquad \boldsymbol{H_1} : p < 0.75.$$

P-value $= P(Z \le -2.77) = 0.0028 < 0.10$ so we reject H_0 and conclude that the data support the claim. Supporting the claim that $p < 0.75$ does not contradict the original conclusion of rejecting the claim that $p = 0.75$. There would be a contradiction if we supported both claims.

5.1.8

a. Since the numerator is $\hat{p} - p_0$, if $z < 0$, then p_0 is larger.
b. In this case they are equal.
c. The P-value in a left-tail test is $P(Z \le z)$. If this probability is greater than 0.5, then z must be positive. By the reasoning in part a., this means that $\hat{p}$ is larger.
d. First assume the test statistic z is negative. Then P-value $= 2 \cdot P(Z \le z) > 0.5$. This means that $P(Z \le z) > 0.25$. By examining the z-score table, we see this is true for approximately $z > -0.675$. By similar arguments, if z is positive, then $z < 0.675$. Therefore, $-0.675 < z < 0.675$.
e. No. A P-value is a probability. Probabilities must be between 0 and 1.
f. The P-value was probably given in scientific notation and the student missed the power of 10. Rounded to 4 decimal places, the P-value is probably 0.0000.

5.1.9

a. The respective values of z are 0.283, -0.566, 2.828, -3.677, and 0 .
b. If $\hat{p} > p_0$, then $z > 0$. If $\hat{p} < p_0$, then $z < 0$. If $\hat{p} = p_0$, then $z = 0$. If $\hat{p}$ is further in value from p_0, then z gets larger in absolute value.

5.2 Testing Claims about a proportion

5.2.1 Let $p =$ the proportion of all users of the allergy medicine that experience drowsiness. The claim is that $p < 0.10$ and the hypothesess are

$$\boldsymbol{H_0} : p = 0.10, \qquad \boldsymbol{H_1} : p < 0.10.$$

We have $n = 75$, $\hat{p} = 3/75 = 0.04$, and $z = \frac{0.4-0.5}{\sqrt{0.5(1-0.5)/75}} = -1.73$. At the 0.05 significance level, the critical value is $z_{0.05} = 1.645$ and the critical region is $(-\infty, -1.645]$. Also, P-value $= P(Z \leq -1.73) = 0.0418$. At the 0.05 significance level, we reject H_0 and conclude that the data support the claim. At the 0.01 significance level, the critical region is $(-\infty, -2.326]$ so that we would not reject H_0. The different significance level yields a different conclusion.

5.2.2

a. Let $p =$ the proportion of all consumers who prefer the Better Bread company's classic white bread over their main competitor's. The claim is that $p > 0.75$ and the hypothesess are

$$\boldsymbol{H_0} : p = 0.75, \qquad \boldsymbol{H_1} : p > 0.75.$$

We have $n = 150$, $\hat{p} = 115/150 = 0.767$, and $z = \frac{0.767-0.75}{\sqrt{0.75(1-0.75)/150}} = 0.48$. At the 0.05 significance level, the critical value is $z_{0.05} = 1.645$ and the critical region is $[1.645, \infty)$. Also, P-value $= P(Z > 0.48) = 0.3156$. At the 0.05 significance level, we do not reject H_0 and conclude that the data do not support the claim.

b. If $p_0 = 0.70$, we have $z = 1.78$ and P-value $= 0.0375$. So we reject H_0 and conclude that the data support the claim.

c. Our hypothesis tests support the second claim, not the first. So it would be more accurate to claim "More than 70% prefer our bread."

5.2.3 Let $p =$ the proportion of all voters in the 2006 Nebraska governor's election that were age 65 and oler. The claim is that $p < 0.25$ and the hypothesess are

$$\boldsymbol{H_0} : p = 0.25, \qquad \boldsymbol{H_1} : p < 0.25.$$

We have $n = 1008$, $\hat{p} = 0.24$, and $z = \frac{0.24-0.25}{\sqrt{0.25(1-0.25)/1008}} = -0.73$. At the 0.05 significance level, the critical value is $z_{0.05} = 1.645$ and the critical region is $[-\infty, -1.645]$. Also, P-value $= P(Z < -0.73) = 0.2327$. We do not reject H_0 and conclude that the data do not support the claim. Thus it would not be appropriate to include this statement in the article.

5.2.4 Let $p =$ the proportion of all cereals that are high-sugar. The claim is that $p > 0.5$ and the hypothesess are

$$\boldsymbol{H_0} : p = 0.5, \qquad \boldsymbol{H_1} : p > 0.5.$$

We have $n = 120$, $\hat{p} = 70/120 = 0.583$, and $z = \frac{0.583-0.5}{\sqrt{0.5(1-0.5)/120}} = 1.82$. At the 0.05 significance level, the critical value is $z_{0.05} = 1.645$ and the critical region is $[1.645, \infty)$. Also, P-value $= P(Z > 1.82) = 0.0344$. We reject H_0 and conclude that the data support the claim.

5.2.5 Let $p =$ the proportion of all ghost crabs with a larger left claw. The claim is that $p > 0.5$ and the hypothesess are

$$\boldsymbol{H_0} : p = 0.5, \qquad \boldsymbol{H_1} : p > 0.5.$$

We have $n = 72$, $\hat{p} = 41/72 = 0.569$, and $z = \frac{0.569-0.5}{\sqrt{0.5(1-0.5)/72}} = 1.18$. At the 0.05 significance level, the critical value is $z_{0.05} = 1.645$ and the critical region is $[1.645, \infty)$. Also, P-value $= P(Z > 1.18) = 0.119$. We do not reject H_0 and conclude that the data do not support the claim. Based on this conclusion, it does not appear that ghost crabs have a genetic disposition to having a larger left claw.

5.2.6 Let $p =$ the proportion of all pool balls that are pocketed behind the rack. The claim is that $p > 0.5$ and the hypothesess are

$$\boldsymbol{H_0} : p = 0.5, \qquad \boldsymbol{H_1} : p > 0.5.$$

We have $n = 130$, $\hat{p} = 75/130 = 0.577$, and $z = \frac{0.577-0.5}{\sqrt{0.5(1-0.5)/130}} = 1.75$. At the 0.05 significance level, the critical value is $z_{0.05} = 1.645$ and the critical region is $[1.645, \infty)$. Also, P-value $= P(Z > 1.75) = 0.0401$. We reject H_0 and conclude that the data support the claim.

5.2.7 Let $p =$ the proportion of all fired FBS coaches that had a winning percentage less than 0.500. The claim is that $p > 0.5$ and the hypothesess are

$$\boldsymbol{H_0} : p = 0.5, \qquad \boldsymbol{H_1} : p > 0.5.$$

We have $n = 39$, $\hat{p} = 30/39 = 0.770$, and $z = \frac{0.770-0.5}{\sqrt{0.5(1-0.5)/39}} = 3.36$. At the 0.05 significance level, the critical value is $z_{0.05} = 1.645$ and the critical region is $[1.645, \infty)$. Also, P-value $= P(Z > 3.36) = 0.0004$. We reject H_0 and conclude that the data support the claim. This result suggests that most coaches who get fired have a losing record. It does not establish a cause-and-effect relationship.

5.2.8 Let $p =$ the proportion of all spins of a penny that land tails up. The claim is that $p = 0.5$ and the hypothesess are

$$\boldsymbol{H_0} : p = 0.5, \qquad \boldsymbol{H_1} : p \neq 0.5.$$

We have $n = 500$, $\hat{p} = 296/500 = 0.592$, and $z = \frac{0.592-0.5}{\sqrt{0.5(1-0.5)/500}} = 4.11$. At the 0.05 significance level, the critical value is $z_{0.05/2} = 1.96$ and the critical region is $(-\infty, -1.96] \cup [1.96, \infty)$. Also, P-value $= 2 \cdot P(Z > 4.11) = 0.0000$. We reject H_0 and conclude that the data support the claim. This suggests that spinning a penny does not produce the same results as flipping a penny.

5.2.9 Let $p =$ the proportion of all ice-cream eating students who choose vanilla over chocolate or twist. The claim is that $p > 0.5$ and the hypothesess are

$$\boldsymbol{H_0} : p = 0.5, \qquad \boldsymbol{H_1} : p > 0.5.$$

We have $n = 56$, $\hat{p} = 35/56 = 0.625$, and $z = \frac{0.625-0.5}{\sqrt{0.5(1-0.5)/56}} = 1.87$. At the 0.05 significance level, the critical value is $z_{0.05} = 1.645$ and the critical region is $[1.645, \infty)$. Also, P-value $= P(Z > 1.87) = 0.0307$. We reject H_0 and conclude that the data support the claim.

5.2.10 Let $p =$ the proportion of all babies born that are girls. The claim is that $p = 0.5$ and the hypothesess are

$$\boldsymbol{H_0} : p = 0.5, \qquad \boldsymbol{H_1} : p \neq 0.5.$$

We have $n = 3465$, $\hat{p} = 1712/3465 = 0.494$, and $z = \frac{0.494-0.5}{\sqrt{0.5(1-0.5)/3465}} = -0.70$. At the 0.05 significance level, the critical value is $z_{0.05/2} = 1.96$ and the critical region is $(-\infty, -1.96] \cup [1.96, \infty)$. Also, P-value $= 2 \cdot P(Z \leq -0.70) = 0.4840$. We do not reject H_0 and conclude that there is not sufficient evidence to reject the claim. Thus our assumption that $P(\text{girl}) = P(\text{boy}) = 0.5$ appears to be reasonable.

5.2.11 Let $p =$ the proportion of red and blue candies that are red. The claim is that $p > 0.5$ and the hypothesess are

$$\boldsymbol{H_0} : p = 0.5, \qquad \boldsymbol{H_1} : p > 0.5.$$

We have $n = 120$, $\hat{p} = 65/120 = 0.542$, and $z = \frac{0.542-0.5}{\sqrt{0.5(1-0.5)/120}} = 0.91$. At the 0.05 significance level, the critical value is $z_{0.05} = 1.645$ and the critical region is $[1.645, \infty)$. Also, P-value $= P(Z > 0.91) = 0.1814$. We do not reject H_0 and conclude that the data do not support the claim.

5.2.12

a. Let $p =$ the proportion of all car crashes that occur within 5 miles of the home of the driver. The claim is that $p < 0.25$ and the hypothesess are

$$\boldsymbol{H_0} : p = 0.25, \qquad \boldsymbol{H_1} : p < 0.25.$$

We have $n = 120$, $\hat{p} = 62/120 = 0.517$, and $z = \frac{0.517-0.25}{\sqrt{0.25(1-0.25)/120}} = 6.75$. At the 0.05 significance level, the critical value is $z_{0.05} = 1.645$ and the critical region is $(-\infty, -1.645]$. Also, P-value $= P(Z < 6.75) = 1$. We do not reject H_0 and conclude that the data do not support the claim. The claim it not at all reasonable.

b. Since the sample proportion is greater than 0.25, there is no way the data could support the claim.

5.2.13

a. The test statistics are $z = 0.89, 1.26, 1.55, 1.78, 2$, respectively. The corresponding P-values are 0.1867, 0.1038, 0.0606, 0.0375, 0.0228. For the first 3 values of n, we do not reject H_0 while we do reject H_0 for the last 2 values.

b. $\lim_{n\to\infty} z = \infty$, $\lim_{n\to\infty}(\text{P-value}) = 0$

c. Larger values of n give smaller P-values, so we would be more confident in rejecting H_0 and supporting the claim.for larger values of n.

5.2.14

a. Using software, $P(X \leq 37) = 0.028732$. This exact P-value is slightly higher than the approximate value found in the example, but it does not change the conclusion.
b. Using software, if X is $b(0.3, 350)$, then $P(X \geq 120) = 0.04666$. To find the approximate P-value, $\hat{p} = 120/350 = 0.343$ so that $z = \frac{0.343-0.3}{\sqrt{0.3(1-0.3)/350}} = 1.75$ and P-value $= P(Z > 1.75) = 0.0401$. The exact P-value is slightly higher than the approximate, but not enough to change the conclusion.
c. Using software, if X is $b(100, 0.3)$, then $p_1 = P(X \leq 24) = 0.11357$. Values of $P(X \geq x)$ for different values of x are shown in the table below. We see that the closest probability less than 0.11357 is 0.07988. Thus $x_2 = 37$ and the exact P-value is $0.11357 + 0.07988 = 0.19345$.

x	35	36	37	38	39
$P(X \geq x)$	0.16286	0.11608	0.07988	0.05305	0.03398

5.2.15 The polling results example is a two-tail test, so we calculate a two-sided confidence interval. We have $n = 350$, $\hat{p} = 0.686$, and the critical value is $z_{0.10/2} = 1.645$ so that the margin of error is

$$E = 1.645\sqrt{\frac{0.686\,(1 - 0.686)}{350}} = 0.041$$

and the confidence interval is $0.686 - 0.041 \leq p \leq 0.686 + 0.041 \;\Rightarrow\; 0.645 \leq p \leq 0.727$. This confidence interval does not contain 0.75, so we reject H_0 which is the same conclusion as in the example.

The M&M example is a left-tail test, so we calculate a one-sided confidence interval that gives an upper bound on p. We have $n = 195$, $\hat{p} = 0.190$, and the critical value is $z_{0.05} = 1.645$ so that the margin of error is

$$E = 1.645\sqrt{\frac{0.190\,(1 - 0.190)}{195}} = 0.046$$

and the confidence interval is $0 \leq p \leq 0.190 + 0.046 \;\Rightarrow\; 0 \leq p \leq 0.236$. This confidence interval does not contain 0.25 so we reject H_0 which is the same conclusion as in the example.

5.2.16

a. We have $n = 100$, $p_0 = 0.45$, and the critical z-value is $z_{0.05} = 1.645$. The critical value of $\hat{p}$ is

$$\hat{p} = 1.645\sqrt{0.45\,(1 - 0.45)\,/100} + 0.45 = 0.5318.$$

We would reject H_0 if $\hat{p} > 0.5318$. Now, $\hat{P}$ is approximately $N(0.6,\, 0.6(1 - 0.6)/100) = N(0.6,\, 0.0024)$, so the power of the test is $P\left(\hat{P} > 0.5318\right) \approx P(Z > -1.39) = 0.9177$. This is a higher power than when $p_0 = 0.5$ and suggests that as p_0 gets further in value from p, the power increases.

b. With $p_0 = 0.5$ and $n = 500$, the critical value of $\hat{p}$ is

$$\hat{p} = 1.645\sqrt{0.5\,(1-0.5)\,/500} + 0.5 = 0.5368.$$

Now, $\hat{P}$ is approximately $N(0.6,\, 0.6(1-0.6)/500) = N(0.6,\, 0.00048)$, so the power of the test is $P\left(\hat{P} > 0.5368\right) \approx P(Z > -2.88) = 0.9980$. This is a higher power than when $p_0 = 0.5$ and $n = 100$ and suggests that as n gets larger, the power increases.

c. With $p_0 = 0.5$, $n = 100$, and $\alpha = 0.2$, the critical z-value is $z_{0.2} = 0.842$. The critical value of $\hat{p}$ is

$$\hat{p} = 0.842\sqrt{0.5\,(1-0.5)\,/100} + 0.5 = 0.5421.$$

Now, $\hat{P}$ has the same distribution as in part a., so the power of the test is $P\left(\hat{P} > 0.421\right) \approx P(Z > -1.18) = 0.8810$. This is a higher power than when $/alpha = 0.05$ as in the example and suggests that as α gets larger, the power increases.

5.2.17 We have $n = 200$, $p_0 = 0.30$, $p = 0.20$, and the critical z-value is $-z_{0.01} = -2.326$. The critical value of $\hat{p}$ is

$$\hat{p} = -2.326\sqrt{0.3\,(1-0.3)\,/200} + 0.3 = 0.2246.$$

We would reject H_0 if $\hat{p} < 0.2246$. Now, $\hat{P}$ is approximately $N(0.2,\, 0.2(1-0.2)/200) = N(0.2,\, 0.0008)$, so the power of the test is $P\left(\hat{P} < 0.2246\right) \approx P(Z < 0.87) = 0.8078$.

5.2.18 We have $n = 150$, $p_0 = 0.75$, $p = 0.70$, and the critical z-value is $-z_{0.05} = -1.645$. The critical value of $\hat{p}$ is

$$\hat{p} = -1.645\sqrt{0.75\,(1-0.75)\,/150} + 0.75 = 0.6918.$$

We would reject H_0 if $\hat{p} < 0.6918$. Now, $\hat{P}$ is approximately $N(0.7,\, 0.7(1-0.7)/150) = N(0.7,\, 0.0014)$, so the power of the test is $P\left(\hat{P} < 0.6918\right) \approx P(Z < -0.22) = 0.4129$. This is a higher power than in the example, so this test would be more likely to detect the change.

5.3 Testing Claims about a Mean

5.3.1

a. Between 0.025 and 0.05
b. Between 0.05 and 0.10
c. Greater than 0.10

5.3.2 The claim is about the mean weight of all 12 oz packages of shredded cheddar, not just the 36 in the sample. The sample mean is an estimate of the population mean. These two means are not the same thing.

5.3.3 Let μ = the mean reading at the freezing point of water of all thermometers from the assembly line. The claim is $\mu = 0$ and the hypotheses are

$$\boldsymbol{H_0} : \mu = 0, \qquad \boldsymbol{H_1} : \mu \neq 0.$$

We have $\sigma = 1$, $n = 75$, and $\bar{x} = 0.12$ so the test statistic is $z = \frac{0.12-0}{1/\sqrt{75}} = 1.04$. At the 0.05 significance level, the critical value is $z_{0.05/2} = 1.96$ and the critical region is $(-\infty, -1.96]\cup[1.96, \infty)$. Also, P-value $= 2 \cdot P(Z > 1.04) = 0.2984$. We do not reject H_0 and conclude that there is not sufficient evidence to reject the claim. Thus it appears that these thermometers do give a reading of 0 °C, on average.

5.3.4 Let μ = the mean square footage of all homes within a five-mile radius of this lake. The claim is $\mu > 1931$ and the hypotheses are

$$\boldsymbol{H_0} : \mu = 1931, \qquad \boldsymbol{H_1} : \mu > 1931.$$

We have $s = 856.8$, $n = 30$, and $\bar{x} = 2421.7$ so the test statistic is $t = \frac{2421.7-1931}{856.8/\sqrt{30}} = 3.137$. At the 0.05 significance level, the critical value is $t_{0.05}(29) = 1.699$ and the critical region is $[1.699, \infty)$. Using software, P-value $= P(T > 3.137) = 0.0019$. We reject H_0 and conclude that the data support the claim. Thus it appears that homes around this lake tend to be larger than a typical home.

5.3.5 Let μ = the mean weight loss of all people on this program. The claim is $\mu < 0$ and the hypotheses are

$$\boldsymbol{H_0} : \mu = 0, \qquad \boldsymbol{H_1} : \mu < 0.$$

We have $s = 10.3$, $n = 50$, and $\bar{x} = -7.7$ so the test statistic is $z = \frac{-7.7-0}{10.3/\sqrt{50}} = -5.29$. At the 0.05 significance level, the critical value is $z_{0.05} = 1.645$ and the critical region is $(-\infty, -1.645]$. Also, P-value $= P(Z < -5.29) = 0.0000$. We reject H_0 and conclude that the data support the claim. Thus it appears that the program is effective.

5.3.6 Let μ = the mean height of all division I offensive linemen. The claim is $\mu > 76$ and the hypotheses are

$$\boldsymbol{H_0} : \mu = 76, \qquad \boldsymbol{H_1} : \mu > 76.$$

We have $s = 2.115$, $n = 40$, and $\bar{x} = 76.875$ so the test statistic is $t = \frac{76.875-76}{2.115/\sqrt{40}} = 2.62$. At the 0.05 significance level, the critical value is $t_{0.05}(39) \approx 1.684$ and the critical region is $[1.684, \infty)$. Using software, P-value $= P(T > 2.62) = 0.0062$. We reject H_0 and conclude that the data support the claim.

5.3.7 Let μ = the mean mass of water in all bottles of this brand. The claim is $\mu > 500$ and the hypotheses are

$$H_0 : \mu = 500, \qquad H_1 : \mu > 500.$$

From the data, we have $s = 0.816$, $n = 4$, and $\bar{x} = 505$ so the test statistic is $t = \frac{505-500}{0.816/\sqrt{4}} = 12.25$. At the 0.05 significance level, the critical value is $t_{0.05}(3) = 2.353$ and the critical region is $[2.353, \infty)$. Using software, P-value $= P(T > 2.62) = 0.0006$. We reject H_0 and conclude that the data support the claim. Thus it appears that the bottles are not under-filled on average.

5.3.8

a. The quantile plot is shown in Figure 5.1. It appears that the population is normally distributed.

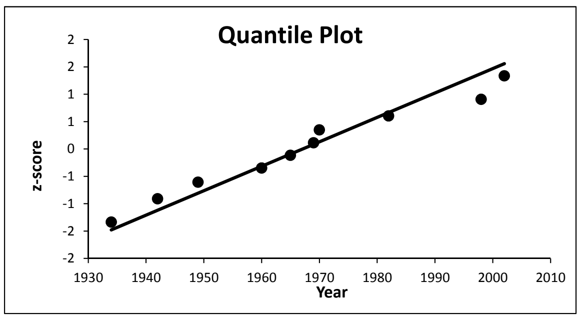


Figure 5.1

b. Since the population appears to be normally distributed and the books were randomly selected, the requirements are met. Let μ = the mean copyright data of all of the library's books on medicine. The claim is $\mu < 1975$ and the hypotheses are

$$H_0 : \mu = 1975, \qquad H_1 : \mu < 1975.$$

From the data, we have $s = 22.3977$, $n = 10$, and $\bar{x} = 1967.1$ so the test statistic is $t = \frac{1967.1-1975}{22.3977/\sqrt{10}} = -1.12$. At the 0.05 significance level, the critical value is $t_{0.05}(9) \approx 1.833$ and the critical region is $(-\infty, -1.833]$. Using software, P-value $= P(T < -1.12) = 0.1459$. We do not reject H_0 and conclude that the data do not support the claim.

5.3.9

a. The percentage savings are 41.8, 68.9, 56.2, 57.6, -4.7, 16.8, 18.3, 22.3, 56.7, and 36.2.
b. Let μ = the mean percentage savings. The claim is $\mu > 20$ and the hypotheses are

$$H_0 : \mu = 20, \qquad H_1 : \mu > 20.$$

From the calculated percentage savings, we have $s = 23.417$, $n = 10$, and $\bar{x} = 37.01$ so the test statistic is $t = \frac{37.01-20}{23.417/\sqrt{10}} = 2.30$. At the 0.05 significance level, the critical value is $t_{0.05}(9) = 1.833$ and the critical region is $[1.833, \infty)$. Using software, P-value $= P(T > 2.30) = 0.0235$. We reject H_0 and conclude that the data support the claim. Thus it appears that purchasing generic products results in substantial savings on average.

5.3.10 The proportions of pages in each magazine containing only advertisements are 0.379, 0.338, 0.337, 0.391, 0.386, 0.374, 0.299, 0.236, 0.341, 0.224, 0.224, 0.18, and 0.176. Let μ = the mean proportions of pages containing only advertisements. The claim is $\mu > 0.25$ and the hypotheses are

$$\boldsymbol{H_0} : \mu = 0.25, \qquad \boldsymbol{H_1} : \mu > 0.25.$$

From the calculated proportions, we have $s = 0.08026$, $n = 13$, and $\bar{x} = 0.2988$ so the test statistic is $t = \frac{0.2988-0.25}{0.08026/\sqrt{13}} = 2.19$. At the 0.05 significance level, the critical value is $t_{0.05}(12) = 1.782$ and the critical region is $[1.782, \infty)$. Using software, P-value $= P(T > 2.19) = 0.0245$. We reject H_0 and conclude that the data support the claim.

5.3.11 The population in the claim is the starters on that one team. He measured all members of this population, so the mean of 66.4 is the population mean he is making a claim about. This this population mean is greater than 64, the claim is proven and there is no need to do a hypothesis test.

5.3.12 As n gets larger, the critical t-values get closer to the critical z-values. Thus for "large" n, the critical regions in both the Z- and T-tests are almost identical. Since the z test statistic with σ replaced by s is identical to t test statistic, both tests will almost certainly yield the same result when n is "large."

5.3.13

a. The test statistic is calculated with the formula $z = \frac{12.05-12}{s/\sqrt{50}}$. The values of z, the P-values, and the final conclusions at the 95% and 99% confidence levels are shown in the table below.

s	0.5	0.2	0.1
z	0.71	1.77	3.54
P-value	0.2389	0.0384	0.0001
95%	not support	support	support
99%	not support	not support	support

b. $\lim_{s\to 0} z = \infty$ and $\lim_{s\to 0}$ (P-value) $= 0$.

c. We would be more confident that the sample data support the claim if s were small because for smaller values of s we get smaller P-values and we would be more confident in rejecting H_0 and supporting the claim.

5.3.14

a. The test statistics for the given values of μ_0 are 7.61, 5.71, 3.86, and 1.9. In each case, we would reject H_0 at the 95% confidence level and conclude that the data support the claim.

b. Because the sample size is so large, $\bar{x}$ is an estremely good estimate of μ. Therefore, we can say with great certainty that μ is very close to 15.05. We don't need any hypothesis test to confirm this.

5.3.15 The formula for the test statistic is $\frac{\bar{x}-40}{s/\sqrt{n}}$. Removing the large outlier would decrease the values of $\bar{x}$ making the value of the numerator in this formula larger in absolute value. This would also decrease s so that the denominator in this formula would be smaller. Assuming n is "large" so that n is not that much different from $n-1$, the test statistic would be larger in absolute value. This would decrease the area to the left of the test statistic, making the P-value smaller.

5.3.16

a. The test statistic has the same value as in the example, $z = 2$. At the 0.05 significance level, the critical value is $z_{0.05} = 1.645$ and the critical region is $[1.645, \infty)$. Also, P-value $= P(Z > 2.00) = 0.0228$. We reject H_0 and conclude that the data support the claim. This is the same conclusion as in the example.

b. For both the Z- and T-tests, the value of the test statistic is $\frac{105-100}{15/\sqrt{36}} = 2$. The results of both tests are summarized the table below. We see that for the T-test we do not reject H_0 while we get the opposite result from the Z-test. Thus these tests do not give the same result.

	Critical value	**Critical region**	**P-value**
T-test	$t_{0.05/2}(35) = 2.030$	$(-\infty, -2.030] \cup [2.030, \infty)$	$2 \cdot P(T > 2) = 0.0533$
Z-test	$z_{0.05/2} = 1.960$	$(-\infty, -1.960] \cup [1.960, \infty)$	$2 \cdot P(Z > 2) = 0.0456$

5.3.17 We have $\mu = 10$, $\sigma = 1$, $\mu_0 = 10.5$, and $n = 15$. The critical value is $-z_{0.05} = -1.645$ so that the critical value of $\bar{x}$ is

$$10.5 - 1.645\frac{1}{\sqrt{15}} = 10.075$$

We will reject H_0 if $\bar{X} < 10.075$. Now, $\bar{X}$ is approximately $N(10, 1/15)$ so that the power of the test is

$$P\left(\bar{X} < 10.075\right) = P(Z < 0.29) = 0.6141.$$

5.4 Comparing Two Proportions

5.4.1 The 2-proportion Z-test is used to compare proportions from two independent populations. In this scenario, we have only one population, the candy. We could rewrite this claim as "the proportion of the candy that is red is greater than 0.5." Thus we see that this is really a 1-proportion test.

5.4.2 The parameters are

$p_1 =$ The proportion of all athletes that take naps on a regular basis, and

$p_2 =$ The proportion of all non-athletes that take naps on a regular basis.

The claim is that $p_1 > p_2$ and the hypotheses are

$$\boldsymbol{H_0} : p_1 = p_2 \qquad \boldsymbol{H_1} : p_1 > p_2.$$

We have $\hat{p}_1 = 0.72$, $n_1 = 75$, $x_1 = 0.72(75) = 54$, $\hat{p}_2 = 0.70$, $n_2 = 80$, and $x_2 = 0.70(80) = 56$ so that $\hat{p} = \frac{54+56}{75+80} = 0.7097$ and the test statistic is

$$z = \frac{0.72 - 0.70}{\sqrt{0.7097\,(1 - 0.7097)\,(1/75 + 1/80)}} = 0.27.$$

At the 0.05 significance level, the critical value is $z_{0.05} = 1.645$ and the critical region is $[1.645, \infty)$. Also, P-value $= P(Z > 0.27) = 0.3936$. We do not reject H_0 and conclude that the data do not support the claim. Based on this result, it does not appear that athletes take naps more often than non-athletes.

5.4.3 The claim is that $p_1 = p_2$ and the hypotheses are

$$\boldsymbol{H_0} : p_1 = p_2 \qquad \boldsymbol{H_1} : p_1 \neq p_2.$$

We have $n_1 = 228 + 255 = 483$, $x_1 = 228$, $\hat{p}_1 = 228/483 = 0.472$, $n_2 = 503 + 495 = 998$, $x_2 = 503$, $\hat{p}_2 = 503/998 = 0.504$, so that $\hat{p} = \frac{228+503}{483+998} = 0.4936$ and the test statistic is

$$z = \frac{0.472 - 0.504}{\sqrt{0.4936\,(1 - 0.4936)\,(1/483 + 1/998)}} = -1.15.$$

At the 0.05 significance level, the critical value is $z_{0.05/2} = 1.96$ and the critical region is $(-\infty, -1.96] \cup [1.96, \infty)$. Also, P-value $= 2 \cdot P(Z < -1.15) = 0.2502$. We do not reject H_0 and conclude that there is not sufficient evidence to reject the claim. Based on this result, there does not appear to be much difference between morning and evening in terms of the proportion of male users.

5.4.4 The parameters are

$$p_1 = \text{The proportion of tails when spinning new quarters, and}$$
$$p_2 = \text{The proportion of tails when spinning old quarters.}$$

The claim is that $p_1 > p_2$ and the hypotheses are

$$\boldsymbol{H_0} : p_1 = p_2 \qquad \boldsymbol{H_1} : p_1 > p_2.$$

We have $n_1 = 296 + 304 = 600$, $x_1 = 304$, $\hat{p}_1 = 304/600 = 0.507$, $n_2 = 351 + 249 = 600$, $x_2 = 249$, $\hat{p}_2 = 249/600 = 0.415$, so that $\hat{p} = \frac{304+249}{600+600} = 0.4608$ and the test statistic is

$$z = \frac{0.507 - 0.415}{\sqrt{0.4608\,(1 - 0.4608)\,(1/600 + 1/600)}} = 3.20.$$

At the 0.05 significance level, the critical value is $z_{0.05} = 1.645$ and the critical region is $[1.645, \infty)$. Also, P-value $= P(Z > 3.20) = 0.0007$. We reject H_0 and conclude that the data support the claim. Based on this result, there does appear to be a significant difference between old and new quarters when spinning them.

5.4.5 The parameters are

$$p_1 = \text{The proportion of blondes who prefer blondes, and}$$
$$p_2 = \text{The proportion of brunettes who prefer brunettes.}$$

The claim is that $p_1 = p_2$ and the hypotheses are

$$\boldsymbol{H_0} : p_1 = p_2 \qquad \boldsymbol{H_1} : p_1 \neq p_2.$$

We have $n_1 = 21 + 7 + 2 + 7 = 37$, $x_1 = 21$, $\hat{p}_1 = 21/37 = 0.5676$, $n_2 = 12 + 32 + 1 + 16 = 61$, $x_2 = 32$, $\hat{p}_2 = 32/61 = 0.5246$, so that $\hat{p} = \frac{21+32}{37+61} = 0.5408$ and the test statistic is

$$z = \frac{0.5676 - 0.5246}{\sqrt{0.5408\,(1 - 0.5408)\,(1/37 + 1/61)}} = 0.41.$$

At the 0.05 significance level, the critical value is $z_{0.05/2} = 1.96$ and the critical region is $(-\infty, -1.96] \cup [1.96, \infty)$. Also, P-value $= 2 \cdot P(Z > 0.41) = 0.6818$. We do not reject H_0 and conclude that there is not sufficient evidence to reject the claim. Based on this result, it appears that students typically prefer members of the opposite sex with the same hair color as themselves.

5.4.6 In both parts a. and b., the parameters are

$$p_1 = \text{The proportion of patients given a placebo who had a heart attack, and}$$
$$p_2 = \text{The proportion of patients given the medicine who had a heart attack.}$$

The claim is that $p_1 = p_2$ and the hypotheses are

$$\boldsymbol{H_0} : p_1 = p_2 \qquad \boldsymbol{H_1} : p_1 \neq p_2.$$

a. We have $n_1 = 75$, $x_1 = 1$, $\hat{p}_1 = 1/75$, $n_2 = 70$, $x_2 = 5$, $\hat{p}_2 = 5/70$, so that $\hat{p} = \frac{1+5}{75+70} = 0.0414$ and the test statistic is

$$z = \frac{1/75 - 5/70}{\sqrt{0.0414\,(1 - 0.0414)\,(1/75 + 1/70)}} = -1.75.$$

At the 0.05 significance level, the critical value is $z_{0.05/2} = 1.96$ and the critical region is $(-\infty, -1.96] \cup [1.96, \infty)$. Also, P-value $= 2 \cdot P(Z \leq -1.75) = 0.0802$. We do not reject H_0 and conclude that there is not sufficient evidence to reject the claim. Thus there does there not appear to be a "statistically significant" difference between the two medicines. This in-and-of itself does not mean mean the new medicine is safe.

b. With $n_2 = 73$ and $x_2 = 8$, we have $\hat{p} = 0.0588$ and $z = -2.49$. The critical region is the same as in part a. and P-value $= 2 \cdot P(Z \leq -2.49) = 0.0128$. We reject H_0 and conclude that there is sufficient evidence to reject the claim. Now there does appear to be a "statistically significant" difference between the two medicines.

5.4.7

a. The parameters are

$$p_1 = \text{The proportion of small tumors successfully treated by treatment A, and}$$
$$p_2 = \text{The proportion of small tumors successfully treated by treatment B.}$$

The claim is that $p_1 > p_2$ and the hypotheses are

$$\boldsymbol{H_0} : p_1 = p_2 \qquad \boldsymbol{H_1} : p_1 > p_2.$$

We have $n_1 = 90$, $x_1 = 84$, $\hat{p}_1 = 84/90 = 0.9333$, $n_2 = 260$, $x_2 = 224$, $\hat{p}_2 = 224/260 = 0.8615$, so that $\hat{p} = \frac{84+224}{90+260} = 0.88$ and the test statistic is

$$z = \frac{0.9333 - 0.8615}{\sqrt{0.88\,(1 - 0.88)\,(1/90 + 1/260)}} = 1.81.$$

At the 0.2 significance level, the critical value is $z_{0.2} = 0.842$ and the critical region is $[0.842, \infty)$. Also, P-value $= P(Z > 1.81) = 0.0351$. We reject H_0 and conclude that the data support the claim.

b. The parameters are the same as in part a. with "small" replaced with "large." The hypotheses are the same. We have $n_1 = 250$, $x_1 = 180$, $\hat{p}_1 = 180/250 = 0.72$, $n_2 = 85$, $x_2 = 54$, $\hat{p}_2 = 54/85 = 0.635$, so that $\hat{p} = \frac{180+54}{250+85} = 0.6985$ and the test statistic is

$$z = \frac{0.72 - 0.635}{\sqrt{0.6985\,(1 - 0.6985)\,(1/250 + 1/85)}} = 1.47.$$

The critical region is the same as part a. and P-value $= P(Z > 1.47) = 0.0708$. We reject H_0 and conclude that the data support the claim. Since we supported both claims, it appears that treatment A is better.

c. The parameters are the same as in part a. with "small" replaced with "all." The claim and alternative hypothesis are $p_1 < p_2$. We have $n_1 = 340$, $x_1 = 264$, $\hat{p}_1 = 264/340 = 0.7765$, $n_2 = 345$, $x_2 = 278$, $\hat{p}_2 = 278/345 = 0.8058$, so that $\hat{p} = \frac{264+278}{340+345} = 0.7912$ and the test statistic is

$$z = \frac{0.7765 - 0.8058}{\sqrt{0.7912\,(1 - 0.7912)\,(1/340 + 1/345)}} = -0.94.$$

The critical region is $(-\infty, -0.842]$ and P-value $= P(Z \le -0.94) = 0.1736$. We reject H_0 and conclude that the data support the claim. This suggests that treatment B is better overall which seems to contradict the conclusion in part b.

5.4.8 The parameters are p_A, p_B, and p_C, the proportion of patients given medication A, B, or C, respectively, who experience drowsiness. In all parts, at the 0.05 significance level, the critical region is $(-\infty, -1.96] \cup [1.96, \infty)$ and and $n_1 = n_2 = 100$.

a. The claim is $p_A = p_B$ so that $x_1 = 12$, $x_2 = 18$. This yields the test statistic $z = -1.19$ and P-value $= 0.2340$. Thus we do not reject H_0 and conclude that there is not a "statistically significant" difference between medications A and B.

b. The claim is $p_B = p_C$ so that $x_1 = 18$, $x_2 = 24$. This yields the test statistic $z = -1.04$ and P-value $= 0.2984$. Thus we do not reject H_0 and conclude that there is not a "statistically significant" difference between medications B and C.

c. The claim is $p_A = p_C$ so that $x_1 = 12$, $x_2 = 24$. This yields the test statistic $z = -2.21$ and P-value $= 0.0272$. Thus we reject H_0 and conclude that there is a "statistically significant" difference between medications A and C.

d. Our calculations so that medication A is "not significantly different" from B and B is "not significantly different" from C, but that A is "significantly different" from C. Therefore, the relation of "not significantly different" is not transitive.

5.4.9 At the 90% confidence level, the critical value is $z_{0.1/2} = 1.645$. In both parts, the parameters are

$$p_1 = \text{The proportion of females who regularly wear seat belts, and}$$
$$p_2 = \text{The proportion of males who regularly wear seat belts.}$$

a. The calculations of each confidence interval are shown in the table below. We see that these intervals overlap, so we informally conclude that there does not appear to be a statistically significant difference between these two populations.

	n	x	$\hat{p}$	E	C.I.
Females	150	90	0.6	0.0658	$0.534 < p_1 < 0.666$
Males	150	75	0.5	0.0672	$0.433 < p_2 < 0.567$

b. The claim is that $p_1 = p_2$ and the alternative hypothesis is $H_1 : p_1 \neq p_2$. For this test, the test statistic is $z = 1.74$, the critical region is $(-\infty, -1.645] \cup [1.645, \infty)$, and P-value $=$ 0.0818. We do not reject H_0 and conclude that there is not a statistically significant difference between these two populations. This is the same conclusion reached in part a.

5.4.10 In both parts, the parameters are

$$p_1 = \text{The proportion of suburban families that do at least one load of laundry a day, and}$$
$$p_2 = \text{The proportion of rural families that do at least one load of laundry a day.}$$

a. The claim and alternative hypothesis are $p_1 > p_2$. We have $x_1 = 550$, $n_1 = 2200$, $x_2 = 450$, and $n_2 = 2000$. This yields a test statistic of $z = 1.90$ and P-value $= 0.0287$. The critical region at the 95% confidence level is $[1.645, \infty)$. We reject H_0 and conclude that there is a "statistically significant" difference between the two populations.

b. Using formulas from section 4.8, the margin of error for the confidence interval is $E = 0.02537$ and the confidence interval is $0 < p_1 - p_2 < 0.05074$. This shows that the difference in the proportions is at most about 0.05, so there is not much practical difference.

5.4.11 We have $n_1 = 350$, $x_1 = 252$, $\hat{p}_1 = 252/350 = 0.72$, $n_2 = 400$, $x_2 = 220$, $p_2 = 220/400 = 0.55$, and $c = 0.1$ so that the test statistic is

$$z = \frac{(0.72 - 0.55) - 0.1}{\sqrt{\frac{0.72\,(1-0.72)}{350} + \frac{0.55\,(1-0.55)}{400}}} = 2.03.$$

At the 95% confidence level, the critical region is $[1.645, \infty)$ and P-value $= P(Z > 2.03) = 0.0212$. We reject H_0 and conclude that the data support the claim.

5.4.12 We have $n_1 = 13 + 21 = 34$, $x_1 = 21$, $\hat{p}_1 = 21/34 = 0.6177$, $n_2 = 67 + 19 = 86$, $x_2 = 19$, $p_2 = 19/86 = 0.2209$, and $c = 0.2$ so that the test statistic is

$$z = \frac{(0.6177 - 0.2209) - 0.2}{\sqrt{\frac{0.6177\,(1-0.6177)}{34} + \frac{0.2209\,(1-0.2209)}{86}}} = 2.08.$$

At the 95% confidence level, the critical region is $[1.645, \infty)$ and P-value $= P(Z > 2.08) = 0.0188$. We reject H_0 and conclude that the data support the claim.

5.5 Comparing Two Variances

5.5.1 The parameter are

$$\sigma_1^2 = \text{The variance of all weights before retooling, and}$$
$$\sigma_2^2 = \text{The variance of all weights after retooling.}$$

The claim is $\sigma_1^2 > \sigma_2^2$, and the hypotheses are

$$\boldsymbol{H_0} : \sigma_1^2 = \sigma_2^2, \qquad \boldsymbol{H_1} : \sigma_1^2 > \sigma_2^2.$$

The test statistic is $f = \frac{1.24^2}{0.54^2} = 5.27$. At the 0.05 significance level, the critical value is $f_{0.05}(40, 60) = 1.59$ and the critical region is $[1.59, \infty)$. The P-value is the area to the right of $f = 5.27$, which is approximately 0. Thus we reject H_0 and conclude that the data support the claim. The retooling appears to have worked.

5.5.2 The parameters are

$$\sigma_1^2 = \text{The variance of all white wines, and}$$
$$\sigma_2^2 = \text{The variance of all red wines.}$$

The claim is $\sigma_1^2 = \sigma_2^2$, and the hypotheses are

$$\boldsymbol{H_0} : \sigma_1^2 = \sigma_2^2, \qquad \boldsymbol{H_1} : \sigma_1^2 \neq \sigma_2^2.$$

The test statistic is $f = \frac{1.55^2}{1.24^2} = 1.56$. At the 0.05 significance level, the critical value is $f_{0.05/2}(30,\ 32) \approx 2.07$ and the critical region is $(-\infty,\ -2.07] \cup [2.07,\ \infty)$. The P-value is twice the area to the right of $f = 1.56$, which is approximately 0.219 found using software. Thus we do not reject H_0 and conclude that there is not sufficient evidence to reject the claim.

5.5.3 The parameters are

$$\sigma_1^2 = \text{The variance of the GPA's of all non-education majors, and}$$
$$\sigma_2^2 = \text{The variance of the GPA's of all education majors.}$$

The claim is $\sigma_1^2 > \sigma_2^2$, and the hypotheses are

$$\boldsymbol{H_0} : \sigma_1^2 = \sigma_2^2, \qquad \boldsymbol{H_1} : \sigma_1^2 > \sigma_2^2.$$

The test statistic is $f = \frac{0.449^2}{0.336^2} = 1.79$. At the 0.05 significance level, the critical value is $f_{0.05}(52,\ 46) \approx 1.64$ and the critical region is $[1.64,\ \infty)$. The P-value is the area to the right of $f = 1.79$, which is approximately 0.0232 found using software. Thus we reject H_0 and conclude that the data support the claim.

5.5.4 The parameters are

$$\sigma_1^2 = \text{The variance of the weights of all catfish in the lake, and}$$
$$\sigma_2^2 = \text{The variance of the weights of all bluegill in the lake.}$$

The claim is $\sigma_1^2 = \sigma_2^2$, and the hypotheses are

$$\boldsymbol{H_0} : \sigma_1^2 = \sigma_2^2, \qquad \boldsymbol{H_1} : \sigma_1^2 \neq \sigma_2^2.$$

The test statistic is $f = \frac{1.02^2}{0.15^2} = 46.24$. At the 0.05 significance level, the critical value is $f_{0.05/2}(17,\ 22) \approx 2.46$ and the critical region is $(-\infty,\ -2.46] \cup [2.46,\ \infty)$. The P-value is twice the area to the right of $f = 46.24$, which is approximately 0. Thus we reject H_0 and conclude that there is sufficient evidence to reject the claim.

5.5.5

a. $f_R = f_{0.025}(7,\ 12) = 3.61$, $f_L = 1/f_{0.025}(12,\ 7) = 1/4.67 = 0.214$
b. $f_R = f_{0.025}(9,\ 20) = 2.84$, $f_L = 1/f_{0.025}(20,\ 9) = 1/3.67 = 0.272$
c. $f_R = f_{0.025}(30,\ 60) = 1.82$, $f_L = 1/f_{0.025}(60,\ 30) = 1/1.94 = 0.515$
d. As the sample sizes get larger, f_r gets smaller and f_L gets larger, so they are closer together. This means the confidence interval gets narrower which means that s_1^2/s_2^2 is a better estimate of of σ_1^2/σ_2^2.

5.5.6

a. The parameters are

$$\sigma_1^2 = \text{The variance of the sodium content of all generic cereals, and}$$
$$\sigma_2^2 = \text{The variance of the sodium content of all brand-name cereals.}$$

Now, $f_R = f_{0.025}(61,\, 73) \approx 1.67$, and $f_L = 1/f_{0.025}(73,\, 61) \approx 1/1.67 = 0.599$ so that an approximate 95% confidence interval estimate of σ_1^2/σ_2^2 is

$$\left[\frac{1}{1.67}\cdot\frac{73.30^2}{60.69^2},\ \frac{1}{0.599}\cdot\frac{73.30^2}{60.69^2}\right] = [0.873,\ 2.436]$$

Using software to get more accurate values of f_R and f_L yields the confidence interval [0.903, 2.373]. Since this interval contains 1, there does not appear to be a significant difference between the population variances.

b. The claim is $\sigma_1^2 = \sigma_2^2$, and the alternative hypothesis is $H_1 : \sigma_1^2 \neq \sigma_2^2$. The test statistic is $f = \frac{73.3^2}{60.69^2} = 1.459$. At the 0.05 significance level, the critical value is $f_{0.05/2}(61,\, 73) \approx 1.67$ and the critical region is $(-\infty,\, -1.67] \cup [1.67,\, \infty)$. The P-value is twice the area to the right of $f = 1.459$, which is approximately 0.1222. Thus we do not reject H_0 and conclude that there is not sufficient evidence to reject the claim of equal variances. This agrees with the conclusion in part a.

5.5.7

a. The parameters are

$$\sigma_1^2 = \text{The variance of the masses of all peanut M\&M's, and}$$
$$\sigma_2^2 = \text{The variance of the masses of all regular M\&M's.}$$

Now, $f_R = f_{0.025}(30,\, 30) = 2.07$, and $f_L = 1/f_{0.025}(30,\, 30) = 1/2.07 = 0.483$ so that a 95% confidence interval estimate of σ_1^2/σ_2^2 is

$$\left[\frac{1}{2.07}\cdot\frac{0.30772^2}{0.04982^2},\ \frac{1}{0.483}\cdot\frac{0.30772^2}{0.04982^2}\right] = [18.43,\ 78.97]$$

Since this interval contains 50, there is not sufficient evidence to reject the claim that the variance of the peanut M&M's is about 50 times as large as that of the regular M&M's.

b. If $n_1 = n_2 = 201$, then $f_R = f_{0.025}(200,\, 200) = 1.32$, and $f_L = 1/1.32 = 0.758$ so that the confidence interval is [28.90, 50.34]. This confidence interval is narrower than that in part a.

5.5.8 The parameters are

$$\sigma_1^2 = \text{The variance of the times of all males, and}$$
$$\sigma_2^2 = \text{The variance of the times of all females.}$$

The claim is $\sigma_1^2 = \sigma_2^2$, and the hypotheses are

$$\boldsymbol{H_0} : \sigma_1^2 = \sigma_2^2, \qquad \boldsymbol{H_1} : \sigma_1^2 \neq \sigma_2^2.$$

From the data we get $n_1 = n_2 = 15$, $s_1 = 11.228$, and $s_2 = 7.448$. The test statistic is $f = \frac{11.228^2}{7.448^2} = 2.27$. At the 0.05 significance level, the critical value is $f_{0.05/2}(14, 14) \approx 2.86$ and the critical region is $(-\infty, -2.86] \cup [2.86, \infty)$. The P-value is twice the area to the right of $f = 2.274$, which is approximately 0.137. Thus we do not reject H_0 and conclude that there is not sufficient evidence to reject the claim. Thus it appears the assumption of equal variances is reasonable.

5.5.9 The parameters and summary statistics are the same as in exercise 5.5.8. Then, $f_R = f_{0.025}(14, 14) \approx 2.86$, and $f_L = 1/f_{0.025}(14, 14) \approx 1/2.86 = 0.350$ so that an approximate 95% confidence interval estimate of σ_1^2/σ_2^2 is

$$\left[\frac{1}{2.86} \cdot \frac{11.228^2}{7.448^2}, \frac{1}{0.483} \cdot \frac{11.228^2}{7.448^2}\right] = [0.795, 6.500]$$

Using software to get more accurate values of f_R and f_L yields the confidence interval [0.763, 6.769].

5.5.10 The parameters are

$$\sigma_1^2 = \text{The variance of the masses of all red candies, and}$$
$$\sigma_2^2 = \text{The variance of the masses of all blue candies.}$$

The claim is $\sigma_1^2 = \sigma_2^2$, and the hypotheses are

$$\boldsymbol{H_0} : \sigma_1^2 = \sigma_2^2, \qquad \boldsymbol{H_1} : \sigma_1^2 \neq \sigma_2^2.$$

From the data we get $n_1 = n_2 = 13$, $s_1 = 0.02238$, and $s_2 = 0.01896$. The test statistic is $f = \frac{0.02238^2}{0.01896^2} = 1.393$. At the 0.10 significance level, the critical value is $f_{0.10/2}(12, 12) = 2.69$ and the critical region is $(-\infty, -2.69] \cup [2.69, \infty)$. The P-value is twice the area to the right of $f = 1.393$, which is approximately 0.5748.

5.5.11 The parameters and summary statistics are the same as in exercise 5.5.10. Then, $f_R = f_{0.05}(12, 12) = 2.69$, and $f_L = 1/f_{0.05}(12, 12) = 1/2.69 = 0.372$ so that a 90% confidence interval estimate of σ_1^2/σ_2^2 is

$$\left[\frac{1}{2.69} \cdot \frac{0.02238^2}{0.01896^2}, \frac{1}{0.372} \cdot \frac{0.02238^2}{0.01896^2}\right] = [0.518, 3.748].$$

5.6 Comparing Two Means

5.6.1 The parameters are

$$\mu_1 = \text{The mean GPA of all education majors at the university, and}$$
$$\mu_2 = \text{The mean GPA of all non-education majors at the university.}$$

The claim is $\mu_1 - \mu_2 > 0$, so we test the hypotheses

$$\boldsymbol{H_0} : \mu_1 - \mu_2 = 0, \qquad \boldsymbol{H_1} : \mu_1 - \mu_2 > 0.$$

We have $\bar{x}_1 = 3.46$, $n_1 = 47$, $s_1 = 0.336$, $\bar{x}_2 = 3.38$, $n_2 = 53$, and $s_2 = 0.449$ so the test statistic is

$$z = \frac{(3.46 - 3.38) - 0}{\sqrt{0.336^2/47 + 0.449^2/53}} = 1.02.$$

At the 0.05 significance level, the critical value is $z_{0.05} = 1.645$ and the critical region is $[1.645, \infty)$. Also, P-value $= P(Z > 1.02) = 0.1539$. Thus we do not reject H_0 and conclude that the data do not support the claim.

5.6.2 The parameters are

$$\mu_1 = \text{The mean percent alcohol of all red wines, and}$$
$$\mu_2 = \text{The mean percent alcohol of all white wines.}$$

The claim is $\mu_1 - \mu_2 > 1$, so we test the hypotheses

$$\boldsymbol{H_0} : \mu_1 - \mu_2 = 1, \qquad \boldsymbol{H_1} : \mu_1 - \mu_2 > 1.$$

We have $\bar{x}_1 = 12.91$, $n_1 = 31$, $s_1 = 1.24$, $\bar{x}_2 = 11.30$, $n_2 = 33$, and $s_2 = 1.55$ so the test statistic is

$$z = \frac{(12.91 - 11.3) - 1}{\sqrt{1.24^2/31 + 1.55^2/33}} = 1.74.$$

At the 0.05 significance level, the critical value is $z_{0.05} = 1.645$ and the critical region is $[1.645, \infty)$. Also, P-value $= P(Z > 1.74) = 0.0409$. Thus we reject H_0 and conclude that the data support the claim.

5.6.3 The parameters are

$$\mu_1 = \text{The mean mood score of all students before chapel, and}$$
$$\mu_2 = \text{The mean mood score of all students after chapel.}$$

The claim is $\mu_1 - \mu_2 = 0$, so we test the hypotheses

$$\boldsymbol{H_0} : \mu_1 - \mu_2 = 0, \qquad \boldsymbol{H_1} : \mu_1 - \mu_2 = 0.$$

We have $\bar{x}_1 = 6.7$, $n_1 = 30$, $s_1 = 2.23$, $\bar{x}_2 = 6.67$, $n_2 = 30$, and $s_2 = 2.14$ so

$$s_p = \sqrt{\frac{(30-1)\,2.23^2 + (30-1)\,2.14^2}{30 + 30 - 2}} = 2.185 \text{ and } t = \frac{(6.7 - 6.67) - 0}{2.185\sqrt{1/30 + 1/30}} = 0.053.$$

At the 0.05 significance level, the critical value is $t_{0.05/2}(30 + 30 - 2) \approx 2$ and the critical region is $(-\infty, -2] \cup [2, \infty)$. Also, P-value $= 2 \cdot P(T > 0.053) = 0.958$, found using software. Thus we do not reject H_0 and conclude that there is not sufficient evidence to reject the claim. Thus there do not appear to be a significant difference between the moods before and after chapel.

5.6.4 The parameters are

$$\mu_1 = \text{The mean volume of all bottles of brand } A\text{, and}$$
$$\mu_2 = \text{The mean volume of all bottles of brand } B.$$

The claim is $\mu_1 - \mu_2 > 5$, so we test the hypotheses

$$\boldsymbol{H_0} : \mu_1 - \mu_2 = 5, \qquad \boldsymbol{H_1} : \mu_1 - \mu_2 > 5.$$

We have $\bar{x}_1 = 510.75$, $n_1 = 4$, $s_1 = 1.25$, $\bar{x}_2 = 505$, $n_2 = 4$, and $s_2 = 0.816$ so

$$s_p = \sqrt{\frac{(4-1)\,1.25^2 + (4-1)\,0.816^2}{4+4-2}} = 1.0555 \text{ and } t = \frac{(510.75 - 505) - 5}{1.0555\sqrt{1/4 + 1/4}} = 1.00.$$

At the 0.05 significance level, the critical value is $t_{0.05}(4+4-2) = 1.943$ and the critical region is $[1.943, \infty)$. Also, P-value $= P(T > 1.00) = 0.178$, found using software. Thus we do not reject H_0 and conclude that the data do not support the claim.

5.6.5 The parameters are

$$\mu_1 = \text{The mean number of items correctly recalled by all boys, and}$$
$$\mu_2 = \text{The mean number of items correctly recalled by all girls.}$$

The claim is $\mu_1 - \mu_2 = 0$, so we test the hypotheses

$$\boldsymbol{H_0} : \mu_1 - \mu_2 = 0, \qquad \boldsymbol{H_1} : \mu_1 - \mu_2 \neq 0.$$

We have $\bar{x}_1 = 9.35$, $n_1 = 40$, $s_1 = 1.61$, $\bar{x}_2 = 10.13$, $n_2 = 40$, and $s_2 = 1.44$ so

$$s_p = \sqrt{\frac{(40-1)\,1.61^2 + (40-1)\,1.44^2}{40+40-2}} = 1.527 \text{ and } t = \frac{(9.35 - 10.13) - 0}{1.527\sqrt{1/40 + 1/40}} = -2.28.$$

At the 0.05 significance level, the critical value is $t_{0.05/2}(40+40-2) \approx 1.990$ and the critical region is $(-\infty, -1.990] \cup [1.990, \infty)$. Also, P-value $= 2 \cdot P(T \leq -2.28) = 0.0253$, found using software. Thus we reject H_0 and conclude that there is sufficient evidence to reject the claim. Thus there does appear to be a significant difference between the mean recall of boys and girls.

5.6.6 The parameters are

$$\mu_1 = \text{The mean average velocity for baseballs hit with a metal bat, and}$$
$$\mu_2 = \text{The mean average velocity for baseballs hit with a wood bat.}$$

The claim is $\mu_1 - \mu_2 > 8$, so we test the hypotheses

$$\boldsymbol{H_0}: \mu_1 - \mu_2 = 8, \qquad \boldsymbol{H_1}: \mu_1 - \mu_2 > 8.$$

We have $\bar{x}_1 = 95.83$, $n_1 = 30$, $s_1 = 15.678$, $\bar{x}_2 = 81.16$, $n_2 = 30$, and $s_2 = 10.279$ so the test statistic and degrees of freedom are

$$t = \frac{(95.83 - 81.16) - 8}{\sqrt{15.678^2/30 + 10.279^2/30}} = 1.95 \text{ and } r = \frac{\left(\frac{15.678^2}{30} + \frac{10.279^2}{30}\right)^2}{\frac{1}{30-1}\left(\frac{15.678^2}{30}\right)^2 + \frac{1}{30-1}\left(\frac{10.279^2}{30}\right)^2} = 50.04$$

At the 0.05 significance level, the critical value is $t_{0.05}(50) = 1.676$ and the critical region is $[1.676, \infty)$. Also, P-value $= P(T > 1.95) = 0.0284$, found using software. Thus we reject H_0 and conclude that the data support the claim.

5.6.7

a. The parameters are

$$\mu_1 = \text{The mean height of all European players, and}$$
$$\mu_2 = \text{The mean height of all players from the Americas.}$$

The claim is $\mu_1 - \mu_2 > 0$, so we test the hypotheses

$$\boldsymbol{H_0}: \mu_1 - \mu_2 = 0, \qquad \boldsymbol{H_1}: \mu_1 - \mu_2 > 0.$$

We have $\bar{x}_1 = 181.26$, $n_1 = 61$, $s_1 = 5.69$, $\bar{x}_2 = 179.83$, $n_2 = 69$, and $s_2 = 7.19$ so the test statistic and degrees of freedom are

$$\frac{(181.26 - 179.83) - 0}{\sqrt{5.69^2/61 + 7.19^2/69}} = 1.26 \text{ and } r = \frac{\left(\frac{5.69^2}{61} + \frac{7.19^2}{69}\right)^2}{\frac{1}{61-1}\left(\frac{5.69^2}{61}\right)^2 + \frac{1}{69-1}\left(\frac{7.19^2}{69}\right)^2} = 126.5$$

The P-value for the Z-test is $2 \cdot P(Z > 1.26) = 0.2076$ and the P-value for the T-test is $2 \cdot P(T > 1.26) = 0.2100$, found using software, where T has 126 degrees of freedom.

b. There is very little difference between the P-values. This suggests that if we have large sample sizes, it does not really matter which test we use.

5.6.8 The parameters are

$$\mu_1 = \text{The mean output of all Vestas, and}$$
$$\mu_2 = \text{The mean output of all Micons.}$$

The claim is $\mu_1 - \mu_2 > 700$, so we test the hypotheses

$$\boldsymbol{H_0}: \mu_1 - \mu_2 = 700, \qquad \boldsymbol{H_1}: \mu_1 - \mu_2 > 700.$$

We have $\bar{x}_1 = 1019.22$, $n_1 = 9$, $s_1 = 9.54$, $\bar{x}_2 = 301.7$, $n_2 = 27$, and $s_2 = 10.5$, so

$$s_p = \sqrt{\frac{(9-1)\,9.54^2 + (27-1)\,10.5^2}{9+27-2}} = 10.282 \text{ and } t = \frac{(1019.22 - 301.7) - 700}{10.282\sqrt{1/27 + 1/9}} = 4.43.$$

At the 0.05 significance level, the critical value is $t_{0.05}(9+27-2) \approx 1.690$ and the critical region is $[1.690, \infty)$. Also, P-value $= P(T > 4.43) \approx 0$, found using software. Thus we reject H_0 and conclude that the data support the claim.

5.6.9 The parameters are

$$\mu_1 = \text{The mean mass of all peanut M\&M's, and}$$
$$\mu_2 = \text{The mean mass of all regular M\&M's.}$$

The claim is $\mu_1 - \mu_2 > 1.25$, so we test the hypotheses

$$\boldsymbol{H_0} : \mu_1 - \mu_2 = 1.25, \qquad \boldsymbol{H_1} : \mu_1 - \mu_2 > 1.25.$$

We have $\bar{x}_1 = 2.370$, $n_1 = 30$, $s_1 = 0.3077$, $\bar{x}_2 = 0.8889$, $n_2 = 30$, and $s_2 = 0.0498$ so the test statistic and degrees of freedom are

$$t = \frac{(2.370 - 0.8889) - 1.25}{\sqrt{0.30772^2/30 + 0.04982^2/30}} = 4.061, \quad r = \frac{\left(\frac{0.3077^2}{30} + \frac{0.0498^2}{30}\right)^2}{\frac{1}{30-1}\left(\frac{0.3077^2}{30}\right)^2 + \frac{1}{30-1}\left(\frac{0.0498^2}{30}\right)^2} = 30.519$$

At the 0.05 significance level, the critical value is $t_{0.05}(30) = 1.697$ and the critical region is $[1.697, \infty)$. Also, P-value $= P(T > 4.061) = 0.0002$, found using software. Thus we reject H_0 and conclude that the data support the claim.

5.6.10 The parameters are

$$\mu_1 = \text{The mean percent of shrink of all tan tanks, and}$$
$$\mu_2 = \text{The mean percent of shrink of all black tanks.}$$

The claim is $\mu_1 - \mu_2 = 0$, so we test the hypotheses

$$\boldsymbol{H_0} : \mu_1 - \mu_2 = 0, \qquad \boldsymbol{H_1} : \mu_1 - \mu_2 \neq 0.$$

From the data, we have $\bar{x}_1 = 2.5113$, $n_1 = 30$, $s_1 = 0.2676$, $\bar{x}_2 = 2.6207$, $n_2 = 30$, and $s_2 = 0.2930$ so

$$s_p = \sqrt{\frac{(30-1)\,0.2676^2 + (30-1)\,0.2930^2}{30+30-2}} = 0.2806, \quad t = \frac{(2.5113 - 2.6207) - 0}{0.2806\sqrt{1/30 + 1/30}} = -1.51.$$

At the 0.05 significance level, the critical value is $t_{0.05/2}(30+30-2) \approx 2$ and the critical region is $(-\infty, -2] \cup [2, \infty)$. Also, P-value $= 2 \cdot P(T \leq -1.51) = 0.1365$, found using software. Thus we do not reject H_0 and conclude that there is not sufficient evidence to reject the claim. Thus the color of the tank does not appear to affect the percent of shrink.

5.6.11 The parameters are

$$\mu_1 = \text{The mean minimum sentence for prisoners age 25 and under, and}$$
$$\mu_2 = \text{The minimum sentence for prisoners over age 25.}$$

The claim is $\mu_1 - \mu_2 < 0$, so we test the hypotheses

$$\boldsymbol{H_0} : \mu_1 - \mu_2 = 0, \qquad \boldsymbol{H_1} : \mu_1 - \mu_2 < 0.$$

From the data, we have $\bar{x}_1 = 4.2109$, $n_1 = 32$, $s_1 = 3.9036$, $\bar{x}_2 = 8.7813$, $n_2 = 32$, and $s_2 = 8.1507$ so the test statistic and degrees of freedom are

$$t = \frac{(4.2109 - 8.7813) - 0}{\sqrt{3.9036^2/30 + 0.04982^2/30}} = -2.86, \quad r = \frac{\left(\frac{3.9036^2}{32} + \frac{8.1507^2}{32}\right)^2}{\frac{1}{32-1}\left(\frac{3.9036^2}{32}\right)^2 + \frac{1}{32-1}\left(\frac{8.1507^2}{32}\right)^2} = 44.51$$

At the 0.05 significance level, the critical value is $-t_{0.05}(44) \approx -1.679$ and the critical region is $(-\infty, -1.679]$. Also, P-value $= P(T \leq -2.86) = 0.0032$, found using software. Thus we reject H_0 and conclude that the data support the claim. Thus the color of the tank does not appear to affect the percent of shrink.

5.6.12 The parameters are

$$\mu_1 = \text{The mean age of mothers who had their first child prior to 1993, and}$$
$$\mu_2 = \text{The mean age of mothers who had their first child in 1993 or later.}$$

The claim is $\mu_1 - \mu_2 < 0$, so we test the hypotheses

$$\boldsymbol{H_0} : \mu_1 - \mu_2 = 0, \qquad \boldsymbol{H_1} : \mu_1 - \mu_2 < 0.$$

From the data we have $\bar{x}_1 = 23.182$, $n_1 = 33$, $s_1 = 3.687$, $\bar{x}_2 = 25.455$, $n_2 = 33$, and $s_2 = 4.796$ so the test statistic is

$$z = \frac{(23.182 - 25.455) - 0}{\sqrt{3.687^2/33 + 4.796^2/33}} = -2.15.$$

At the 0.05 significance level, the critical value is $-z_{0.05} = -1.645$ and the critical region is $(-\infty, -1.645]$. Also, P-value $= P(Z < -2.15) = 0.0158$. Thus we reject H_0 and conclude that the data support the claim. Based on this result, it does appear that mothers now are having their first child at an older age.

5.6.13 The quantile plot is shown in Figure 5.2. Based on this plot, it does appear reasonable to assume that the population of differences is normally distributed.

5.6.14 The $n = 8$ differences are shown in the table below. The mean and standard deviation of these differences are $\bar{x} = 1.6125$ and $s = 1.1307$.

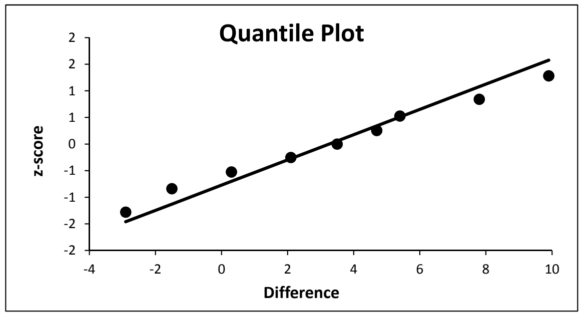


Figure 5.2

Before	53.8	54.2	52.9	53	55.1	58.2	56.7	52
After	52.7	53.8	51.1	52.7	54.2	55.4	53.3	49.8
Before - After	1.1	0.4	1.8	0.3	0.9	2.8	3.4	2.2

Define the parameter

$$\mu_d = \text{the population mean of the difference (Before - After).}$$

The claim is $\mu_d > 0$, and the hypotheses are

$$\boldsymbol{H_0} : \mu_d = 0, \qquad \boldsymbol{H_1} : \mu_d > 0.$$

The test statistic is $t = \frac{1.6125-0}{1.1307/\sqrt{8}} = 4.03$. At the 0.05 significance level, the critical value is $t_{0.05}(8-1) = 1.895$ and the critical region is $[1.895, \infty)$. Also, P-value $= P(T > 4.03) = 0.0025$, found using software. We reject H_0 and conclude that the data support the claim. Based on this result, the training program does appear to have a significant effect on track times.

5.6.15 The $n = 10$ differences are shown in the table below. The mean and standard deviation of these differences are $\bar{x} = 0.6$ and $s = 1.075$.

With	15.5	18.5	22.5	17.5	16.5	17.5	20.5	15.5	19.5	18.5
Without	15.5	17.5	22.5	15.5	15.5	18.5	21.5	14.5	17.5	17.5
With - Without	0	1	0	2	1	-1	-1	1	2	1

Define the parameter

$$\mu_d = \text{the population mean of the difference (With - Without).}$$

The claim is $\mu_d > 0$, and the hypotheses are

$$H_0 : \mu_d = 0, \qquad H_1 : \mu_d > 0.$$

The test statistic is $t = \frac{0.6-0}{1.075/\sqrt{10}} = 1.76$. At the 0.05 significance level, the critical value is $t_{0.05}(10-1) = 1.833$ and the critical region is $[1.833, \infty)$. Also, P-value $= P(T > 1.76) = 0.0561$, found using software. We do not reject H_0 and conclude that the data do not support the claim. Based on this result, shoes do not appear to have a significant affect at the 0.05 significance level.

5.6.16 The $n = 12$ differences are shown in the table below. The mean and standard deviation of these differences are $\bar{x} = 0.4083$ and $s = 0.4562$.

Reported	67	73	67.5	70	68	70	71	68	67.5	66	70	71.5
Measured	66.3	72.8	68	69.4	67.3	70.2	70.9	67.6	67.1	65.3	68.8	70.9
Reported - Measured	0.7	0.2	-0.5	0.6	0.7	-0.2	0.1	0.4	0.4	0.7	1.2	0.6

Define the parameter

$\mu_d =$ the population mean of the difference (Reported - Measured).

The claim is $\mu_d > 0$, and the hypotheses are

$$H_0 : \mu_d = 0, \qquad H_1 : \mu_d > 0.$$

The test statistic is $t = \frac{0.4083-0}{0.4562/\sqrt{12}} = 3.10$. At the 0.05 significance level, the critical value is $t_{0.05}(12-1) = 1.796$ and the critical region is $[1.796, \infty)$. Also, P-value $= P(T > 3.10) = 0.0051$, found using software. We reject H_0 and conclude that the data support the claim. Students may have just been rounding up to the next whole number.

5.6.17 The $n = 6$ differences are shown in the table below. The mean and standard deviation of these differences are $\bar{x} = -0.8833$ and $s = 0.9948$.

Bi-metallic	87.5	82.9	59.2	49.2	42.7	32.9
Infrared	88.7	83.7	61.9	49.1	43.1	33.2
Bi-metallic - Infrared	-1.2	-0.8	-2.7	0.1	-0.4	-0.3

Define the parameter

$\mu_d =$ the population mean of the difference (Bi-metallic - Infared).

We test the claim $\mu_d = 0$, and the hypotheses are

$$H_0 : \mu_d = 0, \qquad H_1 : \mu_d \neq 0.$$

The test statistic is $t = \frac{-0.8833-0}{0.9948/\sqrt{6}} = -2.17$. At the 0.05 significance level, the critical value is $t_{0.05/2}(6-1) = 2.571$ and the critical region is $(-\infty, -2.571] \cup [2.571, \infty)$. Also, P-value $= 2 \cdot P(T < -2.17) = 0.0821$, found using software. We do not reject H_0 and conclude that there is not sufficient evidence to reject the claim. Thus there does not appear to be a significant difference between the temperature readings of the two types of thermometers.

5.6.18 The parameters are

$$\mu_1 = \text{The mean amount of sodium in all generic cereals, and}$$
$$\mu_2 = \text{The mean amount of sodium in all brand-name cereals.}$$

a. Since both sample sizes are greater than 30, we construct Z-intervals. The calculations are summarized in the table below. Since these intervals overlap, there does not appear to be a significant difference between the population means.

	n	$\bar{x}$	s	E	**C.I.**
Generic	62	158.22	73.30	15.31	$142.91 < \mu_1 < 173.5$
Brand-name	74	137.79	60.69	11.6	$126.19 < \mu_2 < 149.39$

b. Since both sample sizes are greater than 30, we perform a Z-test with the alternative hypothesis $H_1 : \mu_1 - \mu_2 \neq 0$. The test statistic is

$$z = \frac{(158.22 - 137.79) - 0}{\sqrt{73.3^2/62 + 60.69^2/74}} = 1.75.$$

At the 0.10 significance level, the critical value is $z_{0.10} = 1.282$ and the critical region is $(-\infty, -1.282] \cup [1.282, \infty)$. Also, P-value $= 2 \cdot P(Z > 1.75) = 0.0802$. Thus we reject H_0 and conclude that there is sufficient evidence to reject the claim. Thus there does appear to be a difference between these two population means. This is a different conclusion reached than in part a.

5.6.19

a. Quantile plots of the data are shown in Figure 5.3. Based on these plots, it appears that both populations are normally distributed.
b. These data have the same summary statistics as in exercise 5.5.10, so we come to the same conclusion as in that exercise: There is not sufficient evidence to reject the claim of equal population variances.
c. The parameters are

$$\mu_1 = \text{The mean mass of all blue candies, and}$$
$$\mu_2 = \text{The mean mass of all red candies.}$$

The claim is $\mu_1 - \mu_2 > 0$, so we test the hypotheses

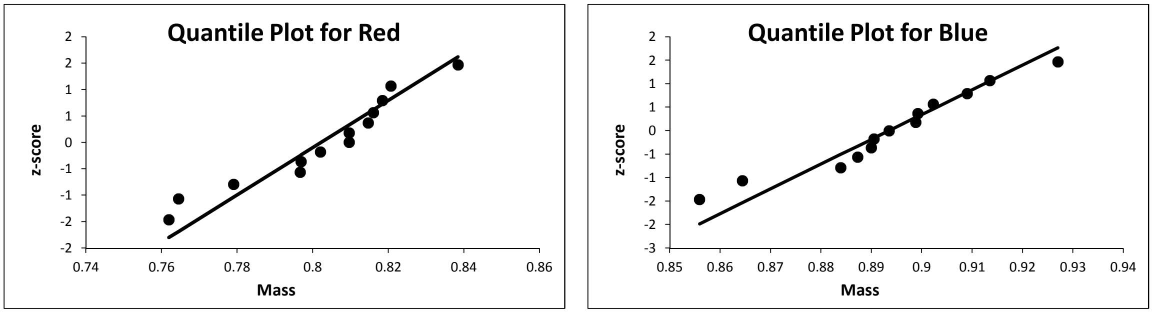


Figure 5.3

$$\boldsymbol{H_0}: \mu_1 - \mu_2 = 0, \qquad \boldsymbol{H_1}: \mu_1 - \mu_2 > 0.$$

From the data we get $n_1 = n_2 = 13$, $\bar{x}_1 = 0.8935$, $s_1 = 0.0190$, $\bar{x}_2 = 0.8022$, and $s_2 = 0.0224$ so

$$s_p = \sqrt{\frac{(13-1)\,0.019^2 + (13-1)\,0.0224^2}{13+13-2}} = 0.0207, \quad t = \frac{(0.8935 - 0.8022) - 0}{0.0207\sqrt{1/13 + 1/13}} = 11.24.$$

At the 95% confidence level, the critical value is $t_{0.05}(13+13-2) = 1.711$ and the critical region is $[1.711, \infty)$. Also, P-value $= P(T > 11.24) = 0.000$. Thus we reject H_0 and conclude that the data support the claim.

5.6.20 From the data, we get $m_1 = 7.77$ and $m_2 = 8.15$. The transformed values are shown in the table below.

Sample 1	5.66	6.62	6.76	7.58	7.66	7.77	7.95	8.38	8.38	8.05	9.13
$\lvert x - m_1 \rvert$	2.11	1.15	1.01	0.19	0.11	0	0.18	0.61	0.61	0.73	1.36
Sample 2	7.41	7.62	7.63	7.73	7.84	8.15	8.26	8.27	8.46	8.54	8.75
$\lvert x - m_2 \rvert$	0.74	0.53	0.52	0.42	0.31	0	0.11	0.12	0.31	0.39	0.6

The parameters are

$\mu_1 =$ The mean of the transformed values in the first population, and
$\mu_2 =$ The mean of the transformed values in the second population.

The claim is $\mu_1 - \mu_2 = 0$, so we test the hypotheses

$$\boldsymbol{H_0}: \mu_1 - \mu_2 = 0, \qquad \boldsymbol{H_1}: \mu_1 - \mu_2 \neq 0.$$