Contents

Preface

This solutions manual provides answers for the even-numbered exercises in *Probability and Statistical Inference*, 8th edition, by Robert V. Hogg and Elliot A. Tanis. Complete solutions are given for most of these exercises. You, the instructor, may decide how many of these answers you want to make available to your students. Note that the answers for the odd-numbered exercises are given in the textbook.

All of the figures in this manual were generated using *Maple*, a computer algebra system. Most of the figures were generated and many of the solutions, especially those involving data, were solved using procedures that were written by Zaven Karian from Denison University. We thank him for providing these. These procedures are available free of charge for your use. They are available on the CD-ROM in the textbook. Short descriptions of these procedures are provided on the "Maple Card" on the CD-ROM. Complete descriptions of these procedures are given in *Probability and Statistics: Explorations with MAPLE*, second edition, 1999, written by Zaven Karian and Elliot Tanis, published by Prentice Hall (ISBN 0-13-021536-8).

Our hope is that this solutions manual will be helpful to each of you in your teaching.

If you find an error or wish to make a suggestion, send these to Elliot Tanis at `tanis@hope.edu` and he will post corrections on his web page, http://www.math.hope.edu/tanis/.

R.V.H.
E.A.T.

Chapter 1

Probability

1.1 Basic Concepts

1.1-2 **(a)** $S = \{$bbb, gbb, bgb, bbg, bgg, gbg, ggb, ggg$\}$;

    (b) $S = \{$female, male$\}$;

    (c) $S = \{000, 001, 002, 003, \ldots, 999\}$.

1.1-4 **(a)**

Clutch size:	4	5	6	7	8	9	10	11	12	13	14
Frequency:	3	5	7	27	26	37	8	2	0	1	1

    (b)
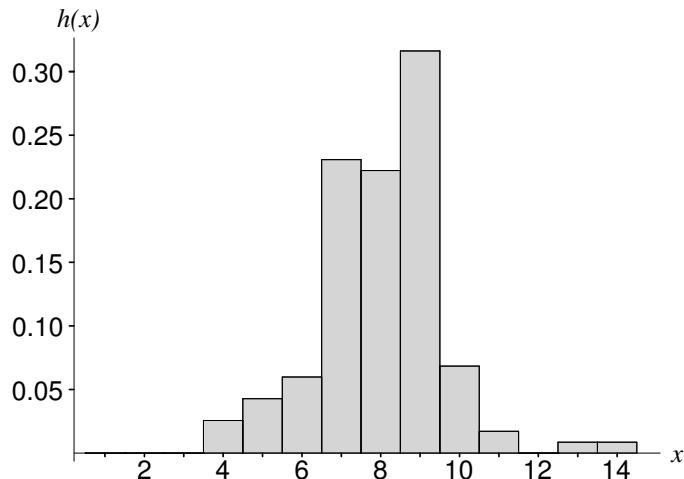
Figure 1.1–4: Clutch sizes for the common gallinule

    (c) 9.

1

1.1-6 **(a)**

No. Boxes:	4	5	6	7	8	9	10	11	12	13	14	15	16	19	24
Frequency:	10	19	13	8	13	7	9	5	2	4	4	2	2	1	1

(b)
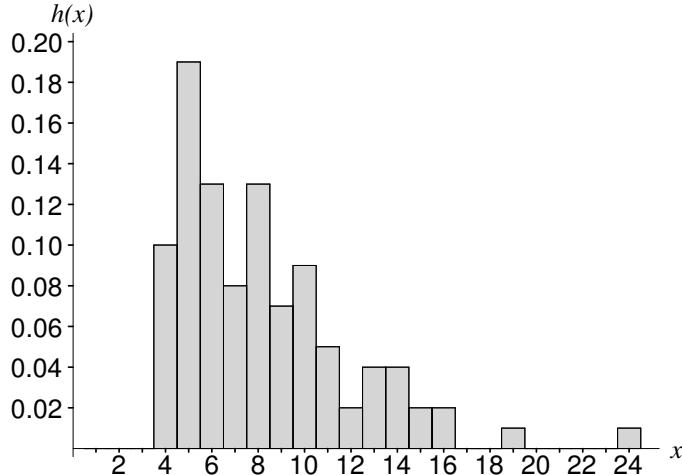
Figure 1.1–6: Number of boxes of cereal

1.1-8 **(a)** $f(1) = \dfrac{2}{10}, f(2) = \dfrac{3}{10}, f(3) = \dfrac{3}{10}, f(4) = \dfrac{2}{10}.$

1.1-10 This is an experiment.

1.1-12 **(a)** $50/204 = 0.245$; $93/329 = 0.283$;

(b) $124/355 = 0.349$; $21/58 = 0.362$;

(c) $174/559 = 0.311$; $114/387 = 0.295$;

(d) Although James' batting average is higher that Hrbek's on both grass and artificial turf, Hrbek's is higher over all. Note the different numbers of at bats on grass and artificial turf and how this affects the batting averages.

1.2   Properties of Probability

1.2-2 Sketch a figure and fill in the probabilities of each of the disjoint sets.

Let $A = \{\text{insure more than one car}\}$, $P(A) = 0.85$.

Let $B = \{\text{insure a sports car}\}$, $P(B) = 0.23$.

Let $C = \{\text{insure exactly one car}\}$, $P(C) = 0.15$.

It is also given that $P(A \cap B) = 0.17$. Since $P(A \cap C) = 0$, it follows that

$P(A \cap B \cap C') = 0.17$. Thus $P(A' \cap B \cap C') = 0.06$ and $P(A' \cap B' \cap C) = 0.09$.

1.2-4 **(a)** $S = \{$HHHH, HHHT, HHTH, HTHH, THHH, HHTT, HTTH, TTHH, HTHT, THTH, THHT, HTTT, THTT, TTHT, TTTH, TTTT$\}$;

(b) **(i)** $5/16$, **(ii)** 0, **(iii)** $11/16$, **(iv)** $4/16$, **(v)** $4/16$, **(vi)** $9/16$, **(vii)** $4/16$.

1.2-6 **(a)** $1/6$;

(b) $P(B) = 1 - P(B') = 1 - P(A) = 5/6$;

(c) $P(A \cup B) = P(S) = 1$.

1.2-8 **(a)** $P(A \cup B) = 0.4 + 0.5 - 0.3 = 0.6;$

(b)
$$
\begin{aligned}
A &= (A \cap B') \cup (A \cap B) \\
P(A) &= P(A \cap B') + P(A \cap B) \\
0.4 &= P(A \cap B') + 0.3 \\
P(A \cap B) &= 0.1;
\end{aligned}
$$

(c) $P(A' \cup B') = P[(A \cap B)'] = 1 - P(A \cap B) = 1 - 0.3 = 0.7.$

1.2-10 Let $A =$ {lab work done}, $B =$ {referral to a specialist},

$P(A) = 0.41$, $P(B) = 0.53$, $P([A \cup B]') = 0.21.$

$$
\begin{aligned}
P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\
0.79 &= 0.41 + 0.53 - P(A \cap B) \\
P(A \cap B) &= 0.41 + 0.53 - 0.79 = 0.15.
\end{aligned}
$$

1.2-12
$$
\begin{aligned}
A \cup B \cup C &= A \cup (B \cup C) \\
P(A \cup B \cup C) &= P(A) + P(B \cup C) - P[A \cap (B \cup C)] \\
&= P(A) + P(B) + P(C) - P(B \cap C) - P[(A \cap B) \cup (A \cap C)] \\
&= P(A) + P(B) + P(C) - P(B \cap C) - P(A \cap B) - P(A \cap C) \\
&\quad + P(A \cap B \cap C).
\end{aligned}
$$

1.2-14 **(a)** $1/3$; **(b)** $2/3$; **(c)** 0; **(d)** $1/2$.

1.2-16 **(a)** $S = \{(1,2), (1,3), (1,4), (1,5), (2,3), (2,4), (2,5), (3,4), (3,5), (4,5)\};$
(b) **(i)** $1/10$; **(ii)** $5/10$.

1.2-18 $P(A) = \dfrac{2[r - r(\sqrt{3}/2)]}{2r} = 1 - \dfrac{\sqrt{3}}{2}.$

1.2-20 Note that the respective probabilities are p_0, $p_1 = p_0/4$, $p_2 = p_0/4^2, \ldots$.

$$
\begin{aligned}
\sum_{k=0}^{\infty} \frac{p_0}{4^k} &= 1 \\
\frac{p_0}{1 - 1/4} &= 1 \\
p_0 &= \frac{3}{4} \\
1 - p_0 - p_1 &= 1 - \frac{15}{16} = \frac{1}{16}.
\end{aligned}
$$

1.3 Methods of Enumeration

1.3-2 $(4)(3)(2) = 24.$

1.3-4 **(a)** $(4)(5)(2) = 40$; **(b)** $(2)(2)(2) = 8.$

1.3-6 **(a)** $4\dbinom{6}{3} = 80;$

(b) $4(2^6) = 256;$

(c) $\dfrac{(4 - 1 + 3)!}{(4 - 1)!3!} = 20.$

1.3-8 $_9P_4 = \dfrac{9!}{5!} = 3024.$

1.3-10 $S =\{$ HHH, HHCH, HCHH, CHHH, HHCCH, HCHCH, CHHCH, HCCHH, CHCHH, CCHHH, CCC, CCHC, CHCC, HCCC, CCHHC, CHCHC, HCCHC, CHHCC, HCHCC, HHCCC $\}$ so there are 20 possibilities.

1.3-12 $3 \cdot 3 \cdot 2^{12} = 36,864.$

1.3-14
$$\binom{n-1}{r} + \binom{n-1}{r-1} = \frac{(n-1)!}{r!(n-1-r)!} + \frac{(n-1)!}{(r-1)!(n-r)!}$$
$$= \frac{(n-r)(n-1)! + r(n-1)!}{r!(n-r)!} = \frac{n!}{r!(n-r)!} = \binom{n}{r}.$$

1.3-16
$$0 = (1-1)^n = \sum_{r=0}^{n}\binom{n}{r}(-1)^r(1)^{n-r} = \sum_{r=0}^{n}(-1)^r\binom{n}{r}.$$

$$2^n = (1+1)^n = \sum_{r=0}^{n}\binom{n}{r}(1)^r(1)^{n-r} = \sum_{r=0}^{n}\binom{n}{r}.$$

1.3-18
$$\binom{n}{n_1, n_2, \ldots, n_s} = \binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n-n_1-\cdots-n_{s-1}}{n_s}$$
$$= \frac{n!}{n_1!(n-n_1)!} \cdot \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!}$$
$$\cdot \frac{(n-n_1-n_2)!}{n_3!(n-n_1-n_2-n_3)!}\cdots\frac{(n-n_1-n_2-\cdots-n_{s-1})!}{n_s!0!}$$
$$= \frac{n!}{n_1!n_2!\ldots n_s!}.$$

1.3-20 (a) $\dfrac{\binom{19}{3}\binom{52-19}{6}}{\binom{52}{9}} = \dfrac{102,486}{351,325} = 0.2917;$

     (b) $\dfrac{\binom{19}{3}\binom{10}{2}\binom{7}{1}\binom{3}{0}\binom{5}{1}\binom{2}{0}\binom{6}{2}}{\binom{52}{9}} = \dfrac{7,695}{1,236,664} = 0.00622.$

1.3-22 $\binom{45}{36} = 886{,}163{,}135.$

1.4   Conditional Probability

1.4-2 (a) $\dfrac{1041}{1456};$

     (b) $\dfrac{392}{633};$

     (c) $\dfrac{649}{823}.$

     (d) The proportion of women who favor a gun law is greater than the proportion of men who favor a gun law.

1.4-4 **(a)** $P(\text{HH}) = \dfrac{13}{52} \cdot \dfrac{12}{51} = \dfrac{1}{17};$

(b) $P(\text{HC}) = \dfrac{13}{52} \cdot \dfrac{13}{51} = \dfrac{13}{204};$

(c) $P(\text{Non-Ace Heart, Ace}) + P(\text{Ace of Hearts, Non-Heart Ace})$

$$= \dfrac{12}{52} \cdot \dfrac{4}{51} + \dfrac{1}{52} \cdot \dfrac{3}{51} = \dfrac{51}{52 \cdot 51} = \dfrac{1}{52}.$$

1.4-6 Let $A = \{3 \text{ or } 4 \text{ kings}\}$, $B = \{2, 3, \text{ or } 4 \text{ kings}\}$.

$$P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{N(A)}{N(B)}$$

$$= \dfrac{\binom{4}{3}\binom{48}{10} + \binom{4}{4}\binom{48}{9}}{\binom{4}{2}\binom{48}{11} + \binom{4}{3}\binom{48}{10} + \binom{4}{4}\binom{48}{9}} = 0.170.$$

1.4-8 Let $H = \{\text{died from heart disease}\}$; $P = \{\text{at least one parent had heart disease}\}$.

$$P(H \mid P') = \dfrac{N(H \cap P')}{N(P')} = \dfrac{110}{648}.$$

1.4-10 **(a)** $\dfrac{3}{20} \cdot \dfrac{2}{19} \cdot \dfrac{1}{18} = \dfrac{1}{1140};$

(b) $\dfrac{\binom{3}{2}\binom{17}{1}}{\binom{20}{3}} \cdot \dfrac{1}{17} = \dfrac{1}{760};$

(c) $\displaystyle\sum_{k=1}^{9} \dfrac{\binom{3}{2}\binom{17}{2k-2}}{\binom{20}{2k}} \cdot \dfrac{1}{20-2k} = \dfrac{35}{76} = 0.4605.$

(d) Draw second. The probability of winning in $1 - 0.4605 = 0.5395$.

1.4-12 $\dfrac{\binom{2}{0}\binom{8}{5}}{\binom{10}{5}} \cdot \dfrac{2}{5} + \dfrac{\binom{2}{1}\binom{8}{4}}{\binom{10}{5}} \cdot \dfrac{1}{5} = \dfrac{1}{5}.$

1.4-14 **(a)** $P(A) = \dfrac{52}{52} \cdot \dfrac{51}{52} \cdot \dfrac{50}{52} \cdot \dfrac{49}{52} \cdot \dfrac{48}{52} \cdot \dfrac{47}{52} = \dfrac{8,808,975}{11,881,376} = 0.74141;$

(b) $P(A') = 1 - P(A) = 0.25859.$

1.4-16 **(a)** It doesn't matter because $P(B_1) = \dfrac{1}{18}$, $P(B_5) = \dfrac{1}{18}$, $P(B_{18}) = \dfrac{1}{18};$

(b) $P(B) = \dfrac{2}{18} = \dfrac{1}{9}$ on each draw.

1.4-18 $\dfrac{3}{5} \cdot \dfrac{5}{8} + \dfrac{2}{5} \cdot \dfrac{4}{8} = \dfrac{23}{40}.$

1.4-20 **(a)** $P(A_1) = 30/100;$

(b) $P(A_3 \cap B_2) = 9/100;$

(c) $P(A_2 \cup B_3) = 41/100 + 28/100 - 9/100 = 60/100;$

(d) $P(A_1 \mid B_2) = 11/41;$

(e) $P(B_1 \mid A_3) = 13/29.$

1.5 Independent Events

1.5-2 **(a)** $\begin{aligned} P(A \cap B) &= P(A)P(B) = (0.3)(0.6) = 0.18; \\ P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 0.3 + 0.6 - 0.18 \\ &= 0.72. \end{aligned}$

(b) $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{0}{0.6} = 0.$

1.5-4 Proof of **(b)**: $\begin{aligned} P(A' \cap B) &= P(B)P(A'|B) \\ &= P(B)[1 - P(A|B)] \\ &= P(B)[1 - P(A)] \\ &= P(B)P(A'). \end{aligned}$

Proof of **(c)**: $\begin{aligned} P(A' \cap B') &= P[(A \cup B)'] \\ &= 1 - P(A \cup B) \\ &= 1 - P(A) - P(B) + P(A \cap B) \\ &= 1 - P(A) - P(B) + P(A)P(B) \\ &= [1 - P(A)][1 - P(B)] \\ &= P(A')P(B'). \end{aligned}$

1.5-6 $\begin{aligned} P[A \cap (B \cap C)] &= P[A \cap B \cap C] \\ &= P(A)P(B)P(C) \\ &= P(A)P(B \cap C). \end{aligned}$

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

$\begin{aligned} P[A' \cap (B \cap C')] &= P(A' \cap C' \cap B) \\ &= P(B)[P(A' \cap C') \mid B] \\ &= P(B)[1 - P(A \cup C \mid B)] \\ &= P(B)[1 - P(A \cup C)] \\ &= P(B)P[(A \cup C)'] \\ &= P(B)P(A' \cap C') \\ &= P(B)P(A')P(C') \\ &= P(A')P(B)P(C') \\ &= P(A')P(B \cap C') \end{aligned}$

$\begin{aligned} P[A' \cap B' \cap C'] &= P[(A \cup B \cup C)'] \\ &= 1 - P(A \cup B \cup C) \\ &= 1 - P(A) - P(B) - P(C) + P(A)P(B) + P(A)P(C) + \\ &\qquad P(B)P(C) - PA)P(B)P(C) \\ &= [1 - P(A)][1 - P(B)][1 - P(C)] \\ &= P(A')P(B')P(C'). \end{aligned}$

1.5-8 $\dfrac{1}{6} \cdot \dfrac{2}{6} \cdot \dfrac{3}{6} + \dfrac{1}{6} \cdot \dfrac{4}{6} \cdot \dfrac{3}{6} + \dfrac{5}{6} \cdot \dfrac{2}{6} \cdot \dfrac{3}{6} = \dfrac{2}{9}.$

1.5-10 **(a)** $\dfrac{3}{4} \cdot \dfrac{3}{4} = \dfrac{9}{16}$;

$\qquad$ **(b)** $\dfrac{1}{4} \cdot \dfrac{3}{4} + \dfrac{3}{4} \cdot \dfrac{2}{4} = \dfrac{9}{16}$;

$\qquad$ **(c)** $\dfrac{2}{4} \cdot \dfrac{1}{4} + \dfrac{2}{4} \cdot \dfrac{4}{4} = \dfrac{10}{16}.$

1.5-12 **(a)** $\left(\dfrac{1}{2}\right)^3 \left(\dfrac{1}{2}\right)^2$;

$\qquad$ **(b)** $\left(\dfrac{1}{2}\right)^3 \left(\dfrac{1}{2}\right)^2$;

$\qquad$ **(c)** $\left(\dfrac{1}{2}\right)^3 \left(\dfrac{1}{2}\right)^2$;

$\qquad$ **(d)** $\dfrac{5!}{3!\,2!} \left(\dfrac{1}{2}\right)^3 \left(\dfrac{1}{2}\right)^2.$

1.5-14 **(a)** $1 - (0.4)^3 = 1 - 0.064 = 0.936$;

$\qquad$ **(b)** $1 - (0.4)^8 = 1 - 0.00065536 = 0.99934464.$

1.5-16 **(a)** $\displaystyle\sum_{k=0}^{\infty} \dfrac{1}{5} \left(\dfrac{4}{5}\right)^{2k} = \dfrac{5}{9}$;

$\qquad$ **(b)** $\dfrac{1}{5} + \dfrac{4}{5} \cdot \dfrac{3}{4} \cdot \dfrac{1}{3} + \dfrac{4}{5} \cdot \dfrac{3}{4} \cdot \dfrac{2}{3} \cdot \dfrac{1}{2} \cdot \dfrac{1}{1} = \dfrac{3}{5}.$

1.5-18 **(a)** 7; **(b)** $(1/2)^7$; **(c)** 63; **(d)** No! $(1/2)^{63} = 1/9{,}223{,}372{,}036{,}854{,}775{,}808.$

1.5-20

n	3	6	9	12	15
(a)	0.7037	0.6651	0.6536	0.6480	0.6447
(b)	0.6667	0.6319	0.6321	0.6321	0.6321

$\qquad$ **(c)** Very little when $n > 15$, sampling with replacement

$\qquad\qquad$ Very little when $n > 10$, sampling without replacement.

$\qquad$ **(d)** Convergence is faster when sampling with replacement.

1.6 Bayes's Theorem

1.6-2 **(a)** $\begin{aligned} P(G) &= P(A \cap G) + P(B \cap G) \\ &= P(A)P(G\,|\,A) + P(B)P(G\,|\,B) \\ &= (0.40)(0.85) + (0.60)(0.75) = 0.79; \end{aligned}$

$\qquad$ **(b)** $\begin{aligned} P(A\,|\,G) &= \dfrac{P(A \cap G)}{P(G)} \\[2mm] &= \dfrac{(0.40)(0.85)}{0.79} = 0.43. \end{aligned}$

1.6-4 Let event B denote an accident and let A_1 be the event that age of the driver is 16–25. Then

$$
\begin{aligned}
P(A_1 \,|\, B) &= \frac{(0.1)(0.05)}{(0.1)(0.05) + (0.55)(0.02) + (0.20)(0.03) + (0.15)(0.04)} \\[2mm]
&= \frac{50}{50 + 110 + 60 + 60} = \frac{50}{280} = 0.179.
\end{aligned}
$$

1.6-6 Let B be the event that the policyholder dies. Let A_1, A_2, A_3 be the events that the deceased is standard, preferred and ultra-preferred, respectively. Then

$$
\begin{aligned}
P(A_1 \,|\, B) &= \frac{(0.60)(0.01)}{(0.60)(0.01) + (0.30)(0.008) + (0.10)(0.007)} \\[2mm]
&= \frac{60}{60 + 24 + 7} = \frac{60}{91} = 0.659; \\[2mm]
P(A_2 \,|\, B) &= \frac{24}{91} = 0.264; \\[2mm]
P(A_3 \,|\, B) &= \frac{7}{91} = 0.077.
\end{aligned}
$$

1.6-8 Let A be the event that the DVD player is under warranty.

$$
\begin{aligned}
P(B_1 \,|\, A) &= \frac{(0.40)(0.10)}{(0.40)(0.10) + (0.30)(0.05) + (0.20)(0.03) + (0.10)(0.02)} \\[2mm]
&= \frac{40}{40 + 15 + 6 + 2} = \frac{40}{63} = 0.635; \\[2mm]
P(B_2 \,|\, A) &= \frac{15}{63} = 0.238; \\[2mm]
P(B_3 \,|\, A) &= \frac{6}{63} = 0.095; \\[2mm]
P(B_4 \,|\, A) &= \frac{2}{63} = 0.032.
\end{aligned}
$$

1.6-10 **(a)** $P(AD) = (0.02)(0.92) + (0.98)(0.05) = 0.0184 + 0.0490 = 0.0674;$

   (b) $P(N \,|\, AD) = \dfrac{0.0490}{0.0674} = 0.727; \quad P(A \,|\, AD) = \dfrac{0.0184}{0.0674} = 0.273;$

   (c) $P(N \,|\, ND) = \dfrac{(0.98)(0.95)}{(0.02)(0.08) + (0.98)(0.95)} = \dfrac{9310}{16 + 9310} = 0.998; \quad P(A \,|\, ND) = 0.002.$

   (d) Yes, particularly those in part (b).

1.6-12 Let $D = \{\text{has the disease}\}$, $DP = \{\text{detects presence of disease}\}$. Then

$$
\begin{aligned}
P(D \,|\, DP) &= \frac{P(D \cap DP)}{P(DP)} \\[2mm]
&= \frac{P(D) \cdot P(DP \,|\, D)}{P(D) \cdot P(DP \,|\, D) + P(D') \cdot P(DP \,|\, D')} \\[2mm]
&= \frac{(0.005)(0.90)}{(0.005)(0.90) + (0.995)(0.02)} \\[2mm]
&= \frac{0.0045}{0.0045 + 0.199} = \frac{0.0045}{0.0244} = 0.1844.
\end{aligned}
$$

1.6-14 Let $D = \{$defective roll$\}$ Then

$$
\begin{aligned}
P(I \mid D) &= \frac{P(I \cap D)}{P(D)} \\[2mm]
&= \frac{P(I) \cdot P(D \mid I)}{P(I) \cdot P(D \mid I) + P(II) \cdot P(D \mid II)} \\[2mm]
&= \frac{(0.60)(0.03)}{(0.60)(0.03) + (0.40)(0.01)} \\[2mm]
&= \frac{0.018}{0.018 + 0.004} = \frac{0.018}{0.022} = 0.818.
\end{aligned}
$$

Chapter 2

Discrete Distributions

2.1 Random Variables of the Discrete Type

2.1-2 **(a)**
$$f(x) = \begin{cases} 0.6, & x = 1, \\ 0.3, & x = 5, \\ 0.1, & x = 10, \end{cases}$$

(b)
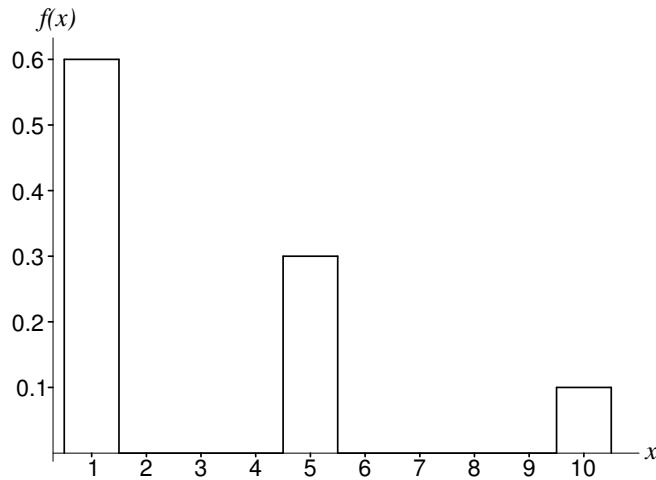
Figure 2.1–2: A probability histogram

2.1-4 **(a)** $f(x) = \dfrac{1}{10}$, $x = 0, 1, 2, \cdots, 10$;

(b) $\mathcal{N}(\{0\})/150 = 11/150 = 0.073$; $\quad \mathcal{N}(\{5\})/150 = 13/150 = 0.087$;
$\mathcal{N}(\{1\})/150 = 14/150 = 0.093$; $\quad \mathcal{N}(\{6\})/150 = 22/150 = 0.147$;
$\mathcal{N}(\{2\})/150 = 13/150 = 0.087$; $\quad \mathcal{N}(\{7\})/150 = 16/150 = 0.107$;
$\mathcal{N}(\{3\})/150 = 12/150 = 0.080$; $\quad \mathcal{N}(\{8\})/150 = 18/150 = 0.120$;
$\mathcal{N}(\{4\})/150 = 16/150 = 0.107$; $\quad \mathcal{N}(\{9\})/150 = 15/150 = 0.100$.

(c)
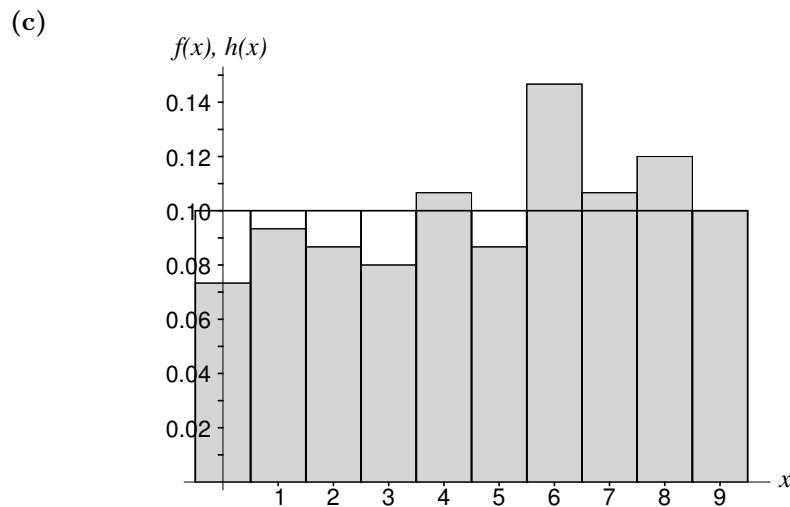
Figure 2.1–4: Michigan daily lottery digits

2.1-6 **(a)** $f(x) = \dfrac{6 - |7 - x|}{36}, \ x = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12.$

(b)
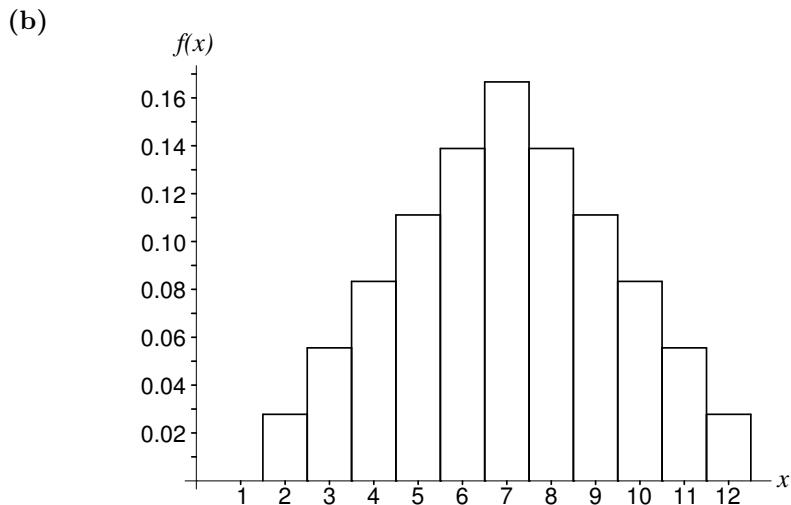
Figure 2.1–6: Probability histogram for the sum of a pair of dice

2.1-8 **(a)** The space of W is $S = \{0, 1, 2, 3, 4, 5, 6, 7\}$.

$$P(W = 0) = P(X = 0, Y = 0) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}, \text{ assuming independence.}$$

$$P(W = 1) = P(X = 0, Y = 1) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8},$$

$$P(W = 2) = P(X = 2, Y = 0) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8},$$

$$P(W = 3) = P(X = 2, Y = 1) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8},$$

$$P(W = 4) = P(X = 0, Y = 4) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8},$$

$$P(W = 5) = P(X = 0, Y = 5) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8},$$

$$P(W = 6) = P(X = 2, Y = 4) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8},$$

$$P(W = 7) = P(X = 2, Y = 5) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}.$$

That is, $f(w) = P(W = w) = \dfrac{1}{8}, \quad w \in S$.

(b)
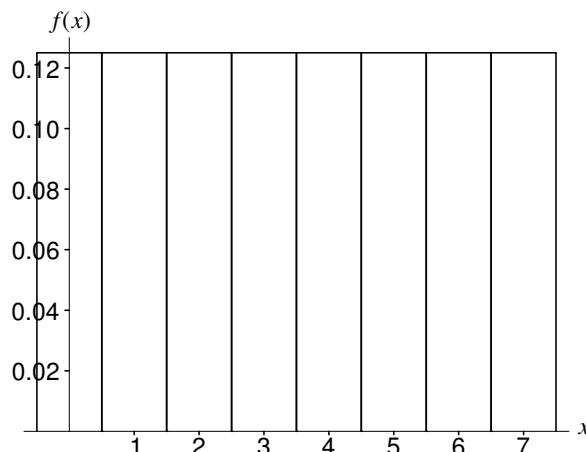
Figure 2.1–8: Probability histogram of sum of two special dice

2.1-10 **(a)** $\dfrac{\binom{3}{1}\binom{47}{9}}{\binom{50}{10}} = \dfrac{39}{98}$;

(b) $\displaystyle\sum_{x=0}^{1} \dfrac{\binom{3}{x}\binom{47}{10-x}}{\binom{50}{10}} = \dfrac{221}{245}$.

2.1-12 $OC(0.04) = \dfrac{\binom{1}{0}\binom{24}{5}}{\binom{25}{5}} + \dfrac{\binom{1}{1}\binom{24}{4}}{\binom{25}{5}} = 1.000;$

$\phantom{\textbf{2.1-12}}\ OC(0.08) = \dfrac{\binom{2}{0}\binom{23}{5}}{\binom{25}{5}} + \dfrac{\binom{2}{1}\binom{23}{4}}{\binom{25}{5}} = 0.967;$

$\phantom{\textbf{2.1-12}}\ OC(0.12) = \dfrac{\binom{3}{0}\binom{22}{5}}{\binom{25}{5}} + \dfrac{\binom{3}{1}\binom{22}{4}}{\binom{25}{5}} = 0.909;$

$\phantom{\textbf{2.1-12}}\ OC(0.16) = \dfrac{\binom{4}{0}\binom{21}{5}}{\binom{25}{5}} + \dfrac{\binom{4}{1}\binom{21}{4}}{\binom{25}{5}} = 0.834.$

**2.1-14** $P(X \geq 1) = 1 - P(X = 0) = 1 - \dfrac{\binom{3}{0}\binom{17}{5}}{\binom{20}{5}} = 1 - \dfrac{91}{228} = \dfrac{137}{228} = 0.60.$

2.1-16 **(a)** Let Y equal the number of H chips that are selected. Then
$X = |Y - (10 - Y)| = |2Y - 10|$ and the p.m.f. of Y is

$$g(y) = \frac{\binom{10}{y}\binom{10}{10-y}}{\binom{20}{10}}, \qquad y = 0, 1, \ldots, 10.$$

The p.m.f. of X is as follows:

$f(0) = g(5)$	$f(2) = 2g(6)$	$f(4) = 2g(7)$	$f(6) = 2g(8)$	$f(8) = 2g(9)$	$f(10) = 2g(10)$
$\dfrac{1}{184{,}756}$	$\dfrac{2025}{92{,}378}$	$\dfrac{22{,}050}{46{,}189}$	$\dfrac{22{,}050}{46{,}189}$	$\dfrac{2025}{92{,}378}$	$\dfrac{1}{92{,}378}$

(b) The mode is equal to 2.

2.1-18 **(a)** $P(2, 1, 6, 10)$ means that 2 is in position 1 so 1 cannot be selected. Thus

$$P(2, 1, 6, 10) = \frac{\binom{1}{0}\binom{1}{1}\binom{8}{5}}{\binom{10}{6}} = \frac{56}{210} = \frac{4}{15};$$

(b) $P(i, r, k, n) = \dfrac{\binom{i-1}{r-1}\binom{1}{1}\binom{n-i}{k-r}}{\binom{n}{k}}.$

2.2 Mathematical Expectation

2.2-2 $E(X) = (-1)\left(\dfrac{4}{9}\right) + (0)\left(\dfrac{1}{9}\right) + (1)\left(\dfrac{4}{9}\right) = 0;$

$E(X^2) = (-1)^2\left(\dfrac{4}{9}\right) + (0)^2\left(\dfrac{1}{9}\right) + (1)^2\left(\dfrac{4}{9}\right) = \dfrac{8}{9};$

$E(3X^2 - 2X + 4) = 3\left(\dfrac{8}{9}\right) - 2(0) + 4 = \dfrac{20}{3}.$

2.2-4 $E(X) = \$\,499(0.001) - \$\,1(0.999) = -\$\,0.50.$

2.2-6
$$1 \;=\; \sum_{x=0}^{6} f(x) = \frac{9}{10} + c\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6}\right)$$

$$c \;=\; \frac{2}{49};$$

$$E(\text{Payment}) \;=\; \frac{2}{49}\left(1\cdot\frac{1}{2} + 2\cdot\frac{1}{3} + 3\cdot\frac{1}{4} + 4\cdot\frac{1}{5} + 5\cdot\frac{1}{6}\right) = \frac{71}{490}\ \text{units.}$$

2.2-8 Note that $\displaystyle\sum_{x=1}^{\infty}\frac{6}{\pi^2 x^2} = \frac{6}{\pi^2}\sum_{x=1}^{\infty}\frac{1}{x^2} = \frac{6}{\pi^2}\frac{\pi^2}{6} = 1$, so this is a p.d.f.

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

and it is well known that the sum of this harmonic series is not finite.

2.2-10 $E(|X - c|) = \dfrac{1}{7}\displaystyle\sum_{x\in S}|x - c|$, where $S = \{1, 2, 3, 5, 15, 25, 50\}$.
When $c = 5$,

$$E(|X - 5|) = \frac{1}{7}\left[(5 - 1) + (5 - 2) + (5 - 3) + (5 - 5) + (15 - 5) + (25 - 5) + (50 - 5)\right].$$

If c is either increased or decreased by 1, this expectation is increased by $1/7$. Thus $c = 5$, the median, minimizes this expectation while $b = E(X) = \mu$, the mean, minimizes $E[(X - b)^2]$. You could also let $h(c) = E(|X - c|)$ and show that $h'(c) = 0$ when $c = 5$.

2.2-12 (1) $\cdot\dfrac{15}{36} + (-1)\cdot\dfrac{21}{36} = \dfrac{-6}{36} = \dfrac{-1}{6};$

$(1)\cdot\dfrac{15}{36} + (-1)\cdot\dfrac{21}{36} = \dfrac{-6}{36} = \dfrac{-1}{6};$

$(4)\cdot\dfrac{6}{36} + (-1)\cdot\dfrac{30}{36} = \dfrac{-6}{36} = \dfrac{-1}{6}.$

2.2-14 **(a)** The average class size is $\dfrac{(16)(25) + (3)(100) + (1)(300)}{20} = 50;$

(b)
$$f(x) = \begin{cases} 0.4, & x = 25, \\ 0.3, & x = 100, \\ 0.3, & x = 300, \end{cases}$$

(c) $E(X) = 25(0.4) + 100(0.3) + 300(0.3) = 130.$

2.3   The Mean, Variance, and Standard Deviation

2.3-2 (a)

$$\mu = E(X)$$

$$= \sum_{x=1}^{3} x \frac{3!}{x!\,(3-x)!} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{3-x}$$

$$= 3\left(\frac{1}{4}\right) \sum_{k=0}^{2} \frac{2!}{k!\,(2-k)!} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{2-k}$$

$$= 3\left(\frac{1}{4}\right)\left(\frac{1}{4} + \frac{3}{4}\right)^2 = \frac{3}{4};$$

$$E[X(X-1)] = \sum_{x=2}^{3} x(x-1) \frac{3!}{x!\,(3-x)!} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{3-x}$$

$$= 2(3)\left(\frac{1}{4}\right)^2 \frac{3}{4} + 6\left(\frac{1}{4}\right)^3$$

$$= 6\left(\frac{1}{4}\right)^2 = 2\left(\frac{1}{4}\right)\left(\frac{3}{4}\right);$$

$$\sigma^2 = E[X(X-1)] + E(X) - \mu^2$$

$$= (2)\left(\frac{3}{4}\right)\left(\frac{1}{4}\right) + \left(\frac{3}{4}\right) - \left(\frac{3}{4}\right)^2$$

$$= (2)\left(\frac{3}{4}\right)\left(\frac{1}{4}\right) + \left(\frac{3}{4}\right)\left(\frac{1}{4}\right) = 3\left(\frac{1}{4}\right)\left(\frac{3}{4}\right);$$

(b) $\mu = E(X)$

$$= \sum_{x=1}^{4} x \frac{4!}{x!\,(4-x)!} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{4-x}$$

$$= 4\left(\frac{1}{2}\right) \sum_{k=0}^{3} \frac{3!}{k!\,(3-k)!} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{3-k}$$

$$= 4\left(\frac{1}{2}\right)\left(\frac{1}{2} + \frac{1}{2}\right)^3 = 2;$$

$$E[X(X-1)] = \sum_{x=2}^{4} x(x-1) \frac{4!}{x!\,(4-x)!} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{4-x}$$

$$= 2(6)\left(\frac{1}{2}\right)^4 + (6)(4)\left(\frac{1}{2}\right)^4 + (12)\left(\frac{1}{2}\right)^4$$

$$= 48\left(\frac{1}{2}\right)^4 = 12\left(\frac{1}{2}\right)^2;$$

$$\sigma^2 = (12)\left(\frac{1}{2}\right)^2 + \frac{4}{2} - \left(\frac{4}{2}\right)^2 = 1.$$

2.3-4 $E[(X-\mu)/\sigma] = (1/\sigma)[E(X) - \mu] = (1/\sigma)(\mu - \mu) = 0;$

$E\{[(X-\mu)/\sigma]^2\} = (1/\sigma^2)E[(X-\mu)^2] = (1/\sigma^2)(\sigma^2) = 1.$

2.3-6 $f(1) = \dfrac{3}{8}, \; f(2) = \dfrac{2}{8}, \; f(3) = \dfrac{3}{8}$

$\mu = 1 \cdot \dfrac{3}{8} + 2 \cdot \dfrac{2}{8} + 3 \cdot \dfrac{3}{8} = 2,$

$\sigma^2 = 1^2 \cdot \dfrac{3}{8} + 2^2 \cdot \dfrac{2}{8} + 3^2 \cdot \dfrac{3}{8} - 2^2 = \dfrac{3}{4}.$

2.3-8 **(a)** $\overline{x} = \dfrac{4}{3} = 1.333;$

(b) $s^2 = \dfrac{88}{69} = 1.275.$

2.3-10 **(a)** $[3, 19, 16, 9];$

(b) $\overline{x} = \dfrac{125}{47} = 2.66, \, s = 0.87;$

(c)
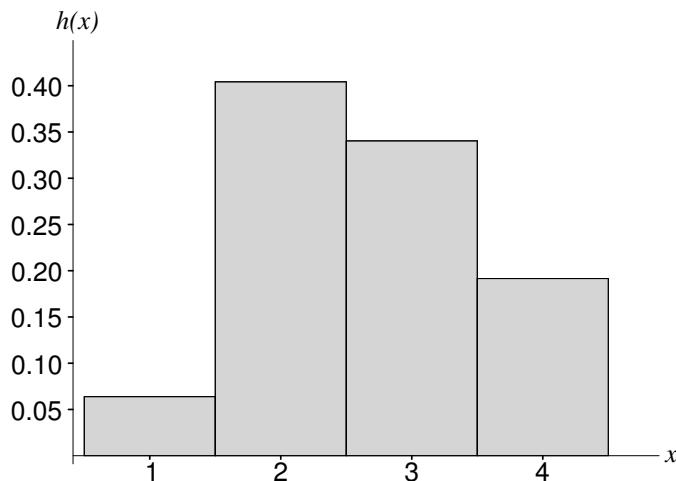
Figure 2.3–10: Number of pets

2.3-12 $\overline{x} = \dfrac{409}{50} = 8.18.$

2.3-14 **(a)** $f(x) = P(X = x) = \dfrac{\dbinom{6}{x}\dbinom{43}{6-x}}{\dbinom{49}{6}}, \qquad x = 0, 1, 2, 3, 4, 5, 6;$

(b) $\mu_X \;=\; \displaystyle\sum_{x=0}^{6} x f(x) = \dfrac{36}{49} = 0.7347,$

$\sigma_X^2 \;=\; \displaystyle\sum_{x=0}^{6} (x - \mu)^2 f(x) = \dfrac{5{,}547}{9{,}604} = 0.5776;$

$\sigma_X \;=\; \dfrac{43}{98}\sqrt{3} = 0.7600;$

(c) $f(0) = \dfrac{435{,}461}{998{,}844} > \dfrac{412{,}542}{998{,}844} = f(1); \quad X = 0$ is most likely to occur.

(d) The numbers are reasonable because

$$(25{,}000{,}000)f(6) = 1.79;$$

$$(25{,}000{,}000)f(5) = 461.25;$$

$$(25{,}000{,}000)f(4) = 24{,}215.49;$$

(e) The respective expected values, $(138)f(x)$, for $x = 0, 1, 2, 3$, are 60.16, 57.00, 18.27, and 2.44, so the results are reasonable. See Figure 2.3-14 for a comparison of the theoretical probability histogram and the histogram of the data.
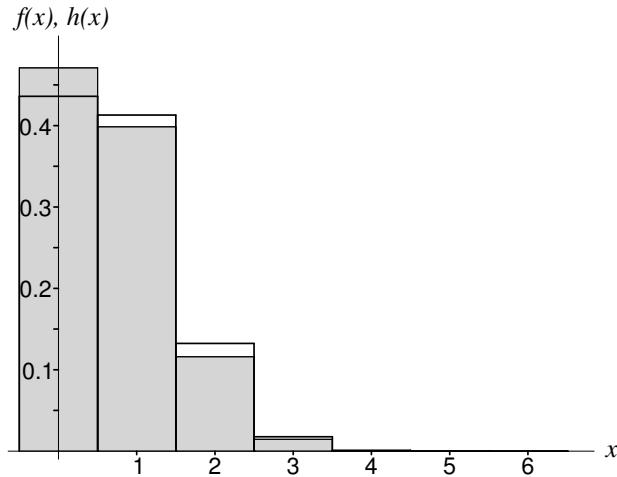
Figure 2.3–14: Empirical (shaded) and theoretical histograms for LOTTO

2.3-16 **(a)** Out of the 75 numbers, first select $x - 1$ of which 23 are selected out of the 24 good numbers on your card and the remaining $x - 1 - 23$ are selected out of the 51 bad numbers. There is now one good number to be selected out of the remaining $75 - (x - 1)$.

(b) The mode is 75.

(c) $\mu = \dfrac{1824}{25} = 72.96.$

(d) $E[X(X+1)] = \dfrac{70{,}224}{13} = 5{,}401.846154.$

(e) $\sigma^2 = \dfrac{46{,}512}{8{,}125} = 5.724554;\ \ \sigma = 2.3926.$

(f) **(i)** $\overline{x} = 72.78$, **(ii)** $s^2 = 8.7187879$, **(iii)** $s = 2.9528$, **(iv)** 5378.34.

(g)
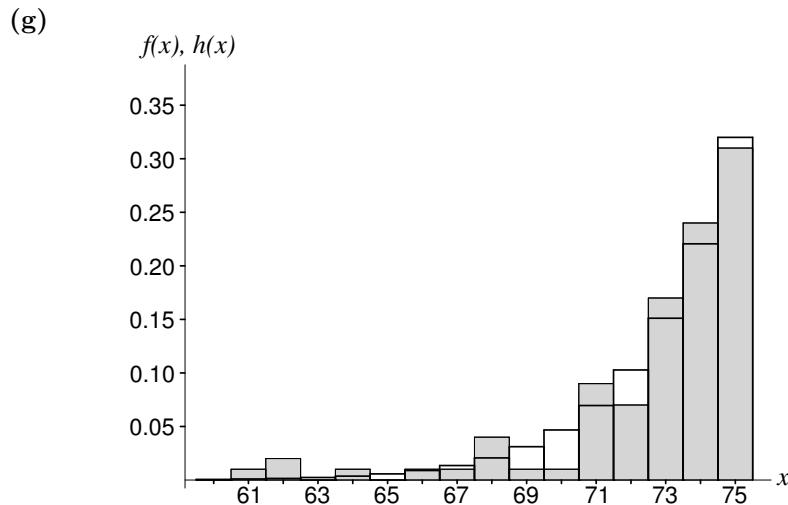
Figure 2.3–16: Bingo "cover-up" comparisons

2.3-18 **(a)** $P(X \geq 1) = \dfrac{2^1}{\dbinom{3}{1}} = \dfrac{2}{3};$

(b) $\displaystyle\sum_{k=1}^{5} P(X \geq k) = P(X = 1) + 2P(X = 2) + \cdots + 5P(X = 5) = \mu;$

(c) $\mu = \dfrac{5{,}168}{3{,}465} = 1.49149;$

(d) In the limit, $\mu = \dfrac{\pi}{2}.$

2.4 Bernoulli Trials and the Binomial Distribution

2.4-2 $f(-1) = \dfrac{11}{18}, \ f(1) = \dfrac{7}{18};$

$\mu = (-1)\dfrac{11}{18} + (1)\dfrac{7}{18} = -\dfrac{4}{18};$

$\sigma^2 = \left(-1 + \dfrac{4}{18}\right)^2 \left(\dfrac{11}{18}\right) + \left(1 + \dfrac{4}{18}\right)^2 \left(\dfrac{7}{18}\right) = \dfrac{77}{81}.$

2.4-4 **(a)** $P(X \leq 5) = 0.6652;$

(b) $P(X \geq 6) = 1 - P(X \leq 5) = 0.3348;$

(c) $P(X \leq 7) - P(X \leq 6) = 0.9427 - 0.8418 = 0.1009;$

(d) $\mu = (12)(0.40) = 4.8, \ \sigma^2 = (12)(0.40)(0.60) = 2.88, \ \ \sigma = \sqrt{2.88} = 1.697.$

2.4-6 **(a)** X is $b(7, \ 0.15);$

(b) **(i)** $P(X \geq 2) \ = \ 1 - P(X \leq 1) = 1 - 0.7166 = 0.2834;$

(ii) $P(X = 1) \ = \ P(X \leq 1) - P(X \leq 0) = 0.7166 - 0.3206 = 0.3960;$

(iii) $P(X \leq 3) \ = \ 0.9879.$

2.4-8 **(a)** $P(X \geq 10) = P(15 - X \leq 5) = 0.5643;$

(b) $P(X \leq 10) = P(15 - X \geq 5) = 1 - P(15 - X \leq 4) = 1 - 0.3519 = 0.6481;$

(c) $P(X = 10) = P(X \geq 10) - P(X \geq 11)$
$$= P(15 - X \leq 5) - P(15 - X \leq 4) = 0.5643 - 0.3519 = 0.2124;$$

(d) X is $b(15, 0.65)$, $15 - X)$ is $b(15, 0.35);$

(e) $\mu = (15)(0.65) = 9.75, \sigma^2 = (15)(0.65)(0.35) = 3.4125; \sigma = \sqrt{3.4125} = 1.847.$

2.4-10 **(a)** $1 - 0.01^4 = 0.99999999;$ **(b)** $0.99^4 = 0.960596.$

2.4-12 **(a)** X is $b(8, \ 0.90);$

(b)   **(i)** $P(X = 8)$ $=$ $P(8 - X = 0) = 0.4305;$

(ii) $P(X \leq 6)$ $=$ $P(8 - X \geq 2)$

$=$ $1 - P(8 - X \leq 1) = 1 - 0.8131 = 0.1869;$

(iii) $P(X \geq 6)$ $=$ $P(8 - X \leq 2) = 0.9619.$

2.4-14 **(a)**
$$f(x) = \begin{cases} 125/216, & x = -1, \\ 75/216, & x = 1, \\ 15/216, & x = 2, \\ 1/216, & x = 3; \end{cases}$$

(b) $\mu \ = \ (-1) \cdot \dfrac{125}{216} + (1) \cdot \dfrac{75}{216} + (2) \cdot \dfrac{15}{216} + (3) \cdot \dfrac{1}{216} = -\dfrac{17}{216};$

$\sigma^2 \ = \ E(X^2) - \mu^2 = \dfrac{269}{216} - \left(-\dfrac{17}{216}\right)^2 = 1.2392;$

$\sigma \ = \ 1.11;$

(c) See Figure 2.4-14.

(d) $\overline{x} \ = \ \dfrac{-1}{100} = -0.01;$

$s^2 \ = \ \dfrac{100(129) - (-1)^2}{100(99)} = 1.3029;$

$s \ = \ 1.14.$

(e)

f(x), h(x)
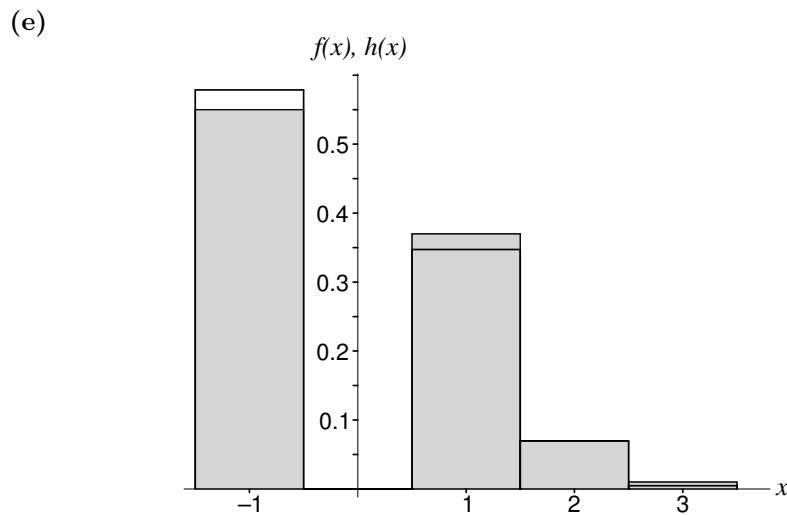
Figure 2.4–14: Losses in chuck-a-luck

2.4-16 Let X equal the number of winning tickets when n tickets are purchased. Then

$$P(X \geq 1) \;=\; 1 - P(X = 0)$$
$$=\; 1 - \left(\frac{9}{10}\right)^n.$$

(a)
$$1 - (0.9)^n \;=\; 0.50$$
$$(0.9)^n \;=\; 0.50$$
$$n \ln 0.9 \;=\; \ln 0.5$$
$$n \;=\; \frac{\ln 0.5}{\ln 0.9} \;=\; 6.58$$
so $n = 7$.

(b)
$$1 - (0.9)^n \;=\; 0.95$$
$$(0.9)^n \;=\; 0.05$$
$$n \;=\; \frac{\ln 0.05}{\ln 0.09} \;=\; 28.43$$
so $n = 29$.

2.4-18 $\dfrac{(0.1)(1 - 0.95^5)}{(0.4)(1 - 0.97^5) + (0.5)(1 - 0.98^5) + (0.1)(1 - 0.95^5)} = 0.178.$

2.4-20 It is given that X is $b(10,\ 0.10)$. We are to find M so that

$P(1000X \leq M) \geq 0.99)$ or $P(X \leq M/1000) \geq 0.99$. From Appendix Table II,

$P(X \leq 4) = 0.9984 > 0.99$. Thus $M/1000 = 4$ or $M = 4000$ dollars.

2.4-22 X is $b(5,\ 0.05)$. The expected number of tests is

$1\,P(X = 0) + 6\,P(X > 0) = 1\,(0.7738) + 6\,(1 - 0.7738) = 2.131.$

2.5   The Moment-Generating Function

2.5-2 **(a)** **(i)** $b(5,\ 0.7)$; **(ii)** $\mu = 3.5, \sigma^2 = 1.05$; **(iii)** 0.1607;

      (b) **(i)** geometric, $p = 0.3$; **(ii)** $\mu = 10/3, \sigma^2 = 70/9$; **(iii)** 0.51;

      (c) **(i)** Bernoulli, $p = 0.55$; **(ii)** $\mu = 0.55, \sigma^2 = 0.2475$; **(iii)** 0.55;

      (d) **(ii)** $\mu = 2.1, \sigma^2 = 0.89$; **(iii)** 0.7;

      (e) **(i)** negative binomial, $p = 0.6, r = 2$; **(ii)** $10/3,\ \sigma^2 = 20/9$; **(iii)** 0.36;

      (f) **(i)** discrete uniform on $1, 2, \ldots, 10$; **(ii)** $5.5,\ 8.25$; **(iii)** 0.2.

2.5-4 **(a)** $f(x) = \left(\dfrac{364}{365}\right)^{x-1}\left(\dfrac{1}{365}\right), \qquad x = 1, 2, 3, \ldots,$

      (b) $\quad \mu \quad = \quad \dfrac{1}{\dfrac{1}{365}} = 365,$

$$\sigma^2 \quad = \quad \dfrac{\dfrac{364}{365}}{\left(\dfrac{1}{365}\right)^2} = 132{,}860,$$

$$\sigma \quad = \quad 364.500;$$

      (c) $\quad P(X > 400) \quad = \quad \left(\dfrac{364}{365}\right)^{400} = 0.3337,$

$$P(X < 300) \quad = \quad 1 - \left(\dfrac{364}{365}\right)^{299} = 0.5597.$$

2.5-6 $P(X \geq 100) = P(X > 99) = (0.99)^{99} = 0.3697.$

2.5-8 $\dbinom{10-1}{5-1}\left(\dfrac{1}{2}\right)^5\left(\dfrac{1}{2}\right)^5 = \dfrac{126}{1024} = \dfrac{63}{512}.$

2.5-10 **(a)** Negative binomial with $r = 10, p = 0.6$ so

$$\mu = \dfrac{10}{0.60} = 16.667, \ \sigma^2 = \dfrac{10(0.40)}{(0.60)^2} = 11.111, \ \sigma = 3.333;$$

      (b) $P(X = 16) = \dbinom{15}{9}(0.60)^{10}(0.40)^6 = 0.1240.$

2.5-12 $P(X > k + j \mid X > k) \quad = \quad \dfrac{P(X > k + j)}{P(X > k)}$

$$= \quad \dfrac{q^{k+j}}{q^k} = q^j = P(X > j).$$

2.5-14 **(b)** $\displaystyle\sum_{x=2}^{\infty} f(x) \quad = \quad \sum_{x=2}^{\infty} \dfrac{1}{\sqrt{5}}\left[\left(\dfrac{1+\sqrt{5}}{2}\right)^{x-1} - \left(\dfrac{1-\sqrt{5}}{2}\right)^{x-1}\right]\left(\dfrac{1}{2^x}\right)$

$$= \quad \dfrac{2}{\sqrt{5}(1+\sqrt{5})}\sum_{x=2}^{\infty}\dfrac{(1+\sqrt{5})^x}{4^x} - \dfrac{2}{\sqrt{5}(1-\sqrt{5})}\sum_{x=2}^{\infty}\dfrac{(1-\sqrt{5})^x}{4^x}$$

$$= \quad \text{(you fill in missing steps)}$$

$$= \quad 1;$$

(c) $E(X) = \sum_{x=2}^{\infty} \frac{x}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2} \right)^{x-1} - \left(\frac{1-\sqrt{5}}{2} \right)^{x-1} \right] \left(\frac{1}{2^x} \right)$

$= \frac{1}{2\sqrt{5}} \sum_{x=1}^{\infty} \left[x \left(\frac{1+\sqrt{5}}{4} \right)^{x-1} - x \left(\frac{1-\sqrt{5}}{4} \right)^{x-1} \right]$

$= \frac{1}{2\sqrt{5}} \left[\frac{1}{(1-(1+\sqrt{5})/4)^2} - \frac{1}{(1-(1-\sqrt{5}/4)^2} \right]$

$=$ (you fill in missing steps)

$= 6;$

(d) $E[X(X-1)] = \sum_{x=2}^{\infty} x(x-1) \frac{1}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2} \right)^{x-1} - \left(\frac{1-\sqrt{5}}{2} \right)^{x-1} \right] \left(\frac{1}{2^x} \right)$

$= \frac{1}{2\sqrt{5}} \sum_{x=2}^{\infty} x(x-1) \left[\left(\frac{1+\sqrt{5}}{4} \right)^{x-1} - \left(\frac{1-\sqrt{5}}{4} \right)^{x-1} \right]$

$= \frac{1}{2\sqrt{5}} \left[\frac{1+\sqrt{5}}{4} \sum_{x=2}^{\infty} x(x-1) \left(\frac{1+\sqrt{5}}{4} \right)^{x-2} - \right.$

$\left. \frac{1-\sqrt{5}}{4} \sum_{x=2}^{\infty} x(x-1) \left(\frac{1-\sqrt{5}}{4} \right)^{x-2} \right]$

$= \frac{1}{2\sqrt{5}} \left[\frac{2 \left(\frac{1+\sqrt{5}}{4} \right)}{\left(1 - \frac{1+\sqrt{5}}{4} \right)^3} - \frac{2 \left(\frac{1-\sqrt{5}}{4} \right)}{\left(1 - \frac{1-\sqrt{5}}{4} \right)^3} \right]$

$=$ (you fill in missing steps)

$= 52;$

$\sigma^2 = E[X(X-1)] + E(X) - \mu^2$
$= 52 + 6 - 36$
$= 22;$

$\sigma = \sqrt{22} = 4.690.$

(e) (i) $P(X \le 3) = \frac{1}{4} + \frac{1}{8} = \frac{3}{8},$

(ii) $P(X \le 5) = 1 - P(X \le 4) = 1 - \frac{1}{4} - \frac{1}{8} - \frac{1}{8} = \frac{1}{2},$

(iii) $P(X = 3) = \frac{1}{8}.$

(f) A simulation question.

2.5-16 Let "being missed" be a success and let X equal the number of trials until the first success. Then $p = 0.01$.

$P(X \le 50) = 1 - 0.99^{50} = 1 - 0.605 = 0.395.$

2.5-18 $M(t) = 1 + \frac{5t}{1!} + \frac{5t^2}{2!} + \frac{5t^3}{3!} + \cdots = e^{5t},$

$f(x) = 1, \quad x = 5.$

2.5-20 (a) $R(t) = \ln(1 - p + pe^t),$

$$R'(t) = \left[\frac{pe^t}{1-p+pe^t}\right]_{t=0} = p,$$

$$R''(t) = \left[\frac{(1-p+pe^t)(pe^t)-(pe^t)(pe^t)}{(1-p+pe^t)^2}\right]_{t=0} = p(1-p);$$

(b) $R(t) = n\ln(1-p+pe^t),$

$$R'(t) = \left[\frac{npe^t}{1-p+pe^t}\right]_{t=0} = np,$$

$$R''(t) = n\left[\frac{(1-p+pe^t)(pe^t)-(pe^t)(pe^t)}{(1-p+pe^t)^2}\right]_{t=0} = np(1-p);$$

(c) $R(t) = \ln p + t - \ln[1-(1-p)e^t],$

$$R'(t) = \left[1+\frac{(1-p)e^t}{1-(1-p)e^t}\right]_{t=0} = 1+\frac{1-p}{p} = \frac{1}{p},$$

$$R''(t) = \left[(-1)\{1-(1-p)e^t\}^2\{-(1-p)e^t\}\right]_{t=0} = \frac{1-p}{p};$$

(d) $R(t) = r\left[\ln p + t - \ln\{1-(1-p)e^t\}\right],$

$$R'(t) = r\left[\frac{1}{1-(1-p)e^t}\right]_{t=0} = \frac{r}{p},$$

$$R''(t) = r\left[(-1)\{1-(1-p)e^t\}^{-2}\{-(1-p)e^t\}\right]_{t=0} = \frac{r(1-p)}{p^2}.$$

2.5-22 $(0.7)(0.7)(0.3) = 0.147.$

2.6 The Poisson Distribution

2.6-2 $\lambda = \mu = \sigma^2 = 3$ so $P(X=2) = 0.423 - 0.199 = 0.224.$

2.6-4
$$3\frac{\lambda^1 e^{-\lambda}}{1!} = \frac{\lambda^2 e^{-\lambda}}{2!}$$
$$e^{-\lambda}\lambda(\lambda - 6) = 0$$
$$\lambda = 6$$
Thus $P(X=4) = 0.285 - 0.151 = 0.134.$

2.6-6 $\lambda = (1)(50/100) = 0.5$, so $P(X=0) = e^{-0.5}/0! = 0.607.$

2.6-8 $np = 1000(0.005) = 5;$

(a) $P(X \le 1) \approx 0.040;$

(b) $P(X=4,5,6) = P(X\le 6) - P(X\le 3) \approx 0.762 - 0.265 = 0.497.$

2.6-10 $\sigma = \sqrt{9} = 3,$

$P(3 < X < 15) = P(X\le 14) - P(X\le 3) = 0.959 - 0.021 = 0.938.$

2.6-12 **(a)** $[17, 47, 63, 63, 49, 28, 21, 11, 1];$

(b) $\bar{x} = 303/100 = 3.03,\quad s^2 = 4,141/1,300 = 3.193,$ yes;

(c)
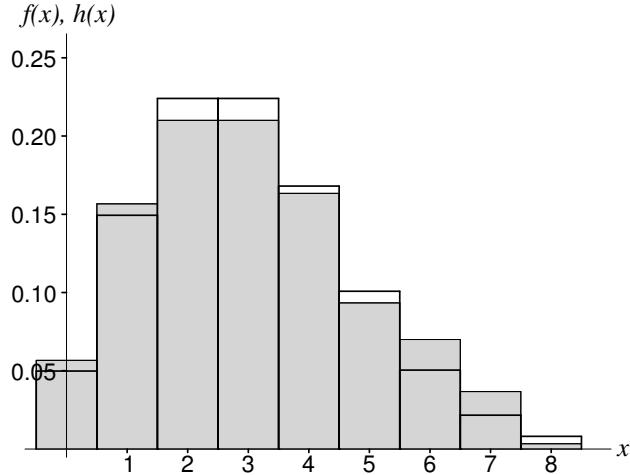
Figure 2.6–12: Background radiation

(d) The fit is very good and the Poisson distribution seems to provide an excellent probability model.

2.6-14 **(a)**
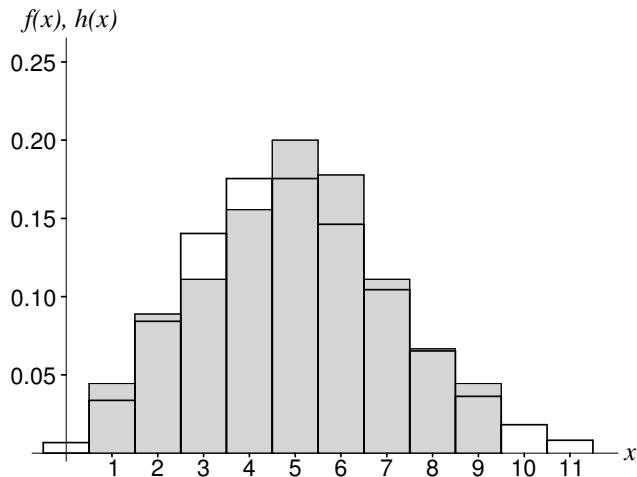
Figure 2.6–14: Green peanut m&m's

(b) The fit is quite good. Also $\bar{x} = 4.956$ and $s^2 = 4.134$ are close to each other.

2.6-16 $OC(p) = P(X \leq 3) \approx \sum_{x=0}^{3} \frac{(400p)^x e^{-400p}}{x!};$

$$OC(0.002) \approx 0.991;$$
$$OC(0.004) \approx 0.921;$$
$$OC(0.006) \approx 0.779;$$
$$OC(0.01) \approx 0.433;$$
$$OC(0.02) \approx 0.042.$$
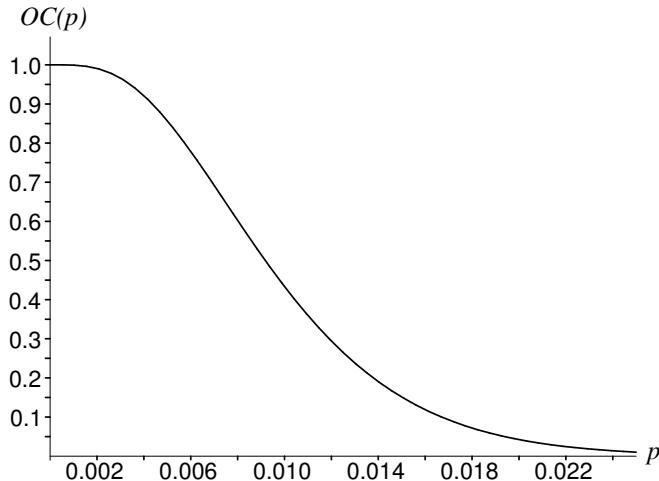
Figure 2.6–16: Operating characteristic curve

2.6-18 Since $E(X) = 0.2$, the expected loss is $(0.02)(\$10,000) = \$2,000$.

2.6-20

$$\frac{\lambda^2 e^{-\lambda}}{2!} = 4 \cdot \frac{\lambda^3 e^{-\lambda}}{3!}$$
$$\lambda^2 e^{-\lambda}[(4/3)\lambda - 1] = 0$$
$$\lambda = 3/4$$
$$\sigma^2 = E(X^2) - \mu^2$$
$$\frac{3}{4} = E(X^2) - \left(\frac{3}{4}\right)^2$$
$$E(X^2) = \frac{9}{16} + \frac{12}{16} = \frac{21}{16}.$$

2.6-22 Using Minitab, **(a)** $\bar{x} = 56.286$, **(b)** $s^2 = 56.205$.

Chapter 3

Continuous Distributions

3.1 Continuous-Type Data

3.1–2 $\bar{x} = 3.58$; $s = 0.5116$.

3.1–4 **(a)** $\bar{x} = 5.833$, $s = 1.661$;

(b) The respective class frequencies are 4, 10, 15, 29, 20, 13, 3, 5, 1;

(c)
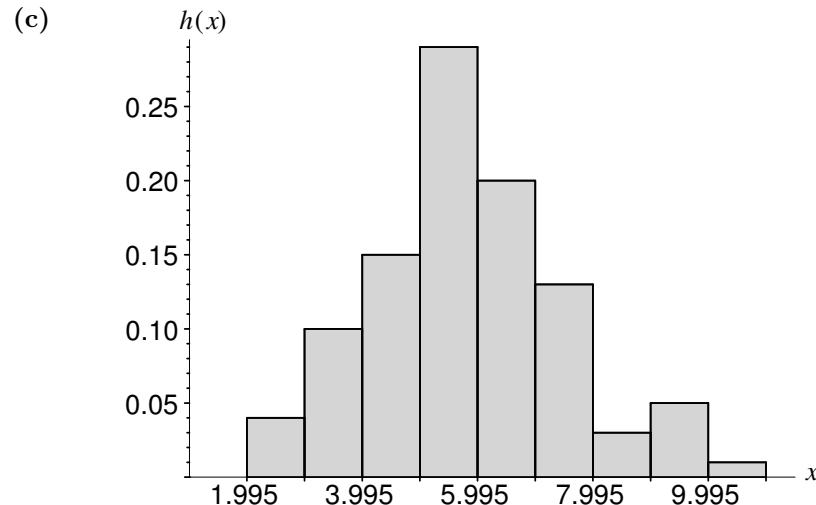
Figure 3.1–4: Weights of laptop computers

3.1–6 (a) The respective class frequencies are 2, 8, 15, 13, 5, 6, 1;
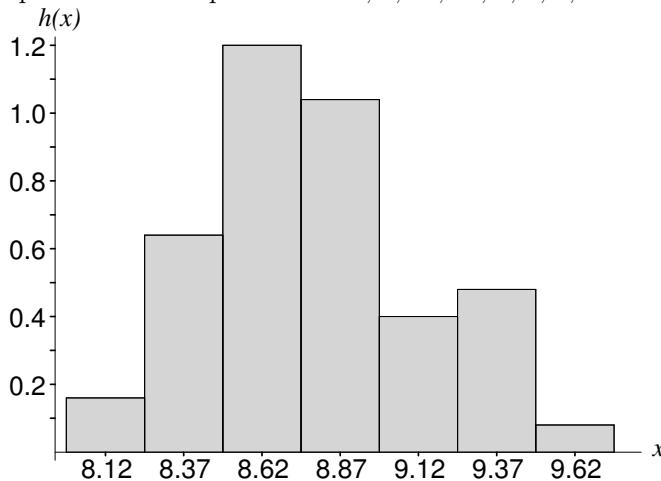
Figure 3.1–6: Weights of nails

(b) $\overline{x} = 8.773$, $\overline{u} = 8.785$, $s_x = 0.365$, $s_u = 0.352$;

(c) $800 * \overline{u} = 7028$, $800 * (\overline{u} + 2 * s_u) = 7591.2$. The answer depends on the cost of the nails as well as the time and distance required if too few nails are purchased.

3.1–8 (a)

Class Interval	Class Limits	Frequency f_i	Class Mark, u_i
(303.5, 307.5)	(304, 307)	1	305.5
(307.5, 311.5)	(308, 311)	5	309.5
(311.5, 315.5)	(312, 315)	6	313.5
(315.5, 319.5)	(316, 319)	10	317.5
(319.5, 323.5)	(320, 323)	11	321.5
(323.5, 327.5)	(324, 327)	9	325.5
(327.5, 331.5)	(328, 331)	7	329.5
(331.5, 335.5)	(332, 335)	1	333.5

(b) $\overline{x} = 320.1$, $s = 6.7499$;

(c)
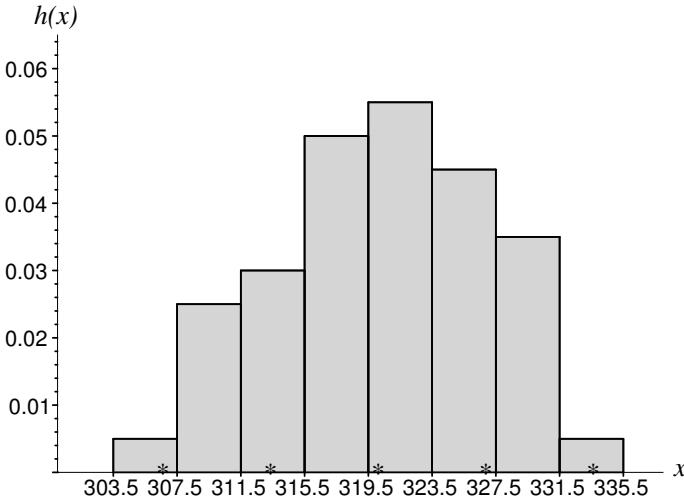
Figure 3.1–8: Melting points of metal alloys

There are 31 observations within one standard deviation of the mean (62%) and 48 observations within two standard deviations of the mean (96%).

3.1–10 **(a)** With the class boundaries $0.5, 5.5, 17.5, 38.5, 163.5, 549.5$, the respective frequencies are $11, 9, 10, 10, 10$.

(b)
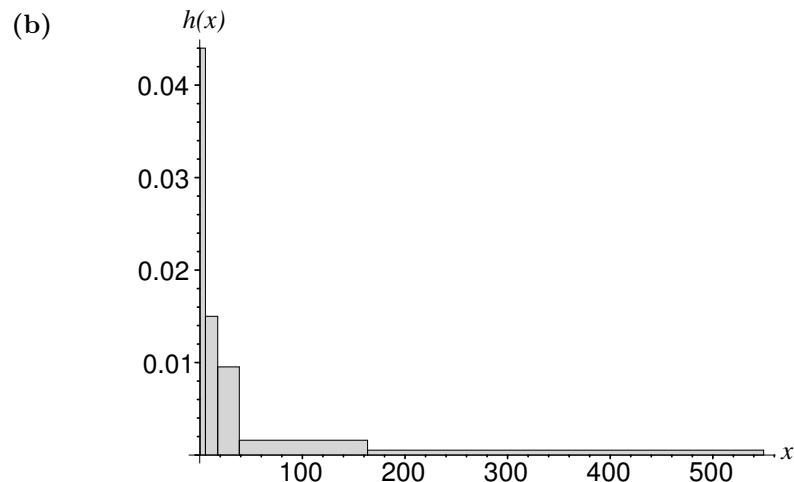
Figure 3.1–10: Mobil home losses

(c) This is a skewed to the right distribution.

3.1–12 **(a)** With the class boundaries $3.5005, 3.5505, 3.6005, \dots, 4.1005$, the respective class frequencies are $4, 7, 24, 23, 7, 4, 3, 9, 15, 23, 18, 2$.

(b)
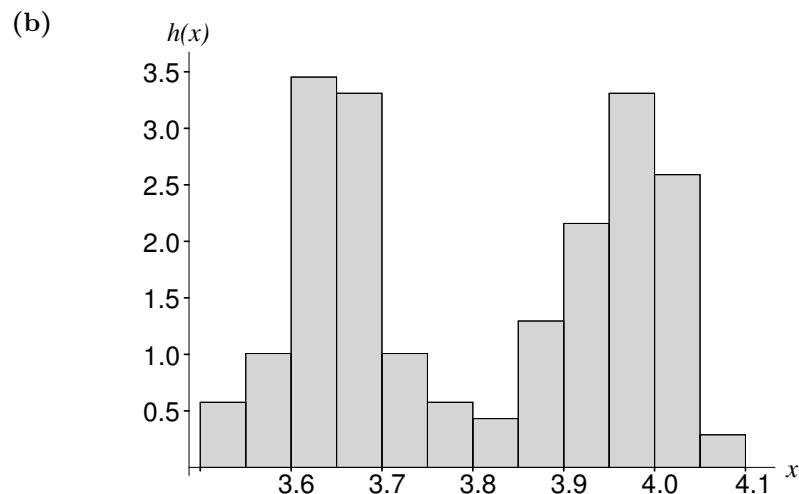
Figure 3.1–12: Weights of mirror parts

(c) This is a bimodal histogram.

3.2   Exploratory Data Analysis

3.2–2 (a)

Stems	Leaves	Freq	Depths
2	20 69 69 69	4	4
3	13 50 50 57 72 90 90 90 90 90	10	14
4	00 20 30 40 60 60 60 77 77 85 90 90 90 90 90	15	29
5	11 12 20 20 20 20 20 20 20 20 20 21 33 33 33 33 33 38 38 40 50 54 58 60 60 73 73 90 96	29	(29)
6	00 06 10 17 20 20 27 28 40 50 50 50 50 50 50 50 51 60 60 80 80	20	42
7	07 10 60 70 85 85 90 90 90 90 90 97 97 97	13	22
8	10 20 60	3	9
9	00 38 38 40 50	5	6
10	10	1	1

(Multiply numbers by 10^{-2}.)

(b) The five-number summary is: 2.20, 4.90, 5.52, 6.60, 10.10.
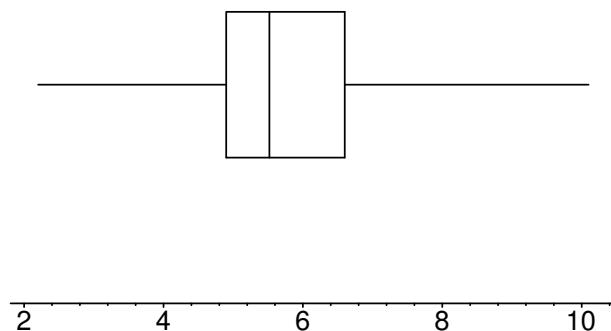
Figure 3.2–2: Box-and-whisker diagram of computer weights

3.2–4 (a) The respective frequencies for the men: 2, 7, 8, 15, 16, 13, 15, 14, 15, 8, 3, 3, 1, 3, 2.
The respective frequencies for the women: 1, 7, 15, 12, 16, 10, 6, 5, 3, 1.

(b)

Men's times

Women's times
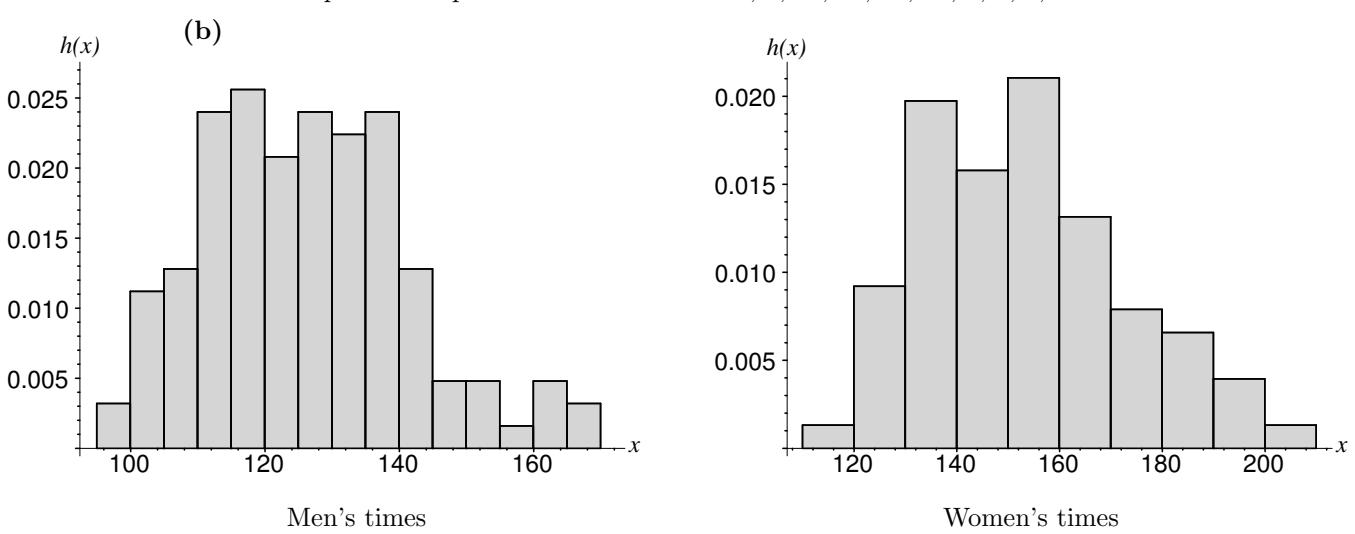
Figure 3.2–4: **(b)** Times for the Fifth Third River Bank Run

(c)

Male Times	Stems	Female Times
84 64	9●	
45 40 32 16 15 14 04	10∗	
97 95 95 88 62 60 52 50	10●	
46 45 37 33 32 32 29 29 28 26 23 19 18 09 08	11∗	
99 95 92 87 85 85 82 82 82 78 69 66 62 57 57 55	11●	81
49 48 41 41 38 30 30 28 23 23 12 03 03	12∗	38
97 94 92 84 80 80 74 74 67 65 62 62 53 53 52	12●	52 53 59 69 84 93
49 47 41 39 30 29 25 20 14 11 06 05 01 00	13∗	01 14 17 22 30 33 34 34 35 43
99 96 87 82 80 78 75 72 72 69 69 69 65 57 51	13●	70 71 73 85 98
46 42 38 31 25 13 01 01	14∗	09 29 38
82 71 57	14●	51 51 55 66 67 81 88 89 98
20 17 13	15∗	01 02 06 08 09 11 14 23 26 29
62	15●	56 70 96 97 98 99
25 12 07	16∗	00 13 22 36
99 67	16●	55 62 81 86 92 99
	17∗	09 12 32 42
	17●	86 88
	18∗	05 31
	18●	61 65 98
	19∗	25
	19●	79 98
	20∗	39

Multiply numbers by 10^{-1}

Table 3.2–4: Back-to-Back Stem-and-Leaf Diagram of Times for the Fifth Third River Bank Run

(d) Five-number summary for the male times: 96.35, 114.55, 125.25, 136.86, 169.90.

Five-number summary for the female times: 118.05, 137.01, 150.72, 167.6325, 203.92.
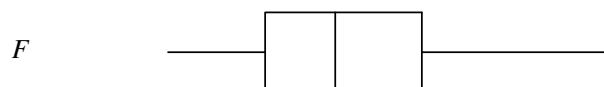
Figure 3.2–4: **(d)** Box-and-whisker diagrams of male and female times

3.2–6 **(a)** The five-number summary is: $\min = 1$, $\widetilde{q}_1 = 6.75$, $\widetilde{m} = 32$, $\widetilde{q}_3 = 90.75$, $\max = 527$.
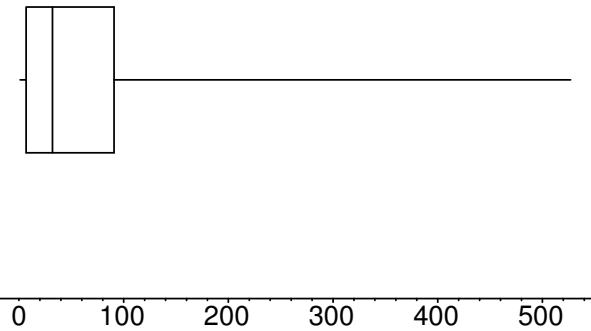
Figure 3.2–6: **(a)** Box-and-whisker diagram of mobile home losses

(b) IQR $= 90.75 - 6.75 = 84$. The inner fence is at 216.75 and the outer fence is at 342.75.

(c)
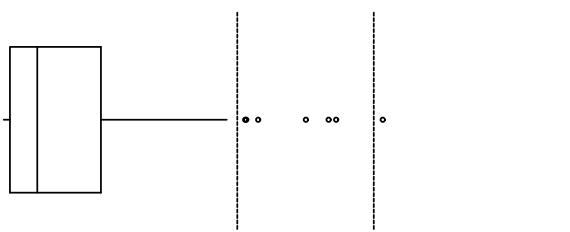
Figure 3.2–6: **(c)** Box-and-whisker diagram of losses with fences and outliers

3.2–8 **(a)**

Stems	Leaves	Freq	Depths
0•	5555555555555555555556666666666666677777778888888999999	53	(53)
1∗	000000111111222334	19	47
1•	5555666677889	13	28
2∗	0111133444	10	15
2•	5	1	5
3∗	4	1	4
3•	5	1	3
4∗		0	2
4•	5	1	2
5∗		0	1
5•	5	1	1

(b) The five-number summary is: $\min = 5$, $\widetilde{q}_1 = 6$, $\widetilde{m} = 9$, $\widetilde{q}_3 = 15$, $\max = 55$.
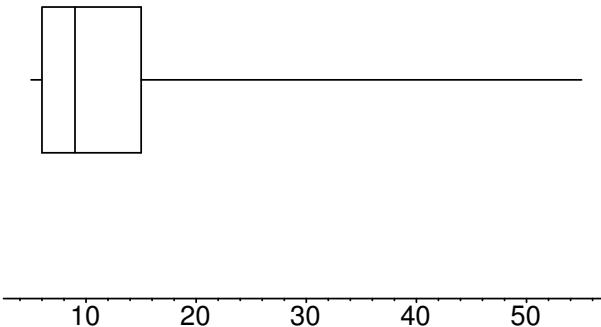
Figure 3.2–8: **(b)** Box-and-whisker diagram of maximum capital

(c) IQR $= 15 - 6 = 9$. The inner fence is at 28.5 and the outer fence is at 42.

(d)
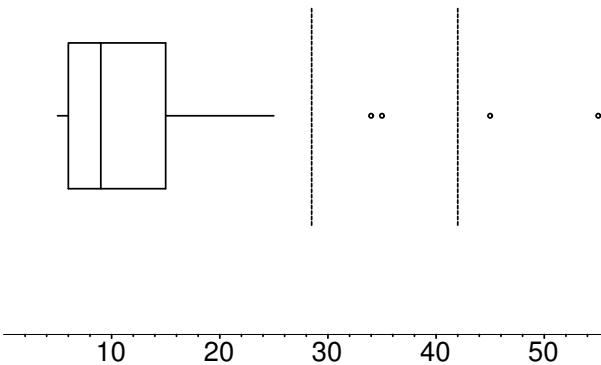
Figure 3.2–8: **(d)** Box-and-whisker diagram of maximum capital with outliers and fences

(e) The 90th percentile is 22.8.

3.2–10 **(a)**

Stems	Leaves	Frequency	Depths
101	7	1	1
102	0 0 0	3	4
103		0	4
104		0	4
105	8 9	2	6
106	1 3 3 6 6 7 7 8 8	9	(9)
107	3 7 9	3	10
108	8	1	7
109	1 3 9	3	6
110	0 2 2	3	3

(Multiply numbers by 10^{-1}.)

Table 3.2–10: Ordered stem-and-leaf diagram of weights of indicator housings

(b)
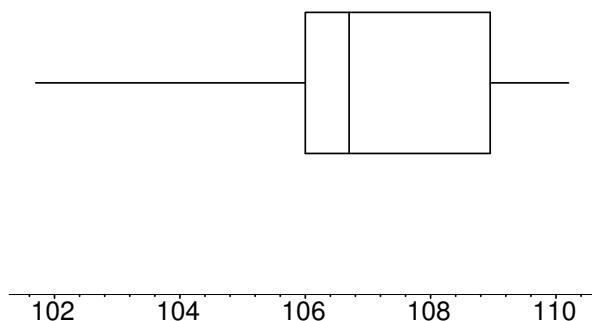
Figure 3.2–10: Weights of indicator housings

min $= 101.7$, $\widetilde{q}_1 = 106.0$, $\widetilde{m} = 106.7$, $\widetilde{q}_3 = 108.95$, max $= 110.2$;

(c) The interquartile range in IQR $= 108.95 - 106.0 = 2.95$. The inner fence is located at $106.7 - 1.5(2.95) = 102.275$ so there are four suspected outliers.

3.2–12 **(a)** With the class boundaries $2.85, 3.85, \ldots, 16.85$ the respective frequencies are 1, 0, 2, 4, 1, 14, 20, 11, 4, 5, 0, 1, 0, 1.

(b)
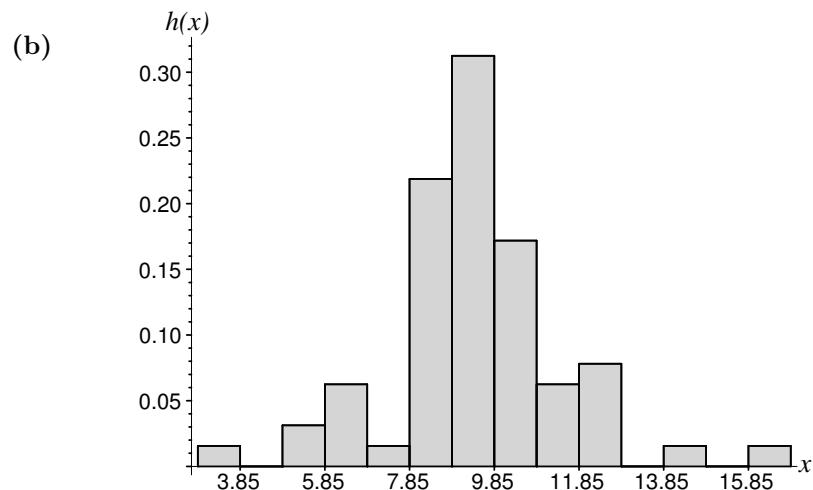
Figure 3.2–12: **(b)** Lead concentrations

(c) $\bar{x} = 9.422$, $s = 2.082$.
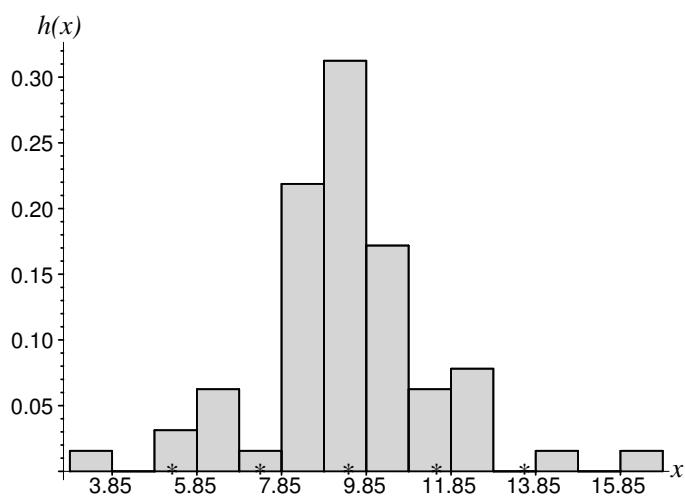
Figure 3.2–12: **(c)** Lead concentrations showing $\bar{x}, \bar{x} \pm s, \bar{x} \pm 2s$

There are 44 ($44/64 = 68.75\%$) within one standard deviation of the mean and 56 ($56/64 = 87.5\%$) within two standard deviations of the mean.

(d)

1976 Leaves	Stems	1977 Leaves
1	2	9
9 2	3	
9	4	
9 9 4 3 2 0 0	5	0 7
9 8 8 7 5 5 4 4 4 4 3 2 2 1 1 0 0 0 0 0	6	3 5 6 8
9 8 6 6 3 2 2 1 0	7	3
7 6 6 5 5 4 3 3 1 1 0 0	8	0 1 1 2 2 2 3 6 7 7 7 8 8 8 9 9
9 7 5 3 2 0	9	1 1 2 3 3 3 3 4 4 4 4 5 5 6 7 8 8 8 9 9 9 9
9 6 1	10	2 2 3 4 5 5 7 9
2	11	0 4 6 9
	12	0 3 4 6
	13	
1	14	8
	15	
	17	7

Multiply numbers by 10^{-1}

Table 3.2–12: Back-to-Back Stem-and-Leaf Diagram of Lead Concentrations
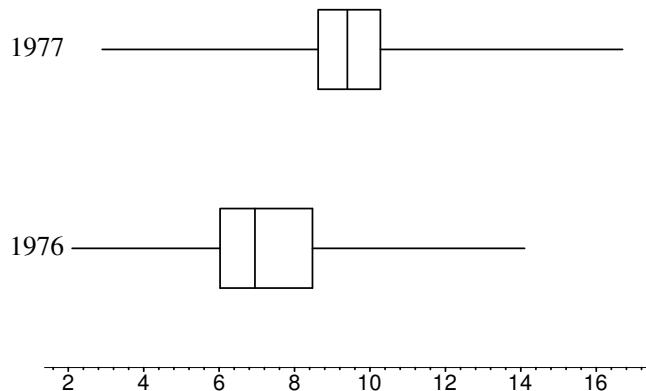
Figure 3.2–12: Box-and-whisker diagrams of 1976 and 1977 lead concentrations

3.3 Random Variables of the Continuous Type

3.3–2 (a) (i) $\displaystyle\int_0^c x^3/4\,dx \;=\; 1$

$$c^4/16 \;=\; 1$$

$$c \;=\; 2;$$

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

$$=\; \int_0^x t^3/4\,dt$$

$$=\; x^4/16,$$

$$F(x) = \begin{cases} 0, & -\infty < x < 0, \\ x^4/16, & 0 \le x < 2, \\ 1, & 2 \le x < \infty. \end{cases}$$
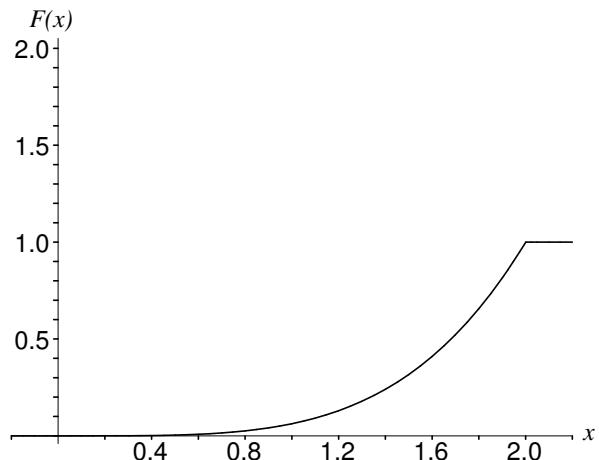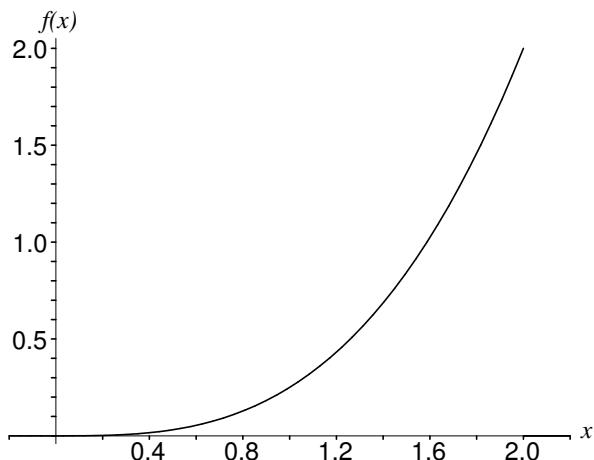
Figure 3.3–2: **(a)** Continuous distribution p.d.f. and c.d.f.

(b) (i) $\displaystyle\int_{-c}^{c} (3/16)x^2 \, dx = 1$

$$c^3/8 = 1$$

$$c = 2;$$

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

$$= \int_{-2}^{x} (3/16)t^2 \, dt$$

$$= \left[\frac{t^3}{16}\right]_{-2}^{x}$$

$$= \frac{x^3}{16} + \frac{1}{2},$$

$$F(x) = \begin{cases} 0, & -\infty < x < -2, \\[2mm] \dfrac{x^3}{16} + \dfrac{1}{2}, & -2 \le x < 2, \\[2mm] 1, & 2 \le x < \infty. \end{cases}$$
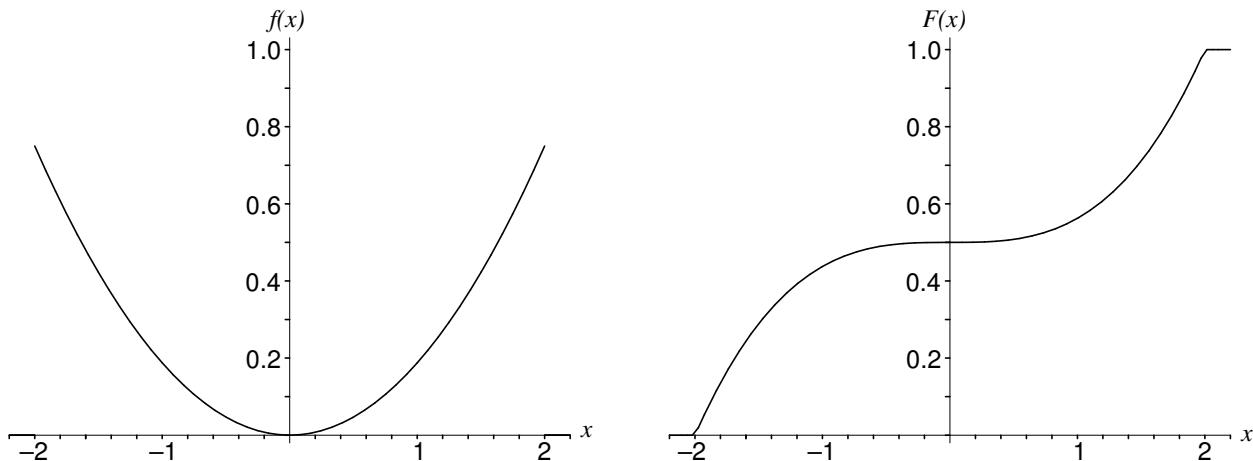
Figure 3.3–2: **(b)** Continuous distribution p.d.f. and c.d.f.

(c) (i) $\displaystyle\int_0^1 \frac{c}{\sqrt{x}}\,dx \;=\; 1$

$$2c \;=\; 1$$

$$c \;=\; 1/2.$$

The p.d.f. in part (c) is unbounded.

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

$$=\; \int_0^x \frac{1}{2\sqrt{t}}\,dt$$

$$=\; \left[\sqrt{t}\,\right]_0^x = \sqrt{x},$$

$$F(x) = \begin{cases} 0, & -\infty < x < 0, \\ \sqrt{x}, & 0 \le x < 1, \\ 1, & 1 \le x < \infty. \end{cases}$$
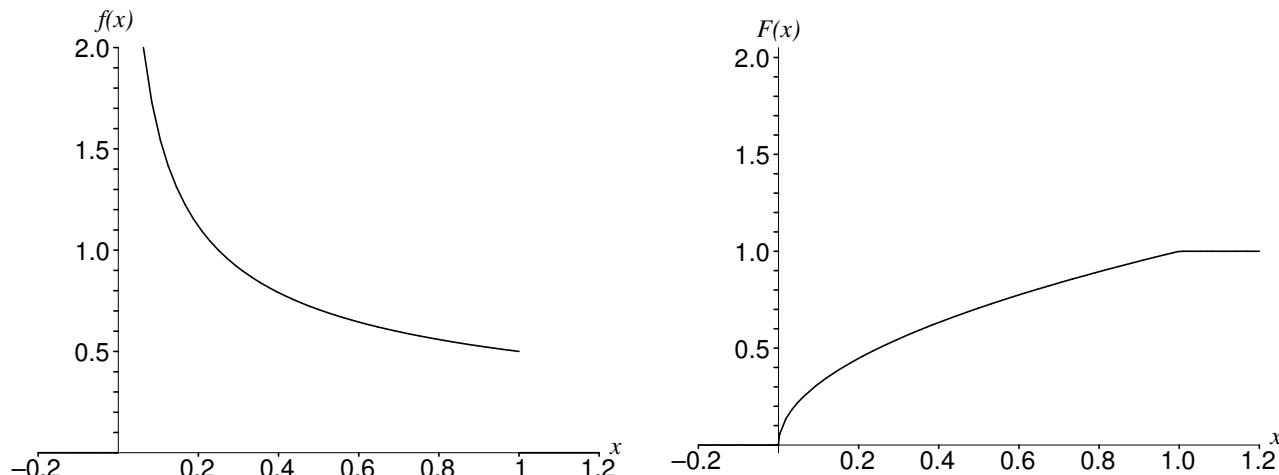
Figure 3.3–2: **(c)** Continuous distribution p.d.f. and c.d.f.

3.3–4 (a)
$$
\begin{aligned}
\mu = E(X) &= \int_0^2 \frac{x^4}{4}\, dx \\[2mm]
&= \left[\frac{x^5}{20}\right]_0^2 = \frac{32}{20} = \frac{8}{5}, \\[2mm]
\sigma^2 = \mathrm{Var}(X) &= \int_0^2 \left(x - \frac{8}{5}\right)^2 \frac{x^3}{4}\, dx \\[2mm]
&= \int_0^2 \left(\frac{x^5}{4} - \frac{4}{5}x^4 + \frac{16}{25}x^3\right) dx \\[2mm]
&= \left[\frac{x^6}{24} - \frac{4x^5}{25} + \frac{4x^4}{25}\right]_0^2 \\[2mm]
&= \frac{64}{24} - \frac{128}{25} + \frac{64}{25} \\[2mm]
&\approx 0.1067, \\[2mm]
\sigma &= \sqrt{0.1067} = 0.3266;
\end{aligned}
$$

(b)
$$
\begin{aligned}
\mu = E(X) &= \int_{-2}^2 \left(\frac{3}{16}\right) x^3\, dx \\[2mm]
&= \left[\frac{3}{64}x^4\right]_{-2}^2 \\[2mm]
&= \frac{48}{64} - \frac{48}{64} = 0, \\[2mm]
\sigma^2 = \mathrm{Var}(X) &= \int_{-2}^2 \left(\frac{3}{16}\right) x^4\, dx \\[2mm]
&= \left[\frac{3}{80}x^5\right]_{-2}^2 \\[2mm]
&= \frac{96}{80} + \frac{96}{80} \\[2mm]
&= \frac{12}{5}, \\[2mm]
\sigma = \sqrt{\frac{12}{5}} &\approx 1.5492;
\end{aligned}
$$

(c) $\mu = E(X) = \int_0^1 \dfrac{x}{2\sqrt{x}}\,dx$

$= \int_0^1 \dfrac{\sqrt{x}}{2}\,dx$

$= \left[\dfrac{x^{3/2}}{3}\right]_0^1 = \dfrac{1}{3},$

$\sigma^2 = \text{Var}(X) = \int_0^1 \left(x - \dfrac{1}{3}\right)^2 \dfrac{1}{2\sqrt{x}}\,dx$

$= \int_0^1 \left(\dfrac{1}{2}x^{3/2} - \dfrac{2}{6}x^{1/2} + \dfrac{1}{18}x^{-1/2}\right)dx$

$= \left[\dfrac{1}{5}x^{5/2} - \dfrac{2}{9}x^{3/2} + \dfrac{1}{9}x^{1/2}\right]_0^1$

$= \dfrac{4}{45},$

$\sigma = \dfrac{2}{\sqrt{45}} \approx 0.2981.$

3.3–6 (a) $M(t) = \int_0^\infty e^{tx}(1/2)x^2 e^{-x}\,dx$

$= \left[-\dfrac{x^2 e^{-x(1-t)}}{2(1-t)} - \dfrac{x e^{-x(1-t)}}{(1-t)^2} - \dfrac{e^{-x(1-t)}}{(1-t)^3}\right]_0^\infty$

$= \dfrac{1}{(1-t)^3}, \qquad t < 1;$

(b) $M'(t) = \dfrac{3}{(1-t)^4}$

$M''(t) = \dfrac{12}{(1-t)^5}$

$\mu = M'(0) = 3$

$\sigma^2 = M''(0) - \mu^2 = 12 - 9 = 3.$

3.3–8 (a) $\displaystyle\int_1^\infty \dfrac{c}{x^2}\,dx = 1$

$\left[\dfrac{-c}{x}\right]_1^\infty = 1$

$c = 1;$

(b) $E(X) = \displaystyle\int_1^\infty \dfrac{x}{x^2}\,dx = [\ln x]_1^\infty,$ which is unbounded.

3.3–10 (a)

$$F(x) = \begin{cases} 0, & -\infty < x < -1, \\ (x^3+1)/2, & -1 \le x < 1, \\ 1, & 1 \le x < \infty. \end{cases}$$
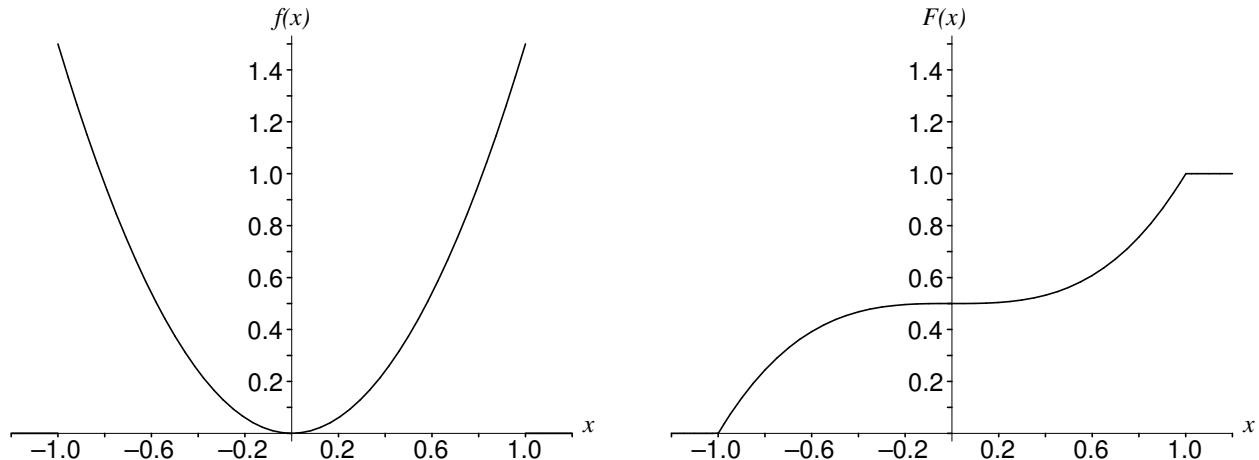
Figure 3.3–10: **(a)** $f(x) = (3/2)x^2$ and $F(x) = (x^3+1)/2$

(b)

$$F(x) = \begin{cases} 0, & -\infty < x < -1, \\ (x+1)/2, & -1 \le x < 1, \\ 1, & 1 \le x < \infty. \end{cases}$$
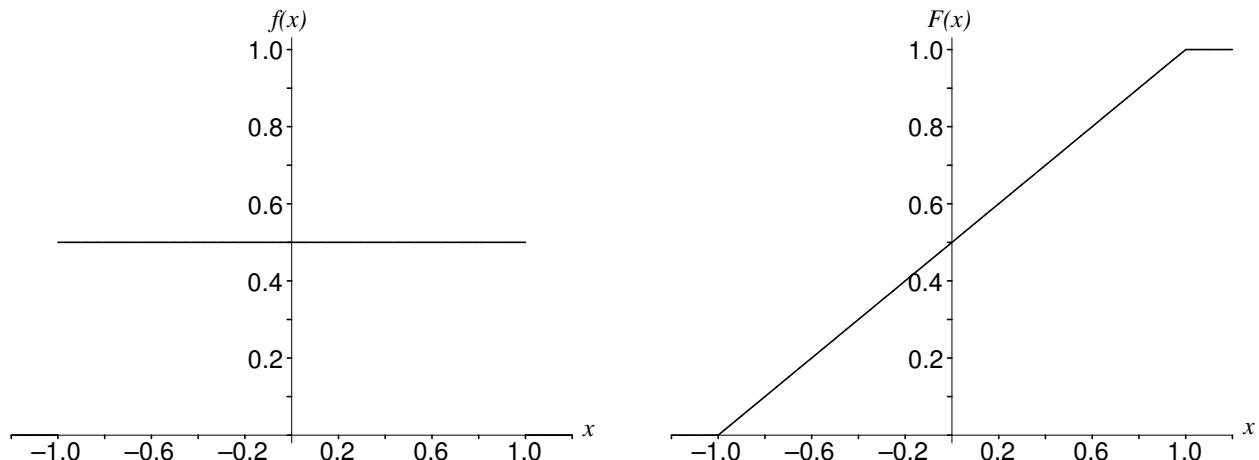
Figure 3.3–10: **(b)** $f(x) = 1/2$ and $F(x) = (x+1)/2$

(c)

$$F(x) = \begin{cases} 0, & -\infty < x < -1, \\ (x+1)^2/2, & -1 \le x < 0, \\ 1 - (1-x)^2/2, & 0 \le x < 1, \\ 1, & 1 \le x < \infty. \end{cases}$$
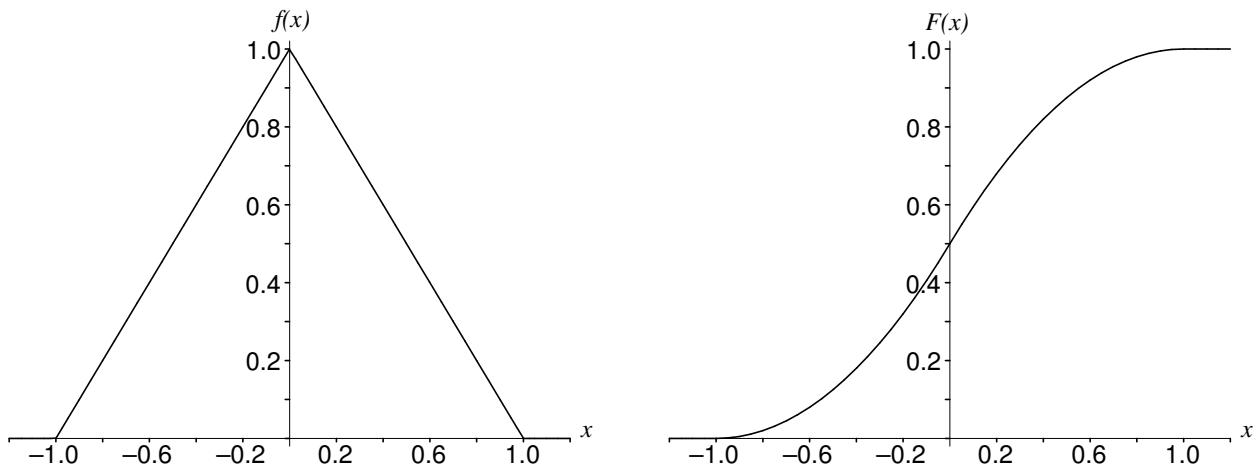
Figure 3.3–10: **(c)** $f(x)$ and $F(x)$ for Exercise 3.3-10(c)

3.3–12 (a) $R'(t) = \dfrac{M'(t)}{M(t)}$; $R'(0) = \dfrac{M'(0)}{M(0)} = M'(0) = \mu$;

(b) $R''(t) = \dfrac{M(t)M''(t) - [M'(t)]^2}{[M(t)]^2}$,

$R''(0) = M''(0) - [M'(0)]^2 = \sigma^2$.

3.3–14 $M(t) = \displaystyle\int_0^\infty e^{tx}(1/10)e^{-x/10}\,dx = \int_0^\infty (1/10)e^{-(x/10)(1-10t)}\,dx$

$= (1 - 10t)^{-1}, \qquad t < 1/10.$

$R(t) = \ln M(t) = -\ln(1 - 10t);$

$R'(t) = 10/(1 - 10t) = 10(1 - 10t)^{-1};$

$R''(t) = 100(1 - 10t)^{-2}.$

Thus $\mu = R'(0) = 10$; $\sigma^2 = R''(0) = 100$.

3.3–16 (b)

$$F(x) = \begin{cases} 0, & -\infty < x \le 0, \\ \dfrac{x}{2}, & 0 < x \le 1, \\ \dfrac{1}{2}, & 1 < x \le 2, \\ \dfrac{x}{2} - \dfrac{1}{2} & 2 \le x < 3, \\ 1, & 3 \le x < \infty; \end{cases}$$
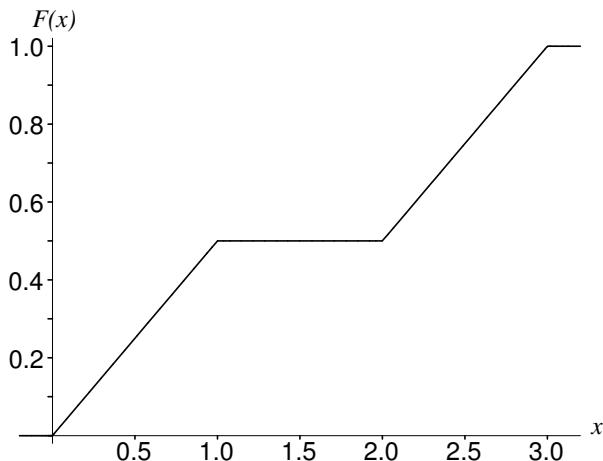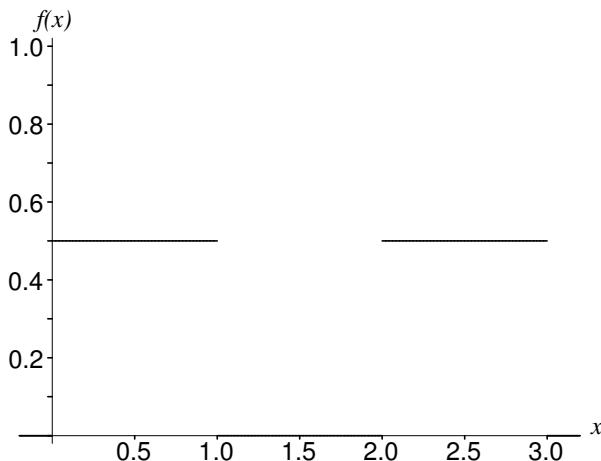
Figure 3.3–16: $f(x)$ and $F(x)$ for Exercise 3.3-16(a)

(c) $\dfrac{q_1}{2} = 0.25$

$\ q_1 = 0.5,$

(d) $1 \le m \le 2,$

(e) $\dfrac{q_3}{2} - \dfrac{1}{2} = 0.75$

$\ \dfrac{q_3}{2} = \dfrac{5}{4}$

$\ q_3 = \dfrac{5}{2}.$

3.3–18 $F(x) = (x+1)^2/4, \qquad -1 < x < 1.$

(a) $F(\pi_{0.64}) = (\pi_{0.64}+1)^2/4 = 0.64$

$\ \pi_{0.64}+1 = \sqrt{2.56}$

$\ \pi_{0.64} = 0.6;$

(b) $(\pi_{0.25}+1)^2/4 = 0.25$

$\ \pi_{0.25}+1 = \sqrt{1.00}$

$\ \pi_{0.25} = 0;$

(c) $(\pi_{0.81}+1)^2/4 = 0.81$

$\ \pi_{0.81}+1 = \sqrt{3.24}$

$\ \pi_{0.81} = 0.8.$

3.3–20 **(a)** $\quad 35c + \left(\dfrac{1}{2}\right)\left(\dfrac{245}{3} - 35\right)(c) \;=\; 1$

$$\left(35 + \dfrac{70}{3}\right)(c) \;=\; 1$$

$$c \;=\; \dfrac{3}{175}$$

(b) $P(X > 65) = \dfrac{1}{2}\left(\dfrac{3}{490}\right)\left(\dfrac{50}{3}\right) = \dfrac{5}{98} = 0.05;$

(c) $\quad \left(\dfrac{3}{490}\right)(m) \;=\; \dfrac{1}{2}$

$$m \;=\; \dfrac{175}{6} \;=\; 29.167.$$

3.3–22 $P(X > 2) = \displaystyle\int_{2}^{\infty} 4x^{3}e^{-x^{4}}\,dx = \left[-e^{-x^{4}}\right]_{2}^{\infty} = e^{-16}.$

3.3–24 **(a)** $P(X > 2000) = \displaystyle\int_{2000}^{\infty} (2x/1000^{2})e^{-(x/1000)^{2}}\,dx = \left[-e^{-(x/1000)^{2}}\right]_{2000}^{\infty} = e^{-4};$

(b) $\quad \left[-e^{-(x/1000)^{2}}\right]_{\pi_{0.75}}^{\infty} \;=\; 0.25$

$$e^{-(\pi_{0.75}/1000)^{2}} \;=\; 0.25$$

$$-(\pi_{0.75}/1000)^{2} \;=\; \ln(0.25)$$

$$\pi_{0.75} \;=\; 1177.41;$$

(c) $\pi_{0.10} = 324.59;$

(d) $\pi_{0.60} = 957.23.$

3.3–26 **(a)** $\quad \displaystyle\int_{0}^{1} x\,dx + \int_{1}^{\infty} \dfrac{c}{x^{3}}\,dx \;=\; 1$

$$\left[\dfrac{x^{2}}{2}\right]_{0}^{1} - \left[\dfrac{c}{2x^{2}}\right]_{1}^{\infty} \;=\; 1$$

$$\dfrac{1}{2} + \dfrac{c}{2} \;=\; 1$$

$$c \;=\; 1;$$

(b) $E(X) = \displaystyle\int_{0}^{1} x^{2}\,dx + \int_{1}^{\infty} \dfrac{1}{x^{2}}\,dx = \dfrac{4}{3};$

(c) the variance does not exist;

(d) $P(1/2 \le X \le 2) = \displaystyle\int_{1/2}^{1} x\,dx + \int_{1}^{2} \dfrac{1}{x^{3}}\,dx = \dfrac{3}{4}.$

3.4   The Uniform and Exponential Distributions

3.4–2 $\mu = 0,\quad \sigma^{2} = 1/3.$ See the figures for Exercise 3.3-10(b).

3.4–4 X is $U(4,5);$

(a) $\mu = 9/2;$ **(b)** $\sigma^{2} = 1/12;$ **(c)** $0.5.$

3.4–6 (a) $\quad P(10 < X < 30) \quad = \quad \displaystyle\int_{10}^{30} \left(\frac{1}{20}\right) e^{-x/20}\, dx$

$\qquad\qquad\qquad\qquad\qquad = \quad \left[-e^{-x/20}\right]_{10}^{30} = e^{-1/2} - e^{-3/2};$

(b) $\quad P(X > 30) \quad = \quad \displaystyle\int_{30}^{\infty} \frac{1}{20} e^{-x/20}\, dx$

$\qquad\qquad\qquad\qquad = \quad \left[-e^{-x/20}\right]_{30}^{\infty} = e^{-3/2};$

(c) $\quad P(X > 40 \,|\, X > 10) \quad = \quad \dfrac{P(X > 40)}{P(X > 10)}$

$\qquad\qquad\qquad\qquad\qquad = \quad \dfrac{e^{-2}}{e^{-1/2}} = e^{-3/2};$

(d) $\quad \sigma^2 = \theta^2 = 400, \quad M(t) = (1 - 20t)^{-1}.$

(e) $\qquad P(10 < X < 30) \quad = \quad 0.383, \quad \text{close to the relative frequency} \quad \dfrac{35}{100},$

$\qquad\qquad\qquad P(X > 30) \quad = \quad 0.223, \quad \text{close to the relative frequency} \quad \dfrac{23}{100},$

$\qquad\qquad P(X > 40 \,|\, X > 10) \quad = \quad 0.223, \quad \text{close to the relative frequency} \quad \dfrac{14}{58} = 0.241.$

3.4–8 (a) $f(x) = \left(\dfrac{2}{3}\right) e^{-2x/3}, \qquad 0 \le x < \infty;$

(b) $P(X > 2) = \displaystyle\int_{2}^{\infty} \frac{2}{3} e^{-2x/3}\, dx = \left[-e^{-2x/3}\right]_{2}^{\infty} = e^{-4/3}.$

3.4–10 (a) Using X for the infected snails and Y for the control snails, $\bar{x} = 84.74$, $s_x = 64.79$, $\bar{y} = 113.1612903$, $s_y = 87.02$;

(b)
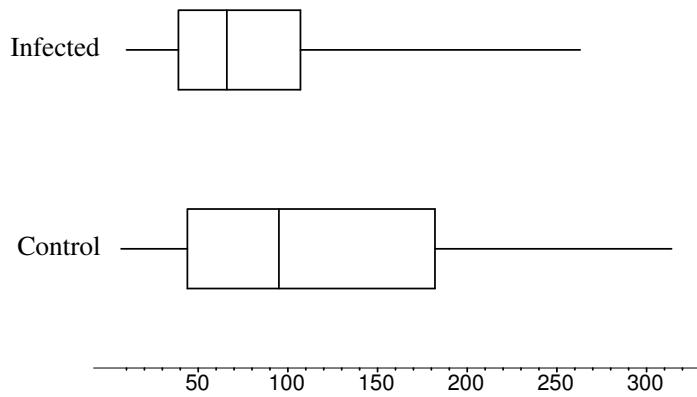
Figure 3.4–10: **(b)** Box-and-whisker diagrams of distances traveled by infected and control snails
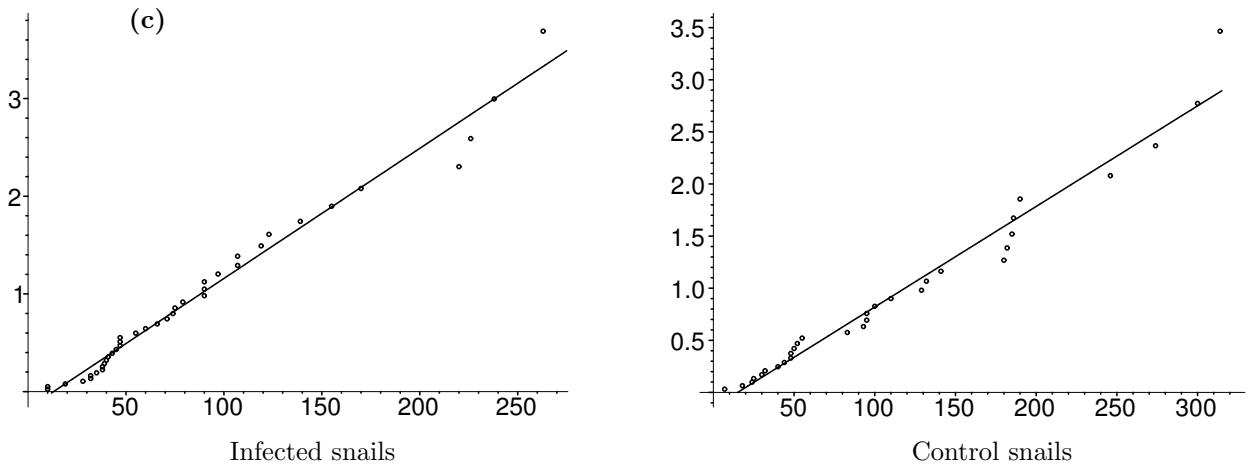
Figure 3.4–10: **(c)** *q-q* plots, exponential quantiles versus ordered infected and control snail times

(d) Possibly;

(e) The control snails move further than the infected snails but the distributions of the two sets of distances are similar.

3.4–12 Let $F(x) = P(X \le x)$. Then

$$P(X > x + y \mid X > x) = P(X > y)$$

$$\frac{1 - F(x+y)}{1 - F(x)} = 1 - F(y).$$

That is, with $g(x) = 1 - F(x)$, $g(x+y) = g(x)g(y)$. This functional equation implies that

$$1 - F(x) = g(x) = a^{cx} = e^{(cx)\ln a} = e^{bx}$$

where $b = c \ln a$. That is, $F(x) = 1 - e^{bx}$. Since $F(\infty) = 1$, b must be negative, say $b = -\lambda$ with $\lambda > 0$. Thus $F(x) = 1 - e^{-\lambda x}$, $0 \le x$, the distribution function of an exponential distribution.

3.4–14
$$E[v(T)] = \int_0^3 100(2^{3-t} - 1)e^{-t/5}/5 \, dt$$

$$= \int_0^3 -20e^{-t/5} dt + 100 \int_0^3 e^{(3-t)\ln 2} e^{-t/5}/5 \, dt$$

$$= -100(1 - e^{-0.6}) + 100 e^{3\ln 2} \int_0^3 e^{-t \ln 2} e^{-t/5}/5 \, dt$$

$$= -100(1 - e^{-0.6}) + 100 e^{3\ln 2} \left[-\frac{e^{-(\ln 2 + 0.2)t}}{\ln 2 + 0.2} \right]_0^3 = 121.734.$$

3.4–16
$$E(\text{profit}) = \int_0^n [x - 0.5(n-x)]\frac{1}{200} \, dx + \int_n^{200} [n - 5(x-n)]\frac{1}{200} \, dx$$

$$= \frac{1}{200}\left[\frac{x^2}{2} + \frac{(n-x)^2}{4} \right]_0^n + \frac{1}{200}\left[6nx - \frac{5x^2}{2} \right]_n^{200}$$

$$= \frac{1}{200}\left[-3.25n^2 + 1200n - 100000 \right]$$

$$\text{derivative} = \frac{1}{200}[-6.5n + 1200] = 0$$

$$n = \frac{1200}{6.5} \approx 185.$$

3.4–18 **(a)** $P(X > 40)$ $=$ $\displaystyle\int_{40}^{\infty} \frac{3}{100} e^{-3x/100}\, dx$

$=$ $\left[-e^{-3x/100}\right]_{40}^{\infty} = e^{-1.2};$

(b) Flaws occur randomly so we are observing a Poisson process.

3.4–20 $F(x)$ $=$ $\displaystyle\int_{-\infty}^{x} \frac{e^{-w}}{(1+e^{-w})^2}\, dw = \frac{1}{1+e^{-x}},$ $-\infty < x < \infty.$

$G(y)$ $=$ $P\left[\dfrac{1}{1+e^{-X}} \le y\right] = P\left[X \le -\ln\left(\dfrac{1}{y} - 1\right)\right]$

$=$ $\dfrac{1}{1 + \left(\dfrac{1}{y} - 1\right)} = y,$ $0 < y < 1,$

the $U(0,1)$ distribution function.

3.4–22 $P(X > 100 \,|\, X > 50) = P(X > 50) = 3/4.$

3.5   The Gamma and Chi-Square Distributions

3.5–2 Either use integration by parts or

$F(x)$ $=$ $P(X \le x)$

$=$ $1 - \displaystyle\sum_{k=0}^{\alpha-1} \frac{(\lambda x)^k e^{-\lambda x}}{k!}.$

Thus, with $\lambda = 1/\theta = 1/4$ and $\alpha = 2,$

$P(X < 5)$ $=$ $1 - e^{-5/4} - \left(\dfrac{5}{4}\right)e^{-5/4}$

$=$ $0.35536.$

3.5–4 The moment generating function of X is $M(t) = (1 - \theta t)^{-\alpha},\; t < 1/\theta.$ Thus

$M'(t)$ $=$ $\alpha\theta(1 - \theta t)^{-\alpha - 1}$

$M''(t)$ $=$ $\alpha(\alpha + 1)\theta^2(1 - \theta t)^{-\alpha - 2}.$

The mean and variance are

μ $=$ $M'(0) = \alpha\theta$

σ^2 $=$ $M''(0) - (\alpha\theta)^2 = \alpha(\alpha + 1)\theta^2 - (\alpha\theta)^2$

$=$ $\alpha\theta^2.$

3.5–6 **(a)** $f(x)$ $=$ $\dfrac{14.7^{100}}{\Gamma(100)} x^{99} e^{-14.7x},$ $0 \le x < \infty,$

μ $=$ $100(1/14.7) = 6.80,\;\; \sigma^2 = 100(1/14.7)^2 = 0.4628;$

(b) $\bar{x} = 6.74,\; s^2 = 0.4617;$

(c) $9/25 = 0.36.$ (See Figure 8.7-2 in the textbook.)

3.5–8 **(a)** W has a gamma distribution with $\alpha = 7$, $\theta = 1/16$.

(b) Using Table III in the Appendix,

$$P(W \leq 0.5) \;=\; 1 - \sum_{k=0}^{6} \frac{8^k e^{-8}}{k!}$$

$$=\; 1 - 0.313 = 0.687,$$

because here $\lambda w = (16)(0.5) = 8$.

3.5–10 $a = 5.226$, $b = 21.03$.

3.5–12 Since the m.g.f. is that of $\chi^2(24)$, we have **(a)** $\mu = 24$; **(b)** $\sigma^2 = 48$; and **(c)** 0.89, using Table IV.

3.5–14 Note that $\lambda = 5/10 = 1/2$ is the mean number of arrivals per minute. Thus $\theta = 2$ and the p.d.f. of the waiting time before the eighth toll is

$$f(x) \;=\; \frac{1}{\Gamma(8)2^8}\, x^{8-1} e^{-x/2}$$

$$=\; \frac{1}{\Gamma\!\left(\dfrac{16}{2}\right) 2^{16/2}}\, x^{16/2-1} e^{-x/2}, \qquad 0 < x < \infty,$$

the p.d.f. of a chi-square distribution with $r = 16$ degrees of freedom. Using Table IV,

$$P(X > 26.30) = 0.05.$$

3.5–16 $P(X > 30.14) = 0.05$ where X denotes a single observation. Let W equal the number out of 10 observations that exceed 30.14. Then the distribution of W is $b(10,\ 0.05)$. Thus

$$P(W = 2) = 0.9885 - 0.9139 = 0.0746.$$

3.5–18 **(a)** $\mu = \displaystyle\int_{80}^{\infty} x \cdot \frac{x - 80}{50^2} e^{-(x-80)/50}\, dx$. Let $y = x - 80$. Then

$$\mu \;=\; 80 + \int_{0}^{\infty} y \cdot \frac{1}{\Gamma(2)50^2} y^{2-1} e^{-y/50}\, dy$$

$$=\; 80 + 2(50) \;=\; 180.$$

$$\mathrm{Var}(X) = \mathrm{Var}(Y) = 2(50^2) = 5000.$$

(b) $$f'(x) \;=\; \frac{1}{50^2} e^{-(x-80)/50} - \frac{x-80}{50^2}\frac{1}{50} e^{-(x-80)/50} \;=\; 0$$

$$50 - x + 80 \;=\; 0$$

$$x \;=\; 130.$$

(c) $$\int_{80}^{200} \frac{x-80}{50^2} e^{-(x-80)/50}\, dx \;=\; \left[-\frac{x-80}{50} e^{-(x-80)/50} - e^{-(x-80)/50} \right]_{80}^{200}$$

$$=\; \frac{-120}{50} e^{-120/50} - e^{-120/50} + 1$$

$$=\; 1 - \frac{17}{5} e^{-12/5} \;=\; 0.6916.$$

3.6   The Normal Distribution

3.6–2 **(a)** 0.3078:   **(b)** 0.4959;

   (c) 0.2711;   **(d)** 0.1646.

3.6–4 **(a)** 1.282;   **(b)** −1.645;

   (c) −1.66;   **(d)** −1.82.

3.6–6 $M(t) = e^{166t + 400t^2/2}$ so

   (a) $\mu = 166$;   **(b)** $\sigma^2 = 400$;

   (c) $P(170 < X < 200) = P(0.2 < Z < 1.7) = 0.3761$;

   (d) $P(148 \leq X \leq 172) = P(-0.9 \leq Z \leq 0.3) = 0.4338$.

3.6–8 We must solve $f''(x) = 0$. We have

$$\ln f(x) = -\ln(\sqrt{3\pi}\,\sigma) - (x - \mu)^2/2\sigma^2,$$

$$\frac{f'(x)}{f(x)} = \frac{-2(x - \mu)}{2\sigma^2}$$

$$\frac{f(x)f''(x) - [f'(x)]^2}{[f(x)]^2} = \frac{-1}{\sigma^2}$$

$$f''(x) = f(x)\left\{\frac{-1}{\sigma^2} + \left[\frac{f'(x)}{f(x)}\right]\right\} = 0$$

$$\frac{(x - \mu)^2}{\sigma^4} = \frac{1}{\sigma^2}$$

$$x - \mu = \pm\sigma \qquad \text{or} \qquad x = \mu \pm \sigma.$$

3.6–10
$$G(y) = P(Y \leq y) = P(aX + b \leq y)$$

$$= P\left(X \leq \frac{y - b}{a}\right) \quad \text{if} \quad a > 0$$

$$= \int_{-\infty}^{(y-b)/a} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}\, dx$$

Let $w = ax + b$ so $dw = a\, dx$. Then

$$G(y) = \int_{-\infty}^{y} \frac{1}{a\sigma\sqrt{2\pi}} e^{-(w-b-a\mu)^2/2a^2\sigma^2}\, dw$$

which is the distribution function of the normal distribution $N(b + a\mu,\ a^2\sigma^2)$. The case when $a < 0$ can be handled similarly.

3.6–12 **(a)**

Stems	Leaves	Frequencies	Depths
11●	8	1	1
12∗	0 3	2	3
12●	5 6	2	5
13∗	1 3 4	3	8
13●	5 5 7 7 7 9	6	14
14∗	0 0 2 2 3 4 4 4	8	22
14●	6 6 7 7 7 8 9 9	8	30
15∗	0 0 0 0 1 1 1 2 3 3 4 4	12	30
15●	5 5 6 7 8 8 8 9	8	18
16∗	0 0 0 2 3 4	6	10
16●	5 5	2	4
17∗	1	1	2
17●	5	1	1

(b)
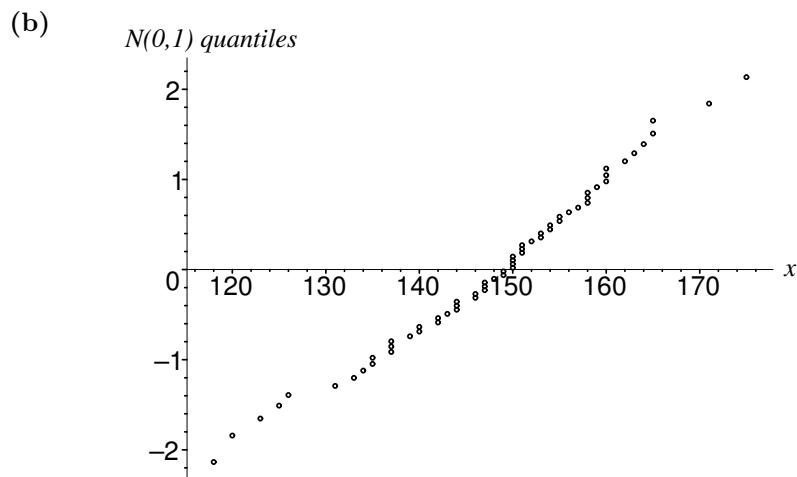
Figure 3.6–12: q-q plot of $N(0,1)$ quantiles versus data quantiles

(c) Yes.

3.6–14 **(a)** $P(X > 22.07) = P(Z > 1.75) = 0.0401;$

(b) $P(X < 20.857) = P(Z < -1.2825) = 0.10.$ Thus the distribution of Y is $b(15,\ 0.10)$ and from Table II in the Appendix, $P(Y \leq 2) = 0.8159.$

3.6–16 X is $N(500, 10000);$ so $[(X - 500)^2/100]^2$ is $\chi^2(1)$ and

$$P\left[2.706 \leq \left(\frac{X - 500}{100}\right)^2 \leq 5.204\right] = 0.975 - 0.900 = 0.075.$$

3.6–18 $G(x)$ $=$ $P(X \leq x)$

$=$ $P(e^Y \leq x)$

$=$ $P(Y \leq \ln x)$

$=$ $\displaystyle\int_{-\infty}^{\ln x} \frac{1}{\sqrt{2\pi}} e^{-(y-10)^2/2} dy = \Phi(\ln x - 10)$

$g(x)$ $=$ $G'(x) = \dfrac{1}{\sqrt{2\pi}} e^{-(\ln x - 10)^2/2} \dfrac{1}{x}, \qquad 0 < x < \infty.$

$P(10{,}000 < X < 20{,}000)$ $=$ $P(\ln 10{,}000 < Y < \ln 20{,}000)$

$=$ $\Phi(\ln 20{,}000 - 10) - \Phi(\ln 10{,}000 - 10)$

$=$ $0.461557 - 0.214863 = 0.246694$ using Minitab.

3.6–20

k	Strengths	$p = k/10$	z_{1-p}	k	Strengths	$p = k/10$	z_{1-p}
1	7.2	0.10	-1.282	6	11.7	0.60	0.253
2	8.9	0.20	-0.842	7	12.9	0.70	0.524
3	9.7	0.30	-0.524	8	13.9	0.80	0.842
4	10.5	0.40	-0.253	9	15.3	0.90	1.282
5	10.9	0.50	0.000				
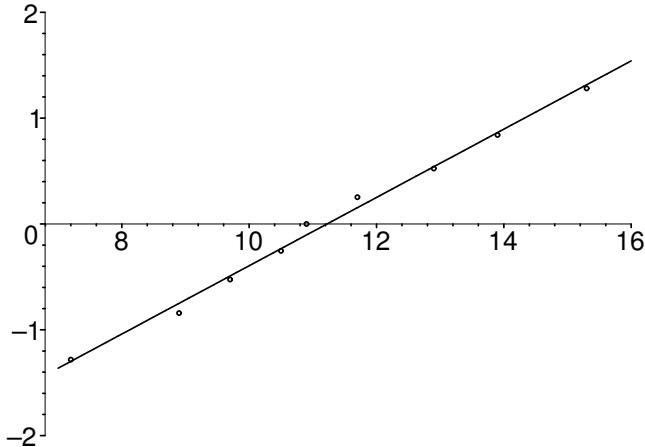
Figure 3.6–20: q-q plot of $N(0,1)$ quantiles versus data quantiles
It seems to be an excellent fit.

3.6–22 The three respective distributions are exponential with $\theta = 4$, $\chi^2(4)$, and $N(4,1)$. Each of these has a mean of $\mu = 4$ and the mean is the first derivative of the moment-generating function evaluated at $t = 0$. Thus the slopes at $t = 0$ are all equal to 4.

3.6–24 (a)
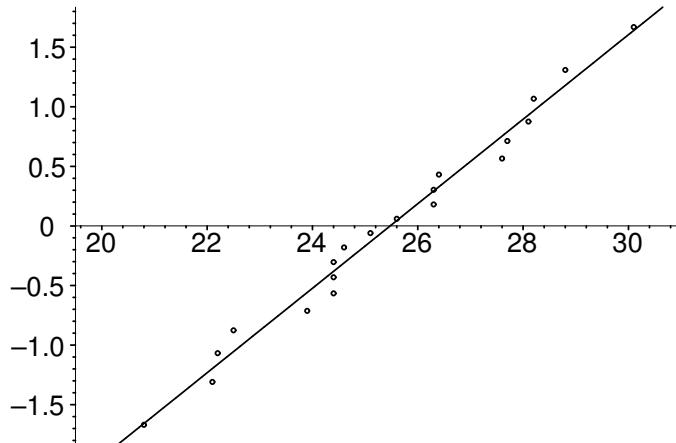
Figure 3.6–24: *q-q* plot of $N(0,1)$ quantiles versus data quantiles

(b) It looks like an excellent fit.

3.6–26 (a) $\bar{x} = 55.95$, $s = 1.78$;

(b)
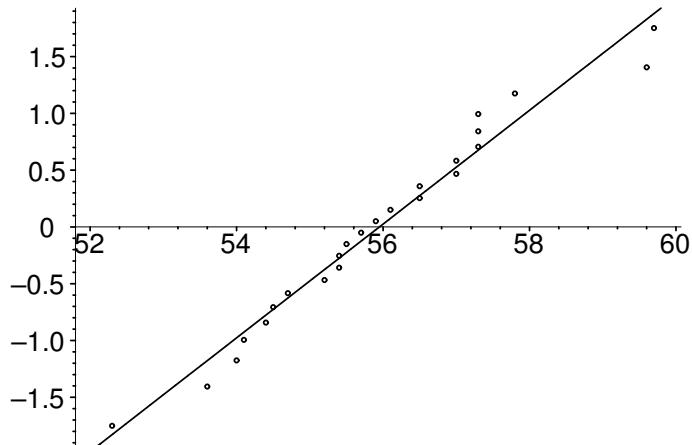
Figure 3.6–26: *q-q* plot of $N(0,1)$ quantiles versus data quantiles

(c) It looks like an excellent fit.

(d) The label weight could actually be a little larger.

3.7   Additional Models

3.7–2 With $b = \ln 1.1$,

$$G(w) \quad = \quad 1 - \exp\left[-\frac{a}{\ln 1.1} e^{w \ln 1.1} + \frac{a}{\ln 1.1}\right]$$

$$G(64) - G(63) \quad = \quad 0.01$$

$$a \quad = \quad 0.00002646 = \frac{1}{37792.19477}$$

$$P(W \leq 71 \mid 70 < W) \quad = \quad \frac{P(70 < W \leq 71)}{P(70 < W)}$$

$$= \quad 0.0217.$$

3.7–4 $\lambda(w) \quad = \quad ae^{bw} + c$

$$H(w) \quad = \quad \int_0^w (ae^{bt} + c)\, dt$$

$$= \quad \frac{a}{b}\left(e^{bw} - 1\right) + cw$$

$$G(w) \quad = \quad 1 - \exp\left[-\frac{a}{b}\left(e^{bw} - 1\right) - cw\right], \qquad 0 < \infty$$

$$g(w) \quad = \quad (ae^{bw} + c)e^{-\frac{a}{b}(e^{bw} - 1) - cw}, \qquad 0 < \infty.$$

3.7–6   **(a)** $1/4 - 1/8 = 1/8$;   **(b)** $1/4 - 1/4 = 0$;

         (c) $3/4 - 1/4 = 1/2$;   **(d)** $1 - 1/2 = 1/2$;

         (e) $3/4 - 3/4 = 0$;   **(f)** $1 - 3/4 = 1/4$.

3.7–8 There is a discrete point of probability at $x = 0$, $P(X = 0) = 1/3$, and $F'(x) = (2/3)e^{-x}$ for $0 < x$. Thus

$$\mu = E(X) \quad = \quad (0)(1/3) + \int_0^\infty x(2/3)e^{-x} dx$$

$$= \quad (2/3)[-xe^{-x} + e^{-x}]_0^\infty = 2/3,$$

$$E(X^2) \quad = \quad (0)^2(1/3) + \int_0^\infty x^2(2/3)e^{-x} dx$$

$$= \quad (2/3)[-x^2 e^{-x} - 2xe^{-x} - 2e^{-x}]_0^\infty = 4/3,$$

so

$$\sigma^2 \quad = \quad \text{Var}(X) = 4/3 - (2/3)^2 = 8/9.$$

3.7–10 $T = \begin{cases} X, & X \le 4, \\ 4, & 4 < X; \end{cases}$

$$
\begin{aligned}
E(T) &= \int_0^4 x \left(\frac{1}{5}\right) e^{-x/5} \, dx + \int_4^\infty 4 \left(\frac{1}{5}\right) e^{-x/5} \, dx \\
&= \left[-xe^{-x/5} - 5e^{-x/5}\right]_0^4 + 4\left[-e^{-x/5}\right]_4^\infty \\
&= 5 - 4e^{-4/5} - 5e^{-4/5} + 4e^{-4/5} \\
&= 5 - 5e^{-4/5} \approx 2.753.
\end{aligned}
$$

3.7–12 (a)
$$
\begin{aligned}
t &= \ln x \\
x &= e^t \\
\frac{dx}{dt} &= e^t \\
g(t) &= f(e^t)\frac{dx}{dt} = e^t e^{-e^t}, \qquad -\infty < t < \infty.
\end{aligned}
$$

(b)
$$
\begin{aligned}
t &= \alpha + \beta \ln w \\
\frac{dt}{dw} &= \frac{\beta}{w} \\
h(w) &= e^{\alpha + \beta \ln w} e^{-e^{\alpha + \beta \ln w}} \left(\frac{\beta}{w}\right) \\
&= \beta w^{\beta - 1} e^\alpha e^{-w^\beta e^\alpha}, \qquad 0 < w < \infty.
\end{aligned}
$$

3.7–14 (a) $((0.03) \int_{2/30}^1 6(1-x)^5 \, dx = 0.0198;$

(b) $E(X) = (0.97)(0) + 0.03 \int_0^1 x6(1-x)^5 \, dx = 0.0042857;$

The expected payment is $E(X) \cdot [\$\,30{,}000] = \$\,128.57.$

3.7–16 $2500m \int_0^1 \frac{1}{10} e^{-x/10} \, dx + (m/2)2500 \int_1^2 \frac{1}{10} e^{-x/10} \, dx = 200$

$$2500m[1 - e^{-1/10}] + 1250m[e^{-1/10} - e^{-2/10}] = 200$$

$$2500m - 1250me^{-1/10} - 1250me^{-2/10} = 200$$

$$m = \frac{4}{50 - 25e^{-1/10} - 25e^{-2/10}}$$

$$= 0.5788.$$

3.7–18 $P(X > x) = \int_x^\infty \left(\frac{t}{4}\right)^3 e^{-(t/4)^4} dt = e^{-(x/4)^4};$

$P(X > 5 \mid X > 4) = \dfrac{P(X > 5)}{P(X > 4)} = \dfrac{e^{-625/256}}{e^{-1}} = e^{-369/256}.$

3.7–20 **(a)** $\displaystyle\int_{40}^{60} \frac{2x}{50^2} e^{-(x/50)^2} dx = \left[-e^{-(x/50)^2}\right]_{40}^{60} = e^{-16/25} - e^{-36/25};$

(b) $P(X > 80) = \left[-e^{-(x/50)^2}\right]_{80}^{\infty} = e^{-64/25}.$

3.7–22 **(a)** $\quad F(y) \;=\; \displaystyle\int_0^y \frac{1}{100} dy = \frac{y}{100}, \qquad 0 < y < 100$

$\qquad\quad e(x) \;=\; \dfrac{\displaystyle\int_x^{100} (y - x) \cdot (1/100)\, dy}{1 - x/100}$

$\qquad\qquad\quad\; =\; \dfrac{1/100}{1 - x/100} \left[\dfrac{(y - x)^2}{2}\right]_x^{100}$

$\qquad\qquad\quad\; =\; \dfrac{1}{100 - x} \dfrac{(100 - x)^2}{2} = \dfrac{100 - x}{2}.$

(b) $\quad F(y) \;=\; \displaystyle\int_{50}^y \frac{50}{t^2} dt = 1 - \frac{50}{y}$

$\qquad\quad e(x) \;=\; \dfrac{\displaystyle\int_x^{\infty} (y - x)(50/y^2)\, dy}{1 - (1 - 50/y)}$

$\qquad\qquad\; =\; \infty.$

Chapter 4

Bivariate Distributions

4.1 Bivariate Distributions

4.1–2

		1	2	3	4	
$\frac{4}{16}$	4	• $\frac{1}{16}$	• $\frac{1}{16}$	• $\frac{1}{16}$	• $\frac{1}{16}$	
$\frac{4}{16}$	3	• $\frac{1}{16}$	• $\frac{1}{16}$	• $\frac{1}{16}$	• $\frac{1}{16}$	
$\frac{4}{16}$	2	• $\frac{1}{16}$	• $\frac{1}{16}$	• $\frac{1}{16}$	• $\frac{1}{16}$	
$\frac{4}{16}$	1	• $\frac{1}{16}$	• $\frac{1}{16}$	• $\frac{1}{16}$	• $\frac{1}{16}$	
		1	2	3	4	
		$\frac{4}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	

(e) Independent, because $f_1(x)f_2(y) = f(x,y)$.

4.1–4

		0	1	2	3	4	5	6	7	8
$\frac{1}{25}$	12									• $\frac{1}{25}$
$\frac{1}{25}$	11									• $\frac{1}{25}$
$\frac{2}{25}$	10							• $\frac{1}{25}$		• $\frac{1}{25}$
$\frac{2}{25}$	9							• $\frac{1}{25}$		• $\frac{1}{25}$
$\frac{3}{25}$	8					• $\frac{1}{25}$		• $\frac{1}{25}$		• $\frac{1}{25}$
$\frac{2}{25}$	7					• $\frac{1}{25}$		• $\frac{1}{25}$		
$\frac{3}{25}$	6			• $\frac{1}{25}$		• $\frac{1}{25}$		• $\frac{1}{25}$		
$\frac{2}{25}$	5			• $\frac{1}{25}$		• $\frac{1}{25}$				
$\frac{3}{25}$	4	• $\frac{1}{25}$		• $\frac{1}{25}$		• $\frac{1}{25}$				
$\frac{2}{25}$	3	• $\frac{1}{25}$		• $\frac{1}{25}$						
$\frac{2}{25}$	2	• $\frac{1}{25}$		• $\frac{1}{25}$						
$\frac{1}{25}$	1	• $\frac{1}{25}$								
$\frac{1}{25}$	0	• $\frac{1}{25}$								
		0	1	2	3	4	5	6	7	8
		$\frac{1}{5}$		$\frac{1}{5}$		$\frac{1}{5}$		$\frac{1}{5}$		$\frac{1}{5}$

(c) Not independent, because $f_1(x)f_2(y) \neq f(x,y)$ and also because the support is not rectangular.

4.1–6 $\dfrac{25!}{7!8!6!4!}(0.30)^7(0.40)^8(0.20)^6(0.10)^4 = 0.00405.$

4.1–8 (a) $f(x,y) = \dfrac{7!}{x!y!(7-x-y)!}(0.78)^x(0.01)^y(0.21)^{7-x-y}, \qquad 0 \le x+y \le 7;$

      (b) X is $b(7,0.78), \qquad x = 0,1,\ldots,7.$

4.1–10 (a) $\begin{aligned} P\left(0 \le X \le \tfrac{1}{2}\right) &= \int_0^{\frac{1}{2}} \int_{x^2}^1 \frac{3}{2}\, dy\, dx \\[2mm] &= \int_0^{\frac{1}{2}} \frac{3}{2}(1-x^2)\, dx = \frac{11}{16}; \end{aligned}$

      (b) $\begin{aligned} P\left(\tfrac{1}{2} \le Y \le 1\right) &= \int_{\frac{1}{2}}^1 \int_0^{\sqrt{y}} \frac{3}{2}\, dx\, dy \\[2mm] &= \int_{\frac{1}{2}}^1 \frac{3}{2}\sqrt{y}\, dy = 1 - \left(\frac{1}{2}\right)^{3/2}; \end{aligned}$

      (c) $\begin{aligned} P\left(\tfrac{1}{2} \le X \le 1,\ \tfrac{1}{2} \le Y \le 1\right) &= \int_{\frac{1}{2}}^1 \int_{\frac{1}{2}}^{\sqrt{y}} \frac{3}{2}\, dx\, dy \\[2mm] &= \int_{\frac{1}{2}}^1 \frac{3}{2}\left(\sqrt{y} - \frac{1}{2}\right) dy \\[2mm] &= \frac{5}{8} - \left(\frac{1}{2}\right)^{3/2}; \end{aligned}$

      (d) $\begin{aligned} P\left(X \ge \tfrac{1}{2}, Y \ge \tfrac{1}{2}\right) &= P\left(\tfrac{1}{2} \le X \le 1,\ \tfrac{1}{2} \le Y \le 1\right) \\[2mm] &= \frac{5}{8} - \left(\frac{1}{2}\right)^{3/2}. \end{aligned}$

      (e) X and Y are dependent.

4.1–12 (a) $\begin{aligned} f_1(x) &= \int_0^1 (x+y)\, dy \\[2mm] &= \left[xy + \frac{1}{2}y^2\right]_0^1 = x + \frac{1}{2}, \qquad 0 \le x \le 10; \\[2mm] f_2(y) &= \int_0^1 (x+y)\, dx = y + \frac{1}{2}, \qquad 0 \le y \le 1; \\[2mm] f(x,y) &= x+y \neq \left(x + \frac{1}{2}\right)\left(y + \frac{1}{2}\right) = f_1(x)f_2(y). \end{aligned}$

      (b) $\mu_X = \int_0^1 x\left(x + \frac{1}{2}\right) dx = \left[\frac{1}{3}x^3 + \frac{1}{4}x^2\right]_0^1 = \frac{7}{12};$

      (c) $\mu_Y = \int_0^1 y\left(y + \frac{1}{2}\right) dy = \frac{7}{12};$

      (d) $\begin{aligned} E(X^2) &= \int_0^1 x^2\left(x + \frac{1}{2}\right) dx = \left[\frac{1}{4}x^4 + \frac{1}{6}x^3\right]_0^1 = \frac{5}{12}, \\[2mm] \sigma_X^2 &= E(X^2) - \mu_X^2 = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144}. \end{aligned}$

(e) Similarly, $\sigma_Y^2 = \dfrac{11}{144}$.

4.1–14 The area of the space is

$$\int_2^6 \int_1^{14-2t_2} dt_1 dt_2 = \int_2^6 (13 - 2t_2)\, dt_2 = 20;$$

Thus

$$\begin{aligned}
P(T_1 + T_2 > 10) &= \int_2^4 \int_{10-t_2}^{14-2t_2} \frac{1}{20} dt_1 dt_2 \\
&= \int_2^4 \frac{4 - t_2}{20} dt_2 \\
&= \left[-\frac{(4 - t_2)^2}{40} \right]_2^4 = \frac{1}{10}.
\end{aligned}$$

4.2   The Correlation Coefficient

4.2–2 (c)
$$\begin{aligned}
\mu_X &= 0.5(0) + 0.5(1) = 0.5, \\
\mu_Y &= 0.2(0) + 0.6(1) + 0.2(2) = 1, \\
\sigma_X^2 &= (0 - 0.5)^2(0.5) + (1 - 0.5)^2(0.5) = 0.25, \\
\sigma_Y^2 &= (0 - 1)^2(0.2) + (1 - 1)^2(0.6) + (2 - 1)^2(0.2) = 0.4, \\
\mathrm{Cov}(X, Y) &= (0)(0)(0.2) + (1)(2)(0.2) + (0)(1)(0.3) + \\
&\qquad (1)(1)(0.3) - (0.5)(1) = 0.2, \\
\rho &= \frac{0.2}{\sqrt{0.25}\sqrt{0.4}} = \sqrt{0.4};
\end{aligned}$$

(d) $y = 1 + \sqrt{0.4}\left(\dfrac{\sqrt{0.4}}{\sqrt{0.25}} \right)(x - 0.5) = 0.6 + 0.8x.$

4.2–4 $E[a_1 u_1(X_1, X_2) + a_2 u_2(X_1, X_2)]$

$$\begin{aligned}
&= \sum_{(x_1, x_2) \in R} \sum [a_1 u_1(x_1, x_2) + a_2 u_2(x_1, x_2)] f(x_1, x_2) \\
&= a_1 \sum_{(x_1, x_2) \in R} \sum u_1(x_1, x_2) f(x_1, x_2) + a_2 \sum_{(x_1, x_2) \in R} \sum u_2(x_1, x_2) f(x_1, x_2) \\
&= a_1 E[u_1(X_1, X_2)] + a_2 E[u_2(X_1, X_2)].
\end{aligned}$$

4.2–6 Note that X is $b(3, 1/6)$, Y is $b(3, 1/2)$ so

(a) $E(X) = 3(1/6) = 1/2;$

(b) $E(Y) = 3(1/2) = 3/2;$

(c) $\mathrm{Var}(X) = 3(1/6)(5/6) = 5/12;$

(d) $\mathrm{Var}(Y) = 3(1/2)(1/2) = 3/4;$

(e) $\mathrm{Cov}(X,Y)$ $=$ $0 + (1)f(1,1) + 2f(1,2) + 2f(2,1) - (1/2)(3/2)$

$\qquad\qquad\qquad = (1)(1/6) + 2(1/8) + 2(1/24) - 3/4$

$\qquad\qquad\qquad = -1/4;$

(f) $\rho = \dfrac{-1/4}{\sqrt{\dfrac{5}{12}\cdot\dfrac{3}{4}}} = \dfrac{-1}{\sqrt{5}}.$

4.2–8 (b)

$\frac{1}{6}$	2	$\bullet \frac{1}{6}$		
$\frac{2}{6}$	1	$\bullet \frac{1}{6}$	$\bullet \frac{1}{6}$	
$\frac{3}{6}$	0	$\bullet \frac{1}{6}$	$\bullet \frac{1}{6}$	$\bullet \frac{1}{6}$
		0	1	2
		$\frac{3}{6}$	$\frac{2}{6}$	$\frac{1}{6}$

(c) $\mathrm{Cov}(X,Y) = (1)(1)\left(\dfrac{1}{6}\right) - \left(\dfrac{2}{3}\right)\left(\dfrac{2}{3}\right) = \dfrac{1}{6} - \dfrac{4}{9} = \dfrac{-5}{18};$

(d) σ_X^2 $=$ $\dfrac{2}{6} + \dfrac{4}{6} - \left(\dfrac{2}{3}\right)^2 = \dfrac{5}{9} = \sigma_Y^2,$

$\qquad \rho$ $=$ $\dfrac{-5/18}{\sqrt{(5/9)(5/9)}} = -\dfrac{1}{2};$

(e) y $=$ $\dfrac{2}{3} - \dfrac{1}{2}\sqrt{\dfrac{5/9}{5/9}}\left(x - \dfrac{2}{3}\right)$

$\qquad y$ $=$ $1 - \dfrac{1}{2}x.$

4.2–10 (a) $f_1(x)$ $=$ $\displaystyle\int_0^x 2\,dy = 2x, \qquad 0 \le x \le 1,$

$\qquad\quad f_2(y)$ $=$ $\displaystyle\int_y^1 2\,dx = 2(1-y), \qquad 0 \le y \le 1;$

(b) $\qquad\quad \mu_X$ $=$ $\displaystyle\int_0^1 2x^2\,dx = \dfrac{2}{3},$

$\qquad\quad \mu_Y$ $=$ $\displaystyle\int_0^1 2y(1-y)\,dy = \dfrac{1}{3},$

$\qquad\quad \sigma_X^2$ $=$ $E(X^2) - (\mu_X)^2 = \displaystyle\int_0^1 2x^3\,dx - \left(\dfrac{2}{3}\right)^2 = \dfrac{1}{2} - \dfrac{4}{9} = \dfrac{1}{18},$

$\qquad\quad \sigma_Y^2$ $=$ $E(Y^2) - (\mu_Y)^2 = \displaystyle\int_0^1 2y^2(1-y)\,dy - \left(\dfrac{1}{3}\right)^2 = \dfrac{1}{6} - \dfrac{1}{9} = \dfrac{1}{18},$

$\quad \mathrm{Cov}(X,Y)$ $=$ $E(XY) - \mu_X\mu_Y = \displaystyle\int_0^1\int_0^x 2xy\,dy\,dx - \left(\dfrac{2}{3}\right)\left(\dfrac{1}{3}\right) = \dfrac{1}{4} - \dfrac{2}{9} = \dfrac{1}{36},$

$\qquad\qquad \rho$ $=$ $\dfrac{1/36}{\sqrt{1/18}\,\sqrt{1/18}} = \dfrac{1}{2};$

(c) $y = \dfrac{1}{3} + \dfrac{1}{2}\sqrt{\dfrac{1/18}{1/18}}\left(x - \dfrac{2}{3}\right) = 0 + \dfrac{1}{2}x.$

4.2–12 **(a)** $\quad f_1(x) \;=\; \displaystyle\int_x^1 8xy\,dy = 4x(1-x^2), \qquad 0 \le x \le 1,$

$\qquad\qquad f_2(y) \;=\; \displaystyle\int_0^y 8xy\,dx = 4y^3, \qquad 0 \le x \le 1;$

(b) $\quad \mu_X \;=\; \displaystyle\int_0^1 x4x(1-x^2)\,dx = \dfrac{8}{15},$

$\qquad\quad \mu_Y \;=\; \displaystyle\int (y * 4y^3\,dy = \dfrac{4}{5},$

$\qquad\quad \sigma_X^2 \;=\; \displaystyle\int_0^1 (x - 8/15)^2 4x(1-x^2)\,dx = \dfrac{11}{225},$

$\qquad\quad \sigma_Y^2 \;=\; \displaystyle\int ((y - 4/5)^2 * 4y^3\,dy = \dfrac{2}{75},$

$\qquad \mathrm{Cov}(X,Y) \;=\; \displaystyle\int_0^1\int_x^1 (x - 8/15)(y - 4/5)8xy\,dy\,dx \;=\; \dfrac{4}{225},$

$\qquad\qquad \rho \;=\; \dfrac{4/225}{\sqrt{(11/225)(2/75)}} = \dfrac{2\sqrt{66}}{33};$

(c) $\; y = \dfrac{20}{33} + \dfrac{4x}{11}.$

4.3 Conditional Distributions

4.3–2

	1	3	
2	$\frac{1}{4}$	$\frac{3}{4}$	$g(x\,\vert\,2)$
1	$\frac{3}{4}$	$\frac{1}{4}$	$g(x\,\vert\,1)$
	1	2	

equivalently, $g(x\,|\,y) = \dfrac{3 - 2|x - y|}{4},$

$x = 1, 2,\ \text{for } y = 1 \text{ or } 2;$

	$h(y\,\vert\,1)$	$h(y\,\vert\,2)$
2	$\frac{1}{4}$	$\frac{3}{4}$
1	$\frac{3}{4}$	$\frac{1}{4}$
	1	2

equivalently, $h(y\,|\,x) = \dfrac{3 - 2|x - y|}{4},$

$y = 1, 2,\ \text{for } x = 1 \text{ or } 2;$

$\mu_{X|1} = 5/4,\ \mu_{X|2} = 7/4,\ \mu_{Y|1} = 5/4,\ \mu_{Y|2} = 7/4;$

$\sigma_{X|1}^2 = \sigma_{X|2}^2 = \sigma_{Y|1}^2 = \sigma_{Y|2}^2 = 3/16.$

4.3–4 **(a)** X is $b(400, 0.75)$;

(b) $E(X) = 300$, $\text{Var}(X) = 75$;

(c) $b(300, 2/3)$;

(d) $E(Y) = 200$, $\text{Var}(Y) = 200/3$.

4.3–6 **(a)** $P(X = 500) = 0.40$, $P(Y = 500) = 0.35$,

$P(Y = 500 \,|\, X = 500) = 0.50$, $P(Y = 100 \,|\, X = 500) = 0.25$;

(b) $\mu_X = 485$, $\mu_Y = 510$, $\sigma_X^2 = 118{,}275$, $\sigma_Y^2 = 130{,}900$;

(c) $\mu_{X|Y=100} = 2400/7$, $\mu_{Y|X=500} = 525$;

(d) $\text{Cov}(X, Y) = 49650$;

(e) $\rho = 0.399$.

4.3–8 **(a)** X and Y have a trinomial distribution with $n = 30$, $p_1 = 1/6$, $p_2 = 1/6$.

(b) The conditional p.d.f. of X, given $Y = y$, is

$$b\left(n - y,\ \frac{p_1}{1 - p_2}\right) = b(30 - y,\ 1/5).$$

(c) Since $E(X) = 5$ and $\text{Var}(X) = 25/6$, $E(X^2) = \text{Var}(X) + [E(X)]^2 = 25/6 + 25 = 175/6$. Similarly, $E(Y) = 5$, $\text{Var}(Y) = 25/6$, $E(Y^2) = 175/6$. The correlation coefficient is

$$\rho = -\sqrt{\frac{(1/6)(1/6)}{(5/6)(5/6)}} = -1/5$$

so

$$E(XY) = -1/5\sqrt{(25/6)(25/6)} + (5)(5) = 145/6.$$

Thus

$$E(X^2 - 4XY + 3Y^2) = \frac{175}{6} - 4\left(\frac{145}{6}\right) + 3\left(\frac{175}{6}\right) = \frac{120}{6} = 20.$$

4.3–10 **(a)** $f(x, y) = 1/[10(10 - x)]$, $\qquad x = 0, 1, \cdots, 9$, $\ y = x, x + 1, \cdots, 9$;

(b) $f_2(y) = \displaystyle\sum_{x=0}^{y} \frac{1}{10(10 - x)}$, $\qquad y = 0, 1, \ldots, 9$;

(c) $E(Y|x) = (x + 9)/2$.

4.3–12 From Example 4.1–10, $\mu_X = \dfrac{1}{3}$, $\mu_Y = \dfrac{2}{3}$, and $E(Y^2) = \dfrac{1}{2}$.

$$E(X^2) = \int_0^1 2x^2(1 - x)\, dx = \frac{1}{6}, \quad \sigma_X^2 = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}, \quad \sigma_Y^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18};$$

$$\text{Cov}(X, Y) = \int_0^1 \int_x^1 2xy\, dy\, dx - \left(\frac{1}{3}\right)\left(\frac{2}{3}\right) = \frac{1}{4} - \frac{2}{9} = \frac{1}{36},$$

so

$$\rho = \frac{1/36}{\sqrt{1/18}\,\sqrt{1/18}} = \frac{1}{2}.$$

4.3–14 (b)

$$
f_1(x) = \begin{cases} \displaystyle\int_0^x 1/8 \, dy &= x/8, & 0 \le x \le 2, \\[2ex] \displaystyle\int_{x-2}^x 1/8 \, dy &= 1/4, & 2 < x < 4, \\[2ex] \displaystyle\int_{x-2}^4 1/8 \, dy &= (6-x)/8, & 4 \le x \le 6; \end{cases}
$$

(c) $f_2(y) = \displaystyle\int_y^{y+2} 1/8 \, dx = 1/4, \qquad 0 \le y \le 4;$

(d)

$$
h(y \mid x) = \begin{cases} 1/x, & 0 \le y \le x, & 0 \le x \le 2, \\ 1/2, & x-2 < y < x, & 2 < x < 4, \\ 1/(6-x), & x-2 \le y \le 4, & 4 \le x \le 6; \end{cases}
$$

(e) $g(x \mid y) = 1/2, \qquad y \le x \le y+2;$

(f)

$$
E(Y \mid x) = \begin{cases} \displaystyle\int_0^x y\left(\frac{1}{x}\right) dy &= \dfrac{x}{2}, & 0 \le x \le 2, \\[2ex] \displaystyle\int_{x-2}^x y \cdot \frac{1}{2} \, dy &= \left[\dfrac{y^2}{4}\right]_{x-2}^x = x-1, & 2 < x < 4, \\[2ex] \displaystyle\int_{x-2}^4 \frac{y}{6-x} \, dy &= \left[\dfrac{y^2}{2(6-x)}\right]_{x-2}^4 = \dfrac{x+2}{2}, & 4 \le x < 6; \end{cases}
$$

(g) $E(X \mid y) = \displaystyle\int_y^{y+2} x \cdot \frac{1}{2} \, dx = \left[\dfrac{x^2}{4}\right]_y^{y+2} = y+1, \qquad 0 \le y \le 4;$
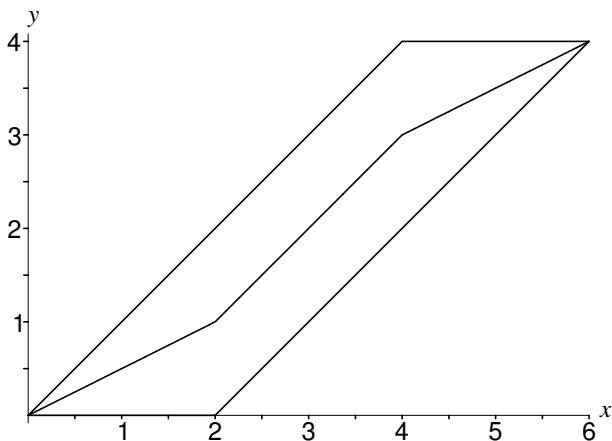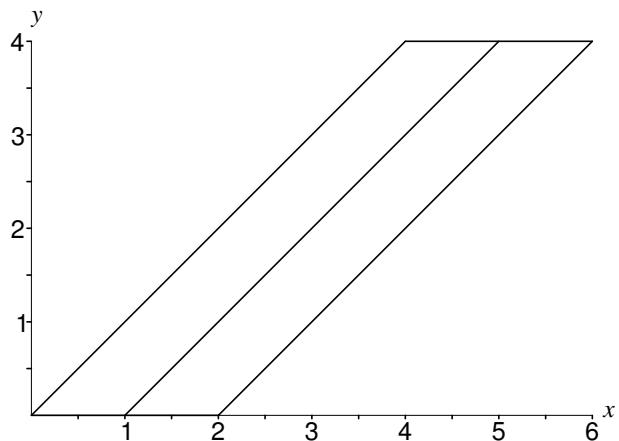
Figure 4.3–14: **(h)** $y = E(Y \mid x)$ **(i)** $x = E(X \mid y)$

4.3–16 (a) $h(y \mid x) = \dfrac{1}{x}, \qquad 0 < y < x, \ \ 0 < x < 1;$

(b) $E(Y \mid x) = \displaystyle\int_0^x \frac{y}{x} \, dy = \frac{x}{2};$

(c) $f(x,y) = h(y \mid x) f_1(x) = \left(\dfrac{1}{x}\right)(1) = \dfrac{1}{x}, \qquad 0 < y < x, \ \ 0 < x < 1;$

(d) $f_2(y) = \int_y^1 \frac{1}{x} \, dx = -\ln y, \qquad 0 < y < 1.$

4.3–18 (a) $f(x, y) = f_1(x)h(y \mid x) = 1 \cdot \frac{1}{x+1} = \frac{1}{x+1}, \qquad 0 < y < x+1, \; 0 < x < 1;$

(b) $E(Y \mid x) = \int_0^{x+1} y\left(\frac{1}{x+1}\right) dy = \left[\frac{y^2}{2(x+1)}\right]_0^{x+1} = \frac{x+1}{2};$

(c)
$$f_2(y) = \begin{cases} \int_0^1 \frac{1}{x+1} \, dx = [\ln(x+1)]_0^1 = \ln 2, & 0 < y < 1, \\[2mm] \int_{y-1}^1 \frac{1}{x+1} \, dx = [\ln(x+1)]_{y-1}^1 = \ln 2 - \ln y, & 1 < y < 2. \end{cases}$$

4.3–20 (a) In order for x, y, and $1 - x - y$ to be the sides of a triangle, it must be true that

$$x + y > 1 - x - y \qquad \text{or} \qquad 2x + 2y > 1;$$

$$x + 1 - x - y > y \qquad \text{or} \qquad y < 1/2;$$

$$y + 1 - x - y > x \qquad \text{or} \qquad x < 1/2.$$
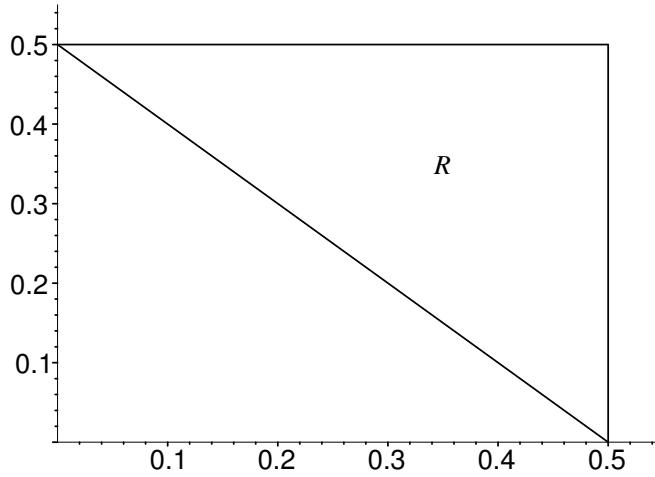
Figure 4.3–20: Set of possible values for x and y

(b) $f(x, y) = \frac{1}{1/8} = 8, \qquad \frac{1}{2} - x < y < \frac{1}{2}, \; 0 < x < \frac{1}{2};$

$$E(T) = \int_0^{1/2} \int_{1/2-x}^{1/2} \frac{1}{4}\sqrt{(2x + 2y - 1)(1 - 2x)(1 - 2y)} \, 8 \, dy \, dx$$

$$= \frac{\pi}{105} = 0.0299;$$

$$\sigma^2 = E(T^2) - [E(T)]^2$$

$$= \int_0^{1/2} \int_{1/2-x}^{1/2} \frac{1}{4}(2x + 2y - 1)(1 - 2x)(1 - 2y) \, 8 \, dy \, dx - \left[\frac{\pi}{105}\right]^2$$

$$= \frac{1}{960} - \frac{\pi^2}{11025} = 0.00014646.$$

(c) $f_1(x) = \int_{1/2-x}^{1/2} 8\,dy = 8x, \qquad 0 < x < \frac{1}{2};$

$$h(y \mid x) = \frac{f(x,y)}{f_1(x)} = \frac{8}{8x} = \frac{1}{x}, \qquad \frac{1}{2} - x < y < \frac{1}{2}, \ \ 0 < x < \frac{1}{2};$$

(d) The distribution function of X is

$$F_1(x) = \int_0^x 8t\,dt = 4x^2.$$

If a is the value of a $U(0,1)$ random variable (a random number), then let $\ a = 4x^2$ and

$$x = (1/2)\sqrt{a}$$

is an observation of X.

The conditional distribution function of Y, given $X = x$, is

$$G(y) = \int_{1/2-x}^{y} \frac{1}{x}\,dt = \frac{y}{x} - \frac{1}{2x} + 1.$$

If b is the value of a $U(0,1)$ random variable (a random number), then solving

$$b = G(y) = \frac{y}{x} - \frac{1}{2x} + 1$$

for y yields

$$y = xb - x + \frac{1}{2}$$

as an observation of Y.

Here is some Maple code for a simulation of the areas of 5000 triangles:

```
> for k from 1 to 5000 do
>    a := rng(); # rng() yields a random number
>    b := rng(); # rng() yields a random number
>    X := sqrt(a)/2;
>    Y := X*b + 1/2 - X;
>    Z := 1 - X - Y;
>    TT(k) := 1/4*sqrt((2*X + 2*Y - 1)*(1 - 2*X)*(1 - 2*Y));
>      # TT(k) finds the area of one triangle
> od:
> T := [seq(TT(k), k = 1 .. 5000)]: # put areas in a sequence
> tbar := Mean(T); # finds the sample mean
> tvar := Variance(T); # finds the sample variance
                 tbar := 0.02992759330
                 tvar := 0.0001469367443
```

(e) X is $U(0, 1/2)$ so $f_1(x) = 2, \ \ 0 < x < 1/2$; The conditional p.d.f. of Y, given $X = x$ is $U(1/2 - x, 1/2)$ so $h(y \mid x) = 1/x, \ \ 1/2 - x < y < 1/2$. Thus the joint p.d.f. of X and Y is

$$f(x,y) = 2\frac{1}{x} = \frac{2}{x}, \qquad \frac{1}{2} - x < y < \frac{1}{2}, \ \ 0 < x < \frac{1}{2}.$$

$$E(T) = \int_0^{1/2} \int_{1/2-x}^{1/2} \frac{1}{4} \sqrt{(2x+2y-1)(1-2x)(1-2y)}\, \frac{2}{x}\, dy\, dx$$

$$= \frac{\pi}{120} = 0.02618;$$

$$\sigma^2 = E(T^2) - [E(T)]^2$$

$$= \int_0^{1/2} \int_{1/2-x}^{1/2} \frac{1}{4}(2x+2y-1)(1-2x)(1-2y)\, \frac{2}{x}\, dy\, dx - \left[\frac{\pi}{120}\right]^2$$

$$= \frac{1}{1152} - \frac{\pi^2}{14400} = 0.00018267.$$

Here is some Maple code to simulate 5000 areas of random triangles:

```
> for k from 1 to 5000 do
>    a := rng();
>    b := rng();
>    X := a/2;
>    Y := X*b + 1/2 - X;
>    Z := 1 - X - Y;
>    TT(k) := 1/4*sqrt((2*X + 2*Y - 1)*(1 - 2*X)*(1 - 2*Y));
> od:
> T := [seq(TT(k), k = 1 .. 5000)]:
> tbar := Mean(T);
> tvar := Variance(T);
                       tbar := 0.02611458560
                       tvar := 0.0001812722807
```

4.4   The Bivariate Normal Distribution

4.4–2 $\displaystyle q(x,y) = \frac{[y - \mu_Y - \rho(\sigma_Y/\sigma_X)(x - \mu_X)]^2}{\sigma_Y^2(1-\rho^2)} + \frac{(x-\mu_X)^2}{\sigma_X^2}$

$$= \frac{1}{1-\rho^2}\left[\frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right.$$

$$\left. + \frac{\rho^2(x-\mu_X)^2}{\sigma_X^2} + (1-\rho^2)\frac{(x-\mu_X)^2}{\sigma_X^2}\right]$$

$$= \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]$$

4.4–4 (a) $E(Y \mid X = 72) = 80 + \dfrac{5}{13}\left(\dfrac{13}{10}\right)(72 - 70) = 81;$

(b) $\mathrm{Var}(Y \mid X = 72) = 169\left[1 - \left(\dfrac{5}{13}\right)^2\right] = 144;$

(c) $P(Y \le 84 \mid X = 72) = P\left(Z \le \dfrac{84 - 81}{12}\right) = \Phi(0.25) = 0.5987.$

4.4–6 (a) $P(18.5 < Y < 25.5) = \Phi(0.8) - \Phi(-1.2) = 0.6730;$

(b) $E(Y \mid x) = 22.7 + 0.78(3.5/4.2)(x - 22.7) = 0.65x + 7.945;$

(c) $\mathrm{Var}(Y \mid x) = 12.25(1 - 0.78^2) = 4.7971;$

(d) $P(18.5 < Y < 25.5 \,|\, X = 23) = \Phi(1.189) - \Phi(-2.007) = 0.8828 - 0.0224 = 0.8604;$

(e) $P(18.5 < Y < 25.5 \,|\, X = 25) = \Phi(0.596) - \Phi(-2.60) = 0.7244 - 0.0047 = 0.7197.$

(f)
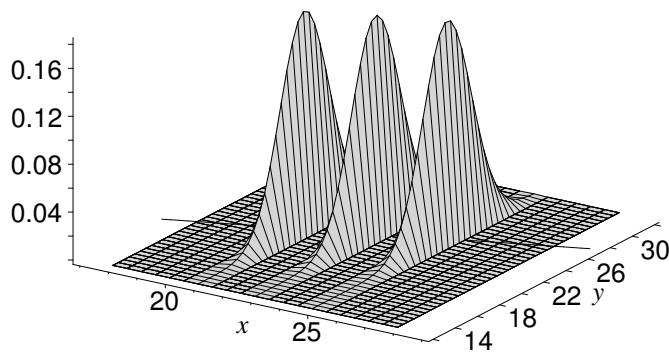
Figure 4.4–6: Conditional p.d.f.s of Y, given $x = 21, 23, 25$

4.4–8 (a) $P(13.6 < Y < 17.2) = \Phi(0.55) - \Phi(-0.35) = 0.3456;$

(b) $E(Y \,|\, x) = 15 + 0(4/3)(x - 10) = 15;$

(c) $\operatorname{Var}(Y \,|\, x) = 16(1 - 0^2) = 16;$

(d) $P(13.6 < Y < 17.2 \,|\, X = 9.1) = 0.3456.$

4.4–10 (a) $P(2.80 \le Y \le 5.35) = \Phi(1.50) - \Phi(0) = 0.4332;$

(b) $E(Y \,|\, X = 82.3) = 2.80 + (-0.57)\left(\dfrac{1.7}{10.5}\right)(82.3 - 72.30) = 1.877;$

$$\operatorname{Var}(Y \,|\, X = 82.3) = 2.89[1 - (-0.57)^2] = 1.9510;$$

$$
\begin{aligned}
P(2.76 \le Y \le 5.34 \,|\, X = 82.3) &= \Phi(2.479) - \Phi(0.632) \\
&= 0.9934 - 0.7363 = 0.2571.
\end{aligned}
$$

4.4–12 (a) $P(0.205 \le Y \le 0.805) = \Phi(1.57) - \Phi(1.17) = 0.0628;$

(b)
$$
\begin{aligned}
\mu_{Y|X=20} &= -1.55 - 0.60\left(\dfrac{1.5}{4.5}\right)(20 - 15) = -2.55 : \\
\sigma^2_{Y|X=20} &= 1.5^2[1 - (-0.60)^2] = 1.44; \\
\sigma_{Y|X=20} &= 1.2;
\end{aligned}
$$

$$P(0.21 \le Y \le 0.81 \,|\, X = 20) = \Phi(2.8) - \Phi(2.3) = 0.0081.$$

4.4–14 (a) $E(Y \,|\, X = 15) = 1.3 + 0.8\left(\dfrac{0.1}{2.5}\right)(15 - 14.1) = 1.3288;$

$$\operatorname{Var}(Y \,|\, X = 15) = 0.1^2(1 - 0.8^2) = 0.0036;$$

$$P(Y > 1.4 \,|\, X = 15) = 1 - \Phi\left(\dfrac{1.4 - 1.3288}{0.06}\right) = 0.031;$$

(b) $E(X \,|\, Y = 1.4) = 14.1 + 0.8\left(\dfrac{2.5}{0.1}\right)(1.4 - 1.3) = 16.1;$

$$\operatorname{Var}(X \,|\, Y = 1.4) = 2.5^2(1 - 0.8^2) = 2.25;$$

$$P(X > 15 \,|\, Y = 1.4) = 1 - \Phi\left(\dfrac{15 - 16.1}{1.5}\right) = 1 - \Phi(-0.7333) = 0.7683.$$

Chapter 5

Distributions of Functions of Random Variables

5.1 Distributions of Functions of a Random Variable

5.1–2 Here $x = \sqrt{y}$, $D_y(x) = 1/2\sqrt{y}$ and $0 < x < \infty$ maps onto $0 < y < \infty$. Thus

$$g(y) = \sqrt{y} \left| \frac{1}{2\sqrt{y}} \right| = \frac{1}{2} e^{-y/2}, \qquad 0 < y < \infty.$$

5.1–4 (a)

$$F(x) = \begin{cases} 0, & x < 0, \\ \int_0^x 2t\,dt = x^2, & 0 \le x < 1, \\ 1, & 1 \le x, \end{cases}$$

(b) Let $y = x^2$; so $x = \sqrt{y}$. Let Y be $U(0,1)$; then $X = \sqrt{Y}$ has the given x-distribution.

(c) Repeat the procedure outlined in part (b) 10 times.

(d) Order the 10 values of x found in part (c), say $x_1 < x_2 < \cdots < x_{10}$ and plot the 10 points $(x_i, \sqrt{i/11})$, $i = 1, 2, \ldots, 10$, where $11 = n + 1$.

5.1–6 It is easier to note that

$$\frac{dy}{dx} = \frac{e^{-x}}{(1 + e^{-x})^2} \qquad \text{and} \qquad \frac{dx}{dy} = \frac{(1 + e^{-x})^2}{e^{-x}}.$$

Say the solution of x in terms of y is given by x^*. Then the p.d.f. of Y is

$$g(y) = \frac{e^{-x^*}}{(1 + e^{-x^*})^2} \left| \frac{(1 + e^{-x^*})^2}{e^{-x^*}} \right| = 1, \qquad 0 < y < 1,$$

as $-\infty < x < \infty$ maps onto $0 < y < 1$. Thus Y is $U(0,1)$.

5.1–8 $x = \left(\dfrac{y}{5}\right)^{10/7}$

$\dfrac{dx}{dy} = \dfrac{10}{7}\left(\dfrac{y}{5}\right)^{3/7}\left(\dfrac{1}{5}\right)$

$f(x) = e^{-x}, \qquad 0 < x < \infty$

$g(y) = e^{-(y/5)^{10/7}}\left(\dfrac{2}{7}\right)\left(\dfrac{1}{5}\right)^{3/7}y^{3/7}$

$\qquad = \dfrac{10/7}{5^{10/7}}y^{3/7}e^{-(y/5)^{10/7}}, \qquad 0 < y < \infty.$

(The reason for writing the p.d.f. in that form is because Y has a Weibull distribution with $\alpha = 10/7$ and $\beta = 5$.)

5.1–10 Since $-1 < x < 3$, we have $0 \le y < 9$.

When $0 < y < 1$, then

$$x_1 = -\sqrt{y}, \quad \dfrac{dx_1}{dy} = \dfrac{-1}{2\sqrt{y}}; \qquad x_2 = \sqrt{y}, \quad \dfrac{dx_2}{dy} = \dfrac{1}{2\sqrt{y}}.$$

When $1 < y < 9$, then

$$x = \sqrt{y}, \quad \dfrac{dx}{dy} = \dfrac{1}{2\sqrt{y}}.$$

Thus

$$g(y) = \begin{cases} \dfrac{1}{4}\cdot\left|\dfrac{-1}{2\sqrt{y}}\right| + \dfrac{1}{4}\cdot\left|\dfrac{1}{2\sqrt{y}}\right| = \dfrac{1}{4\sqrt{y}}, & 0 < y < 1, \\[3mm] \dfrac{1}{4}\cdot\left|\dfrac{1}{2\sqrt{y}}\right| \qquad\qquad = \dfrac{1}{8\sqrt{y}}, & 1 \le y < 9. \end{cases}$$

5.1–12 $E(X) = \displaystyle\int_{-\infty}^{\infty}\dfrac{x}{\pi(1+x^2)}\,dx$

$\qquad = \displaystyle\lim_{a\to-\infty}\left[\dfrac{1}{2\pi}\ln(1+x^2)\right]_a^0 + \lim_{b\to+\infty}\left[\dfrac{1}{2\pi}\ln(1+x^2)\right]_0^b$

$\qquad = \dfrac{1}{2\pi}\left[\displaystyle\lim_{a\to-\infty}\{-\ln(1+a^2)\} + \lim_{b\to+\infty}\ln(1+b^2)\right].$

$E(X)$ does not exist because neither of these limits exists.

5.1–14 X is $N(0,1)$ and $Y = |X|$. Let

$$x_1 = -y, \qquad -\infty < x_1 < 0,$$

$$x_2 = y, \qquad 0 < x_2 < \infty.$$

Then

$$\dfrac{dx_1}{dy} = -1 \qquad\text{and}\qquad \dfrac{dx_2}{dy} = 1.$$

Thus the p.d.f. of Y is

$$g(y) = \dfrac{1}{\sqrt{2\pi}}e^{-(-y^2)}|-1| + \dfrac{1}{\sqrt{2\pi}}e^{-y^2}|1| = \dfrac{2}{\sqrt{2\pi}}e^{-y^2}, \qquad 0 < y < \infty.$$

5.2 Transformations of Two Random Variables

5.2-2 **(a)** The joint p.d.f. of X_1 and X_2 is

$$f(x_1, x_2) = \frac{1}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)2^{(r_1+r_2)/2}}x_1^{r_1/2-1}x_2^{r_2/2-1}e^{-(x_1+x_2)/2},$$

$$0 < x_1 < \infty, \quad 0 < x_2 < \infty.$$

Let $Y_1 = (X_1/r_1)/(X_2/r_2)$ and $Y_2 = X_2$. The Jacobian of the transformation is $(r_1/r_2)y_2$. Thus

$$g(y_1, y_2) = \frac{1}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)2^{(r_1+r_2)/2}}\left(\frac{r_1x_1x_2}{r_2}\right)^{r_1/2-1}x_2^{r_2/2-1}e^{-(y_2/2)(r_1y_1/r_2+1)}\left(\frac{r_1y_2}{r_2}\right),$$

$$0 < y_1 < \infty, \quad 0 < y_2 < \infty.$$

(b) The marginal p.d.f. of Y_1 is $g_1(y_1) = \int_0^\infty g(y_1, y_2)\,dy_2$.

Make the change of variables $w = \dfrac{y_2}{2}\left(\dfrac{r_1y_1}{r_2}+1\right)$. Then

$$g_1(y_1) = \frac{\Gamma\left(\dfrac{r_1+r_2}{2}\right)\left(\dfrac{r_1}{r_2}\right)^{r_1/2}y_1^{r_1/2-1}}{\Gamma\left(\dfrac{r_1}{2}\right)\Gamma\left(\dfrac{r_2}{2}\right)\left(1+\dfrac{r_1y_1}{r_2}\right)^{(r_1+r_2)/2}}\cdot 1, \quad 0 < y_1 < \infty.$$

5.2-4 **(a)** $F_{0.05}(9, 24) = 2.30$;

(b) $F_{0.95}(9, 24) = \dfrac{1}{F_{0.05}(24, 9)} = \dfrac{1}{2.90} = 0.3448$;

(c) $P(W < 0.277) = P\left(\dfrac{1}{W} > \dfrac{1}{0.277}\right) = P\left(\dfrac{1}{W} > 3.61\right) = 0.025$;

$P(0.277 \le W \le 2.70) = P(W \le 2.70) - P(W \le 0.277) = 0.975 - 0.025 = 0.95.$

5.2-6
$$F(w) = P\left(\frac{X_1}{X_1 + X_2} \le w\right), \quad 0 < w < 1$$

$$= \int_0^\infty \int_{(1-w)x_1/w}^\infty \frac{x_1^{\alpha-1}x_2^{\beta-1}e^{-(x_1+x_2)/\theta}}{\Gamma(\alpha)\Gamma(\beta)\theta^{\alpha+\beta}}\,dx_2dx_1$$

$$f(w) = F'(w) = \int_0^\infty \frac{-x_1^{\alpha-1}[(1-w)x_1/w]^{\beta-1}e^{-[x_1+(1-w)x_1/w]/\theta}}{\Gamma(\alpha)\Gamma(\beta)\theta^{\alpha+\beta}}\left(\frac{-1}{w^2}\right)x_1\,dx_1$$

$$= \frac{1}{\Gamma(\alpha)\Gamma(\beta)}\frac{(1-w)^{\beta-1}}{w^{\beta+1}}\int_0^\infty \frac{x_1^{\alpha+\beta-1}e^{-x_1/\theta w}}{\theta^{\alpha+\beta}}\,dx_1$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta)}\frac{(\theta w)^{\alpha+\beta}}{w^{\beta+1}}\frac{(1-w)^{\beta-1}}{\theta^{\alpha+\beta}}$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}w^{\alpha-1}(1-w)^{\beta-1}, \quad 0 < w < 1.$$

5.2-8 **(a)** $E(X)$ $=$ $\displaystyle\int_0^1 x\, \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}\, dx$

$\qquad\qquad\quad = \displaystyle\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+1)}{\Gamma(\alpha)\Gamma(\alpha+\beta+1)} \cdot \int_0^1 \frac{\Gamma(\alpha+1+\beta)}{\Gamma(\alpha+1)\Gamma(\beta)}\, x^{\alpha+1-1}(1-x)^{\beta-1}\, dx$

$\qquad\qquad\quad = \displaystyle\frac{(\alpha)\Gamma(\alpha)\Gamma(\alpha+\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)\Gamma(\alpha)}$

$\qquad\qquad\quad = \displaystyle\frac{\alpha}{\alpha+\beta};$

$\qquad\quad E(X^2) = \displaystyle\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+2)}{\Gamma(\alpha)\Gamma(\alpha+2+\beta)} \int_0^1 \frac{\Gamma(\alpha+2+\beta)}{\Gamma(\alpha+2)\Gamma(\beta)}\, x^{\alpha+2-1}(1-x)^{\beta-1}\, dx$

$\qquad\qquad\quad = \displaystyle\frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}.$

Thus

$\qquad\quad \sigma^2 = \displaystyle\frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)} - \frac{\alpha^2}{(\alpha+\beta)^2} = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}.$

(b) $f(x) = \displaystyle\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}$

$\qquad\quad f'(x) = \displaystyle\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left[(\alpha-1)\, x^{\alpha-2}\,(1-x)^{\beta-1} - (\beta-1)\, x^{\alpha-1}\,(1-x)^{\beta-2} \right].$

Set $f'(x)$ equal to zero and solve for x:

$\qquad\quad \displaystyle\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-2}\,(1-x)^{\beta-2}\, [(\alpha-1)(1-x) - (\beta-1)\, x] \;=\; 0$

$\qquad\qquad\qquad\qquad\qquad\qquad \alpha - \alpha x - 1 + x - \beta x + x \;=\; 0$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\alpha+\beta-2)x \;=\; \alpha - 1$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad x \;=\; \displaystyle\frac{\alpha-1}{\alpha+\beta-2}.$

5.2-10 Use integration by parts two times to show

$\qquad\quad \displaystyle\int_0^p \frac{6!}{3!2!}\, y^3(1-y)^2 dy \;=\; \left[\binom{6}{4} y^4(1-y)^2 + \binom{6}{5} y^5(1-y)^1 + \binom{6}{6} y^6(1-y)^0 \right]_0^p$

$\qquad\qquad\qquad\qquad\qquad\quad = \displaystyle\sum_{y=4}^6 \binom{n}{y} p^y (1-p)^{6-y}.$

5.2-12 **(a)** $w_1 = 2x_1$ and $\displaystyle\frac{dw_1}{dx_1} = 2$. Thus

$\qquad\qquad\qquad\qquad f(x_1) = \displaystyle\frac{2}{\pi(1+4x_1^2)}, \qquad -\infty < x_1 < \infty.$

(b) For $x_2 = y_1 - y_2$, $x_1 = y_2$, $|J| = 1$. Thus

$\qquad\qquad\qquad\qquad g(y_1, y_2) = f(y_2) f(y_1 - y_2), \qquad -\infty < y_i < \infty, \; i = 1, 2.$

(c) $g_1(y_1) = \displaystyle\int_{-\infty}^{\infty} f(y_2) f(y_1 - y_2)\, dy_2.$

(d) $\quad g_1(y_1) \quad = \quad \displaystyle\int_{-\infty}^{\infty} \frac{2}{\pi[1 + 4y_2^2]} \cdot \frac{2}{\pi[1 + 4(y_1 - y_2)^2]}\, dy_2 = \int_{-\infty}^{\infty} h(y_2)\, dy_2$

$$= \quad \frac{4}{\pi^2} \int_{-\infty}^{\infty} \frac{1}{[1 + 2iy_2][1 - 2iy_2]} \cdot \frac{1}{[1 + 2i(y_1 - y_2)][1 - 2i(y_1 - y_2)]}\, dy_2$$

$$= \quad \frac{4}{\pi^2} \int_{-\infty}^{\infty} \frac{1}{2i} \cdot \frac{1}{y_2 - \frac{i}{2}} \cdot \frac{-1}{2i} \cdot \frac{1}{y_2 + \frac{i}{2}} \cdot \frac{-1}{2i} \cdot \frac{1}{y_2 - (y_1 - \frac{i}{2})} \cdot \frac{1}{2i} \cdot \frac{1}{y_2 - (y_1 + \frac{i}{2})}\, dy_2$$

$$= \quad \frac{4(2\pi i)}{\pi^2} \left[\mathrm{Res}\left(h(y_2); y_2 = \frac{i}{2} \right) + \mathrm{Res}\left(h(y_2); y_2 = y_1 + \frac{i}{2} \right) \right]$$

$$= \quad \frac{8\pi i}{\pi^2} \frac{1}{16} \left[\frac{1}{i} \cdot \frac{1}{i - y_1} \cdot \frac{1}{-y_1} + \frac{1}{y_1} \cdot \frac{1}{y_1 + i} \cdot \frac{1}{i} \right]$$

$$= \quad \frac{1}{2\pi} \cdot \frac{1}{y_1} \left[\frac{1}{y_1 - i} + \frac{1}{y_1 + i} \right] = \frac{1}{2\pi} \cdot \frac{1}{y_1} \left[\frac{y_1 + i + y_1 - i}{(y_1 - i)(y_1 + i)} \right]$$

$$= \quad \frac{1}{\pi(1 + y_1^2)}.$$

A *Maple* solution for Exercise 5.2-12:

```
>f := x-> 2/Pi/(1 + 4*x^2);
```

$$f := x-> 2\,\frac{1}{\pi\,(1 + 4\,x^2)}$$

```
>simplify(int(f(y[2])*f(y[1]-y[2]),y[2]=-infinity..infinity));
```

$$\frac{1}{\pi\,(1 + y_1^2)}$$

A *Mathematica* solution for Exercise 5.2-12:

```
In[1]:=
        f[x_] := 2/(Pi*(1 + 4(x)^2))
        g[y1_,y2_] := f[y2]*f[y1-y2]
In[3]:=
        Integrate[g[y1,y2], {y2, -Infinity,Infinity}]
Out[3]=
                1
          -------------
                2
          Pi + Pi y1
```

5.2-14 The joint p.d.f. is

$$h(x, y) = \frac{x}{5^3} e^{-(x+y)/5}, \qquad 0 < x < \infty, \;\; 0 < y < \infty;$$

$$z \;\; = \;\; \frac{x}{y}, \qquad w \;\; = \;\; y$$

$$x \;\; = \;\; zw, \qquad y \;\; = \;\; w$$

The Jacobian is

$$J \quad = \quad \begin{vmatrix} w & z \\ 0 & 1 \end{vmatrix} = w;$$

The joint p.d.f. of Z and W is

$$f(z, w) = \frac{zw}{5^3} e^{-(z+1)w/5} w, \qquad 0 < z < \infty, \ \ 0 < w < \infty;$$

The marginal p.d.f. of Z is

$$
\begin{aligned}
f_1(z) &= \int_0^\infty \frac{zw}{5^3} e^{-(z+1)w/5} w \, dw \\
&= \frac{\Gamma(3)z}{5^3} \left(\frac{5}{z+1}\right)^3 \int_0^\infty \frac{w^{3-1}}{\Gamma(3)(5/[z+1])^3} e^{-w/(5/[z+1])} \, dw \\
&= \frac{2z}{(z+1)^3}, \qquad 0 < z < \infty.
\end{aligned}
$$

5.2-16 $\alpha = 24$, $\beta = 6$, $\gamma = 42$ is reasonable, but other answers around this one are acceptable.

5.3 Several Independent Random Variables

5.3–2 **(a)** $\quad P(X_1 = 2, X_2 = 4) = \left[\dfrac{3!}{2!1!}\left(\dfrac{1}{2}\right)^2\left(\dfrac{1}{2}\right)^1\right]\left[\dfrac{5!}{4!1!}\left(\dfrac{1}{2}\right)^4\left(\dfrac{1}{2}\right)^1\right]$

$$= \frac{15}{2^8} = \frac{15}{256}.$$

(b) $\{X_1 + X_2 = 7\}$ can occur in the two mutually exclusive ways: $\{X_1 = 3, X_2 = 4\}$ and $\{X_1 = 2, X_2 = 5\}$. The sum of the probabilities of the two latter events is

$$\left[\frac{3!}{3!0!}\left(\frac{1}{2}\right)^3\right]\left[\frac{5!}{4!1!}\left(\frac{1}{2}\right)^5\right] + \left[\frac{3!}{2!1!}\left(\frac{1}{2}\right)^3\right]\left[\frac{5!}{5!0!}\left(\frac{1}{2}\right)^5\right] = \frac{5+3}{2^8} = \frac{1}{32}.$$

5.3–4 **(a)** $\quad \left(\displaystyle\int_{0.5}^{1.0} 2e^{-2x_1} \, dx_1\right)\left(\displaystyle\int_{0.7}^{1.2} 2e^{-2x_2} \, dx_2\right)$
$$
\begin{aligned}
&= (e^{-1} - e^{-2})(e^{-1.4} - e^{-2.4}) \\
&= (0.368 - 0.135)(0.247 - 0.091) \\
&= (0.233)(0.156) = 0.036.
\end{aligned}
$$

(b) $E(X_1) = E(X_2) = 0.5$,

$$E[X_1(X_2 - 0.5)^2] = E(X_1)\text{Var}(X_2) = (0.5)(0.25) = 0.125.$$

5.3–6 $\quad E(X) = \displaystyle\int_0^1 x6x(1-x)\,dx = \int_0^1 (6x^2 - 6x^3)\,dx = \left[2x^3 - \left(\frac{3}{2}\right)x^4\right]_0^1 = \frac{1}{2};$

$\quad E(X^2) = \displaystyle\int_0^1 (6x^3 - 6x^4)\,dx = \left[\left(\frac{3}{2}\right)x^4 - \left(\frac{6}{5}\right)x^5\right]_0^1 = \frac{3}{10}.$

Thus

$$\mu_X = \frac{1}{2}; \ \sigma_X^2 = \frac{3}{10} - \frac{1}{4} = \frac{1}{20}, \ \text{and}$$

$$\mu_Y = \frac{1}{2} + \frac{1}{2} = 1; \ \sigma_Y^2 = \frac{1}{20} + \frac{1}{20} = \frac{1}{10}.$$

5.3–8 Let $Y = \max(X_1, X_2)$. Then

$$
\begin{aligned}
G(y) &= [P(X \le y)]^2 \\[2mm]
&= \left[\int_1^y \frac{4}{x^5}\, dx \right]^2 \\[2mm]
&= \left[1 - \frac{1}{y^4} \right]^2, \qquad 1 < y < \infty \\[2mm]
g(y) &= G'(y) \\[2mm]
&= 2 \left(1 - \frac{1}{y^4} \right) \left(\frac{4}{y^5} \right), \qquad 1 < y < \infty; \\[2mm]
E(Y) &= \int_1^\infty y \cdot 2 \left(1 - \frac{1}{y^4} \right) \left(\frac{4}{y^5} \right) dy \\[2mm]
&= \int_1^\infty 8 \left[y^{-4} - y^{-8} \right] dy \\[2mm]
&= \frac{32}{21}.
\end{aligned}
$$

5.3–10 **(a)** $P(X_1 = 1)P(X_2 = 3)P(X_3 = 1) = \left(\frac{3}{4} \right) \left[\left(\frac{3}{4} \right) \left(\frac{1}{4} \right)^2 \right] \left(\frac{3}{4} \right) = \frac{27}{1024}$;

(b) $3P(X_1 = 3, X_2 = 1, X_3 = 1) + 3P(X_1 = 2, X_2 = 2, X_3 = 1) =$

$$
3 \left(\frac{27}{1024} \right) + 3 \left(\frac{27}{1024} \right) = \frac{162}{1024};
$$

(c) $P(Y \le 2) = \left(\frac{3}{4} + \frac{3}{4} \cdot \frac{1}{4} \right)^3 = \left(\frac{15}{16} \right)^3$.

5.3–12 $P(1 < \min X_i) = [P(1 < X_i)]^3 = \left(\int_1^\infty e^{-x}\, dx \right)^3 = e^{-3} = 0.05$.

5.3–14 **(a)**
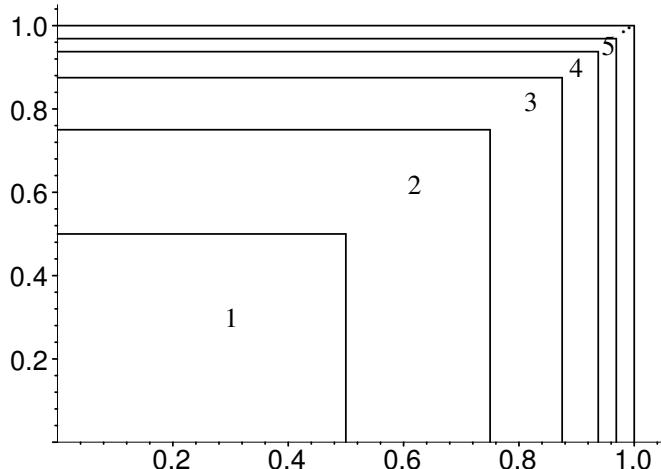
Figure 5.3–14: Selecting points randomly

(b) Note that $P(X = x)$ is the difference of the areas of two squares. Thus

$$
\begin{aligned}
P(X = x) &= \left(1 - \frac{1}{2^x}\right)^2 - \left(1 - \frac{1}{2^{x-1}}\right)^2 \\
&= 1 - \frac{2}{2^x} + \frac{1}{2^{2x}} - 1 + \frac{2}{2^{x-1}} - \frac{1}{2^{2x-2}} \\
&= \frac{-2^{x+1} + 1 + 2^{2+x} - 4}{2^{2x}} \\
&= \frac{2^{x+1} - 3}{2^{2x}} = \frac{2}{2^x} - \frac{3}{2^{2x}}, \qquad x = 1, 2, 3, \ldots.
\end{aligned}
$$

(c)
$$
\begin{aligned}
\sum_{x=1}^{\infty} \frac{2}{2^x} - \frac{3}{2^{2x}} &= \frac{1}{1 - 1/2} - \frac{3/4}{1 - 1/4} \\
&= 2 - 1 = 1;
\end{aligned}
$$

(d)
$$
\begin{aligned}
\mu &= \sum_{x=1}^{\infty} \left[\frac{2x}{2^x} - \frac{3x}{2^{2x}}\right] \\
&= \sum_{x=1}^{\infty} x \left(\frac{1}{2}\right)^{x-1} - \sum_{x=1}^{\infty} \frac{3}{4} x \left(\frac{1}{4}\right)^{x-1} \\
&= \frac{1}{(1 - 1/2)^2} - \frac{3/4}{(1 - 1/4)^2} \\
&= 4 - \frac{4}{3} = \frac{8}{3};
\end{aligned}
$$

(e)
$$
\begin{aligned}
E[X(X - 1)] &= \sum_{x=1}^{\infty} \left[\frac{2x(x-1)}{2^x} - \frac{3x(x-1)}{2^{2x}}\right] \\
&= \frac{1}{4} \sum_{x=2}^{\infty} 2x(x-1)\left(\frac{1}{2}\right)^{x-2} - \frac{1}{16} \sum_{x=2}^{\infty} 3x(x-1)\left(\frac{1}{4}\right)^{x-2} \\
&= \frac{2(2/4)}{(1 - 1/2)^3} - \frac{2(3/16)}{(1 - 1/4)^3} \\
&= 8 - \frac{8}{9} = \frac{64}{9}.
\end{aligned}
$$

So the variance is

$$
\sigma^2 = E[X(X - 1)] + E(X) - \mu^2 = \frac{64}{9} + \frac{8}{3} - \frac{64}{9} = \frac{8}{3}.
$$

5.3–16
$$
\begin{aligned}
P(Y > 1000) &= P(X_1 > 1000)P(X_2 > 1000)P(X_3 > 1000) \\
&= e^{-1}e^{-2/3}e^{-1/2} \\
&= e^{-13/6} = 0.1146.
\end{aligned}
$$

5.3–18
$$
\begin{aligned}
P(\max > 8) &= 1 - P(\max \le 8) \\
&= \left[\sum_{x=0}^{8} \binom{10}{x}(0.7)^x(0.3)^{10-x}\right]^3 \\
&= 1 - (1 - 0.1493)^3 = 0.3844.
\end{aligned}
$$

5.3–20 $\quad G(y) \quad = \quad P(Y \leq y) = P(X_1 \leq y) \cdots P(X_8 \leq y) = [P(X \leq y)]^8$

$$= \quad [y^{10}]^8 = y^{80}, \qquad 0 < y < 1;$$

$$P(0.9999 < Y < 1) = G(1) - G(0.9999) = 1 - 0.9999^{80} = 0.008.$$

5.3–22 Denote the three lifetimes by X_1, X_2, X_3 and let $Y = X_1 + X_2 + X_3$.

$$E(Y) = E(X_1 + X_2 + X_3) = E(X_1) + E(X_2) + E(X_3) = 3 \cdot 2 \cdot 2 = 12.$$

$$\text{Var}(X_1 + X_2 + X_3) = \text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3) = 3 \cdot 2 \cdot 2^2 = 24.$$

5.4 The Moment-Generating Function Technique

5.4–2 $\quad M_Y(t) \quad = \quad E[e^{t(X_1+X_2)}] = E[e^{tX_1}]E[e^{tX_2}]$

$$= \quad (q + pe^t)^{n_1}(q + pe^t)^{n_2} = (q + pe^t)^{n_1+n_2}.$$

Thus Y is $b(n_1 + n_2, p)$.

5.4–4 $\quad E[e^{t(X_1+\cdots+X_n)}] \quad = \quad \prod_{i=1}^{n} E[e^{tX_i}] = \prod_{i=1}^{n} e^{\mu_i(e^t-1)}$

$$= \quad e^{(\mu_1+\mu_2+\cdots+\mu_n)(e^t-1)},$$

the moment generating function of a Poisson random variable with mean $\mu_1 + \mu_2 + \cdots + \mu_n$.

5.4–6 (a) $\quad E[e^{tY}] \quad = \quad E[e^{t(X_1+X_2+X_3+X_4+X_5)}]$

$$= \quad E[e^{tX_1}e^{tX_2}e^{tX_3}e^{tX_4}e^{tX_5}]$$

$$= \quad E[e^{tX_1}]E[e^{tX_2}]E[e^{tX_3}]E[e^{tX_4}]E[e^{tX_5}]$$

$$= \quad \frac{(1/3)e^t}{1-(2/3)e^t}\frac{(1/3)e^t}{1-(2/3)e^t} \cdots \frac{(1/3)e^t}{1-(2/3)e^t}$$

$$= \quad \left[\frac{(1/3)e^t}{1-(2/3)e^t}\right]^5$$

$$= \quad \frac{[(1/3)e^t]^5}{[1-(2/3)e^t]^5}, \qquad t < -\ln(1 - 1/3).$$

(b) So Y has a negative binomial distribution with $p = 1/3$ and $r = 5$.

5.4–8 $\quad E[e^{tW}] \quad = \quad E[e^{t(X_1+X_2+\cdots+X_h)}] = E[e^{tX_1}]E[e^{tX_2}] \cdots E[e^{tX_h}]$

$$= \quad [1/(1-\theta t)]^h = 1/(1-\theta t)^h, \qquad t < 1/\theta,$$

the moment generating function for the gamma distribution with mean $h\theta$.

5.4–10 (a) $E[e^{tX}] = (1/4)(e^{0t} + e^{1t} + e^{2t} + e^{3t});$
(b) $E[e^{tY}] = (1/4)(e^{0t} + e^{4t} + e^{8t} + e^{12t});$
(c) $\quad E[e^{tW}] \quad = \quad E[e^{t(X+Y)}]$

$$= \quad E[e^{tX}]E[e^{tY}]$$

$$= \quad (1/16)(e^{0t} + e^{1t} + e^{2t} + e^{3t})(e^{0t} + e^{4t} + e^{8t} + e^{12t})$$

$$= \quad (1/16)(e^{0t} + e^{1t} + e^{2t} + e^{3t} + \cdots e^{15t});$$

(d) $P(W = x) = 1/16, \qquad w = 0, 1, 2, \ldots, 15.$

5.4–12 (a) $g(w) = \dfrac{1}{12}$, $w = 0, 1, 2, \ldots, 11$, because, for example,

$$P(W = 3) = P(X = 1, Y = 2) = \left(\frac{1}{6}\right)\left(\frac{1}{2}\right) = \frac{1}{12}.$$

(b) $h(w) = \dfrac{1}{36}$, $w = 0, 1, 2, \ldots, 35$, because, for example,

$$P(W = 7) = P(X = 1, Y = 6) = \left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \frac{1}{36}.$$

5.4–14 (a) Let X_1, X_2, X_3 equal the digit that is selected on draw 1, 2, and 3, respectively. Then

$$f(x_i) = 1/10, \qquad x_i = 0, 1, 2, \ldots, 9.$$

Let $W = X_1 + X_2 + X_3$.
$$\begin{aligned}
P(W = 0) &= 1/1000; \\
P(W = 1) &= 3/1000; \\
P(W = 2) &= 6/1000; \\
P(W = 3) &= 10/1000;
\end{aligned}$$

$$\begin{aligned}
\$500 \cdot P(W = 0) - \$1 &= \$500/1000 - \$1 = -50 \text{ cents} \\
\$166 \cdot P(W = 1) - \$1 &= \$498/1000 - \$1 = -50.2 \text{ cents} \\
\$83 \cdot P(W = 2) - \$1 &= \$498/1000 - \$1 = -50.2 \text{ cents};
\end{aligned}$$

(b) $\$50 \cdot P(W = 3) - \$1 = \$500/1000 - \$1 = -50$ cents;

(c) Let $Y = X_1 + X_2 + X_3 + X_4$, the sum in the 4-digit game.
$$\begin{aligned}
P(Y = 0) &= 1/10{,}000; \\
P(Y = 1) &= 1/2{,}500; \\
P(Y = 2) &= 1/1{,}000; \\
P(Y = 3) &= 1/500;
\end{aligned}$$

$$\begin{aligned}
\$5{,}000 \cdot P(Y = 0) - \$1 &= \$5{,}000/10{,}000 - \$1 = -50 \text{ cents} \\
\$1{,}250 \cdot P(Y = 1) - \$1 &= \$1{,}250/2{,}500 - \$1 = -50 \text{ cents} \\
\$500 \cdot P(Y = 2) - \$1 &= \$500/1{,}000 - \$1 = -50 \text{ cents};
\end{aligned}$$

(d) $\$250 \cdot P(Y = 3) - \$1 = \$250/500 - \$1 = -50$ cents.

5.4–16 Let X_1, X_2, X_3 be the number of accidents in weeks 1, 2, and 3, respectively. Then $Y = X_1 + X_2 + X_3$ is Poisson with mean $\lambda = 6$ and

$$P(Y = 7) = 0.744 - 0.606 = 0.138.$$

5.4–18 Let X_1, X_2, X_3, X_4 be the number of sick days for employee i, $i = 1, 2, 3, 4$, respectively. Then $Y = X_1 + X_2 + X_3 + X_4$ is Poisson with mean $\lambda = 8$ and

$$P(Y > 10) = 1 - P(Y \leq 10) = 1 - 0.0816 = 0.184.$$

5.4–20 Let X_i equal the number of cracks in mile i, $i = 1, 2, \ldots, 40$. Then

$$Y = \sum_{i=1}^{40} X_i \quad \text{is Poisson with mean} \quad \lambda = 20.$$

It follows that

$$P(Y < 15) = P(Y \leq 14) = \sum_{y=0}^{14} \frac{20^y e^{-20}}{y!} = 0.1049.$$

The final answer was calculated using Minitab.

5.4–22 $Y = X_1 + X_2 + X_3 + X_4$ has a gamma distribution with $\alpha = 6$ and $\theta = 10$. So

$$P(Y > 90) = \int_{90}^{\infty} \frac{1}{\Gamma(6)10^6} y^{6-1} e^{-y/10} \, dy = 1 - 0.8843 = 0.1157.$$

The final answer was calculated using Minitab.

5.5 Random Functions Associated with Normal Distributions

5.5–2
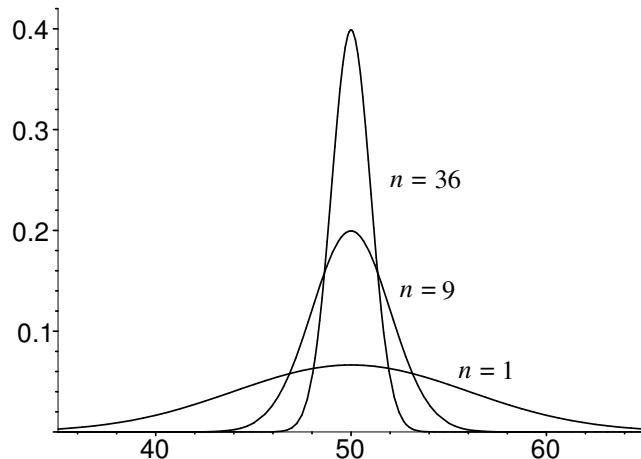
Figure 5.5–2: X is $N(50, 36)$, $\overline{X}$ is $N(50, 36/n)$, $n = 9, 36$

5.5–4 **(a)** $P(X < 6.0171) = P(Z < -1.645) = 0.05$;

  (b) Let W equal the number of boxes that weigh less than 6.0171 pounds. Then W is $b(9, 0.05)$ and $P(W \leq 2) = 0.9916$;

  (c) $P(\overline{X} \leq 6.035) = P\left(Z \leq \dfrac{6.035 - 6.05}{0.02/3}\right)$

$$= P(Z \leq -2.25) = 0.0122.$$

5.5–6 (a)
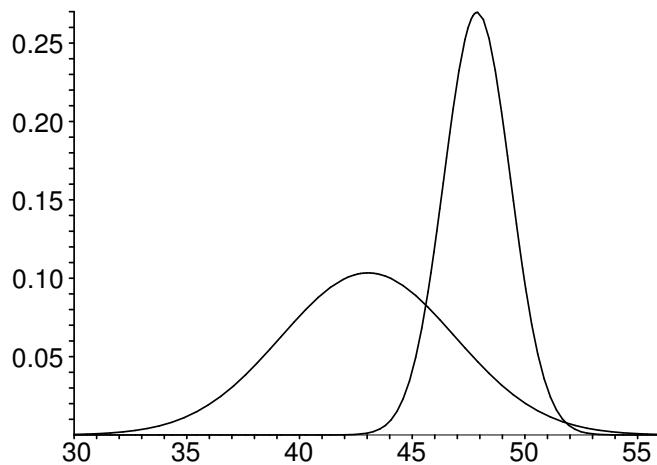
Figure 5.5–6: $N(43.04, 14.89)$ and $N(47.88, 2.19)$ p.d.f.s

(b) The distribution of $X_1 - X_2$ is $N(4.84, 17.08)$. Thus

$$P(X_1 > X_2) = P(X_1 - X_2 > 0) = P\left(Z > \frac{-4.84}{\sqrt{17.08}}\right) = 0.8790.$$

5.5–8 The distribution of Y is $N(3.54, 0.0147)$. Thus

$$P(Y > W) = P(Y - W > 0) = P\left(Z > \frac{-0.32}{\sqrt{0.0147 + 0.092}}\right) = 0.9830.$$
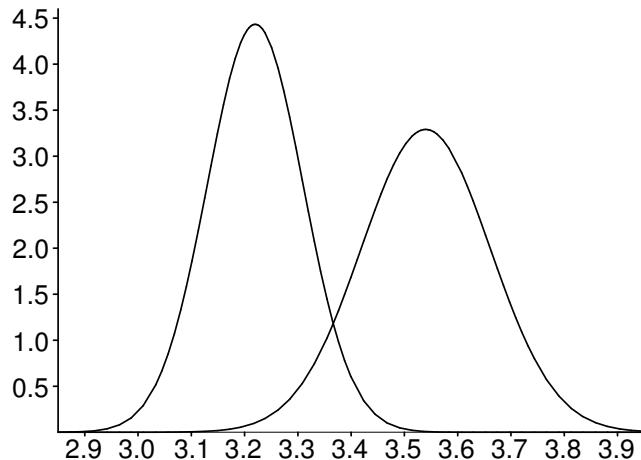
Figure 5.5–8: $N(3.22, 0.09^2)$ and $N(3[1.18], 3[0.07^2])$ p.d.f.s

5.5–10 $X - Y$ is $N(184.09 - 171.93, 39.37 + 50.88)$;

$$P(X > Y) = P\left(\frac{X - Y - 12.16}{\sqrt{90.25}} > \frac{0 - 12.16}{9.5}\right) = P(Z > -1.28) = 0.8997.$$

5.5–12 **(a)** $E(\overline{X}) = 24.5$, $\text{Var}(\overline{X}) = \dfrac{3.8^2}{8} = 1.805$,

$$E(\overline{Y}) = 21.3, \ \text{Var}(\overline{Y}) = \frac{2.7^2}{8} = 0.911;$$

(b) $N(24.5 - 21.3 = 3.2, 1.805 + 0.911 = 2.716)$;

(c) $\begin{aligned} P(\overline{X} > \overline{Y}) &= P(\overline{X} - \overline{Y} > 0) = 1 - \Phi\left(\frac{0 - 3.2}{1.648}\right) \\ &= 1 - \Phi(-1.94) = \Phi(1.94) = 0.9738. \end{aligned}$

5.5–14 Let $Y = X_1 + X_2 + \cdots + X_n$. Then Y is $N(800n, 100^2 n)$. Thus

$$P(Y \geq 10000) = 0.90$$

$$P\left(\frac{Y - 800n}{100\sqrt{n}} \geq \frac{10000 - 800n}{100\sqrt{n}}\right) = 0.90$$

$$-1.282 = \frac{10000 - 800n}{100\sqrt{n}}$$

$$800n - 128.2\sqrt{n} - 10000 = 0.$$

Either use the quadratic formula to solve for $\sqrt{n}$ or use Maple to solve for n. We find that $\sqrt{n} = 3.617$ or $n = 13.08$ so use $n = 14$ bulbs.

5.5–16 The joint p.d.f. is

$$f(x_1, x_2) = \frac{1}{\sqrt{2\pi}} e^{-x_1^2/2} \frac{1}{\Gamma(r/2)2^{r/2}} x_2^{r/2-1} e^{-x_2/2}, \qquad -\infty < x_1 < \infty, \ \ 0 < x_2 < \infty;$$

$$y_1 = x_1/\sqrt{x_2/r}, \qquad y_2 = x_2$$

$$x_1 = y_1\sqrt{y_2/r}, \qquad x_2 = y_2$$

The Jacobian is

$$J = \begin{vmatrix} \sqrt{y_2/r} & y_1(\frac{1}{2})y_2^{-1/2}/\sqrt{r} \\ 0 & 1 \end{vmatrix} = \sqrt{y_2/r};$$

The joint p.d.f. of Y_1 and Y_2 is

$$g(y_1, y_2) = \frac{1}{\sqrt{2\pi}} e^{-y_1^2 y_2/2r} \frac{1}{\Gamma(r/2)2^{r/2}} y_2^{r/2-1} e^{-y_2/2} \frac{\sqrt{y_2}}{\sqrt{r}}, \qquad -\infty < y_1 < \infty, \ \ 0 < y_2 < \infty;$$

The marginal p.d.f. of Y_1 is

$$\begin{aligned} g_1(y_1) &= \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-y_1^2 y_2/2r} \frac{1}{\Gamma(r/2)2^{r/2}} y_2^{r/2-1} e^{-y_2/2} \frac{\sqrt{y_2}}{\sqrt{r}} \, dy_2 \\ &= \frac{\Gamma[(r+1)/2]}{\sqrt{\pi r}\,\Gamma(r/2)} \int_0^\infty \frac{1}{\Gamma[(r+1)/2]2^{(r+1)/2}} y_2^{(r+1)/2-1} e^{-(y_2/2)(1+y_1^2/r)} \end{aligned}$$

Let $u = y_2(1 + y_1^2/r)$. Then $y_2 = \dfrac{u}{1 + y_1^2/r}$ and $\dfrac{dy_2}{du} = \dfrac{1}{1 + y_1^2/r}$. So

$$
\begin{aligned}
g_1(y_1) &= \frac{\Gamma[(r+1)/2]}{\sqrt{\pi r}\,\Gamma(r/2)(1 + y_1^2/r)^{(r+1)/2}} \int_0^\infty \frac{1}{\Gamma[(r+1)/2]2^{(r+1)/2}}\, u^{(r+1)/2-1} e^{-u/2} \\
&= \frac{\Gamma[(r+1)/2]}{\sqrt{\pi r}\,\Gamma(r/2)(1 + y_1^2/r)^{(r+1)/2}}, \qquad -\infty < y_1 < \infty.
\end{aligned}
$$

5.5–18 **(a)** $t_{0.05}(23) = 1.714$;

(b) $t_{0.90}(23) = -t_{0.10}(23) = -1.319$;

(c) $P(-2.069 \leq T \leq 2.500) = 0.99 - 0.025 = 0.965$.

5.5–20 $T = \dfrac{\overline{X} - \mu}{S/\sqrt{9}}$ is t with $r = 9 - 1 = 8$ degrees of freedom.

(a) $t_{0.025}(8) = 2.306$;

(b)
$$
\begin{aligned}
-t_{0.025} &\leq \frac{\overline{X} - \mu}{S/\sqrt{n}} \leq t_{0.025} \\
-t_{0.025}\frac{S}{\sqrt{n}} &\leq \overline{X} - \mu \leq t_{0.025}\frac{S}{\sqrt{n}} \\
-\overline{X} - t_{0.025}\frac{S}{\sqrt{n}} &\leq -\mu \leq -\overline{X} + t_{0.025}\frac{S}{\sqrt{n}} \\
\overline{X} - t_{0.025}\frac{S}{\sqrt{n}} &\leq \mu \leq \overline{X} + t_{0.025}\frac{S}{\sqrt{n}}
\end{aligned}
$$

5.6   The Central Limit Theorem

5.6–2 If
$$
\begin{aligned}
f(x) &= (3/2)x^2, \qquad -1 < x < 1, \\
E(X) &= \int_{-1}^1 x(3/2)x^2\, dx = 0; \\
\operatorname{Var}(X) &= \int_{-1}^1 (3/2)x^4\, dx = \left[\frac{3}{10}x^5\right]_{-1}^1 = \frac{3}{5}.
\end{aligned}
$$

Thus
$$
\begin{aligned}
P(-0.3 \leq Y \leq 1.5) &= P\left(\frac{-0.3 - 0}{\sqrt{15(3/5)}} \leq \frac{Y - 0}{\sqrt{15(3/5)}} \leq \frac{1.5 - 0}{\sqrt{15(3/5)}}\right) \\
&\approx P(-0.10 \leq Z \leq 0.50) = 0.2313.
\end{aligned}
$$

5.6–4
$$
\begin{aligned}
P(39.75 \leq \overline{X} \leq 41.25) &= P\left(\frac{39.75 - 40}{\sqrt{(8/32)}} \leq \frac{\overline{X} - 40}{\sqrt{(8/32)}} \leq \frac{41.25 - 40}{\sqrt{(8/32)}}\right) \\
&\approx P(-0.50 \leq Z \leq 2.50) = 0.6853.
\end{aligned}
$$

5.6–6 **(a)**
$$
\begin{aligned}
\mu &= \int_0^2 x(1 - x/2)\, dx = \left[\frac{x^2}{2} - \frac{x^3}{6}\right]_0^2 = 2 - \frac{4}{3} = \frac{2}{3}; \\
\sigma^2 &= \int_0^2 x^2(1 - x/2)\, dx - \left(\frac{2}{3}\right)^2 \\
&= \left[\frac{x^3}{3} - \frac{x^4}{8}\right]_0^2 - \frac{4}{9} = \frac{2}{9}.
\end{aligned}
$$

(b) $P\left(\dfrac{2}{3} \le \overline{X} \le \dfrac{5}{6}\right) \;=\; P\left(\dfrac{\frac{2}{3}-\frac{2}{3}}{\sqrt{\frac{2}{9}/18}} \le \dfrac{\overline{X}-\frac{2}{3}}{\sqrt{\frac{2}{9}/18}} \le \dfrac{\frac{5}{6}-\frac{2}{3}}{\sqrt{\frac{2}{9}/18}}\right)$

$\approx \;\; P(0 \le Z \le 1.5) = 0.4332.$

5.6–8 (a) $E(\overline{X}) = \mu = 24.43;$

(b) $\text{Var}(\overline{X}) = \dfrac{\sigma^2}{n} = \dfrac{2.20}{30} = 0.0733;$

(c) $P(24.17 \le \overline{X} \le 24.82) \;\approx\; P\left(\dfrac{24.17-24.43}{\sqrt{0.0733}} \le Z \le \dfrac{24.82-24.43}{\sqrt{0.0733}}\right)$

$= \;\; P(-0.96 \le Z < 1.44) = 0.7566.$

5.6–10 Using the normal approximation,

$P(1.7 \le Y \le 3.2) \;=\; P\left(\dfrac{1.7-2}{\sqrt{4/12}} \le \dfrac{Y-2}{\sqrt{4/12}} \le \dfrac{3.2-2}{\sqrt{4/12}}\right)$

$\approx \;\; P(-0.52 \le Z \le 2.078) = 0.6796.$

Using the p.d.f. of Y,

$P(1.7 \le Y \le 3.2) \;=\; \int_{1.7}^{2}[(-1/2)y^3 + 2y^2 - 2y + (2/3)]\,dy$

$+ \int_{2}^{3}[(1/2)y^3 - 4y^2 + 10y - 22/3]\,dy$

$+ \int_{3}^{3.2}[(-1/6)y^3 + 2y^2 - 8y + 32/3]\,dy$

$= \;\; [(-1/8)y^4 + (2/3)y^3 - y^2 + (2/3)y]_{1.7}^{2}$

$+ [(1/8)y^4 - (4/3)y^3 + 5y^2 - (22/3)y]_{2}^{3}$

$+ [(-1/24)y^4 + (2/3)y^3 - 4y^2 + (32/3)y]_{3}^{3.2}$

$= \;\; 0.1920 + 0.4583 + 0.0246 = 0.6749.$

5.6–12 The distribution of $\overline{X}$ is $N(2000, 500^2/25)$. Thus

$$P(\overline{X} > 2050) = P\left(\dfrac{\overline{X}-2000}{500/5} > \dfrac{2050-2000}{500/5}\right) \approx 1 - \Phi(0.50) = 0.3085.$$

5.6–14 $\quad E(X+Y) \;=\; 30 + 50 = 80;$

$\text{Var}(X+Y) \;=\; \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y$

$= \;\; 52 + 64 + 28 = 144;$

$Z \;=\; \sum_{i=1}^{25}(X_i + Y_i) \;\; \text{in approximately} \;\; N(25 \cdot 80, 25 \cdot 144).$

Thus $\;\; P(1970 < Z < 2090) \;=\; P\left(\dfrac{1970-2000}{60} < \dfrac{Z-2000}{60} < \dfrac{2090-2000}{60}\right)$

$\approx \;\; \Phi(1.5) - \Phi(-0.5)$

$= \;\; 0.9332 - 0.3085 = 0.6247.$

5.6–16 Let X_i equal the time between sales of ticket $i-1$ and i, for $i = 1, 2, \ldots, 10$. Each X_i has a gamma distribution with $\alpha = 3$, $\theta = 2$. $Y = \sum_{i=1}^{10} X_i$ has a gamma distribution with parameters $\alpha_Y = 30, \theta_Y = 2$. Thus

$$P(Y \le 60) = \int_{0}^{60} \dfrac{1}{\Gamma(30)2^{30}} y^{30-1} e^{-y/2}\,dy = 0.52428 \;\; \text{using Maple.}$$

The normal approximation is given by

$$P\left(\frac{Y-60}{\sqrt{120}} \le \frac{60-60}{\sqrt{120}}\right) \approx \Phi(0) = 0.5000.$$

5.6–18 We are given that $Y = \sum_{i=1}^{20} X_i$ has mean 200 and variance 80. We want to find y so that

$$P(Y \ge y) \quad < \quad 0.20$$

$$P\left(\frac{Y-200}{\sqrt{80}} > \frac{y-200}{\sqrt{80}}\right) \quad < \quad 0.20;$$

We have that

$$\frac{y-200}{\sqrt{80}} \quad = \quad 0.842$$

$$y \quad = \quad 207.5 \uparrow 208 \text{ days.}$$

5.7   Approximations for Discrete Distributions

5.7–2 **(a)** $P(2 < X < 9) = 0.9532 - 0.0982 = 0.8550;$

(b) $\quad P(2 < X < 9) \quad = \quad P\left(\frac{2.5-5}{2} \le \frac{X - 25(0.2)}{\sqrt{25(0.2)(0.8)}} \le \frac{8.5-5}{2}\right)$

$$\approx \quad P(-1.25 \le Z \le 1.75)$$

$$= \quad 0.8543.$$

5.7–4 $\quad P(35 \le X \le 40) \quad \approx \quad P\left(\frac{34.5-36}{3} \le Z \le \frac{40.5-36}{3}\right)$

$$= \quad P(-0.50 \le Z \le 1.50) = 0.6247.$$

5.7–6 $\mu_X = 84(0.7) = 58.8, \ \text{Var}(X) = 84(0.7)(0.3) = 17.64,$

$$P(X \le 52.5) \approx \Phi\left(\frac{52.5 - 58.8}{4.2}\right) = \Phi(-1.5) = 0.0668.$$

5.7–8 **(a)** $\quad P(X < 20.857) \quad = \quad P\left(\frac{X - 21.37}{0.4} < \frac{20.857 - 21.37}{0.4}\right)$

$$= \quad P(Z < -1.282) = 0.10.$$

(b) The distribution of Y is $b(100, 0.10)$. Thus

$$P(Y \le 5) = P\left(\frac{Y - 100(0.10)}{\sqrt{100(0.10)(0.90)}} \le \frac{5.5 - 10}{3}\right) \approx P(Z \le -1.50) = 0.0668.$$

(c) $\quad P(21.31 \le \overline{X} \le 21.39) \quad \approx \quad P\left(\frac{21.31 - 21.37}{0.4/10} \le Z \le \frac{21.39 - 21.37}{0.4/10}\right)$

$$= \quad P(-1.50 \le Z \le 0.50) = 0.6247.$$

5.7–10 $\quad P(4776 \le X \le 4856) \quad \approx \quad P\left(\frac{4775.5 - 4829}{\sqrt{4829}} \le Z \le \frac{4857.5 - 4829}{\sqrt{4829}}\right)$

$$= \quad P(-0.77 \le Z \le 0.41) = 0.4385.$$

5.7–12 The distribution of Y is $b(1000, 18/38)$. Thus

$$P(Y > 500) \approx P\left(Z \geq \frac{500.5 - 1000(18/38)}{\sqrt{1000(18/38)(20/38)}}\right) = P(Z \geq 1.698) = 0.0448.$$

5.7–14 (a) $E(X) = 100(0.1) = 10$, $\text{Var}(X) = 9$,

$$\begin{aligned} P(11.5 < X < 14.5) &\approx \Phi\left(\frac{14.5 - 10}{3}\right) - \Phi\left(\frac{11.5 - 10}{3}\right) \\ &= \Phi(1.5) - \Phi(0.5) = 0.9332 - 0.6915 = 0.2417. \end{aligned}$$

(b) $P(X \leq 14) - P(X \leq 11) = 0.917 - 0.697 = 0.220$;

(c) $\displaystyle\sum_{x=12}^{14} \binom{100}{x}(0.1)^x(0.9)^{100-x} = 0.2244.$

5.7–16 (a) $E(Y) = 24(3.5) = 84$, $\text{Var}(Y) = 24(35/12) = 70$,

$$P(Y \geq 85.5) \approx 1 - \Phi\left(\frac{85.5 - 84}{\sqrt{70}}\right) = 1 - \Phi(0.18) = 0.4286;$$

(b) $P(Y < 85.5) \approx 1 - 0.4286 = 0.5714$;

(c) $P(70.5 < Y < 86.5) \approx \Phi(0.30) - \Phi(-1.61) = 0.6179 - 0.0537 = 0.5642.$

5.7–18 (a)
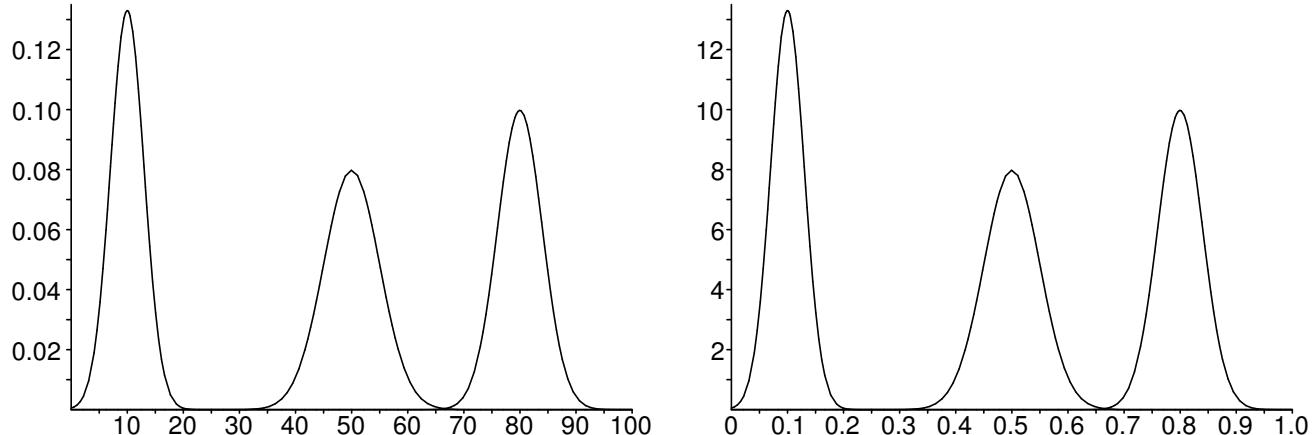
Figure 5.7–18: Normal approximations of the p.d.f.s of Y and $Y/100$, $p = 0.1, 0.5, 0.8$

(b) When $p = 0.1$,

$$P(-1.5 < Y - 10 < 1.5) \approx \Phi\left(\frac{1.5}{3}\right) - \Phi\left(\frac{-1.5}{3}\right) = 0.6915 - 0.3085 = 0.3830;$$

When $p = 0.5$,

$$P(-1.5 < Y - 50 < 1.5) \approx \Phi\left(\frac{1.5}{5}\right) - \Phi\left(\frac{-1.5}{5}\right) = 0.6179 - 0.3821 = 0.2358;$$

When $p = 0.8$,

$$P(-1.5 < Y - 80 < 1.5) \approx \Phi\left(\frac{1.5}{4}\right) - \Phi\left(\frac{-1.5}{4}\right) = 0.6462 - 0.3538 = 0.2924.$$

5.7–20 X is $N(0, 0.5^2)$. The probability that one item exceeds 0.98 in absolute value is

$$\begin{aligned}
P(|X| > 0.98) &= 1 - P(-0.98 \le X \le 0.98) \\
&= 1 - P\left(\frac{-0.98 - 0}{0.5} \le \frac{X - 0}{0.5} \le \frac{0.98 - 0}{0.5}\right) \\
&= 1 - P(-1.96 \le Z \le 1.96) = 1 - 0.95 = 0.05
\end{aligned}$$

If we let Y equal the number out of 100 that exceed 0.98 in absolute value, Y is $b(100, 0.05)$.

(a) Let $\lambda = 100(0.05) = 5$.

$$P(Y \ge 7) = 1 - P(Y \le 6) = 1 - 0.762 = 0.238.$$

(b)
$$\begin{aligned}
P(Y \ge 7) &= P\left(\frac{Y - 5}{\sqrt{100(0.05)(0.95)}} \ge \frac{6.5 - 5}{2.179}\right) \\
&\approx P(Z \ge 0.688) \\
&= 1 - 0.7543 = 0.2447.
\end{aligned}$$

(c) $P(Y \ge 7) = 1 - P(Y \le 6) = 1 - 0.7660 = 0.2340$ using Minitab.

5.7–22 **(a)** Let X equal the number of matches. Then

$$f(x) = \frac{\binom{20}{x}\binom{60}{4 - x}}{\binom{80}{4}}, \qquad x = 0, 1, 2, 3, 4.$$

Thus

$$\begin{aligned}
f(0) &= \frac{97{,}527}{316{,}316} = 0.308 \\
f(1) &= \frac{34{,}220}{79{,}079} = 0.433 \\
f(2) &= \frac{16{,}815}{79{,}079} = 0.218 \\
f(3) &= \frac{3{,}420}{79{,}079} = 0.043 \\
f(4) &= \frac{969}{316{,}316} = 0.003.
\end{aligned}$$

$$\begin{aligned}
EP = E(\text{Payoff}) &= -1f(0) - 1f(1) + 0f(2) + 4f(3) + 54f(4) \\
&= -\frac{9{,}797}{24{,}332} = -0.403; \\
EDP = E(\text{DoublePayoff}) &= -1f(0) - 1f(1) + 1f(2) + 9f(3) + 109f(4) \\
&= \frac{2{,}369}{12{,}166} = 0.195
\end{aligned}$$

(b) The variances and standard deviation for the regular game and the double payoff game, respectively, are

$$\text{Var(Payoff)} = (-1 - EP)^2 f(0) + (-1 - EP)^2 f(1) + (0 - EP)^2 f(2)$$
$$+ (4 - EP)^2 f(3) + (54 - EP)^2 f(4)$$
$$= \frac{78{,}534{,}220{,}095}{7{,}696{,}600{,}912}$$
$$\sigma = 3.1943;$$
$$\text{Var(DoublePayoff)} = (-1 - EDP)^2 f(0) + (-1 - EDP)^2 f(1) +$$
$$(1 - EDP)^2 f(2) + (9 - EDP)^2 f(3) + (109 - EDP)^2 f(4)$$
$$= \frac{78{,}534{,}220{,}095}{1{,}924{,}150{,}228}$$
$$\sigma = 6.3886.$$

(c) Let $Y = \sum_{i=1}^{2000} X_i$, the sum of "winnings" in 2000 repetitions of the regular game. The distribution of Y is approximately

$$N\left(2000\left(-\frac{9{,}797}{24{,}332}\right),\ 2000\left(\frac{78{,}534{,}220{,}095}{7{,}696{,}600{,}912}\right)\right) = N(-805.277,\ 20{,}407.50742).$$

$$P(Y > 0) = P\left(\frac{Y + 805.277}{142.856} > \frac{0.5 + 805.277}{142.856}\right) \approx P(Z > 5.64) = 0.$$

Let $W = \sum_{i=1}^{2000} X_i$, the sum of "winnings" in 2000 repetitions of the double payoff game. The distribution of W is approximately

$$N\left(2000\left(\frac{2{,}369}{12{,}166}\right),\ 2000\left(\frac{78{,}534{,}220{,}095}{1{,}924{,}150{,}228}\right)\right) = N(389.446,\ 81{,}630.02966).$$

$$P(W > 0) = P\left(\frac{W - 389.446}{285.7097} > \frac{0.5 - 389.446}{285.7097}\right) \approx P(Z > -1.3613) = 0.9133.$$

(d) Here are the results of 100 simulations of these two games.

The respective sample means are -803.65 and 392.70. The respective sample variances are 19,354.45202 and 77,417.80808.

Here are box plots comparing the two games.
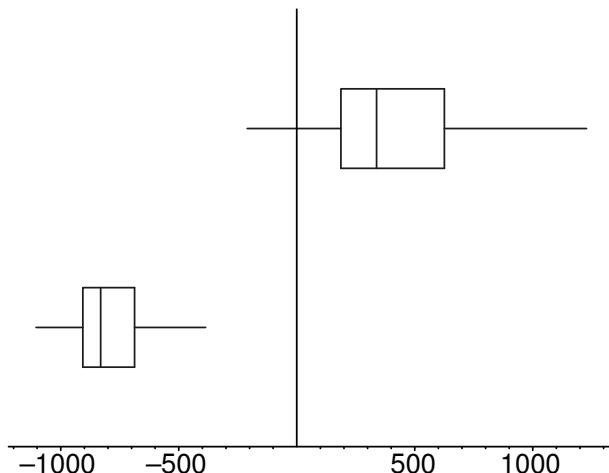
Figure 5.7–22: Box plots of 100 simulations of 2000 plays

Here is a histogram of the 100 simulations of 2000 plays of the regular game.
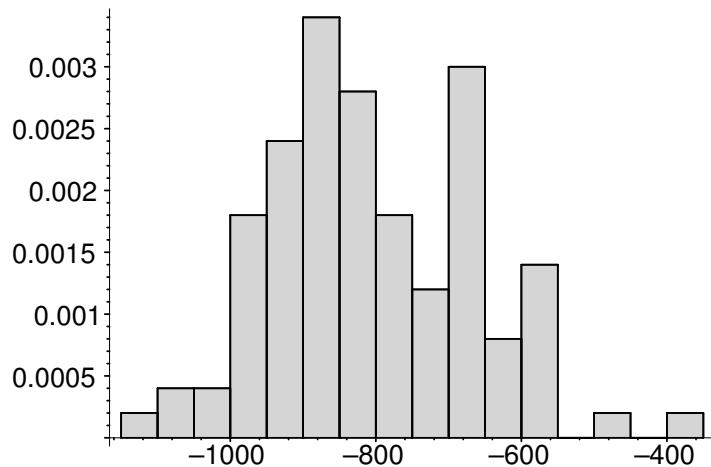
Figure 5.7–22: A histogram of 100 simulations of 2000 plays of the regular game

Here is a histogram of 100 simulations of 2000 plays of the double payoff (promotion) game.
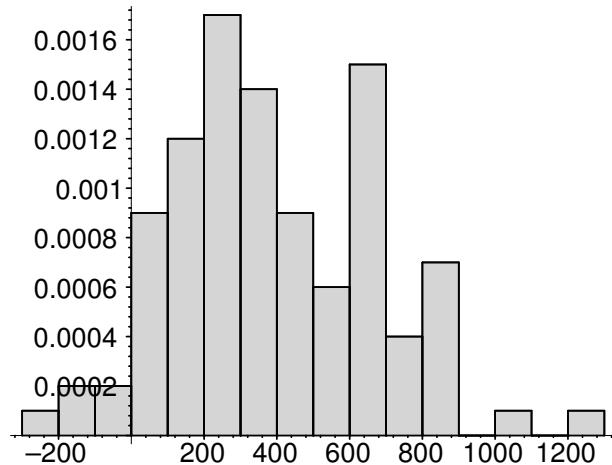

Figure 5.7–22: A histogram of 100 simulations of 2000 plays of the promotion game

Chapter 6

Estimation

6.1 Point Estimation

6.1–2 The likelihood function is

$$L(\theta) = \left[\frac{1}{2\pi\theta}\right]^{n/2} \exp\left[-\sum_{i=1}^{n}(x_i - \mu)^2/2\theta\right], \qquad 0 < \theta < \infty.$$

The logarithm of the likelihood function is

$$\ln L(\theta) = -\frac{n}{2}(\ln 2\pi) - \frac{n}{2}(\ln \theta) - \frac{1}{2\theta}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Setting the first derivative equal to zero and solving for θ yields

$$\frac{d\ln L(\theta)}{d\theta} = -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$$

$$\theta = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Thus

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2.$$

To see that $\widehat{\theta}$ is an unbiased estimator of θ, note that

$$E(\widehat{\theta}) = E\left(\frac{\sigma^2}{n}\sum_{i=1}^{n}\frac{(X_i - \mu)^2}{\sigma^2}\right) = \frac{\sigma^2}{n}\cdot n = \sigma^2,$$

since $(X_i - \mu)^2/\sigma^2$ is $\chi^2(1)$ and hence the expected value of each of the n summands is equal to 1.

6.1–4 (a) $\bar{x} = 394/7 = 56.2857;\quad s^2 = 5452/97 = 56.2062;$

(b) $\widehat{\lambda} = \bar{x} = 394/7 = 56.2857;$

(c) Yes;

(d) $\bar{x}$ is better than s^2 because

$$\text{Var}(\overline{X}) \approx \frac{56.2857}{98} = 0.5743 < 65.8956 = \frac{56.2857[2(56.2857 * 98) + 97]}{98(97)} \approx \text{Var}(S^2).$$

6.1–6 $\widehat{\theta}_1 = \widehat{\mu} = 33.4267;\quad \widehat{\theta}_2 = \widehat{\sigma^2} = 5.0980.$

6.1–8 **(a)**
$$L(\theta) = \left(\frac{1}{\theta^n}\right)\left(\prod_{i=1}^{n} x_i\right)^{1/\theta-1}, \qquad 0 < \theta < \infty$$

$$\ln L(\theta) = -n\ln\theta + \left(\frac{1}{\theta} - 1\right)\ln\prod_{i=1}^{n} x_i$$

$$\frac{d\ln L(\theta)}{d\theta} = \frac{-n}{\theta} - \frac{1}{\theta^2}\ln\prod_{i=1}^{n} x_i = 0$$

$$\widehat{\theta} = -\frac{1}{n}\ln\prod_{i=1}^{n} x_i$$

$$= -\frac{1}{n}\sum_{i=1}^{n}\ln x_i.$$

(b) We first find $E(\ln X)$:

$$E(\ln X) = \int_0^1 \ln x (1/\theta) x^{1/\theta-1}\,dx.$$

Using integration by parts, with $u = \ln x$ and $dv = (1/\theta)x^{1/\theta-1}dx$,

$$E(\ln X) = \lim_{a\to 0}\left[x^{1/\theta}\ln x - \theta x^{1/\theta}\right]_a^1 = -\theta.$$

Thus

$$E(\widehat{\theta}) = -\frac{1}{n}\sum_{i=1}^{n}(-\theta) = \theta.$$

6.1–10 **(a)** $\overline{x} = 1/p$ so $\widetilde{p} = 1/\overline{X} = n/\sum_{i=1}^{n} X_i;$

(b) $\widetilde{p}$ equals the number of successes, n, divided by the number of Bernoulli trials, $\sum_{i=1}^{n} X_i;$

(c) $20/252 = 0.0794.$

6.1–12 **(a)** $E(\overline{X}) = E(Y)/n = np/n = p;$

(b) $\mathrm{Var}(\overline{X}) = \mathrm{Var}(Y)/n^2 = np(1-p)/n^2 = p(1-p)/n;$

(c)
$$\begin{aligned}
E[\overline{X}(1-\overline{X})/n] &= [E(\overline{X}) - E(\overline{X}^2)]/n \\
&= \{p - [p^2 + p(1-p)/n]\}/n = [p(1-1/n) - p^2(1-1/n)]/n \\
&= (1-1/n)p(1-p)/n = (n-1)p(1-p)/n^2;
\end{aligned}$$

(d) From part (c), the constant $c = 1/(n-1).$

6.1–14 **(a)**
$$\begin{aligned}
E(cS) &= E\left\{\frac{c\sigma}{\sqrt{n-1}}\left[\frac{(n-1)S^2}{\sigma^2}\right]^{1/2}\right\} \\[2mm]
&= \frac{c\sigma}{\sqrt{n-1}}\int_0^\infty \frac{v^{1/2}v^{(n-1)/2-1}e^{-v/2}}{\Gamma\left(\dfrac{n-1}{2}\right)2^{(n-1)/2}}\,dv \\[2mm]
&= \frac{c\sigma}{\sqrt{n-1}}\frac{\sqrt{2}\,\Gamma(n/2)}{\Gamma[(n-1)/2)]},
\end{aligned}$$

so $c = \dfrac{\sqrt{n-1}\,\Gamma[(n-1)/2]}{\sqrt{2}\,\Gamma(n/2)};$

(b) When $n = 5$, $c = 8/(3\sqrt{2\pi})$ and when $n = 6$, $c = 3\sqrt{5\pi}/(8\sqrt{2})$.

(c)
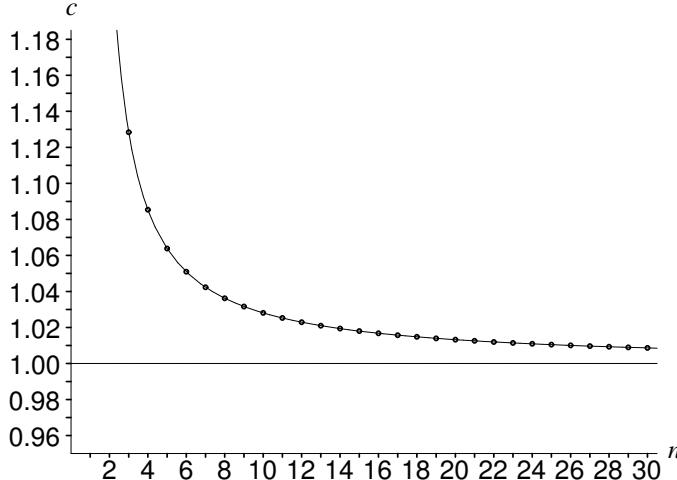
Figure 6.1–14: c as a function of n

We see that

$$\lim_{n \to \infty} c = 1.$$

6.1–16 $\overline{x} = \alpha\theta$, $v = \alpha\theta^2$ so that $\widetilde{\theta} = v/\overline{x}$, $\widetilde{\alpha} = \overline{x}^2/s^2$. For the given data, $\widetilde{\alpha} = 102.4990$, $\widetilde{\theta} = 0.0658$. Note that $\overline{x} = 6.74$, $v = 0.4432$, $s^2 = 0.4617$.

6.1–18 The experiment has a hypergeometric distribution with $n = 8$ and $N = 64$. From the sample, $\overline{x} = 1.4667$. Using this as an estimate for μ we have

$$1.4667 = 8\left(\frac{N_1}{64}\right) \quad \text{implies that} \quad \widetilde{N_1} = 11.73.$$

A guess for the value of N_1 is therefore 12.

6.1–20 **(a)**
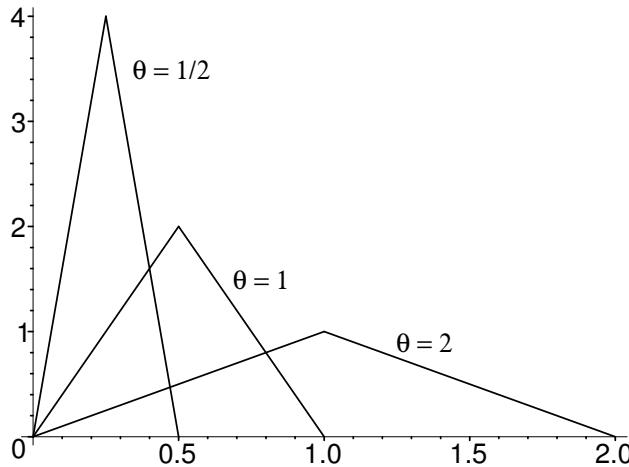
Figure 6.1–20: The p.d.f.s of X for three values of θ

(b) $E(X) = \theta/2$. Thus the method of moments estimator of θ is $\widetilde{\theta} = 2\overline{X}$.

(c) Since $\overline{x} = 0.37323$, a point estimate of θ is $2(0.37323) = 0.74646$.

6.2   Confidence Intervals for Means

6.2–2 **(a)** [77.272, 92.728];   **(b)** [79.12, 90.88];   **(c)** [80.065, 89.935];   **(d)** [81.154, 88.846].

6.2–4 **(a)** $\overline{x} = 56.8$;

   (b) $\left[56.8 - 1.96(2/\sqrt{10}\,),\;\; 56.8 + 1.96(2/\sqrt{10}\,\right] = [55.56, 58.04]$;

   (c) $P(X < 52) = P\left(Z < \dfrac{52 - 56.8}{2}\right) = P(Z < -2.4) = 0.0082.$

6.2–6 $\left[11.95 - 1.96\left(\dfrac{11.80}{\sqrt{37}}\right),\;\; 11.95 + 1.96\left(\dfrac{11.80}{\sqrt{37}}\right)\right] = [8.15,\ 15.75].$

   If more extensive t-tables are available or if a computer program is used, we have

   $$\left[11.95 - 2.028\left(\dfrac{11.80}{\sqrt{37}}\right),\;\; 11.95 + 2.028\left(\dfrac{11.80}{\sqrt{37}}\right)\right] = [8.016,\ 15.884].$$

6.2–8 **(a)** $\overline{x} = 46.42$;
   (b) $46.72 \pm 2.132 s/\sqrt{5}$ or $[40.26,\ 52.58]$.

6.2–10 $\left[21.45 - 1.314\left(\dfrac{0.31}{\sqrt{28}}\right), \infty\right) = [21.373,\ \infty).$

6.2–12 **(a)** $\overline{x} = 3.580$;

   (b) $s = 0.512$;

   (c) $[0,\ 3.580 + 1.833(0.512/\sqrt{10}\,] = [0,\ 3.877].$

6.2–14 **(a)** $\overline{x} = 245.80$, $s = 23.64$, so a 95% confidence interval for μ is

   $$[245.80 - 2.145(23.64)/\sqrt{15}\,,\ 245.80 + 2.145(23.64)/\sqrt{15}\,] = [232.707,\ 258.893];$$

   (b)
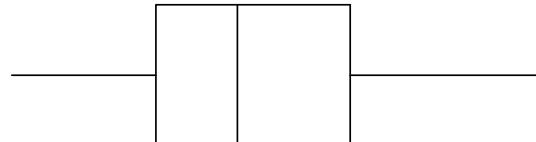
Figure 6.2–14: Box-and-whisker diagram of signals from detectors

   (c) The standard deviation is quite large.

6.2–16 **(a)** $(\overline{x} + 1.96\sigma/\sqrt{5}\,) - (\overline{x} - 1.96\sigma/\sqrt{5}\,) = 3.92\sigma/\sqrt{5} = 1.753\sigma$;
   (b) $(\overline{x} + 2.776 s/\sqrt{5}\,) - (\overline{x} - 2.776 s/\sqrt{5}\,) = 5.552 s/\sqrt{5}.$

   From Exercise 6.2–14 with $n = 5$, $E(S) = \dfrac{\sqrt{2}\,\Gamma(5/2)\sigma}{\sqrt{4}\,\Gamma(4/2)} = \dfrac{3\sqrt{\pi}\sigma}{2^{5/2}} = 0.94\sigma$, so that
   $E[5.552 S/\sqrt{5}\,] = 2.334\sigma.$

6.2–18 $6.05 \pm 2.576(0.02)/\sqrt{1219}$ or [6.049, 6.051].

6.2–20 (a) $\bar{x} = 4.483$, $s^2 = 0.1719$, $s = 0.4146$;

(b) $[4.483 - 1.714(0.4146)/\sqrt{24}), \infty) = [4.338, \infty)$;

(c) yes; construct a *q-q* plot or compare empirical and theoretical distribution functions.

N(4.48, 0.1719) quantiles
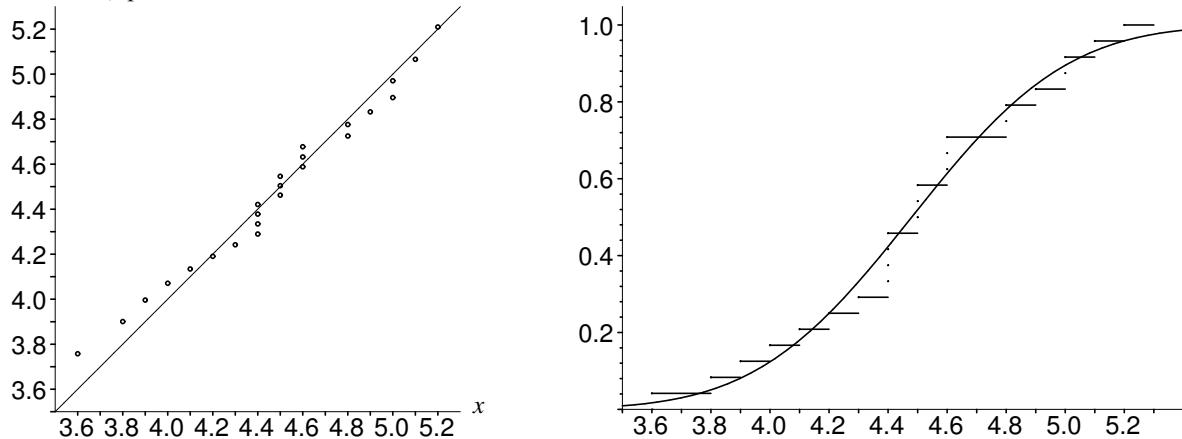
Figure 6.2–20: *q-q* plot and a comparison of empirical and theoretical distribution functions

6.2–22 (a) $\bar{x} = 5.833$, $s = 1.661$;

(b) Using a normal approximation, the 95% confidence interval is

$$[5.833 - 1.96(1.661/10), 5.833 + 1.96(1.661/10)] = [5.507, 6.159].$$

Using $t_{0.025}(99) = 1.98422$, the confidence interval is

$$[5.833 - 1.98422(1.661/10), 5.833 + 1.98422(1.661/10)] = [5.503, 6.163].$$

6.3 Confidence Intervals For Difference of Two Means

6.3–2 $\bar{x} = 539.2$, $s_x^2 = 4{,}948.7$, $\bar{y} = 544.625$, $s_y^2 = 4{,}327.982$, $s_p = 67.481$, $t_{0.05}(11) = 1.796$, so the confidence interval is $[-74.517, 63.667]$.

6.3–4 (a) $\bar{x} - \bar{y} = 1511.714 - 1118.400 = 393.314$;

(b) $s_x^2 = 49{,}669.905$, $s_y^2 = 15{,}297.600$, $r = \lfloor 8.599 \rfloor = 8$, $t_{0.025}(8) = 2.306$, so the confidence interval is $[179.148, 607.480]$.

6.3–6 **(a)** $\bar{x} = 712.25$, $\bar{y} = 705.4375$, $s_x^2 = 29,957.8409$, $s_y^2 = 20,082.1292$, $s_p = 155.7572$, $t_{0.025}(26) = 2.056$. Thus a 95% confidence interval for $\mu_X - \mu_Y$ is $[-115.480, 129.105]$.

(b)
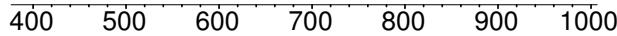
Figure 6.3–6: Box-and-whisker diagrams for butterfat production

(c) No.

6.3–8 **(a)** $\bar{x} = 2.584$, $\bar{y} = 1.564$, $s_x^2 = 0.1042$, $s_y^2 = 0.0428$, $s_p = 0.2711$, $t_{0.025}(18) = 2.101$. Thus a 95% confidence interval for $\mu_X - \mu_Y$ is $[0.7653, 1.2747]$.

(b)

Figure 6.3–8: Box-and-whisker diagrams, wedge on (X) and wedge off (Y)

(c) Yes.

6.3–10 From (a), (b), and (c), we know

(d) $$\frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\dfrac{d\sigma_Y^2}{n} + \dfrac{\sigma_Y^2}{m}}} \div \sqrt{\left[\frac{(n-1)S_X^2}{d\sigma_Y^2} + \frac{(m-1)S_Y^2}{\sigma_Y^2}\right] \Big/ (n+m-2)}$$

has a $t(n+m-2)$ distribution. Clearly, this ratio does not depend upon σ_Y^2; so

$$\bar{x} - \bar{y} \pm t_{\alpha/2}(n+m-2)\sqrt{\frac{(n-1)s_x^2/d + (m-1)s_y^2}{n+m-2}\left(\frac{d}{n} + \frac{1}{m}\right)}$$

provides a $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$.

6.3–12 (a) $\bar{d} = 0.07875$;

(b) $[\bar{d} - 1.7140.25492/\sqrt{24}, \infty) = [-0.0104, \infty)$;

(c) not necessarily.

6.3–14 (a) $\bar{x} = 136.61$, $\bar{y} = 134.87$, $s_x^2 = 3.2972$, $s_y^2 = 1.0957$;

(b) Using Welch with $r = 18$ degrees of freedom, the 95% confidence interval is [0.436, 3.041]. Assuming equal variances with $r = 20$ degrees of freedom, the 95% confidence interval is [0.382, 3.095].

(c) The five-number summary for the X observations is 133.30, 135.625, 136.95, 137.80, 139.40. The five-number summary for the Y observations is 132.70, 134.15, 134.95, 135.825, 136.00.
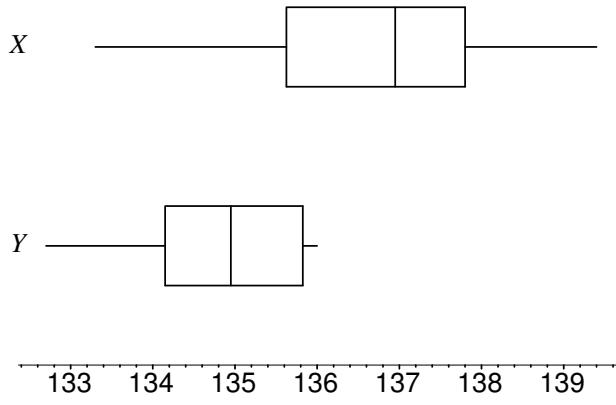
Figure 6.3–14: Hardness of hot (X) and cold (Y) water

(d) The mean for hot seems to be larger than the mean for cold.

6.3–16 $\bar{a} = 31.14$, $\bar{b} = 33.43$, $s_a = 6.12$, $s_b = 7.52$. Assuming normally distributed distributions and equal variances, a 90% confidence interval for the difference of the means is $[-8.82, 4.25]$.

6.4 Confidence Intervals For Variances

6.4–2 For these 9 weights, $\bar{x} = 20.90$, $s = 1.858$.

(a) A point estimate for σ is $s = 1.858$.

(b) $\left[\dfrac{1.858\sqrt{8}}{\sqrt{17.54}}, \dfrac{1.858\sqrt{8}}{\sqrt{2.180}} \right] = [1.255, 3.599]$

or

$\left[\dfrac{1.858\sqrt{8}}{\sqrt{21.595}}, \dfrac{1.858\sqrt{8}}{\sqrt{2.623}} \right] = [1.131, 3.245]$;

(c) $\left[\dfrac{1.858\sqrt{8}}{\sqrt{15.51}}, \dfrac{1.858\sqrt{8}}{\sqrt{2.733}} \right] = [1.334, 3.179]$

or

$\left[\dfrac{1.858\sqrt{8}}{\sqrt{19.110}}, \dfrac{1.858\sqrt{8}}{\sqrt{3.298}} \right] = [1.202, 2.894]$.

6.4–4 **(a)** A point estimate for σ is $s = 0.512$. Note that $s^2 = 0.2618$.

    (b) $\left[\dfrac{9(0.2618)}{19.02}, \dfrac{9(0.2618)}{2.700}\right] = [0.124, \ 0.873]$;

    (c) $\left[\dfrac{3\sqrt{0.2618}}{\sqrt{19.02}}, \dfrac{3\sqrt{0.2618}}{\sqrt{2.700}}\right] = [0.352, \ 0.934]$;

    (d) $\left[\dfrac{3\sqrt{0.2618}}{\sqrt{22.912}}, \dfrac{3\sqrt{0.2618}}{\sqrt{3.187}}\right] = [0.321, \ 0.860]$.

6.4–6 **(a)** Since

$$E(e^{tX}) = (1 - \theta t)^{-1},$$

$$E[e^{t(2X/\theta)}] = [1 - \theta(2t/\theta)]^{-1} = (1 - 2t)^{-2/2},$$

the moment generating function for $\chi^2(2)$. Thus W is the sum of n independent $\chi^2(2)$ variables and so W is $\chi^2(2n)$.

    (b) $P\left(\chi^2_{1-\alpha/2}(2n) \le \dfrac{2\sum_{i=1}^{n} X_i}{\theta} \le \chi^2_{\alpha/2}(2n)\right) = P\left(\dfrac{2\sum_{i=1}^{n} X_i}{\chi^2_{\alpha/2}(2n)} \le \theta \le \dfrac{2\sum_{i=1}^{n} X_i}{\chi^2_{1-\alpha/2}(2n)}\right)$.

Thus, a $100(1-\alpha)\%$ confidence interval for θ is

$$\left[\dfrac{2\sum_{i=1}^{n} x_i}{\chi^2_{\alpha/2}(2n)}, \ \dfrac{2\sum_{i=1}^{n} x_i}{\chi^2_{1-\alpha/2}(2n)}\right].$$

    (c) $\left[\dfrac{2(7)(93.6)}{23.68}, \ \dfrac{2(7)(93.6)}{6.571}\right] = [55.34, \ 199.42]$.

6.4–8 **(a)** $\bar{x} = 84.7436$;

$$\left[\dfrac{2 \cdot 39 \cdot 84.7436}{104.316}, \dfrac{2 \cdot 39 \cdot 84.7436}{55.4656}\right] = [63.365, 119.173];$$

    (b) $\bar{y} = 113.1613$;

$$\left[\dfrac{2 \cdot 31 \cdot 113.1613}{85.6537}, \dfrac{2 \cdot 31 \cdot 113.1613}{42.1260}\right] = [81.9112, 166.5480].$$

6.4–10 A 90% confidence interval for σ_X^2/σ_Y^2 is

$$\left[\dfrac{1}{F_{0.05}(15, 12)}\left(\dfrac{s_x}{s_y}\right)^2, F_{0.05}(12, 15)\left(\dfrac{s_x}{s_y}\right)^2\right] = \left[\dfrac{1}{2.62}\left(\dfrac{0.197}{0.318}\right)^2, 2.48\left(\dfrac{0.197}{0.318}\right)^2\right].$$

So a 90% confidence interval for σ_X/σ_Y is given by the square roots of these values, namely $[0.383, 0.976]$.

6.4–12 **(a)** $\left[\dfrac{1}{3.115}\left(\dfrac{604.489}{329.258}\right), 3.115\left(\dfrac{604.489}{329.258}\right)\right] = [0.589, 5.719]$;

    (b) $[0.77, 2.39]$.

6.4–14 From the restriction, treating b as a function of a, we have

$$g(b)\dfrac{db}{da} - g(a) = 0,$$

or, equivalently,

$$\dfrac{db}{da} = \dfrac{g(a)}{g(b)}.$$

Thus

$$\frac{dk}{da} = s\sqrt{n-1}\left(\frac{-1/2}{a^{3/2}} - \frac{-1/2}{b^{3/2}}\frac{g(a)}{g(b)}\right) = 0$$

requires that

$$a^{3/2}g(a) = b^{3/2}g(b),$$

or, equivalently,

$$a^{n/2}e^{-a/2} = b^{n/2}e^{-b/2}.$$

6.4–16 (a) $\left[\frac{1}{3.01}\left(\frac{29,957.841}{20,082.129}\right), 3.35\left(\frac{29,957.841}{20,082.129}\right)\right] = [0.496, 4.997];$

The F values were found using Table VII and linear interpolation. The right endpoint is 4.968 if $F_{0.025}(15,11) = 3.33$ is used (found using Minitab).

(b) $\left[\frac{1}{2.52}\left(\frac{6.2178}{2.7585}\right), 2.52\left(\frac{6.2178}{2.7585}\right)\right] = [0.894, 5.680];$

Using linear interpolation: $F_{0.025}(19,19) \approx \frac{4(2.46)+2.76}{5} = 2.52;$ using Minitab: $F_{0.025}(19,19) = 2.5265.$

(c) $\left[\frac{1}{4.03}\left(\frac{0.10416}{0.04283}\right), 4.03\left(\frac{0.10416}{0.04283}\right)\right] = [0.603, 9.801].$

6.5 Confidence Intervals For Proportions

6.5–2 (a) $\widehat{p} = \frac{142}{200} = 0.71;$

(b) $\left[0.71 - 1.645\sqrt{\frac{(0.71)(0.29)}{200}}, 0.71 + 1.645\sqrt{\frac{(0.71)(0.29)}{200}}\right] = [0.657, 0.763];$

(c) $\dfrac{0.71 + 1.645^2/400 \pm 1.645\sqrt{0.71(0.29)/200 + 1.645^2/(4\cdot 200^2)}}{1 + 1.645^2/200} = [0.655, 0.760];$

(d) $\widetilde{p} = \frac{142+2}{200+4} = \frac{12}{17} = 0.7059;$

$\left[\frac{12}{17} - 1.645\sqrt{\frac{(12/17)(5/17)}{200}}, \frac{12}{17} - 1.645\sqrt{\frac{(12/17)(5/17)}{200}}\right] = [0.653, 0.759];$

(e) $\left[0.71 - 1.282\sqrt{\frac{(0.71)(0.29)}{200}}, 0\right] = [0.669, 0].$

6.5–4 $\left[0.70 - 1.96\sqrt{\frac{(0.70)(0.30)}{1234}}, 0.70 + 1.96\sqrt{\frac{(0.70)(0.30)}{1234}}\right] = [0.674, 0.726].$

6.5–6 $\left[0.26 - 2.326\sqrt{\frac{(0.26)(0.74)}{5757}}, 0.26 + 2.326\sqrt{\frac{(0.26)(0.74)}{5757}}\right] = [0.247, 0.273].$

6.5–8 (a) $\widehat{p} = \frac{388}{1022} = 0.3796;$

(b) $0.3796 \pm 1.645\sqrt{\frac{(0.3796)(0.6204)}{1022}}$ or $[0.3546, 0.4046].$

6.5–10 **(a)** $0.58 \pm 1.645\sqrt{\dfrac{(0.58)(0.42)}{500}}$ or $[0.544, 0.616]$;

(b) $\dfrac{0.045}{\sqrt{\dfrac{(0.58)(0.42)}{500}}} = 2.04$ corresponds to an approximate 96% confidence level.

6.5–12 **(a)** $\widehat{p}_1 = 206/374 = 0.551$, $\widehat{p}_2 = 338/426 = 0.793$;

(b) $0.551 - 0.793 \quad \pm \quad 1.96\sqrt{\dfrac{(0.551)(0.449)}{374} + \dfrac{(0.793)(0.207)}{426}}$

$-0.242 \quad \pm \quad 0.063$ or $[-0.305, -0.179]$.

6.5–14 **(a)** $\widehat{p}_1 = 28/194 = 0.144$;

(b) $0.144 \pm 1.96\sqrt{(0.144)(0.856)/194}$ or $[0.095, 0.193]$;

(c) $\widehat{p}_1 - \widehat{p}_2 = 28/194 - 11/162 = 0.076$;

(d) $\left[0.076 - 1.645\sqrt{\dfrac{(0.144)(0.856)}{194} + \dfrac{(0.068)(0.932)}{162}},\ 1 \right]$ or $[0.044, 1]$.

6.5–16 $\widehat{p}_1 = 520/1300 = 0.40$, $\widehat{p}_2 = 385/1100 = 0.35$,

$0.40 - 0.35 \pm 1.96\sqrt{\dfrac{(0.40)(0.60)}{1300} + \dfrac{(0.35)(0.65)}{1100}}$ or $[0.011, 0.089]$.

6.5–18 **(a)** $\widehat{p}_A = 170/460 = 0.37$, $\widehat{p}_B = 141/440 = 0.32$,

$0.37 - 0.32 \pm 1.96\sqrt{\dfrac{(0.37)(0.63)}{460} + \dfrac{(0.32)(0.68)}{440}}$ or $[-0.012, 0.112]$;

(b) yes, the interval includes zero.

6.6   Sample Size

6.6–2 $n = \dfrac{(1.96)^2(169)}{(1.5)^2} = 288.5$ so the sample size needed is 289.

6.6–4 $n = \dfrac{(1.96)^2(34.9)}{(0.5)^2} = 537$, rounded up to the nearest integer.

6.6–6 $n = \dfrac{(1.96)^2(33.7)^2}{5^2} = 175$, rounded up to the nearest integer.

6.6–8 If we let $p^* = \dfrac{30}{375} = 0.08$, then $n = \dfrac{1.96^2(0.08)(0.92)}{0.025^2} = 453$, rounded up.

6.6–10 $n = \dfrac{(1.645)^2(0.394)(0.606)}{(0.04)^2} = 404$, rounded up to the nearest integer.

6.6–12 $n = \dfrac{(1.645)^2(0.80)(0.20)}{(0.03)^2} = 482$, rounded up to the nearest integer.

6.6–14 If we let $p^* = \dfrac{686}{1009} = 0.6799$, then $n = \dfrac{2.326^2(0.6799)(0.3201)}{0.025^2} = 1884$, rounded up.

6.6–16 $m = \dfrac{(1.96)^2(0.5)(0.5)}{(0.04)^2} = 601$, rounded up to the nearest integer.

(a) $n = \dfrac{601}{1 + 600/1500} = 430$;

(b) $n = \dfrac{601}{1 + 600/15,000} = 578$;

(c) $n = \dfrac{601}{1 + 600/25,000} = 587$.

6.6–18 For the difference of two proportions with equal sample sizes

$$\varepsilon = z_{\alpha/2}\sqrt{\frac{p_1^*(1 - p_1^*)}{n} + \frac{p_2^*(1 - p_2^*)}{n}}$$

or

$$n = \frac{z_{\alpha/2}^2[p_1^*(1 - p_1^*) + p_2^*(1 - p_2^*)]}{\varepsilon^2}.$$

For unknown p^*,

$$n = \frac{z_{\alpha/2}^2[0.25 + 0.25]}{\varepsilon^2} = \frac{z_{\alpha/2}^2}{2\varepsilon^2}.$$

So $n = \dfrac{1.282^2}{2(0.05)^2} = 329$, rounded up.

## 6.7 A Simple Regression Problem

6.7–2 (a)

x	y	x^2	xy	y^2	$(y - \widehat{y})^2$
2.0	1.3	4.00	2.60	1.69	0.361716
3.3	3.3	10.89	10.89	10.89	0.040701
3.7	3.3	13.69	12.21	10.89	0.027725
2.0	2.0	4.00	4.00	4.00	0.009716
2.3	1.7	5.29	3.91	2.89	0.228120
2.7	3.0	7.29	8.10	9.00	0.206231
4.0	4.0	16.00	16.00	16.00	0.006204
3.7	3.0	13.69	11.10	9.00	0.217630
3.0	2.7	9.00	8.10	7.29	0.014900
2.3	3.0	5.29	6.90	9.00	0.676310
29.0	27.3	89.14	83.81	80.65	1.849254

$$\widehat{\alpha} = \overline{y} = 27.3/10 = 2.73;$$

$$\widehat{\beta} = \frac{83.81 - (29.0)(27.3)/10}{89.14 - (29.0)(29.0)/10} = \frac{4.64}{5.04} = 0.9206;$$

$$\widehat{y} = 2.73 + (4.64/5.04)(x - 2.90)$$

(b)
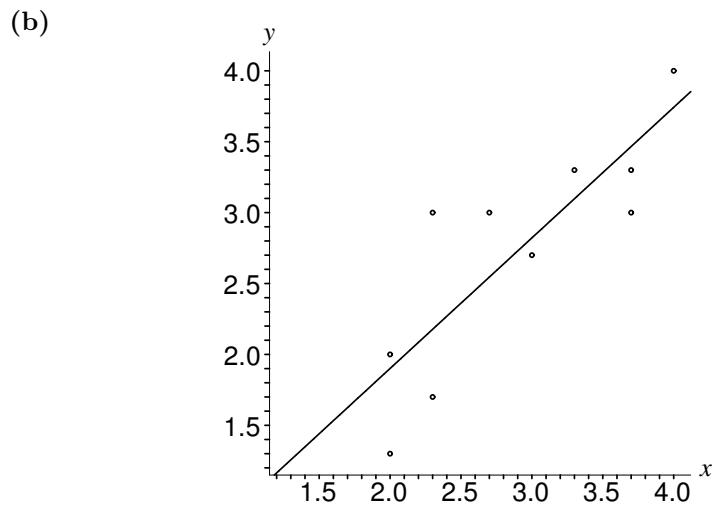
Figure 6.7–2: Earned grade (y) versus predicted grade (x)

(c) $\widehat{\sigma^2} = \dfrac{1.849254}{10} = 0.184925.$

6.7–4 **(a)** $\widehat{y} = 0.9810 + 0.0249x;$

(b)
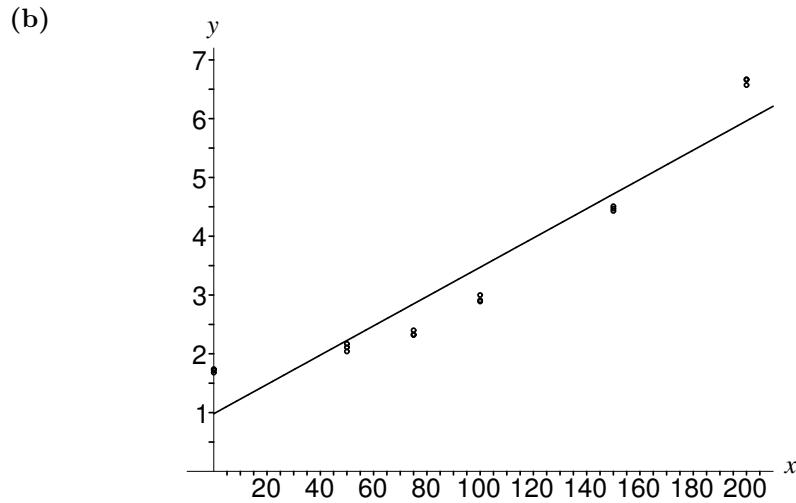
Figure 6.7–4: **(b)** Millivolts (y) versus known concentrations in ppm (x)

(c)
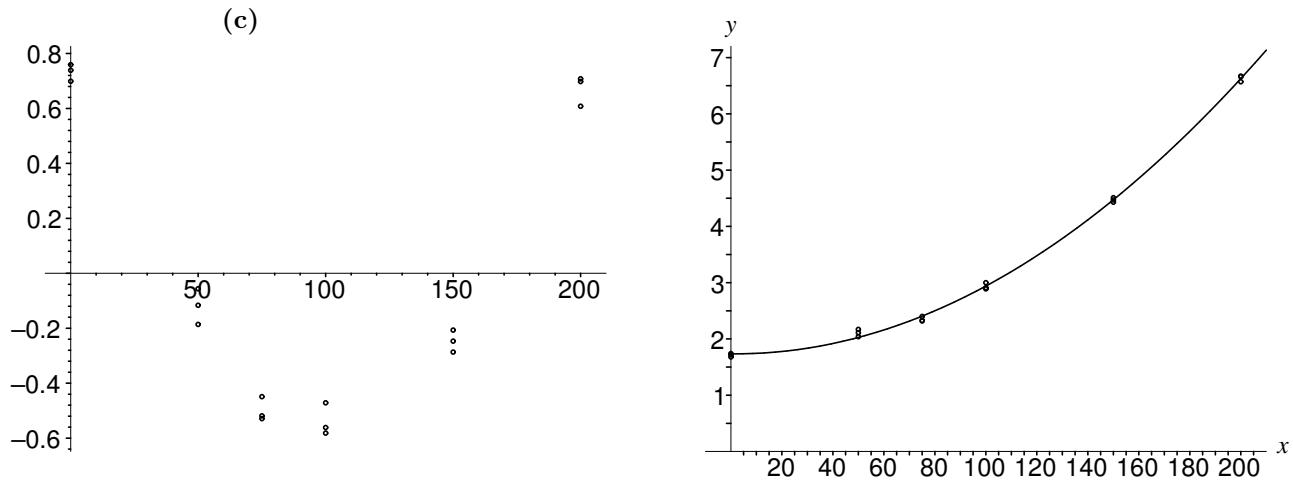
Figure 6.7–4: **(c)** A residual plot along with a quadratic regression line plot (Exercise 6.8–15)

The equation of the quadratic regression line is

$$\widehat{y} = 1.73504 - 0.000377x + 0.000124x^2.$$

6.7–6

$$\sum_{i=1}^{n} [Y_i - \alpha - \beta(x_i - \overline{x})]^2 = \sum_{i=1}^{n} [\{\widehat{\alpha} - \alpha\} + \{\widehat{\beta} - \beta\}\{x_i - \overline{x}\}$$

$$+ \ \{Y_i - \widehat{\alpha} - \widehat{\beta} \sum_{i=1}^{n}(x_i - \overline{x})\}]^2$$

$$= \ n(\widehat{\alpha} - \alpha)^2 + (\widehat{\beta} - \beta)^2 \sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$+ \ \sum_{i=1}^{n} [Y_i - \widehat{\alpha} - \widehat{\beta}(x_i - \overline{x})]^2 + 0.$$

The $+0$ in the above expression is for the three cross product terms and we must still argue that each of these is indeed 0. We have

$$2(\widehat{\alpha} - \alpha)(\widehat{\beta} - \beta) \sum_{i=1}^{n}(x_i - \overline{x}) = 0,$$

$$2(\widehat{\alpha} - \alpha) \sum_{i=1}^{n}[Y_i - \widehat{\alpha} - \widehat{\beta}(x_i - \overline{x})] = 2(\widehat{\alpha} - \alpha)\left[\sum_{i=1}^{n}(Y_i - \overline{Y}) - \widehat{\beta}\sum_{i=1}^{n}(x_i - \overline{x})\right] = 0,$$

$$2(\widehat{\beta} - \beta)\left[\sum_{i=1}^{n}(x_i - \overline{x})(Y_i - \overline{Y}) - \widehat{\beta}\sum_{i=1}^{n}(x_i - \overline{x})^2\right] =$$

$$2(\widehat{\beta} - \beta)\left[\sum_{i=1}^{n}(x_i - \overline{x})(Y_i - \overline{Y}) - \sum_{i=1}^{n}(x_i - \overline{x})(Y_i - \overline{Y})\right] = 0$$

since

$$\widehat{\beta} = \sum_{i=1}^{n}(x_i - \overline{x})(Y_i - \overline{Y}) \Big/ \sum_{i=1}^{n}(x_i - \overline{x})^2 \ .$$

6.7–8 $P\left[\chi^2_{1-\alpha/2}(n-2) \le \dfrac{n\widehat{\sigma^2}}{\sigma^2} \le \chi^2_{\alpha/2}(n-2)\right] = 1 - \alpha$

$$P\left[\dfrac{n\widehat{\sigma^2}}{\chi^2_{\alpha/2}(n-2)} \le \sigma^2 \le \dfrac{n\widehat{\sigma^2}}{\chi^2_{1-\alpha/2}(n-2)}\right] = 1 - \alpha.$$

6.7–10   Recall that $\widehat{\alpha} = 2.73$, $\widehat{\beta} = 4.64/5.04$, $\widehat{\sigma^2} = 0.184925$, $n = 10$. The endpoints for the 95% confidence interval are

$$2.73 \pm 2.306\sqrt{\dfrac{0.184925}{8}} \quad \text{or} \quad [2.379, \ 3.081] \quad \text{for} \ \ \alpha;$$

$$4.64/5.04 \pm 2.306\sqrt{\dfrac{1.84925}{8(5.04)}} \quad \text{or} \quad [0.4268, \ 1.4145] \quad \text{for} \ \ \beta;$$

$$\left[\dfrac{1.84925}{17.54}, \ \ \dfrac{1.84925}{2.180}\right] = [0.105, \ 0.848] \quad \text{for} \ \ \sigma^2.$$

6.7–12 (a) $\begin{aligned}\widehat{\beta} &= \dfrac{(1294) - (110)(121)/12}{(1234) - (110)^2/12} = \dfrac{184.833}{225.667} = 0.819;\end{aligned}$

$\widehat{\alpha} = \dfrac{121}{12} = 10.083;$

$\widehat{y} = 10.083 + \dfrac{184.833}{225.667}\left(x - \dfrac{110}{12}\right)$

$\phantom{\widehat{y}} = 0.819x + 2.575;$

(b)
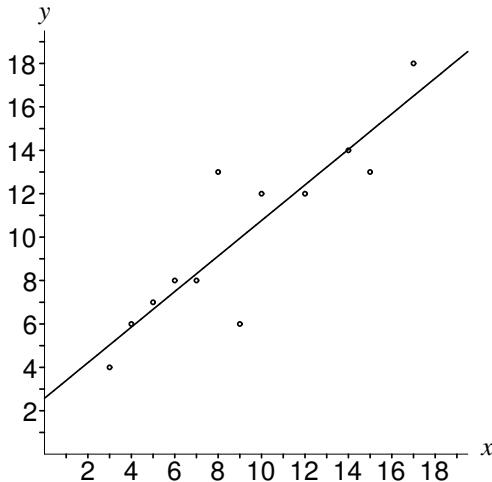
Figure 6.7–12: CO (y) versus tar (x) for 12 brands of cigarettes

(c) $\widehat{\alpha} = 10.083, \ \widehat{\beta} = 0.819,$

$n\widehat{\sigma^2} = 1411 - \dfrac{121^2}{12} - 0.81905(1294) + 0.81905(110)(121)/12 = 39.5289;$

$\widehat{\sigma^2} = \dfrac{39.5289}{12} = 3.294.$

(d) The endpoints for 95% confidence intervals are

$$10.083 \pm 2.228\sqrt{\frac{3.294}{10}} \quad \text{or} \quad [8.804, \ 11.362] \quad \text{for} \quad \alpha;$$

$$0.819 \pm 2.228\sqrt{\frac{39.5289}{10(225.667)}} \quad \text{or} \quad [0.524, \ 1.114] \quad \text{for} \quad \beta;$$

$$\left[\frac{39.5289}{20.48}, \ \frac{39.5289}{3.247}\right] = [1.930, \ 12.174] \quad \text{for} \quad \sigma^2.$$

6.7–14 (a) $\widehat{\alpha} = \dfrac{395}{15} = 26.333,$

$$\widehat{\beta} = \frac{9292 - (346)(395)/15}{8338 - (346)^2/15} = \frac{180.667}{356.933} = 0.506,$$

$$\widehat{y} = 26.333 + \frac{180.667}{356.933}\left(x - \frac{346}{15}\right)$$

$$= 0.506x + 14.657;$$

(b)
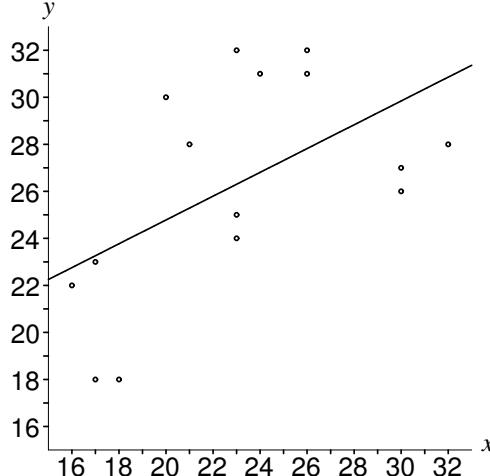
Figure 6.7–14: ACT natural science (y) versus ACT social science (x) scores

(c) $\widehat{\alpha} = 26.33, \ \widehat{\beta} = 0.506,$

$$n\widehat{\sigma^2} = 10,705 - \frac{395^2}{15} - 0.5061636(9292) + 0.5061636(346)(395)/15$$
$$= 211.8861,$$

$$\widehat{\sigma^2} = \frac{211.8861}{15} = 14.126.$$

(d) The endpoints for 95% confidence intervals are

$$26.333 \pm 2.160\sqrt{\frac{14.126}{13}} \quad \text{or} \quad [24.081, \ 28.585] \quad \text{for} \quad \alpha;$$

$$0.506 \pm 2.160\sqrt{\frac{211.8861}{13(356.933)}} \quad \text{or} \quad [0.044, \ 0.968] \quad \text{for} \quad \beta;$$

$$\left[\frac{211.8861}{24.74}, \quad \frac{211.8861}{5.009} \right] = [8.566, \ 42.301] \quad \text{for} \quad \sigma^2.$$

6.7–16 **(a)** $\widehat{y} = 6.919 + 0.8222x$, female front legs versus body lengths on left;

(b) $\widehat{y} = -0.253 + 1.273x$, female back lengths versus front legs on right.
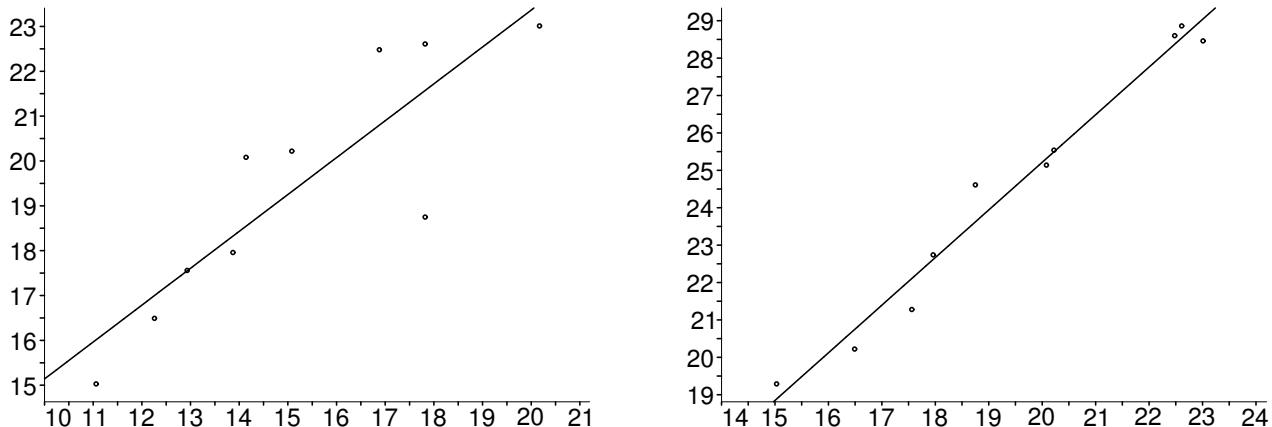

Figure 6.7–16: Female: **(a):** lengths of front legs versus body lengths; **(b):** back versus front legs

(c) $\widehat{y} = 3.996 + 1.703x$, male back legs versus body lengths on left;

**(d)** $\widehat{y} = 0.682 + 1.253x$, male back lengths versus front legs on right.
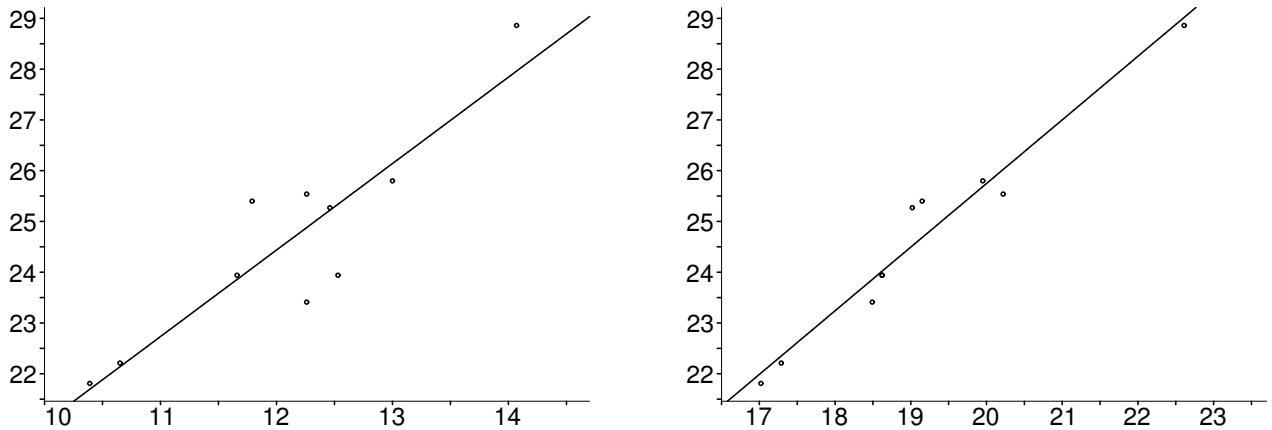
Figure 6.7–16: Male: **(c):** lengths of back legs versus body lengths; **(d):** back versus front legs

6.7–18 (b) The least squares regression line for $y = a + b$ versus b is $\widehat{y} = 1.360 + 1.626b$;

**(c)** $y = \phi x = 1.618x$ is added on the right figure below.
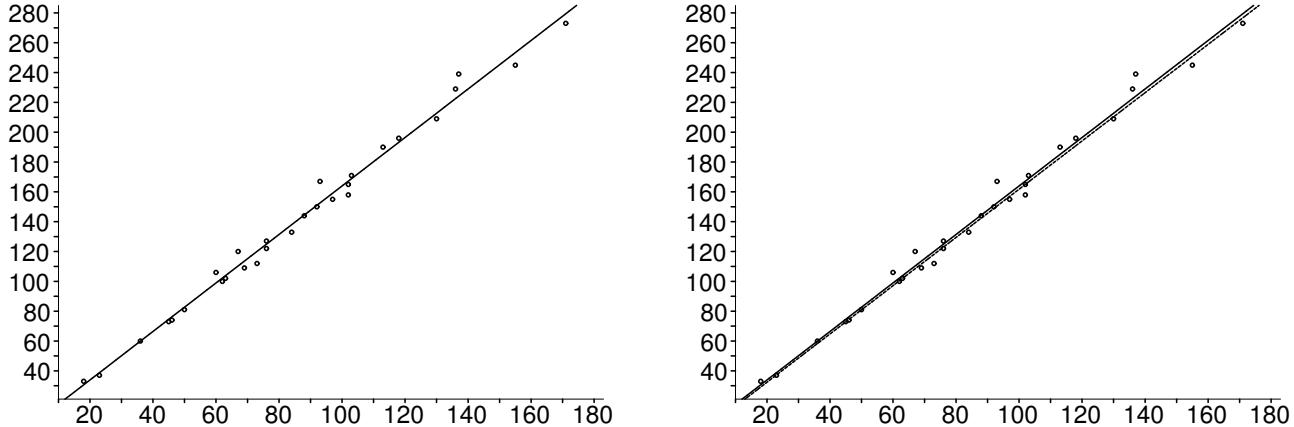
Figure 6.7–18: Scatter plot of $a + b$ versus b with least squares regression line and with $y = \phi x$

(d) The sample mean of the points $(a+b)/b$ is 1.647 which is close to the value of $\phi = 1.618$.

6.8   More Regression

6.8–2 (a) In Exercise 6.7–2 we found that

$$\widehat{\beta} = 4.64/5.04, \qquad n\widehat{\sigma^2} = 1.84924, \qquad \sum_{i=1}^{10}(x_i - \overline{x})^2 = 5.04.$$

So the endpoints for the confidence interval are given by

$$2.73 + \frac{4.64}{5.04}(x - 2.90) \ \pm\ 2.306\sqrt{\frac{1.8493}{8}}\sqrt{\frac{1}{10} + \frac{(x - 2.90)^2}{5.04}},$$

$x = 2:$ $[1.335,\ 2.468]$,

$x = 3:$ $[2.468,\ 3.176]$,

$x = 4:$ $[3.096,\ 4.389]$.

(b) The endpoints for the prediction interval are given by

$$2.73 + \frac{4.64}{5.04}(x - 2.90) \ \pm\ 2.306\sqrt{\frac{1.8493}{8}}\sqrt{1 + \frac{1}{10} + \frac{(x - 2.90)^2}{5.04}},$$

$x = 2:$ $[0.657,\ 3.146]$,

$x = 3:$ $[1.658,\ 3.986]$,

$x = 4:$ $[2.459,\ 5.026]$.
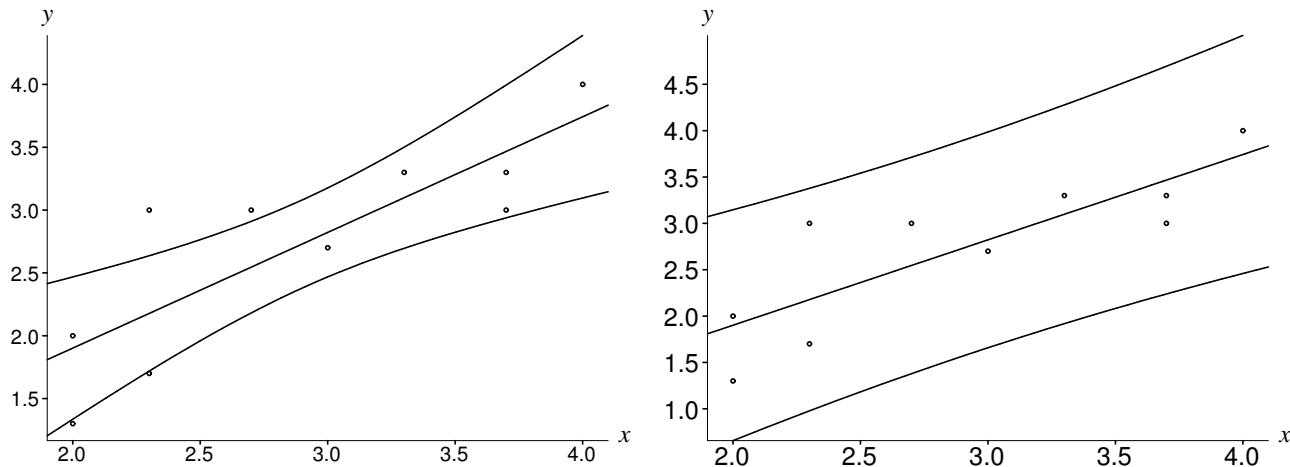
Figure 6.8–2: A 95% confidence interval for $\mu(x)$ and a 95% prediction band for Y

6.8–4 **(a)** In Exercise 6.7–11, we found that

$$\widehat{\beta} = \frac{24.8}{40}, \qquad \widehat{n\sigma^2} = 5.1895, \qquad \sum_{i=1}^{10} (x_1 - \overline{x})^2 = 40,$$

So the endpoints for the confidence interval are given by

$$50.415 + 0.62(x - 56) \pm 1.734\sqrt{\frac{5.1895}{18}}\sqrt{\frac{1}{20} + \frac{(x - 56)^2}{40}}.$$

$x = 54 :$ [48.814, 49.536],

$x = 56 :$ [50.207, 50.623],

$x = 58 :$ [51.294, 52.016].

(b) The endpoints for the prediction interval are given by

$$50.415 + 0.62(x - 56) \pm 1.734\sqrt{\frac{5.1895}{18}}\sqrt{1 + \frac{1}{20} + \frac{(x - 56)^2}{40}},$$

$x = 54 :$ [48.177, 50.173],

$x = 56 :$ [49.461, 51.369],

$x = 58 :$ [50.657, 52.653].
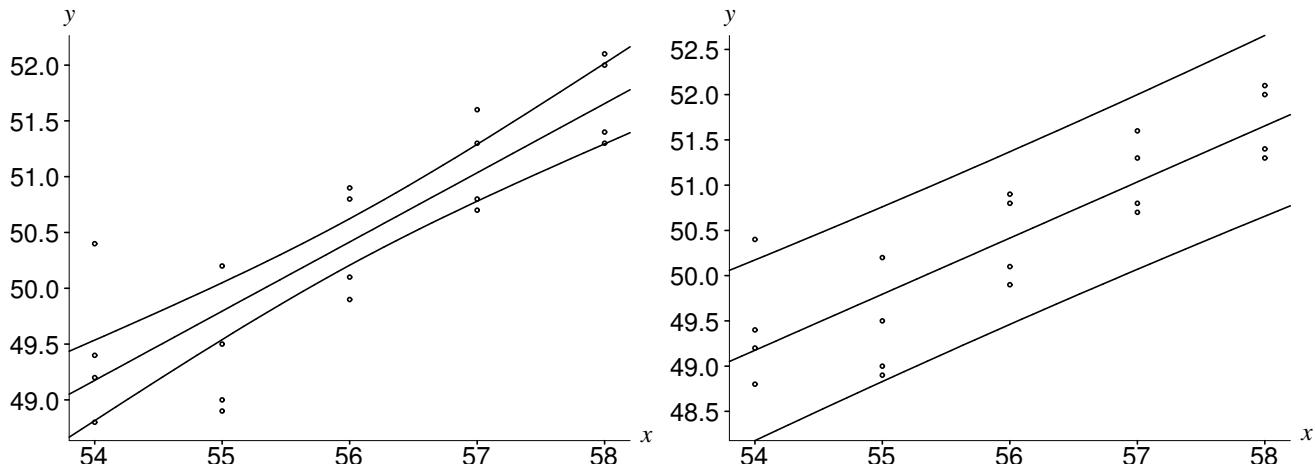
Figure 6.8–4: A 95% confidence interval for $\mu(x)$ and a 95% prediction band for Y

6.8–6 **(a)** For these data,

$$\sum_{i=1}^{10} x_i = 55, \quad \sum_{i=1}^{10} y_i = 9811, \quad \sum_{i=1}^{10} x_i^2 = 385,$$

$$\sum_{i=1}^{10} x_i y_i = 65,550, \quad \sum_{i=1}^{10} y_i^2 = 11,280,031.$$

Thus $\widehat{\alpha} = 9811/10 = 981.1$ and

$$\widehat{\beta} = \frac{65,550 - (55)(9811)/10}{385 - (55)^2/10} = \frac{11589.5}{82.5} = 140.4788.$$

The least squares regression line is

$$\widehat{y} = 981.1 + 140.4788(x - 5.5) = 208.467 + 140.479x.$$

(b)
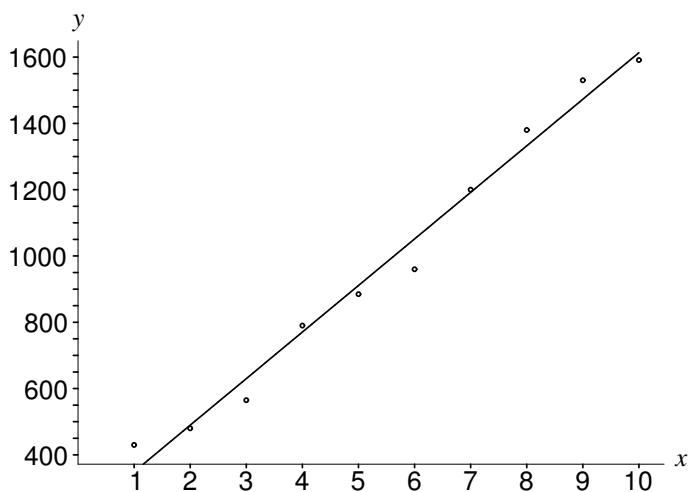
Figure 6.8–6: Number of programs (y) vs. year (x)

(c) 1753.733 ± 160.368 or $[1593.365, 1914.101]$.

6.8–8 Let $K(\beta_1, \beta_2, \beta_3) = \sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_{1i} - \beta_3 x_{2i})^2$. Then

$$
\begin{aligned}
\frac{\partial K}{\partial \beta_1} &= \sum_{i=1}^{n} 2(y_i - \beta_1 - \beta_2 x_{1i} - \beta_3 x_{2i})(-1) = 0; \\
\frac{\partial K}{\partial \beta_2} &= \sum_{i=1}^{n} 2(y_i - \beta_1 - \beta_2 x_{1i} - \beta_3 x_{2i})(-x_{1i}) = 0; \\
\frac{\partial K}{\partial \beta_3} &= \sum_{i=1}^{n} 2(y_i - \beta_1 - \beta_2 x_{1i} - \beta_3 x_{2i})(-x_{2i}) = 0.
\end{aligned}
$$

Thus, we must solve simultaneously the three equations

$$
n\beta_1 + \left(\sum_{i=1}^{n} x_{1i} \right)\beta_2 + \left(\sum_{i=1}^{n} x_{2i} \right)\beta_3 = \sum_{i=1}^{n} y_i
$$

$$
\left(\sum_{i=1}^{n} x_{1i} \right)\beta_1 + \left(\sum_{i=1}^{n} x_{1i}^2 \right)\beta_2 + \left(\sum_{i=1}^{n} x_{1i}x_{2i} \right)\beta_3 = \sum_{i=1}^{n} x_{1i}y_i
$$

$$
\left(\sum_{i=1}^{n} x_{2i} \right)\beta_1 + \left(\sum_{i=1}^{n} x_{1i}x_{2i} \right)\beta_2 + \left(\sum_{i=1}^{n} x_{2i}^2 \right)\beta_3 = \sum_{i=1}^{n} x_{2i}y_i.
$$

We have

$$
12\beta_1 + 4\beta_2 + 4\beta_3 = 23
$$

$$
4\beta_1 + 26\beta_2 + 5\beta_3 = 75
$$

$$
4\beta_1 + 5\beta_2 + 22\beta_3 = 37
$$

so that

$$
\widehat{\beta}_1 = \frac{4373}{5956} = 0.734, \qquad \widehat{\beta}_2 = \frac{3852}{1489} = 2.587, \qquad \widehat{\beta}_3 = \frac{1430}{1489} = 0.960.
$$
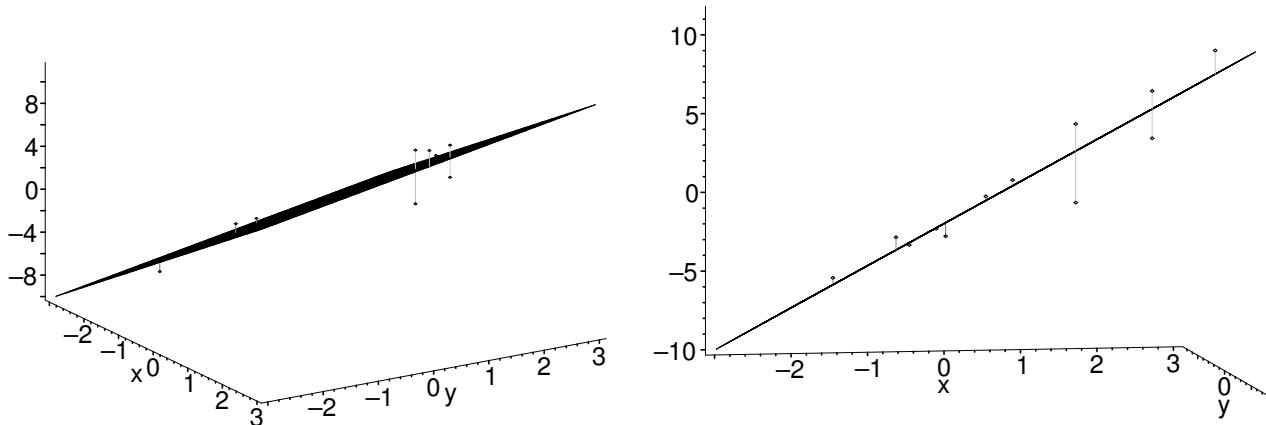
Figure 6.8–8: Two views of the points and the regression plane

6.8–10 (a) and **(b)**
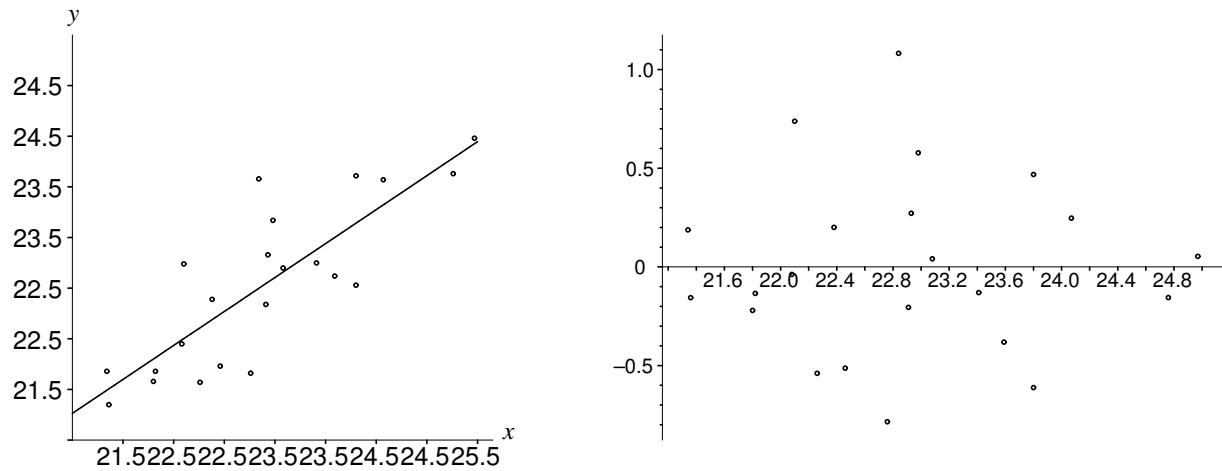
Figure 6.8–10: Swimmer's meet time (y) versus best year time (x) and residual plot

(c) and **d)**
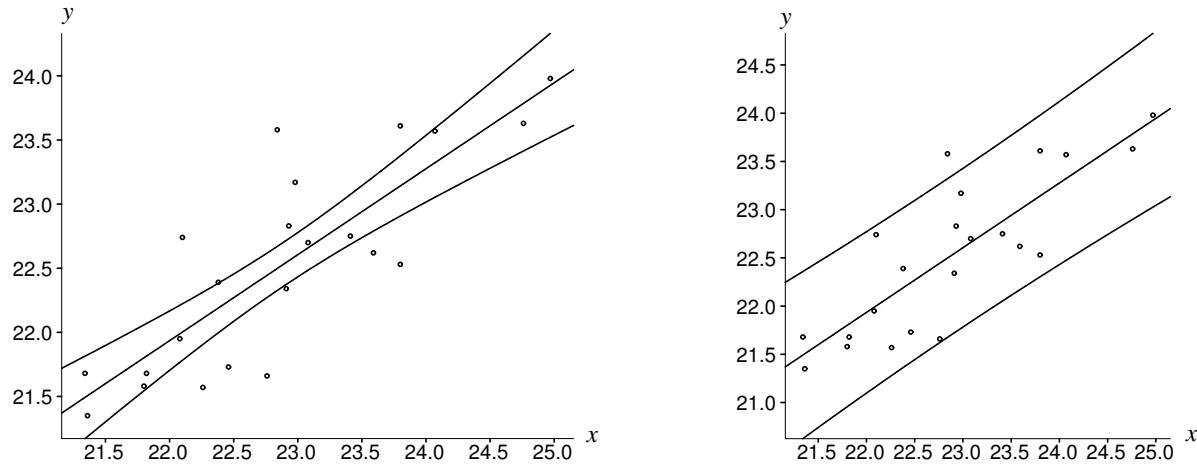
Figure 6.8–10: A 90% confidence interval for $\mu(x)$ and a 90% prediction band for Y

(e)

Parameter	Point Estimates	Confidence Level	Confidence Interval
α	22.5291	0.95	[22.3217, 22.7365]
β	0.6705	0.95	[0.4577, 0.8833]
σ^2	0.1976	0.95	[0.1272, 0.4534]

6.8–12 **(c)** and **(d)**
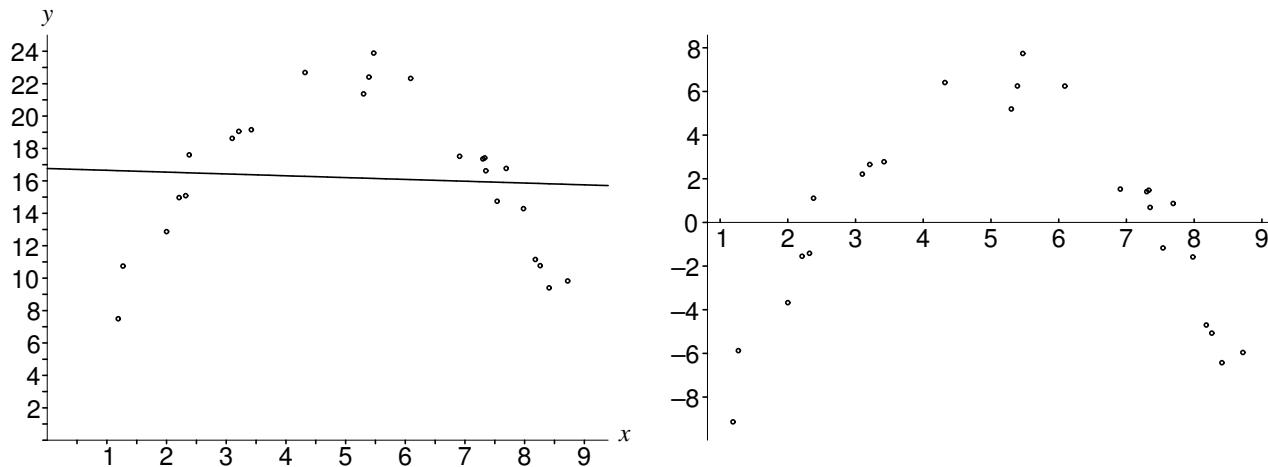
Figure 6.8–12: (y) versus (x) with linear regression line and residual plot

(e) Linear regression is not appropriate. Finding the least-squares quadratic regression line using the raw data yields $\widehat{y} = -1.895 + 9.867x - 0.996x^2$.
(f) and **(g)**
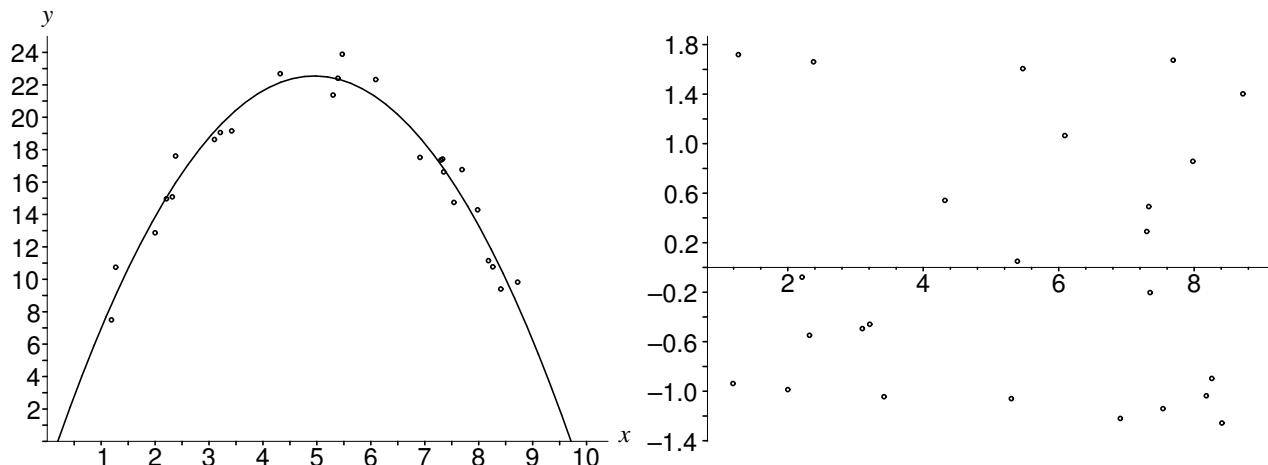
Figure 6.8–12: (y) versus (x) with quadratic regression curve and residual plot

6.8–14 (a)
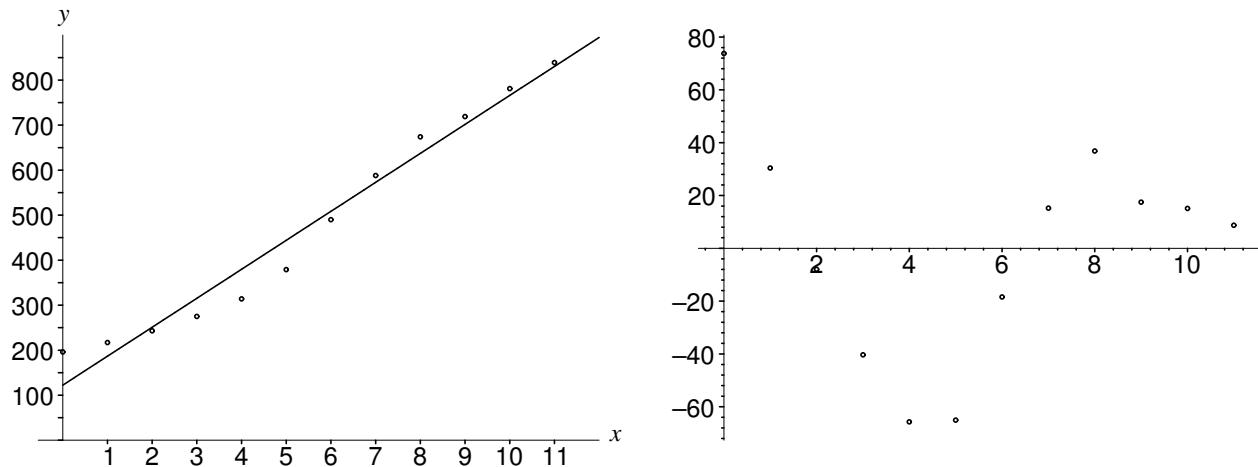
Figure 6.8–14: Number of procedures (y) versus year (x), linear regression and residual plot

(b) Without plotting the data and the residual plot, linear regression seems to be appropriate. However, it is clear that some other polynomial should be used.
(c) and **(d)**
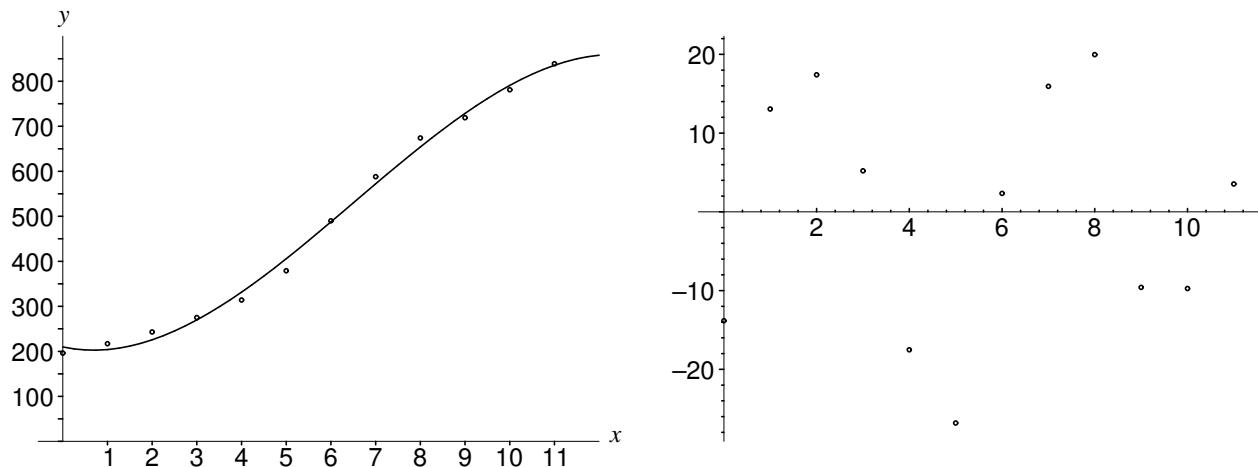
Figure 6.8–14: Number of procedures (y) versus year (x), cubic regression and residual plot

The least squares cubic regression curve is

$$\widehat{y} = 209.8168 - 21.3099x + 16.2631x^2 - 0.8323x^3.$$

Note that the years are $0, 1, 2, \ldots, 11$ rather than $1980, 1981, \ldots, 1991$.

Chapter 7

Tests of Statistical Hypotheses

7.1   Tests about Proportions

7.1–2 **(a)** $C = \{x : x = 0, 1, 2\}$;

(b) $\alpha = P(X = 0, 1, 2; p = 0.6)$

$= (0.4)^4 + 4(0.6)(0.4)^3 + 6(0.6)^2(0.4)^2 = 0.5248$;

$\beta = P(X = 3, 4; p = 0.4)$

$= 4(0.4)^3(0.6) + (0.4)^4 = 0.1792$.

OR

(a′) $C = \{x : x = 0, 1\}$;

(b′) $\alpha = P(X = 0, 1; p = 0.6)$

$= (0.4)^4 + 4(0.6)(0.4)^3 = 0.1792$;

$\beta = P(X = 2, 3, 4; p = 0.4)$

$= 6(0.4)^2(0.6)^2 + 4(0.4)^3(0.6) + (0.4)^4 = 0.5248$.

7.1–4 Using Table II in the Appendix,

(a) $\alpha = P(Y \geq 13; p = 0.40) = 1 - 0.8462 = 0.1538$;

(b) $\beta = P(Y \leq 12; p = 0.60)$

$= P(25 - Y \geq 25 - 12)$ where $25 - Y$ is $b(25, 0.40)$

$= 1 - 0.8462 = 0.1538$.

7.1–6 **(a)** $z = \dfrac{y/n - 1/6}{\sqrt{(1/6)(5/6)/n}} \leq -1.645$;

(b) $z = \dfrac{1265/8000 - 1/6}{\sqrt{(1/6)(5/6)/8000}} = -2.05 < -1.645$, reject H_0.

(c) $[0, \widehat{p} + 1.645\sqrt{\widehat{p}(1 - \widehat{p})/8000}\,] = [0, 0.1648]$, $1/6 = 0.1667$ is not in this interval. This is consistent with the conclusion to reject H_0.

7.1–8 The value of the test statistic is

$$z = \frac{0.70 - 0.75}{\sqrt{(0.75)(0.25)/390}} = -2.280.$$

(a) Since $z = -2.280 < -1.645$, reject H_0.

(b) Since $z = -2.280 > -2.326$, do not reject H_0.

(c) p-value $\approx P(Z \leq -2.280) = 0.0113$. Note that $0.01 < p$-value < 0.05.

7.1–10 (a) H_0: $p = 0.14$; H_1: $p > 0.14$;

(b) $C = \{z : z \geq 2.326\}$ where $z = \dfrac{y/n - 0.14}{\sqrt{(0.14)(0.86)/n}}$;

(c) $z = \dfrac{104/590 - 0.14}{\sqrt{(0.14)(0.86)/590}} = 2.539 > 2.326$

so H_0 is rejected and conclude that the campaign was successful.

7.1–12 (a) $z = \dfrac{y/n - 0.65}{\sqrt{(0.65)(0.35)/n}} \geq 1.96$;

(b) $z = \dfrac{414/600 - 0.65}{\sqrt{(0.65)(0.35)/600}} = 2.054 > 1.96$, reject H_0 at $\alpha = 0.025$.

(c) Since the p-value $\approx P(Z \geq 2.054) = 0.0200 < 0.0250$, reject H_0 at an $\alpha = 0.025$ significance level;

(d) A 95% one-sided confidence interval for p is

$$[0.69 - 1.645\sqrt{(0.69)(0.31)/600}, 1] = [0.659, 1].$$

7.1–14 We shall test H_0: $p = 0.20$ against H_1: $p < 0.20$. With a sample size of 15, if the critical region is $C = \{x : x \leq 1\}$, the significance level is $\alpha = 0.1671$. Because $x = 2$, Dr. X has not demonstrated significant improvement with these few data.

7.1–16 (a) $|z| = \dfrac{|\widehat{p} - 0.20|}{\sqrt{(0.20)(0.80)/n}} \geq 1.96$;

(b) Only 5/54 for which $z = -1.973$ leads to rejection of H_0, so 5% reject H_0.

(c) 5%.

(d) 95%.

(e) $z = \dfrac{219/1124 - 0.20}{\sqrt{(0.20)(0.80)/1124}} = -0.43$, so fail to reject H_0.

7.1–18 (a) Under H_0, $\widehat{p} = (351 + 41)/800 = 0.49$;

$$|z| = \frac{|351/605 - 41/195|}{\sqrt{(0.49)(0.51)\left(\dfrac{1}{605} + \dfrac{1}{195}\right)}} = \frac{|0.580 - 0.210|}{0.0412} = 8.99.$$

Since $8.99 > 1.96$, reject H_0.

(b) $0.58 - 0.21 \quad \pm \quad 1.96\sqrt{\dfrac{(0.58)(0.42)}{605} + \dfrac{(0.21)(0.79)}{195}}$

$0.37 \quad \pm \quad 1.96\sqrt{0.000403 + 0.000851}$

$0.37 \quad \pm \quad 0.07 \quad \text{or} \quad [0.30, 0.44]$.

It is in agreement with (a).

(c) $0.49 \quad \pm \quad 1.96\sqrt{(0.49)(0.51)/800}$

$0.49 \quad \pm \quad 0.035 \quad \text{or} \quad [0.455, 0.525]$.

7.1–20 **(a)** $z = \dfrac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\widehat{p}(1-\widehat{p})(1/n_1 + 1/n_2)}} \geq 1.645;$

(b) $z = \dfrac{0.15 - 0.11}{\sqrt{(0.1325)(0.8675)(1/900 + 1/700)}} = 2.341 > 1.645,$ reject H_0.

(c) $z = 2.341 > 2.326,$ reject H_0.

(d) The p-value $\approx P(Z \geq 2.341) = 0.0096.$

7.1–22 **(a)** $P(\text{at least one match}) = 1 - P(\text{no matches}) = 1 - \dfrac{52}{52}\dfrac{51}{52}\dfrac{50}{52}\dfrac{49}{52}\dfrac{48}{52}\dfrac{47}{52} = 0.259.$

7.1–24 $z = \dfrac{204/300 - 0.73}{\sqrt{(0.73)(0.27)/300}} = \dfrac{-0.05}{0.02563} = -1.95;$

p-value $\approx P(Z < -1.95) = 0.0256 < \alpha = 0.05$ so we reject H_0. That is, the test indicates that there is progress.

7.2 Tests about One Mean and One Variance

7.2–2 **(a)** The critical region is
$$z = \frac{\overline{x} - 13.0}{0.2/\sqrt{n}} \leq -1.96;$$

(b) The observed value of z,
$$z = \frac{12.9 - 13.0}{0.04} = -2.5,$$

is less that -1.96 so we reject H_0.

(c) The p-value of this test is $P(Z \leq -2.50) = 0.0062.$

7.2–4 **(a)** $|t| = \dfrac{|\overline{x} - 7.5|}{s/\sqrt{10}} \geq t_{0.025}(9) = 2.262.$
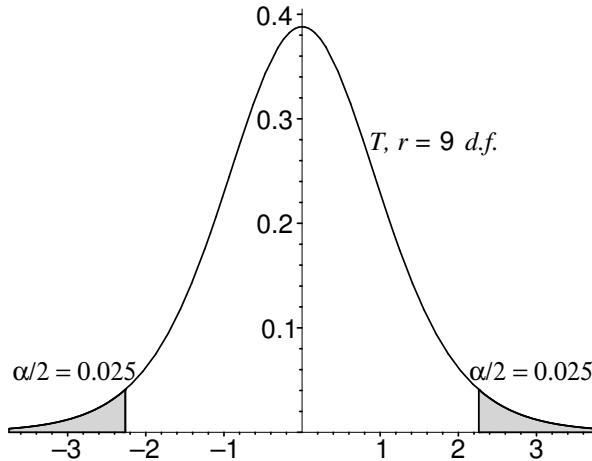
Figure 7.2–4: The critical region is $|t| \geq 2.262$

(b) $|t| = \dfrac{|7.55 - 7.5|}{0.1027/\sqrt{10}} = 1.54 < 2.262,$ do not reject H_0.

(c) A 95% confidence interval for μ is
$$\left[7.55 - 2.262\left(\frac{0.1027}{\sqrt{10}}\right), \; 7.55 + 2.262\left(\frac{0.1027}{\sqrt{10}}\right)\right] = [7.48, \; 7.62].$$

Hence, $\mu = 7.50$ is contained in this interval. We could have obtained the same conclusion from our answer to part (b).

7.2–6 **(a)** H_0: $\mu = 3.4$;

(b) H_1: $\mu > 3.4$;

(c) $t = (\bar{x} - 3.4)/(s/3)$;

(d) $t \geq 1.860$;
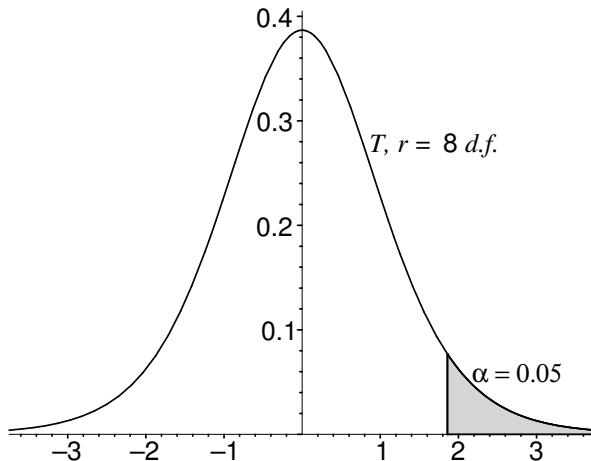
Figure 7.2–6: The critical region is $t \geq 1.860$

(e) $t = \dfrac{3.556 - 3.4}{0.167/3} = 2.802$;

(f) $2.802 > 1.860$, reject H_0;

(g) $0.01 < p\text{-value} < 0.025$, $p\text{-value} = 0.0116$.

7.2–8 **(a)** $t = \dfrac{\bar{x} - 3315}{s/\sqrt{11}} \geq 2.764$;

(b) $t = \dfrac{3385.91 - 3315}{336.32/\sqrt{11}} = 0.699 < 2.764$, do not reject H_0;

(c) $p\text{-value} \approx 0.25$ because $t_{0.25}(10) = 0.700$.

7.2–10 **(a)** $|t| = \dfrac{|\overline{x} - 125|}{s/\sqrt{15}} \geq t_{0.025}(14) = 2.145.$
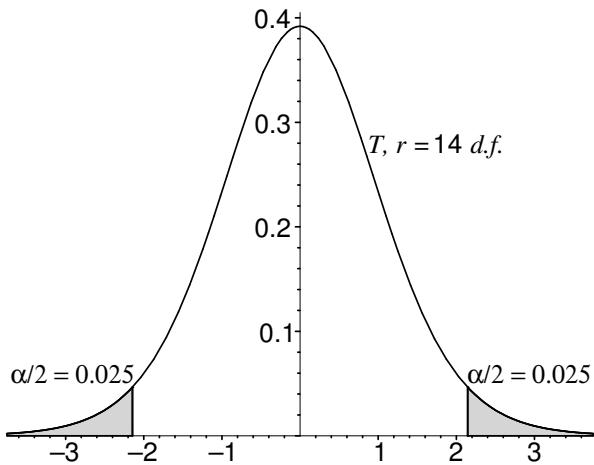
Figure 7.2–10: The critical region is $|t| \geq 2.145$

(b) $|t| = \dfrac{|127.667 - 125|}{9.597/\sqrt{15}} = 1.076 < 2.145,$ do not reject H_0.

7.2–12 **(a)** The test statistic and critical region are given by

$$t = \frac{\overline{x} - 5.70}{s/\sqrt{8}} \geq 1.895.$$

(b) The observed value of the test statistic is

$$t = \frac{5.869 - 5.70}{0.19737/\sqrt{8}} = 2.42.$$

(c) The p-value is a little less than 0.025. Using Minitab, the p-value $= 0.023$.
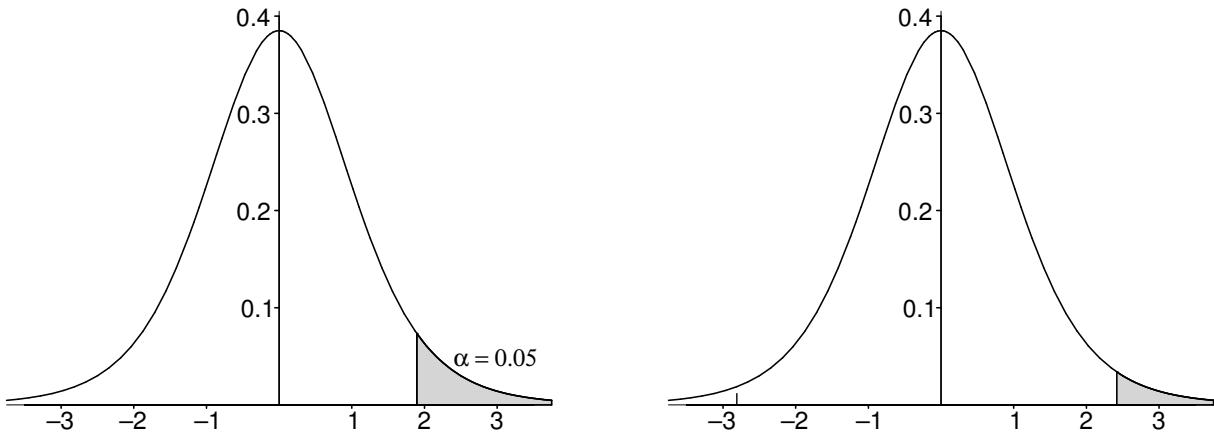
Figure 7.2–12: A $T(7)$ p.d.f. showing the critical region on the left, p-value on the right

7.2–14 The critical region is

$$t = \frac{\bar{d} - 0}{s_d/\sqrt{17}} \geq 1.746.$$

Since $\bar{d} = 4.765$ and $s_d = 9.087$, $t = 2.162 > 1.746$ and we reject H_0.

7.2–16 **(a)** The critical region is

$$t = \frac{\bar{d} - 0}{s_d/\sqrt{20}} \leq -1.729.$$
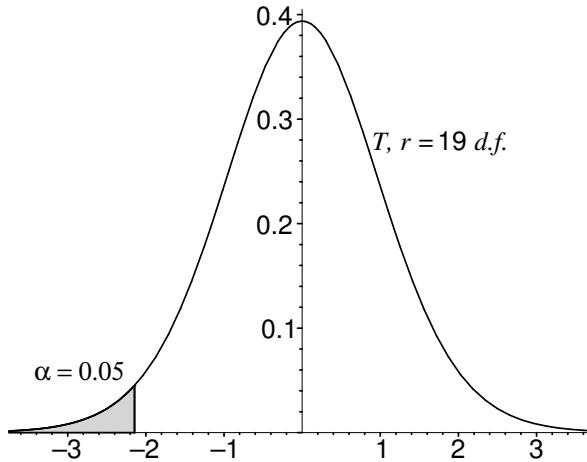
Figure 7.2–20: The critical region is $t \leq -1.729$

(b) Since $\bar{d} = -0.290$, $s_d = 0.6504$, $t = -1.994 < -1.729$, so we reject H_0.

(c) Since $t = -1.994 > -2.539$, we would fail to reject H_0.

(d) From Table VI, $0.025 < p\text{-value} < 0.05$. In fact, $p\text{-value} = 0.0304$.

7.3   Tests of the Equality of Two Means

7.3–2 **(a)** $t = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{15s_x^2 + 12s_y^2}{27}\left(\dfrac{1}{16} + \dfrac{1}{13}\right)}} \leq t_{0.01}(27) = 2.473;$

(b) $t = \dfrac{415.16 - 347.40}{\sqrt{\dfrac{15(1356.75) + 12(692.21)}{27}\left(\dfrac{1}{16} + \dfrac{1}{13}\right)}} = 5.570 > 2.473$, reject H_0.

(c) $c = \dfrac{1356.75}{1356.75 + 692.21} = 0.662,$

$\dfrac{1}{r} = \dfrac{0.662^2}{15} + \dfrac{0.338^2}{12} = 0.0387,$

$r = 25.$

The critical region is therefore $t \geq t_{0.01}(25) = 2.485$. Since $t = 5.570 > 2.485$, we again reject H_0.

7.3–4 **(a)** $t = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{12s_x^2 + 15s_y^2}{27}\left(\dfrac{1}{13} + \dfrac{1}{16}\right)}} \leq -t_{0.05}(27) = -1.703;$

(b) $t = \dfrac{72.9 - 81.7}{\sqrt{\dfrac{(12)(25.6)^2 + (15)(28.3)^2}{27}\left(\dfrac{1}{13} + \dfrac{1}{16}\right)}} = -0.869 > -1.703$, do not reject H_0;

(c) $0.10 < p$-value < 0.25;

7.3–6 (a) Assuming $\sigma_X^2 = \sigma_Y^2$,

$$|t| = \dfrac{|\bar{x} - \bar{y}|}{\sqrt{\dfrac{9s_x^2 + 9s_y^2}{18}\left(\dfrac{1}{10} + \dfrac{1}{10}\right)}} \geq t_{0.025}(18) = 2.101;$$

(b) $|-2.151| > 2.101$, reject H_0;

(c) $0.01 < p$-value < 0.05;

(d)
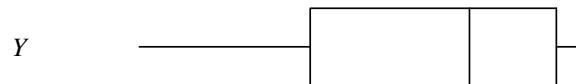
Figure 7.3–6: Box-and-whisker diagram for stud 3 (X) and stud 4 (Y) forces

7.3–8 (a) For these data, $\bar{x} = 1511.7143$, $\bar{y} = 1118.400$, $s_x^2 = 49{,}669.90476$, $s_y^2 = 15297.6000$. If we assume equal variances,

$$t = \dfrac{|\bar{x} - \bar{y}|}{\sqrt{\dfrac{6s_x^2 + 9s_y^2}{15}\left(\dfrac{1}{7} + \dfrac{1}{10}\right)}} = 4.683 > 2.131 = t_{0.025}(15)$$

and we reject $\mu_X = \mu_Y$.

If we use the approximating t and Welch's formula for the number of degrees of freedom given by Equation 6.3-1 in the text, $r = \lfloor 8.599 \rfloor = 8$ degrees of freedom. We then have that $t = 4.683 > t_{0.025}(8) = 2.306$ and we reject H_0.

(b) No and yes so that the answers are compatible.

7.3–10 $t = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{24s_x^2 + 28s_y^2}{52}\left(\dfrac{1}{25} + \dfrac{1}{29}\right)}} = 3.402 > 2.326 = z_{0.01}$,

reject $\mu_X = \mu_Y$.

7.3–12 **(a)** $t = \dfrac{8.0489 - 8.0700}{\sqrt{\dfrac{8(0.00139) + 8(0.00050)}{16}}\sqrt{\dfrac{1}{9} + \dfrac{1}{9}}} = -1.46.$ Since $-1.337 < -1.46 < -1.746,$

$0.05 < p$-value $< 0.10.$ In fact, p-value $= 0.082.$ We would fail to reject H_0 at an $\alpha = 0.05$ significance level but we would reject at $\alpha = 0.10.$

(b) The following figure confirms our answer.
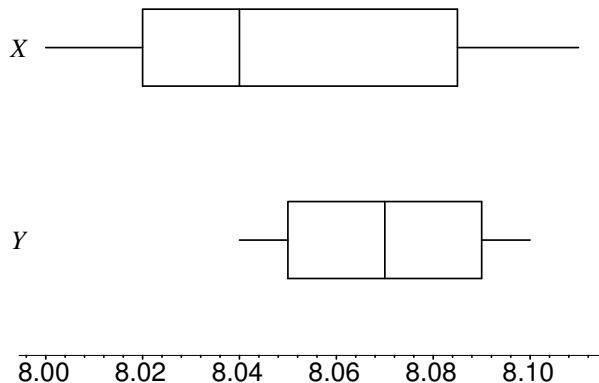
Figure 7.3–12: Box-and-whisker diagram for lengths of columns

7.3–14 $t = \dfrac{4.1633 - 5.1050}{\sqrt{\dfrac{11(0.91426) + 7(2.59149)}{18}}\sqrt{\dfrac{1}{12} + \dfrac{1}{8}}} = -1.648.$ Since $-1.330 < -1.648 < -1.734,$

$0.05 < p$-value $< 0.10.$ In fact, p-value $= 0.058.$ We would fail to reject H_0 at an $\alpha = 0.05$ significance level.

7.3–16 **(a)** $\dfrac{\bar{y} - \bar{x}}{\sqrt{\dfrac{s_y^2}{30} + \dfrac{s_x^2}{30}}} > 1.96;$

(b) $8.98 > 1.96,$ reject $\mu_X = \mu_Y.$

(c) Yes.
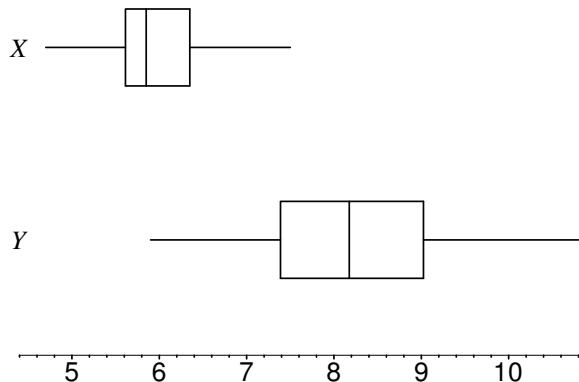
Figure 7.3–16: Lengths of male (X) and female (Y) green lynx spiders

7.3–18 **(a)** For these data, $\bar{x} = 5.9947$, $\bar{y} = 4.3921$, $s_x^2 = 6.0191$, $s_y^2 = 1.9776$. Using the number of degrees of freedom given by Equation 6.3-1 (Welch) we have that $r = \lfloor 28.68 \rfloor = 28$. We have

$$t = \frac{5.9947 - 4.3921}{\sqrt{6.0191/19 + 1.9776/19}} = 2.47 > 2.467 = t_{0.01}(28)$$

so we reject H_0.

(b)
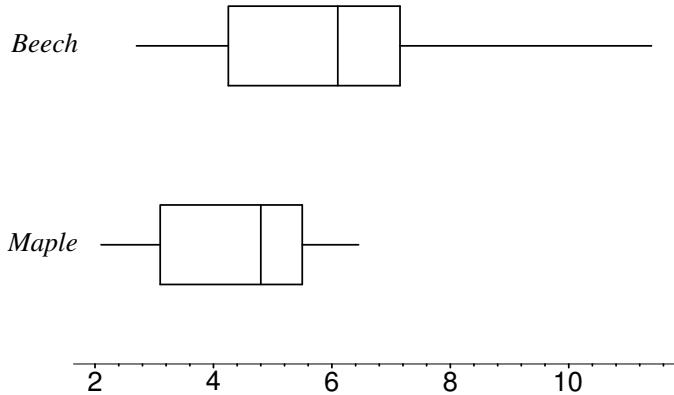
Figure 7.3–18: Tree dispersion distances in meters

7.3–20 **(a)** For these data, $\bar{x} = 5.128$, $\bar{y} = 4.233$, $s_x^2 = 1.2354$, $s_y^2 = 1.2438$. Since $t = 2.46$, we clearly reject H_0. Minitab gives a p-value of 0.01.

(b)
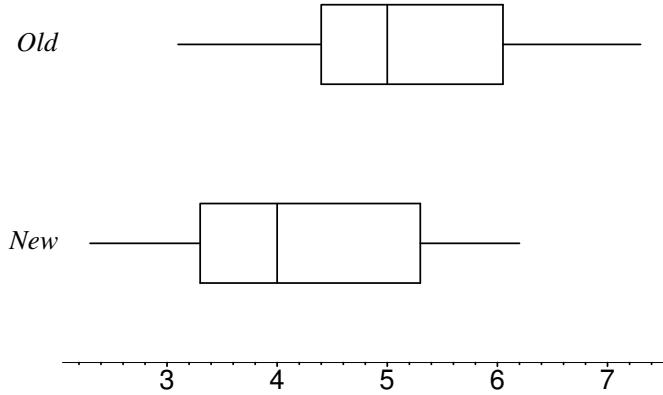
Figure 7.3–20: Times for old procedure and new prodecure

(c) We reject H_0 and conclude that the response times for the new procedure are less than for the old procedure.

7.4 Tests for Variances

7.4–2 **(a)** Reject H_0 if $\chi^2 = \dfrac{10 s_y^2}{525^2} \le \chi^2_{0.95}(10) = 3.940$.

The observed value of the test statistic, $\chi^2 = 4.223 > 3.940$, so we fail to reject H_0.

(b) $F = \dfrac{s_x^2}{s_y^2} = \dfrac{113108.4909}{116388.8545} = 0.9718$ so we clearly accept the equality of the variances.

(c) The critical region is $|t| \geq t_{0.025}(20) = 2.086$.

$$t = \frac{\bar{x} - \bar{y} - 0}{\sqrt{s_x^2/11 + s_y^2/11}} = \frac{3385.909 - 3729.364}{144.442} = -2.378.$$

Since $|-2.378| > 2.086$, we reject the null hypothesis. The p-value for this test is 0.0275.

7.4–4 **(a)** The critical region is

$$\chi^2 = \frac{19s^2}{(0.095)^2} \leq 10.12.$$

The observed value of the test statistic,

$$\chi^2 = \frac{19(0.065)^2}{(0.095)^2} = 8.895,$$

is less than 10.12, so the company was successful.

(b) Since $\chi_{0.975}^2(19) = 8.907$, p-value ≈ 0.025.

7.4–6 $\mathrm{Var}(S^2) = \mathrm{Var}\left(\frac{100}{22} \cdot \frac{22S^2}{100}\right) = \left(\frac{100}{22}\right)^2 (2)(22) = 10{,}000/11$.

7.4–8 $\dfrac{s_x^2}{s_y^2} = \dfrac{9.88}{4.08} = 2.42 < 3.28 = F_{0.05}(12, 8)$, so fail to reject H_0.

7.4–10 $F = \dfrac{9201}{4856} = 1.895 < 3.37 = F_{0.05}(6, 9)$ so we fail to reject H_0.

7.5 One-Factor Analysis of Variance

7.5–2

Source	SS	DF	MS	F	p-value
Treatment	388.2805	3	129.4268	4.9078	0.0188
Error	316.4597	12	26.3716		
Total	704.7402	15			

$F = 4.9078 > 3.49 = F_{0.05}(3, 12)$,   reject H_0.

7.5–4

Source	SS	DF	MS	F	p-value
Treatment	150	2	75	75	0.00006
Error	6	6	1		
Total	156	8			

7.5–6

Source	SS	DF	MS	F	p-value
Treatment	184.8	2	92.4	15.4	0.00015
Error	102.0	17	6.0		
Total	286.8	19			

$F = 15.4 > 3.59 = F_{0.05}(2, 17)$,   reject H_0.

7.5–8 **(a)** $F \geq F_{0.05}(3, 24) = 3.01$;

(b)

Source	SS	DF	MS	F	p-value
Treatment	12,280.86	3	4,093.62	3.455	0.0323
Error	28,434.57	24	1,184.77		
Total	40,715.43	27			

$F = 3.455 > 3.01$, reject H_0;

(c) $0.025 < p\text{-value} < 0.05$.

(d)
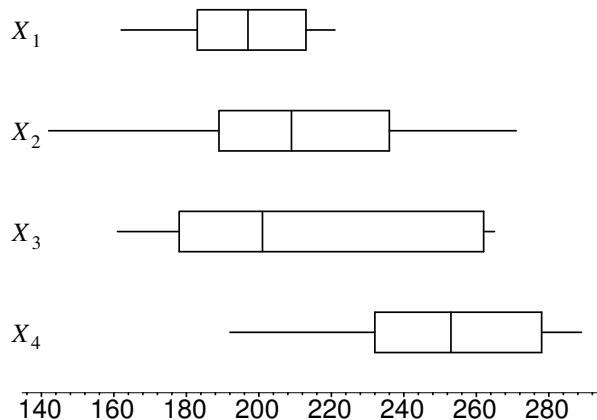
Figure 7.5–8: Box-and-whisker diagrams for cholesterol levels

7.5–10 **(a)** $F \geq F_{0.05}(4, 30) = 2.69$;

(b)

Source	SS	DF	MS	F	p-value
Treatment	0.00442	4	0.00111	2.85	0.0403
Error	0.01157	30	0.00039		
Total	0.01599	34			

$F = 2.85 > 2.69$, reject H_0;

(c)
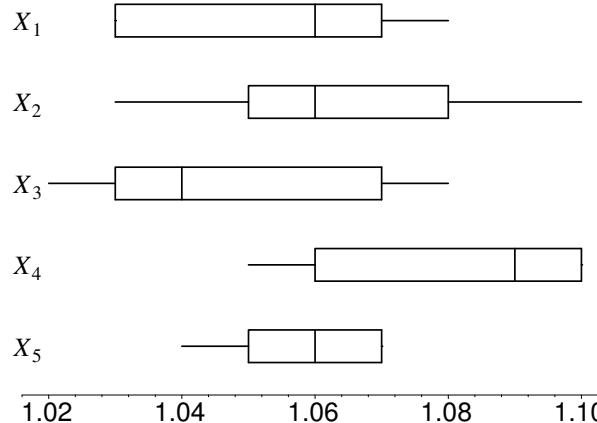
Figure 7.5–10: Box-and-whisker diagrams for nail weights

7.5–12 **(a)** $t = \dfrac{92.143 - 103.000}{\sqrt{\dfrac{6(69.139) + 6(57.669)}{12}\left(\dfrac{1}{7} + \dfrac{1}{7}\right)}} = -2.55 < -2.179$, reject H_0.

$F = \dfrac{412.517}{63.4048} = 6.507 > 4.75$, reject H_0.

The F and the t tests give the same results since $t^2 = F$.

(b) $F = \dfrac{86.3336}{114.8889} = 0.7515 < 3.55$, do not reject H_0.

7.5–14 **(a)**

Source	SS	DF	MS	F	p-value
Treatment	122.1956	2	61.0978	2.130	0.136
Error	860.4799	30	28.6827		
Total	982.6755	32			

$F = 2.130 < 3.32 = F_{0.05}(2, 30)$, fail to reject H_0;

(b)

Figure 7.5–14: Box-and-whisker diagrams for resistances on three days

7.5–16 **(a)**

Source	SS	DF	MS	F	p-value
Worker	1.5474	2	0.7737	1.0794	0.3557
Error	17.2022	24	0.7168		
Total	18.7496	26			

$F = 1.0794 < 3.40 = F_{0.05}(2, 24)$, fail to reject H_0;

(b)
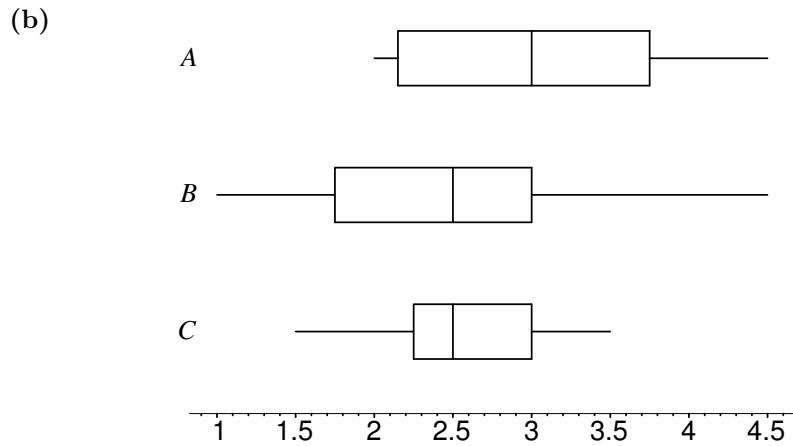
Figure 7.5–16: Box-and-whisker diagrams for workers A, B, and C

The box plot confirms the answer from part (a).

7.6   Two-Factor Analysis of Variance

7.6–2

				$\mu + \alpha_i$
6	3	7	8	6
10	7	11	12	10
8	5	9	10	8
$\mu + \beta_j$ 8	5	9	10	$\mu = 8$

So $\alpha_1 = -2$, $\alpha_2 = 2$, $\alpha_3 = 0$ and $\beta_1 = 0$, $\beta_2 = -3$, $\beta_3 = 1$, $\beta_4 = 2$.

7.6–4 $\displaystyle\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})(X_{ij} - \overline{X}_{i\cdot} - \overline{X}_{\cdot j} + \overline{X}_{\cdot\cdot})$

$$= \sum_{i=1}^{a}(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})\sum_{j=1}^{b}[(X_{ij} - \overline{X}_{i\cdot}) - (\overline{X}_{\cdot j} - \overline{X}_{\cdot\cdot})]$$

$$= \sum_{i=1}^{a}(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})\left\{\sum_{j=1}^{b}(X_{ij} - \overline{X}_{i\cdot}) - \sum_{j=1}^{b}(\overline{X}_{\cdot j} - \overline{X}_{\cdot\cdot})\right\}$$

$$= \sum_{i=1}^{a}(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})(0 - 0) = 0;$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{X}_{\cdot j} - \overline{X}_{\cdot\cdot})(X_{ij} - \overline{X}_{i\cdot} - \overline{X}_{\cdot j} + \overline{X}_{\cdot\cdot}) = 0, \text{ similarly;}$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})(\overline{X}_{\cdot j} - \overline{X}_{\cdot\cdot}) = \left\{\sum_{i=1}^{a}(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})\right\}\left\{\sum_{j=1}^{b}(\overline{X}_{\cdot j} - \overline{X}_{\cdot\cdot})\right\} = (0)(0) = 0 \;.$$

7.6–6

				$\mu + \alpha_i$
6	7	7	12	8
10	3	11	8	8
8	5	9	10	8
$\mu + \beta_j$ 8	5	9	10	$\mu = 8$

So $\alpha_1 = \alpha_2 = \alpha_3 = 0$ and $\beta_1 = 0$, $\beta_2 = -3$, $\beta_3 = 1$, $\beta_4 = 2$ as in Exercise 8.7–2. However, $\gamma_{11} = -2$ because $8 + 0 + 0 + (-2) = 6$. Similarly we obtain the other γ_{ij}'s :

$$
\begin{array}{rrrr}
-2 & 2 & -2 & 2 \\
2 & -2 & 2 & -2 \\
0 & 0 & 0 & 0
\end{array}
$$

7.6–8

Source	SS	DF	MS	F	p-value
Row (A)	99.7805	3	49.8903	4.807	0.021
Col (B)	70.1955	1	70.1955	6.763	0.018
Int(AB)	202.9827	2	101.4914	9.778	0.001
Error	186.8306	18	10.3795		
Total	559.7894	23			

Since $F_{AB} = 9.778 > 3.57$, H_{AB} is rejected. Most statisticians would probably not proceed to test H_A and H_B.

7.6–10

Source	SS	DF	MS	F	p-value
Row (A)	5,103.0000	1	5,103.0000	4.307	0.049
Col (B)	6,121.2857	1	6,121.2857	5.167	0.032
Int(AB)	1,056.5714	1	1,056.5714	0.892	0.354
Error	28,434.5714	24	1,184.7738		
Total	40,715.4286	27			

(a) Since $F = 0.892 < F_{0.05}(1, 24) = 4.26$, do not reject H_{AB};

(b) Since $F = 4.307 > F_{0.05}(1, 24) = 4.26$, reject H_A;

(c) Since $F = 5.167 > F_{0.05}(1, 24) = 4.26$, reject H_B.

7.7   Tests Concerning Regression and Correlation

7.7–2 The critical region is $t_1 \geq t_{0.25}(8) = 2.306$. From Exercise 7.8–2,

$$
\widehat{\beta} = 4.64/5.04 \text{ and } \widehat{n\sigma^2} = 1.84924; \text{ also } \sum_{i=1}^{10} (x_i - \overline{x})^2 = 5.04, \text{ so}
$$

$$
t_1 = \frac{4.64/5.04}{\sqrt{\dfrac{1.84924}{8(5.04)}}} = \frac{0.9206}{0.2142} = 4.299.
$$

Since $t_1 = 4.299 > 2.306$, we reject H_0.

7.7–4 The critical region is $t_1 \geq t_{0.01}(18) = 2.552$. Since

$$\widehat{\beta} = \frac{24.8}{40}, \quad n\widehat{\sigma^2} = 5.1895, \quad \text{and} \quad \sum_{i=1}^{10}(x_1 - \bar{x})^2 = 40,$$

it follows that

$$t_1 = \frac{24.8/40}{\sqrt{\dfrac{5.1895}{18(40)}}} = 7.303.$$

Since $t_1 = 7.303 > 2.552$, we reject H_0. We could also construct the following table. Output like this is given by Minitab.

Source	SS	DF	MS	F	p-value
Regression	15.3760	1	15.3760	53.3323	0.0000
Error	5.1895	18	0.2883		
Total	20.5655	19			

Note that $t_1^2 = 7.303^2 = 53.3338 \approx F = 53.3323$.

7.7–6 For these data, $r = -0.413$. Since $|r| = 0.413 < 0.7292$, do not reject H_0.

7.7–8 Following the suggestion given in the hint, the expression equals

$$(n-1)S_Y^2 - \frac{2Rs_xS_Y}{s_x^2}(n-1)Rs_xS_Y + \frac{R^2s_x^2S_Y^2(n-1)s_x^2}{s_x^2} \quad = \quad (n-1)S_Y^2(1 - 2R^2 + R^2)$$

$$= \quad (n-1)S_Y^2(1 - R^2).$$

7.7–10
$$u(R) \approx u(\rho) + (R-\rho)u'(\rho),$$

$$\mathrm{Var}[u(\rho) + (R-\rho)u'(\rho)] = [u'(\rho)]^2\mathrm{Var}(R)$$

$$= [u'(\rho)]^2\frac{(1-\rho^2)^2}{n} = c, \quad \text{which is free of } \rho,$$

$$u'(\rho) = \frac{k/2}{1-\rho} + \frac{k/2}{1+\rho},$$

$$u(\rho) = -\frac{k}{2}\ln(1-\rho) + \frac{k}{2}\ln(1+\rho) = \frac{k}{2}\ln\left(\frac{1+\rho}{1-\rho}\right).$$

Thus, taking $k = 1$,

$$u(R) = \left(\frac{1}{2}\right)\ln\left[\frac{1+R}{1-R}\right]$$

has a variance almost free of ρ.

7.7–12 (a) $r = -0.4906, |r| = 0.4906 > 0.4258$, reject H_0 at $\alpha = 0.10$;

(b) $|r| = 0.4906 < 0.4973$, fail to reject H_0 at $\alpha = 0.05$.

7.7–14 (a) $r = 0.339, |r| = 0.339 < 0.5325 = r_{0.025}(12)$, fail to reject H_0 at $\alpha = 0.05$;

(b) $r = -0.821 < -0.6613 = r_{0.005}(12)$, reject H_0 at $\alpha = 0.005$;

(c) $r = 0.149, |r| = 0.149 < 0.5325 = r_{0.025}(12)$, fail to reject H_0 at $\alpha = 0.05$.

Chapter 8

Nonparametric Methods

8.1 Chi-Square Goodness of Fit Tests

8.1–2 $q_4 = \dfrac{(224-232)^2}{232} + \dfrac{(119-116)^2}{116} + \dfrac{(130-116)^2}{116} + \dfrac{(48-58)^2}{58} + \dfrac{(59-58)^2}{58}$

$= 3.784.$

The null hypothesis will not be rejected at any reasonable significance level. Note that $E(Q_4) = 4$ when H_0 is true.

8.1–4 $q_3 = \dfrac{(124-117)^2}{117} + \dfrac{(30-39)^2}{39} + \dfrac{(43-39)^2}{39} + \dfrac{(11-13)^2}{13}$

$= 0.419 + 2.077 + 0.410 + 0.308 = 3.214 < 7.815 = \chi^2_{0.05}(3).$

Thus we do not reject the Mendelian theory with these data.

8.1–6 We first find that $\hat{p} = 274/425 = 0.6447$. Using Table II with $p = 0.65$ the hypothesized probabilities are $p_1 = P(X \le 1) = 0.0540$, $p_2 = P(X = 2) = 0.1812$, $p_3 = P(X = 3) = 0.3364$, $p_4 = P(X = 4) = 0.3124$, $p_5 = P(X = 5) = 0.1160$. Thus the respective expected values are 4.590, 15.402, 28.594, 26.554, and 9.860. One degree of freedom is lost because p was estimated. The value of the chi-square goodness of fit statistic is:

$q = \dfrac{(6-4.590)^2}{4.590} + \dfrac{(13-15.402)^2}{15.402} + \dfrac{(30-28.594)^2}{28.594} + \dfrac{(28-26.554)^2}{26.554} + \dfrac{(8-9.860)^2}{9.860}$

$= 1.3065 < 7.815 = \chi^2_{0.05}(3)$

Do not reject the hypothesis that X is $b(5, p)$. The 95% confidence interval for p is

$$0.6447 \pm 1.96\sqrt{(0.6447)(0.3553)/425} \quad \text{or} \quad [0.599, 0.690].$$

The pennies that were used were minted 1998 or earlier. See Figure 8.1-6. Repeat this experiment with similar pennies or with newer pennies and compare your results with those obtained by these students.
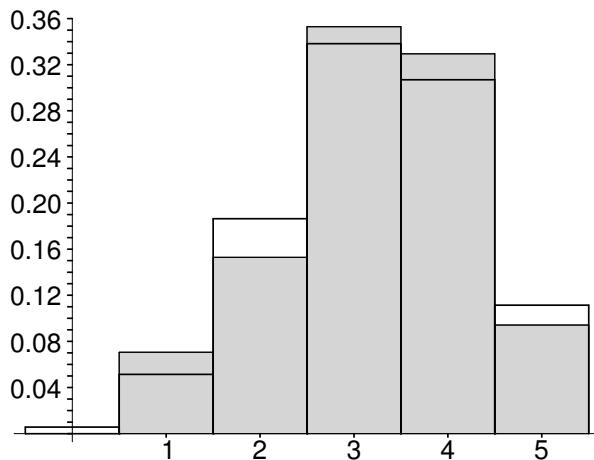
Figure 8.1–6: The $b(5, 0.65)$ probability histogram and the relative frequency histogram (shaded)

8.1–8 The respective probabilities and expected frequencies are 0.050, 0.149, 0.224, 0.224, 0.168, 0.101, 0.050, 0.022, 0.012 and 15.0, 44.7, 67.2, 67.2, 50.4, 30.3, 15.0, 6.6, 3.6. The last two cells could be combined to give an expected frequency of 10.2. From Exercise 3.5–12, the respective frequencies are 17, 47, 63, 63, 49, 28, 21, and 12 giving

$$q_7 = \frac{(17 - 15.0)^2}{15.0} + \frac{(47 - 44.7)^2}{44.7} + \cdots + \frac{(12 - 10.2)^2}{10.2} = 3.841.$$

Since $3.841 < 14.07 = \chi^2_{0.05}(7)$, do not reject. The sample mean is $\overline{x} = 3.03$ and the sample variance is $s^2 = 3.19$ which also supports the hypothesis. The following figure compares the probability histogram with the relative frequency histogram of the data.
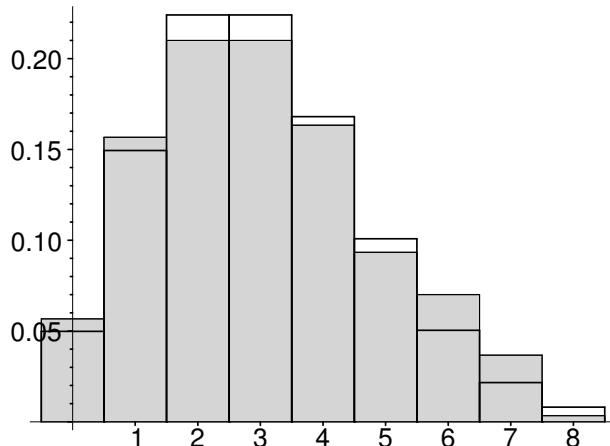
Figure 8.1–8: The Poisson probability histogram, $\lambda = 3$, and relative frequency histogram (shaded)

8.1–10 We shall use 10 sets of equal probability.

A_i	Observed	Expected	q
(0.00, 4.45)	8	9	1/9
[4.45, 9.42)	10	9	1/9
[9.42, 15.05)	9	9	0/9
[15.05, 21.56)	8	9	1/9
[21.56, 29.25)	7	9	4/9
[29.25, 38.67)	11	9	4/9
[38.67, 50.81)	8	9	1/9
[50.81, 67.92)	12	9	9/9
[67.92, 91.17)	10	9	1/9
[91.17, ∞)	7	9	4/9
	90	90	26/9=2.89

Since $2.89 < 15.51 = \chi^2_{0.05}(8)$, we accept the hypothesis that the distribution of X is exponential. Note that one degree of freedom is lost because we had to estimate θ.
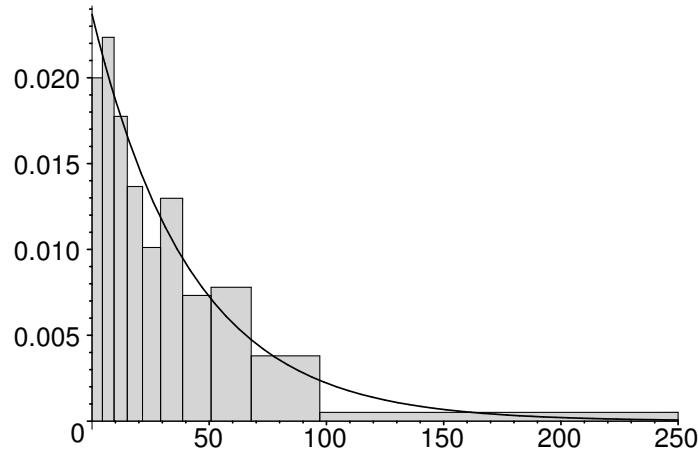
Figure 8.1–10: Exponential p.d.f., $\widehat{\theta} = 42.2$, and relative frequency histogram (shaded)

8.1–12 We shall use 10 sets of equal probability.

A_i	Observed	Expected	q
(−∞, 399.40)	10	9	1/9
[399.40, 437.92)	7	9	4/9
[437.92, 465.71)	9	9	0/9
[465.71, 489.44)	9	9	0/9
[489.44, 511.63)	13	9	16/9
[511.63, 533.82)	8	9	1/9
[533.82, 557.55)	7	9	4/9
[557.55, 585.34)	6	9	9/9
[585.34, 623.86)	11	9	4/9
[623.86, ∞)	10	9	1/9
	90	90	40/9=4.44

Since $4.44 < 14.07 = \chi^2_{0.05}(7)$, we accept the hypothesis that the distribution of X is $N(\mu, \sigma^2)$. Note that 2 degrees of freedom are lost because 2 parameters were estimated.
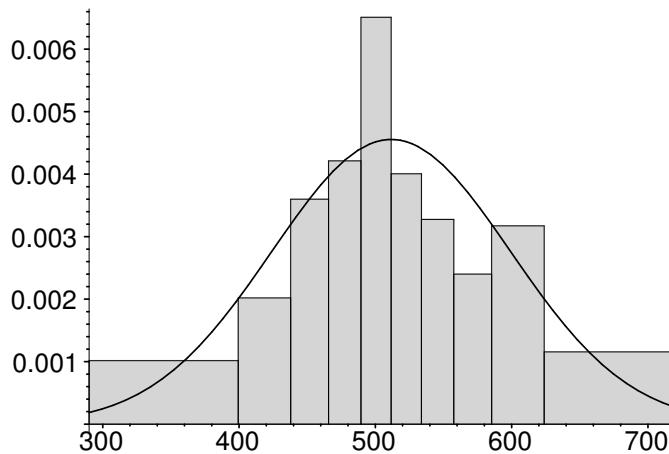
Figure 8.1–12: The $N(511.633, 87.576^2)$ p.d.f. and the relative frequency histogram (shaded)

8.1–14 (a) We shall use 5 classes with equal probability.

A_i	Observed	Expected	q
$[0, 25.25)$	4	6.2	0.781
$[25.25,\ 57.81)$	9	6.2	1.264
$[57.81, 103.69)$	5	6.2	0.232
$[193.69, 182.13)$	6	6.2	0.006
$[182.13, \infty)$	7	6.2	0.103
	31	31.0	2.386

The p-value for $5 - 1 - 1 = 3$ degrees of freedom is 0.496 so we fail to reject the null hypothesis.

(b) We shall use 10 classes with equal probability.

A_i	Observed	Expected	q
$[0, 22.34)$	3	3.9	0.208
$[22.34,\ 34.62)$	3	3.9	0.208
$[34.62, 46.09)$	8	3.9	4.310
$[46.09, 57.81)$	4	3.9	0.003
$[57.81, 70.49)$	2	3.9	0.926
$[70.49, 84.94)$	4	3.9	0.003
$[84.94, 102.45)$	4	3.9	0.003
$[102.45, 125.76)$	4	3.9	0.003
$[125.76, 163.37)$	2	3.9	0.926
$[163.37, \infty)$	5	3.9	0.310
	39	39.0	6.900

The p-value for $10 - 1 = 9$ degrees of freedom is 0.648 so we fail to reject the null hypothesis.

8.1–16 We shall use 10 classes with equal probability. For these data, $\bar{x} = 5.833$ and $s^2 = 2.7598$.

A_i	Observed	Expected	q
[0, 3.704)	8	10	0.4
[3.704, 4.435)	10	10	0.0
[4.435, 4.962)	11	10	0.1
[4.962, 5.412)	20	10	10.0
[5.412, 5.833)	7	10	0.9
[5.833, 6.254)	8	10	0.4
[6.254, 6.704)	12	10	0.4
[6.704, 7.231)	4	10	0.4
[7.231, 7.962)	8	10	0.4
[7.962, 12)	12	10	0.4
	100	100	16.6

The p-value for $10 - 1 - 1 - 1 = 7$ degrees of freedom is 0.0202 so we reject the null hypothesis.

8.2 Contingency Tables

8.2–2 $10.18 < 20.48 = \chi^2_{0.025}(10)$, accept H_0.

8.2–4 In the combined sample of 45 observations, the lower third includes those with scores of 61 or lower, the middle third have scores from 62 through 78, and the higher third are those with scores of 79 and above.

	low	middle	high	Totals
Class U	9	4	2	15
	(5)	(5)	(5)	
Class V	5	5	5	15
	(5)	(5)	(5)	
Class W	1	6	8	15
	(5)	(5)	(5)	
Totals	15	15	15	45

Thus

$$q = 3.2 + 0.2 + 1.8 + 0 + 0 + 0 + 3.2 + 0.2 + 1.8 = 10.4.$$

Since

$$q = 10.4 > 9.488 = \chi^2_{0.05}(4),$$

we reject the equality of these three distributions. (p-value = 0.034.)

8.2–6 $q = 8.410 < 9.488 = \chi^2_{0.05}$, fail to reject H_0. (p-value = 0.078.)

8.2–8 $q = 4.268 > 3.841 = \chi^2_{0.05}(1)$, reject H_0. (p-value = 0.039.)

8.2–10 $q = 7.683 < 9.210 = \chi^2_{0.01}$, fail to reject H_0. (p-value = 0.021.)

8.2–12 (a) $q = 8.006 > 7.815 = \chi^2_{0.05}(3)$, reject H_0.
(b) $q = 8.006 < 9.348 = \chi^2_{0.025}(3)$, fail to reject H_0. (p-value = 0.046.)

8.2–14 $q = 8.792 > 7.378 = \chi^2_{0.025}(2)$, reject H_0. (p-value = 0.012.)

8.2–16 $q = 4.242 < 4.605 = \chi^2_{0.10}(2)$, fail to reject H_0. (p-value = 0.120.)

8.3   Order Statistics

8.3–2 **(a)** The location of the median is $(0.5)(17+1) = 9$, thus the median is

$$\widetilde{m} = 5.2.$$

The location of the first quartile is $(0.25)(17+1) = 4.5$. Thus the first quartile is

$$\widetilde{q}_1 = (0.5)(4.3) + (0.5)(4.7) = 4.5.$$

The location of the third quartile is $(0.75)(17+1) = 13.5$. Thus the third quartile is

$$\widetilde{q}_3 = (0.5)(5.6) + (0.5)(5.7) = 5.65.$$

(b) The location of the 35th percentile is $(0.35)(18) = 6.3$. Thus

$$\widetilde{\pi}_{0.35} = (0.7)(4.8) + (0.3)(4.9) = 4.83.$$

The location of the 65th percentile is $(0.65)(18) = 11.7$. Thus

$$\widetilde{\pi}_{0.65} = (0.3)(5.6) + (0.7)(5.6) = 5.6.$$

8.3–4 $\displaystyle g(y) \;=\; \sum_{k=3}^{5} \left\{ \frac{6!}{k!(6-k)!} (k)[F(y)]^{k-1} f(y)[1 - F(y)]^{6-k} \right.$

$$\left. + \frac{6!}{k!(6-k)!} [F(y)]^k (6-k)[1 - F(y)]^{6-k-1}[-f(y)] \right\} + 6[F(y)]^5 f(y)$$

$$= \frac{6!}{2!3!} [F(y)]^2 f(y)[1 - F(y)]^3 - \frac{6!}{3!2!} [F(y)]^3 [1 - F(y)]^2 f(y)$$

$$+ \frac{6!}{3!2!} [F(y)]^3 f(y)[1 - F(y)]^2 - \frac{6!}{4!1!} [F(y)]^4 [1 - F(y)]^1 f(y)$$

$$+ \frac{6!}{4!1!} [F(y)]^4 f(y)[1 - F(y)]^1 - \frac{6!}{5!0!} [F(y)]^5 [1 - F(y)]^0 f(y) + 6[F(y)]^5 f(y)$$

$$= \frac{6!}{2!3!} [F(y)]^2 [1 - F(y)]^3 f(y), \qquad a < y < b.$$

8.3–6 **(a)** $f(x) \;=\; x, \qquad 0 < x < 1.$ Thus

$$g_1(w) \;=\; n[1 - w]^{n-1}(1), \qquad 0 < w < 1;$$

$$g_n(w) \;=\; n[w]^{n-1}(1), \qquad 0 < w < 1.$$

(b) $\displaystyle E(W_1) \;=\; \int_0^1 (w)(n)(1-w)^{n-1}\, dw$

$$= \left[-w(1-w)^n - \frac{1}{n+1}(1-w)^{n+1} \right]_0^1 = \frac{1}{n+1}.$$

$$E(W_n) \;=\; \int_0^1 (w)(n)w^{n-1}dw = \left[\frac{n}{n+1} w^{n+1} \right]_0^1 = \frac{n}{n+1}.$$

(c) Let $w = w_r$. The p.d.f. of W_r is

$$g_r(w) \;=\; \frac{n!}{(r-1)!(n-r)!} [w]^{r-1}[1-w]^{n-r} \cdot 1$$

$$= \frac{\Gamma(r + n - r + 1)}{\Gamma(r)\Gamma(n - r + 1)} w^{r-1}(1-w)^{n-r+1-1}.$$

Thus W_r has a beta distribution with $\alpha = r,\ \beta = n - r$.

8.3–8 (a)

$$E(W_r^2) = \int_0^1 w^2 \frac{n!}{(r-1)!(n-r)!} w^{r-1}(1-w)^{n-r}\,dw$$

$$= \frac{r(r+1)}{(n+2)(n+1)} \int_0^1 \frac{(n+2)!}{(r+1)!(n-r)!} w^{r+1}(1-w)^{n-r}\,dw$$

$$= \frac{r(r+1)}{(n+2)(n+1)}$$

since the integrand is like that of a p.d.f. of the $(r+2)$th order statistic of a sample of size $n+2$ and hence the integral must equal one.

(b) $\mathrm{Var}(W_r) = \dfrac{r(r+1)}{(n+2)(n+1)} - \dfrac{r^2}{(n+1)^2} = \dfrac{r(n-r+1)}{(n+2)(n+1)^2}.$

8.3–10 (a)

$$1-\alpha = P\left[\chi_{1-\alpha/2}^2(2m) \le \frac{2}{\theta}\left(\sum_{i=1}^m Y_i + (n-m)Y_m\right) \le \chi_{\alpha/2}^2(2m)\right]$$

$$= P\left[\frac{1}{\chi_{1-\alpha/2}(2m)} \ge \frac{\theta}{2(\sum_{i=1}^m Y_i + (n-m)Y_m)} \ge \frac{1}{\chi_{\alpha/2}(2m)}\right]$$

$$= P\left[\frac{2(\sum_{i=1}^m Y_i + (n-m)Y_m)}{\chi_{\alpha/2}} \le \theta \le \frac{2(\sum_{i=1}^m Y_i + (n-m)Y_m)}{\chi_{1-\alpha/2}(2m)}\right]$$

Thus the $100(1-\alpha)\%$ confidence interval is

$$\left[\frac{2(\sum_{i=1}^m y_i + (n-m)y_m)}{\chi_{\alpha/2}(2m)},\ \frac{2(\sum_{i=1}^m y_i + (n-m)y_m)}{\chi_{1-\alpha/2}(2m)}\right].$$

(b) (i) $n = 4$: $\left[\dfrac{89.840}{15.51},\ \dfrac{89.840}{2.733}\right] = [5.792, 32.872];$

    (ii) $n = 5$: $\left[\dfrac{107.036}{18.31},\ \dfrac{107.036}{3.940}\right] = [5.846, 27.1664];$

    (iii) $n = 6$: $\left[\dfrac{113.116}{21.03},\ \dfrac{113.116}{5.226}\right] = [5.379, 21.645];$

    (iv) $n = 7$: $\left[\dfrac{125.516}{23.68},\ \dfrac{125.516}{6.571}\right] = [5.301, 19.102].$

The intervals become shorter as we use more information.

8.3–14 (c) Let $\theta = 1/2$.

$$E(W_1) = E(\overline{X}) = \mu = \frac{1}{2};$$

$$\mathrm{Var}(W_1) = \mathrm{Var}(\overline{X}) = \frac{1/12}{3} = \frac{1}{36};$$

$$E(W_2) = \int_0^1 (w \cdot 6w(1-w)\,dw = \frac{1}{2};$$

$$\mathrm{Var}(W_2) = \int_0^1 (w-1/2)^2 6w(1-w)\,dw = \frac{1}{20};$$

$$E(W_3) = \int_0^1 \int_{w_1}^1 [(w_1+w_3)/2]6(w_3-w_1)\,dw_3 dw_1 = \frac{1}{2};$$

$$\mathrm{Var}(W_3) = \int_0^1 \int_{w_1}^1 [(w_1+w_3)/2]^2 6(w_3-w_1)\,dw_3 dw_1 - \left(\frac{1}{2}\right)^2 = \frac{1}{40}.$$

8.4   Distribution-Free Confidence Intervals for Percentiles

8.4–2 **(a)** $(y_3 = 5.4, y_{10} = 6.0)$ is a 96.14% confidence interval for the median, m.

(b) $(y_1 = 4.8, y_7 = 5.8)$;

$$P(Y_1 < \pi_{0.3} < Y_7) = \sum_{k=1}^{6} \binom{12}{k}(0.3)^k(0.7)^{12-k}$$
$$= 0.9614 - 0.0138 = 0.9476,$$

using Table II with $n = 12$ and $p = 0.30$.

8.4–4 **(a)** $(y_4 = 80.28, y_{11} = 80.51)$ is a 94.26% confidence interval for m.

(b) $(y_6 = 80.32, y_{12} = 80.53)$;

$$\sum_{k=6}^{11} \binom{14}{k}(0.6)^k(0.4)^{14-k} = \sum_{k=3}^{8} \binom{14}{k}(0.4)^k(0.6)^{14-k}$$
$$= 0.9417 - 0.0398 = 0.9019.$$

The interval is $(y_6 = 80.32, y_{12} = 80.53)$.

8.4–6 **(a)** We first find i and j so that $P(Y_i < \pi_{0.25} < Y_j) \approx 0.95$. Let the distribution of W be $b(81, \; 0.25)$. Then

$$P(Y_i < \pi_{0.25} < Y_j) = P(i \le W \le j - 1)$$
$$\approx P\left(\frac{i - 0.5 - 20.25}{\sqrt{15.1875}} \le Z \le \frac{j - 1 + 0.5 - 20.25}{\sqrt{15.1875}}\right).$$

If we let

$$\frac{i - 20.75}{\sqrt{15.1875}} = -1.96 \quad \text{and} \quad \frac{j - 20.75}{\sqrt{15.1875}} = 1.96$$

we find that $i \approx 13$ and $j \approx 28$. Furthermore $P(13 \le W \le 28 - 1) \approx 0.9453$. Also note that the point estimate of $\pi_{0.25}$,

$$\widetilde{\pi}_{0.25} = (y_{20} + y_{21})/2$$

falls near the center of this interval. So a 94.53% confidence interval for $\pi_{0.25}$ is $(y_{13} = 21.0, y_{28} = 21.3)$.

(b) Let the distribution of W be $b(81, 0.5)$. Then

$$P(Y_i < \pi_{0.5} < Y_{82-i}) = P(i \le W \le 81 - i)$$
$$\approx P\left(\frac{i - 0.5 - 40.5}{\sqrt{20.25}} \le Z \le \frac{81 - i + 0.5 - 40.5}{\sqrt{20.25}}\right).$$

If

$$\frac{i - 41}{4.5} = -1.96,$$

then $i = 32.18$ so let $i = 32$. Also

$$\frac{81 - i - 40}{4.5} = 1.96$$

implies that $i = 32$. Furthermore

$$P(Y_{32} < \pi_{0.5} < Y_{50}) = P(32 \le W \le 49) \approx 0.9544.$$

So an approximate 95.44% confidence interval for $\pi_{0.5}$ is $(y_{32} = 21.4, y_{50} = 21.6)$.

(c) Similar to part (a), $P(Y_{54} < \pi_{0.75} < Y_{69}) \approx 0.9453$. Thus a 94.53% confidence interval for $\pi_{0.75}$ is $(y_{54} = 21.6, y_{69} = 21.8)$.

8.4–8 A 95.86% confidence interval for m is ($y_6 = 14.60$, $y_{15} = 16.20$).

8.4–10 **(a)** A point estimate for the medium is $\widetilde{m} = (y_8 + y_9)/2 = (23.3 + 23.4)/2 = 23.35$.

(b) A 92.32% confidence interval for m is ($y_5 = 22.8$, $y_{12} = 23.7$).

8.4–12 **(a)**

Stems	Leaves	Frequency	Depths
3	80	1	1
4	74	1	2
5	20 51 73 73 92	5	7
6	01 31 32 52 57 58 71 74 84 92 95	11	18
7	08 22 36 42 46 57 70 80	8	26
8	03 11 49 51 57 71 82 92 93 93	10	(10)
9	33 40 61	3	24
10	07 09 10 30 31 40 58 75	8	21
11	16 38 41 43 51 55 66	7	13
12	10 22 78	3	6
13	34 44 50	3	3

(b) A point estimate for the median is $\widetilde{m} = (y_{30} + y_{31})/2 = (8.51 + 8.57)/2 = 8.54$.

(c) Let the distribution of W be $b(60, 0.5)$. Then
$$P(Y_i < \pi_{0.5} < Y_{61-i}) \quad = \quad P(i \leq W \leq 60 - i)$$
$$\approx \quad P\left(\frac{i - 0.5 - 30}{\sqrt{15}} \leq Z \leq \frac{60 - i + 0.5 - 30}{\sqrt{15}}\right).$$

If
$$\frac{i - 30.5}{\sqrt{15}} = -1.96$$

then $i \approx 23$. So

$$P(Y_{23} < \pi_{0.5} < Y_{38}) = P(23 \leq W \leq 37) \approx 0.9472.$$

So an approximate 94.72% confidence interval for $\pi_{0.5}$ is

$$(y_{23} = 7.46, \ y_{38} = 9.40).$$

(d) $\widetilde{\pi}_{0.40} = y_{24} + 0.4(y_{25} - y_{24}) = 7.57 + 0.4(7.70 - 7.57) = 7.622$.

(e) Let the distribution of W be $b(60, 0.40)$ then
$$P(Y_i < \pi_{0.40} < Y_j) \quad = \quad P(i \leq W \leq j - 1)$$
$$\approx \quad P\left(\frac{i - 0.5 - 24}{\sqrt{14.4}} \leq Z \leq \frac{j - 1 + 0.5 - 24}{\sqrt{14.4}}\right).$$

If we let $\dfrac{i - 24.5}{\sqrt{14.4}} = -1.645$ and $\dfrac{j - 24.5}{\sqrt{14.4}} = 1.645$ then $i \approx 18$ and $j \approx 31$. Also $P(18 \leq W \leq 31 - 1) = 0.9133$. So an approximate 91.33% confidence interval for $\pi_{0.4}$ is ($y_{18} = 6.95$, $y_{31} = 8.57$).

8.4–14 **(a)** $P(Y_7 < \pi_{0.70}) = \displaystyle\sum_{k=7}^{8} \binom{8}{k}(0.7)^k(0.3)^{8-k} = 0.2553$;

(b) $P(Y_5 < \pi_{0.70} < Y_8) = \displaystyle\sum_{k=5}^{7} \binom{8}{k}(0.7)^k(0.3)^{8-k} = 0.7483$.

8.5   The Wilcoxon Tests

8.5–2 In the following display, those observations that were negative are underlined.

$\lvert x\rvert$:	$\underline{1}$	$\underline{2}$	$\underline{2}$	2	2	$\underline{3}$	$\underline{4}$	4	$\underline{4}$	$\underline{5}$	$\underline{6}$	6
Ranks :	1	3.5	3.5	3.5	3.5	6	8	8	8	10	12	12

$\lvert x\rvert$:	6	7	7	8	11	12	13	$\underline{14}$	14	17	18	21
Ranks :	12	14.5	14.5	16	17	18	19	20.5	20.5	22	23	24

The value of the Wilcoxon statistic is

$$
\begin{aligned}
w \;=\;& -1 - 3.5 - 3.5 - 3.5 + 3.5 - 6 - 8 - 8 - 8 - 10 - 12 + 12 + 12 + \\
& 14.5 + 14.5 + 16 + 17 + 18 + 19 - 20.5 + 20.5 + 22 + 23 + 24 \\
=\;& 132.
\end{aligned}
$$

For a one-sided alternative, the approximate p-value is, using the one-unit correction,

$$
\begin{aligned}
P(W \geq 132) \;=\;& P\left(\frac{W - 0}{\sqrt{24(25)(49)/6}} \geq \frac{131 - 0}{70} \right) \\
\approx\;& P(Z \geq 1.871) = 0.03064.
\end{aligned}
$$

For a two-sided alternative, p-value $= 2(0.03064) = 0.0613$.

8.5–4 In the following display, those observations that were negative are underlined.

$\lvert x\rvert$:	$\underline{0.0790}$	0.5901	$\underline{0.7757}$	$\underline{1.0962}$	$\underline{1.9415}$
Ranks :	1	2	3	4	5

$\lvert x\rvert$:	$\underline{3.0678}$	3.8545	$\underline{5.9848}$	9.3820	$\underline{74.0216}$
Ranks :	6	7	8	9	10

The value of the Wilcoxon statistic is

$$
w = -1 + 2 - 3 - 4 - 5 - 6 + 7 - 8 + 9 - 10 = -19.
$$

Since

$$
\lvert z\rvert = \left\lvert \frac{-19}{\sqrt{10(11)(21)/6}} \right\rvert = 0.968 < 1.96,
$$

we do not reject H_0.

8.5–6 (a) The critical region is given by

$$
w \geq 1.645\sqrt{15(16)(31)/6} = 57.9.
$$

(b) In the following display, those differences that were negative are underlined.

$\lvert x_i - 50\rvert$:	2	$\underline{2}$	2.5	3	4	$\underline{4}$	$\underline{4.5}$	$\underline{6}$	7
Ranks :	1.5	1.5	3	4	5.5	5.5	7	8	9

$\lvert x_i - 50\rvert$:	7.5	8	8	$\underline{14.5}$	15.5	21
Ranks :	10	11.5	11.5	13	14	15

The value of the Wilcoxon statistic is

$$
\begin{aligned}
w &= 1.5 - 1.5 + 3 + 4 + 5.5 - 5.5 - 7 - 8 + 9 + 10 + 11.5 + 11.5 - 13 + 14 + 15 \\
&= 50.
\end{aligned}
$$

Since

$$
z = \frac{50}{\sqrt{15(16)(31)/6}} = 1.420 < 1.645,
$$

or since $w = 50 < 57.9$, we do not reject H_0.

(c) The approximate p-value is, using the one-unit correction,

$$
\begin{aligned}
p\text{-value} &= P(W \geq 50) \\
&\approx P\left(Z \geq \frac{49}{\sqrt{15(16)(31)/6}} \right) = P(Z \geq 1.3915) = 0.0820.
\end{aligned}
$$

8.5–8 The 24 ordered observations, with the x-values underlined and the ranks given under each observation are:

	0.7794	0.7546	0.7565	0.7613	0.7615	0.7701
Ranks :	1	2	3	4	5	6

	0.7712	0.7719	0.7719	0.7720	0.7720	0.7731
Ranks :	7	8.5	8.5	10.5	10.5	12

	0.7741	0.7750	0.7750	0.7776	0.7795	0.7811
Ranks :	13	14.5	14.5	16	17	18

|	0.7815	0.7816	0.7851	0.7870	0.7876	0.7972
Ranks :	19	20	21	22	23	24

(a) The value of the Wilcoxon statistic is

$$
\begin{aligned}
w &= 4 + 10.5 + 12 + 14.5 + 17 + 18 + 19 + 20 + 21 + 22 + 23 + 24 \\
&= 205.
\end{aligned}
$$

Thus

$$
p\text{-value} = P(W \geq 205) \approx P\left(Z \geq \frac{204.5 - 150}{\sqrt{12(12)(25)/12}} \right) = P(Z \geq 3.15) < 0.001
$$

so that we clearly reject H_0.

(b)
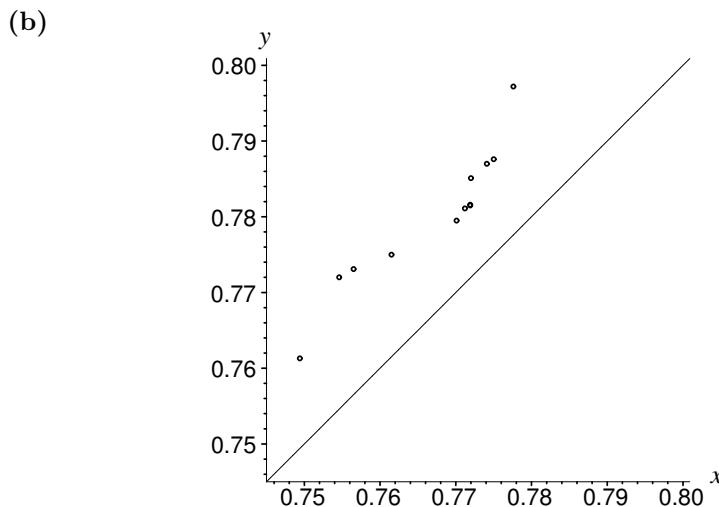
Figure 8.5–8: q-q plot of pill weights, (good, defective) $= (x, y)$

8.5–10 The ordered combined sample with the x observations underlined are:

	67.4	69.3	72.7	73.1	75.9	77.2	77.6	78.9
Ranks:	1	2	3	4	5	6	7	8

	82.5	83.2	83.3	84.0	84.7	86.5	87.5
Ranks:	9	10	11	12	13	14	15

	87.6	88.3	88.6	90.2	90.4	90.4	92.7	94.4	95.0
Ranks:	16	17	18	19	20.5	20.5	22	23	24

The value of the Wilcoxon statistic is

$$w = 4 + 8 + 9 + \cdots + 23 + 24 = 187.5.$$

Since

$$z = \frac{187.5 - 12(25)/2}{\sqrt{12(12)(25)/12}} = 2.165 > 1.645,$$

we reject H_0.

8.5–12 The ordered combined sample with the 48-passenger bus values underlined are:

	104	184	196	197	248	253	260	279
Ranks:	1	2	3	4	5	6	7	8

	300	308	323	331	355	386	393	396
Ranks:	9	10	11	12	13	14	15	16

	414	432	450	452
Ranks:	17	18	19	20

The value of the Wilcoxon statistic is

$$w = 2 + 3 + 4 + 5 + 7 + 8 + 13 + 14 + 15 + 18 + 19 = 108.$$

Since
$$z = \frac{108 - 11(21)/2}{\sqrt{9(11)(21)/12}} = -0.570 > -1.645,$$

we do not reject H_0.

8.5–14 **(a)** Here is the two-sided stem-and-leaf display.

Group A leaves	Stems	Group B leaves
	0	9
7	1	2
	2	1 5 7
3	3	1 2 3 4
6 2	4	4
7 5 1 0	5	3
3 1	6	
1	7	

(b) Here is the ordered combined sample with the Group B values underlined:

	9	12	17	21	25	27	31	32
Ranks :	1	2	3	4	5	6	7	8

	33	33	34	42	44	46	50	51
Ranks :	9.5	9.5	11	12	13	14	15	16

	53	55	57	61	63	71
Ranks :	17	18	19	20	21	22

The value of the Wilcoxon statistic is

$$w = 1 + 2 + 4 + 5 + 6 + 7 + 8 + 9.5 + 11 + 13 + 17 = 83.5.$$

Since
$$z = \frac{83.5 - 126.5}{\sqrt{11(11)(23)/12}} = \frac{-43}{15.2288} = -2.83 < -2.576 = z_{0.005},$$

we reject H_0.

(c) The results of the t-test and the Wilcoxon test are similar.

8.5–16 **(a)** Here is the two-sided stem-and-leaf display.

Young Subjects	Stems	Older Subjects
9	3●	
3	4∗	
9 8 8 6 5	4●	6
0	5∗	3 4
8 8 7 7 7 6 6 6	5●	7 8 9 9
	6∗	2 2
9	6●	5 7
	7∗	2
	7●	
	8∗	1 3
	8●	6 8
	9∗	3

(b) The value of the Wilcoxon statistic, the sum of the ranks for the younger subjects, is $w = 198$. Since

$$z = \frac{198 - 297.5}{29.033} = -3.427,$$

we clearly reject H_0.

(c) The t-test leads to the same conclusion.

8.5–18 **(a)** Using the Wilcoxon statistic, the sum of the ranks for the normal air is 102. Since

$$z = \frac{102 - 126}{\sqrt{168}} = -1.85,$$

we reject the null hypothesis. The p-value is approximately 0.03.

(b) Using a t-statistic, we failed to reject the null hypothesis at an $\alpha = 0.05$ significance level.

(c) For these data, the results are a little different with the Wilcoxon statistic leading to rejection of the null hypothesis while the t-test did not reject H_0.

8.6   Run Test and Test for Randomness

8.6–2 The combined ordered sample is:

13.00	15.50	16.75	17.25	17.50	19.00
y	x	x	x	y	y

19.25	19.75	20.50	20.75	21.50
x	y	x	x	y

22.00	22.50	22.75	23.50	24.75
x	x	y	y	y

For these data, $r = 9$. Also,

$$E(R) = \frac{2(8)(8)}{8 + 8} + 1 = 9$$

so we clearly accept H_0.

8.6–4 x | x | x | x x $x,$ x | x | x x | x $x,$

x | x | x x x | $x,$ x | x x | x | x $x,$

x | x x | x x | $x,$ x | x x x | x | $x,$

x x | x | x | x $x,$ x x | x | x x | $x,$

x x | x x | x | $x,$ x x x | x | x | $x.$

8.6–6 The combined ordered sample is:

-2.0482	-1.5748	-1.2311	-1.0228	-0.8836	-0.8797	-0.7170
x	x	y	y	y	x	x

-0.6684	-0.6157	-0.5755	-0.4907	-0.2051	-0.1019	-0.0297
y	y	y	x	x	y	y

0.1651	0.2893	0.3186	0.3550	0.3781
x	x	x	x	y

0.4056	0.6975	0.7113	0.7377
x	x	x	x

0.7400	0.8479	1.0901	1.1397	1.1748	1.2921	1.7356
y	y	y	y	y	y	x

For these data, the number of runs is $r = 11$. The p-value of this test is

$$p\text{-value} = P(R \leq 11) \approx P\left(Z \leq \frac{11.5 - 16.0}{\sqrt{15(14)/29}}\right) = 0.0473.$$

Thus we would reject H_0 at an $\alpha = 0.0473 \approx 0.05$ significance level.

8.6–8 The median is 22.45. Replacing observations below the median with L and above the median with U, we have

$$L \;\; U \;\; L \;\; U \;\; L \;\; U \;\; L \;\; U \;\; U \;\; U \;\; L \;\; L \;\; L \;\; U \;\; L \;\; U$$

or $r = 12$ runs. Since

$$
\begin{aligned}
P(R \geq 12) &= (2 + 14 + 98 + 294 + 882)/12{,}870 \\
&= 1290/12{,}870 = 0.10
\end{aligned}
$$

and

$$P(R \geq 13) = 408/12{,}870 = 0.0371,$$

we would reject the hypothesis of randomness if $\alpha = 0.10$ but would not reject if $\alpha = 0.0317$.

8.6–10 For these data, the median is 21.55. Replacing lower and upper values with L and U, respectively, gives the following displays:

$$L \;\; U \;\; L \;\; L \;\; U \;\; U \;\; U \;\; U \;\; L \;\; U \;\; L \;\; U \;\; L \;\; L \;\; U \;\; L \;\; U \;\; U \;\; L \;\; U \;\; U \;\; U \;\; U$$

$$L \;\; L \;\; L \;\; U \;\; U \;\; L \;\; U \;\; L \;\; U \;\; L \;\; L \;\; L \;\; U \;\; L \;\; L$$

We see that there are $r = 23$ runs. The value of the standard normal test statistic is

$$z = \frac{23 - 20}{\sqrt{(19)(18)/37}} = 0.987.$$

Thus we would not reject the hypothesis of randomness at any reasonable significance level.

8.6–12 (a) The number of runs is $r = 38$. The p-value of the test is

$$p\text{-value} = P(R \geq 38) \quad \approx \quad P\left(Z \geq \frac{37.5 - 28.964}{\sqrt{(27.964)(26.964)/54.928}}\right)$$

$$= \quad P(Z \geq 2.30) = 0.0107,$$

so we would not reject the hypothesis of randomness in favor of a cyclic effect at $\alpha = 0.01$, but the evidence is strong that the latter might exist. This, however, is not bad.

(b) The different versions of the test were not written in such a way that allowed students to finish earlier on one than on the other.

8.6–14 The number of runs is $r = 30$. The p-value of the test is

$$p\text{-value} = P(R \geq 30) \quad \approx \quad P\left(Z \geq \frac{29.5 - 35.886}{\sqrt{(34.886)(33.886)/69.772}}\right)$$

$$= \quad P(Z \geq 1.55) = 0.9394,$$

so we would not reject the hypothesis of randomness, although there seems to be a tendency of too few runs. A display of the data shows that there is a cyclic effect with long cycles.

8.6–16 The number of runs is $r = 10$. The mean and variance for the run test are

$$\mu = \frac{2(17)(17)}{17 + 17} + 1 = 18;$$

$$\sigma^2 = \frac{(18 - 1)(18 - 2)}{17 + 17 - 1} = \frac{272}{33}.$$

The standard deviation is $\sigma = 2.87$. The p-value for this test is

$$P(R \leq 10) = P\left(\frac{R - 18}{2.87} \leq \frac{10.5 - 18}{2.87}\right) \approx P(Z \leq -2.61) = -0.0045.$$

Thus we reject H_0. The p-value is larger than that for the Wilcoxon test but still clearly leads to reject of the null hypothesis.

8.6–18 The number of runs is $r = 11$ and the mean number of runs is $\mu = 10.6$. Thus the run test would not detect any difference.

8.7   Kolmogorov-Smirnov Goodness of Fit Test

8.7–4 (a)
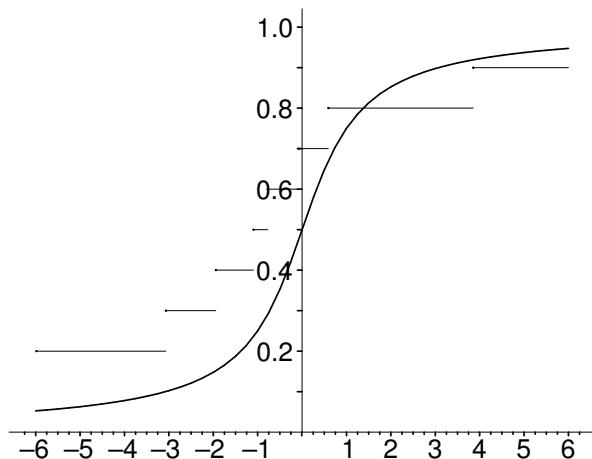
Figure 8.7–4: H_0: X has a Cauchy distribution

(b) $d_{10} = 0.3100$ at $x = -0.7757$. Since $0.31 < 0.37$, we do not reject the hypothesis that these are observations of a Cauchy random variable.

8.7–6
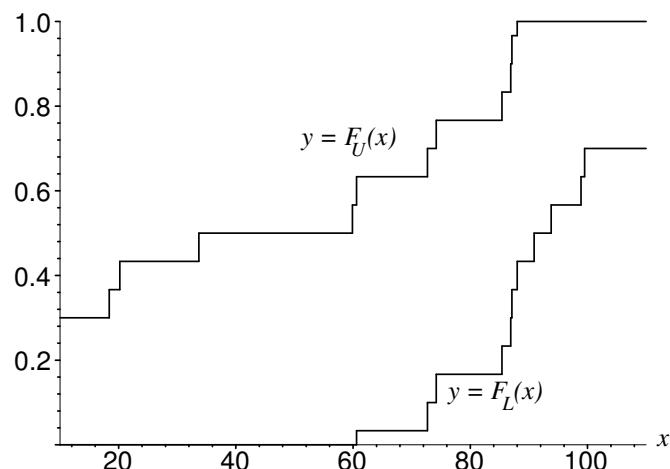
Figure 8.7–6: A 90% confidence band for $F(x)$

8.7–8 The value of the Kolmogorov-Smirnov statistic is 0.0587 which occurs at $x = 21$. We clearly accept the null hypothesis.
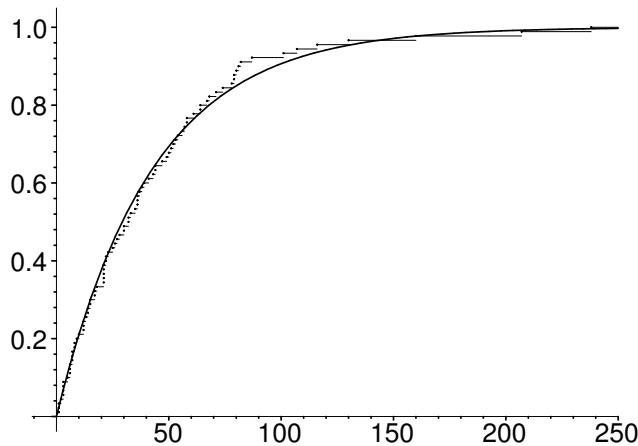
Figure 8.7–8: H_0: X has an exponential distribution

8.7–10 $d_{62} = 0.068$ at $x = 4$ so we accept the hypothesis that X has a Poisson distribution.

8.7–12
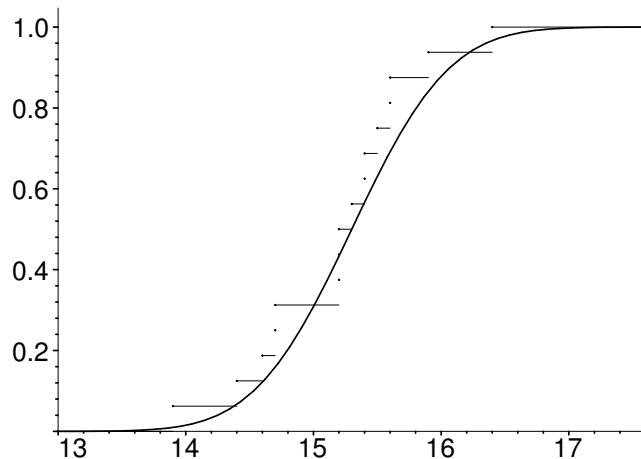
Figure 8.7–12: H_0: X is $N(15.3, 0.6^2)$

$d_{16} = 0.1835$ at $x = 15.6$ so we do not reject the hypothesis that the distribution of peanut weights is $N(15.3, 0.6^2)$.

8.8   Resampling

8.8–2 (a)
```
> read 'C:\\Tanis-Hogg\\Maple Examples\\stat.m':
  with(plots):
  read 'C:\\Tanis-Hogg\\Maple Examples\\HistogramFill.txt':
  read 'C:\\Tanis-Hogg\\Maple Examples\\ScatPlotCirc.txt':
  read 'C:\\Tanis-Hogg\\Maple Examples\\Chapter_05.txt':
  XX := Exercise_5_4_2;
```

$$XX := [12.0, 9.4, 10.0, 13.5, 9.3, 10.1, 9.6, 9.3, 9.1, 9.2, 11.0, 9.1, 10.4, 9.1, 13.3, 10.6]$$

```
> Probs := [seq(1/16, k = 1 .. 16)]:
  XXPDF := zip((XX,Probs)-> (XX,Probs), XX, Probs):
> for k from 1 to 200 do
     X := DiscreteS(XXPDF, 16):
     Svar[k] := Variance(X):
  od:
  Svars := [seq(Svar[k], k = 1 .. 200)]:
> Mean(Svars);
```
$$1.972629584$$

```
> xtics := [seq(0.4*k, k = 1 .. 12)]:
  ytics := [seq(0.05*k, k = 1 .. 11)]:
  P1 := plot([[0,0],[0,0]], x = 0 .. 4.45, y = 0 .. 0.57,
  xtickmarks=xtics, ytickmarks=ytics, labels=['',''])):
  P2 := HistogramFill(Svars,0 .. 4.4, 11):
  display({P1, P2});
```

The histogram is shown in Figure 5.4-2**(ab)**.

(b)
```
> theta := Mean(XX) - 9;
  for k from 1 to 200 do
     Y := ExponentialS(theta,21):
     Svary[k] := Variance(Y):
  od:
  Svarys := [seq(Svary[k], k = 1 .. 200)]:
```

$$\theta := 1.31250000$$

```
> Mean(Svarys);
```
$$1.747515570$$

```
> xtics := [seq(0.4*k, k = 1 .. 14)]:
  ytics := [seq(0.05*k, k = 1 .. 15)]:
  P3 := plot([[0,0],[0,0]], x = 0 .. 5.65, y = 0 .. 0.62,
  xtickmarks=xtics, ytickmarks=ytics, labels=['',''])):
  P4 := HistogramFill(Svarys,0 .. 5.6, 14):
  display({P3, P4});
```

```
> Svars := sort(Svars):
  Svarys := sort(Svarys):
> xtics := [seq(k*0.5, k = 1 .. 18)]:
  ytics := [seq(k*0.5, k = 1 .. 18)]:
  P5 := plot([[0,0],[5.5,5.5]], x = 0 .. 5.4, y = 0 .. 7.4, color=black,
```
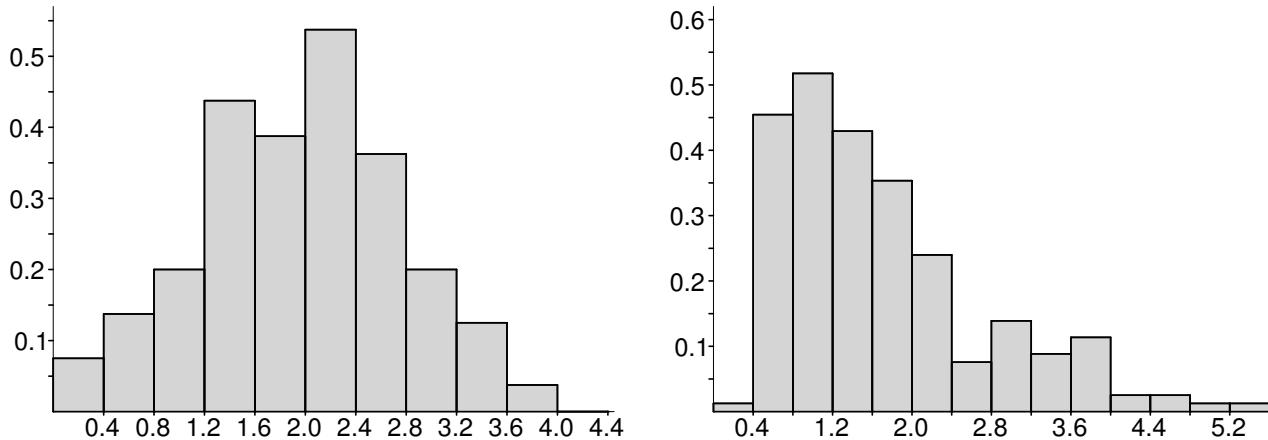
Figure 8.8–2: **(ab)** Histogram of S^2s: Resampling on Left, From Exponential on Right

```
thickness=2, xtickmarks=xtics, ytickmarks=ytics, labels=[``,``]):
P6 := ScatPlotCirc(Svars,Svarys):
display({P5, P6});
```
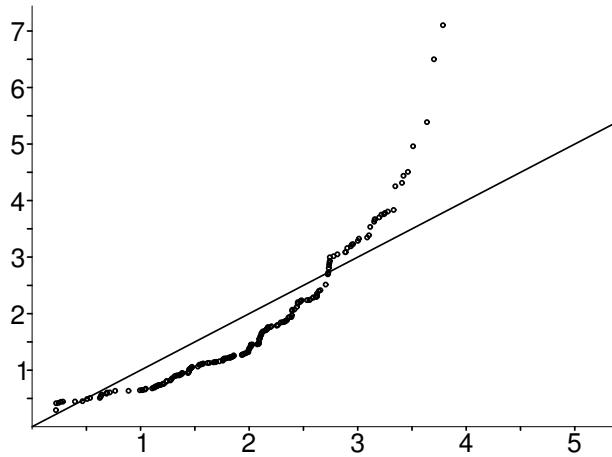

Figure 8.8–2: **(c)** *q-q* Plot of Exponential S^2s Versus Resampling S^2s

Note that the variance of the sample variances from the exponential distribution is greater than the variance of the sample variances from the resampling distribution.

8.8–4 (a)
```
> with(plots):
  read `C:\\Tanis-Hogg\\Maple Examples\\stat.m`:
  read `C:\\Tanis-Hogg\\Maple Examples\\ScatPlotPoint.txt`:
  read `C:\\Tanis-Hogg\\Maple Examples\\EmpCDF.txt`:
  read `C:\\Tanis-Hogg\\Maple Examples\\HistogramFill.txt`:
  read `C:\\Tanis-Hogg\\Maple Examples\\ScatPlotCirc.txt`:
  read `C:\\Tanis-Hogg\\Maple Examples\\Chapter_05.txt`:
  Pairs := Exercise_5_4_4;
  Pairs := [[2.500, 72], [4.467, 88], [2.333, 62], [5.000, 87],
  [1.683, 57], [4.500, 94], [4.500, 91], [2.083, 51], [4.367, 98],
  [1.583, 59], [4.500, 93], [4.550, 86], [1.733, 70], [2.150, 63],
```

```
      [4.400, 91], [3.983, 82], [1.767, 58], [4.317, 97], [1.917, 59],
      [4.583, 90], [1.833, 58], [4.767, 98], [1.917, 55], [4.433, 107],
      [1.750, 61], [4.583, 82], [3.767, 91], [1.833, 65], [4.817, 97],
      [1.900, 52], [4.517, 94], [2.000, 60], [4.650, 84], [1.817, 63],
      [4.917, 91], [4.000, 83], [4.317, 84], [2.133, 71], [4.783, 83],
      [4.217, 70], [4.733, 81], [2.000, 60], [4.717, 91], [1.917, 51],
      [4.233, 85], [1.567, 55], [4.567, 98], [2.133, 49], [4.500, 85],
      [1.717, 65], [4.783, 102], [1.850, 56], [4.583, 86],[1.733, 62]]:
> r := Correlation(Pairs);
```

$$r := .9087434803$$

```
> xtics := [seq(1.4 + 0.1*k, k = 0 .. 37)]:
  ytics := [seq(48 + 2*k, k = 0 .. 31)]:
  P1 := plot([[1.35,47],[1.35,47]], x = 1.35 .. 5.15, y = 47 .. 109,
  xtickmarks = xtics, ytickmarks=ytics, labels=[''','']):
  P2 := ScatPlotCirc(Pairs):
  display({P1, P2});
```
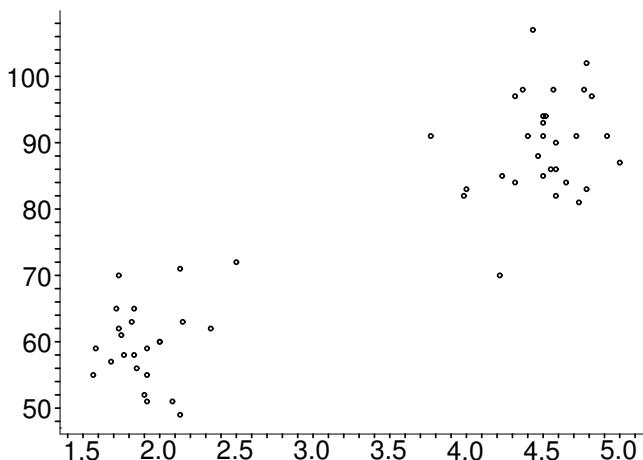

Figure 8.8–4: **(a)** Scatterplot of the 50 Pairs of Old Faithful Data

```
(b) > Probs := [seq(1/54, k = 1 .. 54)]:
      EmpDist := zip((Pairs,Probs)-> (Pairs,Probs), Pairs, Probs):
    > for k from 1 to 500 do
         Samp := DiscreteS(EmpDist, 54);
         RR[k] := Correlation(Samp):
      od:
      R := [seq(RR[k], k = 1 .. 500)]:
      rbar := Mean(R);
```

$$rbar := .9079354926$$

```
(c) > xtics := [seq(0.8 + 0.01*k, k = 0 .. 20)]:
      ytics := [seq(k, k = 1 .. 25)]:
    > P3 := plot([[0.79, 0],[0.79,0]], x = 0.79 .. 1.005,
      y = 0 .. 23.5, xtickmarks=xtics, ytickmarks=ytics, labels=[''','']):
      P4 := HistogramFill(R, 0.8 .. 1, 20):
      display({P3, P4});
```

The histogram is plotted in Figure 5.4-4 **ce**.

(d) Now simulate a random sample of 500 correlation coefficients, each calculated from a sample of size 54 from a bivariate normal distribution with correlation coefficient $r = 0.9087434803$.

```
> for k from 1 to 500 do
    Samp := BivariateNormalS(0,1,0,1,r,54):
    RR[k] := Correlation(Samp):
  od:
  RBivNorm := [seq(RR[k], k = 1 .. 500)]:
  AverageR := Mean(RBivNorm);
```

$$AverageR := .9073168034$$

```
> P5 := plot([[0.79, 0],[0.79,0]], x = 0.79 .. 1.005,
  y = 0 .. 18.5, xtickmarks=xtics, ytickmarks=ytics, labels=['',''']):
  P6 := HistogramFill(RBivNorm, 0.8 .. 1, 20):
  display({P5, P6});
```

(e)
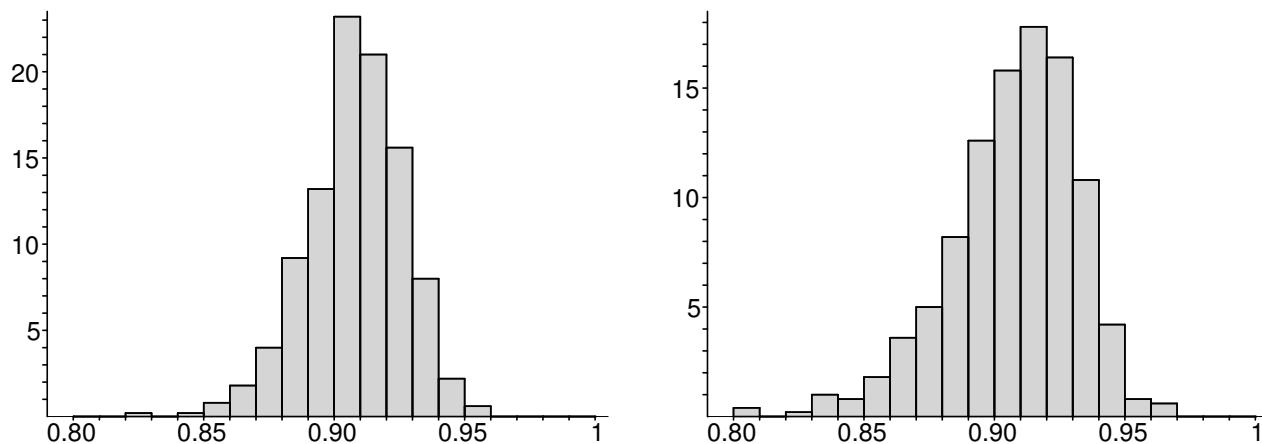
Figure 8.8–4: **(ce)** Histograms of Rs: From Resampling on Left, From Bivariate Normal on Right

(f)
```
> R := sort(R):
  RBivNorm := sort(RBivNorm):
  xtics := [seq(0.8 + 0.01*k, k = 0 .. 20)]:
  ytics := [seq(0.8 + 0.01*k, k = 0 .. 20)]:
  P7 := plot([[0.8, 0.8],[1,1]], x = 0.8 .. 0.97, y = 0.8 .. 0.97,
  color=black, thickness=2, labels=['',''], xtickmarks=xtics,
  ytickmarks=ytics):
  P8 := ScatPlotCirc(R, RBivNorm):
  display({P7, P8});
```
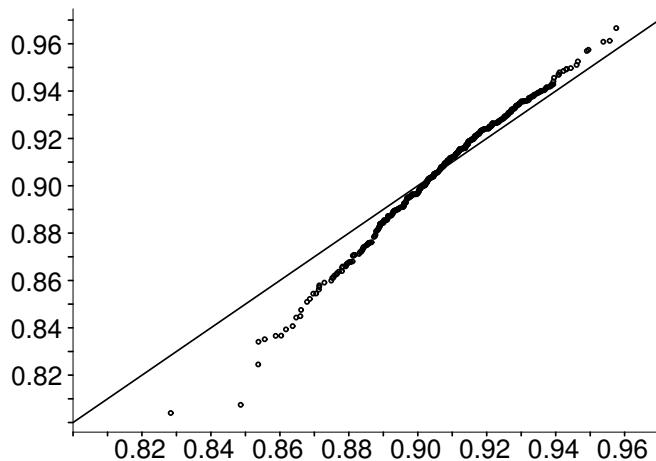

Figure 8.8–4: *q-q* Plot of the Values of R from Bivariate Normal Versus from Resampling

```
> StDev(R);
  StDev(RBivNorm);
```

$$.01852854051$$

$$.02461716901$$

The means are about equal but the standard deviation of the values of R from the bivariate normal distribution is larger than that of the resampling distribution.

8.8–6 (a)
```
> with(plots):
> read 'C:\\Tanis-Hogg\\Maple Examples\\stat.m';
  read 'C:\\Tanis-Hogg\\Maple Examples\\ScatPlotPoint.txt':
  read 'C:\\Tanis-Hogg\\Maple Examples\\EmpCDF.txt':
  read 'C:\\Tanis-Hogg\\Maple Examples\\HistogramFill.txt':
  read 'C:\\Tanis-Hogg\\Maple Examples\\ScatPlotCirc.txt':
  read 'C:\\Tanis-Hogg\\Maple Examples\\Chapter_05.txt':
  Pairs := Exercise_5_4_6;
  Pairs := [[5.4341, 8.4902], [33.2097, 4.7063], [0.4034, 1.8961],
           [1.4137, 0.2996], [17.9365, 3.1350], [4.4867, 6.2089],
           [11.5107, 10.9784], [8.2473, 19.6554], [1.9995, 3.6339],
           [1.8965, 1.7850], [1.7116, 1.1545], [4.4594, 1.2344],
           [0.4036, 0.7260], [3.0578, 19.0489], [21.4049, 4.6495],
           [3.8845, 13.7945], [5.9536, 9.2438], [11.3942, 1.7863],
           [5.4813, 4.3356], [7.0590, 1.15834]]
> r := Correlation(Pairs);
```

$$r := .02267020144$$

```
> xtics := [seq(k, k = 0 .. 35)]:
 ytics := [seq(k, k = 0 .. 35)]:
> P1 := plot([[0,0],[0,0]], x = 0 .. 35.5, y = 0 .. 35.5,
  xtickmarks = xtics, ytickmarks=ytics, labels=['',''']):
  P2 := ScatPlotCirc(Pairs):
  display({P1, P2});
```
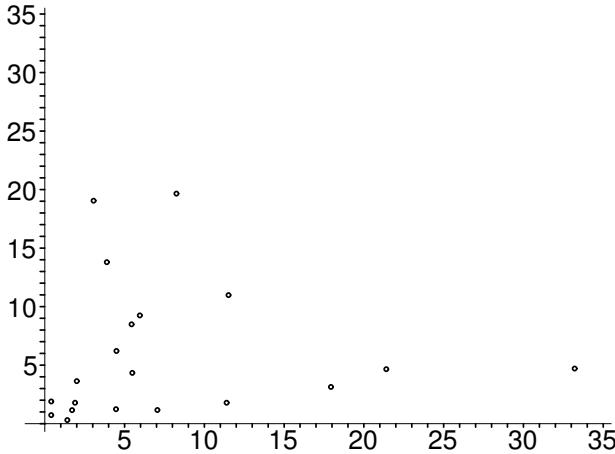

Figure 8.8–6: **(a)** Scatterplot of Paired Data from Two Independent Exponential Distributions

(b)
```
> Probs := [seq(1/20, k = 1 .. 20)]:
  EmpDist := zip((Pairs,Probs)-> (Pairs,Probs), Pairs, Probs):
> for k from 1 to 500 do
    Samp := DiscreteS(EmpDist, 20);
    RR[k] := Correlation(Samp):
  od:
  R := [seq(RR[k], k = 1 .. 500)]:
  rbar := Mean(R);
```

$$rbar := .04691961690$$

```
> Min(R),Max(R);
```

$$-.4224435806, .6607518008$$

```
> xtics := [seq(-0.5 + 0.1*k, k = 0 .. 12)]:
  ytics := [seq(k/2, k = 1 .. 6)]:
> P3 := plot([[0, 0],[0,0]], x = -0.5 .. 0.7, y = 0 .. 2.8,
  xtickmarks=xtics, ytickmarks=ytics, labels=['','']):
  P4 := HistogramFill(R, -0.5 .. 0.7, 12):
  display({P3, P4});
```

The histogram is given in Figure 5.4-6 **(ce)**.

(c) How do these observations compare with a random sample of 500 correlation coefficients, each calculated from a sample of size 20 from a bivariate normal distribution with correlation coefficient $r = 0.02267020145$?

```
> for k from 1 to 500 do
    Samp := BivariateNormalS(0,1,0,1,r,20):
    RR[k] := Correlation(Samp):
  od:
```

```
RBivNorm := [seq(RR[k], k = 1 .. 500)]:
AverageR := Mean(RBivNorm);
Min(RBivNorm),Max(RBivNorm);
```

$$AverageR := .02508989176$$

$$-.6012477460, .6131980318$$

```
> xtics := [seq(-0.7 + 0.1*k, k = 0 .. 14)]:
  ytics := [seq(k/2, k = 1 .. 6)]:
> P5 := plot([[0, 0],[0,0]], x = -0.7 .. 0.7, y = 0 .. 1.8,
  xtickmarks=xtics, ytickmarks=ytics, labels=['',''']):
  P6 := HistogramFill(RBivNorm, -0.7 .. 0.7, 14):
  display({P5, P6});
```

(d)
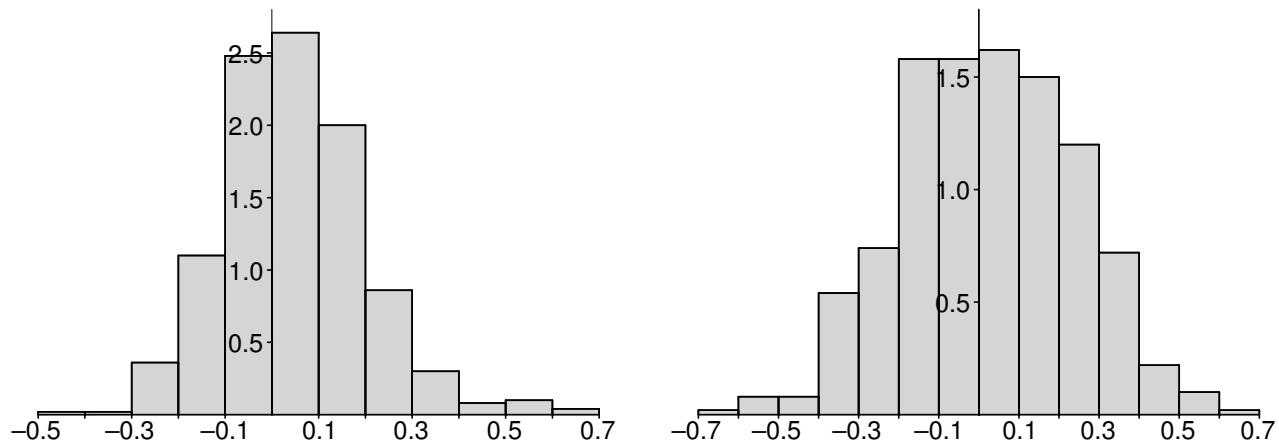
Figure 8.8–6: **(ce)** Histograms of Rs: From Resampling on Left, From Bivariate Normal on Right

(e)
```
> R := sort(R):
  RBivNorm := sort(RBivNorm):
> xtics := [seq(-0.7 + 0.1*k, k = 0 .. 14)]:
  ytics := [seq(-0.7 + 0.1*k, k = 0 .. 14)]:
  P7 := plot([[-0.7, -0.7],[0.7,0.7]], x = -0.7 .. 0.7, y = -0.7 .. 0.7,
  color=black, thickness=2, labels=['',''], xtickmarks=xtics, ytickmarks=ytics):
  P8 := ScatPlotCirc(R, RBivNorm):
  display({P7, P8});


> Mean(R);
  Mean(RBivNorm);
```

$$.04691961692$$

$$.02508989164$$

```
> StDev(R);
  StDev(RBivNorm);
```

$$.1527482117$$

$$.2200346799$$
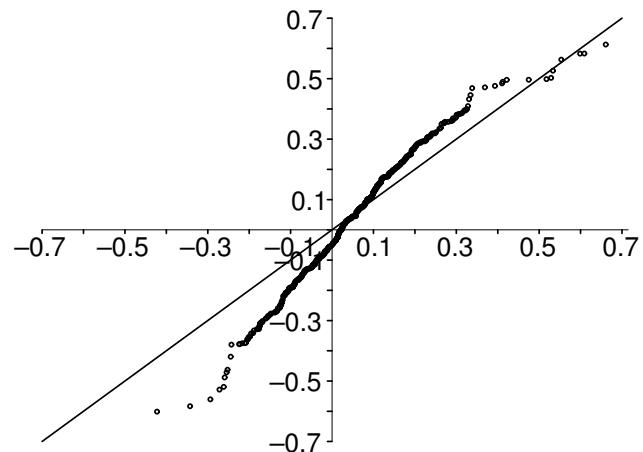
Figure 8.8–6: **(f)** *q-q* Plot of the Values of R from Bivariate Normal Versus from Resampling

The sample mean of the observations of R from the bivariate normal distribution is less than that from the resampling distribution and the standard deviation of the values of R from the bivariate normal distribution is greater than that of the resampling distribution.

Chapter 9

Bayesian Methods

9.1 Subjective Probability

9.1–2 No answer needed.

9.1–4 One solution is 1 to 7 for a bet on A and 5 to 1 for a bet on B.

A bets: for a 7 dollar bet, the bookie gives one back: $30000/7 \times 1 = 4285.71$. So the bookie gives out $4285.71 + 30000 = 34285.71$.

B bets: for a 1 dollar bet, the bookie gives five back: $5000/1 \times 5 = 25000$. So the bookie gives out $25000 + 5000 = 30000$.

9.1–6 Following HINT: before anything, the person has

$$p_1 + \frac{d}{4} + p_2 + \frac{d}{4} - \left(p_3 - \frac{d}{4}\right) = p_1 + p_2 + \frac{3d}{4} - p_3 = -d + \frac{3d}{4} = -\frac{d}{4};$$

that is, the person is down $d/4$ before the start.

1. If A_1 occurs, both win and they exchange units.
2. If A_2 happens, again they exchange units.
3. If neither A_1 nor A_2 occurs, both receive zero; and the person is still down $d/4$ in all three cases.

Thus it is bad for that person to believe that $p_3 > p_1 + p_2$ for it can lead to a Dutch book.

9.1–8 $P(A \cup A') = P(A) + P(A')$ from Theorem 7.1–1. From Exercise 7.1–7, $P(S) = 1$ so that $1 = P(A) + P(A')$. Thus $P(A') = 1 - P(A)$.

9.2   Bayesian Estimation

9.2–2 (a) $g(\tau \mid x_1, x_2, \ldots, x_n) \propto \dfrac{(x_1 x_2 \cdots x_n)^{\alpha - 1}}{[\Gamma(\alpha)]^n} \dfrac{\tau^{n\alpha} \tau^{\alpha_0 - 1} e^{-\tau/\theta_0} e^{-\Sigma x_i/(1/\tau)}}{\Gamma(\alpha_0)\theta_0^{\alpha_0}}$

$ \propto \tau^{\alpha n + \alpha_0 - 1} e^{-(1/\theta_0 + \Sigma x_i)\tau}$

which is $\Gamma\left(n\alpha + \alpha_0, \dfrac{\theta_0}{1 + \theta_0 \Sigma x_i}\right)$.

(b) $E(\tau \mid x_1, x_2, \ldots, x_n) = (n\alpha + \alpha_0)\dfrac{\theta_0}{1 + \theta_0 \overline{X} n}$

$ = \dfrac{\alpha_0 \theta_0}{1 + \theta_0 n \overline{X}} + \dfrac{\alpha n \theta_0}{1 + n\theta_0 \overline{X}}$

$ = \dfrac{n\alpha + \alpha_0}{1/\theta_0 + n\overline{X}}.$

(c) The posterior distribution is $\Gamma(30 + 10, 1/[1/2 + 10\,\overline{x}])$. Select a and b so that $P(a < \tau < b) = 0.95$ with equal tail probabilities. Then

$$\int_a^b \frac{(1/2 + 10\,\overline{x})^{40}}{\Gamma(40)} w^{40-1} e^{-w(1/2 + 10\,\overline{x})} dw = \int_{a(1/2+10\overline{x})}^{b(1/2+10\overline{x})} \frac{1}{\Gamma(40)} z^{39} e^{-z} dz,$$

making the change of variables $w(1/2 + 10\,\overline{x}) = z$. Let $v_{0.025}$ and $v_{0.975}$ be the quantiles for the $\Gamma(40, 1)$ distribution. Then

$$a = \frac{v_{0.025}}{1/2 + 10\,\overline{x}};$$

$$b = \frac{v_{0.975}}{1/2 + 10\,\overline{x}}.$$

It follows that

$$P(a < \tau < b) = 0.95.$$

9.2–4

$$(3\theta)^n (x_1 x_2 \cdots x_n)^2 e^{-\theta \Sigma x_i^3} \cdot \theta^{4-1} e^{-4\theta} \propto \theta^{n+3} e^{-(4 + \Sigma x_i^3)\theta}$$

which is $\Gamma\left(n + 4, \dfrac{1}{4 + \Sigma x_i^3}\right)$. Thus

$$E(\theta \mid x_1, x_2, \ldots, x_n) = \frac{n+4}{4 + \Sigma x_i^3}.$$

9.3 More Bayesian Concepts

7.3–2 $k(x,\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x} \cdot \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}, \qquad x = 0,1,\ldots,n, \;\; 0 < \theta < 1.$

$$
\begin{aligned}
k_1(x) &= \int_0^1 \binom{n}{x}\theta^x(1-\theta)^{n-x} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}\,d\theta \\[2mm]
&= \frac{n!\,\Gamma(\alpha+\beta)\,\Gamma(x+\alpha)\,\Gamma(n-x+\beta)}{x!\,(n-x)!\,\Gamma(\alpha)\,\Gamma(\beta)\,\Gamma(n+\alpha+\beta)}, \qquad x = 0,1,2,\ldots,n.
\end{aligned}
$$

7.3–4
$$
\begin{aligned}
k(x,\theta) &= \int_0^\infty \theta\tau x^{\tau-1}e^{-\theta x^\tau} \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha}\,\theta^{\alpha-1}e^{-\theta/\beta}\,d\theta, \qquad 0 < x < \infty \\[2mm]
&= \frac{\tau x^{\tau-1}}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty \theta^{\alpha+1-1}e^{-(x^\tau+1/\beta)\theta}\,d\theta \\[2mm]
&= \frac{\tau x^{\tau-1}}{\Gamma(\alpha)\beta^\alpha}\,\frac{\Gamma(\alpha+1)}{(x^\tau+1/\beta)^{\alpha+1}}, \qquad 0 < x < \infty \\[2mm]
&= \frac{\alpha\beta\tau x^{\tau-1}}{(\beta x^\tau+1)^{\alpha+1}}, \qquad 0 < x < \infty.
\end{aligned}
$$

7.3–6
$$
g(\theta_1,\theta_2 \mid x_1 = 3, x_2 = 7) \propto \left(\frac{1}{\pi}\right)^2 \frac{\theta_2^2}{[\theta_2^2+(3-\theta_1)^2][\theta_2^2+(7-\theta_1)^2]}.
$$

The figures show the graph of

$$
h(\theta_1,\theta_2) = \frac{\theta_2^2}{[\theta_2^2+(3-\theta_1)^2][\theta_2^2+(7-\theta_1)^2]}.
$$

The second figure shows a contour plot.
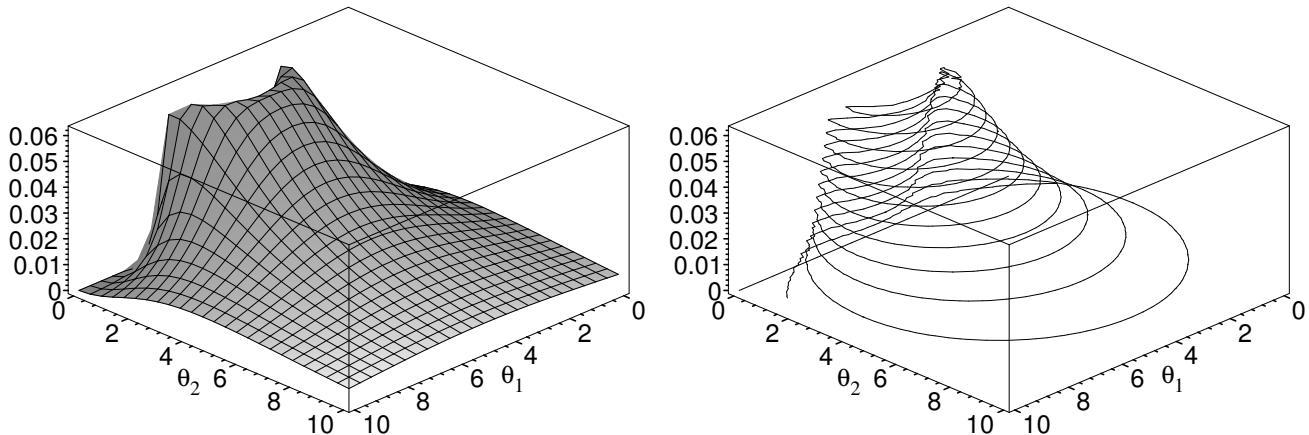
Figure 9.3–6: Graphs to help to see where θ_1 and θ_2 maximize the posterior p.d.f.

Using *Maple*, a solution is $\theta_1 = 5$ and $\theta_2 = 2$. Other solutions satisfy

$$
\theta_2 = \sqrt{-\theta_1^2 + 10\theta_1 - 21}.
$$

Chapter 10

Some Theory

10.1 Sufficient Statistics

10.1–2 The distribution of Y is Poisson with mean $n\lambda$. Thus, since $y = \Sigma\, x_i$,

$$P(X_1 = x_1,\ X_2 = x_2,\ \cdots,\ X_n = x_n \mid Y = y) \;=\; \frac{(\lambda^{\Sigma x_i} e^{-n\lambda})/(x_1!\, x_2! \cdots x_n!)}{(n\lambda)^y\, e^{-n\lambda}/y!}$$

$$= \frac{y!}{x_1!\, x_2! \cdots x_n!\, n^y},$$

which does not depend on λ.

10.1–4 **(a)** $f(x;\theta) = e^{(\theta-1)\ln x + \ln \theta}$, $\qquad 0 < x < 1$, $\qquad 0 < \theta < \infty$;

so $K(x) = \ln x$ and thus

$$Y = \sum_{i=1}^{n} \ln X_i = \ln(X_1 X_2 \cdots X_n)$$

is a sufficient statistic for θ.

(b)
$$L(\theta) \;=\; \theta^n (x_1 x_2 \cdots x_n)^{\theta-1}$$

$$\ln L(\theta) \;=\; n \ln \theta + (\theta - 1)\ln(x_1 x_2 \cdots x_n)$$

$$\frac{d \ln L(\theta)}{d\theta} \;=\; \frac{n}{\theta} + \ln(x_1 x_2 \cdots x_n) = 0.$$

Hence

$$\widehat{\theta} = -n/\ln(X_1 X_2 \cdots X_n),$$

which is a function of Y.

(c) Since $\widehat{\theta}$ is a single valued function of Y with a single valued inverse, knowing the value of $\widehat{\theta}$ is equivalent to knowing the value of Y, and hence it is sufficient.

10.1–6 **(a)**
$$f(x_1, x_2, \ldots, x_n) \;=\; \frac{(x_1 x_2 \cdots x_n)^{\alpha-1} e^{-\Sigma x_i/\theta}}{[\Gamma(\alpha)]^n \theta^{\alpha n}}$$

$$= \left(\frac{e^{-\Sigma x_i/\theta}}{\theta^{\alpha n}} \right)\left(\frac{(x_1 x_2 \cdots x_n)^{\alpha-1}}{[\Gamma(\alpha)]^n} \right).$$

The second factor is free of θ. The first factor is a function of the x_is through $\sum_{i=1}^{n} x_i$ only, so $\sum_{i=1}^{n} x_i$ is a sufficient statistic for θ.

(b) $\ln L(\theta) = \ln(x_1 x_2 \cdots x_n)^{\alpha-1} - \sum_{i=1}^{n} x_i/\theta - \ln[\Gamma(\alpha)]^n - \alpha n \ln \theta$

$$\frac{d \ln L(\theta)}{d\theta} = \sum_{i=1}^{n} x_i/\theta^2 - \alpha\,n/\theta = 0$$

$$\alpha\,n\theta = \sum_{i=1}^{n} x_i$$

$$\widehat{\theta} = \frac{1}{\alpha\,n} \sum_{i=1}^{n} X_i.$$

$Y = \sum_{i=1}^{n} X_i$ has a gamma distribution with parameters $\alpha\,n$ and θ. Hence

$$E(\widehat{\theta}) = \frac{1}{\alpha\,n}(\alpha\,n\,\theta) = \theta.$$

10.1–8
$$E(e^{tZ}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi\theta}}\right)^n e^{-\Sigma x_i^2/(2\theta)} \cdot e^{t\Sigma a_i x_i/\Sigma x_i} dx_1 dx_2 \ldots dx_n.$$

Let $x_i/\sqrt{\theta} = y_i$, $i = 1, 2, \ldots, n$. The Jacobian is $(\sqrt{\theta}\,)^n$. Hence

$$E(e^{tZ}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\sqrt{\theta}\,)^n \left(\frac{1}{\sqrt{2\pi\theta}}\right)^n e^{-\Sigma y_i^2/2} \cdot e^{t\Sigma a_i y_i/\Sigma y_i} dy_1 dy_2 \ldots dy_n$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\Sigma y_i^2/2} \cdot e^{t\Sigma a_i y_i/\Sigma y_i} dy_1 dy_2 \ldots dy_n$$

which is free of θ. Since the distribution of Z is free of θ, Z and $Y = \sum_{i=1}^{n} X_i^2$, the sufficient statistics, are independent.

10.2   Power of a Statistical Test

10.2–2 (a) $K(\mu) = P(\overline{X} \leq 354.05;\ \mu)$

$$= P\left(Z \leq \frac{354.05 - \mu}{2/\sqrt{12}};\ \mu\right)$$

$$= \Phi\left(\frac{354.05 - \mu}{2/\sqrt{12}}\right);$$

(b) $\alpha = K(355) = \Phi\left(\dfrac{354.05 - 355}{2/\sqrt{12}}\right) = \Phi(-1.645) = 0.05;$

(c) $K(354.05) = \Phi(0) = 0.5;$
$K(353.1) = \Phi(1.645) = 0.95.$

(d) $\overline{x} = 353.83 < 354.05$, reject H_0;

(e) p-value $= P(\overline{X} \leq 353.83;\ \mu = 355)$
$= P(Z \leq -2.03) = 0.0212.$

10.2–4 (a) $K(\mu) = P(\overline{X} \geq 83;\ \mu)$

$$= P\left(Z \geq \frac{83 - \mu}{10/5}\right) = 1 - \Phi\left(\frac{83 - \mu}{2}\right);$$
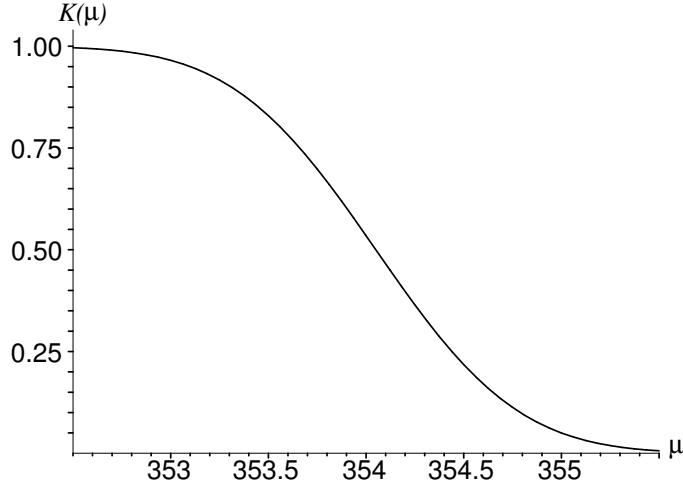
Figure 10.2–2: $K(\mu) = \Phi([354.05 - \mu]/[2/\sqrt{12}])$

(b) $\alpha = K(80) = 1 - \Phi(1.5) = 0.0668;$

(c) $K(80) = \alpha = 0.0668,$

$K(83) = 1 - \Phi(0) = 0.5000,$

$K(86) = 1 - \Phi(-1.5) = 0.9332;$

(d)
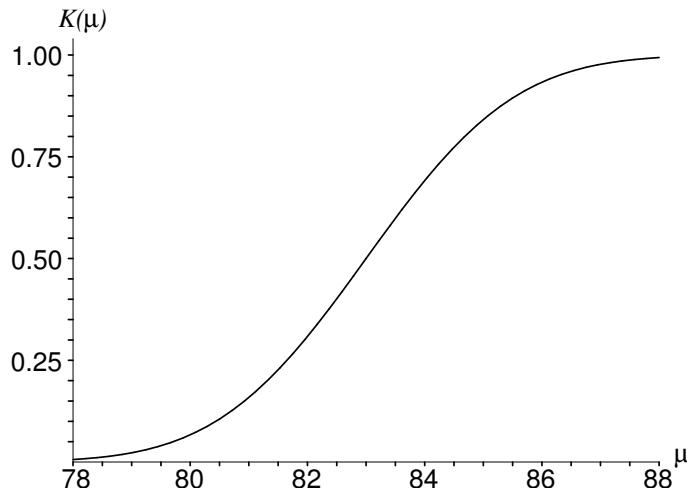
Figure 10.2–4: $K(\mu) = 1 - \Phi([83 - \mu]/2)$

(e) $p\text{-value} = P(\overline{X} \geq 83.41;\ \mu = 80)$

$= P(Z \geq 1.705) = 0.0441.$

10.2–6 (a) $K(\mu) = P(\overline{X} \leq 668.94;\ \mu) = P\left(Z \leq \dfrac{668.94 - \mu}{140/5};\ \mu\right)$

$= \Phi\left(\dfrac{668.94 - \mu}{140/5}\right);$

(b) $\alpha = K(715) = \Phi\left(\dfrac{668.94 - 715}{140/5}\right)$

$= \Phi(-1.645) = 0.05;$

(c) $K(668.94)$ $=$ $\Phi(0) = 0.5;$

$K(622.88)$ $=$ $\Phi(1.645) = 0.95;$

(d)
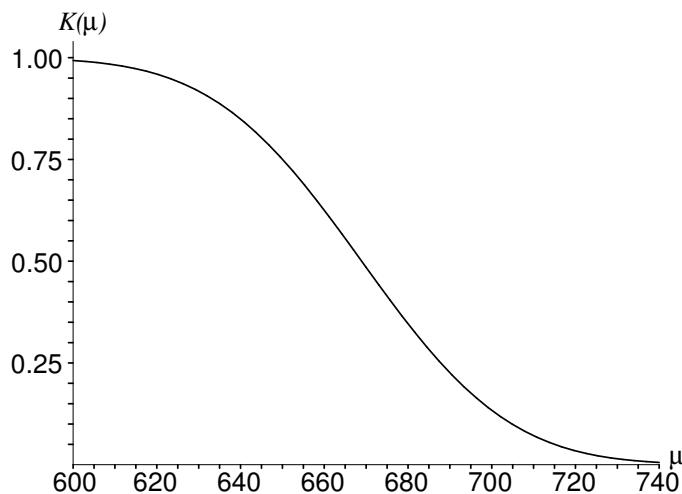
Figure 10.2–6: $K(\mu) = \Phi([668.94 - \mu]/[140/5])$

(e) $\overline{x} = 667.992 < 668.94$, reject H_0;

(f) p-value $=$ $P(\overline{X} \le 667.92; \ \mu = 715)$

$=$ $P(Z \le -1.68) = 0.0465.$

10.2–8 **(a)** and **(b)**
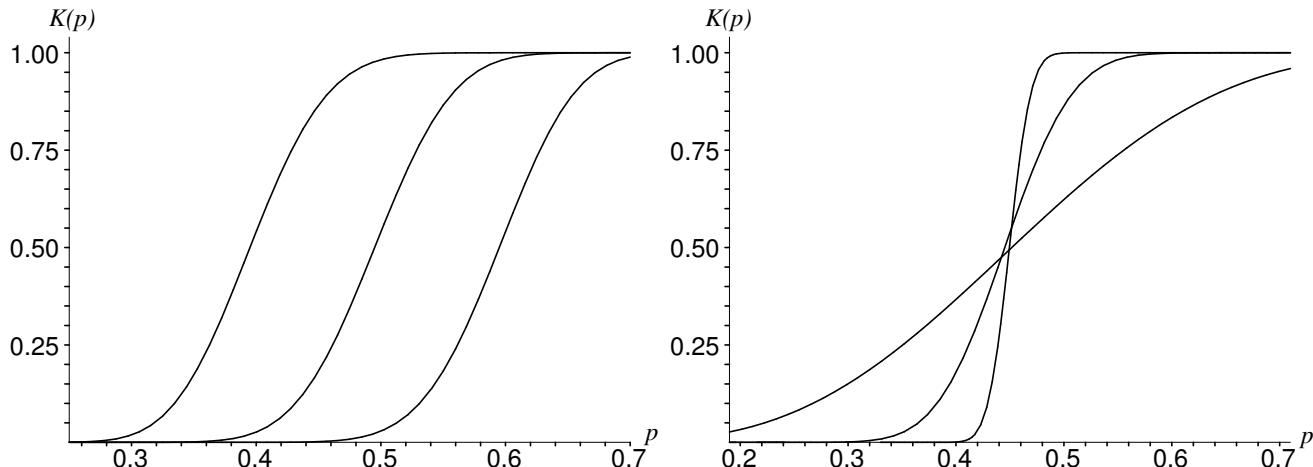
Figure 10.2–8: Power functions corresponding to different critical regions and different sample sizes

10.2–10 Let $Y = \displaystyle\sum_{i=1}^{8} X_i$. Then Y has a Poisson distribution with mean $\mu = 8\lambda$.

(a) α $=$ $P(Y \ge 8; \ \lambda = 0.5) = 1 - P(Y \le 7; \ \lambda) = 0.5$

$=$ $1 - 0.949 = 0.051.$

(b) $K(\lambda) = 1 - \displaystyle\sum_{y=0}^{7} \frac{(8\lambda)^y e^{-8\lambda}}{y!}.$

(c) $\quad K(0.75) \quad = \quad 1 - 0.744 = 0.256,$

$\qquad K(1.00) \quad = \quad 1 - 0.453 = 0.547,$

$\qquad K(1.25) \quad = \quad 1 - 0.220 = 0.780.$
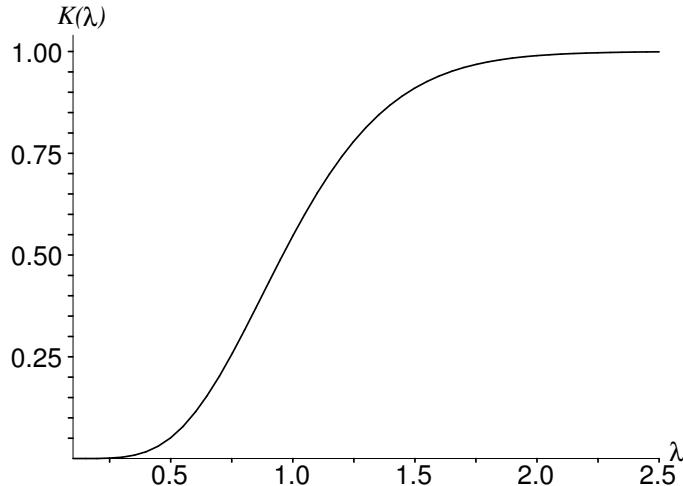
Figure 10.2–10: $K(\lambda) = 1 - P(Y \leq 7; \lambda)$

10.2–12 (a) $\displaystyle\sum_{i=1}^{3} X_i$ has gamma distribution with parameters $\alpha = 3$ and θ. Thus

$$K(\theta) = \int_0^2 \frac{1}{\Gamma(3)\theta^3} x^{3-1} e^{-x/\theta} dx;$$

(b) $\quad K(\theta) = \displaystyle\int_0^2 \frac{x^2 e^{-x/\theta}}{2\theta^3}\, dx \quad = \quad \frac{1}{2\theta^3}\left[-\theta x^2 e^{-x/\theta} - 2\theta^2 x e^{-x/\theta} - 2\theta^3 e^{-x/\theta} \right]_0^2$

$$= \quad 1 - \sum_{y=0}^{2} \frac{(2/\theta)^y}{y!} e^{-2/\theta};$$
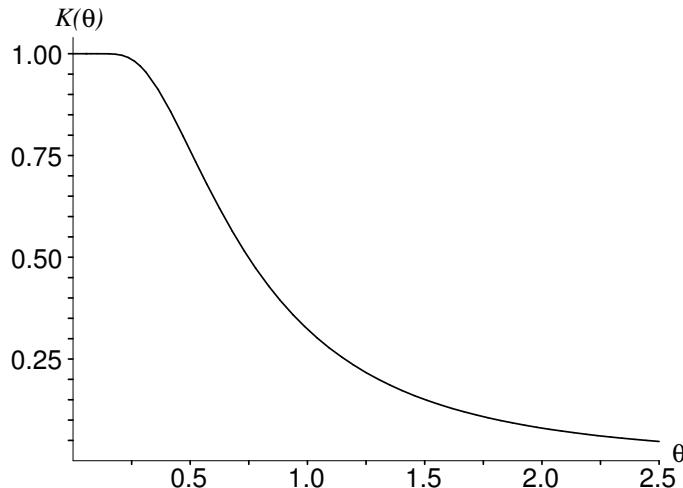
Figure 10.2–12: $K(\theta) = P(\sum_{i=1}^{3} X_i \leq 2)$

(c) $\quad K(2) \quad = \quad 1 - \sum\limits_{y=0}^{2} \dfrac{1^y e^{-1}}{y!} = 1 - 0.920 = 0.080;$

$\qquad\quad K(1) \quad = \quad 1 - 0.677 = 0.323;$

$\qquad K(1/2) \quad = \quad 1 - 0.238 = 0.762;$

$\qquad K(1/4) \quad = \quad 1 - 0.014 = 0.986.$

10.3 Best Critical Regions

10.3–2 (a) $\qquad \dfrac{L(4)}{L(16)} \quad = \quad \dfrac{(1/2\sqrt{2\pi}\,)^n \exp[-\Sigma\, x_i^2/8]}{(1/4\sqrt{2\pi}\,)^n \exp[-\Sigma\, x_i^2/32]}$

$\qquad\qquad\qquad = \quad 2^n \exp[-3\Sigma\, x_i^2/32\,] \le k$

$\qquad -\dfrac{3}{32} \sum\limits_{i=1}^{n} x_i^2 \quad \le \quad \ln k - \ln 2^n$

$\qquad\qquad \sum\limits_{i=1}^{n} x_i^2 \quad \ge \quad -\left(\dfrac{32}{3}\right)(\ln k - \ln 2^n) = c;$

(b) $\quad 0.05 \quad = \quad P\left(\sum\limits_{i=1}^{15} X_i^2 \ge c;\, \sigma^2 = 4\right)$

$\qquad\qquad\quad = \quad P\left(\dfrac{\sum_{i=1}^{15} X_i^2}{4} \ge \dfrac{c}{4};\, \sigma^2 = 4\right)$

$\qquad$ Thus $\dfrac{c}{4} = \chi_{0.05}^2(15) = 25$ and $c = 100.$

(c) $\quad \beta \quad = \quad P\left(\sum\limits_{i=1}^{15} X_i^2 < 100;\, \sigma^2 = 16\right)$

$\qquad\qquad = \quad P\left(\dfrac{\sum_{i=1}^{15} X_i^2}{16} < \dfrac{100}{16} = 6.25\right) \approx 0.025.$

10.3–4 (a) $\quad \dfrac{L(0.9)}{L(0.8)} = \dfrac{(0.9)^{\Sigma x_i}(0.1)^{n-\Sigma x_i}}{(0.8)^{\Sigma x_i}(0.2)^{n-\Sigma x_i}} \quad \le \quad k$

$\qquad\qquad \left[\left(\dfrac{9}{8}\right)\left(\dfrac{2}{1}\right)\right]^{\sum_{1}^{n} x_i}\left[\dfrac{1}{2}\right]^n \quad \le \quad k$

$\qquad\qquad \left(\sum\limits_{i=1}^{n} x_i\right)\ln(9/4) \quad \le \quad \ln k + n\ln 2$

$\qquad\qquad y = \sum\limits_{i=1}^{n} x_i \quad \le \quad \dfrac{\ln k + n\ln 2}{\ln(9/4)} = c.$

$\qquad$ Recall that the distribution of the sum of Bernoulli trials, Y, is $b(n, p)$.

(b) $\quad 0.10 \quad = \quad P[Y \le n(0.85);\, p = 0.9]$

$\qquad\qquad\quad = \quad P\left[\dfrac{Y - n(0.9)}{\sqrt{n(0.9)(0.1)}} \le \dfrac{n(0.85) - n(0.9)}{\sqrt{n(0.9)(0.1)}};\, p = 0.9\right].$

It is true, approximately, that $\dfrac{n(-0.05)}{\sqrt{n}(0.3)} = -1.282$

$$n = 59.17 \text{ or } n = 60.$$

(c) $P[Y > n(0.85) = 51; p = 0.8] = P\left[\dfrac{Y - 60(0.8)}{\sqrt{60(0.8)(0.2)}} > \dfrac{51 - 48}{\sqrt{9.6}}; p = 0.8\right]$

$$\approx P(Z \geq 0.97) = 0.166.$$

(d) Yes.

10.3–6 (a) $0.05 = P\left(\dfrac{\overline{X} - 80}{3/4} \geq \dfrac{c_1 - 80}{3/4}\right)$

$$= 1 - \Phi\left(\dfrac{c_1 - 80}{3/4}\right).$$

Thus

$$\dfrac{c_1 - 80}{3/4} = 1.645$$

$$c_1 = 81.234.$$

Similarly,

$$\dfrac{c_2 - 80}{3/4} = -1.645$$

$$c_2 = 78.766;$$

$$\dfrac{c_3}{3/4} = 1.96$$

$$c_3 = 1.47.$$

(b) $K_1(\mu) = 1 - \Phi([81.234 - \mu]/[3/4]);$

$K_2(\mu) = \Phi([78.766 - \mu]/[3/4];$

$K_3(\mu) = 1 - \Phi([81.47 - \mu]/[3/4]) + \Phi([78.53 - \mu]/[3/4]).$
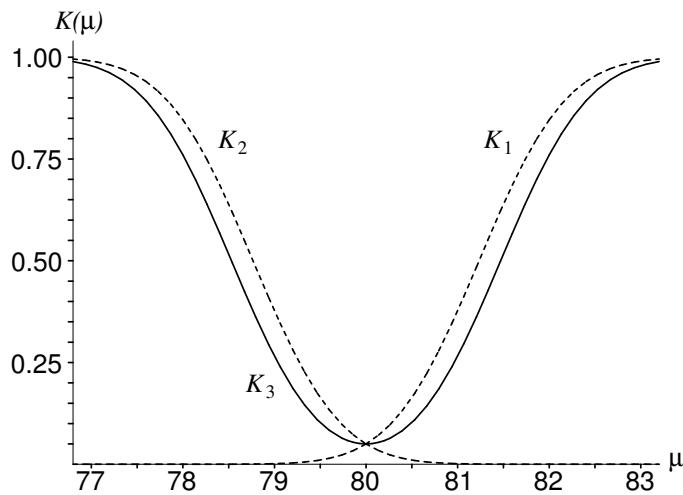
Figure 10.3–6: Three power functions

10.4   Likelihood Ratio Tests

10.4–2   (a) If $\mu \in \omega$ (that is, $\mu \geq 10.35$), then $\widehat{\mu} = \overline{x}$ if $\overline{x} \geq 10.35$, but $\widehat{\mu} = 10.35$ if $\overline{x} < 10.35$.

Thus $\lambda = 1$ if $\overline{x} \geq 10.35$; but, if $\overline{x} < 10.35$, then

$$\lambda = \frac{[1/(0.3)(2\pi)]^{n/2} \exp[-\sum_1^n (x_i - 10.35)^2/(0.06)]}{[1/(0.3)(2\pi)]^{n/2} \exp[-\sum_1^n (x_i - \overline{x})^2/(0.06)]} \quad \leq \quad k$$

$$\exp\left[-\frac{n}{0.06}(\overline{x} - 10.35)^2\right] \quad \leq \quad k$$

$$-\frac{n}{0.06}(\overline{x} - 10.35)^2 \quad \leq \quad \ln k$$

$$\frac{\overline{x} - 10.35}{\sqrt{0.03/n}} \quad \leq \quad \sqrt{-2\ln k} \quad = \quad -z_{0.05}$$

$$= \quad -1.645.$$

(b) $\dfrac{10.31 - 10.35}{\sqrt{0.03/50}} = -1.633 > -1.645$; do not reject H_0.

(c) p-value $= P(Z \leq -1.633) = 0.0513$.

10.4–4   (a) $|z| = \dfrac{|\overline{x} - 59|}{15/\sqrt{n}} \geq 1.96$;

(b) $|z| = \dfrac{|56.13 - 59|}{15/10} = |-1.913| < 1.96$, do not reject H_0;

(c) p-value $= P(|Z| \geq 1.913) = 0.0558$.

10.4–6 $t = \dfrac{324.8 - 335}{40/\sqrt{17}} = -1.051 > -1.337$, do not reject H_0.

10.4–8   In Ω, $\widehat{\mu} = \overline{x}$. Thus,

$$\lambda = \frac{(1/\theta_0)^n \exp[-\sum_1^n x_i/\theta_0]}{(1/\overline{x})^n \exp[-\sum_1^n x_i/\overline{x}]} \quad \leq \quad k$$

$$\left(\frac{\overline{x}}{\theta_0}\right)^n \exp[-n(\overline{x}/\theta_0 - 1)] \quad \leq \quad k.$$

Plotting λ as a function of $w = \overline{x}/\theta_0$, we see that $\lambda = 0$ when $\overline{x}/\theta_0 = 0$, it has a maximum when $\overline{x}/\theta_0 = 1$, and it approaches 0 as $\overline{x}/\theta_0$ becomes large. Thus $\lambda \leq k$ when $\overline{x} \leq c_1$ or $\overline{x} \geq c_2$.

Since the distribution of $\dfrac{2}{\theta_0}\sum_{i=1}^n X_i$ is $\chi^2(2n)$ when H_0 is true, we could let the critical

region be such that we reject H_0 if

$$\frac{2}{\theta_0}\sum_{i=1}^n X_i \leq \chi_{1-\alpha/2}^2(2n) \quad \text{or} \quad \frac{2}{\theta_0}\sum_{i=1}^n X_i \geq \chi_{\alpha/2}^2(2n).$$
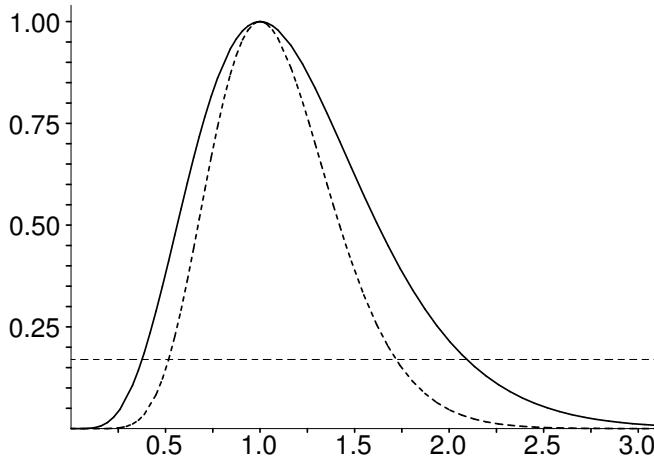
Figure 10.4–8: Likehood functions: solid, $n = 5$; dotted, $n = 10$

10.5 Chebyshev's Inequality and Convergence in Probability

10.5–2 $\text{Var}(X) = 298 - 17^2 = 9$.

(a)
$$P(10 < X < 24) = P(10 - 17 < X - 17 < 24 - 17)$$
$$= P(|X - 17| < 7) \geq 1 - \frac{9}{49} = \frac{40}{49},$$
because $k = 7/3$;

(b) $P(|X - 17| \geq 16) \leq \dfrac{9}{16^2} = 0.035$, because $k = 16/3$.

10.5–4 **(a)** $P\left(\left|\dfrac{Y}{100} - 0.5\right| < 0.08\right) \geq 1 - \dfrac{(0.5)(0.5)}{100(0.08)^2} = 0.609$;

because $k = 0.08/\sqrt{(0.5)(0.5)/100}$;

(b) $P\left(\left|\dfrac{Y}{500} - 0.5\right| < 0.08\right) \geq 1 - \dfrac{(0.5)(0.5)}{500(0.08)^2} = 0.922$;

because $k = 0.08/\sqrt{(0.5)(0.5)/500}$;

(c) $P\left(\left|\dfrac{Y}{1000} - 0.5\right| < 0.08\right) \geq 1 - \dfrac{(0.5)(0.5)}{1000(0.08)^2} = 0.961$,

because $k = 0.08/\sqrt{(0.5)(0.5)/1000}$.

10.5–6 $P(75 < \overline{X} < 85) = P(75 - 80 < \overline{X} - 80 < 85 - 80)$
$$= P(|\overline{X} - 80| < 5) \geq 1 - \frac{60/15}{25} = 0.84,$$

because $k = 5/\sqrt{60/15} = 5/2$.

10.5–8 (a)
$$P(-w < W < w) = 1 - \frac{1}{w^2}$$
$$F(w) - F(-w) = 1 - \frac{1}{w^2}$$
$$F(w) - [1 - F(w)] = 1 - \frac{1}{w^2}$$
$$2F(w) = 2 - \frac{1}{w^2}$$
$$F(w) = 1 - \frac{1}{2w^2}$$

For $w > 1 > 0$, $f(w) = F'(w) = \dfrac{1}{w^3}$.

By symmetry, for $w < -1 < 0$, $f(w) = F'(w) = \dfrac{-1}{w^3}$.

(b)
$$E(W) = \int_{-\infty}^{-} 1 \frac{-w}{w^3}\, dw + \int_{1}^{\infty} \frac{w}{w^3}\, dw$$
$$= -1 + 1 = 0.$$

$E(W^2) = \infty$ so the variance does not exist.

10.6   Limiting Moment-Generating Functions

10.6–2 Using Table III with $\lambda = np = 400(0.005) = 2$, $P(X \leq 2) = 0.677$.

10.6–4 Let $Y = \displaystyle\sum_{i=1}^{n} X_i$, where $X_1, X_2, \ldots, X_n$ are mutually independent $\chi^2(1)$ random variables.
Then $\mu = E(X_i) = 1$ and $\sigma^2 = \text{Var}(X_i) = 2$, $i = 1, 2, \ldots, n$. Hence

$$\frac{Y - n\mu}{\sqrt{n\sigma^2}} = \frac{Y - n}{\sqrt{2n}}$$

has a limiting distribution that is $N(0, 1)$.

10.7   Asymptotic Distributions of Maximum Likelihood Estimators

10.7–2 (a)
$$f(x; p) = p^x (1-p)^{1-x}, \qquad x = 0, 1$$
$$\ln f(x; p) = x \ln p + (1-x) \ln(1-p)$$
$$\frac{\partial \ln f(x; p)}{\partial p} = \frac{x}{p} + \frac{x-1}{1-p}$$
$$\frac{\partial^2 \ln f(x; p)}{\partial p^2} = -\frac{x}{p^2} + \frac{x-1}{(1-p)^2}$$
$$E\left[\frac{X}{p^2} - \frac{X-1}{(1-p)^2}\right] = \frac{p}{p^2} - \frac{p-1}{(1-p)^2} = \frac{1}{p(1-p)}.$$

Rao-Cramér lower bound $= \dfrac{p(1-p)}{n}$.

(b) $\dfrac{p(1-p)/n}{p(1-p)/n} = 1.$

10.7–4 (a)

$$\ln f(x; \theta) = -2 \ln \theta + \ln x - x/\theta$$

$$\frac{\partial \ln f(x; \theta)}{\partial \theta} = -\frac{2}{\theta} + \frac{x}{\theta^2}$$

$$\frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2} = \frac{2}{\theta^2} - \frac{2x}{\theta^3}$$

$$E\left[-\frac{2}{\theta^2} + \frac{2X}{\theta^3}\right] = -\frac{2}{\theta^2} + \frac{2(2\theta)}{\theta^3} = \frac{2}{\theta^2}$$

Rao-Cramér lower bound $= \dfrac{\theta^2}{2n}$.

(b) Very similar to (a); answer $= \dfrac{\theta^2}{3n}$.

(c)

$$\ln f(x; \theta) = -\ln \theta + \left(\frac{1 - \theta}{\theta}\right) \ln x$$

$$\frac{\partial \ln f(x; \theta)}{\partial \theta} = -\frac{1}{\theta} - \frac{1}{\theta^2} \ln x$$

$$\frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2} = \frac{1}{\theta^2} + \frac{2}{\theta^2} \ln x$$

$$E[\ln X] = \int_0^1 \frac{\ln x}{\theta} x^{(1-\theta)/\theta} dx. \quad \text{Let } y = \ln x, \ dy = \frac{1}{x} dx.$$

$$= -\int_0^\infty \frac{y}{\theta} e^{-y(1-\theta)/\theta} e^{-y} dy = -\theta \, \Gamma(2) = -\theta$$

Rao-Cramér lower bound $= \dfrac{1}{n\left(-\dfrac{1}{\theta^2} + \dfrac{2}{\theta^2}\right)} = \dfrac{\theta^2}{n}$.

Chapter 11

Quality Improvement Through Statistical Methods

11.1   Time Sequences

11.1–2
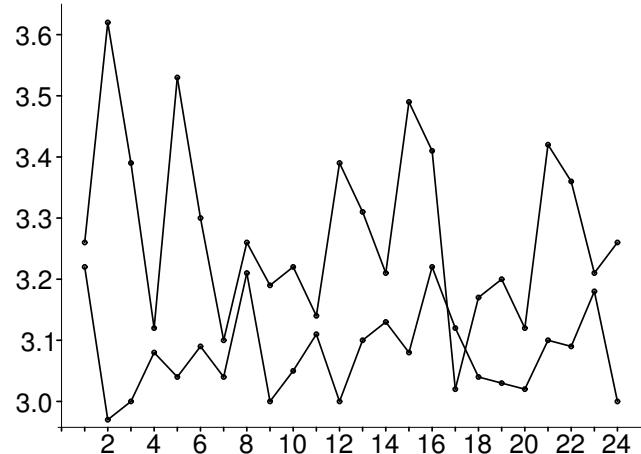
Figure 11.1–2: Apple weights from scales 5 and 6

11.1–4 (a)
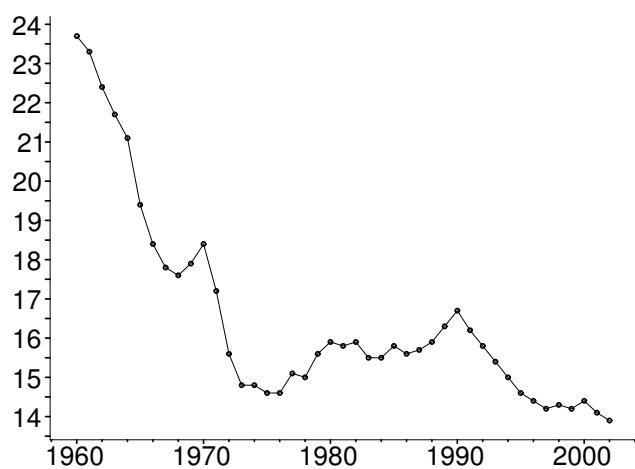
Figure 11.1–4: US birth weights, 1960-1997

(b) From 1960 to the mid 1970's there is a downward trend and then a fairly
steady rate followed by a short upward trend and then another downward trend.

11.1–6 (a) and **(b)**
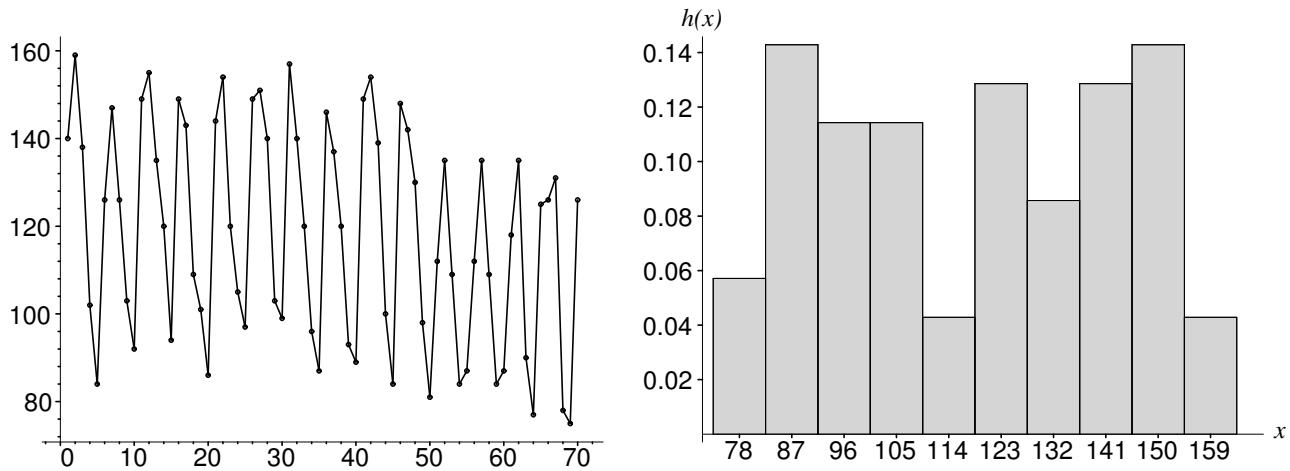
Figure 11.1–6: Force required to pull out studs

(c) The data are cyclic, leading to a bimodal distribution.

11.1–8 (a) and **(b)**
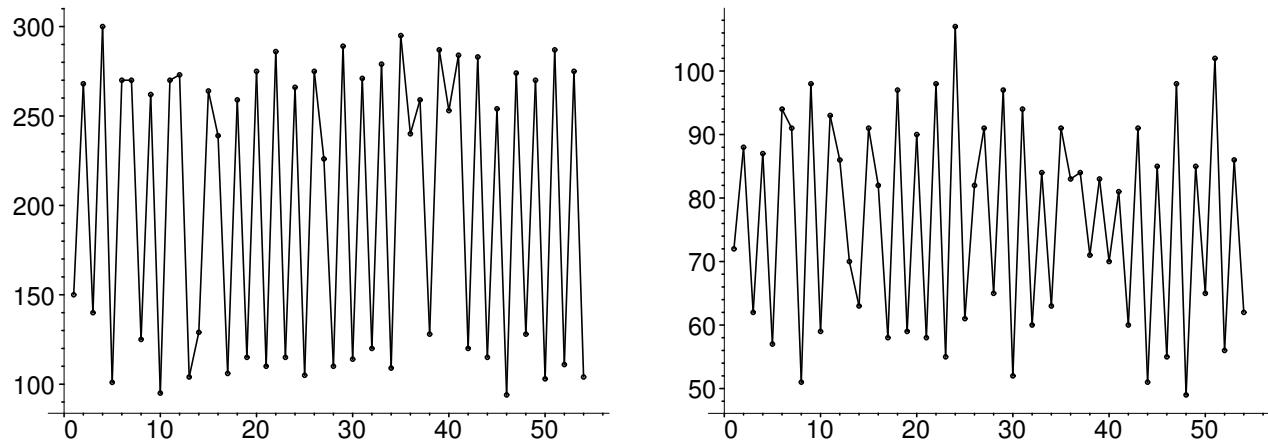
Figure 11.1–8: **(a)** Durations of and **(b)** times between eruptions of Old Faithful Geyser

11.1–10 (a) and **(b)**
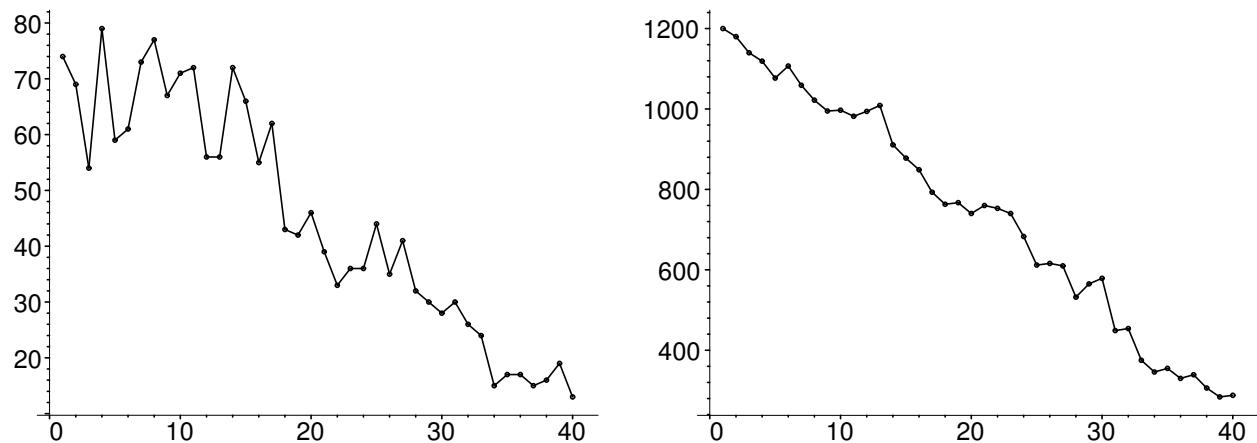
Figure 11.1–10: **(a)** Numbers of users and **(b)** Number of minutes used on each of 40 ports

11.2   Statistical Quality Control

11.2–2 (a) $\bar{\bar{x}} = 67.44, \; \bar{s} = 2.392, \; \overline{R} = 5.88;$

   (b) $\text{UCL} = \bar{\bar{x}} + 1.43(\bar{s}) = 67.44 + 1.43(2.392) = 70.86;$

   $\text{LCL} = \bar{\bar{x}} - 1.43(\bar{s}) = 67.44 - 1.43(2.392) = 64.02;$

   (c) $\text{UCL} = 2.09(\bar{s}) = 2.09(2.392) = 5.00; \; \text{LCL} = 0;$
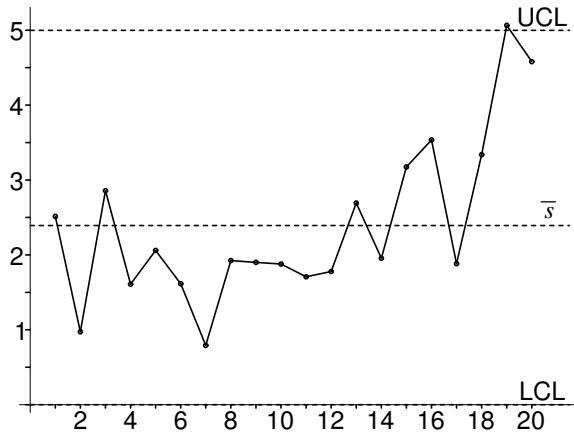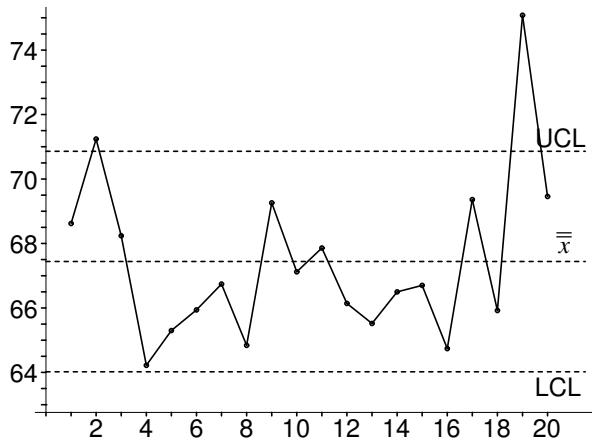
Figure 11.2–2: **(b)** $\bar{x}$-chart using $\bar{s}$ and **(c)** s-chart

   (d) $\text{UCL} = \bar{\bar{x}} + 0.58(\overline{R}) = 67.44 + 0.58(5.88) = 70.85;$

   $\text{LCL} = \bar{\bar{x}} - 0.58(\overline{R}) = 67.44 - 0.58(5.88) = 64.03;$

   (e) $\text{UCL} = 2.11(\overline{R}) = 2.11(5.88) = 12.41; \; \text{LCL} = 0;$
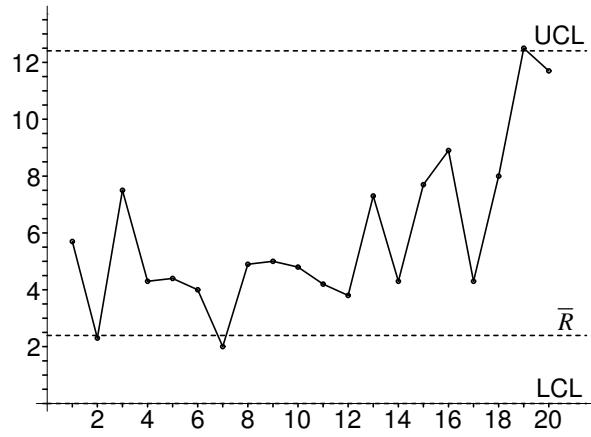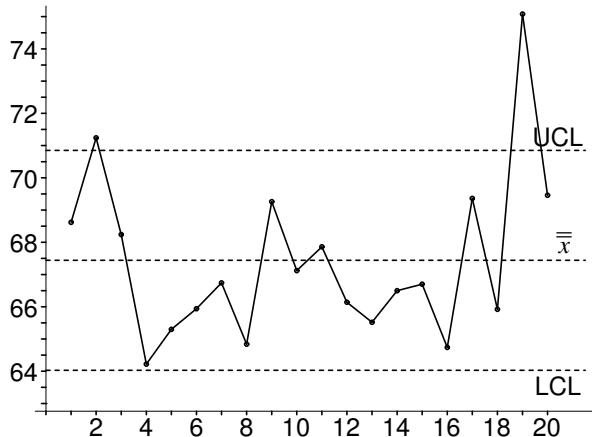
Figure 11.2–2: **(d)** $\bar{x}$-chart using $\overline{R}$ and **(e)** R-chart

   (f) Quite well until near the end.

11.2–4 $\bar{\bar{x}} = 117.141,\ \bar{s} = 1.689,\ \overline{R} = 4.223;$

(a) $\text{UCL} = \bar{\bar{x}} + 0.58(\overline{R}) = 117.141 + 0.58(4.223) = 119.59;$

$\text{LCL} = \bar{\bar{x}} - 0.58(\overline{R}) = 117.141 - 0.58(4.223) = 114.69;$

$\text{UCL} = 2.09(\overline{R}) = 2.11(4.223) = 8.91;\ \text{LCL} = 0;$
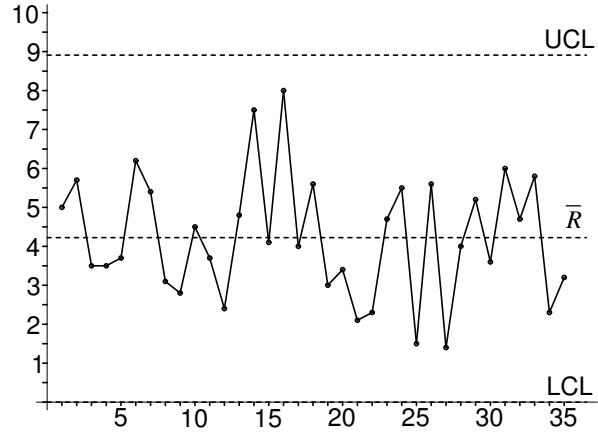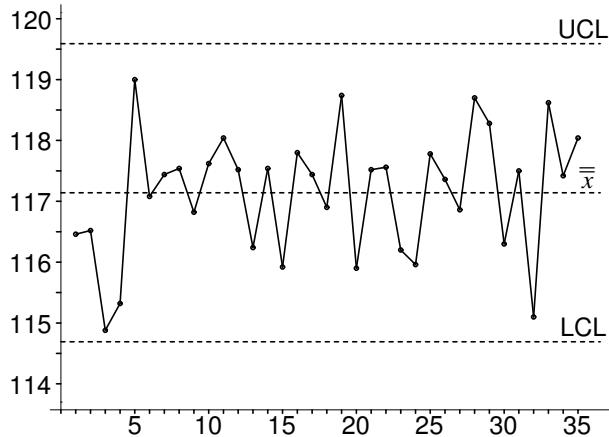
Figure 11.2–4: **(a)** $\bar{x}$-chart using $\overline{R}$ and R-chart

(b) $\text{UCL} = \bar{\bar{x}} + 1.43(\bar{s}) = 117.141 + 1.43(1.689) = 119.56;$

$\text{LCL} = \bar{\bar{x}} - 1.43(\bar{s}) = 117.141 - 1.43(1.689) = 114.73;$

$\text{UCL} = 2.11(\overline{R}) = 2.11(5.88) = 12.41;\ \text{LCL} = 0;$
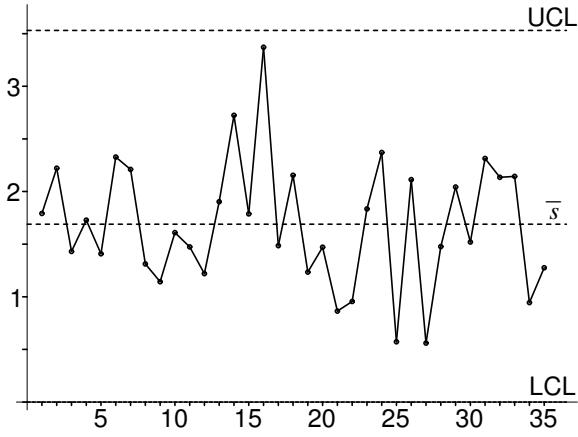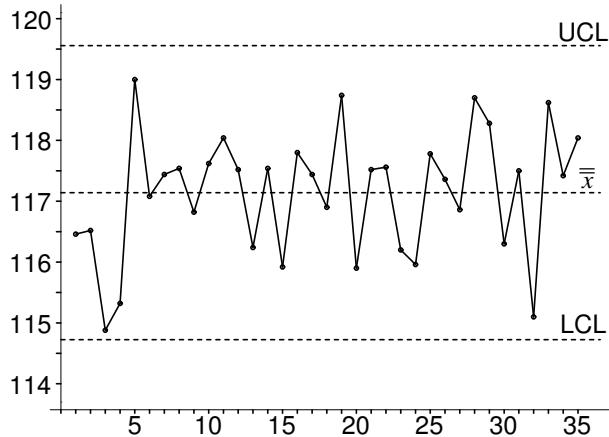
Figure 11.2–4: **(b)** $\bar{x}$-chart using $\bar{s}$ and s-chart

(c) The filling machine seems to be doing quite well.

11.2–6 With $\bar{p} = 0.0254$, $\mathrm{UCL} = \bar{p} + 3\sqrt{\bar{p}(1-\bar{p})/100} = 0.073$;

with $\bar{p} = 0.02$, $\mathrm{UCL} = \bar{p} + 3\sqrt{\bar{p}(1-\bar{p})/100} = 0.062$;

In both cases we see that problems are arising near the end.
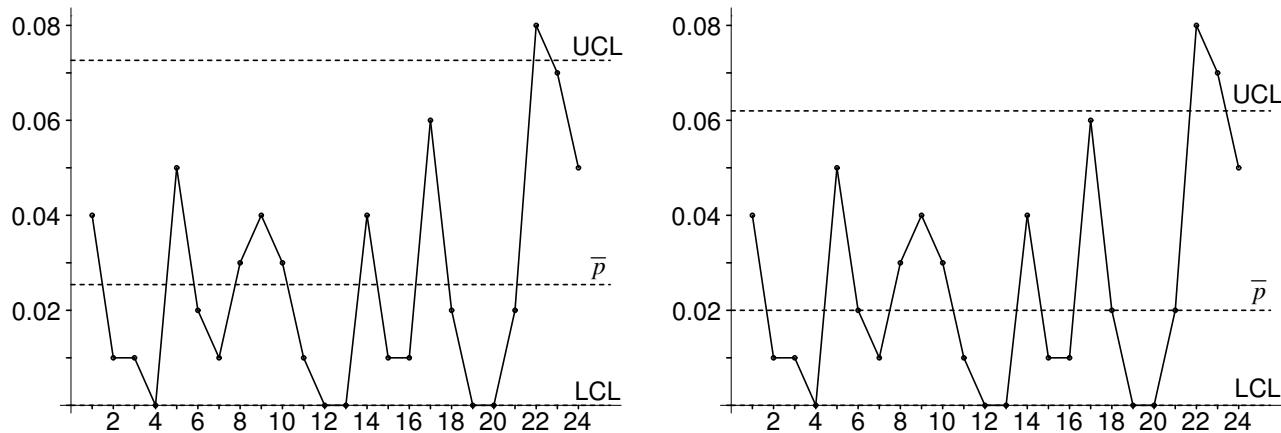
Figure 11.2–6: *p*-charts using $\bar{p} = 0.0254$ and $\bar{p} = 0.02$

11.2–8 **(a)** $\mathrm{UCL} = \bar{c} + 3\sqrt{\bar{c}} = 1.80 + 3\sqrt{1.80} = 5.825$; $\mathrm{LCL} = 0$;
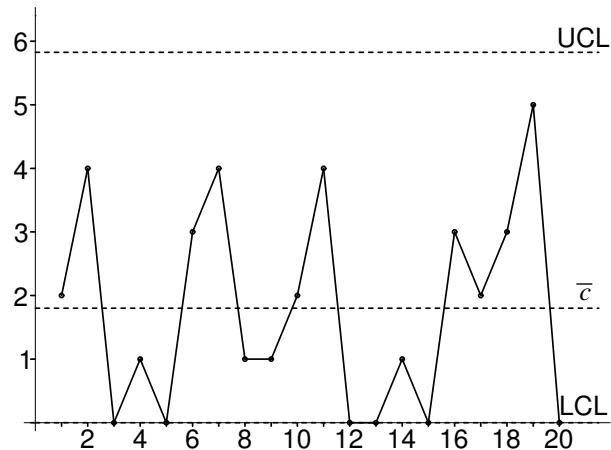

Figure 11.2–8: *c*-chart

(b) The process is in statistical control.

11.3   General Factorial and 2^k Factorial Designs

11.3–4 **(a)** $[A] = -28.4/8 = -3.55$, $[B] = -1.45$, $[C] = 3.2$, $[AB] = -1.525$, $[AC] = -0.525$,

$[BC] = 0.375$, $[ABC] = -1.2$.

(b)

| Identity of Effect	Ordered Effect	Percentile	Percentile from N(0,1)
[A]	−3.550	12.5	−1.15
[AB]	−1.525	25.0	−0.67
[B]	−1.450	37.5	−0.32
[ABC]	−1.200	50.0	0.00
[AC]	−0.525	62.5	0.32
[BC]	0.375	75.0	0.67
[C]	3.20	87.5	1.15

The main effects of temperature (A) and concentration (C) are significant.

Figure 11.3–4: *q-q* plot