Instructor's Manual for

INTRODUCTION TO PROBABILITY AND STATISTICS FOR ENGINEERS AND SCIENTISTS

Fifth Edition

Sheldon M. Ross

Department of Industrial Engineering and Operations Research

University of California, Berkeley

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier

32 Jamestown Road, London NW1 7BY, UK
525 B Street, Suite 1800, San Diego, CA 92101-4495, USA
225 Wyman Street, Waltham, MA 02451, USA
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK

Fifth Edition 2014

Library of Congress Cataloging-in-Publication Data
Application submitted

ISBN 13: 978-0-12-802046-3

For all information on all Elsevier Academic Press publications visit our Web site at *www.elsevierdirect.com*

Table of Contents

Chapter 1

1. Method (c) is probably best, with (e) being the second best.
2. In 1936 only upper middle class and rich people had telephones. Almost all voters have telephones today.
3. No, these people must have been prominent to have their obituaries in the Times; as a result they were probably less likely to have died young than a randomly chosen person.
4. Locations (i) and (ii) are clearly inappropriate; location (iii) is probably best.
5. No, unless it believed that whether a person returned the survey was independent of that person's salary; probably a dubious assumption.
6. No, not without additional information as to the percentages of pedestrians that wear light and that wear dark clothing at night.
7. He is assuming that the death rates observed in the parishes mirror that of the entire country.
8. $12{,}246/.02 = 612{,}300$
9. Use them to estimate, for each present age x, the quantity $A(x)$, equal to the average additional lifetime of an individual presently aged x. Use this to calculate the average amount that will be paid out in annuities to such a person and then charge that person $1 + a$ times that latter amount as a premium for the annuity. This will yield an average profit rate of a per annuity.
10. 64 percent, 10 percent, and 48 percent.

Chapter 2

2. $360/r$ degrees.

6. (d) 3.18
(e) 3
(f) 2
(g) $\sqrt{5.39}$

7. (c) 119.14
(d) 44.5
(e) 144.785

8. Not necessarily. Suppose a town consists of n men and m women, and that a is the average of the weights of the men and b is the average of the weights of the women. Then na and mb are, respectively, the sums of the weights of the men and of the women. Hence, the average weight of all members of the town is

$$\frac{na + mb}{n + m} = ap + b(1 - p)$$

where $p = n/(n + m)$ is the fraction of the town members that are men. Thus, in comparing two towns the result would depend not only on the average of the weights of the men and women in the towns but also their sex proportions. For instance, if town A had 10 men with an average weight of 200 and 20 women with an average weight of 120, while town B had 20 men with an average weight of 180 and 10 women with an average weight of 100, then the average weight of an adult in town A is $200\,\frac{1}{3} + 120\,\frac{2}{3} = \frac{440}{3}$ whereas the average for town B is $180\,\frac{2}{3} + 100\,\frac{1}{3} = \frac{460}{3}$.

10. It implies nothing about the median salaries but it does imply that the average of the salaries at company A is greater than the average of the salaries at company B.

11. The sample mean is 110. The sample median is between 100 and 120. Nothing can be said about the sample mode.

12. (a) 40.904
(d) 8, 48, 64

13. (a) 15.808
(b) 4.395

14. Since $\sum x_i = n\bar{x}$ and $(n-1)s^2 = \sum x_i^2 - n\bar{x}^2$, we see that if x and y are the unknown values, then $x + y = 213$ and

$$x^2 + y^2 = 5(104)^2 + 64 - 102^2 - 100^2 - 105^2 = 22{,}715$$

Therefore,

$$x^2 + (213 - x)^2 = 22{,}715$$

Solve this equation for x and then let $y = 213 - x$.

15. No, since the average value for the whole country is a weighted average where the average wage per state should be weighted by the proportion of all workers who reside in that state.

19. (a) 44.8
(b) 70.45

20. 74, 85, 92

21. (a) 84.9167
(b) 928.6288
(c) 57.5, 95.5, 113.5

25. (a) .3496
(b) .35
(c) .1175
(d) no
(e) $3700/55 = 67.3$ percent

26. (b) 3.72067
(c) .14567

28. Not if both sexes are represented. The weights of the women should be approximately normal as should be the weights of the men, but combined data is probably bimodal.

30. Sample correlation coefficient is .4838

31. No, the association of good posture and back pain incidence does not by itself imply that good posture causes back pain. Indeed, although it does not establish the reverse (that back pain results in good posture) this seems a more likely possibility.

32. One possibility is that new immigrants are attracted to higher paying states because of the higher pay.

33. Sample correlation coefficient is .7429

34. If $y_i = a + bx_i$ then $y_i - \bar{y} = b(x_i - \bar{x})$, implying that

$$\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} = \frac{b}{\sqrt{b^2}} = \frac{b}{|b|}$$

35. If $u_i = a + bx_i$, $v_i = c + dy_i$ then

$$\sum (u_i - \bar{u})(v_i - \bar{v}) = bd \sum (x_i - \bar{x})(y_i - \bar{y})$$

and

$$\sum (u_i - \bar{u})^2 = b^2 \sum (x_i - \bar{x})^2, \quad \sum (v_i - \bar{v})^2 = d^2 \sum (y_i - \bar{y})^2$$

Hence,

$$r_{u,v} = \frac{bd}{|bd|} r_{x,y}$$

36. More likely, taller children tend to be older and that is why they had higher reading scores.

37. Because there is a positive correlation does not mean that one is a cause of the other. There are many other potential factors. For instance, mothers that breast feed might be more likely to be members of higher income families than mothers that do not breast feed.

Chapter 3

1. $S = \{$rr, rb, rg, br, bb, bg, gr, gb, gg$\}$ when done with replacement and $S = \{$rb, rg, br, bg, gr, gb$\}$ when done without replacement, where rb means, for instance, that the first marble is red and the second green.

2. $S = \{$hhh, hht, hth, htt, thh, tht, tth, ttt$\}$. The event $\{$hhh, hht,hth, thh$\}$ corresponds to more heads than tails.

3. (a) $\{7\}$, (b) $\{1, 3, 4, 5, 7\}$, (c) $\{3, 5, 7\}$, (d) $\{1, 3, 4, 5\}$, (e) $\{4, 6\}$, (f) $\{1, 4\}$

4. $EF = \{(1,2), (1,4), (1,6)\}$; $E \cup F = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)$, or any of the 15 possibilities where the first die is not 1 and the second die is odd when the first is even and even when the first is odd.$\}$; $FG = \{(1,4)\}$; $EF^c = \{$any of the 15 possible outcomes where the first die is not 1 and the two dice are not either both even or both odd$\}$; $EFG = FG$.

5. (a) $2^4 = 16$
(b) $\{(1, 1, 0, 0),\ (1, 1, 0, 1),\ (1, 1, 1, 0),\ (1, 1, 1, 1),\ (0, 0, 1, 1),\ (0, 1, 1, 1),\ (1, 0, 1, 1)\}$
(c) $2^2 = 4$

6. (a) EF^cG^c (b) EF^cG (c) $E \cup F \cup G$ (d) $EF \cup EG \cup FG$
(e) EFG (f) $E^cF^cG^c$ (g) $E^cF^c \cup E^cG^c \cup F^cG^c$
(h) $(EFG)^c$ (i) $EFG^c \cup EF^cG \cup E^cFG$ (j) S

7. (a) S (b) 0 (c) E (d) EF (e) $F \cup EG$

9. $1 = EF^cG^c$ $2 = EFG^c$ $3 = E^cFG^c$ $4 = EFG$ $5 = E^cFG$ $6 = E^cF^cG$ $7 = EF^cG$

10. Since $E \subset F$ it follows that $F = E \cup E^cF$ and since E and E^cF are mutually exclusive we have that

$$P(F) = P(E) + P(E^cF) \geq P(E)$$

11. Write $\cup E_i$ as the union of mutually exclusive events as follows:

$$\cup E_i = E_1 \cup E_1^c E_2 \cup E_1^c E_2^c E_3 \cup \cdots \cup E_1^c \quad E_{n-1}^c E_n$$

Now apply Axiom 3 and the results of Problem 10.

12. $1 \geq P(E \cup F) = P(E) + P(F) - P(EF)$

13. (i) Write $E = EF \cup EF^c$ and apply Axiom 3.
(ii) $P(E^cF^c) = P(E^c) - p(E^cF)$ from part (i)
$= 1 - P(E) - [P(F) - P(EF)]$

14. $P(EF^c \cup E^cF) = P(EF^c) + P(E^cF)$
$= P(E) - P(EF) + P(F) - P(EF)$ from Problem 13(i)

15. 84, 84, 21, 21, 120

16. To select r items from a set of n is equivalent to choosing the set of $n - r$ unselected elements.

17.
$$\binom{n-1}{r-1} + \binom{n-1}{r} = \frac{(n-1)!}{(n-r)!(r-1)!} + \frac{(n-1)!}{(n-1-r)!r!}$$
$$= \frac{n!}{(n-r)!r!}\left\{\frac{r}{n} + \frac{n-r}{n}\right\} = \binom{n}{r}$$

18. (a) 1/3 (b) 1/3 (c) 1/15

19. Because the 253 events that persons i and j have the same birthday are not mutually exclusive.

20. $P(\text{smaller of } (A, B) < C) = P(\text{smallest of the 3 is either } A \text{ or } B) = 2/3$

21. (a)
$$\begin{aligned} P(A \cup B) &= P(A \cup B|A)P(A) + P(A \cup B|A^c)P(A^c) \\ &= 1 \cdot P(A) + P(B|A^c)P(A^c) \\ &= .6 + .1(.4) = .64 \end{aligned}$$

(b) Assuming that the events A and B are independent, $P(B|A^c) = P(B)$ and

$$P(AB) = P(A)P(B) = .06$$

22. Chebyshev's inequality yields that at least $1 - 1/4$ of the accountants have salaries between \$90,000 and \$170,000. Consequently, the probability that a randomly chosen accountant will have a salary in this range is at least 3/4. Because a salary above \$160,000 would exceed the sample mean by 1.5 sample standard deviation, it follows from the one-sided Chebyshev inequality that at most $\frac{1}{1+9/4} = 4/13$ of accountants exceed this salary. Hence, the probability that a randomly chosen accountant will have a salary that exceeds this amount is at most 4/13.

23.
$$\begin{aligned} P(RR|\text{red side up}) &= \frac{P(RR, \text{red side up})}{P(\text{red side up})} \\ &= \frac{P(RR)P(\text{red side up}|RR)}{P(\text{red side up})} \\ &= \frac{(1/3)(1)}{1/2} = 2/3 \end{aligned}$$

24. 1/2

25.
$$P(F|CS) = \frac{P(FCS)}{P(CS)}$$
$$= \frac{.02}{.05} = 2/5$$
$$P(CS|F) = \frac{P(FCS)}{P(F)}$$
$$= \frac{.02}{.52} = 1/26$$

26. (a) $\dfrac{248}{500}$

(b) $\dfrac{54/500}{252/500} = \dfrac{54}{252}$

(c) $\dfrac{36/500}{248/500} = \dfrac{36}{248}$

27. Let D_i be the event that ratio i is defective.

$$P(D_2|D_1) = \frac{P(D_1D_2)}{P(D_1)}$$
$$= \frac{P(D_1D_2|A)P(A) + P(D_1D_2|B)P(B)}{P(D_1|A)P(A) + P(D_1|B)(P(B)}$$
$$= \frac{.05^2(1/2) + .01^2(1/2)}{.05(1/2) + .01(1/2)} = 13/300$$

28. (a) $\dfrac{6 \cdot 5 \cdot 4}{6^3} = 5/9$

(b) 1/6 because all orderings are equally likely.

(c) $(5/9)(1/6) = 5/54$

(d) $6^3 = 216$

(e) $\dfrac{6 \cdot 5 \cdot 4}{3 \cdot 2 \cdot 1} = 20$

(f) $20/216 = 5/54$

29.
$$P(A) = P(A|W)P(W) + P(A|W^c)P(W^c) = (.85)(.9) + (.2)(.1) = .785$$
$$P(W^c|D) = \frac{P(W^cD)}{P(D)} = \frac{(.8)(.1)}{.215} = 16/43$$

30. Let N_i be the event that i balls are colored red.

$$P(N_2|R_1R_2) = \frac{P(N_2R_1R_2)}{P(R_1R_2)}$$
$$= \frac{P(R_1R_2|N_2)P(N_2)}{P(R_1R_2|N_0)P(N_0) + P(R_1R_2|N_1)P(N_1) + P(R_1R_2|N_2)P(N_2)}$$
$$= \frac{1(1/4)}{0 + (1/4)(1/2) + 1(1/4)} = 2/3$$
$$P(R_3|R_1R_2) = \frac{P(R_1R_2R_3)}{P(R_1R_2)}$$
$$= \frac{0 + (1/8)(1/2) + 1(1/4)}{3/8} = 5/6$$

31. $$P(D|VR) = \frac{P(DVR)}{P(VR)} = \frac{50/1000}{590/1000} = 5/59$$

32. (a) 1/3 (b) 1/2

33. $P\{S \text{ in second}|S \text{ in first drawer}\} = P\{A\}/P\{S \text{ in first}\}$
$P\{S \text{ in first}\} = P\{S \text{ in first}|A\}1/2 + P\{S \text{ in first}|B\}1/2 = 1/2 + 1/2 \times 1/2 = 3/4$
Thus probability is $1/2 \div 3/4 = 2/3$.

34. $$P(C|E) = \frac{P(E|C)P(C)}{P(E|C)P(C) + P(E|C^c)P(C^c)}$$
$$= \frac{(.268)(.7)}{(.268)(.7) + (.145)(.3)} = .8118$$
$$P(C|E^c) = \frac{P(E^c|C)P(C)}{P(E^c|C)P(C) + P(E^c|C^c)P(C^c)}$$
$$= \frac{(.732)(.7)}{(.732)(.7) + (.865)(.3)} = .6638$$

35. (a) $P\{\text{good}|\text{O}\} = P\{\text{good, O}\}/P\{\text{O}\}$
$= .2P\{\text{O}|\text{good}\}/[P\{\text{O}|\text{good}\}.2 + P\{\text{O}|\text{average}\}.5 + P\{\text{O}|\text{bad}\}.3]$
$= .2 \times .95/[.95 \times .2 + .85 \times .5 + .7 \times .3] = 190/825$

36. (a) $P\{\text{sum is } 7|\text{first is } 4\} = P\{(4,3)\}/P\{\text{first is } 4\} = \frac{1/36}{1/6} = 1/6 = P\{\text{sum is } 7\}$.
(b) Same argument as in (a).

37. (a) $p_5[1 - (1 - p_1p_2)(1 - p_3p_4)]$
(b) Conditioning on whether or not circuit 3 closes yields the answer

$$p_3[(1-(1-p_1)(1-p_2)][1-(1-p_4)(1-p_5)]+(1-p_3)[1-(1-p_1p_4)(1-p_2p_5)]$$

38. $1-P(\text{at most 1 works}) = 1-Q_1Q_2Q_3Q_4-P_1Q_2Q_3Q_4-Q_1P_2Q_3Q_4-Q_1Q_2P_3Q_4-Q_1Q_2Q_3P_4$; where $Q_1 = 1 - P_1$.

39. (a) $1/8 + 1/8 = 1/4$
(b) $P(F \cup L) = P(F) + P(L) - P(FL) = 1/4 + 1/4 - 2/32 = 7/16$
(c) 6/32, since there are 6 outcomes that give the desired result.

40. Let N_i be the event that outcome i never occurs. Then

$$P(N_1 \cup N_2) = .5^n + .8^n - .3^n$$

Hence, the desired answer is $1 - .5^n + .8^n - .3^n$

41. Let W_1 be the event that component 1 works. Then,

$$P(W_1|F) = \frac{P(W_1F)}{P(F)} = \frac{P(F|W_1)(1/2)}{1-(1/2)^n} = \frac{1/2}{1-(1/2)^n}$$

42. 1: (a) $1/2 \times 3/4 \times 1/2 \times 3/4 \times 1/2 = 9/128$
(b) $1/2 \times 3/4 \times 1/2 \times 3/4 \times 1/2 = 9/128$
(c) 18/128
(d) $1 - P(\text{resembles first or second}) = 1 - [9/128 + 9/128 - P(\text{resembles both})]$
$= 110/128$

2: (a) $1/2 \times 1/2 \times 1/2 \times 1/2 \times 1/2 = 1/32$
(b) 1/32 (c) 1/16 (d) $1 - 2/32 = 15/16$

43. Prisoner A's probability of being executed remains equal to 1/3 provided the jailer is equally likely to answer either B or C when A is the one to be executed. To see this suppose that the jailer tells A that B is to be set free. Then

$$\begin{aligned} P\big\{A \text{ to be executed} \mid \text{jailer says } B\big\} &= P\{A \text{ executed}, B\}/P\{B\} \\ &= \frac{P\{B|A \text{ executed}\}1/3}{P\{B|A \text{ exec.}\}1/3 + P\{B|C \text{ exec.}\}1/3} \\ &= 1/6 + (1/6 + 1/3) = 1/3 \end{aligned}$$

44. Since brown is dominant over blue the fact that you have blue eyes means that both your parents have one brown and one blue gene. Thus the desired probability is 1/4.

45. (a) Call the desired probability p_A. Then $p_A = \frac{p^3}{p^3+(1-p)^3}$
(b) Conditioning on which team is ahead gives the result

$$p_A(1-(1-p)^4) + (1-p_A)(1-p^4)$$

(c) Let W be the event that team that wins the first game also wins the series. Now, imagine that the teams continue to play even after the series winner is decided. Then the team that won the first game will be the winner of the series if and only if that team wins at least 3 of the next 6 games played. (For if they do they would get to 4 wins before the other team, and if they did not then the other team would reach 4 wins first.) Hence,

$$P(W) = \sum_{i=3}^{6} \binom{6}{i} (1/2)^i (1/2)^{6-i} = \frac{20+15+6+1}{64} = \frac{21}{32}$$

46. Let 1 be the card of lowest value, 2 be the card of next higher value, and 3 be the card of highest value.

(a) 1/3, since the first card is equally likely to be any of the 3 cards.

(b) You will accept the highest value card if the cards appear in any of the orderings;

$$1,3,2 \quad \text{or} \quad 2,3,1 \quad \text{or} \quad 2,1,3$$

Thus, with probability 3/6 you will accept the highest valued card.

47. $.2 + .3 = .5, \qquad .2 + .3 - (.2)(.3) = .44, \qquad .2(.3)(.4) = .024, \qquad 0$

48. Let C be the event that the woman has breast cancer. Then

$$\begin{aligned} P(C|pos) &= \frac{P(C, pos)}{P(pos)} \\ &= \frac{P(pos|C)P(C)}{P(pos|C)P(C) + P(pos|C^c)P(C^c)} \\ &= \frac{.9(.02)}{.9(.02) + .1(.98)} \\ &= \frac{18}{116} \end{aligned}$$

49. Let C be the event that the household is from California and let O be the event that it earns over 250, 000. Then

$$\begin{aligned} P(C|O) &= \frac{P(CO)}{P(O)} \\ &= \frac{P(O|C)P(C)}{P(O|C)P(C) + P(O|C^c)P(C^c)} \\ &= \frac{.063(.12)}{.063(.12) + .033(.88)} = .2066 \end{aligned}$$

50.

$$\begin{aligned} P(A \cup B) &= P(A \cup B|A)P(A) + P(A \cup B|A^c)P(A^c) \\ &= P(A) + P(B|A^c)P(A^c) = .6 + .1(.4) = .64 \end{aligned}$$

51. The only way in which it would not be smaller than the value on card C is for card C to have the smallest of the 3 values, which is 1/3. Hence, the desired probability is 2/3.

Chapter 4

1. $P_1 = 5/10, P_2 = 5/10 \times 5/9 = .2778, P_3 = 5/10 \times 4/9 \times 5/8 = .1389.$

   $P_4 = 5/10 \times 4/9 \times 3/8 \times 5/7 = .0595, P_5 = 5/10 \times 4/9 \times 3/8 \times 2/7 \times 5/6 = .0198,$

   $P_6 = 5/10 \times 4/9 \times 3/8 \times 2/7 \times 1/6 = .0040,$ where $P_i = P(X = i)$.

2. $n - 2i, i = 0, 1, \ldots, n$

3. $P\{X = 3 - 2i\} = P\{i \text{ tails}\} = P_i,$ where $P_0 = 1/8, P_1 = 3/8, P_2 = 3/8, P_3 = 1/8.$

4. (b) $1 - F(1/2) = 3/4$ (c) $F(4) - F(2) = 1/12$ (d) $\lim_{h \to 0} F(3 - h) = 11/12$
   (e) $F(1) - \lim_{h \to 0} F(1 - h) = 2/3 - 1/2 = 1/6$

5. (a) $c \int_0^1 x^3 dx = 1 \Rightarrow c = 4$
   (b) $4 \int_{.4}^{.8} x^3 dx = .8^4 - .4^4 = .384$

6. Note first that since $\int f(x)dx = 1$, it follows that $\lambda = 1/100$; therefore, $\int_{50}^{150} f(x)dx = e^{-1/2} - e^{-3/2} = .3834$. Also, $\int_0^{100} f(x)dx = 1 - e^{-1} = .6321$.

7. The probability that a given radio tube will last less than 150 hours is $\int_0^{150} f(x)dx = 1 - 2/3 = 1/3$. Therefore, the probability desired is $\binom{5}{2}(1/3)^2(2/3)^3 = .3292$

8. Since the density must integrate to 1: c = 2 and $P\{X > 2\} = e^{-4} = .0183$.

9. With $p(i, j) = P(N_1 = i, N_2 = j)$

$$\begin{aligned}
p(1,1) &= (3/5)(2/4) = 3/10 \\
p(1,2) &= (3/5)(2/4)(2/3) = 2/10 \\
p(1,3) &= (3/5)(2/4)(1/3) = 1/10 \\
p(2,1) &= (2/5)(3/4)(2/3) = 2/10 \\
p(2,2) &= (2/5)(3/4)(1/3) = 1/10 \\
p(3,1) &= (2/5)(1/4) = 1/10 \\
p(i,j) &= 0 \quad \text{otherwise}
\end{aligned}$$

10. (a) Show that the multiple integral of the joint density equals 1.
    (b) $\int_0^2 f(x, y)dy = 12x^2/7 + 6x/7$
    (c) $\int_0^1 \int_0^x f(x, y)dy\,dx = \int_0^1 (6x^3/7 + 3x^3/14)dx = 15/56.$

11. $P\{M \le x\} = P\{X_1 \le x, X_2 \le x, \ldots, X_n \le x\} = \prod_{i=1}^{n} P\{X_i \le x\} = x^n$.
Differentiation yields that the probability density function is $nx^{n-1}, 0 \le x \le 1$.

12. (i) Integrate over all y to obtain $f_X(x) = xe^{-x}$
(ii) Integrate the joint density over all x to obtain $f_Y(y) = e^{-y}$ (since $\int xe^{-x}\,dx = 1$).
(iii) Yes since the joint density is equal to the product of the individual densities.

13. (i) $\int_x^1 2dy = 2(1-x), 0 < x < 1$.
(ii) $\int_0^y 2dx = 2y, 0 < y < 1$.
(iii) No, since the product of the individual densities is not equal to the joint density.

14. $f_X(x) = k(x)\int 1(y)dy$, and $f_Y(y) = 1(y)\int k(x)dx$. Hence, since $1 = \int\int f(x,y)dydx = \int 1(y)dy \int 1(x)dx$. we can write $f(x,y) = f_X(x)f_Y(y)$ which proves the result.

15. Yes because only in Problem 12 does the joint density factor.

16. (i) $P(X + Y \le a) = \iint\limits_{x+y\le a} f(x,y)dxdy$

$$= \iint\limits_{x\le a-y} f_X(x)f_Y(y)dxdy = \int F_X(a-y)f_Y(y)dy$$

(ii) $P\{X \le Y\} = \iint\limits_{x<y} f_X(x)dx f_Y(y)dy = \int F_X(y)f_Y(y)dy$.

17. $P\{W \le a\} = \iint\limits_{x^2y\le a} f_I(x)f_R(y)dydx$

$$= \int_0^{\sqrt{a}}\int_0^1 + \int_{\sqrt{a}}^1\int_0^{a/x^2} = 3a - 2a^{2/3} + a^2\int_{\sqrt{a}}^1 6x(1-x)/x^4dx$$

The density is now obtained upon differentiation:

$$f_W(a) = 3 - 3a^{1/2} + 2a\int_{\sqrt{a}}^1 6(1-x)/x^3dx - 3(1-a^{1/2})$$

19. (a) $f_Y(y) = \int_0^1 f(x,y)dx = 2/7 + 3y/14, 0 < y < 2$. Hence,

$$f_{x|y}(x|y) = \frac{12x^2 + 6xy}{4+3y}, \quad 0 < x < 1.$$

20. $P_{X|Y}(x|y) = P_X(x)$ is equivalent to $P_{X,Y}(x,y)/P_Y(y) = P_X(x)$ which is the condition (3.8). The verification in (b) is similar.

21. $E[X] = 5/10 + 5/9 + 5/12 + 5/21 + 25/252 + 1/42 = 1.833$

22. $E[X] = 0$ by symmetry.

23. The expected score of a meteorologist who says that it will rain with probability p is

$$E = p^*[1 - (1-p)^2] + (1-p^*)[1-p^2]$$

Differentiation yields that

$$\frac{dE}{dp} = 2p^*(1-p) - 2p(1-p^*).$$

Setting the above equal to 0 yields that the maximal (since the second derivative is negative) value is attained when $p = p^*$.

24. If the company charges c, then

$$E[\text{profit}] = c - Ap$$

Therefore, $E[\text{profit}] = .1A$ when $c = A(p + .1)$.

25. (a) $E[X]$, because the randomly chosen student is more likely to have been on a bus carrying a large number of students than on one with a small number of students.

(b) $E[X] = 40(40/148) + 33(33/148) + 25(25/148) + 50(50/148) \approx 39.28$
$E[Y] = (40 + 33 + 25 + 50)/4 = 37$

26. Let X denote the number of games played.

$$E[X] = 2[p^2 + (1-p)^2] + 3[2p(1-p)] = 2 + 2p(1-p)$$

Differentiating this and setting the result to 0 gives that the maximizing value of p is such that

$$2 = 4p$$

27. Since f is a density it integrates to 1 and so $a + b/3 = 1$. In addition $3/5 = E[X] = \int_0^1 x(a + bx^2)dx = a/2 + b/4$. Hence, $a = 3/5$ and $b = 6/5$.

28. $E[X] = \alpha^2 \int x^2 e^{-\alpha x}\,dx = \int y^2 e^{-y} dt/\alpha = 2/\alpha$ (upon integrating by parts twice).

29. $E[\text{Max}] = \int_0^1 xnx^{n-1}dx = n/(n+1)$ where we have used the result of Problem 11. $P\{\text{Min} \le x\} = 1 - P\{\text{Min} > x\} = 1 - \prod P\{X_i > x\} = 1 - (1-x)^n, 0 < x < 1$. Hence, the density of Min is $n(1-x)^{n-1}, 0 < x < 1$; and so

$$E[\text{Min}] = \int_0^1 nx(1-x)^{n-1}dx = \int_0^1 n(1-y)y^{n-1}dy = 1/(n+1).$$

30. $P\{X^n \le x\} = P\{X \le x^{1/n}\} = x^{1/n}$. Hence, $f_{X^n}(x) = x^{(1/n-1)}/n$, and so $E[X^n] = \frac{1}{n}\int_0^1 x^{1/n}dx = 1/(n+1)$. Proposition 5.1 directly yields that $E[X^n] = \int_0^1 x^n dx = 1/(n+1)$.

31. $E[\text{cost}] = \dfrac{1}{2}\int_0^2 (40 + 30\sqrt{x})dx = 40 + 10 \times 2^{3/2} = 68.284$

32. (a) $E[4 + 16X + 16X^2] = 164$
(b) $E[X^2 + X^2 + 2X + 1] = 21$

33. Let $X_i = \begin{matrix} 1 & \text{if the } i\text{th ball chosen is white} \\ 0 & \text{otherwise} \end{matrix}$.
Now $E[X_i] = P\{X_i = 1\} = 17/40$ and so $E[X] = 170/40$.
Suppose the white balls are arbitrarily numbered before the selection and let

$$Y_i = \begin{matrix} 1 & \text{if white ball number } i \text{ is selected,} \\ 0 & \text{otherwise.} \end{matrix}$$

Now $E[Y_i] = P\{Y_i = 1\} = 10/40$ and so $E[X] = 170/40$.

34. (a) Since

$$F(x) = \int_0^x e^{-x} dx = 1 - e^{-x}$$

if follows that

$$1/2 = 1 - e^{-m} \quad \text{or} \quad m = \log(2)$$

(a) In this case, $F(x) = x, 0 \leq x \leq 2$; hence $m = 1/2$.

35. Using the expression given in the hint yields that

$$\begin{aligned} \frac{d}{dc}E[|X - c|] &= cf(c) + F(c) - cf(c) - cf(c) - [1 - F(c)] + cf(c) \\ &= 2F(c) - 1 \end{aligned}$$

Setting equal to 0 and solving gives the result.

36. As $F(x) = 1 - e^{-2x}$, we see that

$$p = 1 - e^{-2m_p} \quad \text{or} \quad m_p = -\frac{1}{2}\log(1 - p)$$

37. $E[X_i] = \binom{198}{50}/\binom{200}{50} = \frac{150 \times 149}{200 \times 199}$. Hence $E[\sum X_i] = 75 \times 149/199 = 56.156$.

38. Let X_i equals 1 if trial i is a success and let it equal 0 otherwise. Then $X = \sum X_i$ and so $E[X] = \sum E[X_i] = np$. Also $\text{Var}(X) = \sum \text{Var}(X_i) = np(1 - p)$ since the variance of $\text{Var}(X_i) = E[X_i^2] - (E[X_i])^2 = p - p^2$. Independence is needed for the variance but not for the expectation (since the expected value of a sum is always the sum of the expected values but the corresponding result for variances requires independence).

39. $E[X] = (1 + 2 + 3 + 4)/4 = 10/4$. $E[X^2] = (1 + 4 + 9 + 16)/4 = 30/4$, and $\text{Var}(X) = 1.25$

40. $p_1 + p_2 + P_3 = 1$ $p_1 + 2p_2 + 3p_3 = 2$ and the problem is to minimize and maximize $p_1 + 4p_2 + 9p_3 = P_1 + 4(1 - 2p_1) + 2p_1 + 4$. Clearly, the maximum is obtained when $p_1 = 1/2$ — the largest possible value of p_1 since $p_3 = p_1$ — (and $p_2 = 0, p_3 = 1/2$) and the minimum when $p_1 = 0$ (and $p_2 = 1, p_3 = 0$).

41. Let X_i denote the number that appear on the ith flip. Then $E[X_i] = 21/6$. $E[X_i^2] = 91/6$, and $\text{Var}(X_i) = 91/6 - 49/4 = 35/12$. Therefore,

$$E\left[\sum X_i\right] = 3 \times 21/6 = 21/2; \text{Var}\left(\sum X_i\right) = 35/4.$$

42. $0 \le \text{Var}(X) = E[X]^2 - (E[X])^2$. Equality when the variance is 0 (that is, when X is constant with probability 1).

43.
$$E[X] = \int_8^9 x(x-8)dx + \int_9^{10} x(10-x)dx$$
$$E[X^2] = \int_8^9 x^2(x-8)dx + \int_9^{10} x^2(10-x)dx \text{ and } \text{Var}(X) = E[X^2] - (E[X])^2$$
$$E[\text{Profit}] = -\int_8^{8.25} (x/15 + .35)f(x)dx + \int_{8.25}^{10} (2 - x/15 - .35)f(x)dx$$

44. (a)
$$\begin{aligned} f_{X_1}(x) &= 3\int_0^{1-x} (x+y)dy \\ &= 3x(1-x) + 3(1-x)^2/2 \\ &= \frac{3}{2}(1-x^2), \quad 0 < x < 1, \end{aligned}$$

with the same density for X_2.
(b) $E[X_i] = 3/8$, $\text{Var}(X_i) = 1/5 - (3/8)^2 = 19/64$

45.
$$P_{X_1}(i) = \begin{matrix} 3/16, & i = 0 \\ 1/8, & i = 1 \\ 5/16, & i = 2 \\ 3/8, & i = 3 \end{matrix} \qquad P_{X_2}(i) = \begin{matrix} 1/2 & i = 1 \\ 1/2 & i = 2 \end{matrix}$$

$E[X_1] = 30/16$ $\text{Var}(X_1) = 19/4 - (15/8)^2 = 1.234$, $E[X_2] = 3/2$ $\text{Var}(X_2) = .25$

46.
$$\begin{aligned} E[X_1X_2] &= 3\int_0^1 \int_0^{1-x} xy(x+y)dydx \\ &= 3\int_0^1 x\int_0^{1-x} (xy + y^2)dy \\ &= 3\int_0^1 x(x(1-x)^2/2 + (1-x)^3/3))dx \end{aligned}$$

$$
\begin{aligned}
&= \frac{3}{2}\int_0^1 x^2(1-x)^2 dx + \int_0^1 x(1-x)^3 dx \\
&= \frac{1}{20} + \frac{1}{20} = \frac{1}{10}
\end{aligned}
$$

Hence,

$$
\text{Corr}(X_1, X_2) = \frac{1/10 - 9/64}{19/64} = -26/190
$$

47. $\text{Cov}(aX, Y) = E[aXY] - E[aX]E[Y] = aE[XY] - aE[X]E[Y] = a\text{Cov}(X, Y)$

48. $$\begin{aligned}\text{Cov}\left(\sum_{i=1}^{n} X_i, Y\right) &= \text{Cov}\left(\sum_{i=1}^{n-1} X_i, Y\right) + \text{Cov}(X_n, Y) \text{ by Lemma 7.1} \\ &= \sum_{i=1}^{n-1} \text{Cov}(X_i, Y) + \text{Cov}(X_n, Y) \text{ by the induction hypothesis}\end{aligned}$$

49. $0 \leq \text{Var}(X/\sigma_x + Y/\sigma_y) = 1 + 1 + 2\text{Cov}(X, Y)/\sigma_x\sigma_y$ since $\text{Var}(X/\sigma_x) = 1$ which yields that $-1 \leq \text{Corr}(X, Y)$. The fact that $\text{Corr}(X, Y) \leq 1$ follows in the same manner using the second inequality. If $\text{Corr}(X, Y) = 1$ then $0 = \text{Var}(X/\sigma_x - Y/\sigma_x)$ implying that $X/\sigma_x - Y/\sigma_y = c$ or $Y = a + bX$, where $b = \sigma y/\sigma x$. The result for $\text{Corr}(X, Y) = -1$ is similarly show.

50. If N_1 is large, then a large number of trials result in outcome1, implying that there are fewer possible trials that can result outcome 2. Hence, intuitively, N_1 and N_2 are negatively correlated.

$$
\begin{aligned}
\text{Cov}(N_1, N_2) &= \sum_{i=1}^{n} \sum_{j=i}^{n} \text{Cov}\left(X_i, Y_j\right) \\
&= \sum_{i=1}^{n} \text{Cov}\left(X_i, Y_i\right) + \sum_{i=1}^{n} \sum_{j \neq i} \text{Cov}\left(X_i, Y_j\right) \\
&= \sum_{i=1}^{n} \text{Cov}(X_i, Y_i) \\
&= \sum_{i=1}^{n} \left(E\left[X_i Y_i\right] - E\left[X_i\right] E\left[Y_i\right]\right) \\
&= \sum_{i=1}^{n} \left(-E\left[X_i\right] E\left[Y_i\right]\right) \\
&= -np_1p_2
\end{aligned}
$$

where the third equality follows since X_i and Y_j are independent when $i \neq j$, and the next to last equality because $X_iY_i = 0$.

51. $E[X_iX_j] = P\{X_i = X_j = 1\} = P\{X_i = 1\}P\{X_j = 1|X_i = 1\} = \frac{1}{n}\frac{1}{n-1}$. Hence, $\text{Cov}(X_iX_j) = [n(n-1)]^{-1} - 1/n^2 = [n^2(n-1)]^{-1}$, for $i \neq j$; and since $\text{Var}(X_i) = \dfrac{1}{n}(1 - 1/n) = (n-1)/n^2$ we see that $\text{Var}(X) = (n-1)/n + 2\dbinom{n}{2}[n^2(n-1)]^{-1} = 1$.

52. $\text{Cov}(X_1 - X_2, X_1 + X_2) = \text{Cov}(X_1, X_1) - \text{Cov}(X_2, X_1) + \text{Cov}(X_1, X_2) - \text{Cov}(X_2, X_2) = 0$ since $\text{Cov}(X_1, X_1) = \text{Cov}(X_2, X_2)$ and $\text{Cov}(X_1, X_2) = \text{Cov}(X_2, X_1)$

53. $\phi(t) = \int e^{tx}e^{-x}dx = \int e^{-(1-t)x}dx = (1-t)^{-1}$
$\phi^1(t) = (1-t)^{-2}$ and so $E[X] = 1$
$\phi^2(t) = 2(1-t)^{-3}$ and so $E[X^2] = 2$. Hence, $\text{Var}(X) = 1$.

54. $E[e^{tX}] = \int_0^1 e^{tx}\,dx = (e^t - 1)/t = 1 + t/2! + t^2/3! + \cdots + t^n/(\text{n}+1)! + \cdots$. From this it is easy to see that nth derivative evaluated at $t = 0$ is equal to $1/(n+1) = E[X^n]$.

55. $P\{0 \leq X \leq 40\} = 1 - P\{|X - 20| > 20\} \geq 1 - 1/20$ by Chebyshev's inequality.

56. (a) 75/85 by Markov's inequality.
(b) it is greater than or equal to 3/4 by the Chebyshev's inequality.
(c) $P\{|\bar{X} - 75| > 75\} \leq \text{Var}(\bar{X})/25 = (25/n)/25 = 1/n$. So $n = 10$ would suffice.

57. $P(X \leq x) = P\left(Y \leq \frac{x-a}{b}\right) = P(a + bY \leq x)$
Therefore, X has the same distribution as $a + bY$, giving the results:
(a) $E(X) = a + bE[Y]$ (b) $\text{Var}(X) = b^2\text{Var}[Y]$

Chapter 5

1. $\binom{4}{2}(3/5)^3(2/5)^2 + \binom{4}{3}(3/5)^3(2/5) + (3/5)^4 = 513/625 = .8208$

2. $\binom{5}{2}(.2)^3(.8)^2 + \binom{5}{4}(.2)^4(.8) + (.2)^5 = .0579$

3. $\binom{10}{7} \cdot 7^7 \cdot 3^3 = .2668$

4. $\binom{4}{3}(3/4)^3(1/4) = 27/64$

5. Need to determine when

$$6p^2(1-p)^2 + 4p^3(1-p) + p^4 > 2p(1-p) + p^2$$

Algebra shows that this is equivalent to

$$(p-1)^2(3p-2) > 0$$

showing that the 4 engine plane is better when $p > 2/3$.

6. Since

$$E(X) = np = 7, \quad \text{Var}(X) = np(1-p) = 2.1$$

it follows that $p = .7$, $n = 10$. Hence,

$$P\{X = 4\} = \binom{10}{4}(.7)^4(.3)^6, \quad P\{X > 12\} = 0$$

7. Let X denote the number of successes and $Y = n - X$, the number of failures, in n independent trials each of which is a success with probability p. The result follows by noting that X and Y are both binomial with respective parameters $(n \cdot p)$ and $(n \cdot 1 - p)$.

8. $P\{X = k+1\} = \dfrac{n!}{(n-k-1)!(k+1)!}p^{k+1}(1-p)^{n-k-1} = \dfrac{n!}{(n-k)!k!}p^k(1-p)^{n-k}\dfrac{n-k}{k+1}\dfrac{p}{1-p}$. From this we see that $P\{X = k+1\} \geq P\{X = k\}$ if $p(n-k) \geq (1-p)(k+1)$ which is equivalent to $np \geq k+1-p$ or $k+1 \leq (n+1)p$.

9. $\sum\limits_{i=0}^{n} e^{ti}\binom{n}{i}p^i(1-p)^{n-i} = \sum\limits_{i=0}^{n}\binom{n}{i}(pe^t)^i(1-p)^{n-i} = (pe^t + 1 - p)^n$. The first 2 derivatives evaluated at $t = 0$ are np and $n(n-1)p^2 + np$ which gives that the mean is np and the variance $np(1-p)$.

10. (a) approximation $= .1839$ exact $= .1938$
(b) approximation $= .3679$ exact $= .3487$
(c) approximation $= .0723$ exact $= .0660$

11. (a) $1 - e^{-1/2} = .3935$ (b) $\frac{1}{2}e^{-1/2} = .3033$ (c) $1 - e^{-1/2} - \frac{1}{2}e^{-1/2} = .0902$

12.
$$P\{\text{beneficial}|0\text{ colds}\} = \frac{P\{0\text{ colds}|\text{beneficial}\}P\{\text{beneficial}\}}{P\{0|\text{ben}\}P\{\text{ben}\} + P\{0|\text{not ben}\}P\{\text{not ben}\}}$$
$$= \frac{e^{-2}3/4}{e^{-2}3/4 + e^{-3}1/4} = 3e^{-2}/(3e^{-2} + e^{-3}) = .8908$$

13. With $\lambda = 121.95$
(a) $1 - \sum_{i=0}^{129} e^{-\lambda}\lambda^i/i!$
(b) $\sum_{i=0}^{100} e^{-\lambda}\lambda^i/i!$

14. Assuming that each person's birthday is equally likely to be any of the 365 days.
(a) $1 - \exp\{-80,000/365^2\}$
(b) $1 - \exp\{-80,000/365\}$

15. Say that trial i is a success if the ith card turned over results in a match. Because each card results in a match with probability 4/52, the Poisson paradigm says that the number of matches should approximately be Poisson distributed with mean 4, yielding that the approximate probability of winning is $P(\text{no matches}) \approx e^{-4} = .0183$.

16. Exact $= 1 - \sum_{i=0}^{3}\binom{1000}{i}(.001)^i(.999)^{1000-i} = .01891$. Approximate $= 1 - e^{-1} - e^{-1} - \frac{1}{2}e^{-1} = .01899$

17. $P\{X = i\}/P\{X = i-1\} = \lambda/i \geq 1$ when $i \leq \lambda$.

18. $\dfrac{\binom{80}{10} + \binom{80}{9}\binom{20}{1}}{\binom{100}{10}} = .3630$

19. $P\{X = i\}/P\{X = i-1\} = \dfrac{(n-i+1)(k-i+1)}{i(m-k+i)}$

20. (a) $(1-p)^{k-1}p$
(b) $E[X] = p\sum k(1-p)^{k-1} = p/p^2 = 1/p$
(c) $\binom{k-1}{r-1}p^{r-1}(1-p)^{k-r}p$
(d) Using the hint, each Y_i is geometric with parameter p and, by part (b) mean $1/p$.

21. For $a < x < b$, $P\{a + (b-a)U < x\} = P\{U < (x-a)/(b-a)\} = (x-a)/(b-a)$

22. 2/3, 1/3

23. (a) $1 - \phi(-5/6) = \phi(5/6) = .7977$ (b) $\phi(1) - \phi(-1) = 2\phi(1) - 1 = .6827$
(c) $\phi(-1/3) = .3695$ (d) $\phi(10/6) = .9522$ (e) $1 - \phi(1) = .1587$

24. $(\Phi(1))^5 = .4215$, $10(\Phi(1.4))^2(1\Phi(1.4))^3 = .0045$

25. Let $p = P\{\text{rain exceeds } 50\} = 1 - \phi(2.5) = .00621$. Answer is $6p^2(1-p)^2 = .00023$

26. Let W_i be the demand in week $i, i = 1, 2$. Then,

$$\begin{aligned}P(W_i < 1100) &= P\left(\frac{W_i - 1000}{200} < \frac{1100 - 1000}{200}\right)\\ &= P(Z < .5)\\ &= \Phi(.5 = .6915)\end{aligned}$$

Hence, for (a)

$$P(W_1 < 1100,\ W_2 < 1100) = P(W_1 < 1100)P(W_2 < 1100) = (.6915)^2 = .4782$$

Using that $W_1 + W_2$ is normal with parameters

$$E[W_1 + W_2] = 2000, \quad \mathrm{Var}(W_1 + W_2) = (200)^2 + (200)^2 = 8 \times 10^4$$

we obtain

$$\begin{aligned}P(W_1 + W_2 > 2000) &= P\left(\frac{W_1 + W_2 - 2000}{\sqrt{8 \times 10^4}} > \frac{2200 - 2000}{\sqrt{8 \times 10^4}}\right)\\ &= P(Z < 2/\sqrt{8})\\ &= 1 - .7601 = .2399\end{aligned}$$

27. $.95 = P\{X > L\} = 1 - \phi\left(\frac{L-2000}{85}\right)$ or $\phi([L - 2000]/85) = .05$ implying that $(L - 2000)/85 = -1.64$ or $L = 1860.6$

28. $P\{|X - 1.2| > .01\} = 2P\{Z > .01/.005\} = 2(1 - \phi(2)) = .0456$.

29. (a) Make the change of variables $y = x/\sigma$
(b) $I^2 = \int_0^{2\pi} \int e^{-r^2/2} r dr d\theta = 2\pi$

30. $P\{X \le x\} = P\{\log X \le \log x\} = \phi([\log x - \mu]/\sigma)$

31. Let μ and σ^2 be the mean and variance. With X being the salary of a randomly chosen physician, and with $Z = (X - \mu)/\sigma$ being a standard normal, we are given that

$$.25 = P(X < 180) = P(Z < \frac{180 - \mu}{\sigma})$$

and that

$$.25 = P(X > 320) = P(Z > \frac{320 - \mu}{\sigma})$$

Because $P(Z < -.675) = P(Z > .675) \approx .25$, we see that

$$\frac{180 - \mu}{\sigma} = -.675, \quad \frac{320 - \mu}{\sigma} = .675$$

giving that $\mu = 250, \sigma = 70/.675 = 103.70$. Hence,

(a) $P(X < 250) = .5$

(b)
$$\begin{aligned} P(260 < X < 300) &= P(10/103.7 < Z < 50/103.7) \\ &= P(.096 < Z < .482) \approx .147 \end{aligned}$$

32. (a) $\frac{70-60}{20} < \frac{62-55}{10}$ so your percentile score was higher on the statistics exam.

(b) $P(X(econ) < 70) = P(Z < .5) = .6915$

(c) $P(X(stat) < 62) = P(Z < .7) = .7580$

33. (a) $.01 = P(-X > v) = P(X < -v) = P(Z < \frac{-v-10}{7})$. Because $P(Z < -2.33) = .01$ we have $v = 6.31$.

(b) $.01 = P(-X > v) = P(Z > \frac{v+\mu}{\sigma})$. Becuase $P(Z > 2.33) = .01$, this yields $v = -\mu + 2.33\sigma$.

34. (a) $1 - \Phi(1.86/8.7) = .415$

(b) $1 - \Phi\left(\dfrac{84 - 80.28}{8.7\sqrt{2}}\right) = .381$

(c) $1 - \Phi\left(\dfrac{126 - 120.42}{8.7\sqrt{3}}\right) = .356$

35. (a) $\Phi(-1.5/2.4) = .266$

(b) $\Phi(5.5/2.4) = .989$

(c) $\Phi(5.5/2.4) - \Phi(-1.5/2.4) = .723$

(d) $\Phi(7.5/2.4) = .999$

(e) $1 - \Phi\left(\dfrac{132 - 129}{2.4\sqrt{2}}\right) = .188$

(f) $1 - \Phi\left(\dfrac{264 - 158}{4.8}\right) = .106$

36. $\dfrac{x - 100}{14.2} = z_{.01} = 2.58 \rightarrow x = 100 + (2.58)(1.42) = 136.64$

37. (a) $e^{-1} = .3679$ (b) $e^{-1/2} = .6065$

38. $e^{-10/8} = .2865$

39. (i) $e^{-1} = .3679$ (ii) $P\{X > 30|X > 10\} = 1/4 + 3/4 = 1/3$

40. (a) The time of the nth event.

(b) The nth event occurs before or at time t is equivalent to saying that n or more events occur by time t.

(c) $P\{S_n \le t\} = P\{N(t) \ge n\} = 1 - \sum_{j=0}^{n-1} e^{-\lambda t}(\lambda t)^j/j!$

41. (a) $1 - e^{-2.5} - 2.5e^{-2.5} = .7127$

(b) $e^{-15/4} = .0235$ by the independent increment assumption

(c) $P\{N(3/4) \geq 4 | N(1/2) \geq 2\} = P\{N(3/4) \geq 4,\ N(1/2) \geq 2\}/P\{N(1/2) \geq 2\}$

$$= \frac{P\{N(1/2)=2, N(3/4)-N(1/2)\geq 2\}+P\{N(1/2)=3, N(3/4)-N(1/2)\geq 1\}+P\{N(1/2)\geq 4\}}{P\{N(1/2)\geq 2\}}$$

$$= \frac{(e^{-5/2}(5/2)^2/2)[1-e^{-5/4}-5/4e^{-5/4}]+e^{-2}(5/2)^3/6[1-e^{-5/4}]+1-\sum_{i=0}^{3} e^{-5/2}(5/2)^i/i!}{1-e^{-5/2}-5/2e^{-5/2}}$$

42. $(X/2)^2 + (Y/2)^2 = D^2/4$ is chi-square with 2 degrees of freedom. Hence, $P\{D > 3.3\} = P\{D^2/4 > (3.3)^2/4 = .2531$

43. .5770, .6354

44. .3504

45. Upon making the suggested substitution we see that

$$\Gamma(1/2) = \text{sqr}(2) \int_0 e^{-y^2/2} dy = 2\text{sqr}(\pi) P\{N(0.1) > 0\} = \text{sqr}(\pi)$$

46. .1732, .9597, .6536

47. $T = \dfrac{N(0.1)}{\text{sqr}(X_n/n)}$ where X_n is chi-square with n degrees of freedom. Therefore,

$$T^2 = N^2(0.1)(X_n/n)$$

which is F with 1 and n degrees of freedom.

48. X is a normal random variable with mean a and variance b^2.

Chapter 6

1. $E[\bar{X}_2] = E[\bar{X}_3] = 1.8,\ \text{Var}(\bar{X}_2) = .78,\ \text{Var}(\bar{X}_3) = .52$

2. If X_i is the ith roll, then $E[X_i] = 7/2$, $\text{Var}(X_i) = 35/12$. Hence, with $X = \sum_i X_i$, it follows that $E[X] = 35$, $\text{Var}(X) = 350/12$. Hence, by the central limit theorem

$$\begin{aligned} P\{30 \le X \le 40\} &= P\{29.5 \le X \le 40.5\} \\ &\approx P\left\{\frac{29.5 - 35}{\sqrt{350/12}} \le Z \le \frac{40.5 - 35}{\sqrt{350/12}}\right\} \\ &= \Phi(1.02) - \Phi(-1.02) = .6922 \end{aligned}$$

3. $E[S] = 8$, $\text{Var}(S) = 16/12$. By the central limit theorem

$$P\{S > 10\} \approx 1 - \Phi(2/\sqrt{4/3}) = .042$$

4. If W is your winnings on a single play, then $E[W] = 35/38 - 37/38 = -1/19 = -.0526$, $\text{Var}(W) = 33.21$.
(a) $1 - (37/38)^{34} = .596$

$$P\{S > 0\} = P\{S > .5\} \approx 1 - \Phi\left(\frac{.5 + .0526n}{\sqrt{33.21n}}\right)$$

The preceding is $1 - \Phi(.29) = .386$ when $n = 1000$, and $1 - \Phi(2.89) = .002$ when $n = 100,000$.

5. Let S be the amount of snow over the next 50 days.

(a) $P\{S < 80\} \approx \Phi\left(\dfrac{80 - 50(1.5)}{.3\sqrt{50}}\right) = \Phi(2.357) = .9908$

The preceding assumes that the daily amounts of snow are independent, a dubious assumption.

6. If R is the sum of the roundoff errors then R has mean 0 and variance 50/12. Therefore,

$$P\{|R| > 3\} = 2P\{R > 3\} \approx 2[1 - \Phi(3/\sqrt{50/12})] = .141$$

7. Imagine that we continue to roll forever, and let S be the sum of the first 140 rolls.

$$P\{S \le 400.5\} \approx \Phi\left(\frac{400.5 - 140(3.5)}{\sqrt{140(35/12)}}\right) = \Phi(-4.43) \approx 0$$

8. Let T be the lifetime (in weeks) of 12 batteries.

$$P\{T < 52\} \approx \Phi\left(\frac{52-60}{1.5\sqrt{52}}\right) = \Phi(-.739) = .230$$

9. (a) $P\{\bar{X} < 104\} \approx \Phi(16/20) = .788$
(b) $.788 - \Phi(-8/20) = .443$

10. $1 - \Phi(9/.3) = 0$

11. $1 - \Phi(25\sqrt{n}/80)$

12. (a) $\Phi(25/15) - \Phi(-25/15) = .9044$
(b) $\Phi(40/15) - \Phi(-40/15) = .9924$
(c) 1/2, since the amount by which the average score of the smaller exceeds that of the larger is a normal random variable with mean 0.
(d) the smaller one

14. $P\{X < 199.5\} \approx \Phi(-50.5/\sqrt{3000/16}) = \Phi(-3.688) = .0001$

15. (a) no
(b) they are both binomial
(c) $X = X_A + X_B$ (d) Since X_A is binomial with parameters (32, .5), and X_B is binomial with parameters (28, .7) it follows the X is approximately distributed as the sum of two independent normals, with respective parameters (16, 8) and (19.6, 5.88). Therefore, X is approximately normal with mean 35.6 and variance 13.88. Hence,

$$P\{X > 39.5\} \approx 1 - \Phi(3.9/\sqrt{13.88}) = .148$$

16. Since the sum of independent Poisson random variables remains a Poisson random variable, it has the same distribution as the sum of n independent Poisson random variables with mean λ/n. To 3 decimal places, the exact probability is .948; the normal approximation without the continuity correction is .945, and with the correction is .951.

17. The actual probability is .5832; the Poisson approximation is .5830 and the normal approximation is .566.

18. (a) $P\{S^2/\sigma \leq 1.8\} = P\{\chi^2_4 \leq 7.2\}$
(b) $P\{3.4 \leq \chi^2_4 \leq 4.2\}$

20. Using that $9S_1^2/4$ and $4S_2^2/2$ are chi squares with respective degrees of freedom 9 and 4 shows that $S_1^2/(2S_2^2)$ is an F random variable with degrees of freedom 9 and 4. Hence,

$$P\{S_2^2 > S_1^2\} = P\{S_1^2/(2S_2^2) < 1/2\} = P\{F_{9,4} < 1/2\}$$

21. .5583

22. Using the disk gives the answers: .6711, .6918, .9027, .99997

23. The exact answers are .0617, .9735

24. The exact answers are .9904, .0170

25. X, the number of men that rarely eat breakfast is approximately a normal random variable with mean $300(.42) = 126$ and variance $300(.42)(.58) = 73.08$ whereas Y, the number of men that rarely eat breakfast, is approximately a normal random variable with mean $300(.454) = 136.2$ and variance $300(.454)(.546) = 74.3652$. Hence, $X - Y$ is approximately normal with mean -10.2 and variance 147.4452, implying that

$$P\{X - Y > 0\} \approx 1 - \Phi\left(\frac{10.2}{\sqrt{147.5542}}\right) = \Phi(-.84) = .2005$$

27. $(.851)^5 = .4463$, $(.645)^5 = .1116$

28. Using that $120/\sqrt{144} = 10$ gives the following
(a) $1 - \Phi(-1) = \Phi(1) = .8413$
(b) 1/2
(c) $1 - \Phi(2) = \Phi(-2) = .0227$
(d) $1 - \Phi(3.3) = \Phi(-3.3) = .0005$

29. $1 - \Phi(1.4\sqrt{12}/3.2) = 1 - \Phi(1.516) = .0648$

Chapter 7

1. $f(x_1 \ldots x_n) = e^{n\theta} \exp\{-\sum x_i\} = ce^{n\theta}$, $\theta < x_i$, $i = 1, \ldots, n$; Thus, f is 0, otherwise maximized when θ is as large as possible — that is, when $\theta = \min x_i$. Hence, the maximum likelihood estimator is min x_i

2. $\log[f(x_1, \ldots, x_n)] = \log\left[\theta^{2n} \prod_{i=1}^{n} x_i e^{-\theta x_i}\right]$

$$= 2n\log(\theta) + \sum_{i=1}^{n} \log(x_i) - \theta \sum_{i=1}^{n} x_i$$

Therefore, $(\partial/\partial\theta)f = 2n/\theta - \sum_{i=1}^{n} x_i$. Setting equal to 0 gives the maximum likelihood estimator $\hat{\theta} = 2n/\sum_{i=1}^{n} x_i$

3. $f(x_1 \ldots x_n) = c(\sigma^2)^{-n/2} \exp\left(-\sum(x_i - \mu)^2/2\sigma^2\right)$
$\log(f(x)) = -n/2 \log\sigma^2 - \sum(x_i - \mu)^2/2\sigma^2$
$\dfrac{\mathrm{d}}{\mathrm{d}\sigma^2} \log f(x) = \dfrac{-n}{2\sigma^2} + \sum(x_i - \mu)^2/2\sigma^4$
Equating to 0 shows that the maximum likelihood estimator of σ^2 is $\sum(x_i - \mu)^2/n$. Its mean is σ^2.

4. The joint density is

$$f(x_1, \ldots, x_n) = \lambda^n a^{n\lambda}(x_1 \cdots x_n)^{-(\lambda+1)}, \quad \min_i x_i \geq a$$

and is 0 otherwise. Because this is increasing in a for $a \leq \min x_i$ and is then 0, $m = \min x_i$ is the maximum likelihood estimator for a. The maximum likelihood estimate of λ is the value that maximizes $\lambda^n m^{n\lambda}(x_1 \cdots x_n)^{-(\lambda+1)}$. Taking logs gives

$$n\log(\lambda) + n\lambda\log(m) - (\lambda + 1)\log(x_1 \cdots x_n)$$

Differentiating, setting equal to 0 and solving for λ, gives that its maximum likelihood estimator is $\frac{n}{log(x_1 \cdots x_n) - n\log(m)}$.

5. $f(x_1, \ldots, x_n, y_1, \ldots, y_n, w_1, \ldots, w_n)$
$= (2\pi\sigma^2)^{3n/2} e^{-\sum_{i=1}^{n}[(x_i - \mu_1)^2 + (y_i - \mu_2)^2 + (w_i - \mu_1 - \mu_2)^2]/(2\sigma^2)}$
$$\log[f(\text{data})] = \tfrac{3n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{n}\left[(x_i - \mu_1)^2 + (y_i - \mu_2)^2 + (w_i - \mu_1 - \mu_2)^2\right]/(2\sigma^2)$$

yielding

$$\frac{\partial}{\partial \mu_1} f = -\sum_{i=1}^{n} [(x_i - \mu_1) + (w_i - \mu_1 - \mu_2)]/\sigma^2$$

and

$$\frac{\partial}{\partial \mu_2} f = -\sum_{i=1}^{n} [(y_i - \mu_2) + (w_i - \mu_1 - \mu_2)]/\sigma^2$$

Setting equal to 0 gives

$$\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} w_i = 2n\mu_1 + n\mu_2 \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} w_i = n\mu_1 + 2n\mu_2$$

yielding $\hat{\mu}_1 = \dfrac{2\sum x_i + \sum w_i - \sum y_i}{3n}, \qquad \hat{\mu}_2 = \dfrac{2\sum y_i + \sum w_i - \sum x_i}{3n}$

6. The average of the distances is 150.456, and that of the angles is 40.27. Using these estimates the length of the tower, call it T, is estimated as follows:

$$T = X \tan(\theta) \approx 127.461$$

7. With $Y = \log(X)$, then $X = e^Y$. Because Y is normal with parameters μ and σ^2

$$E[X] = E[e^Y] = e^{\mu+\sigma^2/2}, \quad E[X^2] = E[e^{2Y}] = e^{2\mu+2\sigma^2}$$

giving that

$$\text{Var}(X) = e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2}$$

(c) Taking the sample mean and variance of the logs of the data, yields the estimates that $\hat{\mu} = 3.7867$, $\hat{\sigma^2} = .0647$. Hence, the estimate of $E[X]$ is $e^{\hat{\mu}+\hat{\sigma^2}/2} = 45.561$.

8. $\bar{X} = 3.1502$

(a) $3.1502 \pm 1.96(.1)/\sqrt{5} = (3.0625, \ 3.2379)$

(b) $3.1502 \pm 12.58(.1)/\sqrt{5} = (3.0348, \ 3.2656)$

9. $\bar{X} = 11.48$

(a) $11.48 \pm 1.96(.08)/\sqrt{10} = 11.48 \pm .0496$

(b) $(-\infty, 11.48 + 1.645(.08)/\sqrt{10}) = (-\infty, \ 11.5216)$

(c) $(11.48 - 1.645(.08)/\sqrt{10}, \infty) = (11.4384, \infty)$

10. $74.6 \pm 1.645(11.3)/9 = 74.6 \pm 2.065 = (72.535, \ 76.665)$

11. (a) Normal with mean 0 and variance $1 + 1/n$

(b) With probability .9, $-1.64 < (X_{n+1} - \bar{X}_n)/\sqrt{1 + 1/n} < 1.64$. Therefore, with 90 percent confidence, $X_{n+1} \in \bar{X}_n \pm 1.64\sqrt{1 + 1/n}$.

12. $P\{\sqrt{n}(\mu - \bar{X})/\sigma < z_\alpha\} = 1 - \alpha$ and so $P\{\mu < \bar{X} + z_\alpha \sigma/\sqrt{n}\} = 1 - \alpha$

13. $1.2 \pm z_{.005}0.2/\sqrt{20}$ or $(1.0848,\ 1.3152)$

14. $1.2 \pm t_{.005,19}.2/\sqrt{20}$ or $(1.0720,\ 1.3280)$

15. $1.2 \pm t_{.01,19}.2/\sqrt{20} = 1.31359$

16. The size of the confidence interval will be $2t_{\alpha/2,n-1}S_n/\sqrt{n} \simeq 2z_{a/2}\sigma/\sqrt{n}$ for n large. First take a subsample of size 30 and use its sample standard deviation, call it σ_e, to estimate σ. Now choose the total sample size n (= 30 + additional sample size) to be such that $2z_{\alpha/2}\sigma_e/\sqrt{n} \leq A$. The final confidence interval should now use all n values.

17. Run program 7-3-1. The 95 percent interval is (331.0572, 336.9345), whereas the 99 percent interval is (330.0082, 337.9836).

18. (a) (128.14, 138.30)
(b) $(-\infty, 129.03)$
(c) $(137.41, \infty)$

19. Using that $t_{.05,8} = 1.860$, shows that, with 95 percent confidence, the mean price is above

$$\begin{aligned} 122,000 - 1.860(12,000)/3 &= 129,440 \\ &\ 114,560 \end{aligned}$$

20. $2,200 \pm 1.753(800)/4 = 2,200 \pm 350.6$

21. Using that $t_{.025,99} = 1.985$, show that, with 95 percent confidence, the mean score is in the interval $320 \pm 1.985(16)/10 = 320 \pm 3.176$

22. $330.2 \pm 2.094(15.4)/\sqrt{20}$, $\quad 330.2 \pm 2.861(15.4)/\sqrt{20}$ where the preceding used that $t_{.025,19} = 2.094$, $t_{.005,19} = 2.861$.

23. $1220 \pm 1.968(840)/\sqrt{300}$, since $t_{.025,299} = 1.968$

24. $1220 \pm 1.284(840)/\sqrt{300}$, since $t_{.10,299} = 1.284$

26. (a) (2013.9, 2111.6), (b) (1996.0, 2129.5) (c) 2022.4

27. (93.94, 103.40)

28. (.529, .571)

29. $E[N] = e$

30. (10.08, 13.05)

31. (85, 442.15, 95, 457.85)

32. $X_{n+1} - \bar{X}_n$ is normal with mean 0 and variance $\sigma^2 + \sigma^2/n$. Hence, $E[(X_{n+1} - \bar{X}_n)^2] = \sigma^2(1 + 1/n)$.

33. (3.382, 6.068)

34. 3.007

36. 32.23, (12.3, 153.83, 167.2), 69.6

37. (.00206, .00529)

38. $\frac{2}{3}S_1^2 + \frac{1}{3}S_2^2 = 4$

39. .008

40. Use the fact that nT^2/σ^2 is chi-square with n degrees of freedom, where $T^2 = \sum_{i=1}^{n}(X_i - \mu)^2/n$. This gives that $\sqrt{n}(\bar{X} - \mu)/T$ is a t-random variable with n degrees of freedom. The confidence interval is $(\bar{X} - t_{\alpha_2,n}T/\sqrt{n}, \bar{X} + t_{\alpha_2,n}T/\sqrt{n})$. The additional degree of freedom is like having an extra observation.

41. $(-22.84, 478.24)$, $(20.91, \infty)$, $(-\infty, 434.49)$

42. $(-11.18, -8.82)$

43. $(-11.12, -8.88)$

44. $(-74.97, 41.97)$

45. Using that

$$\frac{S_y^2/\sigma_2^2}{S_x^2/\sigma_1^2}$$

has an F-distribution with parameters $m-1$ and $n-1$ gives the following $100(1-\alpha)$ percent confidence interval for σ_1^2/σ_2^2

$$\left(F_{1-\alpha/2,m-1,n-1}S_x^2/S_y^2, \quad F_{\alpha/2,m-1,n-1}S_x^2/S_y^2\right)$$

47. (a) $.396 \pm .024$ $[.372, .42]$
(b) $.389 \pm .037$ $[.352, .426]$

48. $.17 \pm .018$, $.107 \pm .015$

49. $.5106 \pm .010$, $.5106 \pm .013$

50. 1692

51. no

52. 2401

53. $z_{.005}\sqrt{79 \times 61/140}/140 = .108$

54. $.17 \pm \sqrt{17 \times .83}/100 = (.096, .244)$, $(.073, .267)$

55. .67 An additional sample of size 2024 is needed.

57. (21.1, 74.2)

58. Since

$$P\left\{2\sum X_i/\theta > \chi^2_{1-\alpha,2n}\right\} = 1-\alpha$$

it follows that the lower confidence interval is given by

$$\theta < 2\sum X_i \Big/ \chi^2_{1-\alpha,2n}$$

Similarly, a $100(1-\alpha)$ percent upper confidence interval for θ is

$$\theta > 2\sum X_i \Big/ \chi^2_{\alpha,2n}$$

60. Since $\text{Var}[(n-1)S_x^2/\sigma^2] = 2(n-1)$ it follows that $\text{Var}(S_x^2) = 2\sigma^4/(n-1)$ and similarly $\text{Var}(S_y^2) = 2\sigma^4/(n-1)$. Hence, using Example 5.5b which shows that the best weights are inversely proportional to the variances, it follows that the pooled estimator is best.

61. As the risk of d_1 is 6 whereas that of d_2 is also 6 they are equally good.

62. Since the number of accidents over the next 10 days is Poisson with mean 10λ it follows that $P\{83|\lambda\} = e^{-10\lambda}(10\lambda)^{83}/83!$. Hence,

$$f(\lambda|83) = \frac{P\{83|\lambda\}e^{-\lambda}}{\int P\{83|\lambda\}e^{-\lambda}d\lambda} = c\lambda^{83}e^{-11\lambda}$$

where c does not depend on λ. Since this is the gamma density with parameters 84.11 it has mean $84/11 = 7.64$ which is thus the Bayes estimate. The maximum likelihood estimate is 8.3. (The reason that the Bayes estimate is smaller is that it incorporates our initial belief that λ can be thought of as being the value of an exponential random variable with mean 1.)

63. $$\begin{aligned} f(\lambda|x_1\ldots x_n) &= f(x_1\ldots x_n|\lambda)g(\lambda)/c \\ &= c\lambda^n e^{-\lambda\sum x_i}e^{-\lambda}\lambda^2 \\ &= c\lambda^{n+2}e^{-\lambda}(1+\sum x_i) \end{aligned}$$

where $c \times p(x_1\ldots x_n)$ does not depend on λ. Thus we see that the posterior distribution of λ is the gamma distribution with parameters $n+3.1+\sum x_i$: and so the Bayes estimate is $(n+3)/(1+\sum x_i)$, the mean of the posterior distribution. In our problem this yields the estimate 23/93.

64. The posterior density of p is, from Equation (5.5.2)
$f(p|\text{data}) = 11!p^i(1-p)^{10-i}/1!(10-i)!$ where i is the number of defectives in the sample of 10. In all cases the desired probability is obtained by integrating this density from p equal 0 to p equal .2. This has to be done numerically as the above does not have a closed form integral.

65. The posterior distribution is normal with mean $80/89(182) + 9/89(200) = 183.82$ and variance $36/89 = .404$. Therefore, with probability $.95, \theta \in 183.82 \pm z_{.025}\text{sqr}(.404)$. That is, $\theta \in (182.57, 185.07)$ with probability .95.

Chapter 8

1. (a) The null hypothesis should be the defendant is innocent.
(b) The significance level should be relatively small, say $\alpha = .01$.

2. If the selection was random, then the data would constitute a sample of size 25 from a normal population with mean 32 and standard deviation 4. Hence, with Z being a standard normal

$$\begin{aligned} p\text{-value} &= P_{H_0}\left\{|\bar{X} - 32| > 1.6\right\} \\ &= P_{H_0}\left\{\frac{|\bar{X} - 32|}{4/5} > 2\right\} \\ &= P\{|Z| > 2\} \\ &= .046 \end{aligned}$$

Thus the hypothesis that the selection was random is rejected at the 5 percent level of significance.

3. Since $\sqrt{n}/\sigma = .4$, the relevant p-values are
(a) $P_{H_0}\{|\bar{X} - 50| > 2.5\} = P\{|Z| > 1\} = .3174$

(b) $P_{H_0}\{|\bar{X} - 50| > 5\} = P\{|Z| > 2\} = .0455$

(c) $P_{H_0}\{|\bar{X} - 50| > 7.5\} = P\{|Z| > 3\} = .0027$

4. $\bar{X} = 8.179$ p-value $= 2[1 - \phi(3.32)] = .0010$ Rejection at both levels

5. $\bar{X} = 199.125$ p-value $= \phi(-.502) = .3078$ Acceptance at both levels

6. $\bar{X} = 72.015$ p-value of the test that the mean is 70 when the standard deviation is 3 is given by p-value $= 2[1 - \phi(3.138)] = .0017$ Rejection at the 1% level of significance.

7. (a) Reject if $|\bar{X} - 8.20|\sqrt{n}/.02 > 1.96$
(b) Using (8.3.7) need $n = 6$
(c) Statistic in $(a) = 13.47$ and so reject
(d) probability $\simeq 1 - \phi(-12.74) \simeq 1$

8. If $\mu_1 < \mu_0$ then $\phi[\sqrt{n}(\mu_0 - \mu_1)/\sigma + z_{\alpha/2}] > \phi(z_{\alpha/2}) = 1 - \alpha/2 \simeq 1$. Thus, from (8.3.5)

$$1 - \phi[\sqrt{n}(\mu_0 - \mu_1)/\sigma - z_{\alpha/2}] \simeq \beta$$

and so

$$\sqrt{n}(\mu_0 - \mu_1)/\sigma - z_{\alpha/2} \simeq z_\beta$$

9. The null hypothesis should be that the mean time is greater than or equal to 10 minutes.

10. p-value $= P_{7.6}\{\bar{X} \leq 7.2\} = P\left\{Z \leq \frac{4}{1.2}(-.4)\right\} = P\{Z > 1.33\} = .0913$. Thus the hypothesis is rejected at neither the 1 nor the 5 percent level of significance.

11. The p-values are as follows.
(a) $P_{100}\{\bar{X} \geq 105\} = P\{Z \geq 5(\sqrt{20}/5)\} = P\{Z > 4.47\} \approx 0$
(b) $P_{100}\{\bar{X} \geq 105\} = P\{Z \geq 5(\sqrt{20}/10)\} = P\{Z > 2.236\} = .0127$
(c) $P_{100}\{\bar{X} \geq 105\} = P\{Z \geq 5(\sqrt{20}/15)\} = P\{Z > 1.491\} = .068$

12. Testing the null hypothesis that the mean number of cavities is at the least 3 gives

$$\begin{aligned} p\text{-value} &= P_3\{\bar{X} \leq 2.95\} \\ &= P_3\{\sqrt{n}(\bar{X} - 3) \leq -.05\sqrt{n}\} \\ &= P\{Z > .05(50)\} = .0062 \end{aligned}$$

Thus we can conclude that the new toothpaste results, on average, in fewer than 3 cavities per child. However, since it also suggests that the mean drop is of the order of .05 cavities, it is probably not large enough to convince most users to switch.

13. With T_{24} being a t-random variable with 24 degrees of freedom

$$p\text{-value} = P\{|T_{24}| \geq 5|19.7 - 20|/1.3\} = 2P\{T_{24} \geq 1.154\} = .26$$

14. With T_{24} being a t-random variable with 35 degrees of freedom

$$p\text{-value} = P\{|T_{35}| \geq 6|22.5 - 24|/3.1\} = 2P\{T_{35} \geq 2.903\} = .0064$$

15. With T_{27} being a t-random variable with 27 degrees of freedom

$$p\text{-value} = P\{|T_{27}| \geq \sqrt{28}|1 - .8|/.3\} = 2P\{T_{27} \geq 3.528\} = .0016$$

17. The p-value of the test of the null hypothesis that the mean temperature is equal to 98.6 versus the alternative that it exceeds this value is

$$p\text{-value} = P\{T_{99} \geq 10(98.74 - 98.6)/1.1\} = P\{T_{99} \geq 1.273\} = .103$$

Thus, the data is not strong enough to verify, even at the 10 percent level, the claim of the scientist.

20. p-value $= p\{T_9 < -3.25\} = .005$

21. p-value $= P\{T_{17} < -1.107\} = .142$

22. p-value $= .019$, rejecting the hypothesis that the mean is less than or equal to 80.

23. No, it would have to have been greater than .192 to invalidate the claim.

24. p-value $= P\{T_{15} < -1.847\} = .04$

25. no, yes.

26. The data neither prove nor disprove the manufacture's claim. The p-value obtained when the claim is the alternative hypothesis is .237.

27. Yes, the test statistic has value 4.8, giving a p-value near 0.

28. p-value $= P\{|Z| > .805\} = .42$

29. .004, .018, .092

30. p-value $= 2P\{T_{13} > 1.751\} = .1034$

31. p-value $= 2P\{T_{11} > .437\} = .67$

32. p-value $= P\{T_{10} > 1.37\} = .10$

33. p-value = .019

34. yes, p-value $= .004$

35. p-value $= P\{T_{30} > 1.597\} = .06$ The professor's claim, although strengthened by the data, has not been proven at, say, the 5 percent level of significance.

36. p-value = .122

37. p-value = .025

38. The value of the test statistics is 1.15, not enough to reject the null hypothesis.

39. The value of the test statistic is 8.2, giving a p-value approximately equal to 0.

40. The value of the test statistic is .87, with a resulting p-value of .39.

41. p-value (test statistic $= 7.170$)

42. p-value $= 2P\{T_9 > 2.333\} = .044$ The hypothesis of no change is rejected at the 5% level of significance.

43. For a 2-sided test of no effect p-value $= 2P\{T_7 > 1.263\} = .247$ and we cannot conclude on the basis of the presented data that jogging effects pulse rates.

44. Reject at the α level of significance if $(n-1)S^2/\sigma_0^2 > X^2_{\alpha,n-1}$. Equivalently, if the values of $(n-1)S^2/\sigma_0^2$ is v then the p-value $= P\{X^2_{n-1} > v\}$.

45. Reject at the α level if $\sum(X_1 - \mu)^2/\sigma_0^2 > X^2_{\alpha,n}$

46. Test the null hypothesis that $\sigma \geq .1$. The values of the test statistic is $(n-1)S^2/.01 = 49 \times .0064/.01 = 31.36$ and so p-value $= P\{X^2_{49} < 31.36\} = .023$. Hence, the hypothesis that $\sigma \geq 1$ is rejected and the apparatus can be utilized.

47. Test $H_0 : \sigma \geq .4$ $9S^2/(.4)^2 = 9.2525 \times 10^4$ p-value $= P\{X^2_9 < .000925\} < .0001$. Hence, the null hypothesis that the standard deviation is as large as .4 is rejected and so the new method should be adopted.

48. $S_1^2/S_2^2 = .53169$ p-value $= 2P\{F_{7.7} < .53169\} = .42$ and so the hypothesis of equal variances is accepted.

49. $S_1^2/S_2^2 = 14.053$ p-value $= 2P\{F_{5.6} > 14.053\} = .006$ and the hypothesis of equal variances is rejected.

50. $\sigma_y^2 S_x^2 + \sigma_x^2 S_y^2$ has an F-distribution with $n-1$ and $m-1$ degrees of freedom. Hence, under H_0, $P\{S_x^2/S_y^2 > F_{\alpha,n-1,m-1}\} \leq P\{F_{n-1,m-1} > F_{\alpha,n-1,m-1}\} = \alpha$ and so the test is to reject if $S_x^2/S_y^2 > F_{\alpha,n-1,m-1}$ or, equivalently, we could compute S_x^2/S_y^2, call its value v, and determine p-value $= P\{F_{n-1,m-1} > v\}$.

51. Test $H_0 : \sigma^2_{\text{in}} \leq \sigma^2_{\text{out}}$ against $H_1 : \sigma^2_{\text{in}} > \sigma^2_{\text{out}}$ $S^2_{\text{out}}/S^2_{\text{in}} = .4708$ p-value $= P\{F_{74.74} < .4708\} = 7.5 \times 10^{-4}$ by 3-8-3-a and so conclude that the variability is greater on the inner surface.

52. The test statistic has value 5, giving a p-value approximately 0.

53. The test statistic has value 1.43 which is not large enough to reject the null hypothesis that the probability of stroke is unchanged

54. p-value $= P\{\text{Bin}(50, .72) \geq 42\} = .036$

55. (a) No, since p-value $= P\{\text{Bin}(100, .5) \geq 56\} = .136$
(b) No, since p-value $= P\{\text{Bin}(120, .5) \geq 68\} = .085$
(c) No, since p-value $= P\{\text{Bin}(110, .5) \geq 62\} = .107$
(d) Yes, since p-value $= P\{\text{Bin}(330, .5) \geq 186\} = .012$

56. (a) If the probability that a birth results in twins is .0132 then the mean number of twin births will be 13.2 with a variance equal to 13.02576. As the standard deviation is 3.609. Because a normal random variable would be greater than its mean by at least 1.96 of its standard deviations is .025 it would seem that 6 or fewer twin births would result in rejection. An exact calculation yields that $P(\text{Bin}(1000, .0132) \leq 6) = .02235$, and so the null hypothesis would be rejected if there were 6 or fewer births.
(b) When the null hypothesis is true the exact probability of getting at least 21 twin births is .02785. Because $.02785 + .02235 \approx .05$, the test can be to reject when either there are 6 or fewer or 21 or more twins births. Thus, for X

being a binomial with paramters (1000, .0180, te answer, to 4 decimal places, is $P(X \geq 21) + P(X \leq 6) = .2840 + .0008 = .2848$.

57. The claim is believable at neither level, since

$$p\text{-value} = \mathrm{P}\{\mathrm{Bin}(200, .45) \geq 70\} = .003$$

58. p-value $= 2P\{\mathrm{Bin}(50, 3/4) \geq 42\} = .183$

59. $p\text{-value} = 2P\left\{Z > \dfrac{41.5 - 150/4}{\sqrt{150/16}}\right\} = .19$

60. Using the Fisher-Irwin conditional test, the p-value is twice the probability that a hypergeometric random variable X, equal to the number of red balls chosen when a sample of 83 balls is randomly chosen from a collection of 84 red and 72 blue balls, is at most 44. Because

$$E[X] = 83(84)/156 = 44.69$$

and

$$\sqrt{\mathrm{Var}(X)} = \sqrt{83 \cdot \frac{84}{156}\left[1 - \frac{82}{155}\right]} = 4.59$$

it is clear that the p-value is quite large and so the null hypothesis would not be rejected.

(b) We need test that $p = .5$ when a total of 156 trials resulted in 84 successes and 72 failures. Wit X being a binomial random variable with parameters $n = 156, p = .5$, the p-value is given by

$$\begin{aligned} p\text{-value} &= 2P(X \geq 84) \\ &= 2P(X \geq 83.5) \\ &= 2P\left(\frac{X - 78}{\sqrt{39}} \geq \frac{83.5 - 78}{\sqrt{39}}\right) \\ &\approx 2P(Z \geq .8807) \\ &\approx .38 \end{aligned}$$

Thus the data is consistent with the claim that the determination of the treatment to be given to each patient was made in a totally random fashion.

62. $$\frac{(n_1 - i)!i!(n_2 - k + i)!(k - i)!}{(n_1 - i - 1)!(i + 1)!(n_2 - k + i + 1)!(k - i - 1)!} = \frac{(n_1 - i)(k - i)}{(i + 1)(n_2 - k + i + 1)}$$

63. Let $Y = X_1/n_1 + X_2/n_2$. Then $E[Y] = p_1 + p_2$ and $\mathrm{Var}(Y) = p_1(1 - p_1)/n_1 + p_2(1 - p_2)/n_2$. By the normal approximation, to the binomial it follows that Y is approximately normally distributed and so (a) follows. Part (b) follows since the proposed estimate of $p_1 = p_2$ is just the proportion of the $\mathrm{n}_1 + \mathrm{n}_2$ trials that result in successes.

64. p-value $= 2[1 - \phi(1.517443)] = .129$

65. p-value $= P\{|Z| > 2.209\} = .027$, indicating that the way in which the information was presented made a difference.

66. (a) Assuming independence of the two samples, the value of the normal approximation test statistic is 1.57, giving

$$p\text{-value} = 2P(Z > 1.57) = .1164$$

(b) The value of the normal approximation test statistic is .552, giving

$$p\text{-value} = 2P(Z > .522) = .1602$$

67. The value of the normal approximation test statistic is .375. Thus, the hypothesis cannot be rejected any reasonable significance level.

68. p-value $= 2P\{Po(416) \geq 431\} = .47$

69. The p-value is $P(X \geq 27)$ where X is Poisson with mean 6.7. Because the standard deviation of X is $\sqrt{6.7}$, X would have to exceed its mean by about 6 of its standard deviations, which has a miniscule probability of occurring.

70. p-value $= 2P\{\text{Bin}375,\ 3/11) \geq 119\} = .063$

71. The scientist should try to match her samples, so that for each smoker there is a nonsmoker of roughly the same age.

72. No because the researcher will only be considering stocks that have been around for the past 20 years, and is thus ignoring those that were in play 20 years ago but have since gone bust.

Chapter 9

1. $y = 2.464 + 1.206x$

2. $y = 206.74 - 2.376x$; the estimated response at $x = 25$ is 147.34

3. $y = .0072 + .0117x$; the estimated response at $x = 32$ is .0448

4. $y = 2826.1 + 12246.2x$; the estimated response at $x = .43$ is 2439.8

5. $y = 2.64 + 11.80x$; the estimated response at $x = 7$ is 85.22

6. 57.1 percent

8. $A = \sum Y_i/n - B\bar{x}$

$\text{Var}(A) = \text{Var}\left(\sum Y_i/n\right) + \bar{x}^2\text{Var}(B) - \frac{2\bar{x}}{n}\text{Cov}\left(\sum Y_i \cdot B\right)$

Now, $\text{Cov}\left(Y_j,\ B\right) = \text{Cov}\left(Y_j,\ \dfrac{1}{c}\sum(x_1 - \bar{x})Y_i\right)$ where $c = \sum x_i^2 - n\bar{x}^2$.

$= (x_j - \bar{x})\sigma^2/c$

Hence,

$$\text{Cov}\left(\sum Y_j,\ B\right) = \sum(x_j - \bar{x})\sigma^2/c = 0$$

and so

$$\text{Var}(A) = \sigma^2/n + \frac{\bar{x}^2\sigma^2}{\sum x_i^2 - n\bar{x}^2} = \frac{\sigma^2\sum x_i^2}{n\left[\sum x_i^2 - n\bar{x}^2\right]}$$

9. (a) $SS_R/8 = 105659.1$ (b) $(SS_R/X^2_{.05,8},\ SS_R/X^2_{.95,8}) = (54518, 308521)$

10. $SS_R = \sum\left(Y_j - \bar{Y} + \bar{Y} - A - Bx_j\right)^2$

$= S_{YY} + \sum\left(\bar{Y} - A - Bx_j\right)^2 + 2\sum\left(Y_j - \bar{Y}\right)\left(\bar{Y} - A - Bx_j\right)$

$= S_{YY} + B^2S_{xx} - 2BS_{xY}$ since $A = \bar{Y} - B\bar{x}$

$= S_{YY} - S_{xY}^2/S_{xx}$ since $B = S_{xY}/S_{xx}$

11. $y = 46.44 + .0481x$; p-value when testing $\beta = 0$ is $P\{|T_{12} > 2.8\} = .016$. Confidence interval for α is (42.93, 49.95)

12. The p-value when testing $\beta = 0$ is $P\{|T_{10}| > 1.748\} = .11\}$, not small enough to establish the hypothesis.

14. .239, .265, .283

15. The very fine and very poor landings might just have been chance results; the following outcomes would then be more normal even without any verbal remarks.

16. It follows from the text that

$$Z = \frac{(A-\alpha)\sqrt{nS_{xx}}}{\sigma\sqrt{\sum x_i^2}}$$

has a standard normal distribution, and thus $z/\left(\sqrt{SS_R/[(n-2)\sigma^2]}\right)$ has a t-distribution with $n-2$ degrees of freedom.

17. (b) $y = -.664 + .553x$
(c) p-value $= P\{|T_8| > 2.541\} = .035$
(d) 12.603
(e) (11.99, 13.22)

18. 510.081, 517.101, 518.661, 520.221

26. (b) $y = -22.214 + .9928x$ (c) p-value $= 2P\langle T_8 > 2.73\rangle = .026$ (d) 2459.7
(e) $2211 \pm 10.99xt_{.025,8} = (2186.2, 2236.9)$

28. $y = -3.6397 + 4.0392x$ at $x = 1.52$ $y = 2.4998$
95 percent confidence interval $= 2.4998 \pm .00425$

29. (a) $d/dB\sum(Y_i - Bx_i)^2 = -2\sum x_i(Y_i - Bx_i)$. Equating to 0 yields that the least squares estimator is $B = \sum x_iY_i\Big/\sum x_i^2$

(b) B is normal with mean $E[B] = \sum x_iE[Y_i]\Big/\sum x_i^2 = \beta$ (since $E[Y] = \beta x_i$) and variance $\text{Var}(B) = \sum x_i^2\text{Var}(Y_i)\Big/\left(\sum x_i^2\right)^2 = \sigma^2\Big/\sum x_i^2$

(c) $SS_R = \sum(Y_i - Bx_i)^2$ has a chi-square distribution with $n-1$ degrees of freedom.

(d) $\text{Sqr}\left(\sum x_i^2\right)(B-\beta_0)/\sigma$ has a unit normal distribution when $\beta = \beta_0$ and so $V \equiv \text{Sqr}\left(\sum x_i^2\right)(B-\beta_0)\ \big/\text{Sqr}[SS_R/(n-1)]$ has a t-distribution with $n-1$ degrees of freedom when $\beta = \beta_0$. Hence, if the observed value of $|V|$ is $V = v$ then p-value $= 2P(T_{n-1} > v)$ where T_{n-1} has a t-distribution with $n-1$ degrees of freedom.

(e) $Y - Bx_0$ is normal with mean 0 and variance $\sigma^2 + x_0^2\sigma^2\Big/\sum x_i^2$ and so $-t_{\alpha/2,n-1} < \dfrac{Y - Bx_0}{\text{Sqr}\left[\left(1+x_0^2\Big/\sum x_i^2\right)SS_R/(n-1)\right]} < t_{\alpha_2,n-1}$ with confidence $100(1-\alpha)$

31. (a) $A = 68.5846$ $B = .4164$ (b) p-value $< 10^{-4}$ (c) 144.366 ± 4.169
(e) R = .7644

32. Take logs to obtain $\log S = \log A - m \log N$ or $\log N = (1/m) \log A - (1/m) \log S$. Fitting this with a regression line with $\log N = 55.59 - 14.148 \log S$ which yields that $m = .0707$ and $A = 50.86$.

33. Taking logs and using Program 9-2 yields the estimates $\log t = 3.1153 \quad \log s = .0924$ or $t = 22.540$ and $s = 1.097$

34. Taking logs and letting time be the independent variable yields, upon running Program 9-2, the estimates $\log a = .5581$ or $a = 1.7473$ and $b = .0239$. The predicted value after 15 hours is 1.22.

35. Using the results of Problem 21a on the model $\log(1 - P) = -\alpha t$ yields the estimate $\alpha = 1.05341$. Solving the equation $1/2 = e^{-\alpha t}$ yields $t = \log 2/\alpha = .658$.

36. With Y being the bacterial count and x the days since inoculation $Y = 64777e^{.1508x}$

37. The normal equations are

$$\begin{aligned} 9.88 &= 10\alpha + 55\beta + 385\gamma \\ 50.51 &= 55\alpha + 385\beta + 3025\gamma \\ 352.33 &= 385\alpha + 3025\beta + 25333\gamma \end{aligned}$$

which yield the solution: $\alpha = 1.8300 \quad \beta = -.3396 \quad \gamma = .0267$

39. (a) $y = -46.54051 + 32.02702x$

40. $y = .5250839 + 12.14343x$ at $\quad x = 7 \quad y = 85.52909$

41. $y = 20.23334 + 3.93212x$ using ordinary least squares
$y = 20.35098 + 3.913405x$ using weighted least squares

42. (a) The weighted least squares fit is $y = -4.654 + .01027x$ at $x = 3500$ $y = 31.29$
(b) The variance stabilizing transformation yields the least squares solution $\sqrt{y} = 2.0795 + .00098x$ at $\quad x = 3500 \quad y = 30.35$

43. Peak Discharge $= 150.1415 + .362051x_1 - 3163.567x_2$

44. $y = -1.061 + .252x_1 + 3.578 \times 10^{-4}x_2$

45. $y = -606.77 + 59.14x_1 - 111.64x_2 + 14.00x_3 - 19.25x_4 \quad SS_R = 1973396$

46. $\log(\text{survival}) = 7.956696 - 1.204655x_1 - .02250433x_2 \quad SS_R/9 = 2.453478 =$ est. of σ^2

47. (a) $y = -2.8277 + 5.3707x_1 + 9.8157x_2 + .448147x_3 \quad SS_R = 201.97$

(b) p-value $= 2P(T_{11} > .7635) = .46$
(c) p-value $= 2P(T_{11} > .934) = .37$
(d) p-value $= 2P(T_{11} > 1.66) = .125$

48. (a) $y = 177.697 + 1.035x_1 + 10.721x_2$
(b) 238.03 ± 3.94

49. (a) $y = 1108.68 + 8.64x_1 + .26x_2 - .71x_3$
(b) $SS_R/6 = 520.67$
(c) 2309.6 ± 28.8

50. A prediction interval is always larger than the corresponding confidence interval for the mean since it has to take into account not only the variance in the estimate of the mean but also the variance of an observation. For instance, if one had an infinite number of observations then the confidence interval for a mean response would shrink to a single point whereas a prediction interval for an observation would still involve σ^2 the variance of an observation even when its mean is known.

51. (a) $y = 5.239 + 5.697x_1 + 9.550x_2$
(b) $SS_R/9 = 68.82$
(c) 225.70 ± 20.07

52. (a) $y = 6.144 - 3.764 \times 10^{-2}x_1 + 8.504 \times 10^{-2}x_2$
(b) p-value $= 2P\{T_9 > 12.4\} \approx 0$
(c) $4.645 \pm .615$

53. $y = 28.210 + .116x_1 + .566x_2$
(a) p-value of "$\beta_1 = 0$" $= 2P\{T_6 > .2487\} = .81$
(b) 135.41 ± 17.24 or $(118.17, 152.65)$

Chapter 10

1. F-statistic $= .048737$ p-value $= .954$

2. F-statistic $= .32$ p-value $= .727$

3. The resulting test statistics would have complicated dependencies.

4. $F = 10.204$ p-value = .00245

5. $F = 7.4738$ p-value = .0043

6. $\sum_{i=1}^{n}(X_i - \mu)^2/\sigma^2 = \sum(X_i - \bar{X})^2/\sigma^2 + n(\bar{X} - \mu)^2/\sigma^2$. As the first term of the right side of the equality sign has $n - 1$ and the second 1 degree of freedom the result follows.

7. The value of the test statistics is 1.332, with a corresponding p-value of .285; thus the hypothesis is not rejected at either significance level.

8. Since $S_i^2 = \sum_{j=1}^{n}(X_{ij} - X_{i.})^2/(n - 1)$, it follows that

$$SS_w = \sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} - X_{i.})^2 = (n - 1)\sum_{i=1}^{m} S_i^2$$

9. The result of Problem 8 shows that

$$SS_w = 9[24 + 23.2 + 17.1] = 578.7$$

Since a simple computation yields that $SS_b = 388.866$, the value of the test statistic is

$$T = \frac{388.866/2}{578.7/27} = 9.072$$

As $F_{.05,2,27} = 3.35$ the hypothesis is rejected.

10. The value of the test statistic is 5.08, with a corresponding p-value of .01; thus the hypothesis is rejected at the 5 percent significance level.

11. The value of the test statistic is 5.140. Since $F_{.01,4,60} = 3.65$ the hypothesis of equal levels is rejected even at the 1 percent level of significance.

13. The value of the test statistic is .1666, with a corresponding p-value of .849; thus the hypothesis of equal fat content is not rejected.

14. The value of the test statistic is .07, with a corresponding p-value of .934; thus, the data is consistent with the hypothesis.

17. 37 for both parts.

19. Use a two-way analysis of variance model.

20.

$$\begin{aligned}
\mu &= 31.545\\
\hat{\alpha}_1 &= .180\\
\hat{\alpha}_2 &= -1.1295\\
\hat{\alpha}_3 &= .130\\
\hat{\alpha}_4 &= .205\\
\hat{\alpha}_5 &= .780\\
\hat{\beta}_1 &= 3.075\\
\hat{\beta}_2 &= -.065\\
\hat{\beta}_3 &= -1.505\\
\hat{\beta}_4 &= -1.505
\end{aligned}$$

The p-value for the hypothesis that the season is irrelevant is .027 (test statistic value 5.75), and the p-value for the hypothesis that the year has no effect is .56 (test statistic value .793); hence the first hypothesis is rejected and the second accepted at the 5 percent level.

21. The p-value for the hypothesis that the methods of extraction are equivalent is .001, and the p-value for the hypothesis that storage conditions have no effect is .017; hence both hypotheses are rejected at the 5 percent level.

22. (a) –2, –4.67, 3.33
(b) 6.25, 7.75

The p-value for the hypothesis that the detergent used has no effect is .011 (test statistic value 9.23), and the p-value for the hypothesis that the machine used had no effect is .0027 (test statistic value 18.29); hence both hypotheses are rejected at the 5 percent level.

23. F-stat for rows $= .3798$ $\quad p$-value $= .706$
F-stat for columns $= 11.533$ $\quad p$-value $= .0214$

24. F-stat for rows $= 2.643$ p-value $= .1001$
F-stat for columns $= .0155$ p-value $= .9995$
F-stat for interaction $= 2.5346$ p-value $= .1065$

25. F-stat for rows $= .0144$ p-value $= .9867$
F-stat for columns $= 34.0257$ p-value $< .0001$
F-stat for interaction $= 2.7170$ p-value $= .0445$

26. F-stat for rows $= 4.8065$ p-value $= .028$
F-stat for columns $= 50.406$ p-value $< .0001$
F-stat for interactions $= 3.440$ p-value $= .0278$

27. F-stat for rows $= 11.0848$ p-value $= .0003$
F-stat for columns $= 11.1977$ p-value $= .0003$
F-stat for interactions $= 7.0148$ p-value $= .00005$

28. F-stat for rows $= .3815$ p-value $= .5266$
F-stat for columns $= .3893$ p-value $= .7611$
F-stat for interactions $= .1168$ p-value $= .9497$

(d) Using an F-statistic to see if there is a placebo effect yields the value 11.8035 for the statistic; with a corresponding p-value of .0065. This, the hypothesis of no placebo effect is rejected.

Chapter 11

1. $T = .8617$ p-value $= .648$ accept

2. $T = 2.1796$ p-value $= .824$ accept

3. $T = 15.955$ p-value $= .143$

4. $T = 2.1144$ p-value $= .55$

5. $T = 23.13$ p-value $= .00004$

6. $T = 43.106$ p-value $= .0066$

7. $T = 37.709$ using 6 regions p-value $< .00001$

8. $TS = 4.063$, p-value $= .131$

9. $TS = 4.276$, p-value $= .639$

10. $TS = .0016$, p-value $= .968$
    The probability that a fit as good or better than obtained would occur by chance is .032.

13. $TS = 19.295$, p-value $= .0017$

16. $T = 3.4286$ p-value $= .052$

17. $T = 6.8571$ p-value $= .007$

18. $T = 4327.9$ p-value $< 10^{-6}$

19. $T = 16.4858$ p-value $= .0003$

20. $TS = 1.250$, p-value $= .535$

21. $TS = .186$, p-value $= .666$

22. $TS = 9.442$, p-value $= .024$

23. $TS = 5.526$, p-value $= .063$

24. $TS = 27.370$, p-value $= .00005$

Chapter 12

1. p-value $= 2P\{\text{Bin}(18, .5) \leq 5\} = .096$

2. p-value $= i$

3. (a) p-value $= .056$
(b) p-value $= 7.8 \times 10^{-5}$
(c) p-value $= 1.12 \times 10^{-9}$

4. yes, p-value $= .0028$

5. p-value $= .6$

6. (a) p-value $= .29$
(b) the normal approximation gives p-value $= .0498$

7. p-value in 21 $= 2P\{T \leq 23\} = .0047$
p-value in 22 $= 2P\{T \leq 15\} = .742$

8. (a) p-value $= 2P\{\text{Bin}(11, .5) \leq 2\} = .0654$ (b) p-value $= 2P\{T \leq 13\} = .0424$. Thus, at the 5% level we would accept when the sign test and reject when the using the signed rank test.

9. p-value using sign $= .02$ p-value using sign rank $= .0039$ engineer's claim is upheld.

10. Using sign rank p-value $= 2P\{T \leq 5\} = .0195$ so equivalence is rejected

11. (a) Determine the number of data values less than m_0. If this value is k then p-value $= P\{\text{Bin}(n, .5) \geq k$
(b) Let T be the sign rank statistic. If $T = t$ then p-value $= P_{m0}\{T \leq t\} = P_n(t)$

12. $T = 36$ p-value $= .699$

13. $T = 66$ p-value $= .866$

14. normal approximation gives p-value $= 2P\{Z > .088\} = .93$

15. (b) $2P(Z > 1.26) = .2076$

18. The value of the test statistic is 7587.5 giving that

$$p-value \approx P(\chi_2^2 \geq 4.903) = .0862$$

19. $R = 11 \quad p\text{-value} = .009$

20. sample median $= 122 \quad R = 9 \quad p\text{-value} = .14$

21. Yes, but since you would not necessarily have an equal number of 1's and

Chapter 13

1. Control limits are $35 \pm 9/\sqrt{5}$ which give LCL $= 30.98$ UCL $= 39.03$. Subgroup 3 falls outside these limits.

2. Suppose the mean jumps to 16.2. The probability that the next subgroup falls outside is approximately $P\{X > 14 + 6/\sqrt{5}\} = P\{Z > (6/\sqrt{5}2.2)/(2/\sqrt{5})\} = 1 - \phi(.54) = .2946$. On average, it will take a geometric distributed number of subgroups with mean $1/.2946 = 3.39$ to detect a shift. The result is the same when the mean falls by 2.2.

4. (a) $\overline{\overline{X}} = 14.288$ $\bar{S} = .1952$ LCL $= 14.01$ UCL $= 14.57$
(b) The estimate of σ is $.1952/.94 = .2077$. Hence with $\mu = 14.288$, $\sigma = .2077$ $P\{13.85 < X < 14.75\} = \phi(2.212) - \phi(-2.115) = .969$

5. $\overline{\overline{X}} = 36.02$ $\bar{S} = 4.29$ For $\bar{X}$: LCL $= 29.9$ UCL $= 42.14$
For S: LCL $= 0$ UCL $= 8.96$

6. LCL $= 0$ UCL $= .4077$

7. $\overline{\overline{X}} = 36.23$ $\bar{S} = 5.43$
(a) LCL $= 28.48$ UCL $= 43.98$
(b) LCL $= 0$ UCL $= 11.34$
(c) yes
(e) $P\{25 < X < 45\} = \phi(1.52) - \phi(-1.94) = .91$ when X is normal with mean 36.23 and standard deviation $5.43/.94 (= 5.777)$.

8. $\overline{\overline{X}} = 422$ $\bar{S} = 16.667$ $c(4) = .9213$
(a) LCL $= 394.86$ UCL $= 449.14$
(b) LCL $= 0$ UCL $= 37.77$
(c) $16.667/.9213 = 18.09$
(d) $P\{X < 400\} = \phi(-1.216) = .11$

9. (a) LCL $= 394.37$ UCL $= 467.63$ LCL $= 0$ UCL $= 28.5$
(b) $22.5/.9213 = 24.4$
(c) $P\{400 < X < 460\} = \phi(1.19 - \phi(-1.27) = .78$
(d) $P\{\bar{X} > 467.63\} = P\{Z > -(467.63 - 491)/12.2\} = P\{Z > 1.9\} = .97$

11. $\overline{\overline{X}} = 19.4$ $\bar{S} = 1.7$ $1.7/c(6) = 1.79$
(a) LCL $= 17.2$ UCL $= 21.6$ LCL $= .051$ UCL $= 3.349$
(b) $P\{15 < X < 23\} = \phi(2.011) - \phi(-2.46) = .975$

12. The estimate of p, the probability that assembly is defective, is .0445. Thus, the mean number in a sample of size 100 should fall within $4.45 \pm 3 \times \text{sqr}(4.45 \times .9555)$; which means that there should never be more than 10 defectives, which is satisfied.

13. The estimate of p is $.072 =$ number of defects $+$ number of units. Thus if n are produced on a day then the number of defects should be within $np \pm \text{sqr}\{np(1-p)\}$ where p is taken equal to .072. The data all satisfy this criteria.

14. Control limits are $20 \pm 3\text{sqr}(20 \times .96)$ which gives $\text{UCL} = 33.1$. The desired probability is thus $P\{\text{Bin}(500, .08) \geq 34\} = .86$.

15. $\overline{X} = 90.8$ As $3\sqrt{90.8} = 28.59$ $\text{LCL} = 62.2$ $\text{UCL} = 119.5$. The first 2 data points fall outside. Eliminating them and recomputing gives $\overline{X} = 85.23$, $3\sqrt{85.23} = 27.70$ and so $\text{LCL} = 57.53$ $\text{UCL} = 112.93$. As all points fall within, these limits can be used for future production.

16. $\overline{X} = 3.76$ and so $\text{UCL} = 9.57$. The process appears to have been out of control when number 14 was produced.

Chapter 14

1. $\lambda(t) = \alpha\beta t^{\beta-1}$

2. $P\{Z > t\} = P\{X > t,\ Y > t\} = [1 - F_X(t)][1 - F_Y(t)]$

$$\lambda_2(t) = \frac{[1 - F_X(t)]f_Y(t) + [1 - F_Y(t)]f_X(t)}{[1 - F_X(t)][1 - F_Y(t)]} = \lambda_y(t) + \lambda_x(t)$$

3.
$$\begin{aligned} P\{40 \text{ year old smoker reaches age } t\} &= \exp\left\{-\int_{40}^{t} \lambda(y)dy\right\} \\ &= \exp\left\{-[.027(t-40) + .025(t-40)^5/50000]\right\} \\ &= \begin{cases} .726 & \text{if } t = 50 \\ .118 & \text{if } t = 60 \\ .004 & \text{if } t = 65 \\ .000002 & \text{if } t = 70 \end{cases} \end{aligned}$$

4. $1 - F(t) = e^{-t^4}/4$ (a) $.018316$ (b) $.6109$ (c) $\int_{t>0} e^{-t^4}/4\, dt = 1.277$
(d) $\exp\left\{-\int_1^2 s^3 ds\right\} = e^{-15/4} = .0235$

5. (b) Using the hint

$$\begin{aligned} \lambda(t) &= \left[\int_{s>t} e^{-\lambda(s-t)}(s/t)^{\alpha-1}ds\right]^{-1} \\ &= \left[\int_{s>0} e^{-\lambda u}(1 + u/t)^{\alpha-1}du\right]^{-1} \quad \text{by the substitution } u = s - t \end{aligned}$$

As the integrand in the above is decreasing in t when $\alpha - 1 > 0$ and increasing otherwise the result follows.

6. $f(x) = 1/(b-a),\ a < x < b \quad F(x) = (x-a)/(b-a),\ a < x < b$

$$\lambda(x) = \frac{1/(b-a)}{(b-x)/(b-a)} = 1/(b-x), \qquad a < x < b$$

8. $\tau = 1541.5$ (a) $\tau/10 = 154.15$ (b) $\left(3083/\chi^2_{.025,20}, \chi^2_{.975,20}\right)$ (c) $3083/\chi^2_{.95,20}$
(d) p-value $= 2p\{\chi^2_{20} > 41.107\} = .007$ so reject at $\alpha = .1$.

9. The null hypothesis should be rejected either if $P_{\theta_0}\{2T/\theta_0 \leq v\} \leq \alpha/2$ or if $P_{\theta_0}\{2T/\theta_0 \geq v\} \leq \alpha/2$. As, under H_0, $2T/\theta_0$ has a chi-square distribution with $2r$ degrees of freedom it follows that the hypothesis should be rejected if

$$\text{Min}(P\{\chi^2_{2r} \leq v\},\ 1 - P\left\{\chi^2_{2r} \geq v\right\}) \leq \alpha/2$$

or, equivalently, if

$$\alpha \geq 2\text{Min}(P\{\chi^2_{2r} \leq v\},\ 1 - P\{\chi^2_{2r} \leq v\})$$

10. $\tau = 71.51 \quad 2\tau/10 = 14.302 \quad p\text{-value} = 2P\{\chi^2_{16} < 14.302\} = .85$

11. (a) $10\sum_{11}^{20} 1/j$ (b) $100\sum_{11}^{20} 1/j^2$

12. $20\log\left(\frac{n}{n-9}\right) = 3$ or $n/(n-9) = e^{.15}$ or $n = 9e^{.15}/(e^{.15} - 1)$ which yield that $n = 65$

13. (a) 300/16 (b) p-value $= .864$

14. Let $X_1, X_2, \ldots$ be independent exponentials with mean 1 and think of them as being the interarrival times of a Poisson process with rate 1. Let $N(t)$ be the number of events of this process by time t. Then

$$\begin{aligned} P\{N(x/2) \geq n\} &= P\{X_1 + \ldots + X_n \leq x/2\} = P\{\text{Gamma}(n, 1) \leq x/2\} \\ &= P\left\{\frac{1}{2}\chi^2_{2n} \leq x/2\right\} \end{aligned}$$

which proves the result since $N(x/2)$ has a Poisson distribution with mean $x/2$.

15. Let the data be the times of failure $x_1, \ldots, x_k$ with $k = r$ meaning that the test stopped at the rth failure and $k < r$ meaning that it ended at time T. Since the lifetimes are $x_1 - x_0,\ x_2 - x_1, \ldots, x_k - x_{k-1}$ (where $x_0 = 0$) and in addition when $k < r$ there is an additional lifetime that exceeds $T - x_k$. The likelihood can be written as

$$L(x_1, \ldots, x_k) = \begin{array}{ll} \prod_{i=1}^{r} 1/\theta e^{-(x_i - x_{i-1})/\theta} = \theta^{-r} e^{-x_1/\theta} & \text{if } k = r \\ \prod_{i=1}^{k} 1/\theta e^{-(x_i - x_{i-1})/\theta} e^{-(T - x_k)/\theta} = \theta^{-r} e^{-T/\theta} & \text{if } k < r \end{array}$$

Hence,

$$\log L = \begin{array}{ll} -r\log\theta - x_r/\theta & \text{if } k = r \\ -r\log\theta - T/\theta & \text{if } k < r \end{array}$$

Differentiation now yields that the maximum likelihood estimate is x_r/r when $k = r$ and T/k when $k < r$. In either case this is equal to the total time on test divided by the number of observed failures.

16. $\text{Log } L = -r\log\theta - \left(\sum x_i + \sum y_i\right)/\theta + \log K$
$\frac{d}{d\theta}\text{Log } L = -r/\theta + \left(\sum x_i + \sum y_i\right)/\theta^2$ and the result follows upon setting equal to 0 and solving.

17. Total time on test $= 5 \times 86 + 4(128 - 86) + 3(153 - 128) + 2(197 - 153) = 761$. MLE $= 761/9 = 84.556$

18. $702.8/12 = 58.567$

19. $10/861 = .0116$

20. $13/732.8 = .0177$

21. $\bar{X}/\bar{Y} = 1.376 \quad p\text{-value} = 2P\{F_{7.7} > 1.376\} = .684$

22. (a) $2r_i/\theta_i$, $i = 1, 2$ have chi-square distributions with $2r_i$ degrees of freedom respectively. Hence, when the means are equal $(\tau_1/r_1)/(\tau_2/r_2)$ has an F-distribution with r_1 and r_2 degrees of freedom.

(b) $.7\tau_1/\tau_2 = 2.461 \quad p\text{-value} - 2P\{F_{20.14} > 2.461\} = 0.89$

23. $E[X] = \int (x/\alpha)^{1/\beta} x dx$ upon making the suggested substitution.

24. $E[X^2] = \int (x/\alpha)^{2/\beta} x dx$ by the same substitution as in Problem 23. Now use $\text{Var}(X) = E[X^2] - (E[X])^2$.

26. $P\{\alpha X^\beta \le x\} = P\{X \le (x/\alpha)^{1/\beta}\} = 1 - \exp\{-\alpha(x/a)\} = 11 - e^{-x}$

27. $P\{(-(1/\alpha)\log U)^{1/\beta} < x\} = P\{-(1/\alpha)\log U < x^\beta\} = P\{U > e^{-\alpha x^\beta}\} = 1 - e^{-\alpha x^\beta}$

28. (a) $P\{F(X) < a\} = P\{X < F^{-1}(a)\} = F(F^{-1}(a)) = a, \quad 0 < a < 1$

(b) $P\{1 - F(X) < a\} = P\{F(X) > 1 - a\} = a$ from part (a)

29. (a) In order for the ith smallest of n random variables to be equal to t $i - 1$ must be less than t one equal to t and $n - i$ greater than t. Since there are $n!/(i-1)!(n-i)!$ choices of these 3 sets the result follows.

(b) It follows from (a) that $\int_0^1 t^{i-1}(1-t)^{n-i}dt = (n-i)!(i-1)!/n!$. Hence, by substituting $i + 1$ for i and $n + 1$ for n we see that $\int_0^1 t^i(1-t)^{n-i}dt = (n-i)!i!(n+1)!$ and so

$$E[U_{(i)}] = \frac{n!(n-i)!i!}{(n-i)!(i-1)!(n+1)!} = \frac{i}{n+1}$$

(c) Since $F(X)$ is uniform (0.1) the result follows from (b) since $F(X_{(i)})$ has the same distribution as the ith smallest of a set of n uniform (0.1) random variables.

30. $P\{-\log U < x\} = P\{U > e^{-x}\} = 1 - e^{-x}$

Using this the left side of (10.5.7) would equal the expected time of the ith failure when n exponentials with rate 1 are simultaneously put on test. But this is equal to the mean time until the first failure $(1/n)$ plus the mean time between the first and second failure $(1/(n - 1))$ plus ... plus the mean time between the $(i - 1)$st and ith failure $(1/[n - (i - 1)])$.

Chapter 15

1. If $x_0 = 4$, and

$$x_n = 3\,x_{n-1} \quad \text{mod } 7$$

then $x_1, \ldots, x_{10}$ are

$$5, 1, 3, 2, 6, 4, 5, 1, 3, 2, 6$$

2. It is immediate for $n = 2$. So, the permutation before the interchange is equally likely to be either $P_1 = 1, 2, 3$ or $P_2 = 2, 1, 3$. So, with F being the final permutation

$$P(F = 1, 2, 3) = P(F = 1, 2, 3|P_1)P(P_1) = (1/3)(1/2) = 1/6$$

$$P(F = 2, 1, 3) = P(F = 2, 1, 3|P_2)P(P_2) = (1/3)(1/2) = 1/6$$

$$P(F = 1, 3, 2) = P(F = 1, 3, 2|P_1)P(P_1) = (1/3)(1/2) = 1/6$$

$$P(F = 2, 3, 1) = P(F = 2, 3, 1|P_2)P(P_2) = (1/3)(1/2) = 1/6$$

$$P(F = 3, 1, 2) = P(F = 3, 1, 2|P_2)P(P_2) = (1/3)(1/2) = 1/6$$

$$P(F = 3, 2, 1) = P(F = 3, 2, 1|P_1)P(P_1) = (1/3)(1/2) = 1/6$$

3. (a) The estimator is $\frac{\bar{X}_n}{\bar{Y}_n}$.

(b) Suppose the observed data is $X_i = x_i$, $Y_i = y_i$, $i = 1, \ldots, n$. Let $\bar{x} = \sum_{i=1}^{n} x_i/n$ and $\bar{y} = \sum_{i=1}^{n} y_i/n$. Estimate the mean square error by

$$MSE_e = E_e\left[\left(\sum_{i=1}^{n} X_i \Big/ \sum_{i=1}^{n} Y_i - \bar{x}/\bar{y}\right)^2\right]$$

where the random vectors $(X_i,\ Y_i)$, $i = 1, \ldots, n$ are independent and have common mass function

$$P(X_i = x_j,\ Y_i = y_j) = 1/n, \quad j = 1, \ldots, n$$

The quantity MSE_e can then be estimated by a simulation.

4. (a) We need to compute $\text{Var}\left(\sum_{i=1}^{2}(X_i - \bar{X})^2\right)$, where X_1 and X_2 are independent and equally likely to be either 1 or 3. Consequently,

$$P\left(\sum_{i=1}^{2}(X_i - \bar{X})^2 = 0\right) = P(X_1 = X_2) = 1/2$$

$$P\left(\sum_{i=1}^{2}(X_i - \bar{X})^2 = 2\right) = P(X_1 \neq X_2) = 1/2$$

Hence,

$$E\left[\sum_{i=1}^{2}(X_i - \bar{X})^2\right] = 1$$

and

$$E\left[\left(\sum_{i=1}^{2}(X_i - \bar{X})^2\right)^2\right] = 0(1/2) + 4(1/2) = 2$$

giving that

$$\mathrm{Var}\left(\sum_{i=1}^{2}(X_i - \bar{X})^2\right) = 1$$

(b) We want to compute $\mathrm{Var}\left(\sum_{i=1}^{15}(X_i - \bar{X})^2\right)$ when $X_i,\ i = 1, \ldots, 15$ are independent with each equally likely to be any of the 15 given data values. A simulation yields that this is approximately equal to 33.20.

5. Estimate p by $P\left(\sum_{i=1}^{8} X_i/8 < 8\right)$, when $X_i,\ i = 1, \ldots, 8$, are independent with each equally likely to be any of the 8 given data values. A simulation yields that this is approximately equal to 0.525.

6. The value of the test statistic is $T = \sum_j jX_j = 4582$. Under the null hypothesis that all orderings are equally likely

$$E_{H_0}[T] = 4334, \sqrt{\mathrm{Var}_{H_0}(T)} = 81.44$$

Thus, using the normal approximation

$$p\text{-value} \approx P\left(Z \geq \frac{4582 - 4334}{81.44}\right) = P(Z \geq 3.045) < .0013$$

Hence, the data strongly support the hypothesis that the student improved as the semester progressed.

7. The value of the test statistic is $T = \sum_j jX_j = 840$. Under the null hypothesis that all orderings are equally likely $E_{H_0}[T] = 840$, showing that the p-value is approximately .5, which is not a validation of the player's reputation.

8. The value of the test statistic T, equal to the sum of the group 1 lifetimes, is $T = 1389$. Under the null hypothesis

$$E_{H_0}[T] = 1391.5, \quad \sqrt{\mathrm{Var}_{H_0}(T)} = 112.44$$

Hence,

$$p\text{-value} = 2P_{H_0}(T \le 1389) \approx 2P\left(Z \le \frac{1389 - 1391.5}{112.44}\right) \approx 1$$

9. The value of the test statistic is $T = 402$. Under the null hypothesis

$$E_{H_0}[T] = 385.6, \sqrt{\mathrm{Var}_{H_0}(T)} = 32.85$$

Thus, in a two-sided test

$$p\text{-value} = 2P_{H_0}(T \ge 402) \approx 2P(Z \ge .499) = .617$$

10. The value of the test statistic is $T = 826$. Under the null hypothesis

$$E_{H_0}[T] = 1450, \sqrt{\mathrm{Var}_{H_0}(T)} = 266.12$$

Thus,

$$p\text{-value} = P_{H_0}(T \le 826) \approx P(Z \le -2.35) = .0094$$

11. Use that

$$P(X = i + 1) = \frac{\lambda}{i + 1} P(X = i), \quad i \ge 0$$

This gives

(1) Set $I = 0$, $P = F = e^{-\lambda}$
(2) Generate U
(3) If $U \le F$, set $X = I$ and stop.
(4) $I = I + 1,\ P = P * \frac{\lambda}{I},\ F = F + P$
(5) Go to Step 3

12. For X being geometric with parameter p

$$F(n) = P(X \le n) = 1 - P(X > n) = 1 - (1 - p)^n$$

Therefore, the inverse transform method is to generate a random number U and set $X = n$ if

$$F(n - 1) \le U < F(n)$$

which is equivalent to

$$(1 - p)^{n-1} \ge 1 - U > (1 - p)^n$$

or, upon taking logarithms,

$$(n-1)\log(1-p) \geq \log(1-U) > n\log(p)$$

or

$$n-1 \leq \frac{\log(1-U)}{\log(1-p)} < n$$

Thus, X is the smallest integer larger than $\frac{\log(1-U)}{\log(1-p)}$.

Another method is to continually generate random numbers $U_1, \ldots$, stopping and setting

$$X = \min(n : U_n \leq p)$$

13. Use the inverse transform method. To begin, we have

$$F(x) = \frac{e^x - 1}{e - 1}, \quad 0 < x < 1$$

Thus, if $X = F^{-1}(U)$, then

$$U = F(x) = \frac{e^X - 1}{e - 1}$$

or

$$X = \log(1 + (e-1)U)$$

So, generate a random number U and set $X = \log(1 + (e-1)U)$.

14. Using the inverse transform method, generate a random number U and set $X = U^{1/n}$.

15. Use inverse transform. If $X = F^{-1}(U)$, then $U = F(X)$. Thus,

$$X^2 + X - 2U = 0$$

or

$$X = \frac{-1 \pm \sqrt{1+8U}}{2}$$

Because $X \geq 0$, this yields that $X = \left(\sqrt{1+8U} - 1\right)/2$.