INSTRUCTOR'S

SOLUTIONS MANUAL

A BRIEF COURSE IN

MATHEMATICAL STATISTICS

Elliot A. Tanis and Robert V. Hogg

Hope College University of Iowa

August 22, 2006

Contents

Preface

This solutions manual provides answers for the even-numbered exercises in *A Brief Course in Mathematical Statistics* by Elliot A. Tanis and Robert V. Hogg. 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 a "CD" that can be downloaded from Prentice-Hall. Short descriptions of these procedures are provided on the "Maple Card" on the "CD". 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/.

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

Chapter 1

Probability

1.1 Basic Concepts

1.1-2 **(a)** $O = \{$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.

1.1-4 **(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) $P(A' \cup B') = P[(A \cap B)'] = 1 - P(A \cap B) = 1 - 0.3 = 0.7$.

1.1-6 **(a)**
$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ 0.7 &= 0.4 + 0.5 - P(A \cap B) \\ P(A \cap B) &= 0.2; \end{aligned}$$

(b)
$$\begin{aligned} P(A' \cup B') = P[(A \cap B)'] &= 1 - P(A \cap B) \\ &= 1 - 0.2 \\ &= 0.8. \end{aligned}$$

1.1-8 **(a)** $O = \{(1,2), (1,3), (1,4), (1,5), (2,3), (2,4), (2,5), (3,4), (3,5), (4,5)\}$;

(b) **(i)** 1/10; **(ii)** 5/10.

1.1-10 $P(A) = \dfrac{2[r - r(\sqrt{3}/2)]}{2r} = 1 - \dfrac{\sqrt{3}}{2}$.

1.1-12
$$\begin{aligned} A \cup B \cup C &= A \cup (B \cup C) \\ P(A \cup B \cup C) &= P(A) + P(B \cup C) - P[A \cap (B \cup C)] \\ &= P(A) + P(B) + P(C) - P(B \cap C) - P[(A \cap B) \cup (A \cap C)] \\ &= P(A) + P(B) + P(C) - P(B \cap C) - P(A \cap B) - P(A \cap C) \\ &\quad + P(A \cap B \cap C). \end{aligned}$$

1.2 Methods of Enumeration

1.2-2 $(4)(3)(2) = 24$.

1.2-4 **(a)** $(4)(5)(2) = 40$; **(b)** $(2)(2)(2) = 8$.

1.2-6 $O =\{$ FFF, FFRF, FRFF, RFFF, FFRRF, FRFRF, RFFRF, FRRFF, RFRFF, RRFFF, RRR, RRFR, RFRR, FRRR, RRFFR, RFRFR, FRRFR, RFFRR, FRFRR, FFRRR $\}$ so there are 20 possibilities.

1.2-8
$$\binom{n-1}{r} + \binom{n-1}{r-1} = \frac{(n-1)!}{r!(n-1-r)!} + \frac{(n-1)!}{(r-1)!(n-r)!}$$
$$= \frac{(n-r)(n-1)! + r(n-1)!}{r!(n-r)!} = \frac{n!}{r!(n-r)!} = \binom{n}{r}.$$

1.2-10
$$0 = (1-1)^n = \sum_{r=0}^{n} \binom{n}{r}(-1)^r(1)^{n-r} = \sum_{r=0}^{n} (-1)^r \binom{n}{r}.$$
$$2^n = (1+1)^n = \sum_{r=0}^{n} \binom{n}{r}(1)^r(1)^{n-r} = \sum_{r=0}^{n} \binom{n}{r}.$$

1.3 Conditional Probability

1.3-2 **(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.

1.3-4 **(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) $P(\text{Non-Ace Heart, Ace}) + P(\text{Ace of Hearts, Non-Heart Ace})$
$$= \frac{12}{52} \cdot \frac{4}{51} + \frac{1}{52} \cdot \frac{3}{51} = \frac{51}{52 \cdot 51} = \frac{1}{52}.$$

1.3-6 **(a)** $\dfrac{8}{14} \cdot \dfrac{7}{13} = \dfrac{56}{182}$;

(b) $\dfrac{6}{14} \cdot \dfrac{5}{13} = \dfrac{30}{182}$;

(c) $2\left(\dfrac{8}{14} \cdot \dfrac{6}{13}\right) = \dfrac{96}{182}$ or $1 - \left[\dfrac{56}{182} + \dfrac{30}{182}\right] = \dfrac{96}{182}$.

1.3-8
$$\frac{\binom{2}{0}\binom{8}{5}}{\binom{10}{5}} \cdot \frac{2}{5} + \frac{\binom{2}{1}\binom{8}{4}}{\binom{10}{5}} \cdot \frac{1}{5} = \frac{1}{5}.$$

1.3-10 **(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.

1.3-12 $\dfrac{3}{5} \cdot \dfrac{5}{8} + \dfrac{2}{5} \cdot \dfrac{4}{8} = \dfrac{23}{40}$.

1.4 Independent Events

1.4-2 **(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) $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{0}{0.6} = 0$.

1.4-4
$$\begin{aligned} P[A \cap (B \cap C)] &= P[A \cap B \cap C] \\ &= P(A)P(B)P(C) \\ &= P(A)P(B \cap C). \end{aligned}$$

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

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

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

1.4-6 $\dfrac{1}{6} \cdot \dfrac{2}{6} \cdot \dfrac{3}{6} + \dfrac{1}{6} \cdot \dfrac{4}{6} \cdot \dfrac{3}{6} + \dfrac{5}{6} \cdot \dfrac{2}{6} \cdot \dfrac{3}{6} = \dfrac{2}{9}$.

1.4-8 **(a)** $\dfrac{3}{4} \cdot \dfrac{3}{4} = \dfrac{9}{16}$;

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

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

1.4-10 **(a)** $1 - (0.4)^3 = 1 - 0.064 = 0.936$;

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

1.4-12 **(a)** 7; **(b)** $(1/2)^7$; **(c)** 63; **(d)** No! $(1/2)^{63} = 1/9{,}223{,}372{,}036{,}854{,}775{,}808$.

1.5 Bayes's Theorem

1.5-2 **(a)**
$$\begin{aligned} P(G) &= P(A \cap G) + P(B \cap G) \\ &= P(A)P(G \,|\, A) + P(B)P(G \,|\, B) \\ &= (0.40)(0.85) + (0.60)(0.75) = 0.79; \end{aligned}$$

(b)
$$\begin{aligned} P(A \,|\, G) &= \frac{P(A \cap G)}{P(G)} \\ &= \frac{(0.40)(0.85)}{0.79} = 0.43. \end{aligned}$$

1.5-4 Let event B denote an accident and let A_1 be the event that age of the driver is 16–25. Then

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

1.5-6 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

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

1.5-8 Let A be the event that the VCR is under warranty.

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

1.5-10 **(a)** $P(AD) = (0.02)(0.92) + (0.98)(0.05) = 0.0184 + 0.0490 = 0.0674$;

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

(c) $P(N \mid ND) = \dfrac{(0.98)(0.95)}{(0.02)(0.08) + (0.98)(0.95)} = \dfrac{9310}{16 + 9310} = 0.998;$

$P(A \mid ND) = 0.002.$

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

Chapter 2

Discrete Distributions

2.1 Discrete Probability Distributions

2.1-2 **(a)**

$$f(x) = \begin{cases} \dfrac{9}{13}, & x = 0, \\ \dfrac{1}{13}, & x = 1, 2, 3, 4; \end{cases}$$

(b) $P(X \geq 1) = \dfrac{4}{13}$;

(c) $\mu = 0\left(\dfrac{9}{13}\right) + 1\left(\dfrac{1}{13}\right) + 2\left(\dfrac{1}{13}\right) + 3\left(\dfrac{1}{13}\right) + 4\left(\dfrac{1}{13}\right) = \dfrac{10}{13}$;

(d) $\sigma^2 = 0^2\left(\dfrac{9}{13}\right) + 1^2\left(\dfrac{1}{13}\right) + 2^2\left(\dfrac{1}{13}\right) + 3^2\left(\dfrac{1}{13}\right) + 4^2\left(\dfrac{1}{13}\right) - \left(\dfrac{10}{13}\right)^2 = \dfrac{290}{169}$;

(e)

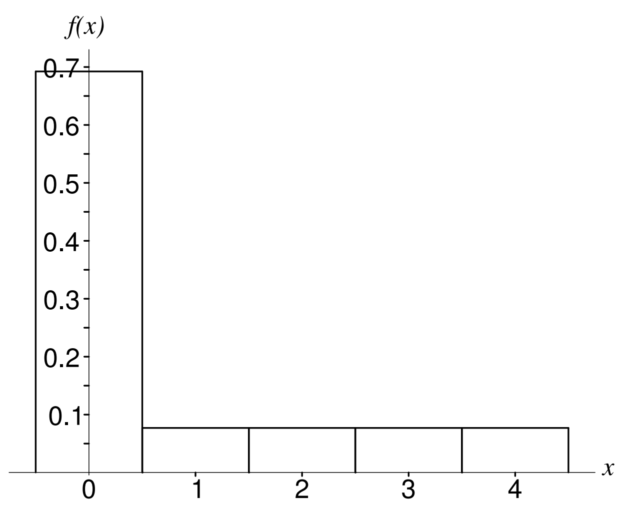


Figure 2.1–2: A Probability Histogram

2.1-4 **(a)** $\mu = 13/3, \sigma^2 = 20/9$;

(b) $\mu = 3, \sigma^2 = 6$;

(c) $\mu = 11/7, \sigma^2 = 19/49$;

(d) $\mu = 0, \sigma^2 = 4/5$.

2.1-6 **(a)** The respective relative frequencies are $0.38, 0.27, 0.21, 0.14$;

(b)

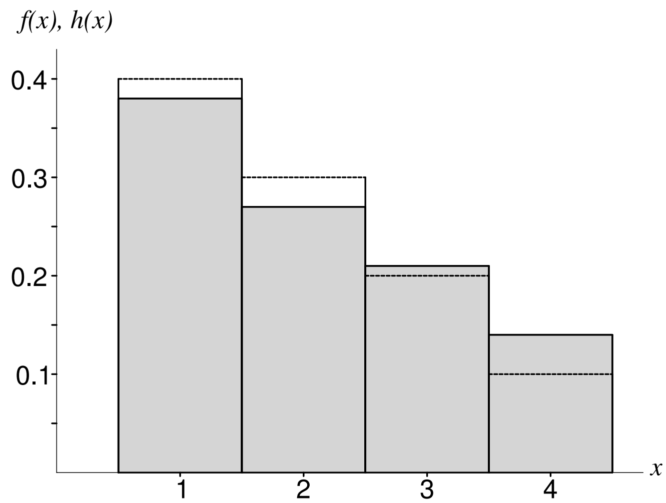


Figure 2.1–6: Relative Frequency Histogram $h(x)$ (shaded), Probability Histogram $f(x)$

(c) $\mu = 2$, $\overline{x} = 2.11$; $\sigma^2 = 1$, $s^2 = 1.149$.

2.1-8 **(a)** $f(x) = \dfrac{13-2x}{36}, \qquad x = 1, 2, 3, 4, 5, 6;$

(b)

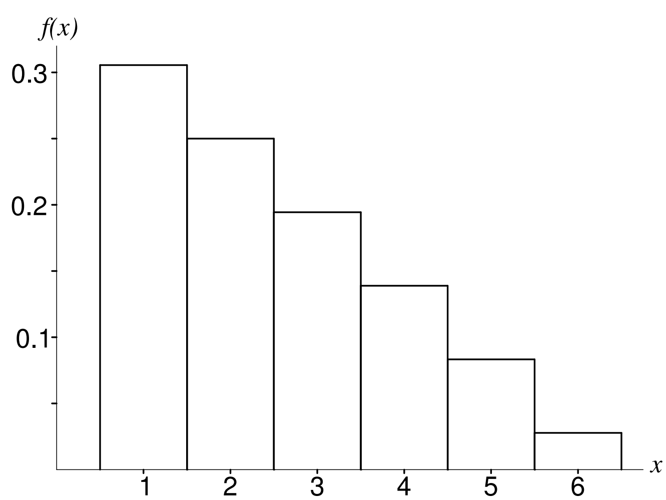


Figure 2.1–8: Probability histogram for the smaller of a pair of dice

(c)

$$g(y) = \begin{cases} \dfrac{6}{36}, & y = 0, \\ \dfrac{12-2y}{36}, & y = 1, 2, 3, 4, 5; \end{cases}$$

(d)

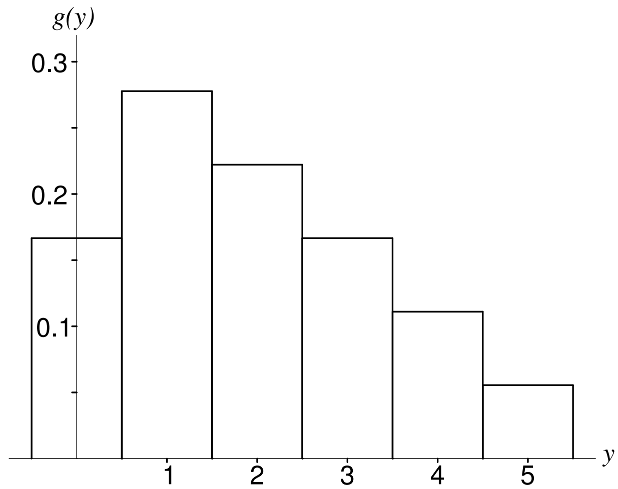


Figure 2.1–8: Probability histogram for the range of a pair of dice

2.1-10 **(a)** $\dfrac{365}{365} \cdot \dfrac{364}{365} \cdots \dfrac{365-r+1}{365} \cdot \dfrac{(365-r)!}{(365-r)!} = \dfrac{365!}{(365-r)!365^r}$;

(b) $P(r) = 1 - \dfrac{365!}{(365-r)!365^r}$;

(c)

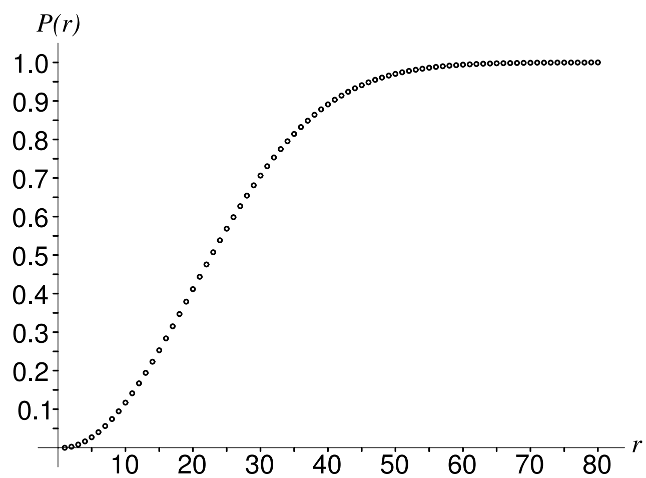


Figure 2.1–10: A Plot of $[r, P(r)]$

(d) **(i)** $r = 32$, **(ii)** $r = 41$;

(e)

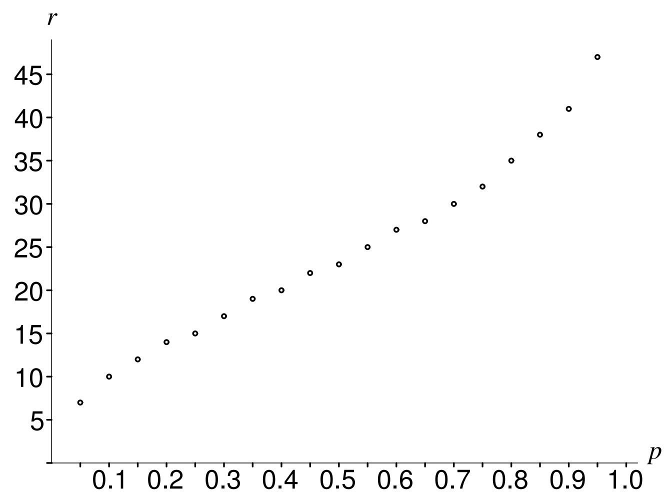


Figure 2.1–10: A Plot of $[p, r]$

2.2 Expectations

2.2-2 $E[X(11-X)] = \sum_{x=1}^{10} x(11-x)(1/10) = 22.$

2.2-4

$$\begin{aligned} 1 &= \sum_{x=0}^{6} f(x) = \frac{9}{10} + c\left(\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}\right) \\ c &= \frac{2}{49}; \\ E(\text{Payment}) &= \frac{2}{49}\left(1\cdot\frac{1}{2}+2\cdot\frac{1}{3}+3\cdot\frac{1}{4}+4\cdot\frac{1}{5}+5\cdot\frac{1}{6}\right) = \frac{71}{490} \text{ units.} \end{aligned}$$

2.2-6 **(a)** $E\left(\dfrac{1}{X}\right) = \sum_{x=1}^{5}\left(\dfrac{1}{x}\right)\left(\dfrac{1}{5}\right) = \dfrac{137}{300} = 0.4567;$

(b)

$$\begin{aligned} \int_1^{101} \frac{1}{x}\,dx &\le \sum_{x=1}^{100} \frac{1}{x} \le 1 + \int_1^{100} \frac{1}{x}\,dx \\ \ln 101 &\le \sum_{x=1}^{100} \frac{1}{x} \le 1 + \ln 100 \\ \frac{\ln 101}{100} &\le \sum_{x=1}^{100}\left(\frac{1}{x}\right)\left(\frac{1}{100}\right) \le \frac{1+\ln 100}{100} \\ 0.04615 &\le \sum_{x=1}^{100}\left(\frac{1}{x}\right)\left(\frac{1}{100}\right) \le 0.05605 \end{aligned}$$

Note that $(0.04615 + 0.05605)/2 = 0.05110$;

(c) Using *Maple*, $\sum_{x=1}^{100}\left(\dfrac{1}{x}\right)\left(\dfrac{1}{100}\right) = 0.05187.$

2.2-8 $E(Y) = E\left(\dfrac{X-\mu}{\sigma}\right) = \dfrac{1}{\sigma}[E(X)-\mu] = 0;$

$$\text{Var}(Y) = \text{Var}\left(\frac{X-\mu}{\sigma}\right) = \frac{E[(X-\mu)^2]}{\sigma^2} = 1.$$

2.2-10 **(a)**

$$\begin{aligned} E(X) &= (-1)p + 0(1-p) + p = 0 \\ E(X^2) &= 1p + 0^2(1-p) + 1p = 2p = \sigma^2 \\ \sigma^4 &= 4p^2 \\ E(X^4) &= (-1)^4 p + 0^4(1-p) + 1^4 p = 2p \\ kurtosis &= \frac{2p}{4p^2} = \frac{1}{2p} \end{aligned}$$

(b) When $p = 1/3, 1/5, 1/10, 1/100$, the kurtosis equals 3/2, 5/2, 5, 50, respectively;

(c)

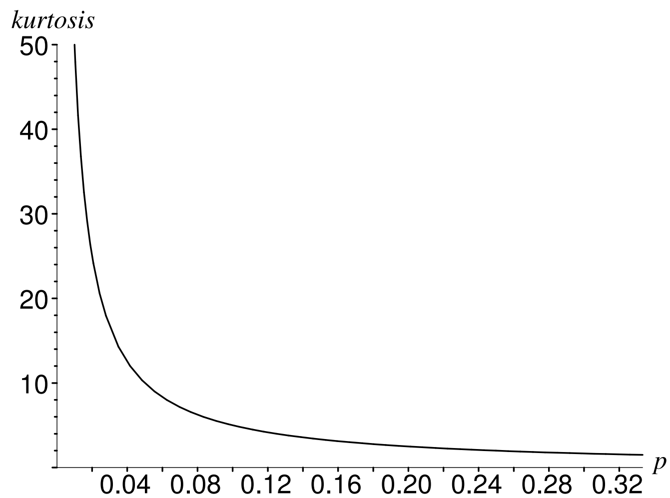


Figure 2.2–10: Kurtosis as a Function of p

2.2-12 $P(X \geq 1) = 5/9$ so $P(X = 0) = 4/9$;

$P(X = 0) = \binom{2}{0} p^0 (1-p)^2 = (1-p)^2 = 4/9$ so $p = 1/3$;

$P(Y \geq 1) = 1 - P(Y = 0) = 1 - (1/3)^0 (2/3)^4 = 65/81$.

2.3 Special Discrete Distributions

2.3-2 X is $b(25, 0.70)$, thus $25 - X$ is $b(25, 0.30)$. Using Table II

(a) $P(X \geq 13) = P(25 - X \leq 25 - 13) = 0.9825$;

(b)
$$\begin{aligned} P(X \leq 11) &= P(25 - X \geq 25 - 11) \\ &= 1 - P(25 - X \leq 13) \\ &= 1 - 0.9940 = 0.0060; \end{aligned}$$

(c)
$$\begin{aligned} P(X = 12) &= P(25 - X = 13) \\ &= P(25 - X \leq 13) - P(25 - X \leq 12) \\ &= 0.9940 - 0.9825 = 0.0115; \end{aligned}$$

(d) $\mu = 25(7/10) = 17.5$, $\sigma^2 = 25(7/10)(3/10) = 5.25$, $\sigma = \sqrt{5.25}$.

2.3-4 **(a)** X is $b(15, 0.2)$,

(b) $\mu = 15(0.2) = 3, \quad \sigma^2 = 15(0.2)(0.8) = 2.4, \quad \sigma = \sqrt{2.4} = 1.549;$

(c) $P(X \geq 5) = 1 - P(X \leq 4) = 1 - 0.8358 = 0.1642.$

2.3-6 $P(X = 0 \,|\, Red) = 125/162 = 0.7716,\ P(X = 0) = 25/36 = 0.6944.$

2.3-8 **(a)** $(4/5)^3(1/5)$;

(b) $f(x) = \left(\frac{4}{5}\right)^{x-1}\left(\frac{1}{5}\right), \qquad x = 1, 2, 3, \ldots;$

$$\mu = \frac{1}{p} = 5,\ \sigma^2 = \frac{1-p}{p^2} = 20.$$

2.3-10 Use the Poisson distribution with $\lambda = np = 1000(0.005) = 5$.

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

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

2.3-12 With $\lambda = 1.5$, $P(X \leq 1) = 0.558$.

2.3-14 With $\lambda = 3$, we want to find x so that $P(X > x) \leq 0.01$ or $P(X \leq x) \geq 0.99$. From Table III we see that $P(X \leq 7) = 0.988$ and $P(X \leq 8) = 0.996$, so let $x = 8$.

2.3-16
$$\begin{aligned} P(X > k + j \,|\, X > k) &= \frac{P(X > k + j)}{P(X > k)} \\ &= \frac{q^{k+j}}{q^k} = q^j = P(X > j). \end{aligned}$$

2.3-18 $1 + \dfrac{1}{3/4} + \dfrac{1}{2/4} + \dfrac{1}{1/4} = \dfrac{25}{3}.$

2.3-20 $f(x) = \dfrac{\binom{3}{x}\binom{2}{2-x}}{\binom{5}{2}}, \qquad x = 0, 1, 2;$

$$\mu = 2\left(\frac{3}{5}\right) = \frac{6}{5};$$

$$\sigma^2 = 2\left(\frac{3}{5}\right)\left(\frac{2}{5}\right)\left(\frac{5-2}{5-1}\right) = \frac{9}{25};$$

The expected payoff is $E[X + 4(2 - X)] = E[8 - 3X] = 4.40$. At a charge of \$4.75, the player can expect to lose 35 cents per play on the average.

2.4 Estimation

2.4-2

$$\begin{aligned}
L(p) &= \prod_{i=1}^{n}\binom{x_i-1}{r-1}p^r(1-p)^{x_i-r}\\
\ln L(p) &= \ln\left[\prod_{i=1}^{n}\binom{x_i-1}{r-1}\right]nr\ln p+\left(\sum_{i=1}^{n}x_i-nr\right)\ln(1-p)\\
\frac{d\ln L(p)}{dp} &= \frac{nr}{p}-\frac{\sum_{i=1}^{n}x_i-nr}{1-p}=0\\
nr-nrp-p\sum_{i=1}^{n}x_i+nrp &= 0\\
\widehat{p} &= \frac{nr}{\displaystyle\sum_{i=1}^{n}X_i}
\end{aligned}$$

Note that $\widehat{p}$ equals the number of successes divided by the number of Bernoulli trials.

2.4-4 **(a)** $[17, 47, 63, 63, 49, 28, 21, 11, 1]$;

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

(c)

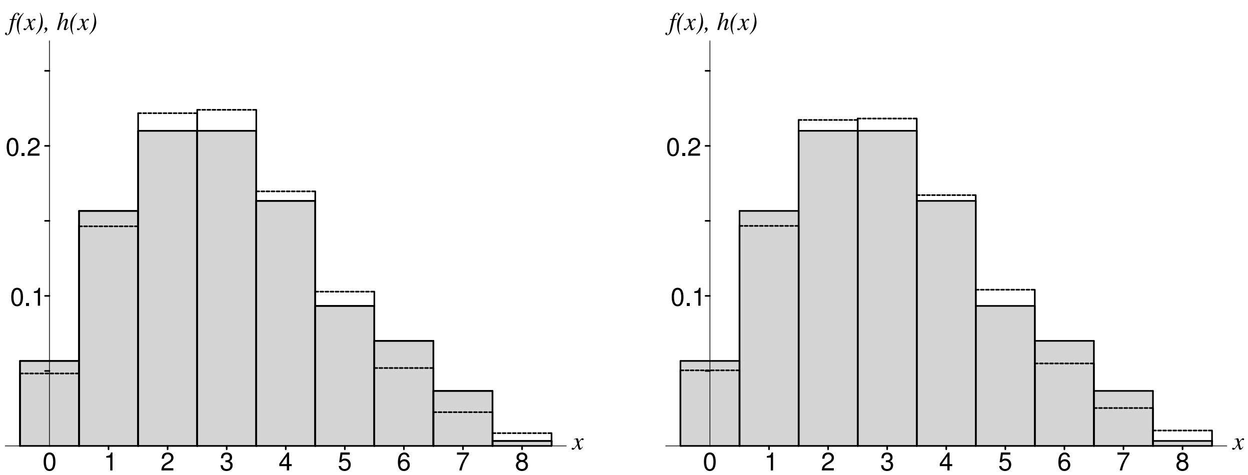


Figure 2.4–4: Background Radiation: Poisson Fit on Left, Translated Negative Binomial Fit on Right

(d) The Poisson distribution seems to provide an excellent probability model with $\lambda = 3.03$. The translated negative binomial distribution with $p = \overline{x}/s^2$ and $r = 57$ also provides an excellent fit.

2.4-6 **(a)** For these data, $\overline{x} = \dfrac{347}{62} = 5.597$, $s^2 = 3.4905$.

(b)

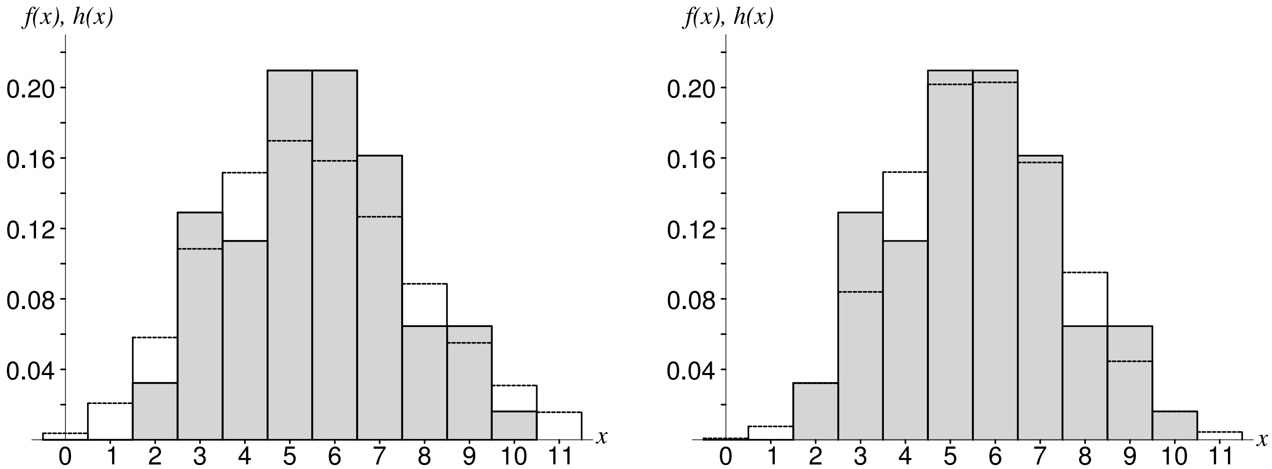


Figure 2.4–6: Number Chocolate Chips per Cookie: Poisson Fit on Left, Binomial Fit on Right

For the Poisson distribution, $\lambda = \overline{x} = 5.597$.

For the binomial distribution, $p = 1 - \dfrac{s^2}{\overline{x}} = 0.376$ and $m = \lceil \overline{x}/p \rceil = 15$.

2.5 Linear Functions of Independent Random Variables

2.5-2
$$\begin{aligned} \mathrm{Var}(Y) &= \mathrm{Var}(3X_2 - X_1) \ = \ 9\mathrm{Var}(X_2) + \mathrm{Var}(X_1) \ = \ 25 \\ 9(2) + k &= 25 \\ k &= \mathrm{Var}(X_1) = 7. \end{aligned}$$

2.5-4
$$\begin{aligned} \mu_Y &= E(X_1X_2) = E(X_1)E(X_2) \ = \ \mu_1\mu_2; \\ \mathrm{Var}(Y) &= E[(X_1X_2 - \mu_1\mu_2)^2] \\ &= E(X_1^2X_2^2) - 2\mu_1\mu_2E(X_1X_2) + \mu_1^2\mu_2^2 \\ &= (\sigma_1^2 + \mu_1^2)(\sigma_2^2 + \mu_2^2) - \mu_1^2\mu_2^2 \\ &= \sigma_1^2\sigma_2^2 + \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2. \end{aligned}$$

2.5-6 **(a)** $P(Y \le y) = P(X_1 \le y)P(X_2 \le y)\cdots P(X_8 \le y) = [1 - (1/2)^y]^8$;

(b) $P(Y = y) = P(Y \le y) - P(Y \le y-1) = [1 - (1/2)^y]^8 - [1 - (1/2)^{y-1}]^8, \quad y = 1, 2, \ldots$;

(c) $E(Y) = \sum_{y=1}^{\infty} y\{[1 - (1/2)^y]^8 - [1 - (1/2)^{y-1}]^8\}$;

(d)

n	$E(Y)$
2	2.6667
4	3.5048
8	4.4211
16	5.3774
32	6.3552
64	7.3440
128	8.3384

(e) Here are the results of 1,000 simulation of this experiment. For these data, $n = 8$ and $\overline{x} = 4.449$.

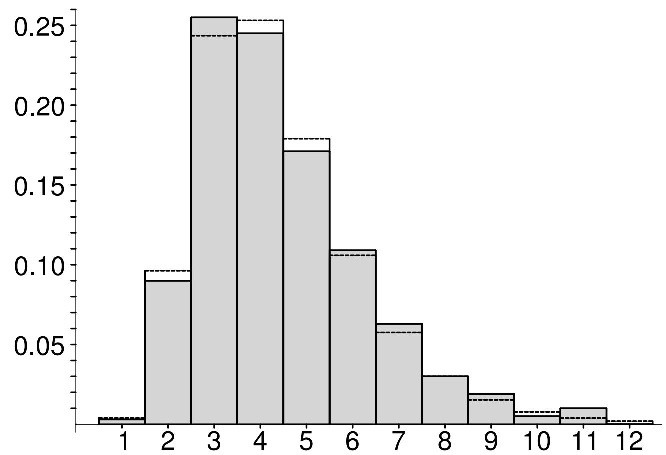


Figure 2.5–6: Number of Flips to Get All Heads with $n = 8$ Fair Coins

2.5-8 $6.05 \pm 2(0.02)/\sqrt{1219}$ or $[6.049,\ 6.051]$.

2.5-10 $0.70 \pm 2\sqrt{(0.70)(0.30)/1234}$ or $[0.674,\ 0.726]$.

2.6 Multivariate Discrete Distributions

2.6-2
$$\begin{aligned}
E[(X_1-\mu_1)(X_2-\mu_2)] &= E[X_1X_2 - \mu_1X_2 - \mu_2X_1 + \mu_1\mu_2] \\
&= E(X_1X_2) - \mu_1E(X_2) - \mu_2E(X_1) + \mu_1\mu_2 \\
&= E(X_1X_2) - \mu_1\mu_2 - \mu_2\mu_1 + \mu_1\mu_2 \\
&= E(X_1X_2) - \mu_1\mu_2.
\end{aligned}$$

2.6-4
$$\begin{aligned}
h(v) &= \sigma_1^2 + 2v\rho\sigma_1\sigma_2 + v^2\sigma_2^2 \;=\; av^2 + bv + c \;\ge\; 0, \\
b^2 - 4ac &= (2\rho\sigma_1\sigma_2)^2 - 4\sigma_1^2\sigma_2^2 \;=\; 4\sigma_1^2\sigma_2^2(\rho^2 - 1) \;\le\; 0,
\end{aligned}$$

so $\rho^2 \le 1$ and $-1 \le \rho \le 1$.

2.6-6
$$\begin{aligned}
\mu_Y &= (4)(5) \;=\; 20, \\
\sigma_Y^2 &= (4)(6) + (6)(2)(0.1)(\sqrt{6})\sqrt{6} \;=\; 24 + 7.2 \;=\; 31.2.
\end{aligned}$$

2.6-8

2	$\frac{1}{4}$	$\frac{3}{4}$	$g(x_1 \mid 2)$
1	$\frac{3}{4}$	$\frac{1}{4}$	$g(x_1 \mid 1)$
	1	2	

equivalently, $g(x_1 \mid x_2) = \dfrac{3 - 2|x_1 - x_2|}{4}$,

$x_1 = 1, 2,$ for $x_2 = 1$ or 2;

	$h(x_2\,\vert\,1)$	$h(x_2\,\vert\,2)$
2	$\frac{1}{4}$	$\frac{3}{4}$
1	$\frac{3}{4}$	$\frac{1}{4}$
	1	2

equivalently, $h(x_2\,|\,x_1) = \dfrac{3-2|x_1-x_2|}{4}$,

$x_2 = 1, 2,$ for $x_1 = 1$ or 2;

$\mu_{X_1|1} = 5/4,\ \mu_{X_1|2} = 7/4,\ \mu_{X_2|1} = 5/4,\ \mu_{X_2|2} = 7/4;$

$\sigma^2_{X_1|1} = \sigma^2_{X_1|2} = \sigma^2_{X_2|1} = \sigma^2_{X_2|2} = 3/16.$

2.6-10 $r = -0.4906.$

Chapter 3

Continuous Distributions

3.1 Descriptive Statistics and EDA

3.1–2 **(a)**

Class Interval	Class Limits	Frequency f_i	Class Mark, u_i
(93.555, 101.555)	(93.56, 101.55)	5	97.555
(101.555, 109.555)	(101.56, 109.55)	11	105.555
(109.555, 117.555)	(109.56, 117.55)	22	113.555
(117.555, 125.555)	(117.56, 125.55)	26	121.555
(125.555, 133.555)	(125.56, 133.55)	22	129.555
(133.555, 141.555)	(133.56, 141.55)	22	137.555
(141.555, 149.555)	(141.56, 149.55)	8	145.555
(149.555, 157.555)	(149.56, 157.55)	4	153.555
(157.555, 165.555)	(157.56, 165.55)	3	161.555
(165.555, 173.555)	(165.56, 173.55)	2	169.555

(b)

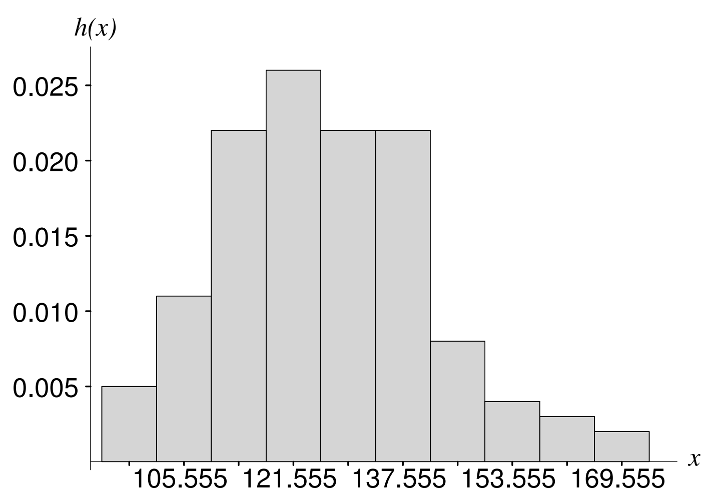


Figure 3.1–2: Old Kent River Bank Run Times

(c) The histogram is skewed slightly to the right.

3.1–4 **(a)** $\overline{x} = 1.335$, $s^2 = 0.003971$, $s = 0.0630$;

(b)

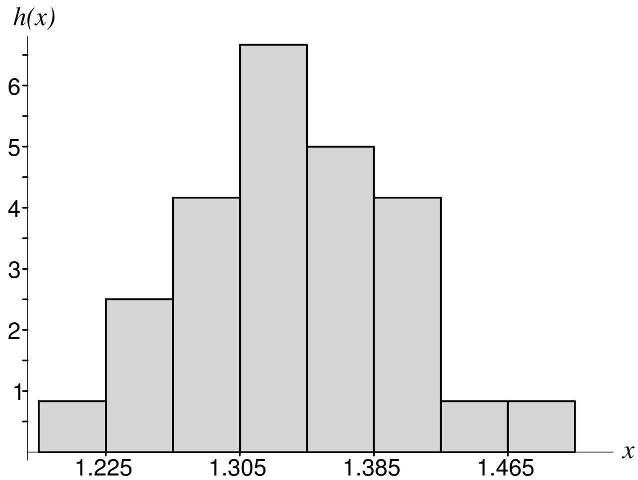


Figure 3.1–4: Diameters of Grains of Soil

3.1–6 **(a)**

Stems	Leaves	Frequency	Depths
20f	5	1	1
20s	6 6 7 7	4	5
20•	8 8 9 9 9	5	10
21∗	0 0 0 0 0 1 1	7	17
21t	2 2 2 2 2 3 3 3 3 3 3 3 3	13	30
21f	4 4 4 4 4 4 4 5 5 5 5 5 5 5 5	15	(15)
21s	6 6 6 6 6 6 6 6 6 6 6 6 7 7 7 7 7 7 7 7 7 7 7	23	36
21•	8 8 8 8 8 8 9 9 9 9	10	13
22∗	0 0 0	3	3

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

(b) **(i)** $\widetilde{q}_1 = \frac{1}{2}(21.2 + 21.2) = 21.2$; $\widetilde{q}_2 = 21.5$; $\widetilde{q}_3 = \frac{1}{2}(21.7 + 21.7) = 21.7$;

(ii) $\widetilde{\pi}_{0.60} = (0.8)21.6 + (0.2)21.6 = 21.6$;

(iii) $\widetilde{\pi}_{0.15} = (0.7)21.0 + (0.3)21.0 = 21.0$;

(c)

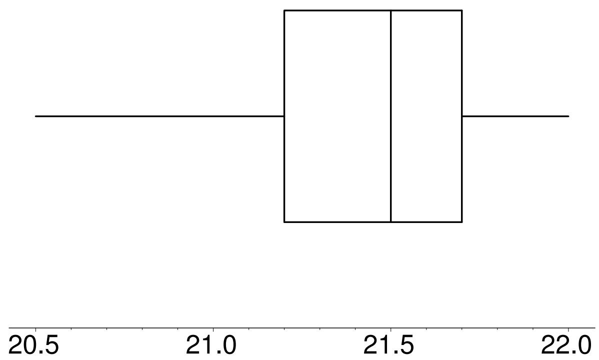


Figure 3.1–6: Weights of Mints

3.1–8 **(a)**

Stems	Leaves	Freq	Depths
0•	55555555555555555555666666666666666777777788888888999999	53	(53)
1∗	0000001111111222334	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) 5, 6, 9, 15, 55;

(c)

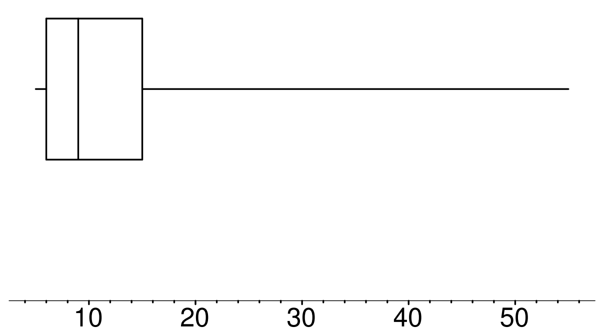


Figure 3.1–8: A Players's Maximum Capital Before Losing $5

(d) 22.8.

3.2 Continuous Probability Distributions

3.2–2 $A_2 \subset A_1'$; so $P(A_2) \le P(A_1') = 1 - 1/4 = 3/4$.

3.2–4 **(a)**

$$\begin{aligned} f'(x) &= 60x^2(1-x) - 20x^3 = 0, \quad x \neq 0 \\ 3(1-x) - x &= 0 \end{aligned}$$

Thus $x = 3/4$ is the mode.

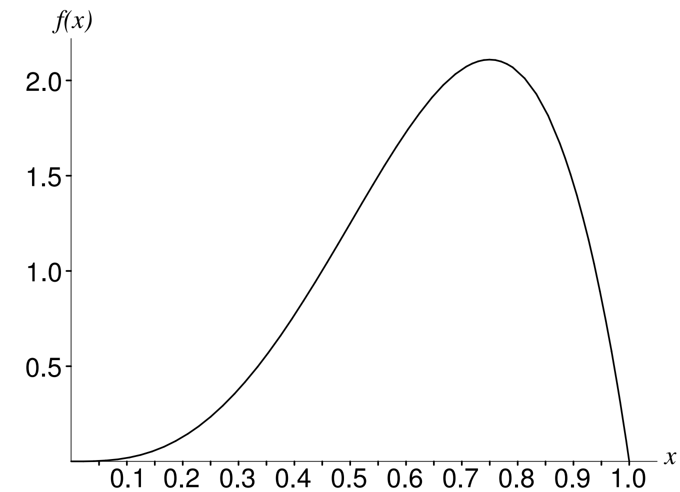


Figure 3.2–4: $f(x) = 20x^3(1-x)$

(b)

$$\begin{aligned} f'(x) &= \frac{x^2 e^{-x}}{2} - \frac{x^3 e^{-x}}{6} = 0, \quad x \neq 0 \\ 3 - x &= 0 \end{aligned}$$

Thus $x = 3$ is the mode.

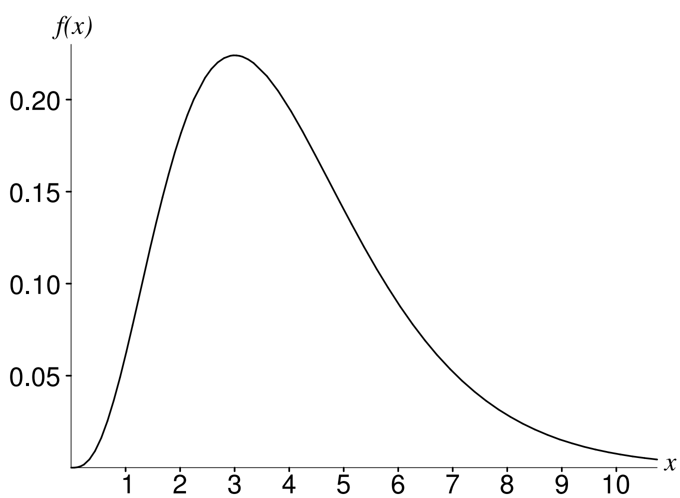


Figure 3.2–4: $f(x) = x^3 e^{-x}/6$

3.2–6 **(a)**

$$\mu = \int_{-2}^{2} \frac{x^4+8x}{32}\,dx = \left[\frac{x^5/5+4x^2}{32}\right]_{-2}^{2} = \frac{2}{5};$$

$$\sigma^2 = \int_{-2}^{2} \frac{x^5+8x^2}{32}\,dx - \mu^2 = \frac{4}{3} - \frac{4}{25} = \frac{88}{75};$$

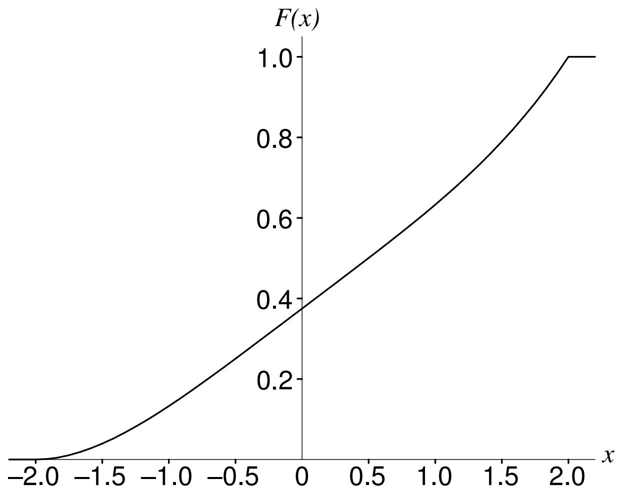


Figure 3.2–6: $F(x) = x^4/128 + x/4 + 3/8$

(b)

$$\mu = \int_{-2}^{4} \frac{x^2+2x}{18}\,dx = \left[\frac{x^3/3+x^2}{18}\right]_{-2}^{4} = 2;$$

$$\sigma^2 = \int_{-2}^{4} \frac{x^3+2x^2}{18}\,dx - 2^2 = \left[\frac{x^4/4+2x^3/3}{18}\right]_{-2}^{4} - 4 = 2;$$

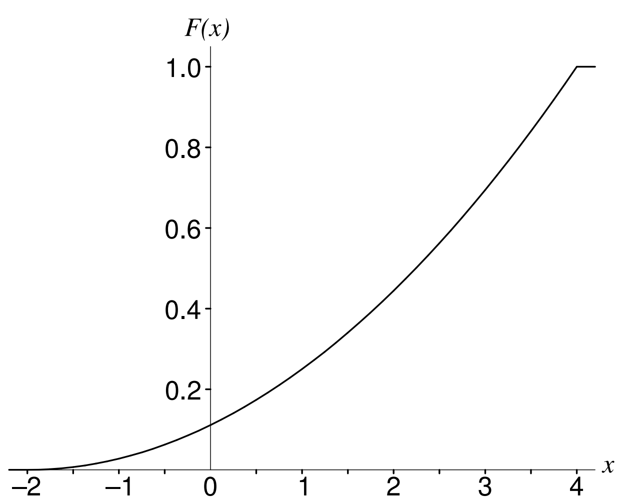


Figure 3.2–6: $F(x) = x^2/36 + x/9 + 1/9$

3.2–8 Recall that a measure of skewness is defined by

$$\frac{E[(X-\mu)^3]}{\{E[(X-\mu)^2]\}^{3/2}} = \frac{E[(X-\mu)^3]}{(\sigma^2)^{3/2}} = \frac{E[(X-\mu)^3]}{\sigma^3}.$$

A measure of kurtosis is defined by

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{E[(X-\mu)^4]}{\sigma^4}.$$

(a) When $f(x) = 1,\ 0 < x < 1$, $\mu = 1/2$, $\sigma^2 = 1/12$, skewness equals 0, kurtosis equals 9/5.

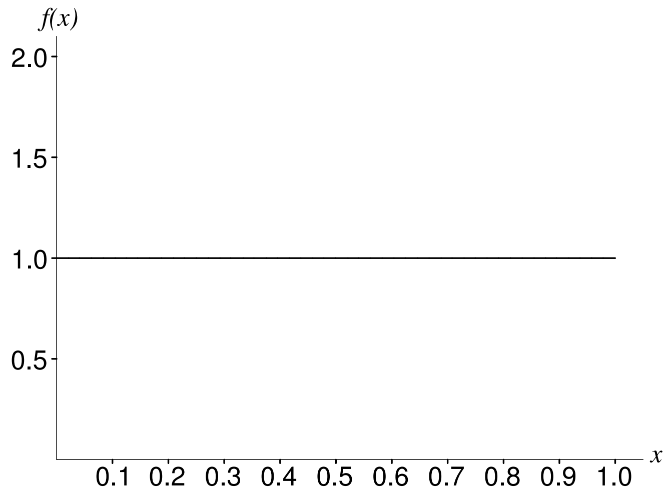


Figure 3.2–8: **(a)** $f(x) = 1$

(b) When $f(x) = 2x,\ 0 < x < 1$, $\mu = 2/3$, $\sigma^2 = 1/18$, skewness equals $-2^{3/2}/5$, kurtosis equals 12/5.

(c) When $f(x) = 2(1-x),\ 0 < x < 1$, $\mu = 1/3$, $\sigma^2 = 1/18$, skewness equals $2^{3/2}/5$, kurtosis equals 12/5.

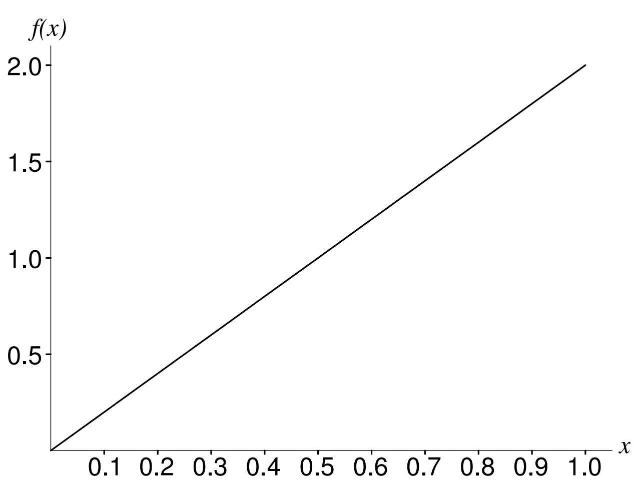
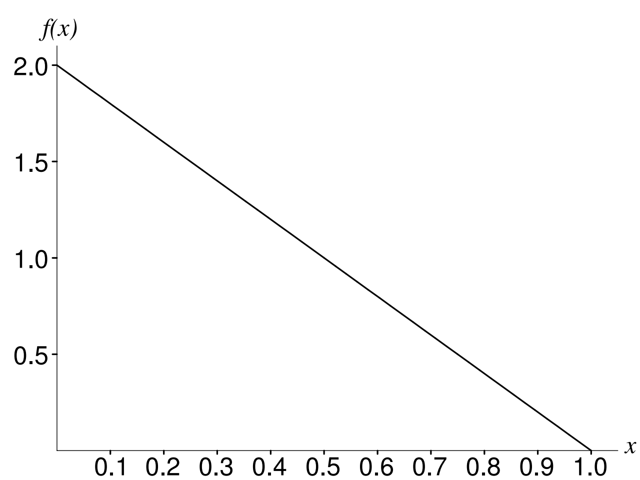


Figure 3.2–8: **(b)** $f(x) = 2x$



(c) $f(x) = 2(1-x)$

3.2–10 $F(x) = (x+1)^2/4, \qquad -1 < x < 1.$

(a)
$$\begin{aligned} F(\pi_{0.64}) = (\pi_{0.64}+1)^2/4 &= 0.64 \\ \pi_{0.64}+1 &= \sqrt{2.56} \\ \pi_{0.64} &= 0.6; \end{aligned}$$

(b)
$$\begin{aligned} (\pi_{0.25}+1)^2/4 &= 0.25 \\ \pi_{0.25}+1 &= \sqrt{1.00} \\ \pi_{0.25} &= 0; \end{aligned}$$

(c)
$$\begin{aligned} (\pi_{0.81}+1)^2/4 &= 0.81 \\ \pi_{0.81}+1 &= \sqrt{3.24} \\ \pi_{0.81} &= 0.8. \end{aligned}$$

3.2–12 **(a)** $\overline{x} = 1.3134;$

(b) $s = 0.5220;$

(c) The respective frequencies are 1, 6, 7, 6, 8, 10, 9, 13, 20, 20.

(d)

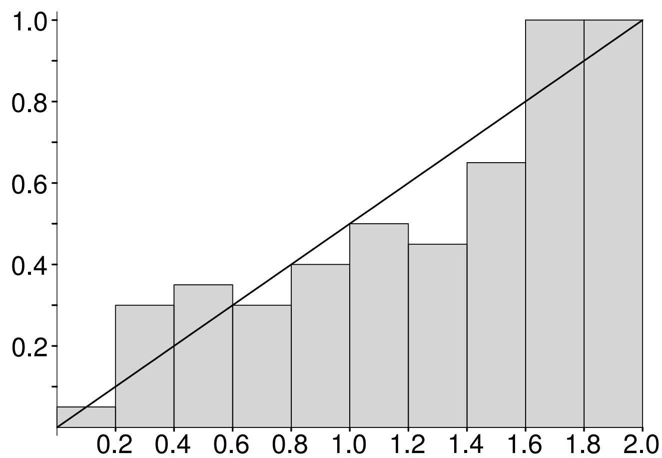


Figure 3.2–12: Relative Frequency Histogram with $f(x) = x/2$ Superimposed

(e) $\mu = \displaystyle\int_0^2 x\left(\frac{x}{2}\right)dx = \left[\frac{x^3}{6}\right]_0^2 = \frac{4}{3};$

$$\begin{aligned} \sigma^2 &= \int_0^2 x^2\left(\frac{x}{2}\right)dx - \left(\frac{4}{3}\right)^2 \\ &= \left[\frac{x^4}{8}\right]_0^2 - \frac{16}{9} = \frac{2}{9}. \end{aligned}$$

3.3 Special Continuous Distributions

3.3–2 $\mu = 0, \quad \sigma^2 = 1/3.$

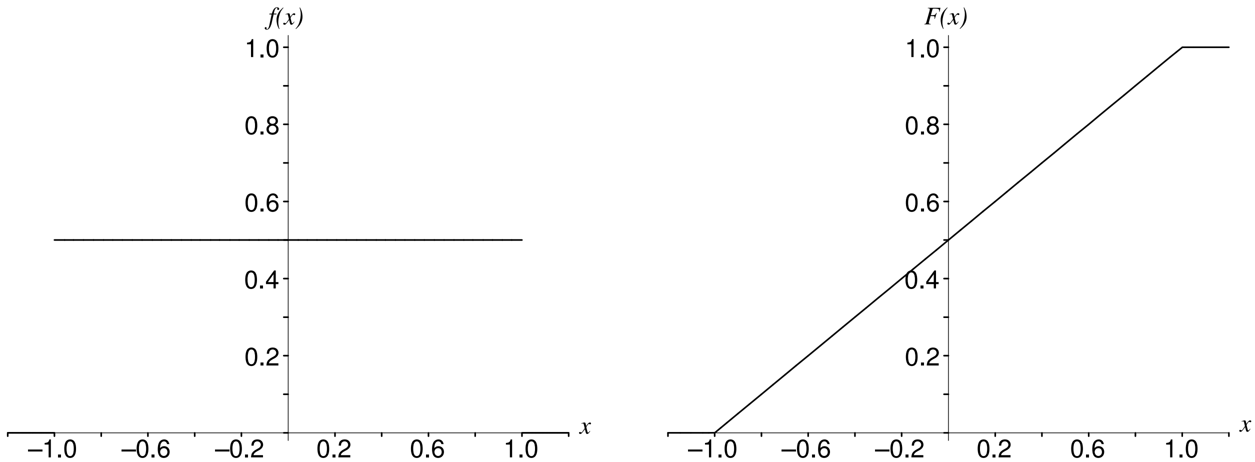


Figure 3.3–2: $f(x) = 1/2$ and $F(x) = (x + 1)/2$

3.3–4 **(a)**

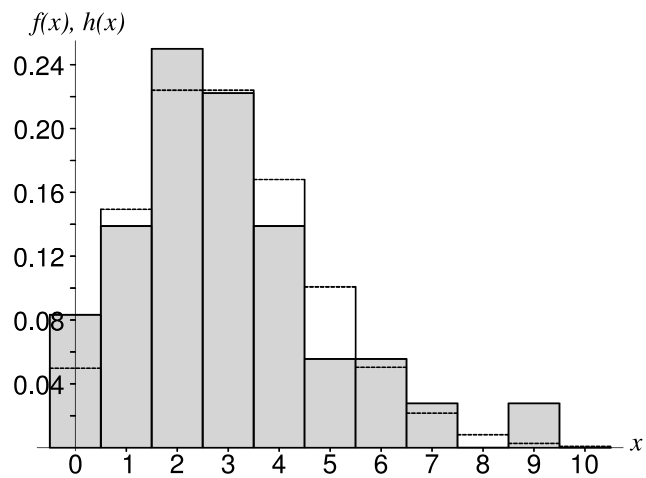


Figure 3.3–4: Relative Frequency Histogram with Poisson p.d.f., $\lambda = 3$, Superimposed

(b) $\overline{x} = 717/35 = 20.4857$, $s = 20.3437$;

(c) $P(X > 15) = e^{-15/20} = 0.4724$. The proportion of observations that is greater than 15 is $49/105 = 0.4667$.

(d) $P(X > 45.5 \,|\, 30.5) = e^{-45.5/20}/e^{-30.5/20} = 0.4724$. The proportion of observations that are greater than 45.5 among the observations that are greater than 30.5 is $12/26 = 0.4615$.

3.3–6 **(a)** $F(x) = 1 - e^{-x/\theta}$, $1/4 = 1 - e^{-x/\theta}$; so $q_1 = \theta \ln(4/3) \approx 0.2877\,\theta$;

(b) $\theta - \theta \ln(4/3) = \theta[1 + \ln(3/4)] \approx 0.7123\,\theta$;

(c) $q_3 = \theta \ln(4) \approx 1.3863\,\theta$;

(d) $q_3 - \theta = \theta[\ln(4) - 1] \approx 0.3863\,\theta$.

3.3–8 From Table IV, $c = 0.831$ and $d = 12.83$.

3.3–10
$$\begin{aligned} G(y) &= P(Y \le y) = P(-2\ln X \le y) \\ &= P(X \ge e^{-y/2}) = 1 - \int_0^{e^{-y/2}} 1\,dx \\ &= 1 - e^{-y/2}, \qquad 0 < y < \infty. \end{aligned}$$

$$g(y) = G'(y) = (1/2)e^{-y/2}, \qquad 0 < y < \infty.$$

This is the p.d.f. for a $\chi^2(2)$ distribution.

3.3–12
$$\begin{aligned} \int_0^\infty \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}\,dx &= \int_0^\infty \frac{1}{\Gamma(\alpha)\theta^\alpha} (\theta y)^{\alpha-1} e^{-y} \theta\,dy \\ &= \frac{1}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-y}\,dy \\ &= \frac{1}{\Gamma(\alpha)} \Gamma(\alpha) = 1. \end{aligned}$$

3.4 The Normal Distribution

3.4–2 Using Table V: **(a)** 2.326; **(b)** -2.576; **(c)** 1.67; **(d)** -2.17.

3.4–4 **(a)** $P\left(\dfrac{600-650}{25} \le \dfrac{X-650}{25} < \dfrac{660-650}{25}\right) = \Phi(0.4) - \Phi(-2) = 0.6326$;

(b) $P\left(-\dfrac{c}{25} < \dfrac{X-650}{25} < \dfrac{c}{25}\right) = 0.9544$; so $c/25 = 2$ and thus $c = 50$.

3.4–6

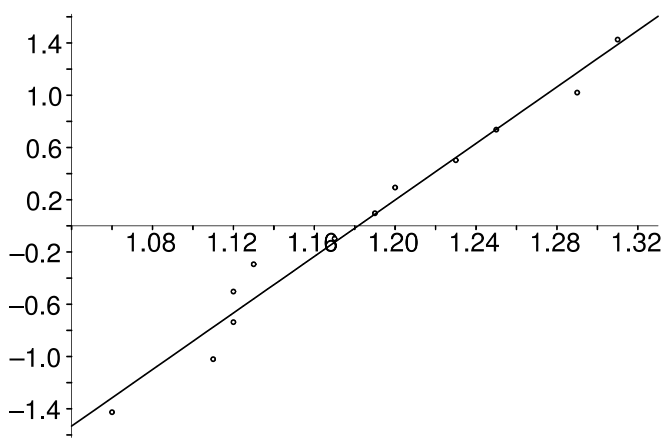


Figure 3.4–6: q-q Plot, $N(0,1)$ Quantiles Versus Weights

3.4–8 **(a)**

Stems	Leaves	Frequency	Depths
7•	938	1	1
8∗	032 089	2	3
8t	222 268 383	3	6
8f	442 490 528 572	4	(4)
8s	674 734 786	3	9
8•	850 873 920	3	6
9∗	069 150	2	3
9t	243	1	1

(Multiply numbers by 10^{-4}.)

(b)

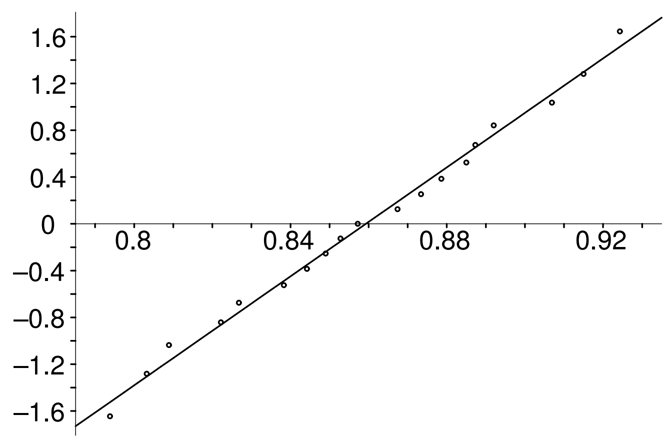


Figure 3.4–8: q-q Plot, $N(0,1)$ Quantiles Versus Weights

(c) Yes. The q-q plot is linear and the stem-and-leaf diagram is "mound shaped."

3.4–10 We must solve $f''(x) = 0$. We have

$$\begin{aligned}
\ln f(x) &= -\ln(\sqrt{2\pi}\,\sigma) - (x-\mu)^2/2\sigma^2, \\
\frac{f'(x)}{f(x)} &= \frac{-2(x-\mu)}{2\sigma^2} \\
\frac{f(x)f''(x) - [f'(x)]^2}{[f(x)]^2} &= \frac{-1}{\sigma^2} \\
f''(x) &= f(x)\left\{\frac{-1}{\sigma^2} + \left[\frac{f'(x)}{f(x)}\right]^2\right\} = 0 \\
\frac{(x-\mu)^2}{\sigma^4} &= \frac{1}{\sigma^2} \\
x - \mu &= \pm\sigma \qquad \text{or} \qquad x = \mu \pm \sigma.
\end{aligned}$$

3.4–12
$$\begin{aligned}E(|X-\mu|) &= 2\int_{\mu}^{\infty}(x-\mu)\,\frac{1}{\sigma\sqrt{2\pi}}\,e^{-(x-\mu)^2/(2\sigma^2)}\,dx\\ &= 2\int_0^{\infty}\sigma y\,\frac{1}{\sqrt{2\pi}}\,e^{-y^2/2}\,dy, \quad \text{using} \quad y=(x-\mu)/\sigma\\ &= \frac{2\sigma}{\sqrt{2\pi}}\left[-e^{-y^2/2}\right]_0^{\infty}\\ &= \sigma\sqrt{2/\pi}.\end{aligned}$$

3.4–14
$$\begin{aligned}\mu &= \int_0^{\infty} x\,\frac{2}{\sqrt{2\pi}}\,e^{-x^2/2}\,dx = \sqrt{\frac{2}{\pi}}\left[-e^{-x^2/2}\right]_0^{\infty} = \sqrt{\frac{2}{\pi}}\\ \sigma^2 &= \int_0^{\infty} x^2\,\frac{2}{\sqrt{2\pi}}\,e^{-x^2/2}\,dx - \frac{2}{\pi} = \int_{-\infty}^{\infty} x^2\,\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}\,dx - \frac{2}{\pi} = 1-\frac{2}{\pi}.\end{aligned}$$

3.5 Estimation in the Continuous Case

3.5–2 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

$$\begin{aligned}\frac{d\ln L(\theta)}{d\theta} &= -\frac{n}{2\theta}+\frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i-\mu)^2 = 0\\ \theta &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2.\end{aligned}$$

Thus

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2.$$

To see that $\widehat{\theta}$ is an unbiased estimator of $\theta=\sigma^2$, note that

$$E(\widehat{\theta}) = E\left(\frac{\theta}{n}\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\theta}\right) = \frac{\theta}{n}\cdot n = \theta = \sigma^2,$$

since $(X_i-\mu)^2/\theta$ is $\chi^2(1)$ and hence the expected value of each of the n summands is equal to 1.

3.5–4 **(a)**

$$\begin{aligned}
L(\theta) &= \left(\frac{1}{\theta^n}\right)\left(\prod_{i=1}^n x_i\right)^{1/\theta-1}, \qquad 0<\theta<\infty \\
\ln L(\theta) &= -n\ln\theta + \left(\frac{1}{\theta}-1\right)\ln\prod_{i=1}^n x_i \\
\frac{d\ln L(\theta)}{d\theta} &= \frac{-n}{\theta} - \frac{1}{\theta^2}\ln\prod_{i=1}^n x_i = 0 \\
\widehat{\theta} &= -\frac{1}{n}\ln\prod_{i=1}^n x_i \\
&= -\frac{1}{n}\sum_{i=1}^n \ln x_i.
\end{aligned}$$

(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

$$E(\widehat{\theta}\,) = -\frac{1}{n}\sum_{i=1}^n(-\theta) = \theta.$$

3.5–6

$$\begin{aligned}
f(x) &= F'(x) = \theta x^{-\theta-1}, \qquad 1<x<\infty \\
L(\theta) &= \theta^n(x_1x_2\cdots x_n)^{-\theta-1}, \qquad 1<x_i<\infty,\quad i=1,2,\ldots,n \\
\ln L(\theta) &= n\ln\theta - (\theta+1)\ln(x_1x_2\cdots x_n) \\
D_\theta[\ln L(\theta)] &= \frac{n}{\theta} - \ln(x_1x_2\cdots x_n) = 0 \\
\widehat{\theta} &= \frac{n}{\ln(X_1X_2\cdots X_n)} \quad \text{because} \quad \frac{d^2[\ln L(\theta)]}{d\theta^2} < 0.
\end{aligned}$$

3.5–8 $0.5 = (2)(4)/\sqrt{n}$; so $\sqrt{n}=16$ and $n=256$.

3.6 The Central Limit Theorem

3.6–2 If

$$\begin{aligned}
f(x) &= (3/2)x^2, \qquad -1<x<1, \\
E(X) &= \int_{-1}^1 x(3/2)x^2\,dx = 0; \\
\mathrm{Var}(X) &= \int_{-1}^1 (3/2)x^4\,dx = \left[\frac{3}{10}x^5\right]_{-1}^1 = \frac{3}{5}.
\end{aligned}$$

Thus

$$\begin{aligned}
P(-0.3\le Y\le 1.5) &= P\left(\frac{-0.3-0}{\sqrt{15(3/5)}} \le \frac{Y-0}{\sqrt{15(3/5)}} \le \frac{1.5-0}{\sqrt{15(3/5)}}\right) \\
&\approx P(-0.10\le Z\le 0.50) = 0.2313.
\end{aligned}$$

3.6–4 **(a)**
$$\mu = \int_0^2 x(1-x/2)\,dx = \left[\frac{x^2}{2}-\frac{x^3}{6}\right]_0^2 = 2-\frac{4}{3}=\frac{2}{3};$$
$$\sigma^2 = \int_0^2 x^2(1-x/2)\,dx - \left(\frac{2}{3}\right)^2 = \left[\frac{x^3}{3}-\frac{x^4}{8}\right]_0^2 - \frac{4}{9} = \frac{2}{9}.$$

(b)
$$P\left(\frac{2}{3}\le \overline{X}\le\frac{5}{6}\right) = P\left(\frac{\frac{2}{3}-\frac{2}{3}}{\sqrt{\frac{2}{9}/18}}\le\frac{\overline{X}-\frac{2}{3}}{\sqrt{\frac{2}{9}/18}}\le\frac{\frac{5}{6}-\frac{2}{3}}{\sqrt{\frac{2}{9}/18}}\right) \approx P(0\le Z\le 1.5)=0.4332.$$

3.6–6 **(a)** $E(\overline{X})=\mu=24.43;$

(b) $\operatorname{Var}(\overline{X})=\dfrac{\sigma^2}{n}=\dfrac{2.20}{30}=0.0733;$

(c)
$$P(24.17\le\overline{X}\le 24.82) \approx P\left(\frac{24.17-24.43}{\sqrt{0.0733}}\le Z\le\frac{24.82-24.43}{\sqrt{0.0733}}\right) = P(-0.96\le Z<1.44)=0.7566.$$

3.6–8 Using the normal approximation,
$$P(1.7\le Y\le 3.2) = P\left(\frac{1.7-2}{\sqrt{4/12}}\le\frac{Y-2}{\sqrt{4/12}}\le\frac{3.2-2}{\sqrt{4/12}}\right) \approx P(-0.52\le Z\le 2.078)=0.6796.$$
Using the p.d.f. of Y,
$$\begin{aligned}P(1.7\le Y\le 3.2) &= \textstyle\int_{1.7}^2[(-1/2)y^3+2y^2-2y+(2/3)]\,dy\\ &\quad+\textstyle\int_2^3[(1/2)y^3-4y^2+10y-22/3]\,dy\\ &\quad+\textstyle\int_3^{3.2}[(-1/6)y^3+2y^2-8y+32/3]\,dy\\ &= [(-1/8)y^4+(2/3)y^3-y^2+(2/3)y]_{1.7}^2\\ &\quad+[(1/8)y^4-(4/3)y^3+5y^2-(22/3)y]_2^3\\ &\quad+[(-1/24)y^4+(2/3)y^3-4y^2+(32/3)y]_3^{3.2}\\ &= 0.1920+0.4583+0.0246=0.6749.\end{aligned}$$

3.6–10 The distribution of $\overline{X}$ is $N(2000, 500^2/25)$. Thus
$$P(\overline{X}>2050)=P\left(\frac{\overline{X}-2000}{500/5}>\frac{2050-2000}{500/5}\right)\approx 1-\Phi(0.50)=0.3085.$$

3.6–12 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
$$P(Y\le 60)=\int_0^{60}\frac{1}{\Gamma(30)2^{30}}y^{30-1}e^{-y/2}\,dy=0.52428 \text{ using Maple.}$$
The normal approximation is given by
$$P\left(\frac{Y-60}{\sqrt{120}}\le\frac{60-60}{\sqrt{120}}\right)\approx\Phi(0)=0.5000.$$

3.6–14 We are given that $Y = \sum_{i=1}^{20} X_i$ has mean 200 and variance 80. We want to find y so that

$$\begin{aligned} P(Y \geq y) &< 0.20 \\ P\left(\frac{Y-200}{\sqrt{80}} > \frac{y-200}{\sqrt{80}}\right) &< 0.20; \end{aligned}$$

We have that

$$\begin{aligned} \frac{y-200}{\sqrt{80}} &= 0.842 \\ y &= 207.5 \uparrow 208 \text{ days.} \end{aligned}$$

3.7 Approximations for Discrete Distributions

3.7–2 **(a)** $P(2 < X < 9) = 0.9532 - 0.0982 = 0.8550$;

(b)
$$\begin{aligned} P(2 < X < 9) &= P\left(\frac{2.5-5}{2} \leq \frac{X - 25(0.2)}{\sqrt{25(0.2)(0.8)}} \leq \frac{8.5-5}{2}\right) \\ &\approx P(-1.25 \leq Z \leq 1.75) \\ &= 0.8543. \end{aligned}$$

3.7–4
$$\begin{aligned} P(35 \leq X \leq 40) &\approx P\left(\frac{34.5-36}{3} \leq Z \leq \frac{40.5-36}{3}\right) \\ &= P(-0.50 \leq Z \leq 1.50) = 0.6247. \end{aligned}$$

3.7–6 $\mu_X = 84(0.7) = 58.8$, $\text{Var}(X) = 84(0.7)(0.3) = 17.64$,

$$P(X \leq 52.5) \approx \Phi\left(\frac{52.5-58.8}{4.2}\right) = \Phi(-1.5) = 0.0668.$$

3.7–8 **(a)**
$$\begin{aligned} P(X < 20.857) &= P\left(\frac{X-21.37}{0.4} < \frac{20.857-21.37}{0.4}\right) \\ &= P(Z < -1.282) = 0.10. \end{aligned}$$

(b) The distribution of Y is $b(100, 0.10)$. Thus

$$P(Y \leq 5) = P\left(\frac{Y - 100(0.10)}{\sqrt{100(0.10)(0.90)}} \leq \frac{5.5-10}{3}\right) \approx P(Z \leq -1.50) = 0.0668.$$

(c)
$$\begin{aligned} P(21.31 \leq \overline{X} \leq 21.39) &\approx P\left(\frac{21.31-21.37}{0.4/10} \leq Z \leq \frac{21.39-21.37}{0.4/10}\right) \\ &= P(-1.50 \leq Z \leq 0.50) = 0.6247. \end{aligned}$$

3.7–10
$$\begin{aligned} P(4776 \leq X \leq 4856) &\approx P\left(\frac{4775.5-4829}{\sqrt{4829}} \leq Z \leq \frac{4857.5-4829}{\sqrt{4829}}\right) \\ &= P(-0.77 \leq Z \leq 0.41) = 0.4385. \end{aligned}$$

3.7–12 The distribution of Y is $b(1000, 18/38)$. Thus

$$P(Y > 500) \approx P\left(Z \geq \frac{500.5 - 1000(18/38)}{\sqrt{1000(18/38)(20/38)}}\right) = P(Z \geq 1.698) = 0.0448.$$

3.7–14 **(a)** $E(X) = 100(0.1) = 10,\ \mathrm{Var}(X) = 9,$

$$\begin{aligned} P(11.5 < X < 14.5) &\approx \Phi\left(\frac{14.5-10}{3}\right) - \Phi\left(\frac{11.5-10}{3}\right) \\ &= \Phi(1.5) - \Phi(0.5) = 0.9332 - 0.6915 = 0.2417. \end{aligned}$$

(b) $P(X \le 14) - P(X \le 11) = 0.917 - 0.697 = 0.220;$

(c) $\displaystyle\sum_{x=12}^{14} \binom{100}{x} (0.1)^x (0.9)^{100-x} = 0.2244.$

3.7–16 **(a)** $E(Y) = 24(3.5) = 84,\ \mathrm{Var}(Y) = 24(35/12) = 70,$

$$P(Y \ge 85.5) \approx 1 - \Phi\left(\frac{85.5-84}{\sqrt{70}}\right) = 1 - \Phi(0.18) = 0.4286;$$

(b) $P(Y < 85.5) \approx 1 - 0.4286 = 0.5714;$

(c) $P(70.5 < Y < 86.5) \approx \Phi(0.30) - \Phi(-1.61) = 0.6179 - 0.0537 = 0.5642.$

3.7–18 **(a)**

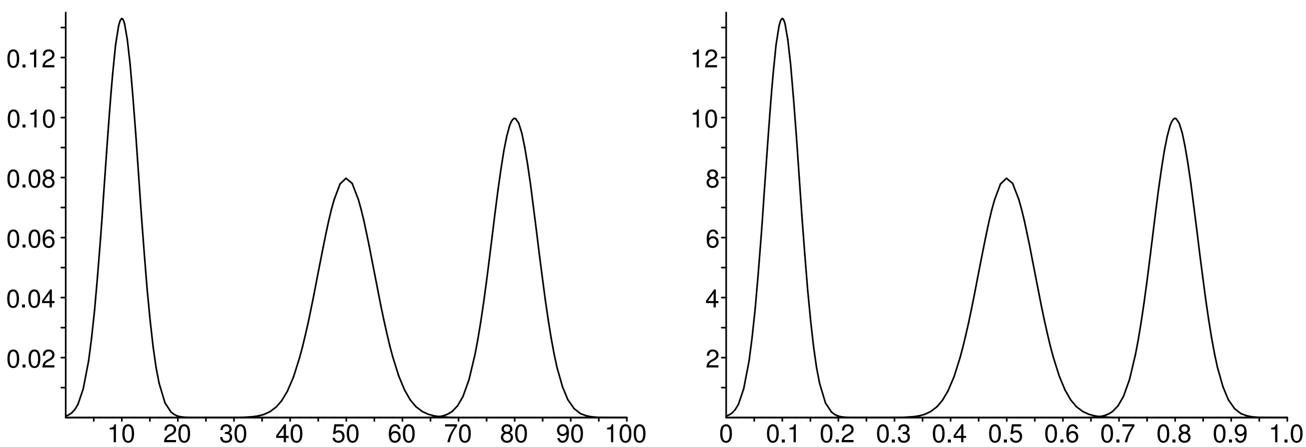


Figure 3.7–18: Normal Approximations of the p.d.f.s of Y and $Y/100$, $p = 0.1, 0.5, 0.8$

(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 $p = 0.8$,

$$P(-1.5 < Y - 80 < 1.5) \approx \Phi\left(\frac{1.5}{4}\right) - \Phi\left(\frac{-1.5}{4}\right) = 0.6462 - 0.3538 = 0.2924.$$

3.7–20

$$\begin{aligned} P(X > 35) &= P\left(\frac{X-25}{5} > \frac{35.5-25}{5}\right) \\ &\approx 1 - \Phi(2.1) = 0.0179. \end{aligned}$$

Note that $P(X > 35) = 0.0225$ using Minitab.

3.7–22 **(a)** Y has a Poisson distribution with mean 30.

(b)
$$\begin{aligned} P(Y \le 25) &= P\left(\frac{Y-30}{\sqrt{30}} \le \frac{25.5-30}{\sqrt{30}}\right) \\ &\approx \Phi(-0.8216) = 0.2057. \end{aligned}$$

Using Minitab, $P(Y \le 25) = 0.2084$.

Chapter 4

Applications of Statistical Inference

4.1 Summary of Necessary Theoretical Results

4.1–2 **(a)** $\mu_Y = \mu_1 + \mu_2$; thus $14 = 3 + \mu_2$, $\mu_2 = 11$; guess $\chi^2(11)$;

(b) From Table IV, $0.99 - 0.01 = 0.98$.

4.1–4 $Y = X_1 + X_2 + X_3$, $\mu_Y = 6 + 4 + 5 = 15$, $\sigma_Y^2 = 4 + 4 + 9 = 17$, so the distribution of Y is $N(15, 17)$.

$$P(Y < 20) = P\left(\frac{Y - 15}{\sqrt{17}} < \frac{5}{\sqrt{17}}\right) = \Phi(1.213) = 0.8875, \quad \text{approximately.}$$

4.1–6 **(a)** $\overline{X} - \overline{Y}$ is the difference of two independent normal random variables with mean $\mu_X - \mu_Y$ and variance $\sigma^2/n + \sigma^2/m = \sigma^2(1/n + 1/m)$; so $\overline{X} - \overline{Y}$ is $N[\mu_X - \mu_Y, \sigma^2(1/n + 1/m)]$.

(b) $(n-1)S_X^2/\sigma^2$ is $\chi^2(n-1)$ and $(m-1)S_Y^2/\sigma^2$ is $\chi^2(m-1)$ and they are independent; so $(n-1)S_X^2/\sigma^2 + (m-1)S_Y^2/\sigma^2$ is $\chi^2(n+m-1)$.

(c) From (a) the numerator of T is $N(0, 1)$ and from (b) the denominator of T is the square root of a chi-square variable divided by its degrees of freedom. Since $(\overline{X}, \overline{Y})$ is independent of (S_X^2, S_Y^2), the numerator and denominator of T are independent. By definition, T has a Students's t distribution with $n + m - 2$ degrees of freedom.

4.2 Confidence Intervals using χ^2, F, and T

4.2–2 **(a)** $\overline{x} = 273.04$, $s^2 = 3155.54$;

(b) $\left[\dfrac{24(3155.54)}{36.42}, \dfrac{24(3155.54)}{13.85}\right] = [2079.4333, 5468.0838]$;

(c) $[\sqrt{2079.4333}, \sqrt{5468.0838}] = [45.60, 73.95]$;

(d) The assumption of normality seems to be valid by considering the q-q plot in Figure 4.2–2.

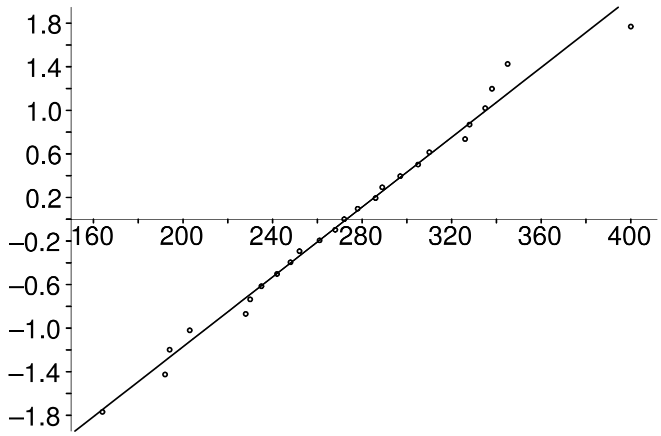


Figure 4.2–2: A *q*-*q* Plot of the Quantiles for $N(0,1)$ Versus Ordered Sample Values

4.2–4 **(a)** $s_x^2/s_y^2 = 0.0040/0.0076 = 0.5263$;

(b) $\left[\dfrac{1}{F_{0.025}(9,8)}\dfrac{s_x^2}{s_y^2},\ F_{0.025}(8,9)\dfrac{s_x^2}{s_y^2}\right] = \left[\left(\dfrac{1}{4.36}\right)(0.5263),\ 4.10(0.5263)\right] = [0.121, 2.158]$.

4.2–6 From the restriction, treating b as a function of a, we have

$$g(b)\frac{db}{da} - g(a) = 0,$$

or, equivalently,

$$\frac{db}{da} = \frac{g(a)}{g(b)}.$$

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,

$$a^{n/2}e^{-a/2} = b^{n/2}e^{-b/2}.$$

4.2–8 **(a)** $\overline{x} = 3.243$;

(b) $s^2 = 0.2372, s = 0.487$;

(c) $[3.243 - 1.796 * 0.487/\sqrt{12}, 3.243 + 1.796 * 0.487/\sqrt{12}] = [2.991, 3.495]$.

4.2–10 **(a)** $(\,\overline{x} + 1.96\sigma/\sqrt{5}\,) - (\,\overline{x} - 1.96\sigma/\sqrt{5}\,) = 3.92\sigma/\sqrt{5} = 1.753\sigma$;

(b) $(\,\overline{x} + 2.776s/\sqrt{5}\,) - (\,\overline{x} - 2.776s/\sqrt{5}\,) = 5.552s/\sqrt{5}$.

$$\begin{aligned} E(S) &= E\left\{\frac{\sigma}{\sqrt{n-1}}\left[\frac{(n-1)S^2}{\sigma^2}\right]^{1/2}\right\} \\ &= \frac{\sigma}{\sqrt{n-1}}\int_0^\infty \frac{v^{1/2}v^{(n-1)/2-1}e^{-v/2}}{\Gamma\left(\dfrac{n-1}{2}\right)2^{(n-1)/2}}\,dv \\ &= \frac{\sigma}{\sqrt{n-1}}\,\frac{\sqrt{2}\,\Gamma(n/2)}{\Gamma[(n-1)/2)]}\,. \end{aligned}$$

With $n = 5$,

$$E(S) = \frac{\sqrt{2}\,\Gamma(5/2)\sigma}{\sqrt{4}\,\Gamma(4/2)} = \frac{3\sqrt{\pi}\sigma}{2^{5/2}} = 0.94\sigma, \text{ so that } E[5.552S/\sqrt{5}\,] = 2.334\sigma.$$

4.2–12 $\left[65.7 - 68.2 \pm 2.485\sqrt{\dfrac{11(16)+14(9)}{25}}\sqrt{\dfrac{1}{12}+\dfrac{1}{15}}\right]$ or $[-5.845, 0.845]$.

4.2–14 **(a)** $\overline{x} = 2.584$, $\overline{y} = 1.564$, $s_x^2 = 0.1042$, $s_y^2 = 0.0428$, $s_p = 0.2711$, $t_{0.025}(18) = 2.101$. Thus a 95% confidence interval for $\mu_X - \mu_Y$ is [0.7653, 1.2747].

(b)

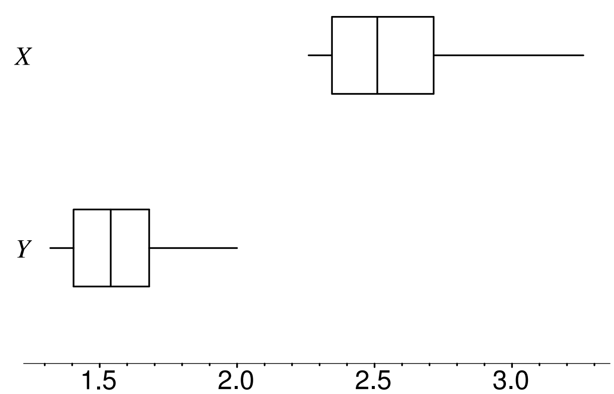


Figure 4.2–14: Box-and-Whisker Diagrams, Wedge On (X) and Wedge Off (Y)

(c) Yes.

4.2–16 $(1.96)\sqrt{10}/\sqrt{n} = 1/2$; $\sqrt{n} = (1.96)(2)\sqrt{10}$; $n = 153.6 \approx 154$.

4.3 Confidence Intervals and Tests of Hypotheses

4.3–2 **(a)** $z = \dfrac{\overline{x} - 170}{10/\sqrt{25}} \ge 1.645$;

(b) $1.260 < 1.645$, do not reject H_0.

4.3–4 $\alpha = P\left(\dfrac{\overline{X} - 530}{90/6} \le \dfrac{510.77 - 530}{90/6};\ \mu = 530\right) = P(Z \le -1.282) = 0.10.$

4.3–6 **(a)** $\overline{x} = 667.92 < 668.94$, reject H_0;

(b) $\alpha = P\left(\dfrac{\overline{X} - 715}{140/5} \le \dfrac{668.94 - 715}{140/5}\right) = P(Z \le -1.645) = 0.05.$

4.4 Basic Tests Concerning One Parameter

4.4-2 **(a)** H_0: $\mu = 3.4$;

(b) H_1: $\mu > 3.4$;

(c) $t = (\overline{x} - 3.4)/(s/3)$;

(d) $t \geq 1.860$;

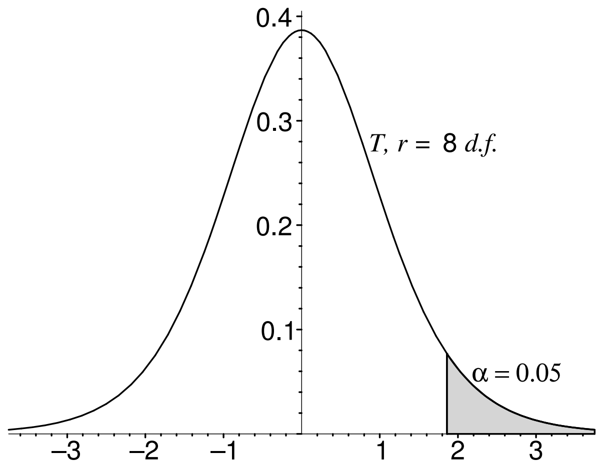


Figure 4.4–2: The Critical Region is $t \geq 1.860$

(e) $t = \dfrac{3.556 - 3.4}{0.167/3} = 2.802$;

(f) $2.802 > 1.860$, reject H_0;

(g) $0.01 < p\text{-value} < 0.025$, $p\text{-value} = 0.0116$.

4.4-4 $z = \dfrac{2.07 - 0}{\sqrt{84.63/51}} = 1.607 < 1.645$, do not reject H_0.

4.4-6 **(a)** $\chi^2 = 24s^2/140^2, \chi^2 \geq 36.42$;

(b) $\chi^2 = \dfrac{24(23,827.4933)}{140^2} = 29.18 < 36.42$, do not reject H_0;

(c) $[0,\ 667.920 + 2.054(140/5)] = [0, 725.432]$.

4.4-8 $\alpha = P(X = 2, 3;\ p = 1/3) = 3\left(\dfrac{1}{3}\right)^2\left(\dfrac{2}{3}\right) + \left(\dfrac{1}{3}\right)^3 = \dfrac{7}{27}$;

$\beta = P(X = 0, 1;\ p = 2/3) = \left(\dfrac{1}{3}\right)^3 + 3\left(\dfrac{1}{3}\right)^2\left(\dfrac{2}{3}\right) = \dfrac{7}{27}$.

4.4-10 Using Table II in the Appendix,

(a) $\alpha = P(Y \geq 13;\ p = 0.40) = 1 - 0.8462 = 0.1538$;

(b)
$$\begin{aligned} \beta &= P(Y \leq 12;\ p = 0.60) \\ &= P(25 - Y \geq 25 - 12) \qquad \text{where } 25 - Y \text{ is } b(25,\ 0.40) \\ &= 1 - 0.8462 = 0.1538. \end{aligned}$$

(c) $p\text{-value} = P(Y \geq 15;\ p = 0.40) = 1 - P(Y \leq 14;\ p = 0.40) = 1 - 0.9656 = 0.0344$.

4.4-12 **(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.

4.5 Tests of the Equality of Two Parameters

4.5–2 **(a)** $t = \dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{12s_x^2 + 15s_y^2}{27}\left(\dfrac{1}{13} + \dfrac{1}{16}\right)}} \le -t_{0.05}(27) = -1.703;$

(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) $0.10 < p\text{-value} < 0.25$;

(d) $\dfrac{s_x^2}{s_y^2} = \dfrac{(25.6)^2}{(28.3)^2} = 0.818 < 2.96 = F_{0.025}(12, 15),$

$\dfrac{s_y^2}{s_x^2} = 1.222 < 3.18 = F_{0.025}(15, 12),$

do not reject equality of variances.

4.5–4 **(a)** Assuming $\sigma_X^2 = \sigma_Y^2$,

$$|t| = \frac{|\,\overline{x} - \overline{y}\,|}{\sqrt{\dfrac{9s_x^2 + 9s_y^2}{18}\left(\dfrac{1}{10} + \dfrac{1}{10}\right)}} \ge t_{0.025}(18) = 2.101;$$

(b) $|-2.151\,| > 2.101$, reject H_0;

(c) $0.01 < p\text{-value} < 0.05$;

(d)

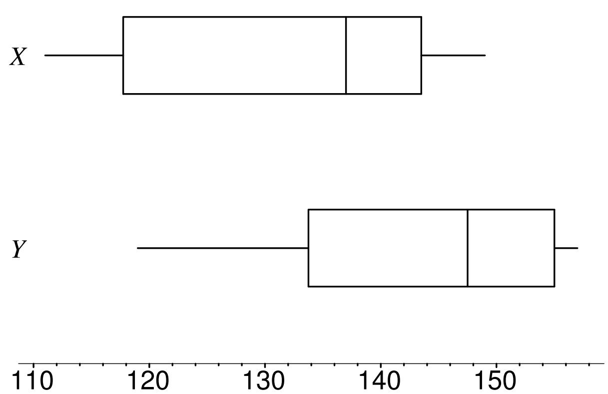


Figure 4.5–4: Box-and-Whisker Diagram for Stud 3 (X) and Stud 4 (Y) Forces

(e) $1.318 < 4.03 = F_{0.025}(9, 9)$, $0.759 < 4.03 = F_{0.025}(9, 9)$, do not reject $\sigma_X^2 = \sigma_Y^2$.

4.5–6 **(a)** $t = \dfrac{80.464 - 80.378 - 0}{\sqrt{\dfrac{9(0.0316) + 9(0.0195)}{18}}\sqrt{\dfrac{1}{10} + \dfrac{1}{10}}} = 1.203.$

$0.20 < p\text{–value} < 0.50$, $p\text{-value} = 0.2446$;

(b) $F = \dfrac{0.0316}{0.0195} = 1.621$, do not reject H_0;

(c)

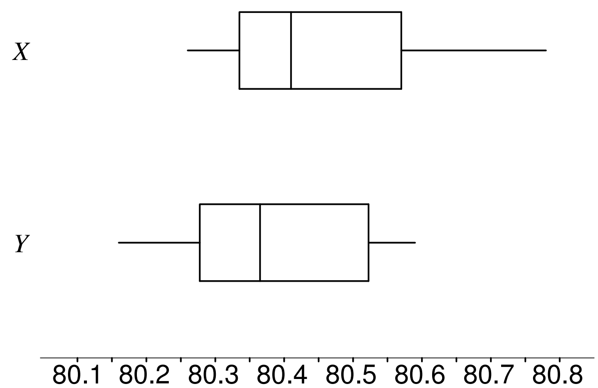


Figure 4.5–6: Box-and-Whisker Diagram for X and Y Weights

This figure confirms the results.

4.5–8 **(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)
$$\begin{aligned} 0.58 - 0.21 &\quad\pm\quad 1.96\sqrt{\frac{(0.58)(0.42)}{605} + \frac{(0.21)(0.79)}{195}} \\ 0.37 &\quad\pm\quad 1.96\sqrt{0.000403 + 0.000851} \\ 0.37 &\quad\pm\quad 0.07 \quad \text{or} \quad [0.30, 0.44]. \end{aligned}$$

It is in agreement with (a).

4.6 Simple Linear Regression

4.6–2 **(a)** $\widehat{y} = 125.7333 + (0.5118)(30.3067/3.4600)(x - 26.6000) = 6.4871 + 4.4829x$;

(b)

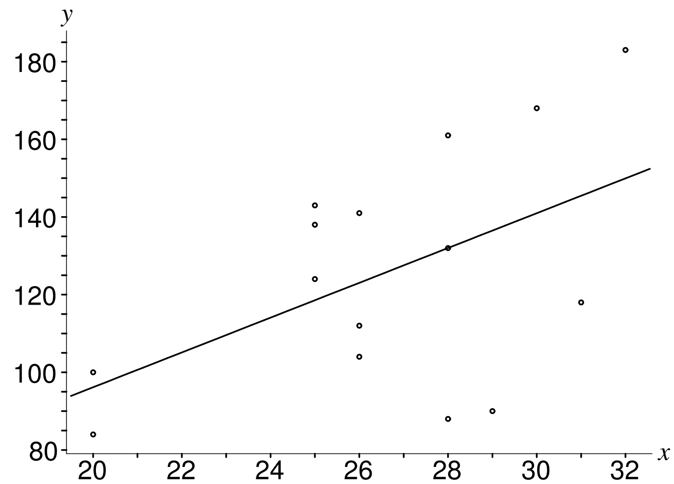


Figure 4.6–2: Final Exam in Calculus (y) Versus ACT Score (x)

4.6–4 Solve for α:

$$\begin{aligned}
-t_{\gamma/2}(n{-}2) \;&\le\; \frac{\widehat{\alpha} - \alpha}{\sqrt{\widehat{\sigma^2}/(n-2)}} \;\le\; t_{\gamma/2}(n{-}2)\\
-t_{\gamma/2}(n{-}2)\sqrt{\widehat{\sigma^2}/(n-2)} \;&\le\; \widehat{\alpha} - \alpha \;\le\; t_{\gamma/2}(n{-}2)\sqrt{\widehat{\sigma^2}/(n-2)}\\
\widehat{\alpha} - t_{\gamma/2}(n{-}2)\sqrt{\widehat{\sigma^2}/(n-2)} \;&\le\; \alpha \;\le\; \widehat{\alpha} + t_{\gamma/2}(n{-}2)\sqrt{\widehat{\sigma^2}/(n-2)}
\end{aligned}$$

4.6–6

$$\begin{aligned}
\widehat{\alpha} &= \overline{y} = 86.80\\
\widehat{\beta} &= \frac{\displaystyle\sum_{i=1}^{n} y_i(x_i - \overline{x})}{\displaystyle\sum_{i=1}^{n} (x_i - \overline{x})^2} = 1.01568\\
\widehat{\sigma^2} &= \frac{1}{n}\sum_{i=1}^{n} [\,y_i - \widehat{\alpha} - \widehat{\beta}(x_i - \overline{x})]^2 = 17.9998.
\end{aligned}$$

A 95% confidence interval for α is $[83.341, 90.259]$;

A 95% confidence interval for β is $[0.478, 1.553]$;

A 95% confidence interval for σ^2 is $[10.265, 82.578]$.

4.6–8 $$T = \frac{\widehat{\beta} - 0}{\sqrt{\dfrac{n\widehat{\sigma^2}}{(n-2)\sum_{i=1}^{n}(x_i - \overline{x}\,)^2}}} = \frac{1.0157}{\sqrt{179.9981/3316}} = 4.3595 > 2.306.$$

Reject H_0.

4.6–10
$$\begin{aligned}\sum_{i=1}^{n}[Y_i-\alpha-\beta(x_i-\overline{x})]^2 &= \sum_{i=1}^{n}[\{\widehat{\alpha}-\alpha\}+\{\widehat{\beta}-\beta\}\{x_i-\overline{x}\} \\ &\quad + \{Y_i-\widehat{\alpha}-\widehat{\beta}\sum_{i=1}^{n}(x_i-\overline{x})\}]^2 \\ &= n(\widehat{\alpha}-\alpha)^2+(\widehat{\beta}-\beta)^2\sum_{i=1}^{n}(x_i-\overline{x})^2 \\ &\quad + \sum_{i=1}^{n}[Y_i-\widehat{\alpha}-\widehat{\beta}(x_i-\overline{x})]^2+0.\end{aligned}$$

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

$$\widehat{\beta}=\sum_{i=1}^{n}(x_i-\overline{x})(Y_i-\overline{Y})\Bigg/\sum_{i=1}^{n}(x_i-\overline{x})^2 .$$

4.6–12 **(b)** The least squares regression line is $1.3596+1.6260x$. It is superimposed on the figure on the left.

(c) The line $y=\phi x=1.6180x$ is superimposed on the figure on the right.

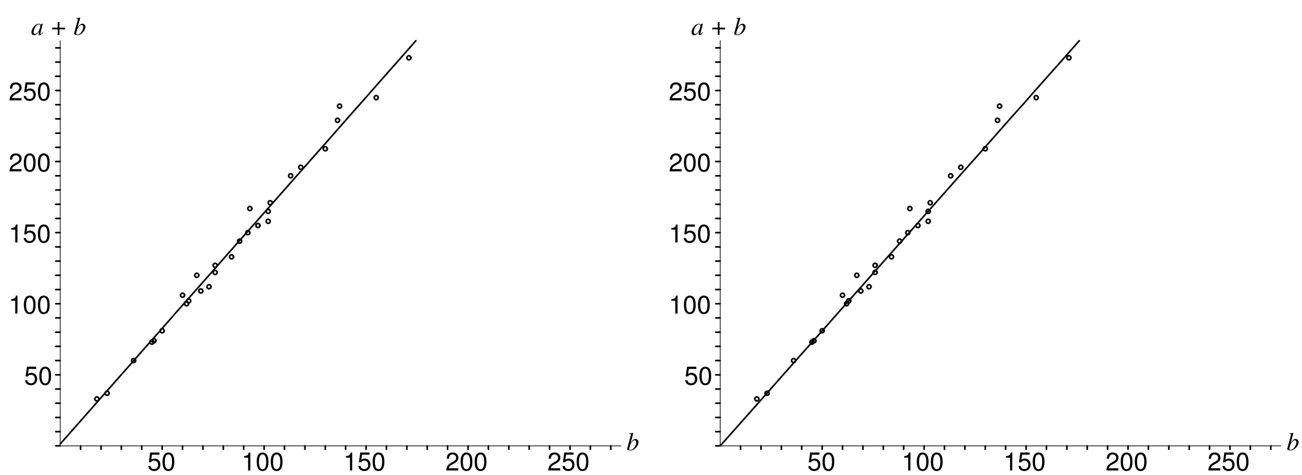


Figure 4.6–12: Scatterplots with Least Squares Regression Line on Left and $y=\phi x$ on Right

(d) The sample mean of the points $(a+b)/b$ is 1.6471 which is close to the "Golden Ratio" $\phi=(1+\sqrt{5}\,)/2=1.6180$.

4.7 More on Linear Regression

4.7–2 **(a)** We shall let $x = b$ and $y = a + b$. From Section 4.6 we have the following:

$$\widehat{\alpha} = \overline{y},$$

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n} y_i(x_i - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2},$$

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}[\,y_i - \widehat{\alpha} - \widehat{\beta}(x_i - \overline{x})]^2$$

From Section 9.3, the endpoints for a 95% confidence interval for $\mu(b) = \mu(x)$ are

$$\widehat{\alpha} + \widehat{\beta}(x - \overline{x}\,) \pm c\,t_{\gamma/2}(n–2)$$

where

$$c = \sqrt{\frac{n\widehat{\sigma^2}}{n-2}}\sqrt{\frac{1}{n} + \frac{(x - \overline{x}\,)^2}{\sum_{i=1}^{n}(x_i - \overline{x}\,)^2}}\,.$$

For our data, $\widehat{\alpha} = 140.6897$, $\widehat{\beta} = 85.6897$, and $\widehat{\sigma^2} = 35.6749$.

b	c	endpoints
50	1.5917	[79.3926, 85.9251]
90	1.1571	[145.3238, 150.0727]
130	1.7861	[209.0726, 216.4026]

(b) From Section 9.3, the endpoints for a 95% prediction interval for $y = a + b$ are given by the following:

$$\widehat{\alpha} + \widehat{\beta}(x_{n+1} - \overline{x}\,) \pm d\,t_{\gamma/2}(n–2)$$

where

$$d = \sqrt{\frac{n\widehat{\sigma^2}}{n-2}}\sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \overline{x}\,)^2}{\sum_{i=1}^{n}(x_i - \overline{x}\,)^2}}.$$

For our data we have the following:

b	d	endpoints
50	6.3915	[69.5435, 95.7742]
90	6.2973	[134.7761, 160.6204]
130	6.4426	[199.5173, 225.9579]

Here are the *Maple* commands that were used to solve this exercise.

```
> a := [38,28,56,56,24,77,40,46,15,39,53,14,102,51,58,39,
  49,58,68,93,90,63,31,74,102,46,28,78,79]:
  b := [62,46,102,88,36,113,69,60,18,63,67,23,171,76,97,73,
  84,92,103,136,155,102,50,93,137,76,45,118,130]:
  apb := a + b:
  read ‘C:\\Tanis-Hogg\\Maple Examples\\RegAnal.txt‘;
```

```
> C := ([.05, .05, .05]):
  RegAnal(b,apb,C);
------------------------------------------------------------

                 Point      Confidence       Confidence

Parameter      Estimates       Level          Interval

------------------------------------------------------------

alpha          140.6897          .95      [138.3311,143.0482]

beta             1.6260          .95       [1.5627,1.6893]

sigma^2         35.6749          .95       [23.9515,70.9906]

------------------------------------------------------------
```

We shall let $x = b$ and let $y = apb$.

```
> x := b:
  y := apb;
> alphahat := evalf(Mean(y));
```

$$alphahat := 140.6896552$$

```
> xbar := evalf(Mean(x));
```

$$xbar := 85.68965517$$

```
> betahat := sum(y[i]*(x[i] - xbar), i = 1 .. 29)/(28*Variance(x));
```

$$betahat := 1.625984165$$

```
> sigmasqhat := sum((y[i] - alphahat - betahat*(x[i] - xbar))^2, i = 1
  .. 29)/29;
```

$$sigmasqhat := 35.67493366$$

Now find 95% confidence intervals for $\mu(b)$ when $x = b = 50, 90, 130$.

```
> for k from 1 to 3 do
    xx := 50 + (k-1)*40;
    c := sqrt(29*sigmasqhat/27)*sqrt(1/29 +
      (xx - xbar)^2/(28*Variance(x)));
    alphahat + betahat*(xx - xbar) - 2.052*c, alphahat +
      betahat*(xx - xbar) + 2.052*c;
  od;
```

$$xx := 50$$

$$c := 1.591742099$$

$$79.39258625,\ 85.92509583$$

$$xx := 90$$

$$c := 1.157141862$$

$$145.3237525,\ 150.0726627$$

$$xx := 130$$

$$c := 1.786071833$$

$$209.0725548,\ 216.4025936$$

Now find the 95% prediction intervals for a + b when b = 50, 90, 130

```
> for k from 1 to 3 do
    xx := 50 + (k-1)*40;
    d := sqrt(29*sigmasqhat/27)*sqrt(1 + 1/29 +
      (xx - xbar)^2/(28*Variance(x)));
    alphahat + betahat*(xx - xbar) - 2.052*d, alphahat +
       betahat*(xx - xbar)+ 2.052*d;
  od;
```

$$xx := 50$$

$$d := 6.391491549$$

$$69.54350038,\ 95.77418170$$

$$xx := 90$$

$$d := 6.297340602$$

$$134.7760647,\ 160.6203505$$

$$xx := 130$$

$$d := 6.442637187$$

$$199.5172827,\ 225.9578657$$

4.7–4 **(a)** From Section 4.6 we have the following:

$$\widehat{\alpha} = \overline{y},$$

$$\widehat{\beta} = \frac{\displaystyle\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})}{\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\displaystyle\sum_{i=1}^{n} y_i(x_i - \overline{x})}{\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2},$$

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}[\,y_i - \widehat{\alpha} - \widehat{\beta}(x_i - \overline{x})]^2$$

From Section 9.3, the endpoints for a 95% confidence interval for $\mu(x)$ are

$$\widehat{\alpha} + \widehat{\beta}(x - \overline{x}\,) \pm c\, t_{\gamma/2}(n–2)$$

where

$$c = \sqrt{\frac{n\widehat{\sigma^2}}{n-2}}\sqrt{\frac{1}{n} + \frac{(x - \overline{x}\,)^2}{\sum_{i=1}^{n}(x_i - \overline{x}\,)^2}}\,.$$

For our data, $\widehat{\alpha} = 26.3333$, $\widehat{\beta} = 0.5062$, $\overline{x} = 23.0667$ and $\widehat{\sigma^2} = 14.1258$.

x	c	endpoints
17	1.6635	[19.6694, 26.8558]
20	1.2313	[22.1215, 27.4407]
23	1.0425	[24.0478, 28.5514]
26	1.2164	[25.1908, 30.4454]
29	1.6414	[25.7912, 32.8820]

(b) From Section 9.3, the endpoints for a 90% prediction interval for Y are given by

$$\widehat{\alpha} + \widehat{\beta}(x_{n+1} - \overline{x}\,) \pm d\, t_{\gamma/2}(n–2)$$

where

$$d = \sqrt{\frac{n\widehat{\sigma^2}}{n-2}}\sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \overline{x}\,)^2}{\sum_{i=1}^{n}(x_i - \overline{x}\,)^2}}.$$

For our data we have the following:

b	d	endpoints
17	4.3665	[15.5296, 30.9957]
20	4.2208	[17.3061, 32.2561]
23	4.1696	[18.9152, 33.6840]
26	4.2165	[20.3507, 35.2854]
29	4.3581	[21.6184, 37.0548]

Figure 4.7-4 shows a 95% confidence band for $\mu(x)$ on the left and a 90% prediction band for Y on the right.

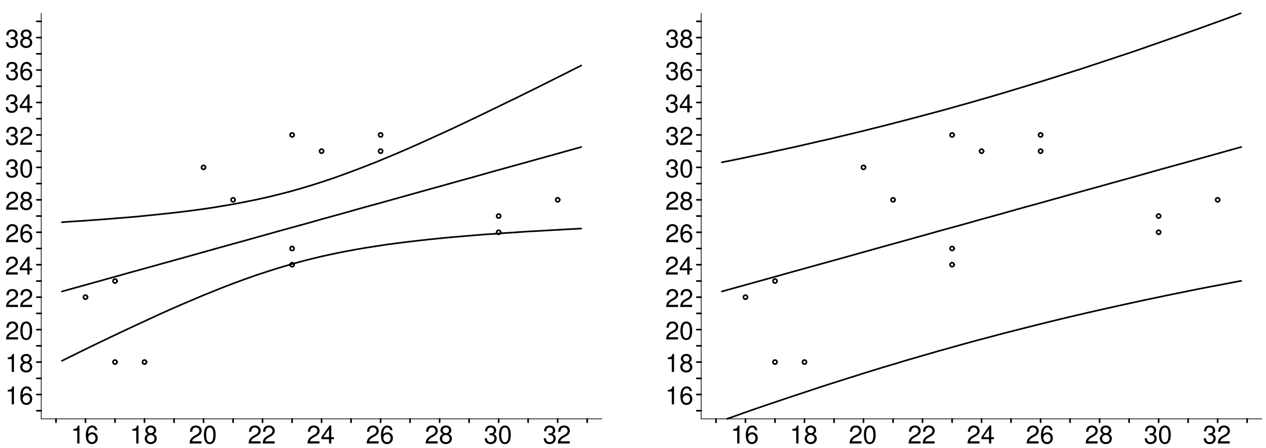


Figure 4.7–4: A 95% Confidence Band for $\mu(x)$ on Left, A 90% Prediction Band for Y on Right

4.7–6 Figure 4.7-6:(ab) shows the scatterplot with the least squares regression line.

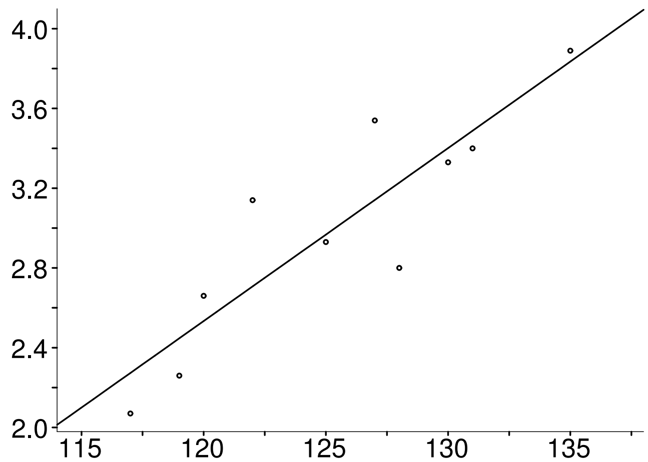


Figure 4.7–6: (ab) Scatterplot and Regression Line for Amount Bet (y) Versus Attendance (x)

Figure 4.7-6(cd) shows a 95% confidence band for $\mu(x)$ on the left and a 90% prediction band for Y on the right.

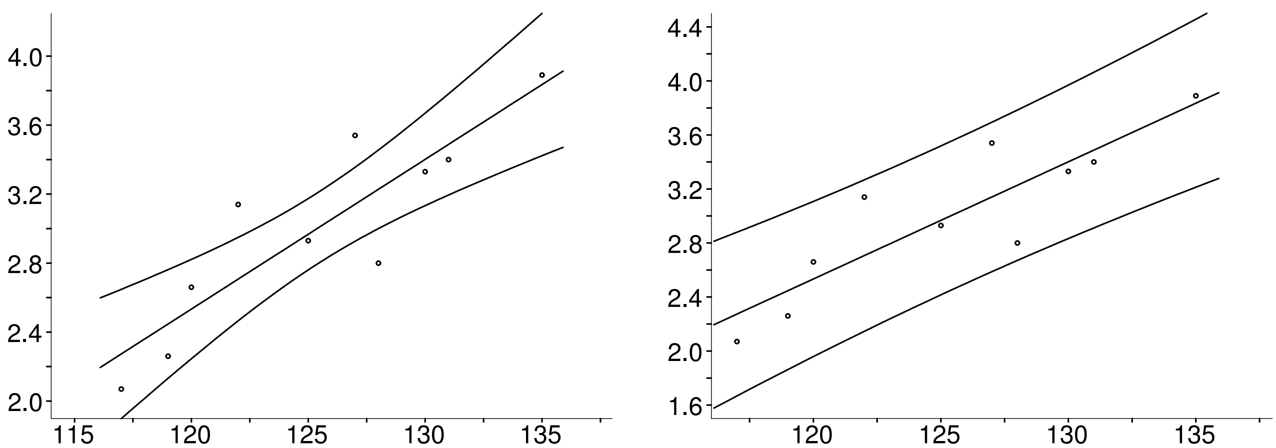


Figure 4.7–6: (cd) 95% Confidence Band for $\mu(x)$ on Left, A 90% Prediction Band for Y on Right

(d) To find a 90% prediction interval for the amount bet when the attendance is 12,000 we need the following information:

$$\widehat{\alpha} = \overline{y} = 3.002,$$
$$\overline{x} = 125.400,$$
$$\widehat{\beta} = 0.0868,$$
$$\widehat{\sigma^2} = 0.0639,$$
$$d = 0.3090.$$

The endpoints are

$$3.002 + 0.0868(120 - 125.400) \pm 0.3090(1.860$$

or

$$[1.959,\ 3.108].$$

4.8 One-Factor Analysis of Variance

4.8–2

Source	SS	DF	MS	F	p-value
Treatment	150	2	75	75	0.00006
Error	6	6	1		
Total	156	8			

4.8–4 Here is a Minitab solution.

```
One-way ANOVA: Observations versus Treatment

Analysis of Variance for Observations
Source     DF        SS        MS        F        P
Treatment   4    475.76    118.94    14.76    0.000
Error      20    161.20      8.06
Total      24    636.96
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  ------+---------+---------+---------+
15          5     9.800     3.347  (-----*----)
20          5    15.400     3.130              (----*----)
25          5    17.600     2.074                  (----*----)
30          5    21.600     2.608                          (----*----)
35          5    10.800     2.864    (-----*----)
                                   ------+---------+---------+---------+
Pooled StDev =    2.839                 10.0      15.0      20.0      25.0
```

Note that $14.76 > 4.43$ so H_0 is clearly rejected. Minitab listed the p-value as 0.

4.8–6 **(a)** An $\alpha = 0.05$ critical region is $F \geq 4.07$.

(b) Here is a Minitab solution.

```
One-way ANOVA: Observations versus Treatment

Analysis of Variance for Observations
Source     DF        SS        MS        F        P
Treatment   3      3215      1072     4.11    0.049
Error       8      2088       261
Total      11      5303
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  ---------+---------+---------+-------
1           3    242.00     22.07         (--------*-------)
2           3    271.33     15.01                     (--------*-------)
3           3    245.33     13.50           (-------*--------)
4           3    225.67     12.22   (-------*--------)
                                   ---------+---------+---------+-------
Pooled StDev =    16.16                   225       250       275
```

Since $4.11 > 4.07$ we would reject H_0.

(c) If $\alpha = 0.025$, the critical region is $F \geq 5.42$. But $4.11 < 5.42$ so we would fail to reject H_0.

(d) The p-value given by Minitab is p-value $= 0.049$.

4.8–8 **(a)** $$t = \frac{92.143 - 103.000}{\sqrt{\dfrac{6(69.139) + 6(57.669)}{12}\left(\dfrac{1}{7} + \dfrac{1}{7}\right)}} = -2.55 < -2.179, \text{ reject } H_0.$$

$$F = \frac{412.517}{63.4048} = 6.507 > 4.75, \text{ reject } H_0.$$

The F and the t tests give the same results since $t^2 = F$.

(b) $$F = \frac{86.3336}{114.8889} = 0.7515 < 3.55, \text{ do not reject } H_0.$$

4.8–10 **(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)

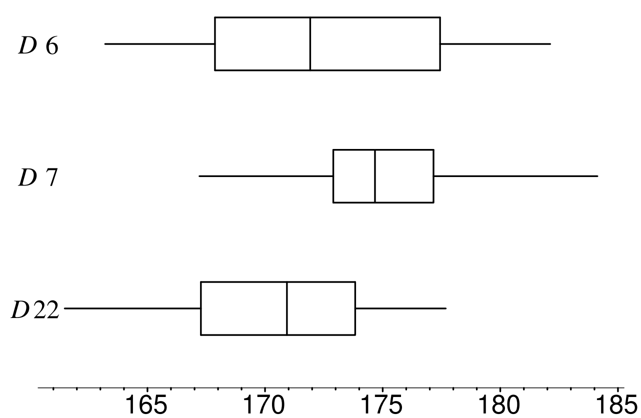


Figure 4.8–10: Box-and-whisker diagrams for resistances on three days

4.9 Distribution-Free Confidence and Tolerance Intervals

4.9–2 **(a)** $$P(Y_2 < \pi_{0.5} < Y_5) = \sum_{k=2}^{4} \binom{6}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{6-k} = 0.8906 - 0.1094 = 0.7812;$$

(b) $$P(Y_1 < \pi_{0.25} < Y_4) = \sum_{k=1}^{3} \binom{6}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{6-k} = 0.9624 - 0.1780 = 0.7844;$$

(c) $$\begin{aligned} P(Y_4 < \pi_{0.9} < Y_6) &= \sum_{k=4}^{5} \binom{6}{k} \left(\frac{9}{10}\right)^k \left(\frac{1}{10}\right)^{6-k} \\ &= \sum_{k=1}^{2} \binom{6}{k} \left(\frac{1}{10}\right)^k \left(\frac{9}{10}\right)^{6-k} \\ &= 0.9842 - 0.5314 = 0.4528. \end{aligned}$$

4.9–4 **(a)** $(y_4 = 80.28,\ y_{11} = 80.51)$ is a 94.26% confidence interval for m.

(b) $(y_6 = 80.32,\ y_{12} = 80.53)$;

$$\begin{aligned} P(Y_6 < \pi_{0.6} < Y_{12}) &= \sum_{k=6}^{11} \binom{14}{k}(0.6)^k(0.4)^{14-k} \\ &= \sum_{k=3}^{8} \binom{14}{k}(0.4)^k(0.6)^{14-k} \\ &= 0.9417 - 0.0398 = 0.9019. \end{aligned}$$

4.9–6 **(a)**

Stems	Leaves	Frequency	Depths
0	09 11 13 24 34 51 57 75 79 98	10	10
1	36 52 61 98	4	(4)
2	15 38 44 72 88	5	11
3	15 15 50	3	6
4	45	1	3
5		0	2
6	20	1	2
7	23	1	1

(b) $\widetilde{\pi}_{0.25} = 54, \widetilde{m} = 161, \widetilde{\pi}_{0.75} = 301.5$;

(c) **(i)** (13, 98), 89.66%; **(ii)** (79, 244), 89.22%; **(iii)** (238, 455), 89.66%.

(d) $[142.7, 267.1]$, 89.22%; $[141.2, 268.6]$, 90%; they are quite different because data are skewed (exponential) rather than normal.

4.9–8 **(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

$$\begin{aligned} P(Y_i < \pi_{0.5} < Y_{61-i}) &= P(i \le W \le 60 - i) \\ &\approx P\left(\frac{i - 0.5 - 30}{\sqrt{15}} \le Z \le \frac{60 - i + 0.5 - 30}{\sqrt{15}}\right). \end{aligned}$$

If

$$\frac{i - 30.5}{\sqrt{15}} = -1.96$$

then $i \approx 23$. So

$$P(Y_{23} < \pi_{0.5} < Y_{38}) = P(23 \le W \le 37) \approx 0.9472.$$

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

$$\begin{aligned} P(Y_i < \pi_{0.40} < Y_j) &= P(i \le W \le j-1) \\ &\approx P\left(\frac{i - 0.5 - 24}{\sqrt{14.4}} \le Z \le \frac{j - 1 + 0.5 - 24}{\sqrt{14.4}}\right). \end{aligned}$$

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 $P(18 \le W \le 31 - 1) = 0.9133$. So an approximate 91.33% confidence interval for $\pi_{0.4}$ is $(y_{18} = 6.95,\ y_{31} = 8.57)$.

4.9–10 We want to find the smallest n so that

$$\gamma = P[F(Y_n) - F(Y_1) \ge 0.90] \ge 0.95.$$

We have

$$\begin{aligned} \gamma &= \int_{0.90}^{1} \frac{\Gamma(n+1)}{\Gamma(n-1)\Gamma(2)} v^{n-2}(1-v)^{2-1}\,dv \\ &= 1 - \sum_{k=n-1}^{n} \binom{n}{k} (0.9)^k (0.1)^{n-k} \\ &= 1 - \frac{n!}{(n-1)!1!}(0.9)^{n-1}(0.1)^1 - \frac{n!}{n!0!}(0.9)^n(0.1)^0 \\ &= 1 - n(0.9)^{n-1}(0.1) - (0.9)^n \ \ge\ 0.95. \end{aligned}$$

Now

$$\begin{aligned} 1 - 45(0.9)^{44}(0.1) - (0.1)^{45} &= 0.947, \\ 1 - 46(0.9)^{45}(0.1) - (0.1)^{46} &= 0.952. \end{aligned}$$

Thus $n = 46$.

4.10 Chi-Square Goodness of Fit Tests

4.10–2

$$\begin{aligned} q_4 &= \frac{(224-232)^2}{232} + \frac{(119-116)^2}{116} + \frac{(130-116)^2}{116} + \frac{(48-58)^2}{58} + \frac{(59-58)^2}{58} \\ &= 3.784. \end{aligned}$$

The null hypothesis will not be rejected at any reasonable significance level. Note that $E(Q_4) = 4$ when H_0 is true.

4.10–4

$$\begin{aligned} q_3 &= \frac{(124-117)^2}{117} + \frac{(30-39)^2}{39} + \frac{(43-39)^2}{39} + \frac{(11-13)^2}{13} \\ &= 0.419 + 2.077 + 0.410 + 0.308 = 3.214 < 7.815 = \chi^2_{0.05}(3). \end{aligned}$$

Thus we do not reject the Mendelian theory with these data.

4.10–6

```
> read `C:\\Tanis-Hogg\\Maple Examples\\stat.m`:
  read `C:\\Tanis-Hogg\\Maple Examples\\ProbHistBdash.txt`:
  read `C:\\Tanis-Hogg\\Maple Examples\\HistogramFill.txt`:
```

```
read `C:\\Tanis-Hogg\\Maple Examples\\Data\\Chapter_04.txt`:
read `C:\\Tanis-Hogg\\Maple Examples\\Chapter_04.txt`:
X := Exercise_4_10_6G;
```

$$X := [[0,0],[1,6],[2,13],[3,30],[4,28],[5,8]]$$

```
m := 0:
for k from 2 to 6 do
  for j from 1 to X[k,2] do
    m := m + 1:
    XX[m] := X[k,1]:
  od:
od:
Y := [seq(XX[m], m = 1 .. 85)];
```

$Y := [1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,$
$3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4,$
$4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5]$

```
> xbar := Mean(X);
```

$$xbar := \frac{274}{85}$$

```
> phat := xbar/5;
  phat := evalf(%);
```

$$phat := \frac{274}{425}$$

$$phat := .6447058824$$

Now find the expected values of each possible outcome.

```
> for k from 0 to 5 do
    Exp[k] := evalf(85*BinomialPDF(5,phat,k));
  od;
```

$$Exp_0 := .4812370010$$

$$Exp_1 := 4.366190009$$

$$Exp_2 := 15.84551076$$

$$Exp_3 := 28.75278113$$

$$Exp_4 := 26.08696038$$

$$Exp_5 := 9.467320723$$

Here are the observed values.

```
> for k from 1 to 6 do
    Obs[k-1] := X[k,2];
  od;
```

$$Obs_0 := 0$$

$$Obs_1 := 6$$

$$Obs_2 := 13$$

$$Obs_3 := 30$$

$$Obs_4 := 28$$

$$Obs_5 := 8$$

Now find the value of the chi-square goodness of fit statistic and the p-value.

```
> k := 'k':
  q := sum((Obs[k] - Exp[k])^2/Exp[k], k = 0 .. 5);
  pvalue := 1 - ChisquareCDF(4,q);
```

$$q := 2.025401328$$

$$pvalue := .7310866940$$

We should really group the first two classes.

```
> Exp[1] := Exp[0] + Exp[1];
  Obs[1] := Obs[0] + Obs[1];
```

$$Exp_1 := 4.847427010$$

$$Obs_1 := 6$$

```
> qg := sum((Obs[k] - Exp[k])^2/Exp[k], k = 1 .. 5);
  pvalue := 1 - ChisquareCDF(3,qg);
```

$$qg := 1.206846824$$

$$pvalue := .7513626336$$

We also give a graphical comparison.

```
> xtics := [$0 .. 5]:
  ytics := [seq(0.05*k, k = 1 .. 7)]:
  P0 := plot([[0,0],[0,0]], x = -0.7 .. 5.7, xtickmarks=xtics,
  ytickmarks=ytics, labels=[``,``]):
  P1 := ProbHistBdash(BinomialPDF(5,phat,x), 0 .. 5):
  P2 := HistogramFill(Y, -0.5 .. 5.5, 6):
  display(\{P0,P1, P2\});
```

Here are the endpoints for a confidence interval for p calculated by *Maple*.

```
> [evalf(phat - 1.96*sqrt(phat*(1-phat)/425)),
  evalf(phat + 1.96*sqrt(phat*(1-phat)/425))];
```

$$[.5992032595, .6902085053]$$

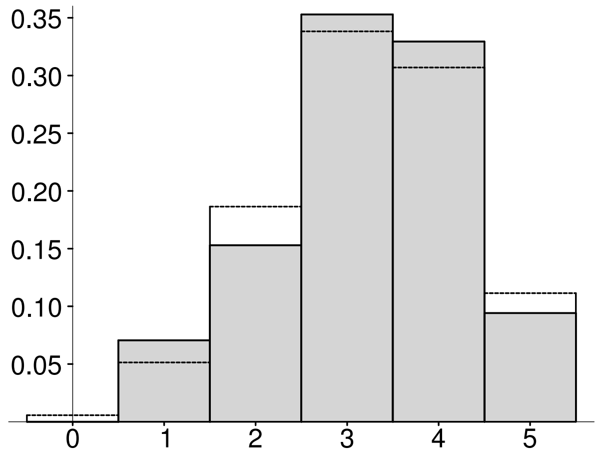


Figure 4.10–6: $b(5, 0.645)$ Probability Histogram (dashed), Relative Frequency Histogram (Shaded)

4.10–8 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.

4.11 Contingency Tables

4.11–2 In the combined sample of 30 observations, the lower third includes those with scores of 55 or lower, the middle third have scores from 56 through 75.5, and the higher third are those with scores of 75.5 and above.

	low	middle	high	Totals
Class U	7 (5)	4 (5)	4 (5)	15
Class V	3 (5)	6 (5)	6 (5)	15
Totals	10	10	10	30

Thus

$$q = 0.8 + 0.2 + 0.2 + 0.8 + 0.2 + 0.2 = 2.4.$$

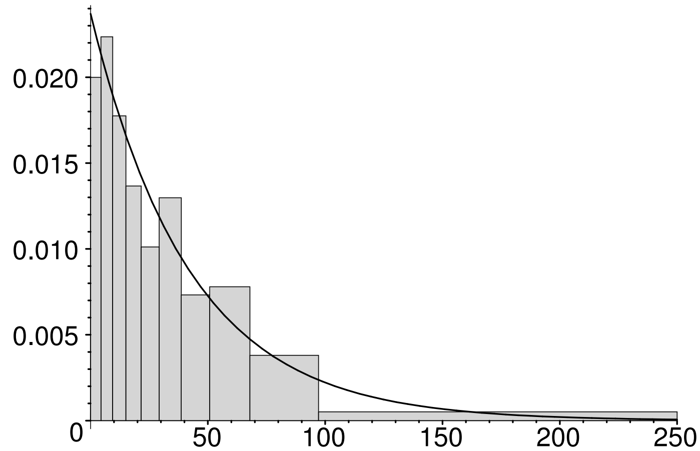


Figure 4.10–8: Exponential p.d.f. , $\widehat{\theta} = 42.2$, and relative frequency histogram (shaded)

Since

$$q = 2.4 < 5.991 = \chi^2_{0.05}(2),$$

we do not reject the equality of these two distributions. (p-value = 0.3012.)

4.11–4

Gender	Favor	Oppose	No Opinion	Totals
Male	262.00 (279.50)	231.00 (216.07)	10.00 (7.43)	503
Female	302.00 (284.50)	205.00 (219.93)	5.00 (7.57)	512
Totals	564	436	15	1015

Thus

$$q = 1.096 + 1.032 + 0.886 + 1.076 + 1.014 + 0.871 = 5.975.$$

Since $5.975 < \chi^2_{0.05}(2) = 5.991$, we do not reject the null hypothesis; however, p-value ≈ 0.05.

4.11–6 Here is a Minitab solution.

```
Expected counts are printed below observed counts

    Age      <180  180-210     >210    Totals

   < 50         5       11        9        25
             4.50     7.00    13.50

   >=50         4        3       18        25
             4.50     7.00    13.50

   Totals       9       14       27        50

Chi-Sq =  0.056 +  2.286 +  1.500 +
```

```
          0.056 +  2.286 +  1.500 = 7.683
DF = 2, P-Value = 0.021
2 cells with expected counts less than 5.0
```

Note that $q = 7.683 < 9.210 = \chi^2_{0.01}$ so we fail to reject H_0.

4.11–8 Minitab solutions are given.

(a)

```
Expected counts are printed below observed counts

Age          Newspaper        TV    Radio    Total

Under 35            30        68       10      108
                 47.47     47.72    12.81

35-54               61        79       20      160
                 70.33     70.70    18.98

Over 54             98        43       21      162
                 71.20     71.58    19.21

Total              189       190       51      430

Chi-Sq =  6.429 +  8.618 +  0.616 +
          1.237 +  0.975 +  0.055 +
         10.083 + 11.412 +  0.166 = 39.591
DF = 4, P-Value = 0.000
```

Since $39.591 > 9.488$, reject hypothesis of independence.

(b)

```
Expected counts are printed below observed counts

Gender    Newspaper        TV    Radio    Total

Male             92       108       19      219
              96.48     96.48    26.03

Female           97        81       32      210
              92.52     92.52    24.97

Total           189       189       51      429

Chi-Sq =  0.208 +  1.375 +  1.901 +
          0.217 +  1.434 +  1.982 = 7.117
DF = 2, P-Value = 0.028
```

Since $7.117 > 5.991$, reject hypothesis of independence.

(c)

```
Expected counts are printed below observed counts

Education      Newspaper       TV    Radio    Total

Grade School          45       22        6       73
                   32.22    32.39     8.40

High School           94      115       30      239
                  105.47   106.04    27.49

College               49       52       13      114
                   50.31    50.58    13.11

Total                188      189       49      426

Chi-Sq =  5.073 +  3.331 +  0.684 +
          1.248 +  0.758 +  0.229 +
          0.034 +  0.040 +  0.001 = 11.399
DF = 4, P-Value = 0.022
```

Since $11.399 > 9.488$, reject hypothesis of independence.

Chapter 5

Computer Oriented Techniques

5.1 Computation of Statistics

5.1–2 To find these probabilities you use

Calc → Probability Distributions → Poisson

and input the necessary information.

(a)
```
Cumulative Distribution Function
Poisson with mu = 17.6000
         x    P( X <= x )
     18.00        0.5996
```

(b)
```
Probability Density Function
Poisson with mu = 17.6000
         x     P( X = x )
     17.00        0.0953
```

5.1–4 In Minitab we can find characteristics of these data by putting these data in two columns with a column header x for first set of data and a column header of y for the second set of data.

(a) Input

Stat → Basic Statistics → Display Descriptive Statistics

and select x and y.

```
Descriptive Statistics: x, y

Variable    N     Mean   Median   TrMean    StDev  SE Mean
x         150    4.473    5.000    4.470    3.099    0.253
y          96    3.938    4.000    3.930    2.639    0.269

Variable       Minimum    Maximum         Q1         Q3
x                0.000      9.000      2.000      7.000
```

(b) Now select

Stat → Tables → Tally

and select x and y. The following tallies are given.

```
Tally for Discrete Variables: x, y

  x  Count      y  Count
  0     21      0      8
  1     15      1     13
  2     16      2     15
  3     13      3     10
  4      6      4     12
  5     15      5     10
  6     13      6      4
  7     17      7      9
  8     17      8     15
  9     17     N=     96
 N=    150
```

(c) You can obtain histograms by inputting

Graph → Histogram

and selecting x and y.

5.1–6 To find the quartiles for the Weibull distribution using Minitab, select

Calc → Probability Distributions → Weibull

and then select "Inverse cumulative probability" and input the two parameters, 3 and 4. Now input the constant 0.25. Repeat this inputting the constants 0.5 and 0.75. Minitab will then give you the following:

```
Inverse Cumulative Distribution Function

Weibull with first shape parameter = 3.00000 and second = 4.00000

P( X <= x )          x
    0.2500      2.6406

P( X <= x )          x
    0.5000      3.5400

P( X <= x )          x
    0.7500      4.4601
```

5.1–8 To find probabilities related to the χ^2 distribution, select

Calc → Probability Distributions → Chi-Square

and input the necessary information.

```
Cumulative Distribution Function

Chi-Square with 23 DF

         x    P( X <= x )
   23.0000         0.5392
```

```
Inverse Cumulative Distribution Function

Chi-Square with 23 DF

P( X <= x )          x
    0.0700       13.8892

P( X <= x )          x
    0.9200       33.0616
```

5.2 Computer Algebra Systems

5.2–2 **(a)** X is $N(0,1)$, Y^2+Z^2 is $\chi^2(2)$, and the numerator and the denominator are independent. So T is a $N(0,1)$ random variable divided by the square root of an independent χ^2 random variable divided by its degrees of freedom, yielding a T random variable with 2 degrees of freedom.

(b) Note that the numerator and denominator are not independent in this case.
The joint p.d.f. of X and Y is

```
> g := exp(-x^2/2)/sqrt(2*Pi)*exp(-y^2/2)/sqrt(2*Pi);
```

$$g := \frac{1}{2}\,\frac{e^{(-x^2/2)}\,e^{(-y^2/2)}}{\pi}$$

```
> interface(showassumed=0):
  assume(-sqrt(2) < u):
  additionally(u < sqrt(2)):
```

First define the distribution function for positive y values.

```
> F := int(Int(g, x = -u*y/sqrt(2-u^2) .. u*y/sqrt(2-u^2)),
  y = 0 .. infinity);
```

$$F := \int_0^{\infty}\int_{-\frac{u\,y}{\sqrt{2-u^2}}}^{\frac{u\,y}{\sqrt{2-u^2}}}\frac{1}{2}\,\frac{e^{(-x^2/2)}\,e^{(-y^2/2)}}{\pi}\,dx\,dy$$

```
> f := simplify(diff(F, u));
```

$$f := \frac{1}{\sqrt{2-u^2}\,\pi}$$

Now define the distribution function for negative y values.

```
> F := int(Int(g, x = u*y/sqrt(2-u^2) .. -u*y/sqrt(2-u^2)),
  y = -infinity .. 0);
```

$$F := \int_{-\infty}^{0}\int_{\frac{u\,y}{\sqrt{2-u^2}}}^{-\frac{u\,y}{\sqrt{2-u^2}}}\frac{1}{2}\,\frac{e^{(-x^2/2)}\,e^{(-y^2/2)}}{\pi}\,dx\,dy$$

```
> f := simplify(diff(F, u));
```

$$f := \frac{1}{\sqrt{2-u^2}\,\pi}$$

Note that we obtain the same p.d.f. in both cases. It is interesting to note that the mean and variance of U are 0 and 1, respectively. We first show that f is indeed a p.d.f.

```
> int(f, u = -sqrt(2) .. sqrt(2));
```

$$1$$

```
> mu := int(u*f, u = -sqrt(2) .. sqrt(2));
```

$$\mu := 0$$

```
> var := int(u^2*f, u = -sqrt(2) .. sqrt(2));
```

$$var := 1$$

(c) Now find the distribution function and p.d.f. of $V = U\hat{}2/2$. For $0 < v < 1$,

$$G(v) = P(V \leq v) = P(U^2/2 \leq v) = P(U^2 \leq 2v) = P(-\sqrt{2v} \leq U \leq \sqrt{2v}).$$

Using *Maple*,

```
> G := Int(1/Pi/sqrt(2 - u^2), u = -sqrt(2*v) .. sqrt(2*v));
```

$$G := \int_{-\sqrt{2}\,\sqrt{v}}^{\sqrt{2}\,\sqrt{v}} \frac{1}{\sqrt{2-u^2}\,\pi}\,du$$

```
g := simplify(diff(G, v));
```

$$> g := \frac{1}{\pi\,\sqrt{v}\,\sqrt{1-v}}$$

or

```
> G := int(1/Pi/sqrt(2 - u^2), u = -sqrt(2*v) .. sqrt(2*v));
```

$$G := \frac{2\arcsin(\sqrt{v})}{\pi}$$

```
> g := simplify(diff(G, v));
```

$$g := \frac{1}{\pi\,\sqrt{v}\,\sqrt{1-v}}$$

Note that the beta p.d.f. with $\alpha = 1/2$ and $\beta = 1/2$ is

```
> GAMMA(1/2 + 1/2)/GAMMA(1/2)/GAMMA(1/2)*v^(1/2 - 1)*(1 - v)^(1/2 -
1);
```

$$\frac{1}{\pi\,\sqrt{v}\,\sqrt{1-v}}$$

Thus we see that $V = U\hat{}2/2$ has a beta distribution with $\alpha = 1/2$ and $\beta = 1/2$.

5.2–4 **(a)** Let X, Y be a random sample of size 2 from a Cauchy distribution. The joint p.d.f. of X and Y is

```
> g := 1/Pi/(1 + x^2)*1/Pi/(1 + y^2);
```

$$g := \frac{1}{\pi^2\,(1+x^2)\,(y^2+1)}$$

(b) Let $W = (X + Y)/2$. The distribution function of W is

```
> F := int(Int(g, y = 0 .. 2*w - x), x = -infinity .. infinity);
```

$$F := \int_{-\infty}^{\infty}\int_{0}^{2\,w-x} \frac{1}{\pi^2\,(1+x^2)\,(y^2+1)}\,dy\,dx$$

(c) The p.d.f. of W is the derivative of F with respect to w.

```
> f := simplify(diff(F, w));
```

$$f := \frac{-I}{\pi\,(-1 + w\,I)\,(-w + I)}$$

```
> f := factor(evalc(f));
```

$$f := \frac{1}{\pi\,(w^2 + 1)}$$

Note that the p.d.f. of W is the the p.d.f. of the Cauchy distribution.

(d) It follows that the distribution of the sample mean for any size sample from the Cauchy distribution is Cauchy.

5.2–6 We first note that

$$P(Y \le y) = P\left(\frac{1}{1 + e^{-X}} \le y\right) = P[X \le -\ln(1/y - 1)].$$

Using *Maple* we have

```
> G := Int(exp(-u)/(1 + exp(-u))^2, u = -infinity .. -ln(1/y - 1));
```

$$G := \int_{-\infty}^{-\ln(\frac{1}{y}-1)} \frac{e^{(-u)}}{(1 + e^{(-u)})^2}\,du$$

```
> g := diff(G, y);
```

$$g := 1$$

Note that $0 < y < 1$ so that the distribution of Y is $U(0, 1)$.

5.2–8 We first note that

$$G(y) = P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}).$$

Using *Maple* we have

```
> G := Int(1/Pi/(1 + x^2), x = -sqrt(y) .. sqrt(y));
```

$$G := \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\pi\,(1 + x^2)}\,dx$$

```
> g := diff(G, y);
```

$$g := \frac{1}{\sqrt{y}\,\pi\,(1 + y)}$$

We must note that $0 \le y < \infty$.

5.3 Simulation

5.3–2 This solution is given in the hint.

5.3–4
```
> read `C:\\Tanis-Hogg\\Maple Examples\\stat.m`:
  read `C:\\Tanis-Hogg\\Maple Examples\\HistogramFill.txt`:
> f := 2*x;
```

$$f := 2\,x$$

```
> mu := int(x*f, x = 0 .. 1);
  evalf(%);
```

$$\mu := \frac{2}{3}$$

$$.6666666667$$

```
> var := int((x - mu)^2*f, x = 0 .. 1);
  evalf(%);
  sigma := evalf(sqrt(var));
```

$$var := \frac{1}{18}$$

$$.05555555556$$

$$\sigma := .2357022604$$

```
> for j from 1 to 1000 do
    for k from 1 to 18 do
      XX[k] := evalf(sqrt(rng())):
    od:
    X := [seq(XX[k], k = 1 .. 18)]:
    Xbar[j] := Mean(X):
  od:
  Xbars := [seq(Xbar[j], j = 1 .. 1000)]:

> Mean(Xbars);
```

$$.6687477451$$

```
> Variance(Xbars);
```

$$.003402774975$$

```
> evalf(var/18);
```

$$.003086419753$$

```
> Min(Xbars), Max(Xbars);
```

$$.4379161342,\ .8285015133$$

```
> xtics := [seq(0.4 + k*0.05, k = 0 .. 10)]:
  ytics := [seq(0.5*k, k = 1 .. 14)]:
> P1 := plot(NormalPDF(mu, var/18, x), x = 0.4 .. 0.9, y = 0 .. 7.4,
  color=black, thickness=2, xtickmarks=xtics, ytickmarks=ytics,
  labels=[``,``]):
  P2 := HistogramFill(Xbars, 0.40 .. 0.90, 20):
  display({P1, P2});
```

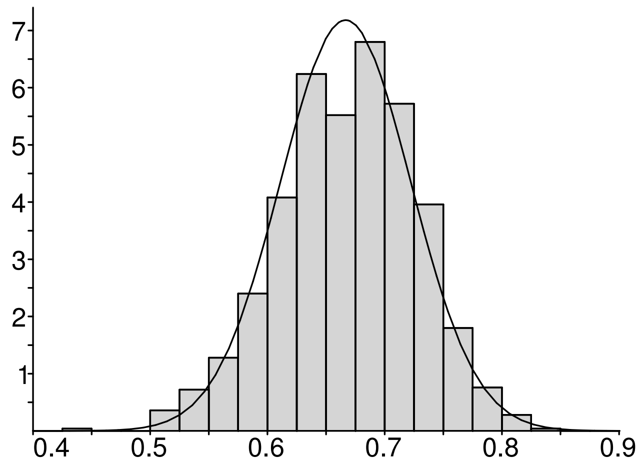


Figure 5.3–4: A Histogram of 1,000 $\overline{x}$s and the $N(2/3, 1/18/18)$ p.d.f.

5.3–6 **(a)**
```
> read `C:\\Tanis-Hogg\\Maple Examples\\stat.m`:
> f := CauchyPDF(x);
```

$$f := \frac{1}{\pi\,(1 + x^2)}$$

```
> for k from 1 to 9 do
    w := Pi*rng() - Pi/2:
    XX[k] := evalf(tan(w)):
  od:
  X := [seq(XX[k], k = 1 .. 9)];
```

$X := [-2.069229847, 0.6070884640, 2.266579631, -8.875162342, -1.422504696, 2.415775942, -0.02463755435, 0.3228193440, 31.41479916]$

(b)
```
> xbar := Mean(X);
```

$$xbar = 2.737280900$$

```
> X := sort(X);
```

$X = [-8.875162342, -2.069229847, -1.422504696, -0.02463755435, 0.3228193440, 0.6070884640, 2.266579631, 2.415775942, 31.41479916]$

```
> Median(X);
```

$$0.3228193440$$

Generally the median will be closer to 0. The mean is influenced by outliers.

(c)
```
> numgt1 := 0:  # number greater than 1
  numgt5 := 0:  # number greater than 5
  numgt10 := 0: # number greater than 10
  for k from 1 to 1000 do
    w := Pi*rng() - Pi/2:
    X := evalf(tan(w)):
      if X > 1 then numgt1 := numgt1 + 1 fi:
      if X > 5 then numgt5 := numgt5 + 1 fi:
      if X > 10 then numgt10 := numgt10 + 1 fi:
    od:
```

```
> numgt1;
  evalf(numgt1/1000);
  evalf(int(f, x = 1 .. infinity));
```

$$247$$

$$0.2470000000$$

$$0.2500000000$$

```
> numgt5;
  evalf(numgt5/1000);
  evalf(int(f, x = 5 .. infinity));
```

$$62$$

$$0.06200000000$$

$$0.06283295825$$

```
> numgt10;
  evalf(numgt10/1000);
  evalf(int(f, x = 10 .. infinity));
```

$$33$$

$$0.03300000000$$

$$0.03172551759$$

5.3–8 **(a)**
```
> read `C:\\Tanis-Hogg\\Maple Examples\\stat.m`:
  read `C:\\Tanis-Hogg\\Maple Examples\\HistogramFill.txt`:
  read `C:\\Tanis-Hogg\\Maple Examples\\ProbHistBdash.txt`:
> Craps();
```

$$[1, 7]$$

```
> Craps();
```

$$[0, 10, 8, 9, 4, 6, 7]$$

```
> for k from 1 to 1000 do
    Y := Craps():
    NN[k] := nops(Y) - 1:
  od:
  N := [seq(NN[k], k = 1 .. 1000)]:
  nbar := evalf(Mean(N));
  svar := evalf(Variance(N));
```

$$nbar := 3.500000000$$

$$svar := 9.917917918$$

(b) We now show that the theoretical answers given in the exercise are correct. Enter the p.d.f. (without the special case, $f(1)$.

```
> f := 2*((27/36)^(x-2)*(3/36)*(9/36) +
  (26/36)^(x-2)*(4/36)*(10/36) + (25/36)^(x-2)*(5/36)*(11/36));
```

$$f := \frac{1}{24}\,(\frac{3}{4})^{(x-2)} + \frac{5}{81}\,(\frac{13}{18})^{(x-2)} + \frac{55}{648}\,(\frac{25}{36})^{(x-2)}$$

Now calculate the mean and variance directly using Maple's sum command.

```
> mu := 12/36 + sum(x*f, x = 2 .. infinity);
  evalf(%);
  var := 12/36*(1 - mu)^2 + sum((x - mu)^2*f, x = 2 .. infinity);
  evalf(%);
```

$$\mu := \frac{557}{165}$$

$$3.375757576$$

$$var := \frac{245672}{27225}$$

$$9.023764922$$

(c) We first find the minimum and maximum of the observations of N and then construct a histogram.

```
> Min(N),Max(N);
```

$$1,\ 20$$

```
> f := x-> 2*((27/36)^(x-2)*(3/36)*(9/36) +
  (26/36)^(x-2)*(4/36)*(10/36) + (25/36)^(x-2)*(5/36)*(11/36));
```

$$f := x \to \frac{1}{24}\,(\frac{3}{4})^{(x-2)} + \frac{5}{81}\,(\frac{13}{18})^{(x-2)} + \frac{55}{648}\,(\frac{25}{36})^{(x-2)}$$

```
> probs := [12/36,seq(f(x), x = 2 .. 20)]:
  domain := [seq(k, k = 1 .. 20)]:
  pdf := zip( (x,y) -> (x,y), domain, probs):
> ytics := [seq(0.02*k, k = 1 .. 17)]:
  P1 := ProbHistBdash(pdf):
  P2 := HistogramFill(N, 0.5 .. 20.5, 20):
  P3 := plot([[0,0],[0,0]], x = 0 .. 20.7, y = 0 .. 0.345,
  xtickmarks=domain, ytickmarks=ytics, labels=[``,``]):
  display({P1, P2,P3});
```

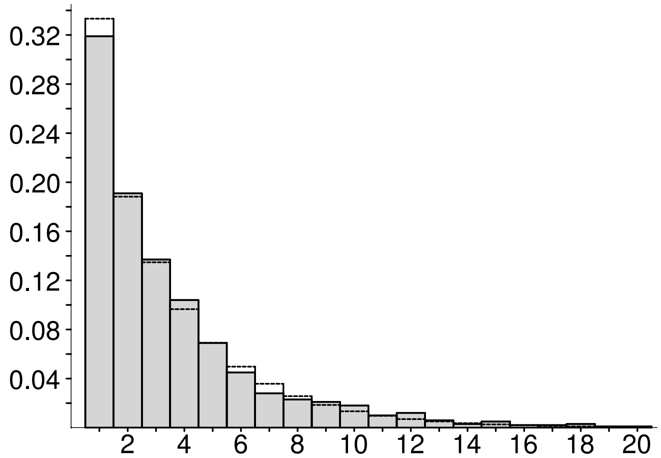


Figure 5.3–8: A Histogram of 1,000 Observations of N and the p.d.f. of N

5.3–10 (f)

```
> with(plots):
  read `C:\\Tanis-Hogg\\Maple Examples\\stat.m`:
  read `C:\\Tanis-Hogg\\Maple Examples\\BoxPlotB.txt`:
> with(combinat, randperm):
> for jj from 1 to 7 do
  N1[jj] := jj*50: #Number marked and released into population
  n := 400 - N1[jj]: #Number recaptured in the second catch
  N := 1000: #Size of population
  # Y[k] is the number of marked that are recaptured
  # Nhat = N1[jj] * n/Y[k] is the mle of N on the kth repetition
  m := 200: #Number of repetitions
  pop := [1$i=1..N1[jj], 0$j=1..N-N1[jj]]:
  randomize():
  for k from 1 to m do
    perm := randperm(pop):
    Y[k] := sum(perm[j], j = 1 .. n):
    popest[k] := floor(N1[jj]*n/Y[k]):
  od:
  Nhat[jj] := [seq(popest[k], k = 1 .. m)]:
  Nbar[jj] := evalf(Mean(Nhat[jj]));
  s[jj] := evalf(StDev(Nhat[jj]));
  M[jj] := Median(Nhat[jj]);
  od:
```

Here are the respective values of $N1$, the means, standard deviations, and medians.

```
> N1s := [seq(N1[jj], jj = 1 .. 7)];
  Nbars := [seq(Nbar[jj], jj=1 .. 7)];
  sds := [seq(s[jj], jj=1 .. 7)];
  Medians := [seq(evalf(M[jj]), jj=1 .. 7)];
```

$$N1s := [50, 100, 150, 200, 250, 300, 350]$$

$$Nbars := [1044.855000, 1018.665000, 1023.505000, 1007.090000, 1032.090000, 1038.540000, 1074.435000]$$

$$sds := [204.0255772, 156.1866966, 135.3647964, 131.1009121, 145.4538967, 168.1731532, 247.4283527]$$

$$Medians := [1029., 1000., 1013., 1000., 1013., 1000., 1029.]$$

```
> xtics := [seq(550 + k*50, k = 0 .. 35)]:
  ytics := [0,10]:
> P1 := plot([[550,0],[550,0]], x = 550 .. 2250, y = 0 .. 7,
  xtickmarks=xtics, ytickmarks=ytics, labels=[‘‘,‘‘]):
  P2 := BoxPlot(Nhat[1], Nhat[2],Nhat[3], Nhat[4], Nhat[5], Nhat[6], Nhat[7]):
  txt1 := textplot([550, 0.6, ‘50‘], font=[TIMES,ROMAN,12]):
  txt2 := textplot([550, 1.6, ‘100‘], font=[TIMES,ROMAN,12]):
  txt3 := textplot([550, 2.6, ‘150‘], font=[TIMES,ROMAN,12]):
  txt4 := textplot([550, 3.6, ‘200‘], font=[TIMES,ROMAN,12]):
  txt5 := textplot([550, 4.6, ‘250‘], font=[TIMES,ROMAN,12]):
  txt6 := textplot([550, 5.6, ‘300‘], font=[TIMES,ROMAN,12]):
  txt7 := textplot([550, 6.6, ‘350‘], font=[TIMES,ROMAN,12]):
  display({P1, P2, txt1,txt2,txt3,txt4,txt5,txt6,txt7});
```

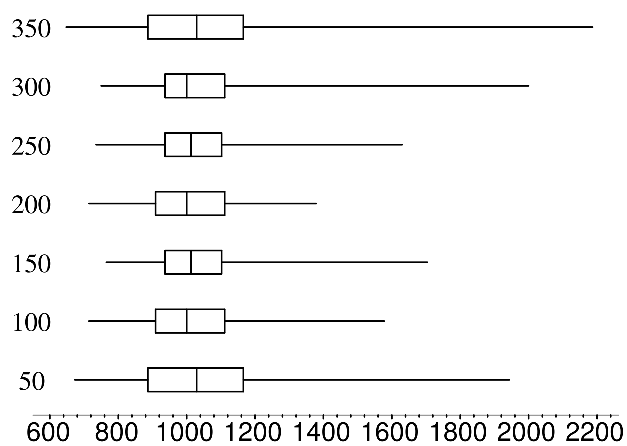


Figure 5.3–10: A Boxplot Comparison of the Different Combinations

Note that the medians are all close to 1000 but the standard deviations vary with $n = 200$ having the smallest standard deviation in this particular simulation.

5.3–12 **(a)**

```
> read ‘C:\\Tanis-Hogg\\Maple Examples\\stat.m‘:
  read ‘C:\\Tanis-Hogg\\Maple Examples\\HistogramFill.txt‘:
  read ‘C:\\Tanis-Hogg\\Maple Examples\\ScatPlotCirc.txt‘:
> randomize():
> for k from 1 to 500 do
    X := NormalS(1,1,7):
```

```
  TT[k] := evalf((Mean(X) - 1)/(StDev(X)/sqrt(7))):
od:
T := [seq(TT[k], k = 1 .. 500)]:
Min(T),Max(T);
tbar := Mean(T);
mu := 0;
svar := Variance(T);
var := 6/4;
```

$$-4.675793531,\ 4.590200174$$

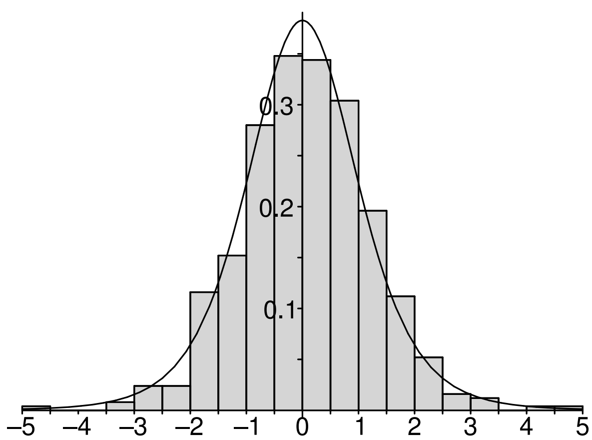
$$tbar := .07803562026$$

$$\mu := 0$$

$$svar := 1.372265305$$

$$var := 3/2$$

```
> xtics := [seq(-5 + 0.5*k, k = 0 .. 20)]:
  ytics := [seq(0.05*k, k = 1 .. 9)]:
> P1 := plot(TPDF(6,x), x = -5 .. 5, color=black, thickness=2,
  xtickmarks=xtics, ytickmarks=ytics, labels=[``,``]):
  P2 := HistogramFill(T, -5 .. 5, 20):
  display({P1, P2});
> quantiles := [seq(TP(6, k/501), k = 1 .. 500)]:
> T := sort(T):
> xtics := [seq(-5 + k*.5, k = 0 .. 20)]:
  ytics := xtics:
> P1 := plot([[-5,-5],[5,5]], x = -5 .. 5, y = -5 .. 5, color=black,
  thickness=2, xtickmarks=xtics, ytickmarks=ytics, labels=[``,``]):
  P2 := ScatPlotCirc(T, quantiles):
  display({P1, P2});
```



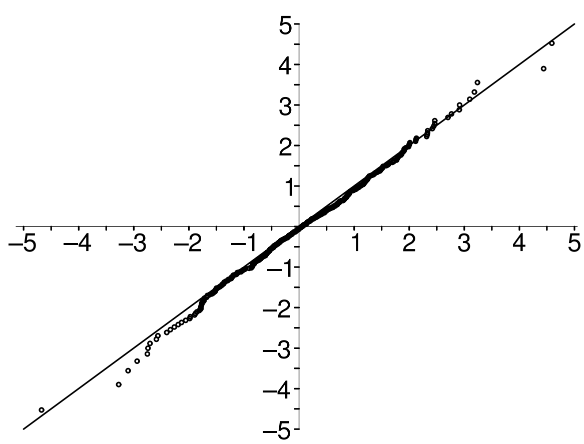


Figure 5.3–12: **(a)** Histogram and q-q Plot When Sampling from $N(1, 1)$ Distribution

(b) Sample from an exponential distribution with mean 1 and look at the values of T.

```
> for k from 1 to 500 do
    X := ExponentialS(1,7):
```

```
    TT[k] := evalf((Mean(X) - 1)/(StDev(X)/sqrt(7))):
  od:
  T := [seq(TT[k], k = 1 .. 500)]:
  Min(T),Max(T);
  tbar := Mean(T);
  mu := 0;
  svar := Variance(T);
  var := 6/4;
```

$$-9.827071016,\ 5.197207278$$

$$tbar := -.5360861658$$

$$\mu := 0$$

$$svar := 2.698085637$$

$$var := 3/2$$

```
> xtics := [seq(-7 + 0.5*k, k = 0 .. 28)]:
  ytics := [seq(0.05*k, k = 1 .. 9)]:
> P1 := plot(TPDF(6,x), x = -7 .. 7, color=black, thickness=2,
  xtickmarks=xtics, ytickmarks=ytics, labels=[``,``]):
  P2 := HistogramFill(T, -7 .. 7, 28):
  display({P1, P2});
> quantiles := [seq(TP(6, k/501), k = 1 .. 500)]:
  T := sort(T):
> xtics := [seq(-10 + k, k = 0 .. 18)]:
  ytics := xtics:
> P1 := plot([[-7,-7],[7,7]], x = -7 .. 7, y = -7 .. 7, color=black,
  thickness=2, xtickmarks=xtics, ytickmarks=ytics, labels=[``,``]):
  P2 := ScatPlotCirc(T, quantiles):
  display({P1, P2});
```

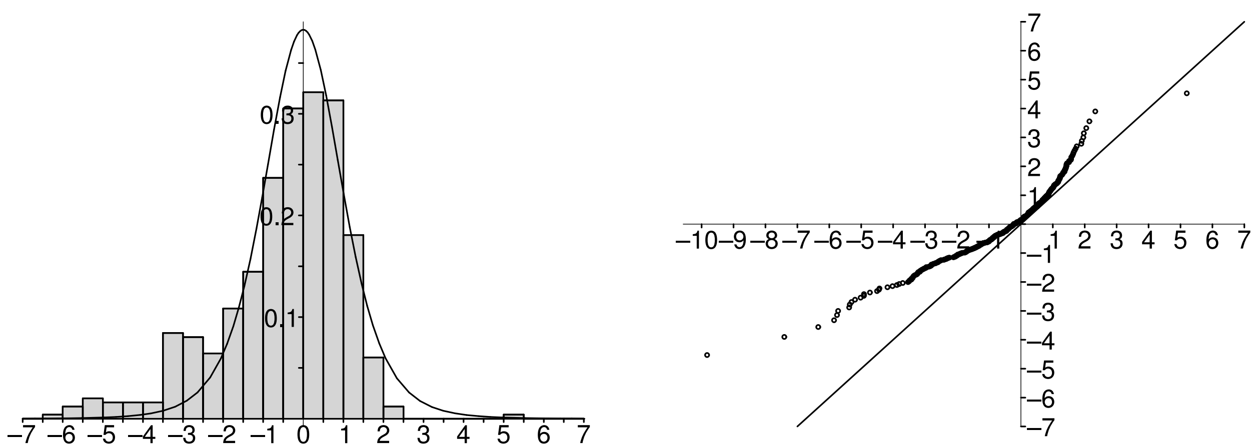


Figure 5.3–12: **(b)** Histogram and *q*-*q* Plot When Sampling from the Exponential Distribution

(c) Sample from a Cauchy distribution. There is a procedure for simulating a sample of size 7 from a Cauchy distribution.

```
>  CauchyS(7);
```

$$[-.03111542788, .5863792198, -.8740796048, -1.782216018, -.9964616806, -29.80581525, -1.875688958]$$

```
> for k from 1 to 500 do
    X := CauchyS(7):
    TT[k] := evalf(Mean(X)/(StDev(X)/sqrt(7))):
  od:
  T := [seq(TT[k], k = 1 .. 500)]:
  Min(T),Max(T);
  tbar := Mean(T);
  mu := 0;
  svar := Variance(T);
  var := 6/4;
```

$$-3.301465693, 3.431873453$$

$$tbar := .02355344944$$

$$\mu := 0$$

$$svar := 1.195258003$$

$$var := 3/2$$

```
> P1 := plot(TPDF(6,x), x = -4 .. 4, color=black, thickness=2,
  xtickmarks=xtics, ytickmarks=ytics, labels=[``,``]):
  P2 := HistogramFill(T, -4 .. 4, 16):
  display({P1, P2});
> T := sort(T):
  P1 := plot([[-4,-4],[4,4]], x = -4 .. 4, y = -4 .. 4, color=black,
  thickness=2, xtickmarks=xtics, ytickmarks=ytics, labels=[``,``]):
  P2 := ScatPlotCirc(T, quantiles):
  display({P1, P2});
```

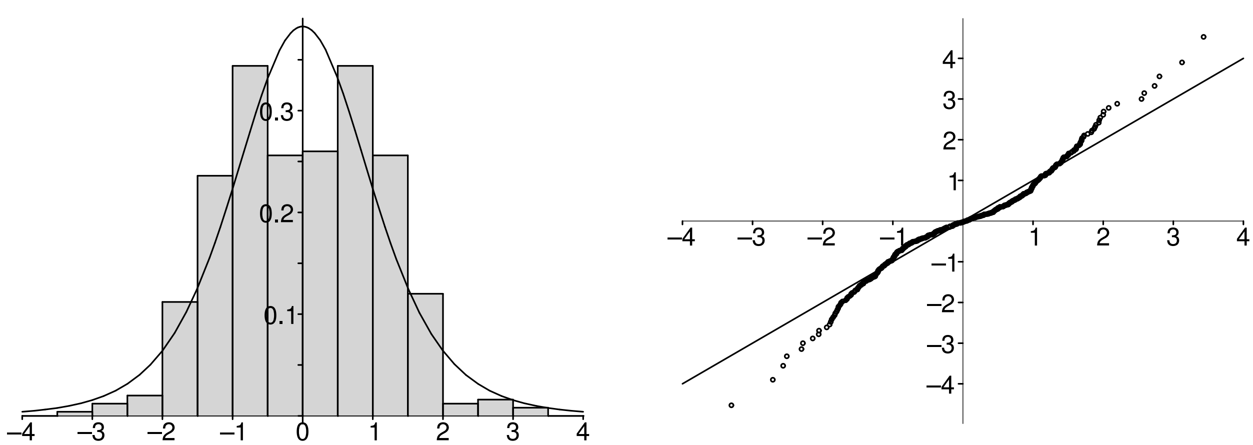


Figure 5.3–12: **(c)** Histogram and *q*-*q* Plot When Sampling from the Cauchy Distribution

5.3–14 **(a)**
```
> read ‘C:\\Tanis-Hogg\\Maple Examples\\stat.m‘:
  read ‘C:\\Tanis-Hogg\\Maple Examples\\HistogramFill.txt‘:}{%
> randomize():
  for k from 1 to 200 do
    X := NormalS(10,1,16):
    Svar[k] := evalf(Variance(X)):
  od:
  Svars := [seq(Svar[k], k = 1 .. 200)]:
  Mean(Svars);
  Variance(Svars);
  Percentile(Svars,0.05),Percentile(Svars,0.95);
```

$$.9994231685$$

$$.1435011407$$

Here is the answer for part **(c)** for the data in part a.

$$.5178685367, 1.805670407$$

```
> chisqs := 15*Svars:
  Mean(chisqs);
  Variance(chisqs);
```

$$14.99134750$$

$$32.28775779$$

```
> xtics := [seq(k, k = 0 .. 36)]:
  ytics := [seq(0.01*k, k = 1 .. 9)]:
> P1 := plot(ChisquarePDF(15,x), x = 0 .. 36,y = 0 .. 0.095, color=black,
  thickness=2, xtickmarks=xtics, ytickmarks=ytics, labels=[‘‘,‘‘]):
  P2 := HistogramFill(chisqs, 0 .. 36, 18):
  display({P1, P2});
```

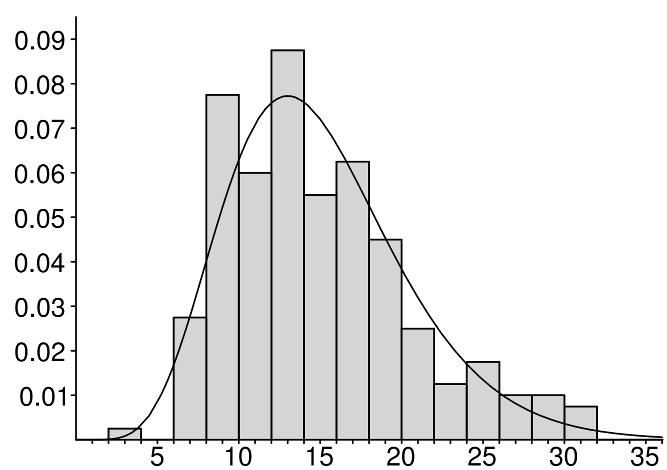


Figure 5.3–14: **(a)** Histogram of 200 Values of $15S^2$ When Sampling from $N(10, 1)$ Distribution

(b) Sample from a shifted exponential.

```
> randomize():
  for k from 1 to 200 do
    X := [seq(9 - ln(1 - rng()), k = 1 .. 16)]:
    Svar[k] := evalf(Variance(X)):
  od:
  Svars := [seq(Svar[k], k = 1 .. 200)]:
  Mean(Svars);
  Variance(Svars);
  Percentile(Svars,0.05),Percentile(Svars,0.95);
```

$$1.057524443$$

$$.6302476533$$

Here is the answer for part **(c)** for the data in part b.

$$.2698434433, 2.734101427$$

```
> chisqs := 15*Svars:
  Mean(chisqs);
  Variance(chisqs);
```

$$15.86286662$$

$$141.8057222$$

```
> xtics := [seq(3*k, k = 0 .. 20)]:
  ytics := [seq(0.01*k, k = 1 .. 8)]:
> P1 := plot(ChisquarePDF(15,x), x = 0 .. 60,y = 0 .. 0.085, color=black,
  thickness=2, xtickmarks=xtics, ytickmarks=ytics, labels=[``,``]):
  P2 := HistogramFill(chisqs, 0 .. 60, 20):
  display({P1, P2});
```

$$WARNING: \ There \ are, 2, \ data \ points \ out \ of \ plot \ range.$$

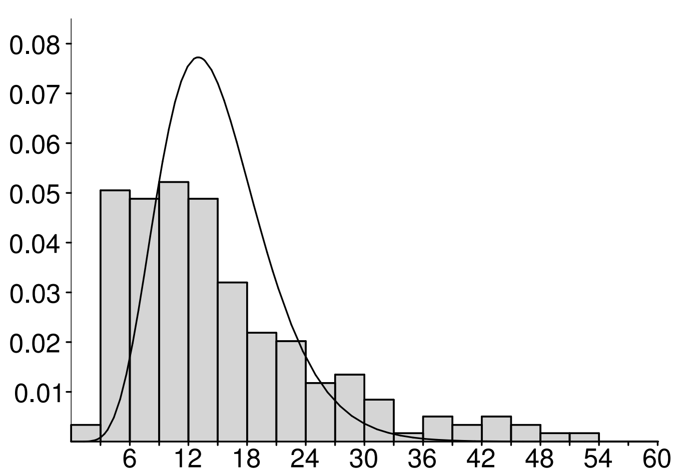


Figure 5.3–14: **(b)** Histogram of 200 Values of $15S^2$, Sampling from a Shifted Exponential

5.3–16 (This exercise is a simulation for the material in the last paragraph of this section. It is not in the text.) In this exercise we simulate observations from a bivariate normal distribution.

(a) Use Equations 5.3-3 and 5.3-4 to simulate 200 observations of the pair of bivariate normal random variables, (X_1, X_2), for which the means are equal to zero, the variances are equal to one, and the correlation coefficient is $\rho = 0.6$.

(b) Use Equations 5.3-5 to transform these observations of (X_1, X_2) into 200 observations of (Y_1, Y_2) for which $E(Y_1) = 5.8$, $E(Y_2) = 5.3$, $\sigma_1 = \sigma_2 = 0.2$, and $\rho = 0.6$.

(c) Construct relative frequency histograms of each of the sets of observations of Y_1 and Y_2 and superimpose the $N(5.8, 0.2^2)$ and $N(5.3, 0.2^2)$ p.d.f.s, respectively.

(d) Construct *q-q* plots of the quantiles of the $N(5.8, 0.2^2)$ and $N(5.3, 0.2^2)$ versus the order statistics of the observations of Y_1 and Y_2, respectively.

(e) Construct a scatterplot of the observations of (Y_1, Y_2) along with the least squares regression line.

(f) Calculate the sample correlation coefficient of the observations of (Y_1, Y_2).

(g) Does the simulation agree with the theory?

5.3–16 (a)

```
> read ‘C:\\Tanis-Hogg\\Maple Examples\\stat.m‘:
  read ‘C:\\Tanis-Hogg\\Maple Examples\\HistogramFill.txt‘:
  read ‘C:\\Tanis-Hogg\\Maple Examples\\ScatPlotCirc.txt‘:
> rho := 0.6:
  for k from 1 to 200 do
    U1 := rng():
    U2 := rng():
    Z1 := evalf(sqrt(-2*ln(U1))*cos(2*Pi*U2)):
    Z2 := evalf(sqrt(-2*ln(U1))*sin(2*Pi*U2)):
    X1[k] := evalf(sqrt(1 - rho^2)*Z1 + rho*Z2):
    X2[k] := Z2:
  od:
> X1s := [seq(X1[k], k = 1 .. 200)]:
  X2s := [seq(X2[k], k = 1 .. 200)]:
> x1bar := Mean(X1s);
  x2bar := Mean(X2s);
```

$$x1bar := .005103390290$$

$$x2bar := -.03892763303$$

```
> svarx1 := Variance(X1s);
  svarx2 := Variance(X2s);
```

$$svarx1 := 1.091815753$$

$$svarx2 := 1.084120871$$

```
> r := Correlation(X1s, X2s);
```

$$r := .6096305247$$

(b)

```
> for k from 1 to 200 do
    Y1[k] := 0.2*X1[k] + 5.8:
    Y2[k] := 0.2*X2[k] + 5.3:
  od:
> Y1s := [seq(Y1[k], k = 1 .. 200)]:
```

```
  Y2s := [seq(Y2[k], k = 1 .. 200)]:
> y1bar := Mean(Y1s);
  y2bar := Mean(Y2s);
```

$$y1bar := 5.801020665$$

$$y2bar := 5.292214470$$

```
> y1sd := StDev(Y1s);
  y2sd := StDev(Y2s);
```

$$y1sd := .2089804155$$

$$y2sd := .2082423453$$

(c)
```
> xtics := [seq(5.2 + k*0.05, k = 0 .. 24)]:
  ytics := [seq(0.2*k, k = 1 .. 12)]:
> P1 := plot(NormalPDF(5.8,0.2^2,x), x = 5.15 .. 6.45,
  y = 0 .. 2.25, color=black, thickness=2, xtickmarks=xtics,
  ytickmarks=ytics, labels=[``,``]):
  P2 := HistogramFill(Y1s, 5.2 .. 6.4, 18):
  display({P1, P2});
> xtics := [seq(4.7 + k*0.05, k = 0 .. 24)]:
  ytics := [seq(0.2*k, k = 1 .. 11)]:
> P5 := plot(NormalPDF(5.3,0.2^2,x), x = 4.65 .. 5.95,
  y = 0 .. 2.15, color=black, thickness=2, xtickmarks=xtics,
  ytickmarks=ytics, labels=[``,``]):
  P6 := HistogramFill(Y2s, 4.7 .. 5.9, 18):
  display({P5, P6});
```

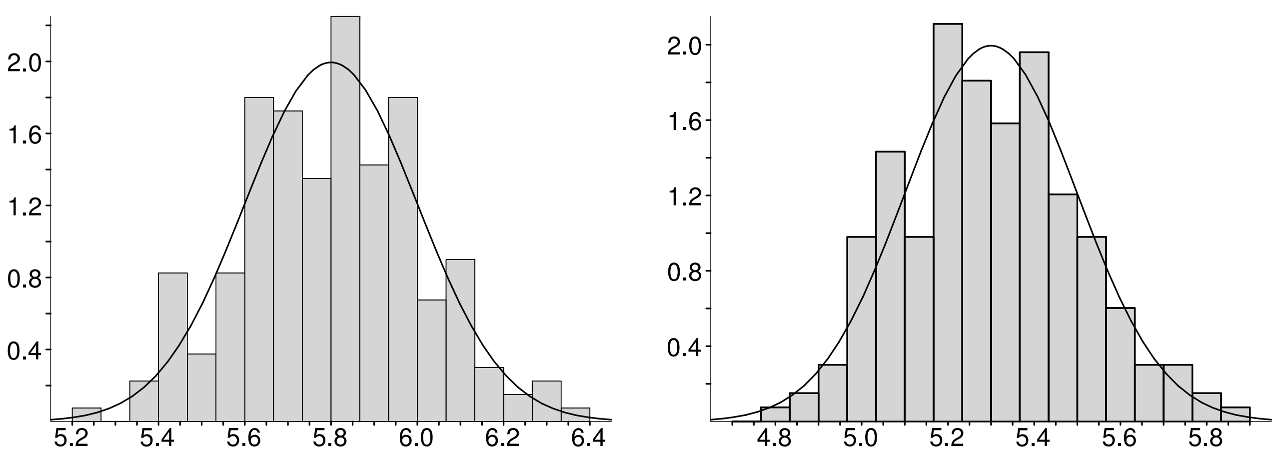


Figure 5.3–16: **(c)** $N(5.8, 0.2^2)$ on Left, $N(5.3, 0.2^2)$ on Right

(d)
```
> Quantiles1 := [seq(NormalP(5.8, 0.2^2, k/201), k = 1 .. 200)]:
  Y1sOrdered := sort(Y1s):
  xtics := [seq(5.2 + k*0.1, k = 0 .. 12)]:
  ytics := xtics:
  P3 := plot([[5.15, 5.15],[6.45,6.45]], x = 5.15 .. 6.45,
  y = 5.15 .. 6.45, color = black, thickness=2, xtickmarks=xtics,
  ytickmarks=ytics, labels=[``,``]):
  P4 := ScatPlotCirc(Y1sOrdered, Quantiles1):
```

```
  display({P3, P4});
> Quantiles2 := [seq(NormalP(5.3, 0.2^2, k/201), k = 1 .. 200)]:
  Y2sOrdered := sort(Y2s):
> xtics := [seq(4.7 + k*0.1, k = 0 .. 12)]:
  ytics := xtics:
> P7 := plot([[4.65, 4.65],[5.95,5.95]], x = 4.65 .. 5.95,
  y = 4.65 .. 5.95, color = black, thickness=2, xtickmarks=xtics,
  ytickmarks=ytics, labels=[``,``]):
  P8 := ScatPlotCirc(Y2sOrdered, Quantiles2):
  display({P7, P8});
```

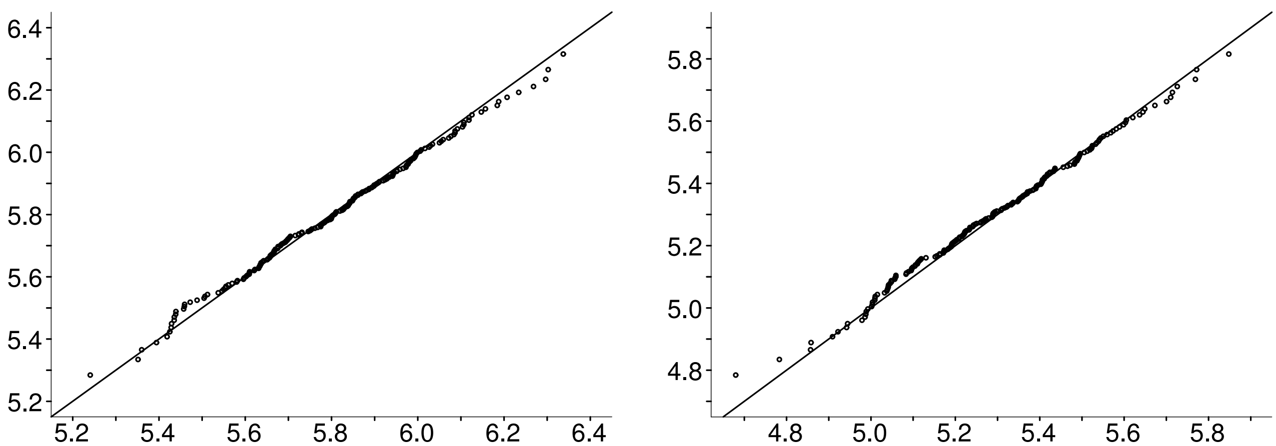


Figure 5.3–12: **(d)** q-q Plots, $N(5.8, 0.2^2)$ on Left, $N(5.3, 0.2^2)$ on Right

(e)
```
> f := LinReg(Y1s, Y2s, x);
```

$$f := 1.768218351 + .6074786357\,x$$

```
> xtics := [seq(5.2 + k*0.1, k = 0 .. 12)]:
  ytics := [seq(4.7 + k*0.1, k = 0 .. 12)]:
> P9 := plot(f,x = 5.15 .. 6.45, y = 4.65 .. 5.95, color=black, thickness=2,
  xtickmarks=xtics, ytickmarks=ytics):
  P10 := ScatPlotCirc(Y1s, Y2s):
  display({P9, P10});
```

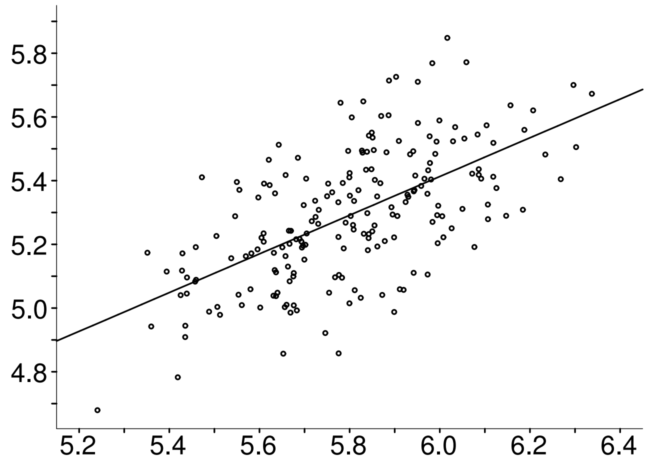


Figure 5.3–16: A Scatterplot of Observations of (Y_1, Y_2) with Least Squares Regression Line

(f)
```
> r := Correlation(Y1s, Y2s);
```

$$r := .6096317131$$

(g) The simulation agrees with the theory.

5.4 Resampling

5.4–2 **(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});
```

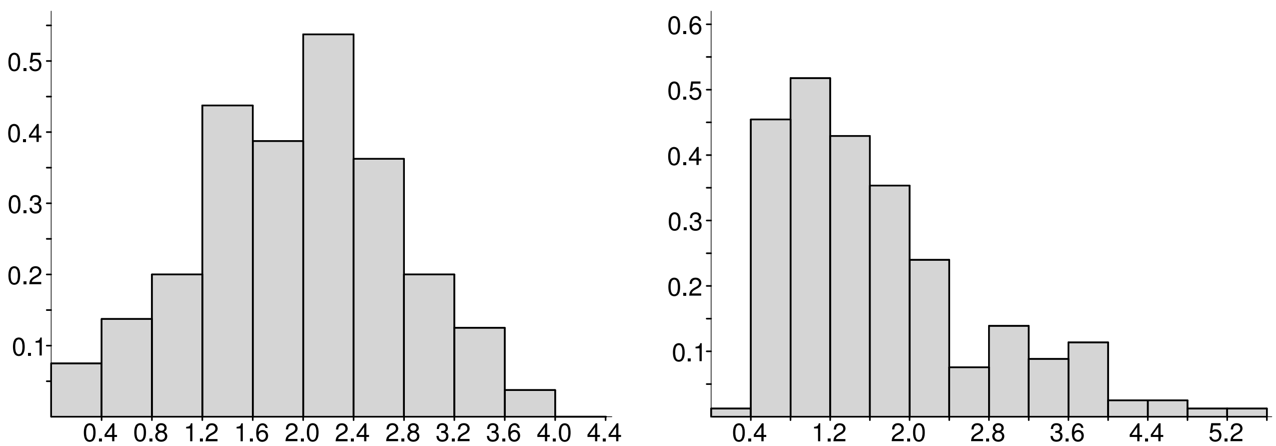


Figure 5.4–2: **(ab)** Histogram of S^2s: Resampling on Left, From Exponential on Right

```
> 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,
  thickness=2, xtickmarks=xtics, ytickmarks=ytics, labels=[‘‘,‘‘]):
  P6 := ScatPlotCirc(Svars,Svarys):
  display({P5, P6});
```

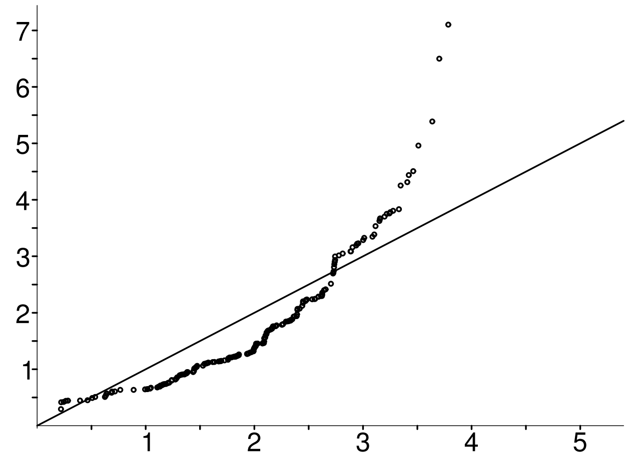


Figure 5.4–2: **(c)** q-q Plot of Exponential S^2s Versus Resampling S^2s

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.

5.4–4 **(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});
```

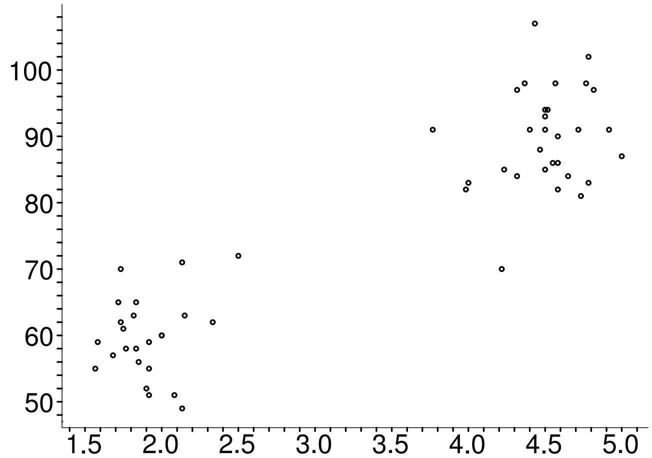


Figure 5.4–4: **(a)** Scatterplot of the 50 Pairs of Old Faithful Data

(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)

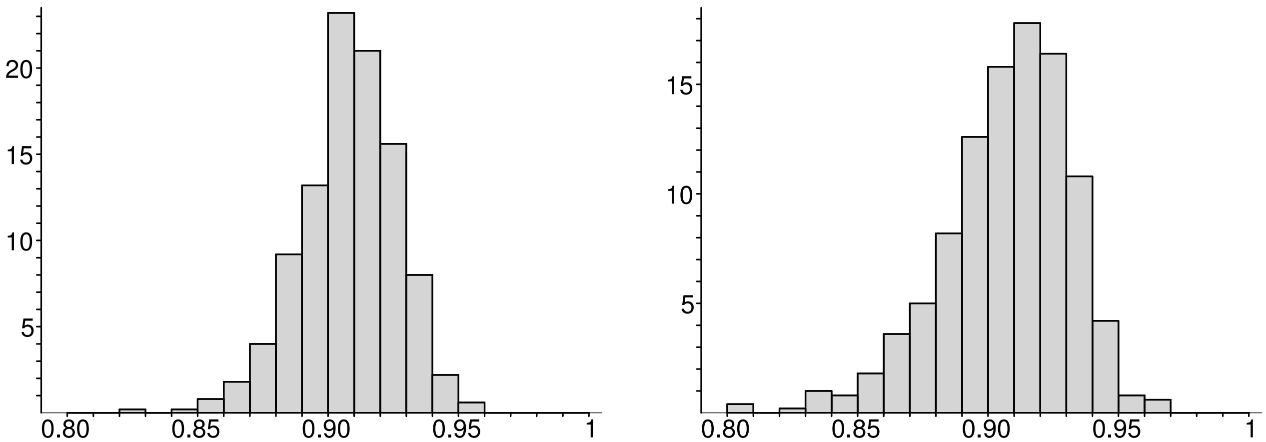


Figure 5.4–4: **(ce)** Histograms of Rs: From Resampling on Left, From Bivariate Normal on Right

(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});
```

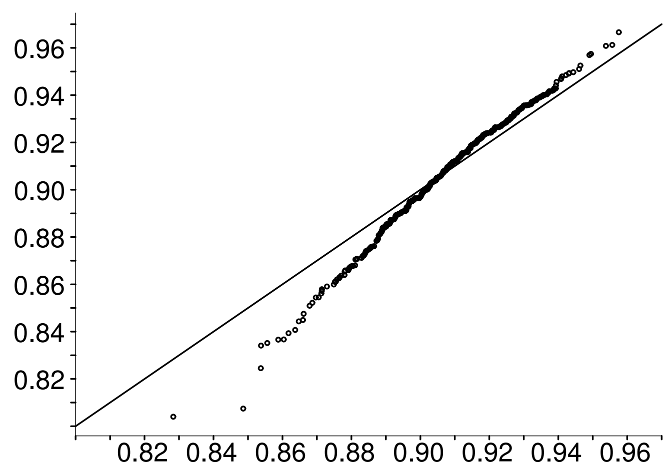


Figure 5.4–4: q-q Plot of the Values of R from Bivariate Normal Versus from Resampling

```
> 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.

5.4–6 **(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});
```

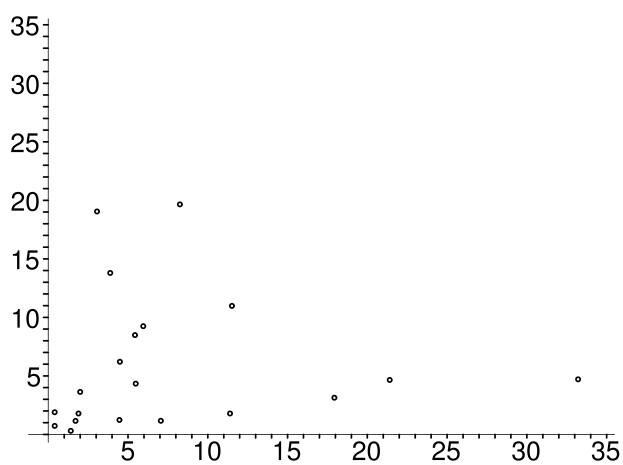


Figure 5.4–6: **(a)** Scatterplot of Paired Data from Two Independent Exponential Distributions

(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)

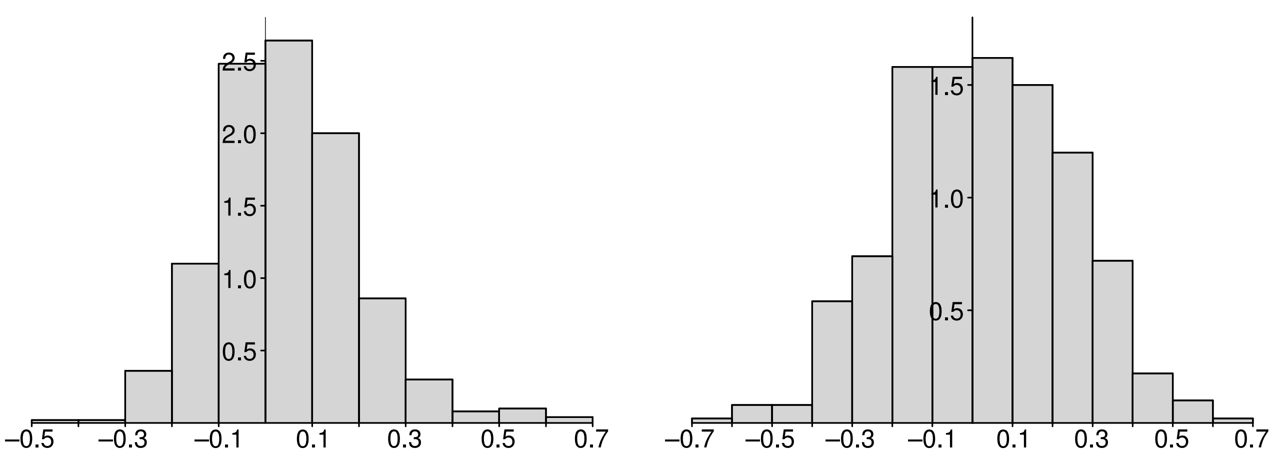


Figure 5.4–6: **(ce)** Histograms of Rs: From Resampling on Left, From Bivariate Normal on Right

(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});
```

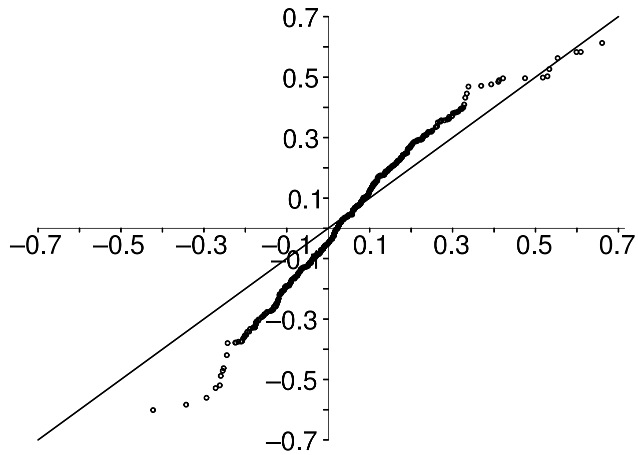


Figure 5.4–6: **(f)** q-q Plot of the Values of R from Bivariate Normal Versus from Resampling

```
> Mean(R);
  Mean(RBivNorm);
```

.04691961692

.02508989164

```
> StDev(R);
  StDev(RBivNorm);
```

.1527482117

.2200346799

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 6

Sampling Distribution Theory

6.1 Moment-Generating Function Technique

6.1–2 **(a)** $\mu = 0\left(\frac{44}{120}\right) + 1\left(\frac{45}{120}\right) + 2\left(\frac{20}{120}\right) + 3\left(\frac{10}{120}\right) + 5\left(\frac{1}{120}\right) = 1;$

$$\sigma^2 = 0^2\left(\frac{44}{120}\right) + 1^2\left(\frac{45}{120}\right) + 2^2\left(\frac{20}{120}\right) + 3^2\left(\frac{10}{120}\right) + 5^2\left(\frac{1}{120}\right) - 1^2 = 1;$$

(b) $P(X \ge 1) = 1 - P(X = 0) = \frac{76}{120} = \frac{19}{30}$

(c)

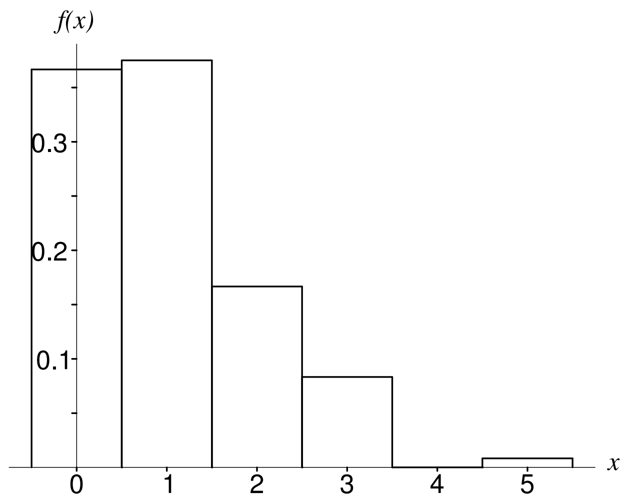


Figure 6.1–2: Probability Histogram of the p.m.f. of X

6.1–4
$$\begin{aligned} M(t) &= \sum_{x=0}^{\infty} e^{tx}\frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!} \\ &= e^{-\lambda}e^{\lambda e^t} = e^{\lambda(e^t-1)}; \end{aligned}$$

$$
\begin{aligned}
R(t) &= \ln M(t) = \lambda(e^t - 1), \\
R'(t) &= \lambda e^t, \\
R''(t) &= \lambda e^t, \\
\mu &= R'(0) = \lambda, \\
\sigma^2 &= R''(0) = \lambda.
\end{aligned}
$$

6.1–6 **(a)** Gamma $(\alpha = 1, \theta = 4)$;
(b) Gamma $(\alpha = 1, \theta = 1/2)$;
(c) Gamma $(\alpha = 10, \theta = 5)$;
(d) Gamma $(\alpha = 12, \theta = 2)$ or $\chi^2(24)$;
(e) $N(-7, 16)$.

6.1–8
```
> interface(showassumed=0):
  assume(t < 1):
> f := x*exp(-x);
```

$$f := x\, e^{(-x)}$$

```
> M := int(exp(t*x)*f, x = 0 .. infinity);
```

$$M := \frac{1}{(t-1)^2}$$

```
> Mprime := diff(M, t);
```

$$Mprime := -\frac{2}{(t-1)^3}$$

```
> Mdoubleprime := diff(Mprime, t);
```

$$Mdoubleprime := \frac{6}{(t-1)^4}$$

```
> mu := subs(t = 0, Mprime);
```

$$\mu := 2$$

```
> var := subs(t = 0, Mdoubleprime) - mu^2;
```

$$var := 2$$

6.2 M.G.F. of Linear Functions

6.2–2 **(a)**

$$
\begin{aligned}
M(t) &= E[e^{t(X_1+X_2+X_3)}] = E[e^{tX_1}]\, E[e^{tX_2}]\, E[e^{tX_3}] \\
&= e^{2(e^t-1)}\, e^{1(e^t-1)}\, e^{4(e^t-1)} = e^{7(e^t-1)}
\end{aligned}
$$

(b) Poisson$(\lambda = 7)$;
(c) $P(3 \le Y \le 9) = P(Y \le 9) - P(Y \le 2) = 0.830 - 0.030 = 0.800$.

6.2–4 **(a)**
$$\begin{aligned} M(t) &= E[e^{t(X_1+X_2+X_3+X_4+X_5)}] = E[e^{tX_1}]\,E[e^{tX_2}]\,E[e^{tX_3}]]\,E[e^{tX_4}]\,E[e^{tX_5}] \\ &= \left[\frac{e^t/3}{1-2e^t/3}\right]^5 = \frac{(e^t/3)^5}{(1-2e^t/3)^5}; \end{aligned}$$

(b) This is the moment-generating function for the negative binomial distribution with $r = 5$ and $p = 1/3$.

6.2–6 **(a)**
$$\begin{aligned} E(e^{tW}) &= E[e^{t(X+Y)}] = E[e^{tX}]\,E[e^{tY}] \\ &= (1/12)(e^{2t}+2e^{3t}+3e^{4t}+3e^{5t}+2e^{6t}+e^{7t}) \end{aligned}$$

(b)

w	2	3	4	5	6	7
$P(W=w)$	1/12	2/12	3/12	3/12	2/12	1/12

6.2–8
$$\begin{aligned} M(t) &= E[e^{t(X_1+X_2+\cdots+X_n)}] \\ &= E[e^{tX_1}]\,E[e^{tX_2}]\,\cdots E[e^{tX_n}] \\ &= (1-2t)^{-r_1/2}\,(1-2t)^{-r_2/2}\,\cdots(1-2t)^{-r_n/2} \\ &= (1-2t)^{-(r_1+r_2+\cdots+r_n)/2} \end{aligned}$$

which is the m.g.f. of $\chi^2(r_1+r_2+\cdots+r_n)$.

6.2–10 Using $\chi^2(16)$, $P\left(\dfrac{796.2}{100} \le \dfrac{\sum_{i=1}^{16}(X_i-50)^2}{100} \le \dfrac{2630}{100}\right) = 0.95 - 0.05 = 0.90;$

6.2–12
$$\begin{aligned} E[e^{t(X_1-X_2+n_2)}] &= E[e^{tX_1}]\,E[e^{-tX_2}]\,e^{tn_2} \\ &= \left(\frac{1}{2}+\frac{1}{2}\,e^t\right)^{n_1}\left(\frac{1}{2}+\frac{1}{2}\,e^{-t}\right)^{n_2}\,e^{tn_2} \\ &= \left(\frac{1}{2}+\frac{1}{2}\,e^t\right)^{n_1}\left(\frac{1}{2}+\frac{1}{2}\,e^t\right)^{n_2} = \left(\frac{1}{2}+\frac{1}{2}\,e^t\right)^{n_1+n_2} \end{aligned}$$

which is the m.g.f. of $b(n_1+n_2, 1/2)$.

6.2–14 **(a)**
```
> M := product(((7-i)/6)*exp(t)/(1 -(i-1)/6*exp(t)), i = 1 .. 6);
```

$$M := \frac{5}{324}\,\frac{(e^t)^6}{(1-\frac{1}{6}\,e^t)\,(1-\frac{1}{3}\,e^t)\,(1-\frac{1}{2}\,e^t)\,(1-\frac{2}{3}\,e^t)\,(1-\frac{5}{6}\,e^t)}$$

(b)
```
> Mprime := diff(M, t):
  mu := eval(subs(t=0, Mprime));
```

$$\mu := \frac{147}{10}$$

```
> mu := sum(1/((7-i)/6), i = 1 .. 6);
```

$$\mu := \frac{147}{10}$$

(c)
```
> Mdoubleprime := diff(Mprime,t):
  var := eval(subs(t=0, Mdoubleprime)) - mu^2;
```

$$var := \frac{3899}{100}$$

```
> var := sum((i/6)/((6-i)/6)^2, i = 1 .. 5);
```

$$var := \frac{3899}{100}$$

(d) Here are the results of a simulation using *Maple*.

```
> for k from 1 to 1000 do
    XX[k] := 1: # The first roll is always a new face
    for j from 2 to 6 do
      p := (7 - j)/6: # probability of new face after seeing j-1 faces
      XX[k] := XX[k] + op(GeometricS(p,1)): # gives observation of
           geometric random variable with probability of success p
    od:
  od:
> Y := [seq(XX[k], k = 1 .. 1000)]:}{%
  ybar := evalf(Mean(Y));
  mu := evalf(mu);
```

$$ybar := 14.81000000$$

$$\mu := 14.70000000$$

```
> svar := evalf(Variance(Y));
  theovar := evalf(var);
```

$$svar := 40.91681682$$

$$theovar := 38.99000000$$

```
> sd := evalf(StDev(Y));
  sigma := sigma;
```

$$sd := 6.396625424$$

$$\sigma := 6.244197306$$

6.2–16

$$\begin{aligned} E(e^{tY}) &= E[e^{tX_1}]\,E[e^{tX_2}]; \\ (1-2t)^{-r/2} &= (1-2t)^{-r_1/2}\,E[e^{tX_2}]; \\ E[e^{tX_2}] &= (1-2t)^{-(r-r_1)/2} \end{aligned}$$

which is the m.g.f. of $\chi^2(r - r_1)$.

6.3 Limiting Moment-Generating Functions

6.3–2 $np = 1000(0.005) = 5$;

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

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

6.3–4 **(a)** Y is $\chi^2(18)$;

(b) $P\left(\dfrac{Y-18}{\sqrt{36}} \le \dfrac{9.390-18}{6}\right) \approx P(Z \le -1.435) = 0.0756$;

$P\left(\dfrac{Y-18}{\sqrt{36}} \le \dfrac{34.80-18}{6}\right) \approx P(Z \le 2.8) = 0.9974$.

6.3–6 $\mu = (72)(1/3) = 24; \qquad \sigma^2 = (72)(1/3)(2/3) = 16.$

$$\begin{aligned} P(22 \le Y \le 28) &= P(21.5 \le Y \le 28.5) = \Phi\left(\frac{28.5 - 24}{4}\right) - \Phi\left(\frac{21.5 - 24}{4}\right) \\ \Phi(1.125) - \Phi(-0.625) &= 0.8697 - 0.2660 = 0.6037. \end{aligned}$$

If you had used the binomial distribution, the answer is $0.8691 - 0.2690 = 0.6001$.

6.3–8
$$\begin{aligned} E[e^{t(X-n)/\sqrt{n}}] &= e^{-\sqrt{n}\,t}E[e^{(t/\sqrt{n})X}] \\ &= e^{-\sqrt{n}\,t}e^{n(e^{t/\sqrt{n}}-1)} \\ &= e^{-\sqrt{n}\,t}e^{n(t/\sqrt{n} + (t/\sqrt{n})^2/2! + (t_1/\sqrt{n})^3/3!)}, \\ &\qquad \text{where } t_1 \text{is between } t \text{ and } 0, \\ &= e^{t^2/2 + (t_1/\sqrt{n})/6} \end{aligned}$$

which in the limit equals $e^{t^2/2}$.

6.4 Use of Order Statistics in Non-regular Cases

6.4–2
```
> read `C:\\Tanis-Hogg\\Maple Examples\\stat.m`:
  with(plots):
> randomize():
  n := 30: # Number of observations of the shifted exponential random variable
  m := 50: # Number of confidence intervals that are simulated
  for k from 1 to m do
    for j from 1 to n do
      XX[j] := evalf(5 - ln(1 - rng())):
    od:
    X := [seq(XX[j], j = 1 .. n)]:
    CIY1[k] := [evalf(Min(X) + ln(0.05)/30), Min(X)]: # Confidence intervals
                                                        using minimum
    CIxbar[k] := [evalf(Mean(X) - 1 - 2*StDev(X)/sqrt(n)), evalf(Mean(X)- 1 +
      2*StDev(X)/sqrt(n))]: # Confidence intervals using xbar and s
  od:
> CIY1s := [seq(CIY1[k], k = 1 .. m)]:
  CIxbars := [seq(CIxbar[k], k = 1 .. m)]:
```

These confidence intervals will now be plotted.

```
> xtics := [seq(4.85 + 0.05*k, k = 0 .. 6)]:}{%
  P1 := ConfIntPlot(CIY1s, 5): # plots the intervals based on min.
  P2 := plot([[0, 0],[0,0]], x = 4.85 .. 5.15, xtickmarks=xtics, labels=[``,``]):
  display(\{P1, P2\} );
```

This plot is in Figure 6.4-2 on the left.

```
> xtics := [seq(4 + 0.5*k, k = 0 .. 5)]:
  P3 := ConfIntPlot(CIxbars, 5): # plots the intervals based on xbar and s
  P4 := plot([[0, 0],[0,0]], x = 4.3 .. 6.2, xtickmarks=xtics, labels=[``,``]):
display(\{P3, P4\} );
```

This plot is in Figure 6.4-2 on the right.

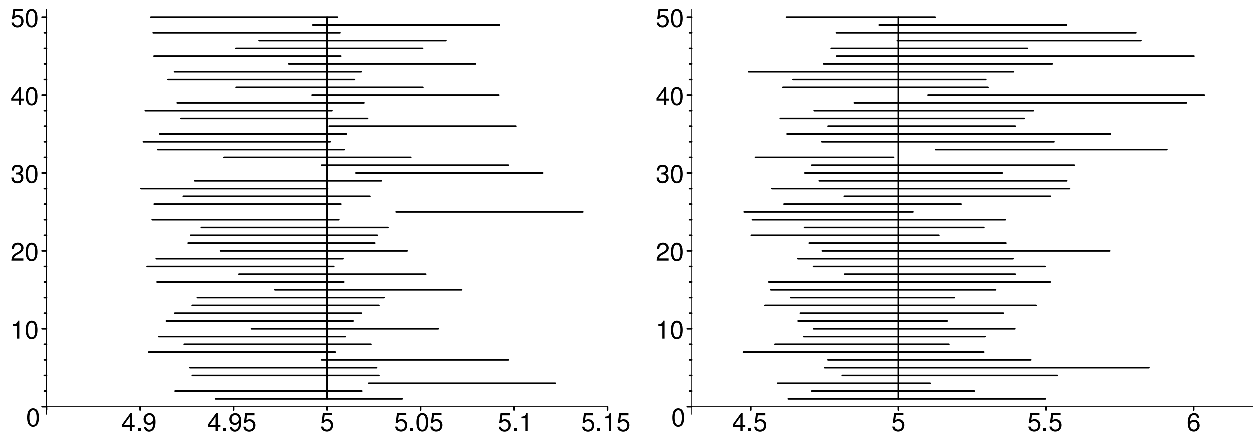


Figure 6.4–2: y_1 Intervals on Left, $(\overline{x},\, s)$ Intervals on Right

We now find the average lengths of the two sets of intervals.

```
> y1avlen := ConfIntAvLen(CIY1s); #finds the lengths of y1-intervals
  xbaravlen := ConfIntAvLen(CIxbars); #finds average lengths of xbar intervals
```

$$y1avlen := .09985774200$$

$$xbaravlen := .7581737928$$

We now find the number of confidence intervals that actually contain $\theta = 5$.

```
> y1success := ConfIntSuc(CIY1s, 5); # counts number that contain theta
  xbarsuccess := ConfIntSuc(CIxbars, 5); # counts number that contain theta
```

$$y1success := 46$$

$$xbarsuccess := 47$$

Note that all of the confidence intervals using y_1 have the same length, $-\ln(0.05)/30 = 0.0998577$ while the confidence intervals that use $\overline{x}$ and s vary in length because s changes from sample to sample. The scale of the two figures is quite different so note that the latter intervals are much longer than those based on y_1. Different simulations give different results but approximately 95% of each set of intervals should cover $\theta = 5$.