Instructor Solution Manual

Probability and Statistics for Engineers and Scientists (3rd Edition)

Anthony Hayter

Instructor Solution Manual

This instructor solution manual to accompany the third edition of

"Probability and Statistics for Engineers and Scientists" by Anthony Hayter

provides worked solutions and answers to *all* of the problems given in the textbook. The student solution manual provides worked solutions and answers to only the odd-numbered problems given at the end of the chapter sections. In addition to the material contained in the student solution manual, this instructor manual therefore provides worked solutions and answers to the even-numbered problems given at the end of the chapter sections together with all of the supplementary problems at the end of each chapter.

Contents

1 Probability Theory **7**
- 1.1 Probabilities . . . 7
- 1.2 Events . . . 9
- 1.3 Combinations of Events . . . 13
- 1.4 Conditional Probability . . . 16
- 1.5 Probabilities of Event Intersections . . . 22
- 1.6 Posterior Probabilities . . . 28
- 1.7 Counting Techniques . . . 32
- 1.9 Supplementary Problems . . . 37

2 Random Variables **49**
- 2.1 Discrete Random Variables . . . 49
- 2.2 Continuous Random Variables . . . 54
- 2.3 The Expectation of a Random Variable . . . 58
- 2.4 The Variance of a Random Variable . . . 62
- 2.5 Jointly Distributed Random Variables . . . 68
- 2.6 Combinations and Functions of Random variables . . . 77
- 2.8 Supplementary Problems . . . 86

3 Discrete Probability Distributions **95**
- 3.1 The Binomial Distribution . . . 95
- 3.2 The Geometric and Negative Binomial Distributions . . . 99
- 3.3 The Hypergeometric Distribution . . . 102
- 3.4 The Poisson Distribution . . . 105
- 3.5 The Multinomial Distribution . . . 107
- 3.7 Supplementary Problems . . . 109

4 Continuous Probability Distributions **113**
- 4.1 The Uniform Distribution . . . 113
- 4.2 The Exponential Distribution . . . 116
- 4.3 The Gamma Distribution . . . 119
- 4.4 The Weibull Distribution . . . 121
- 4.5 The Beta Distribution . . . 123
- 4.7 Supplementary Problems . . . 125

5 The Normal Distribution **129**
5.1 Probability Calculations using the Normal Distribution . . . 129
5.2 Linear Combinations of Normal Random Variables . . . 135
5.3 Approximating Distributions with the Normal Distribution . . . 140
5.4 Distributions Related to the Normal Distribution . . . 144
5.6 Supplementary Problems . . . 148

6 Descriptive Statistics **157**
6.1 Experimentation . . . 157
6.2 Data Presentation . . . 159
6.3 Sample Statistics . . . 161
6.6 Supplementary Problems . . . 164

7 Statistical Estimation and Sampling Distributions **167**
7.2 Properties of Point Estimates . . . 167
7.3 Sampling Distributions . . . 170
7.4 Constructing Parameter Estimates . . . 176
7.6 Supplementary Problems . . . 177

8 Inferences on a Population Mean **183**
8.1 Confidence Intervals . . . 183
8.2 Hypothesis Testing . . . 189
8.5 Supplementary Problems . . . 196

9 Comparing Two Population Means **205**
9.2 Analysis of Paired Samples . . . 205
9.3 Analysis of Independent Samples . . . 209
9.6 Supplementary Problems . . . 218

10 Discrete Data Analysis **225**
10.1 Inferences on a Population Proportion . . . 225
10.2 Comparing Two Population Proportions . . . 232
10.3 Goodness of Fit Tests for One-way Contingency Tables . . . 240
10.4 Testing for Independence in Two-way Contingency Tables . . . 246
10.6 Supplementary Problems . . . 251

11 The Analysis of Variance **263**
11.1 One Factor Analysis of Variance . . . 263
11.2 Randomized Block Designs . . . 273
11.4 Supplementary Problems . . . 281

12 Simple Linear Regression and Correlation **287**
12.1 The Simple Linear Regression Model . . . 287
12.2 Fitting the Regression Line . . . 289
12.3 Inferences on the Slope Parameter $\hat{\beta}_1$ . . . 292
12.4 Inferences on the Regression Line . . . 296
12.5 Prediction Intervals for Future Response Values . . . 298
12.6 The Analysis of Variance Table . . . 300
12.7 Residual Analysis . . . 302

12.8 Variable Transformations . . . 303
12.9 Correlation Analysis . . . 305
12.11 Supplementary Problems . . . 306

13 Multiple Linear Regression and Nonlinear Regression 317
13.1 Introduction to Multiple Linear Regression . . . 317
13.2 Examples of Multiple Linear Regression . . . 320
13.3 Matrix Algebra Formulation of Multiple Linear Regression . . . 322
13.4 Evaluating Model Accuracy . . . 327
13.6 Supplementary Problems . . . 328

14 Multifactor Experimental Design and Analysis 333
14.1 Experiments with Two Factors . . . 333
14.2 Experiments with Three or More Factors . . . 336
14.3 Supplementary Problems . . . 340

15 Nonparametric Statistical Analysis 343
15.1 The Analysis of a Single Population . . . 343
15.2 Comparing Two Populations . . . 347
15.3 Comparing Three or More Populations . . . 350
15.4 Supplementary Problems . . . 354

16 Quality Control Methods 359
16.2 Statistical Process Control . . . 359
16.3 Variable Control Charts . . . 361
16.4 Attribute Control Charts . . . 363
16.5 Acceptance Sampling . . . 364
16.6 Supplementary Problems . . . 365

17 Reliability Analysis and Life Testing 367
17.1 System Reliability . . . 367
17.2 Modeling Failure Rates . . . 369
17.3 Life Testing . . . 372
17.4 Supplementary Problems . . . 374

Chapter 1

Probability Theory

1.1 Probabilities

1.1.1 $\mathcal{S}$ = {(head, head, head), (head, head, tail), (head, tail, head), (head, tail, tail), (tail, head, head), (tail, head, tail), (tail, tail, head), (tail, tail, tail)}

1.1.2 $\mathcal{S}$ = {0 females, 1 female, 2 females, 3 females, ..., n females}

1.1.3 $\mathcal{S} = \{0,1,2,3,4\}$

1.1.4 $\mathcal{S}$ = {January 1, January 2, ... , February 29, ... , December 31}

1.1.5 $\mathcal{S}$ = {(on time, satisfactory), (on time, unsatisfactory), (late, satisfactory), (late, unsatisfactory)}

1.1.6 $\mathcal{S}$ = {(red, shiny), (red, dull), (blue, shiny), (blue, dull)}

1.1.7 (a) $\frac{p}{1-p} = 1 \;\Rightarrow\; p = 0.5$

(b) $\frac{p}{1-p} = 2 \;\Rightarrow\; p = \frac{2}{3}$

(c) $p = 0.25 \;\Rightarrow\; \frac{p}{1-p} = \frac{1}{3}$

1.1.8 $0.13 + 0.24 + 0.07 + 0.38 + P(V) = 1 \;\Rightarrow\; P(V) = 0.18$

1.1.9 $0.08 + 0.20 + 0.33 + P(IV) + P(V) = 1 \;\Rightarrow\; P(IV) + P(V) = 1 - 0.61 = 0.39$

Therefore, $0 \leq P(V) \leq 0.39$.

If $P(IV) = P(V)$ then $P(V) = 0.195$.

1.1.10 $P(I) = 2 \times P(II)$ and $P(II) = 3 \times P(III) \;\Rightarrow\; P(I) = 6 \times P(III)$

Therefore,

$P(I) + P(II) + P(III) = 1$

so that

$(6 \times P(III)) + (3 \times P(III)) + P(III) = 1.$

Consequently,

$P(III) = \frac{1}{10}$, $P(II) = 3 \times P(III) = \frac{3}{10}$

and

$P(I) = 6 \times P(III) = \frac{6}{10}$.

1.2 Events

1.2.1 (a) $0.13 + P(b) + 0.48 + 0.02 + 0.22 = 1 \;\Rightarrow\; P(b) = 0.15$

(b) $A = \{c, d\}$ so that $P(A) = P(c) + P(d) = 0.48 + 0.02 = 0.50$

(c) $P(A') = 1 - P(A) = 1 - 0.5 = 0.50$

1.2.2 (a) $P(A) = P(b) + P(c) + P(e) = 0.27$ so $P(b) + 0.11 + 0.06 = 0.27$
and hence $P(b) = 0.10$

(b) $P(A') = 1 - P(A) = 1 - 0.27 = 0.73$

(c) $P(A') = P(a) + P(d) + P(f) = 0.73$ so $0.09 + P(d) + 0.29 = 0.73$
and hence $P(d) = 0.35$

1.2.3 Over a four year period including one leap year, the number of days is

$(3 \times 365) + 366 = 1461$.

The number of January days is $4 \times 31 = 124$

and the number of February days is $(3 \times 28) + 29 = 113$.

The answers are therefore $\frac{124}{1461}$ and $\frac{113}{1461}$.

1.2.4 $\mathcal{S} = \{1,\, 2,\, 3,\, 4,\, 5,\, 6\}$

Prime $= \{1,\, 2,\, 3,\, 5\}$

All the events in $\mathcal{S}$ are equally likely to occur and each has a probability of $\frac{1}{6}$

so that

$P(\text{Prime}) = P(1) + P(2) + P(3) + P(5) = \frac{4}{6} = \frac{2}{3}$.

1.2.5 See Figure 1.10.

The event that the score on at least one of the two dice is a prime number consists of the following 32 outcomes:

{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4) (2,5), (2,6), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,5), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,5)}

Each outcome in $\mathcal{S}$ is equally likely to occur with a probability of $\frac{1}{36}$ so that

$P(\text{at least one score is a prime number}) = 32 \times \frac{1}{36} = \frac{32}{36} = \frac{8}{9}$.

The complement of this event is the event that neither score is a prime number which includes the following four outcomes:

{(4,4), (4,6), (6,4), (6,6)}

Therefore, $P(\text{neither score prime}) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{9}$.

1.2.6 In Figure 1.10 let (x, y) represent the outcome that the score on the red die is x and the score on the blue die is y. The event that the score on the red die is *strictly greater* than the score on the blue die consists of the following 15 outcomes:

{(2,1), (3,1), (3,2), (4,1), (4,2), (4,3), (5,1), (5,2), (5,3), (5,4), (6,1), (6,2), (6,3), (6,4), (6,5)}

The probability of each outcome is $\frac{1}{36}$ so the required probability is $15 \times \frac{1}{36} = \frac{5}{12}$.

This probability is less than 0.5 because of the possibility that both scores are equal.

The complement of this event is the event that the red die has a score *less than or equal* to the score on the blue die which has a probability of $1 - \frac{5}{12} = \frac{7}{12}$.

1.2.7 $P(\spadesuit \text{ or } \clubsuit) = P(A\spadesuit) + P(K\spadesuit) + \ldots + P(2\spadesuit) + P(A\clubsuit) + P(K\clubsuit) + \ldots + P(2\clubsuit)$
$= \frac{1}{52} + \ldots + \frac{1}{52} = \frac{26}{52} = \frac{1}{2}$

1.2.8 $P(\text{draw an ace}) = P(A\spadesuit) + P(A\clubsuit) + P(A\diamondsuit) + P(A\heartsuit)$
$= \frac{1}{52} + \frac{1}{52} + \frac{1}{52} + \frac{1}{52} = \frac{4}{52} = \frac{1}{13}$

1.2.9 (a) Let the four players be named A, B, C, and T for Terica, and let the notation (X, Y) indicate that player X is the winner and player Y is the runner up.

The sample space consists of the 12 outcomes:

$\mathcal{S}$ = {(A,B), (A,C), (A,T), (B,A), (B,C), (B,T), (C,A), (C,B), (C,T), (T,A), (T,B), (T,C)}

The event *'Terica is winner'* consists of the 3 outcomes {(T,A), (T,B), (T,C)}.

Since each outcome in $\mathcal{S}$ is equally likely to occur with a probability of $\frac{1}{12}$ it follows that

$P(\text{Terica is winner}) = \frac{3}{12} = \frac{1}{4}$.

(b) The event *'Terica is winner or runner up'* consists of 6 out of the 12 outcomes so that

$P(\text{Terica is winner or runner up}) = \frac{6}{12} = \frac{1}{2}$.

1.2.10 (a) See Figure 1.24.
P(Type I battery lasts longest)
$= P((II, III, I)) + P((III, II, I))$
$= 0.39 + 0.03 = 0.42$

(b) P(Type I battery lasts shortest)
$= P((I, II, III)) + P((I, III, II))$
$= 0.11 + 0.07 = 0.18$

(c) P(Type I battery does not last longest)
$= 1 - P$(Type I battery lasts longest)
$= 1 - 0.42 = 0.58$

(d) P(Type I battery last longer than Type II)
$= P((II, I, III)) + P((II, III, I)) + P((III, II, I))$
$= 0.24 + 0.39 + 0.03 = 0.66$

1.2.11 (a) See Figure 1.25.
The event *'both assembly lines are shut down'* consists of the single outcome {(S,S)}.
Therefore,
P(both assembly lines are shut down) $= 0.02$.

(b) The event *'neither assembly line is shut down'* consists of the outcomes {(P,P), (P,F), (F,P), (F,F)}.
Therefore,
P(neither assembly line is shut down)
$= P((P, P)) + P((P, F)) + P((F, P)) + P((F, F))$
$= 0.14 + 0.2 + 0.21 + 0.19 = 0.74.$

(c) The event *'at least one assembly line is at full capacity'* consists of the outcomes {(S,F), (P,F), (F,F), (F,S), (F,P)}.
Therefore,
P(at least one assembly line is at full capacity)
$= P((S, F)) + P((P, F)) + P((F, F)) + P((F, S)) + P((F, P))$
$= 0.05 + 0.2 + 0.19 + 0.06 + 0.21 = 0.71.$

(d) The event *'exactly one assembly line at full capacity'* consists of the outcomes {(S,F), (P,F), (F,S), (F,P)}.
Therefore,
P(exactly one assembly line at full capacity)
$= P((S, F)) + P((P, F)) + P((F, S)) + P((F, P))$

$= 0.05 + 0.20 + 0.06 + 0.21 = 0.52.$

The complement of *'neither assembly line is shut down'* is the event *'at least one assembly line is shut down'* which consists of the outcomes

{(S,S), (S,P), (S,F), (P,S), (F,S)}.

The complement of *'at least one assembly line is at full capacity'* is the event *'neither assembly line is at full capacity'* which consists of the outcomes

{(S,S), (S,P), (P,S), (P,P)}.

1.2.12 The sample space is

$\mathcal{S}$ = {(H,H,H), (H,T,H), (H,T,T), (H,H,T), (T,H,H), (T,H,T), (T,T,H), (T,T,T)}

with each outcome being equally likely with a probability of $\frac{1}{8}$.

The event *'two heads obtained in succession'* consists of the three outcomes

{(H,H,H), (H,H,T), (T,H,H)}

so that $P(\text{two heads in succession}) = \frac{3}{8}$.

1.3 Combinations of Events

1.3.1 The event A contains the outcome 0 while the empty set does not contain any outcomes.

1.3.2 (a) See Figure 1.55.
$P(B) = 0.01 + 0.02 + 0.05 + 0.11 + 0.08 + 0.06 + 0.13 = 0.46$

(b) $P(B \cap C) = 0.02 + 0.05 + 0.11 = 0.18$

(c) $P(A \cup C) = 0.07+0.05+0.01+0.02+0.05+0.08+0.04+0.11+0.07+0.11 = 0.61$

(d) $P(A \cap B \cap C) = 0.02 + 0.05 = 0.07$

(e) $P(A \cup B \cup C) = 1 - 0.03 - 0.04 - 0.05 = 0.88$

(f) $P(A' \cap B) = 0.08 + 0.06 + 0.11 + 0.13 = 0.38$

(g) $P(B' \cup C) = 0.04 + 0.03 + 0.05 + 0.11 + 0.05 + 0.02 + 0.08 + 0.04 + 0.11 + 0.07$
$+ 0.07 + 0.05 = 0.72$

(h) $P(A \cup (B \cap C)) = 0.07 + 0.05 + 0.01 + 0.02 + 0.05 + 0.08 + 0.04 + 0.11 = 0.43$

(i) $P((A \cup B) \cap C) = 0.11 + 0.05 + 0.02 + 0.08 + 0.04 = 0.30$

(j) $P(A' \cup C) = 0.04 + 0.03 + 0.05 + 0.08 + 0.06 + 0.13 + 0.11 + 0.11 + 0.07 + 0.02$
$+ 0.05 + 0.08 + 0.04 = 0.87$
$P(A' \cup C)' = 1 - P(A' \cup C) = 1 - 0.87 = 0.13$

1.3.4 (a) $A \cap B = \{$females with black hair$\}$

(b) $A \cup C' = \{$all females and any man who does not have brown eyes$\}$

(c) $A' \cap B \cap C = \{$males with black hair and brown eyes$\}$

(d) $A \cap (B \cup C) = \{$females with either black hair or brown eyes or both$\}$

1.3.5 Yes, because a card must be drawn from either a red suit or a black suit but it cannot be from both at the same time.

No, because the ace of hearts could be drawn.

1.3.6 $P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq 1$

so that

$P(B) \leq 1 - 0.4 + 0.3 = 0.9.$

Also, $P(B) \geq P(A \cap B) = 0.3$

so that

$0.3 \leq P(B) \leq 0.9.$

1.3.7 Since $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

it follows that

$P(B) = P(A \cup B) - P(A) + P(A \cap B)$

$= 0.8 - 0.5 + 0.1 = 0.4.$

1.3.8 $\mathcal{S} = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$ where each outcome is equally likely with a probability of $\frac{1}{6}$.

The events A, B, and B' are $A = \{2,\ 4,\ 6\}$, $B = \{1,\ 2,\ 3,\ 5\}$ and $B' = \{4,\ 6\}$.

(a) $A \cap B = \{2\}$ so that $P(A \cap B) = \frac{1}{6}$

(b) $A \cup B = \{1, 2, 3, 4, 5, 6\}$ so that $P(A \cup B) = 1$

(c) $A \cap B' = \{4, 6\}$ so that $P(A \cap B') = \frac{2}{6} = \frac{1}{3}$

1.3.9 Yes, the three events are mutually exclusive because the selected card can only be from one suit.

Therefore,

$P(A \cup B \cup C) = P(A) + P(B) + P(C) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}.$

A' is the event *'a heart is not obtained'* (or similarly the event *'a club, spade, or diamond is obtained'*) so that B is a subset of A'.

1.3.10 (a) $A \cap B = \{A\heartsuit,\ A\diamondsuit\}$

(b) $A \cup C = \{A\heartsuit,\ A\diamondsuit,\ A\clubsuit,\ A\spadesuit,\ K\heartsuit,\ K\diamondsuit,\ K\clubsuit,\ K\spadesuit,\ Q\heartsuit,\ Q\diamondsuit,\ Q\clubsuit,\ Q\spadesuit,$
$J\heartsuit,\ J\diamondsuit,\ J\clubsuit,\ J\spadesuit\}$

(c) $B \cap C' = \{A\heartsuit, 2\heartsuit,\ \ldots\ ,\ 10\heartsuit,\ A\diamondsuit,\ 2\diamondsuit,\ \ldots\ ,\ 10\diamondsuit\}$

(d) $B' \cap C = \{K\clubsuit,\ K\spadesuit,\ Q\clubsuit,\ Q\spadesuit,\ J\clubsuit,\ J\spadesuit\}$
$A \cup (B' \cap C) = \{A\heartsuit,\ A\diamondsuit,\ A\clubsuit,\ A\spadesuit,\ K\clubsuit,\ K\spadesuit,\ Q\clubsuit,\ Q\spadesuit,\ J\clubsuit,\ J\spadesuit\}$

1.3.11 Let the event O be an on time repair and let the event S be a satisfactory repair.

It is known that $P(O \cap S) = 0.26$, $P(O) = 0.74$ and $P(S) = 0.41$.

We want to find $P(O' \cap S')$.

Since the event $O' \cap S'$ can be written $(O \cup S)'$ it follows that

$P(O' \cap S') = 1 - P(O \cup S)$

$= 1 - (P(O) + P(S) - P(O \cap S))$

$= 1 - (0.74 + 0.41 - 0.26) = 0.11.$

1.3.12 Let R be the event that a red ball is chosen and let S be the event that a shiny ball is chosen.

It is known that $P(R \cap S) = \frac{55}{200}$, $P(S) = \frac{91}{200}$ and $P(R) = \frac{79}{200}$.

Therefore, the probability that the chosen ball is either shiny or red is

$P(R \cup S) = P(R) + P(S) - P(R \cap S)$

$= \frac{79}{200} + \frac{91}{200} - \frac{55}{200}$

$= \frac{115}{200} = 0.575.$

The probability of a dull blue ball is

$P(R' \cap S') = 1 - P(R \cup S)$

$= 1 - 0.575 = 0.425.$

1.3.13 Let A be the event that the patient is male, let B be the event that the patient is younger than thirty years of age, and let C be the event that the patient is admitted to the hospital.

It is given that $P(A) = 0.45$, $P(B) = 0.30$, $P(A' \cap B' \cap C) = 0.15$,

and $P(A' \cap B) = 0.21$.

The question asks for $P(A' \cap B' \cap C')$.

Notice that

$P(A' \cap B') = P(A') - P(A' \cap B) = (1 - 0.45) - 0.21 = 0.34$

so that

$P(A' \cap B' \cap C') = P(A' \cap B') - P(A' \cap B' \cap C) = 0.34 - 0.15 = 0.19.$

1.4 Conditional Probability

1.4.1 See Figure 1.55.

(a) $P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0.02+0.05+0.01}{0.02+0.05+0.01+0.11+0.08+0.06+0.13} = 0.1739$

(b) $P(C \mid A) = \frac{P(A \cap C)}{P(A)} = \frac{0.02+0.05+0.08+0.04}{0.02+0.05+0.08+0.04+0.018+0.07+0.05} = 0.59375$

(c) $P(B \mid A \cap B) = \frac{P(B \cap (A \cap B))}{P(A \cap B)} = \frac{P(A \cap B)}{P(A \cap B)} = 1$

(d) $P(B \mid A \cup B) = \frac{P(B \cap (A \cup B))}{P(A \cup B)} = \frac{P(B)}{P(A \cup B)} = \frac{0.46}{0.46+0.32-0.08} = 0.657$

(e) $P(A \mid A \cup B \cup C) = \frac{P(A \cap (A \cup B \cup C))}{P(A \cup B \cup C)} = \frac{P(A)}{P(A \cup B \cup C)} = \frac{0.32}{1-0.04-0.05-0.03} = 0.3636$

(f) $P(A \cap B \mid A \cup B) = \frac{P((A \cap B) \cap (A \cup B))}{P(A \cup B)} = \frac{P(A \cap B)}{P(A \cup B)} = \frac{0.08}{0.7} = 0.1143$

1.4.2 $A = \{1, 2, 3, 5\}$ and $P(A) = \frac{4}{6} = \frac{2}{3}$

$P(5 \mid A) = \frac{P(5 \cap A)}{P(A)} = \frac{P(5)}{P(A)} = \frac{\left(\frac{1}{6}\right)}{\left(\frac{2}{3}\right)} = \frac{1}{4}$

$P(6 \mid A) = \frac{P(6 \cap A)}{P(A)} = \frac{P(\emptyset)}{P(A)} = 0$

$P(A \mid 5) = \frac{P(A \cap 5)}{P(5)} = \frac{P(5)}{P(5)} = 1$

1.4.3 (a) $P(A\heartsuit \mid \text{red suit}) = \frac{P(A\heartsuit \cap \text{red suit})}{P(\text{red suit})} = \frac{P(A\heartsuit)}{P(\text{red suit})} = \frac{\left(\frac{1}{52}\right)}{\left(\frac{26}{52}\right)} = \frac{1}{26}$

(b) $P(\text{heart} \mid \text{red suit}) = \frac{P(\text{heart} \cap \text{red suit})}{P(\text{red suit})} = \frac{P(\text{heart})}{P(\text{red suit})} = \frac{\left(\frac{13}{52}\right)}{\left(\frac{26}{52}\right)} = \frac{13}{26} = \frac{1}{2}$

(c) $P(\text{red suit} \mid \text{heart}) = \frac{P(\text{red suit} \cap \text{heart})}{P(\text{heart})} = \frac{P(\text{heart})}{P(\text{heart})} = 1$

(d) $P(\text{heart} \mid \text{black suit}) = \frac{P(\text{heart} \cap \text{black suit})}{P(\text{black suit})} = \frac{P(\emptyset)}{P(\text{black suit})} = 0$

(e) $P(\text{King} \mid \text{red suit}) = \frac{P(\text{King} \cap \text{red suit})}{P(\text{red suit})} = \frac{P(K\heartsuit,\, K\diamondsuit)}{P(\text{red suit})} = \frac{\left(\frac{2}{52}\right)}{\left(\frac{26}{52}\right)} = \frac{2}{26} = \frac{1}{13}$

(f) $P(\text{King} \mid \text{red picture card}) = \frac{P(\text{King} \cap \text{red picture card})}{P(\text{red picture card})}$

$$= \frac{P(K\heartsuit,\ K\diamondsuit)}{P(\text{red picture card})} = \frac{\left(\frac{2}{52}\right)}{\left(\frac{6}{52}\right)} = \frac{2}{6} = \frac{1}{3}$$

1.4.4 $P(A)$ is smaller than $P(A \mid B)$.

Event B is a necessary condition for event A and so conditioning on event B increases the probability of event A.

1.4.5 There are 54 blue balls and so there are $150 - 54 = 96$ red balls.

Also, there are 36 shiny, red balls and so there are $96 - 36 = 60$ dull, red balls.

$P(\text{shiny} \mid \text{red}) = \frac{P(\text{shiny} \cap \text{red})}{P(\text{red})} = \frac{\left(\frac{36}{150}\right)}{\left(\frac{96}{150}\right)} = \frac{36}{96} = \frac{3}{8}$

$P(\text{dull} \mid \text{red}) = \frac{P(\text{dull} \cap \text{red})}{P(\text{red})} = \frac{\left(\frac{60}{150}\right)}{\left(\frac{96}{150}\right)} = \frac{60}{96} = \frac{5}{8}$

1.4.6 Let the event O be an on time repair and let the event S be a satisfactory repair.

It is known that $P(S \mid O) = 0.85$ and $P(O) = 0.77$.

The question asks for $P(O \cap S)$ which is

$P(O \cap S) = P(S \mid O) \times P(O) = 0.85 \times 0.77 = 0.6545$.

1.4.7 (a) It depends on the weather patterns in the particular location that is being considered.

(b) It increases since there are proportionally more black haired people among brown eyed people than there are in the general population.

(c) It remains unchanged.

(d) It increases.

1.4.8 Over a four year period including one leap year the total number of days is

$(3 \times 365) + 366 = 1461$.

Of these, $4 \times 12 = 48$ days occur on the first day of a month and so the probability that a birthday falls on the first day of a month is

$\frac{48}{1461} = 0.0329$.

Also, $4 \times 31 = 124$ days occur in March of which 4 days are March 1st.

Consequently, the probability that a birthday falls on March 1st. conditional that it is in March is

$\frac{4}{124} = \frac{1}{31} = 0.0323$.

Finally, $(3 \times 28) + 29 = 113$ days occur in February of which 4 days are February 1st.

Consequently, the probability that a birthday falls on February 1st. conditional that it is in February is

$\frac{4}{113} = 0.0354$.

1.4.9 (a) Let A be the event that *'Type I battery lasts longest'* consisting of the outcomes {(III, II, I), (II, III, I)}.

Let B be the event that *'Type I battery does not fail first'* consisting of the outcomes {(III,II,I), (II,III,I), (II,I,III), (III,I,II)}.

The event $A \cap B = \{(\text{III,II,I}), (\text{II,III,I})\}$ is the same as event A.

Therefore,

$P(A \mid B) = \frac{P(A\cap B)}{P(B)} = \frac{0.39+0.03}{0.39+0.03+0.24+0.16} = 0.512$.

(b) Let C be the event that *'Type II battery fails first'* consisting of the outcomes {(II,I,III), (II,III,I)}.

Thus, $A \cap C = \{(II, III, I)\}$ and therefore

$P(A \mid C) = \frac{P(A\cap C)}{P(C)} = \frac{0.39}{0.39+0.24} = 0.619$.

(c) Let D be the event that *'Type II battery lasts longest'* consisting of the outcomes {(I,III,II), (III,I,II)}.

Thus, $A \cap D = \emptyset$ and therefore

$P(A \mid D) = \frac{P(A\cap D)}{P(D)} = 0$.

(d) Let E be the event that *'Type II battery does not fail first'* consisting of the outcomes {(I,III,II), (I,II,III), (III,II,I), (III,I,II)}.

Thus, $A \cap E = \{(III, II, I)\}$ and therefore

$P(A \mid E) = \frac{P(A\cap E)}{P(E)} = \frac{0.03}{0.07+0.11+0.03+0.16} = 0.081$.

1.4.10 See Figure 1.25.

(a) Let A be the event *'both lines at full capacity'* consisting of the outcome {(F,F)}.

Let B be the event *'neither line is shut down'* consisting of the outcomes {(P,P), (P,F), (F,P), (F,F)}.

Thus, $A \cap B = \{(F, F)\}$ and therefore

$P(A \mid B) = \frac{P(A\cap B)}{P(B)} = \frac{0.19}{(0.14+0.2+0.21+0.19)} = 0.257$.

(b) Let C be the event *'at least one line at full capacity'* consisting of the outcomes {(F,P), (F,S), (F,F), (S,F), (P,F)}.

Thus, $C \cap B = \{(F, P), (F, F), (P, F)\}$ and therefore

$P(C \mid B) = \frac{P(C\cap B)}{P(B)} = \frac{0.21+0.19+0.2}{0.74} = 0.811$.

(c) Let D be the event that *'one line is at full capacity'* consisting of the outcomes {(F,P), (F,S), (P,F), (S,F)}.
Let E be the event *'one line is shut down'* consisting of the outcomes {(S,P), (S,F), (P,S), (F,S)}.
Thus, $D \cap E = \{(F,S),(S,F)\}$ and therefore
$P(D \mid E) = \frac{P(D\cap E)}{P(E)} = \frac{0.06+0.05}{0.06+0.05+0.07+0.06} = 0.458.$

(d) Let G be the event that *'neither line is at full capacity'* consisting of the outcomes {(S,S), (S,P), (P,S), (P,P)}.
Let H be the event that *'at least one line is at partial capacity'* consisting of the outcomes {(S,P), (P,S), (P,P), (P,F), (F,P)}.
Thus, $F \cap G = \{(S,P),(P,S),(P,P)\}$ and therefore
$P(F \mid G) = \frac{P(F\cap G)}{P(G)} = \frac{0.06+0.07+0.14}{0.06+0.07+0.14+0.2+0.21} = 0.397.$

1.4.11 Let L, W and H be the events that the length, width and height respectively are within the specified tolerance limits.

It is given that $P(W) = 0.86$, $P(L \cap W \cap H) = 0.80$, $P(L \cap W \cap H') = 0.02$, $P(L' \cap W \cap H) = 0.03$ and $P(W \cup H) = 0.92$.

(a) $P(W \cap H) = P(L \cap W \cap H) + P(L' \cap W \cap H) = 0.80 + 0.03 = 0.83$
$P(H) = P(W \cup H) - P(W) + P(W \cap H) = 0.92 - 0.86 + 0.83 = 0.89$
$P(W \cap H \mid H) = \frac{P(W\cap H)}{P(H)} = \frac{0.83}{0.89} = 0.9326$

(b) $P(L \cap W) = P(L \cap W \cap H) + P(L \cap W \cap H') = 0.80 + 0.02 = 0.82$
$P(L \cap W \cap H \mid L \cap W) = \frac{P(L\ \cap\ W\ \cap\ H)}{P(L\ \cap\ W)} = \frac{0.80}{0.82} = 0.9756$

1.4.12 Let A be the event that the gene is of *'type A'*, and let D be the event that the gene is *'dominant'*.

$P(D \mid A') = 0.31$

$P(A' \cap D) = 0.22$

Therefore,

$P(A) = 1 - P(A')$

$= 1 - \frac{P(A'\cap D)}{P(D|A')}$

$= 1 - \frac{0.22}{0.31} = 0.290$

1.4.13 (a) Let E be the event that the *'component passes on performance'*, let A be the event that the *'component passes on appearance'*, and let C be the event that the *'component passes on cost'*.
$P(A \cap C) = 0.40$

$P(E \cap A \cap C) = 0.31$
$P(E) = 0.64$
$P(E' \cap A' \cap C') = 0.19$
$P(E' \cap A \cap C') = 0.06$
Therefore,
$P(E' \cap A' \cap C) = P(E' \cap A') - P(E' \cap A' \cap C')$
$= P(E') - P(E' \cap A) - 0.19$
$= 1 - P(E) - P(E' \cap A \cap C) - P(E' \cap A \cap C') - 0.19$
$= 1 - 0.64 - P(A \cap C) + P(E \cap A \cap C) - 0.06 - 0.19$
$= 1 - 0.64 - 0.40 + 0.31 - 0.06 - 0.19 = 0.02$

(b) $P(E \cap A \cap C \mid A \cap C) = \frac{P(E \cap A \cap C)}{P(A \cap C)}$
$= \frac{0.31}{0.40} = 0.775$

1.4.14 (a) Let T be the event *'good taste'*, let S be the event *'good size'*, and let A be the event *'good appearance'*.
$P(T) = 0.78$
$P(T \cap S) = 0.69$
$P(T \cap S' \cap A) = 0.05$
$P(S \cup A) = 0.84$
Therefore,
$P(S \mid T) = \frac{P(T \cap S)}{P(T)} = \frac{0.69}{0.78} = 0.885.$

(b) Notice that
$P(S' \cap A') = 1 - P(S \cup A) = 1 - 0.84 = 0.16.$
Also,
$P(T \cap S') = P(T) - P(T \cap S) = 0.78 - 0.69 = 0.09$
so that
$P(T \cap S' \cap A') = P(T \cap S') - P(T \cap S' \cap A) = 0.09 - 0.05 = 0.04.$
Therefore,
$P(T \mid S' \cap A') = \frac{P(T \cap S' \cap A')}{P(S' \cap A')} = \frac{0.04}{0.16} = 0.25.$

1.4.15 $P(\text{delay}) = (P(\text{delay} \mid \text{technical problems}) \times P(\text{technical problems}))$

$+ (P(\text{delay} \mid \text{no technical problems}) \times P(\text{no technical problems}))$

$= (1 \times 0.04) + (0.33 \times 0.96) = 0.3568$

1.4.16 Let S be the event that a chip *'survives 500 temperature cycles'* and let A be the event that the chip was *'made by company A'*.

$P(S) = 0.42$

$P(A \mid S') = 0.73$

Therefore,
$P(A' \cap S') = P(S') \times P(A' \mid S') = (1 - 0.42) \times (1 - 0.73) = 0.1566.$

1.5 Probabilities of Event Intersections

1.5.1 (a) $P(\text{both cards are picture cards}) = \frac{12}{52} \times \frac{11}{51} = \frac{132}{2652}$

(b) $P(\text{both cards are from red suits}) = \frac{26}{52} \times \frac{25}{51} = \frac{650}{2652}$

(c) $P(\text{one card is from a red suit and one is from black suit})$
$= (P(\text{first card is red}) \times P(\text{2nd card is black} \mid \text{1st card is red}))$
$+ (P(\text{first card is black}) \times P(\text{2nd card is red} \mid \text{1st card is black}))$
$= \left(\frac{26}{52} \times \frac{26}{51}\right) + \left(\frac{26}{52} \times \frac{26}{51}\right) = \frac{676}{2652} \times 2 = \frac{26}{51}$

1.5.2 (a) $P(\text{both cards are picture cards}) = \frac{12}{52} \times \frac{12}{52} = \frac{9}{169}$
The probability increases with replacement.

(b) $P(\text{both cards are from red suits}) = \frac{26}{52} \times \frac{26}{52} = \frac{1}{4}$
The probability increases with replacement.

(c) $P(\text{one card is from a red suit and one is from black suit})$
$= (P(\text{first card is red}) \times P(\text{2nd card is black} \mid \text{1st card is red}))$
$+ (P(\text{first card is black}) \times P(\text{2nd card is red} \mid \text{1st card is black}))$
$= \left(\frac{26}{52} \times \frac{26}{52}\right) + \left(\frac{26}{52} \times \frac{26}{52}\right) = \frac{1}{2}$
The probability decreases with replacement.

1.5.3 (a) No, they are not independent.
Notice that
$P((ii)) = \frac{3}{13} \neq P((ii) \mid (i)) = \frac{11}{51}$.

(b) Yes, they are independent.
Notice that
$P((i) \cap (ii)) = P((i)) \times P((ii))$
since
$P((i)) = \frac{1}{4}$
$P((ii)) = \frac{3}{13}$
and
$P((i) \cap (ii)) = P(\text{first card a heart picture} \cap (ii))$
$+ P(\text{first card a heart but not a picture} \cap (ii))$
$= \left(\frac{3}{52} \times \frac{11}{51}\right) + \left(\frac{10}{52} \times \frac{12}{51}\right) = \frac{153}{2652} = \frac{3}{52}$.

(c) No, they are not independent.
Notice that
$P((ii)) = \frac{1}{2} \neq P((ii) \mid (i)) = \frac{25}{51}$.

(d) Yes, they are independent.
Similar to part (b).

(e) No, they are not independent.

1.5.4 P(all four cards are hearts) $= P$(lst card is a heart)
$\times P$(2nd card is a heart | lst card is a heart)
$\times P$(3rd card is a heart | 1st and 2nd cards are hearts)
$\times P$(4th card is a heart | 1st, 2nd and 3rd cards are hearts)
$= \frac{13}{52} \times \frac{12}{51} \times \frac{11}{50} \times \frac{10}{49} = 0.00264$

P(all 4 cards from red suits) $= P$(1st card from red suit)
$\times P$(2nd card is from red suit | lst card is from red suit)
$\times P$(3rd card is from red suit | 1st and 2nd cards are from red suits)
$\times P$(4th card is from red suit | 1st, 2nd and 3rd cards are from red suits)
$= \frac{26}{52} \times \frac{25}{51} \times \frac{24}{50} \times \frac{23}{49} = 0.055$

P(all 4 cards from different suits) $= P$(1st card from any suit)
$\times P$(2nd card not from suit of 1st card)
$\times P$(3rd card not from suit of 1st or 2nd cards)
$\times P$(4th card not from suit of 1st, 2nd, or 3rd cards)
$= 1 \times \frac{39}{51} \times \frac{26}{50} \times \frac{13}{49} = 0.105$

1.5.5 P(all 4 cards are hearts) $= (\frac{13}{52})^4 = \frac{1}{256}$
The probability increases with replacement.

P(all 4 cards are from red suits) $= (\frac{26}{52})^4 = \frac{1}{16}$
The probability increases with replacement.

P(all 4 cards from different suits) $= 1 \times \frac{39}{52} \times \frac{26}{52} \times \frac{13}{52} = \frac{3}{32}$
The probability decreases with replacement.

1.5.6 The events A and B are independent so that $P(A \mid B) = P(A)$, $P(B \mid A) = P(B)$, and $P(A \cap B) = P(A)P(B)$.

To show that two events are independent it needs to be shown that one of the above three conditions holds.

(a) Recall that
$P(A \cap B) + P(A \cap B') = P(A)$
and

$P(B) + P(B') = 1.$
Therefore,
$P(A \mid B') = \frac{P(A \cap B')}{P(B')}$
$= \frac{P(A) - P(A \cap B)}{1 - P(B)}$
$= \frac{P(A) - P(A)P(B)}{1 - P(B)}$
$= \frac{P(A)(1 - P(B))}{1 - P(B)}$
$= P(A).$

(b) Similar to part (a).

(c) $P(A' \cap B') + P(A' \cap B) = P(A')$
so that
$P(A' \cap B') = P(A) - P(A' \cap B) = P(A) - P(A')P(B)$
since the events A' and B are independent.
Therefore,
$P(A' \cap B') = P(A)(1 - P(B)) = P(A')P(B').$

1.5.7 The only way that a message will not get through the network is if both branches are closed at the same time. The branches are independent since the switches operate independently of each other.

Therefore,

P(message gets through the network)

$= 1 - P$(message cannot get through the top branch or the bottom branch)

$= 1 - (P$(message cannot get through the top branch)

$\times$ P(message cannot get through the bottom branch)).

Also,

P(message gets through the top branch) $= P$(switch 1 is open $\cap$ switch 2 is open)

$= P$(switch 1 is open) $\times$ P(switch 2 is open)

$= 0.88 \times 0.92 = 0.8096$

since the switches operate independently of each other.

Therefore,

P(message cannot get through the top branch)

$= 1 - P$(message gets through the top branch)

$= 1 - 0.8096 = 0.1904.$

Furthermore,

P(message cannot get through the bottom branch)

$= P(\text{switch 3 is closed}) = 1 - 0.9 = 0.1.$

Therefore,

$P(\text{message gets through the network}) = 1 - (0.1 \times 0.1904) = 0.98096.$

1.5.8 Given the birthday of the first person, the second person has a different birthday with a probability $\frac{364}{365}$.

The third person has a different birthday from the first two people with a probability $\frac{363}{365}$, and so the probability that all three people have different birthdays is

$1 \times \frac{364}{365} \times \frac{363}{365}$.

Continuing in this manner the probability that n people all have different birthdays is therefore

$\frac{364}{365} \times \frac{363}{365} \times \frac{362}{365} \times \ldots \times \frac{366-n}{365}$

and

$P(\text{at least 2 people out of } n \text{ share the same birthday})$

$= 1 - P(n \text{ people all have different birthdays})$

$= 1 - \left(\frac{364}{365} \times \frac{363}{365} \times \ldots \frac{366-n}{365}\right)$.

This probability is equal to 0.117 for $n = 10$,

is equal to 0.253 for $n = 15$,

is equal to 0.411 for $n = 20$,

is equal to 0.569 for $n = 25$,

is equal to 0.706 for $n = 30$,

and is equal to 0.814 for $n = 35$.

The smallest values of n for which the probability is greater than 0.5 is $n = 23$.

Note that in these calculations it has been assumed that birthdays are equally likely to occur on any day of the year, although in practice seasonal variations may be observed in the number of births.

1.5.9 $P(\text{no broken bulbs}) = \frac{83}{100} \times \frac{82}{99} \times \frac{81}{98} = 0.5682$

$P(\text{one broken bulb}) = P(\text{broken, not broken, not broken})$

$+ P(\text{not broken, broken, not broken}) + P(\text{not broken, not broken, broken})$

$= \left(\frac{17}{100} \times \frac{83}{99} \times \frac{82}{98}\right) + \left(\frac{83}{100} \times \frac{17}{99} \times \frac{82}{98}\right) + \left(\frac{83}{100} \times \frac{82}{99} \times \frac{17}{98}\right) = 0.3578$

$P(\text{no more than one broken bulb in the sample})$

$= P(\text{no broken bulbs}) + P(\text{one broken bulb})$

$= 0.5682 + 0.3578 = 0.9260$

1.5.10 $P(\text{no broken bulbs}) = \frac{83}{100} \times \frac{83}{100} \times \frac{83}{100} = 0.5718$

$P(\text{one broken bulb}) = P(\text{broken, not broken, not broken})$
$+ P(\text{not broken, broken, not broken}) + P(\text{not broken, not broken, broken})$
$= \left(\frac{17}{100} \times \frac{83}{100} \times \frac{83}{100}\right) + \left(\frac{83}{100} \times \frac{17}{100} \times \frac{83}{100}\right) + \left(\frac{83}{100} \times \frac{83}{100} \times \frac{17}{100}\right) = 0.3513$

$P(\text{no more than one broken bulb in the sample})$
$= P(\text{no broken bulbs}) + P(\text{one broken bulb})$
$= 0.5718 + 0.3513 = 0.9231$

The probability of finding no broken bulbs increases with replacement, but the probability of finding no more than one broken bulb decreases with replacement.

1.5.11 $P(\text{drawing 2 green balls})$
$= P(\text{1st ball is green}) \times P(\text{2nd ball is green} \mid \text{1st ball is green})$
$= \frac{72}{169} \times \frac{71}{168} = 0.180$

$P(\text{two balls same color})$
$= P(\text{two red balls}) + P(\text{two blue balls}) + P(\text{two green balls})$
$= \left(\frac{43}{169} \times \frac{42}{168}\right) + \left(\frac{54}{169} \times \frac{53}{168}\right) + \left(\frac{72}{169} \times \frac{71}{168}\right) = 0.344$

$P(\text{two balls different colors}) = 1 - P(\text{two balls same color})$
$= 1 - 0.344 = 0.656$

1.5.12 $P(\text{drawing 2 green balls}) = \frac{72}{169} \times \frac{72}{169} = 0.182$

$P(\text{two balls same color})$
$= P(\text{two red balls}) + P(\text{two blue balls}) + P(\text{two green balls})$
$= \left(\frac{43}{169} \times \frac{43}{169}\right) + \left(\frac{54}{169} \times \frac{54}{169}\right) + \left(\frac{72}{169} \times \frac{72}{169}\right) = 0.348$

$P(\text{two balls different colors}) = 1 - P(\text{two balls same color})$
$= 1 - 0.348 = 0.652$

The probability that the two balls are green increases with replacement while the probability of drawing two balls of different colors decreases with replacement.

1.5.13 $P(\text{same result on both throws}) = P(\text{both heads}) + P(\text{both tails})$
$= p^2 + (1-p)^2 = 2p^2 - 2p + 1 = 2(p-0.5)^2 + 0.5$
which is minimized when $p = 0.5$ (a fair coin).

1.5.14 $P(\text{each score is obtained exactly once})$
$= 1 \times \frac{5}{6} \times \frac{4}{6} \times \frac{3}{6} \times \frac{2}{6} \times \frac{1}{6} = \frac{5}{324}$
$P(\text{no sixes in seven rolls}) = \left(\frac{5}{6}\right)^7 = 0.279$

1.5.15 (a) $\left(\frac{1}{2}\right)^5 = \frac{1}{32}$

(b) $1 \times \frac{5}{6} \times \frac{4}{6} = \frac{5}{9}$

(c) $P(BBR) + P(BRB) + P(RBB)$
$= \left(\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\right) + \left(\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\right) + \left(\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\right)$
$= \frac{3}{8}$

(d) $P(BBR) + P(BRB) + P(RBB)$
$= \left(\frac{26}{52} \times \frac{25}{51} \times \frac{26}{50}\right) + \left(\frac{26}{52} \times \frac{26}{51} \times \frac{25}{50}\right) + \left(\frac{26}{52} \times \frac{26}{51} \times \frac{25}{50}\right)$
$= \frac{13}{34}$

1.5.16 $1 - (1 - 0.90)^n \geq 0.995$
is satisfied for $n \geq 3$.

1.5.17 Claims from clients in the same geographical area would not be independent of each other since they would all be affected by the same flooding events.

1.5.18 (a) $P(\text{system works}) = 0.88 \times 0.78 \times 0.92 \times 0.85 = 0.537$

(b) $P(\text{system works}) = 1 - P(\text{no computers working})$
$= 1 - ((1 - 0.88) \times (1 - 0.78) \times (1 - 0.92) \times (1 - 0.85)) = 0.9997$

(c) $P(\text{system works}) = P(\text{all computers working})$
$+ P(\text{computers 1,2,3 working, computer 4 not working})$
$+ P(\text{computers 1,2,4 working, computer 3 not working})$
$+ P(\text{computers 1,3,4 working, computer 2 not working})$
$+ P(\text{computers 2,3,4 working, computer 1 not working})$
$= 0.537 + (0.88 \times 0.78 \times 0.92 \times (1 - 0.85)) + (0.88 \times 0.78 \times (1 - 0.92) \times 0.85)$
$+ (0.88 \times (1 - 0.78) \times 0.92 \times 0.85) + ((1 - 0.88) \times 0.78 \times 0.92 \times 0.85)$
$= 0.903$

1.6 Posterior Probabilities

1.6.1 (a) The following information is given:
$P(\text{disease}) = 0.01$
$P(\text{no disease}) = 0.99$
$P(\text{positive blood test} \mid \text{disease}) = 0.97$
$P(\text{positive blood test} \mid \text{no disease}) = 0.06$
Therefore,
$P(\text{positive blood test}) = (P(\text{positive blood test} \mid \text{disease}) \times P(\text{disease}))$
$+ \ (P(\text{positive blood test} \mid \text{no disease}) \times P(\text{no disease}))$
$= (0.97 \times 0.01) + (0.06 \times 0.99) = 0.0691.$

(b) $P(\text{disease} \mid \text{positive blood test})$
$= \frac{P(\text{positive blood test} \cap \text{disease})}{P(\text{positive blood test})}$
$= \frac{P(\text{positive blood test} \mid \text{disease}) \times P(\text{disease})}{P(\text{positive blood test})}$
$= \frac{0.97 \times 0.01}{0.0691} = 0.1404$

(c) $P(\text{no disease} \mid \text{negative blood test})$
$= \frac{P(\text{no disease} \cap \text{negative blood test})}{P(\text{negative blood test})}$
$= \frac{P(\text{negative blood test} \mid \text{no disease}) \times P(\text{no disease})}{1 - P(\text{positive blood test})}$
$= \frac{(1-0.06)\times 0.99}{(1-0.0691)} = 0.9997$

1.6.2 (a) $P(\text{red}) = (P(\text{red} \mid \text{bag 1}) \times P(\text{bag 1})) + (P(\text{red} \mid \text{bag 2}) \times P(\text{bag 2}))$
$+ \ (P(\text{red} \mid \text{bag 3}) \times P(\text{bag 3}))$
$= \left(\frac{1}{3} \times \frac{3}{10}\right) + \left(\frac{1}{3} \times \frac{8}{12}\right) + \left(\frac{1}{3} \times \frac{5}{16}\right) = 0.426$

(b) $P(\text{blue}) = 1 - P(\text{red}) = 1 - 0.426 = 0.574$

(c) $P(\text{red ball from bag 2}) = P(\text{bag 2}) \times P(\text{red ball} \mid \text{bag 2})$
$= \frac{1}{3} \times \frac{8}{12} = \frac{2}{9}$

$P(\text{bag 1} \mid \text{red ball}) = \frac{P(\text{bag 1} \cap \text{red ball})}{P(\text{red ball})}$
$= \frac{P(\text{bag 1}) \times P(\text{red ball} \mid \text{bag 1})}{P(\text{red ball})}$
$= \frac{\frac{1}{3} \times \frac{3}{10}}{0.426} = 0.235$

$P(\text{bag 2} \mid \text{blue ball}) = \frac{P(\text{bag 2} \cap \text{blue ball})}{P(\text{blue ball})}$
$= \frac{P(\text{bag 2}) \times P(\text{blue ball} \mid \text{bag 1})}{P(\text{blue ball})}$

$= \frac{\frac{1}{3}\times\frac{4}{12}}{0.574} = 0.194$

1.6.3 (a) $P(\text{Section I}) = \frac{55}{100}$

(b) $P(\text{grade is A})$
$= (P(\text{A} \mid \text{Section I}) \times P(\text{Section I})) + (P(\text{A} \mid \text{Section II}) \times P(\text{Section II}))$
$= \left(\frac{10}{55} \times \frac{55}{100}\right) + \left(\frac{11}{45} \times \frac{45}{100}\right) = \frac{21}{100}$

(c) $P(\text{A} \mid \text{Section I}) = \frac{10}{55}$

(d) $P(\text{Section I} \mid \text{A}) = \frac{P(\text{A} \cap \text{Section I})}{P(\text{A})}$
$= \frac{P(\text{Section I})\times P(\text{A} \mid \text{Section I})}{P(\text{A})}$
$= \frac{\frac{55}{100}\times\frac{10}{55}}{\frac{21}{100}} = \frac{10}{21}$

1.6.4 The following information is given:

$P(\text{Species 1}) = 0.45$

$P(\text{Species 2}) = 0.38$

$P(\text{Species 3}) = 0.17$

$P(\text{Tagged} \mid \text{Species 1}) = 0.10$

$P(\text{Tagged} \mid \text{Species 2}) = 0.15$

$P(\text{Tagged} \mid \text{Species 3}) = 0.50$

Therefore,

$P(\text{Tagged}) = (P(\text{Tagged} \mid \text{Species 1}) \times P(\text{Species 1}))$

$+ (P(\text{Tagged} \mid \text{Species 2}) \times P(\text{Species 2})) + (P(\text{Tagged} \mid \text{Species 3}) \times P(\text{Species 3}))$

$= (0.10 \times 0.45) + (0.15 \times 0.38) + (0.50 \times 0.17) = 0.187.$

$P(\text{Species 1} \mid \text{Tagged}) = \frac{P(\text{Tagged} \cap \text{Species 1})}{P(\text{Tagged})}$
$= \frac{P(\text{Species 1})\times P(\text{Tagged} \mid \text{Species 1})}{P(\text{Tagged})}$
$= \frac{0.45\times 0.10}{0.187} = 0.2406$

$P(\text{Species 2} \mid \text{Tagged}) = \frac{P(\text{Tagged} \cap \text{Species 2})}{P(\text{Tagged})}$
$= \frac{P(\text{Species 2})\times P(\text{Tagged} \mid \text{Species 2})}{P(\text{Tagged})}$
$= \frac{0.38\times 0.15}{0.187} = 0.3048$

$P(\text{Species 3} \mid \text{Tagged}) = \frac{P(\text{Tagged} \cap \text{Species 3})}{P(\text{Tagged})}$

$= \frac{P(\text{Species 3}) \times P(\text{Tagged} \mid \text{Species 3})}{P(\text{Tagged})}$

$= \frac{0.17 \times 0.50}{0.187} = 0.4545$

1.6.5 (a) $P(\text{fail}) = (0.02 \times 0.77) + (0.10 \times 0.11) + (0.14 \times 0.07) + (0.25 \times 0.05)$
$= 0.0487$

$P(C \mid \text{fail}) = \frac{0.14 \times 0.07}{0.0487} = 0.2012$

$P(D \mid \text{fail}) = \frac{0.25 \times 0.05}{0.0487} = 0.2567$

The answer is $0.2012 + 0.2567 = 0.4579$.

(b) $P(A \mid \text{did not fail}) = \frac{P(A) \times P(\text{did not fail} \mid A)}{P(\text{did not fail})}$
$= \frac{0.77 \times (1-0.02)}{1-0.0487} = 0.7932$

1.6.6 $P(C) = 0.15$
$P(W) = 0.25$
$P(H) = 0.60$
$P(R \mid C) = 0.30$
$P(R \mid W) = 0.40$
$P(R \mid H) = 0.50$
Therefore,
$P(C \mid R') = \frac{P(R'|C)P(C)}{P(R'|C)P(C)+P(R'|W)P(W)+P(R'|H)P(H)}$
$= \frac{(1-0.30)\times 0.15}{((1-0.30)\times 0.15)+((1-0.40)\times 0.25)+((1-0.50)\times 0.60)}$
$= 0.189$

1.6.7 (a) $P(C) = 0.12$
$P(M) = 0.55$
$P(W) = 0.20$
$P(H) = 0.13$
$P(L \mid C) = 0.003$
$P(L \mid M) = 0.009$
$P(L \mid W) = 0.014$
$P(L \mid H) = 0.018$
Therefore,
$P(H \mid L) = \frac{P(L|H)P(H)}{P(L|C)P(C)+P(L|M)P(M)+P(L|W)P(W)+P(L|H)P(H)}$
$= \frac{0.018\times 0.13}{(0.003\times 0.12)+(0.009\times 0.55)+(0.014\times 0.20)+(0.018\times 0.13)}$
$= 0.224$

(b) $P(M \mid L') = \frac{P(L'|M)P(M)}{P(L'|C)P(C)+P(L'|M)P(M)+P(L'|W)P(W)+P(L'|H)P(H)}$
$= \frac{0.991\times0.55}{(0.997\times0.12)+(0.991\times0.55)+(0.986\times0.20)+(0.982\times0.13)}$
$= 0.551$

1.6.8 (a) $P(A) = 0.12$
$P(B) = 0.34$
$P(C) = 0.07$
$P(D) = 0.25$
$P(E) = 0.22$
$P(M \mid A) = 0.19$
$P(M \mid B) = 0.50$
$P(M \mid C) = 0.04$
$P(M \mid D) = 0.32$
$P(M \mid E) = 0.76$
Therefore,
$P(C \mid M) = \frac{P(M|C)P(C)}{P(M|A)P(A)+P(M|B)P(B)+P(M|C)P(C)+P(M|D)P(D)+P(M|E)P(E)}$
$= \frac{0.04\times0.07}{(0.19\times0.12)+(0.50\times0.34)+(0.04\times0.07)+(0.32\times0.25)+(0.76\times0.22)}$
$= 0.0063$

(b) $P(D \mid M') = \frac{P(M'|D)P(D)}{P(M'|A)P(A)+P(M'|B)P(B)+P(M'|C)P(C)+P(M'|D)P(D)+P(M'|E)P(E)}$
$= \frac{0.68\times0.25}{(0.81\times0.12)+(0.50\times0.34)+(0.96\times0.07)+(0.68\times0.25)+(0.24\times0.22)}$
$= 0.305$

1.7 Counting Techniques

1.7.1 (a) $7! = 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 5040$

(b) $8! = 8 \times 7! = 40320$

(c) $4! = 4 \times 3 \times 2 \times 1 = 24$

(d) $13! = 13 \times 12 \times 11 \times \ldots \times 1 = 6{,}227{,}020{,}800$

1.7.2 (a) $P_2^7 = \frac{7!}{(7-2)!} = 7 \times 6 = 42$

(b) $P_5^9 = \frac{9!}{(9-5)!} = 9 \times 8 \times 7 \times 6 \times 5 = 15120$

(c) $P_2^5 = \frac{5!}{(5-2)!} = 5 \times 4 = 20$

(d) $P_4^{17} = \frac{17!}{(17-4)!} = 17 \times 16 \times 15 \times 14 = 57120$

1.7.3 (a) $C_2^6 = \frac{6!}{(6-2)! \times 2!} = \frac{6 \times 5}{2} = 15$

(b) $C_4^8 = \frac{8!}{(8-4)! \times 4!} = \frac{8 \times 7 \times 6 \times 5}{24} = 70$

(c) $C_2^5 = \frac{5!}{(5-2)! \times 2!} = \frac{5 \times 4}{2} = 10$

(d) $C_6^{14} = \frac{14!}{(14-6)! \times 6!} = 3003$

1.7.4 The number of full meals is $5 \times 3 \times 7 \times 6 \times 8 = 5040$.

The number of meals with just soup or appetizer is $(5 + 3) \times 7 \times 6 \times 8 = 2688$.

1.7.5 The number of experimental configurations is $3 \times 4 \times 2 = 24$.

1.7.6 (a) Let the notation (2,3,1,4) represent the result that the player who finished 1st in tournament 1 finished 2nd in tournament 2, the player who finished 2nd in tournament 1 finished 3rd in tournament 2, the player who finished 3rd in tournament 1 finished 1st in tournament 2, and the player who finished 4th in tournament 1 finished 4th in tournament 2.

Then the result (1,2,3,4) indicates that each competitor received the same ranking in both tournaments.

Altogether there are 4! = 24 different results, each equally likely, and so this single result has a probability of $\frac{1}{24}$.

(b) The results where no player receives the same ranking in the two tournaments are:
(2,1,4,3), (2,3,4,1), (2,4,1,3), (3,1,4,2), (3,4,1,2) (3,4,2,1), (4,1,2,3), (4,3,1,2), (4,3,2,1)
There are nine of these results and so the required probability is $\frac{9}{24} = \frac{3}{8}$.

1.7.7 The number of rankings that can be assigned to the top 5 competitors is
$P_5^{20} = \frac{20!}{15!} = 20 \times 19 \times 18 \times 17 \times 16 = 1{,}860{,}480$.

The number of ways in which the best 5 competitors can be chosen is
$C_5^{20} = \frac{20!}{15! \times 5!} = 15504$.

1.7.8 (a) $C_3^{100} = \frac{100!}{97! \times 3!} = \frac{100 \times 99 \times 98}{6} = 161700$

(b) $C_3^{83} = \frac{83!}{80! \times 3!} = \frac{83 \times 82 \times 81}{6} = 91881$

(c) $P(\text{no broken lightbulbs}) = \frac{91881}{161700} = 0.568$

(d) $17 \times C_2^{83} = 17 \times \frac{83 \times 82}{2} = 57851$

(e) The number of samples with 0 or 1 broken bulbs is
$91881 + 57851 = 149732$.
$P(\text{sample contains no more than 1 broken bulb}) = \frac{149732}{161700} = 0.926$

1.7.9 $C_k^{n-1} + C_{k-1}^{n-1} = \frac{(n-1)!}{k!(n-1-k)!} + \frac{(n-1)!}{(k-1)!(n-k)!} = \frac{n!}{k!(n-k)!}\left(\frac{n-k}{n} + \frac{k}{n}\right) = \frac{n!}{k!(n-k)!} = C_k^n$

This relationship can be interpreted in the following manner.

C_k^n is the number of ways that k balls can be selected from n balls. Suppose that one ball is red while the remaining $n-1$ balls are blue. Either all k balls selected are blue or one of the selected balls is red. C_k^{n-1} is the number of ways k blue balls can be selected while C_{k-1}^{n-1} is the number of ways of selecting the one red ball and $k-1$ blue balls.

1.7.10 (a) The number of possible 5 card hands is $C_5^{52} = \frac{52!}{47! \times 5!}$ =2,598,960.

(b) The number of ways to get a hand of 5 hearts is $C_5^{13} = \frac{13!}{8! \times 5!} = 1287$.

(c) The number of ways to get a flush is $4 \times C_5^{13} = 4 \times 1{,}287 = 5148$.

(d) $P(\text{flush}) = \frac{5148}{2{,}598{,}960} = 0.00198$.

(e) There are 48 choices for the fifth card in the hand and so the number of hands containing all four aces is 48.

(f) $13 \times 48 = 624$

(g) $P(\text{hand has four cards of the same number or picture}) = \frac{624}{2{,}598{,}960} = 0.00024$.

1.7.11 There are $n!$ ways in which n objects can be arranged in a line. If the line is made into a circle and rotations of the circle are considered to be indistinguishable, then there are n arrangements of the line corresponding to each arrangement of the circle. Consequently, there are $\frac{n!}{n} = (n-1)!$ ways to order the objects in a circle.

1.7.12 The number of ways that six people can sit in a line at a cinema is $6! = 720$.

See the previous problem.

The number of ways that six people can sit around a dinner table is $5! = 120$.

1.7.13 Consider 5 blocks, one block being Andrea and Scott and the other four blocks being the other four people. At the cinema these 5 blocks can be arranged in 5! ways, and then Andrea and Scott can be arranged in two different ways within their block, so that the total number of seating arrangements is $2 \times 5! = 240$.

Similarly, the total number of seating arrangements at the dinner table is $2 \times 4! = 48$.

If Andrea refuses to sit next to Scott then the number of seating arrangements can be obtained by subtraction. The total number of seating arrangements at the cinema is $720 - 240 = 480$ and the total number of seating arrangements at the dinner table is $120 - 48 = 72$.

1.7.14 The total number of arrangements of n balls is $n!$ which needs to be divided by $n_1!$ because the rearrangements of the n_1 balls in box 1 are indistinguishable, and similarly it needs to be divided by $n_2! \ldots n_k!$ due to the indistinguishable rearrangements possible in boxes 2 to k.

When $k = 2$ the problem is equivalent to the number of ways of selecting n_1 balls (or n_2 balls) from $n = n_1 + n_2$ balls.

1.7.15 (a) Using the result provided in the previous problem the answer is $\frac{12!}{3! \times 4! \times 5!} = 27720$.

(b) Suppose that the balls in part (a) are labelled from 1 to 12. Then the positions of the three red balls in the line (where the places in the line are labelled 1 to

12) can denote which balls in part (a) are placed in the first box, the positions of the four blue balls in the line can denote which balls in part (a) are placed in the second box, and the positions of the five green balls in the line can denote which balls in part (a) are placed in the third box. Thus, there is a one-to-one correspondence between the positioning of the colored balls in part (b) and the arrangements of the balls in part (a) so that the problems are identical.

1.7.16 $\frac{14!}{3!\times 4!\times 7!} = 120120$

1.7.17 $\frac{15!}{3!\times 3!\times 3!\times 3!\times 3!} = 168{,}168{,}000$

1.7.18 The total number of possible samples is C_{12}^{60}.

(a) The number of samples containing only items which have either excellent or good quality is C_{12}^{43}.
Therefore, the answer is
$\frac{C_{12}^{43}}{C_{12}^{60}} = \frac{43}{60} \times \frac{42}{59} \ldots \times \frac{32}{49} = 0.0110.$

(b) The number of samples that contain three items of excellent quality, three items of good quality, three items of poor quality and three defective items is
$C_3^{18} \times C_3^{25} \times C_3^{12} \times C_3^{5}$ =4,128,960,000.
Therefore, the answer is
$\frac{4{,}128{,}960{,}000}{C_{12}^{60}} = 0.00295.$

1.7.19 The ordering of the visits can be made in $10! = 3{,}628{,}800$ different ways.

The number of different ways the ten cities be split into two groups of five cities is $C_5^{10} = 252$.

1.7.20 $\binom{26}{2} \times \binom{26}{3} = 845000$

1.7.21 (a) $\frac{\binom{39}{8}}{\binom{52}{8}} = \frac{39}{52} \times \frac{38}{51} \times \frac{37}{50} \times \frac{36}{49} \times \frac{35}{48} \times \frac{34}{47} \times \frac{33}{46} \times \frac{32}{45} = 0.082$

(b) $\frac{\binom{13}{2} \times \binom{13}{2} \times \binom{13}{2} \times \binom{13}{2}}{\binom{52}{8}} = 0.049$

1.7.22 $$\frac{\binom{5}{2} \times \binom{30}{4} \times \binom{5}{2}}{\binom{40}{8}} = 0.0356$$

1.9 Supplementary Problems

1.9.1 $\mathcal{S} = \{\ 1,\ 1.5,\ 2,\ 2.5,\ 3,\ 3.5,\ 4,\ 4.5,\ 5,\ 5.5,\ 6\}$

1.9.2 If the four contestants are labelled A, B, C, D and the notation (X, Y) is used to indicate that contestant X is the winner and contestant Y is the runner up, then the sample space is:

$\mathcal{S} = \{(A, B), (A, C), (A, D), (B, A), (B, C), (B, D),$

$(C, A), (C, B), (C, D), (D, A), (D, B), (D, C)\}$

1.9.3 One way is to have the two team captains each toss the coin once. If one obtains a head and the other a tail, then the one with the head wins (this could just as well be done the other way around so that the one with the tail wins, as long as it is decided beforehand). If both captains obtain the same result, that is if there are two heads or two tails, then the procedure could be repeated until different results are obtained.

1.9.4 See Figure 1.10.

There are 36 equally likely outcomes, 16 of which have scores differing by no more than one.

Therefore,

$P(\text{the scores on two dice differ by no more than one}) = \frac{16}{36} = \frac{4}{9}$.

1.9.5 The number of ways to pick a card is 52.

The number of ways to pick a diamond picture card is 3.

Therefore,

$P(\text{picking a diamond picture card}) = \frac{3}{52}$.

1.9.6 With replacement:

$P(\text{drawing two hearts}) = \frac{13}{52} \times \frac{13}{52} = \frac{1}{16} = 0.0625$

Without replacement:

$P(\text{drawing two hearts}) = \frac{13}{52} \times \frac{12}{51} = \frac{3}{51} = 0.0588$

The probability decreases without replacement.

1.9.7 $A = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (3, 1)\}$

$B = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$

(a) $A \cap B = \{(1,1),(2,2)\}$
$P(A \cap B) = \frac{2}{36} = \frac{1}{18}$

(b) $A \cup B = \{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1),(3,3),(4,4),(5,5),(6,6)\}$
$P(A \cup B) = \frac{10}{36} = \frac{5}{18}$

(c) $A' \cup B = \{(1,1),(1,4),(1,5),(1,6),(2,2),(2,3),(2,4),(2,5),(2,6),$
$(3,2),(3,3),(3,4),(3,5),(3,6),(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),$
$(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$
$P(A' \cup B) = \frac{32}{36} = \frac{8}{9}$

1.9.8 See Figure 1.10.

Let the notation (x,y) indicate that the score on the red die is x and that the score on the blue die is y.

(a) The event *'the sum of the scores on the two dice is eight'*
consists of the outcomes:
$\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$

Therefore,
$P(\text{red die is 5} \mid \text{sum of scores is 8})$
$= \frac{P(\text{red die is 5} \cap \text{sum of scores is 8})}{P(\text{sum of scores is 8})}$
$= \frac{\left(\frac{1}{36}\right)}{\left(\frac{5}{36}\right)} = \frac{1}{5}.$

(b) $P(\text{either score is 5} \mid \text{sum of scores is 8}) = 2 \times \frac{1}{5} = \frac{2}{5}$

(c) The event *'the score on either die is 5'*
consists of the 11 outcomes:
$\{(1,5),(2,5),(3,5),(4,5),(5,5),(6,5),(5,6),(5,4),(5,3),(5,2),(5,1)\}$

Therefore,
$P(\text{sum of scores is 8} \mid \text{either score is 5})$
$= \frac{P(\text{sum of scores is 8} \cap \text{either score is 5})}{P(\text{either score is 5})}$
$= \frac{\left(\frac{2}{36}\right)}{\left(\frac{11}{36}\right)} = \frac{2}{11}.$

1.9.9 $P(A) = P(\text{either switch 1 or 4 is open or both})$
$= 1 - P(\text{both switches 1 and 4 are closed})$
$= 1 - 0.15^2 = 0.9775$

$P(B) = P(\text{either switch 2 or 5 is open or both})$

$= 1 - P(\text{both switches 2 and 5 are closed})$
$= 1 - 0.15^2 = 0.9775$

$P(C) = P(\text{switches 1 and 2 are both open}) = 0.85^2 = 0.7225$
$P(D) = P(\text{switches 4 and 5 are both open}) = 0.85^2 = 0.7225$

If $E = C \cup D$ then
$P(E) = 1 - (P(C') \times P(D'))$
$= 1 - (1 - 0.85^2)^2 = 0.923.$

Therefore,
$P(\text{message gets through the network})$
$= (P(\text{switch 3 is open}) \times P(A) \times P(B)) + (P(\text{switch 3 closed}) \times P(E))$
$= (0.85 \times (1 - 0.15^2)^2) + (0.15 \times (1 - (1 - 0.85^2)^2)) = 0.9506.$

1.9.10 The sample space for the experiment of two coin tosses consists of the four equally likely outcomes:
$\{(H, H), (H, T), (T, H), (T, T)\}$
Three out of these four outcomes contain at least one head, so that
$P(\text{at least one head in two coin tosses}) = \frac{3}{4}.$

The sample space for four tosses of a coin consists of $2^4 = 16$ equally likely outcomes of which the following 11 outcomes contain at least two heads:
$\{(HHTT), (HTHT), (HTTH), (THHT), (THTH), (TTHH),$
$(HHHT), (HHTH), (HTHH), (THHH), (HHHH)\}$
Therefore,
$P(\text{at least two heads in four coin tosses}) = \frac{11}{16}$
which is smaller than the previous probability.

1.9.11 (a) $P(\text{blue ball}) = (P(\text{bag } 1) \times P(\text{blue ball} \mid \text{bag } 1))$
$+ (P(\text{bag } 2) \times P(\text{blue ball} \mid \text{bag } 2))$
$+ (P(\text{bag } 3) \times P(\text{blue ball} \mid \text{bag } 3))$
$+ (P(\text{bag } 4) \times P(\text{blue ball} \mid \text{bag } 4))$
$= \left(0.15 \times \frac{7}{16}\right) + \left(0.2 \times \frac{8}{18}\right) + \left(0.35 \times \frac{9}{19}\right) + \left(0.3 \times \frac{7}{11}\right) = 0.5112$

(b) $P(\text{bag } 4 \mid \text{green ball}) = \frac{P(\text{green ball} \cap \text{bag } 4)}{P(\text{green ball})}$
$= \frac{P(\text{bag } 4) \times P(\text{green ball} \mid \text{bag } 4)}{P(\text{greenball})}$
$= \frac{0.3 \times 0}{P(\text{green ball})} = 0$

(c) $P(\text{bag 1} \mid \text{blue ball}) = \frac{P(\text{bag 1}) \times P(\text{blue ball} \mid \text{bag 1})}{P(\text{blue ball})}$

$= \frac{0.15 \times \frac{7}{16}}{0.5112} = \frac{0.0656}{0.5112} = 0.128$

1.9.12 (a) $\mathcal{S} = \{1, 2, 3, 4, 5, 6, 10\}$

(b) $P(10) = P(\text{score on die is 5}) \times P(\text{tails})$
$= \frac{1}{6} \times \frac{1}{2} = \frac{1}{12}$

(c) $P(3) = P(\text{score on die is 3}) \times P(\text{heads})$
$= \frac{1}{6} \times \frac{1}{2} = \frac{1}{12}$

(d) $P(6) = P(\text{score on die is 6}) + (P(\text{score on die is 3}) \times P(\text{tails}))$
$= \frac{1}{6} + (\frac{1}{6} \times \frac{1}{2})$
$= \frac{1}{4}$

(e) 0

(f) $P(\text{score on die is odd} \mid \text{6 is recorded})$
$= \frac{P(\text{score on die is odd} \cap \text{6 is recorded})}{P(\text{6 is recorded})}$
$= \frac{P(\text{score on die is 3}) \times P(\text{tails})}{P(\text{6 is recorded})}$
$= \frac{\left(\frac{1}{12}\right)}{\left(\frac{1}{4}\right)} = \frac{1}{3}$

1.9.13 $5^4 = 625$

$4^5 = 1024$

In this case $5^4 < 4^5$, and in general $n_2^{n_1} < n_1^{n_2}$ when $3 \leq n_1 < n_2$.

1.9.14 $\frac{20!}{5! \times 5! \times 5! \times 5!} = 1.17 \times 10^{10}$

$\frac{20!}{4! \times 4! \times 4! \times 4 \times 4!} = 3.06 \times 10^{11}$

1.9.15 $P(X = 0) = \frac{1}{4}$

$P(X = 1) = \frac{1}{2}$

$P(X = 2) = \frac{1}{4}$

$P(X = 0 \mid \text{white}) = \frac{1}{8}$

$P(X = 1 \mid \text{white}) = \frac{1}{2}$

$P(X = 2 \mid \text{white}) = \frac{3}{8}$

$P(X = 0 \mid \text{black}) = \frac{1}{2}$
$P(X = 1 \mid \text{black}) = \frac{1}{2}$
$P(X = 2 \mid \text{black}) = 0$

1.9.16 Let A be the event that *'the order is from a first time customer'* and let B be the event that *'the order is dispatched within one day'*.
It is given that $P(A) = 0.28$, $P(B \mid A) = 0.75$, and $P(A' \cap B') = 0.30$.

Therefore,
$P(A' \cap B) = P(A') - P(A' \cap B')$
$= (1 - 0.28) - 0.30 = 0.42$

$P(A \cap B) = P(A) \times P(B \mid A)$
$= 0.28 \times 0.75 = 0.21$

$P(B) = P(A' \cap B) + P(A \cap B)$
$= 0.42 + 0.21 = 0.63$

and
$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$
$= \frac{0.21}{0.63} = \frac{1}{3}$.

1.9.17 It is given that
$P(\text{Puccini}) = 0.26$
$P(\text{Verdi}) = 0.22$
$P(\text{other composer}) = 0.52$
$P(\text{female} \mid \text{Puccini}) = 0.59$
$P(\text{female} \mid \text{Verdi}) = 0.45$
and
$P(\text{female}) = 0.62$.

(a) Since
$P(\text{female}) = (P(\text{Puccini}) \times P(\text{female} \mid \text{Puccini}))$
$+ (P(\text{Verdi}) \times P(\text{female} \mid \text{Verdi}))$
$+ (P(\text{other composer}) \times P(\text{female} \mid \text{other composer}))$
it follows that
$0.62 = (0.26 \times 0.59) + (0.22 \times 0.45) + (0.52 \times P(\text{female} \mid \text{other composer}))$
so that
$P(\text{female} \mid \text{other composer}) = 0.7069$.

(b) $P(\text{Puccini} \mid \text{male}) = \frac{P(\text{Puccini}) \times P(\text{male} \mid \text{Puccini})}{P(\text{male})}$
$= \frac{0.26 \times (1-0.59)}{1-0.62} = 0.281$

1.9.18 The total number of possible samples is C_{10}^{92}.

(a) The number of samples that do not contain any fibers of polymer B is C_{10}^{75}.
Therefore, the answer is
$\frac{C_{10}^{75}}{C_{10}^{92}} = \frac{75}{92} \times \frac{74}{91} \ldots \times \frac{66}{83} = 0.115.$

(b) The number of samples that contain exactly one fiber of polymer B is $17 \times C_9^{75}$.
Therefore, the answer is
$\frac{17 \times C_9^{75}}{C_{10}^{92}} = 0.296.$

(c) The number of samples that contain three fibers of polymer A, three fibers of polymer B, and four fibers of polymer C is
$C_3^{43} \times C_3^{17} \times C_4^{32}$.
Therefore, the answer is
$\frac{C_3^{43} \times C_3^{17} \times C_4^{32}}{C_{10}^{92}} = 0.042.$

1.9.19 The total number of possible sequences of heads and tails is $2^5 = 32$, with each sequence being equally likely. Of these, sixteen don't include a sequence of three outcomes of the same kind.

Therefore, the required probability is
$\frac{16}{32} = 0.5.$

1.9.20 (a) Calls answered by an experienced operator that last over five minutes.

(b) Successfully handled calls that were answered either within ten seconds or by an inexperienced operator (or both).

(c) Calls answered after ten seconds that lasted more than five minutes and that were not handled successfully.

(d) Calls that were either answered within ten seconds and lasted less than five minutes, or that were answered by an experienced operator and were handled successfully.

1.9.21 (a) $\frac{20!}{7! \times 7! \times 6!} = 133,024,320$

(b) If the first and the second job are assigned to production line I, the number of assignments is
$\frac{18!}{5!\times 7!\times 6!} = 14,702,688.$

If the first and the second job are assigned to production line II, the number of assignments is
$\frac{18!}{7!\times 5!\times 6!} = 14,702,688.$

If the first and the second job are assigned to production line III, the number of assignments is

$\frac{18!}{7!\times 7!\times 4!} = 10,501,920.$

Therefore, the answer is
$14,702,688 + 14,702,688 + 10,501,920 = 39,907,296.$

(c) The answer is $133,024,320 - 39,907,296 = 93,117,024.$

1.9.22 (a) $\frac{\binom{13}{3}}{\binom{52}{3}} = \frac{13}{52} \times \frac{12}{51} \times \frac{11}{50} = 0.0129$

(b) $\frac{\binom{4}{1}\times\binom{4}{1}\times\binom{4}{1}}{\binom{52}{3}} = \frac{12}{52} \times \frac{8}{51} \times \frac{4}{50} = 0.0029$

1.9.23 (a) $\frac{\binom{48}{4}}{\binom{52}{4}} = \frac{48}{52} \times \frac{47}{51} \times \frac{46}{50} \times \frac{45}{49} = 0.719$

(b) $\frac{\binom{4}{1}\times\binom{48}{3}}{\binom{52}{4}} = \frac{4\times 4\times 48\times 47\times 46}{52\times 51\times 50\times 49} = 0.256$

(c) $\left(\frac{1}{52}\right)^3 = \frac{1}{140608}$

1.9.24 (a) True

(b) False

(c) False

(d) True

(e) True

(f) False

(g) False

1.9.25 Let W be the event that *'the team wins the game'*
and let S be the event that *'the team has a player sent off'*.

$P(W) = 0.55$
$P(S') = 0.85$
$P(W \mid S') = 0.60$

Since
$P(W) = P(W \cap S) + P(W \cap S')$
$= P(W \cap S) + (P(W \mid S') \times P(S'))$
it follows that
$0.55 = P(W \cap S) + (0.60 \times 0.85)$.

Therefore,
$P(W \cap S) = 0.04$.

1.9.26 (a) Let N be the event that the machine is *'new'*
and let G be the event that the machine has *'good quality'*.

$P(N \cap G') = \frac{120}{500}$
$P(N') = \frac{230}{500}$

Therefore,
$P(N \cap G) = P(N) - P(N \cap G')$
$= 1 - \frac{230}{500} - \frac{120}{500} = \frac{150}{500} = 0.3$.

(b) $P(G \mid N) = \frac{P(N \cap G)}{P(N)}$
$= \frac{0.3}{1 - \frac{230}{500}} = \frac{5}{9}$

1.9.27 (a) Let M be the event *'male'*,
let E be the event *'mechanical engineer'*,
and let S be the event *'senior'*.

$P(M) = \frac{113}{250}$
$P(E) = \frac{167}{250}$
$P(M' \cap E') = \frac{52}{250}$
$P(M' \cap E \cap S) = \frac{19}{250}$

Therefore,
$P(M \mid E') = 1 - P(M' \mid E')$
$= 1 - \frac{P(M' \cap E')}{P(E')}$
$= 1 - \frac{52}{250-167} = 0.373.$

(b) $P(S \mid M' \cap E) = \frac{P(M' \cap E \cap S)}{P(M' \cap E)}$
$= \frac{P(M' \cap E \cap S)}{P(M') - P(M' \cap E')}$
$= \frac{19}{250-113-52} = 0.224$

1.9.28 (a) Let T be the event that *'the tax form is filed on time'*,
let S be the event that *'the tax form is from a small business'*,
and let A be the event that *'the tax form is accurate'*.

$P(T \cap S \cap A) = 0.11$
$P(T' \cap S \cap A) = 0.13$
$P(T \cap S) = 0.15$
$P(T' \cap S \cap A') = 0.21$

Therefore,
$P(T \mid S \cap A) = \frac{P(T \cap S \cap A)}{P(S \cap A)}$
$= \frac{P(T \cap S \cap A)}{P(T \cap S \cap A) + P(T' \cap S \cap A)}$
$= \frac{0.11}{0.11+0.13} = \frac{11}{24}.$

(b) $P(S') = 1 - P(S)$
$= 1 - P(T \cap S) - P(T' \cap S)$
$= 1 - P(T \cap S) - P(T' \cap S \cap A) - P(T' \cap S \cap A')$
$= 1 - 0.15 - 0.13 - 0.21 = 0.51$

1.9.29 (a) $P(\text{having exactly two heart cards}) = \frac{C_2^{13} \times C_2^{39}}{C_4^{52}} = 0.213$

(b) $P(\text{having exactly two heart cards and exactly two club cards})$
$= \frac{C_2^{13} \times C_2^{13}}{C_4^{52}} = 0.022$

(c) $P(\text{having 3 heart cards} \mid \text{no club cards})$
$= P(\text{having 3 heart cards from a reduced pack of 39 cards})$
$= \frac{C_3^{13} \times C_1^{26}}{C_4^{39}} = 0.09$

1.9.30 (a) $P(\text{passing the first time}) = 0.26$
$P(\text{passing the second time}) = 0.43$

$P(\text{failing the first time and passing the second time})$
$= P(\text{failing the first time}) \times P(\text{passing the second time})$
$= (1 - 0.26) \times 0.43 = 0.3182$

(b) $1 - P(\text{failing both times}) = 1 - (1 - 0.26) \times (1 - 0.43) = 0.5782$

(c) $P(\text{passing the first time} \mid \text{moving to the next stage})$
$= \frac{P(\text{passing the first time and moving to the next stage})}{P(\text{moving to the next stage})}$
$= \frac{0.26}{0.5782} = 0.45$

1.9.31 The possible outcomes are $(6, 5, 4, 3, 2)$, $(6, 5, 4, 3, 1)$, $(6, 5, 4, 2, 1)$, $(6, 5, 3, 2, 1)$, $(6, 4, 3, 2, 1)$, and $(5, 4, 3, 2, 1)$.
Each outcome has a probability of $\frac{1}{6^5}$ so that the required probability is
$\frac{6}{6^5} = \frac{1}{6^4} = \frac{1}{1296}$.

1.9.32 $P(\text{at least one uncorrupted file}) = 1 - P(\text{both files corrupted})$
$= 1 - (0.005 \times 0.01) = 0.99995$

1.9.33 Let C be the event that *'the pump is operating correctly'*
and let L be the event that *'the light is on'*.

$P(L \mid C') = 0.992$
$P(L \mid C) = 0.003$
$P(C) = 0.996$

Therefore, using Bayes theorem
$P(C' \mid L) = \frac{P(L \mid C')P(C')}{P(L \mid C')P(C') + P(L \mid C)P(C)}$
$= \frac{0.992 \times 0.004}{(0.992 \times 0.004) + (0.003 \times 0.996)} = 0.57.$

1.9.34 $$\frac{\binom{4}{2} \times \binom{4}{2} \times \binom{4}{3} \times \binom{4}{3}}{\binom{52}{10}} = \frac{1}{27{,}465{,}320}$$

1.9.35 (a) $$\frac{\binom{7}{3}}{\binom{11}{3}} = \frac{7}{11} \times \frac{6}{10} \times \frac{5}{9} = \frac{7}{33}$$

(b) $$\frac{\binom{7}{1} \times \binom{4}{2}}{\binom{11}{3}} = \frac{14}{55}$$

1.9.36 (a) The probability of an infected person having strain A is $P(A) = 0.32$.
The probability of an infected person having strain B is $P(B) = 0.59$.
The probability of an infected person having strain C is $P(C) = 0.09$.

$P(S \mid A) = 0.21$
$P(S \mid B) = 0.16$
$P(S \mid C) = 0.63$

Therefore, the probability of an infected person exhibiting symptoms is
$P(S) = (P(S \mid A) \times P(A)) + (P(S \mid B) \times P(B)) + (P(S \mid C) \times P(C))$
$= 0.2183$
and
$P(C \mid S) = \frac{P(S \mid C) \times P(C)}{P(S)}$
$= \frac{0.63 \times 0.09}{0.2183} = 0.26.$

(b) $P(S') = 1 - P(S) = 1 - 0.2183 = 0.7817$
$P(S' \mid A) = 1 - P(S \mid A) = 1 - 0.21 = 0.79$

Therefore,
$P(A \mid S') = \frac{P(S' \mid A) \times P(A)}{P(S')}$
$= \frac{0.79 \times 0.32}{0.7817} = 0.323.$

(c) $P(S') = 1 - P(S) = 1 - 0.2183 = 0.7817$

Chapter 2

Random Variables

2.1 Discrete Random Variables

2.1.1 (a) Since
$0.08 + 0.11 + 0.27 + 0.33 + P(X = 4) = 1$
it follows that
$P(X = 4) = 0.21$.

(c) $F(0) = 0.08$
$F(1) = 0.19$
$F(2) = 0.46$
$F(3) = 0.79$
$F(4) = 1.00$

2.1.2

x_i	-4	-1	0	2	3	7
p_i	0.21	0.11	0.07	0.29	0.13	0.19

2.1.3

x_i	1	2	3	4	5	6	8	9	10
p_i	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$
$F(x_i)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{8}{36}$	$\frac{10}{36}$	$\frac{14}{36}$	$\frac{16}{36}$	$\frac{17}{36}$	$\frac{19}{36}$

x_i	12	15	16	18	20	24	25	30	36
p_i	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$F(x_i)$	$\frac{23}{36}$	$\frac{25}{36}$	$\frac{26}{36}$	$\frac{28}{36}$	$\frac{30}{36}$	$\frac{32}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	1

2.1.4 (a)

x_i	0	1	2
p_i	0.5625	0.3750	0.0625

(b)

x_i	0	1	2
$F(x_i)$	0.5625	0.9375	1.000

(c) The value $x = 0$ is the most likely.

Without replacement:

x_i	0	1	2
p_i	0.5588	0.3824	0.0588
$F(x_i)$	0.5588	0.9412	1.000

Again, $x = 0$ is the most likely value.

2.1.5

x_i	-5	-4	-3	-2	-1	0	1	2	3	4	6	8	10	12
p_i	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{5}{36}$	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{3}{36}$	$\frac{3}{36}$	$\frac{3}{36}$
$F(x_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{6}{36}$	$\frac{9}{36}$	$\frac{12}{36}$	$\frac{14}{36}$	$\frac{19}{36}$	$\frac{20}{36}$	$\frac{24}{36}$	$\frac{27}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	1

2.1.6 (a)

x_i	-6	-4	-2	0	2	4	6
p_i	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

(b)

x_i	-6	-4	-2	0	2	4	6
$F(x_i)$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$	$\frac{5}{8}$	$\frac{6}{8}$	$\frac{7}{8}$	1

(c) The most likely value is $x = 0$.

2.1.7 (a)

x_i	0	1	2	3	4	6	8	12
p_i	0.061	0.013	0.195	0.067	0.298	0.124	0.102	0.140

(b)

x_i	0	1	2	3	4	6	8	12
$F(x_i)$	0.061	0.074	0.269	0.336	0.634	0.758	0.860	1.000

(c) The most likely value is 4.

P(not shipped) = P(X $\leq$ 1) = 0.074

2.1.8

x_i	-1	0	1	3	4	5
p_i	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
$F(x_i)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	1

2.1.9

x_i	1	2	3	4
p_i	$\frac{2}{5}$	$\frac{3}{10}$	$\frac{1}{5}$	$\frac{1}{10}$
$F(x_i)$	$\frac{2}{5}$	$\frac{7}{10}$	$\frac{9}{10}$	1

2.1.10 Since
$\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$
it follows that
$P(X = i) = \frac{6}{\pi^2 i^2}$
is a possible set of probability values.

However, since
$\sum_{i=1}^{\infty} \frac{1}{i}$
does not converge, it follows that
$P(X = i) = \frac{c}{i}$
is not a possible set of probability values.

2.1.11 (a) The state space is $\{3, 4, 5, 6\}$.

(b) $P(X = 3) = P(MMM) = \frac{3}{6} \times \frac{2}{5} \times \frac{1}{4} = \frac{1}{20}$

$P(X = 4) = P(MMTM) + P(MTMM) + P(TMMM) = \frac{3}{20}$

$P(X = 5) = P(MMTTM) + P(MTMTM) + P(TMMTM)$

$+ \ P(MTTMM) + P(TMTMM) + P(TTMMM) = \frac{6}{20}$

Finally,
$P(X = 6) = \frac{1}{2}$
since the probabilities sum to one, or since the final appointment made is equally likely to be on a Monday or on a Tuesday.

$P(X \leq 3) = \frac{1}{20}$
$P(X \leq 4) = \frac{4}{20}$
$P(X \leq 5) = \frac{10}{20}$
$P(X \leq 6) = 1$

2.2 Continuous Random Variables

2.2.1 (a) Continuous

(b) Discrete

(c) Continuous

(d) Continuous

(e) Discrete

(f) This depends on what level of accuracy to which it is measured.
It could be considered to be either discrete or continuous.

2.2.2 (b) $\int_4^6 \frac{1}{x\ln(1.5)}\,dx = \frac{1}{\ln(1.5)} \times [\ln(x)]_4^6$

$= \frac{1}{\ln(1.5)} \times (\ln(6) - \ln(4)) = 1.0$

(c) $P(4.5 \le X \le 5.5) = \int_{4.5}^{5.5} \frac{1}{x\ln(1.5)}\,dx$

$= \frac{1}{\ln(1.5)} \times [\ln(x)]_{4.5}^{5.5}$

$= \frac{1}{\ln(1.5)} \times (\ln(5.5) - \ln(4.5)) = 0.495$

(d) $F(x) = \int_4^x \frac{1}{y\ln(1.5)}\,dy$

$= \frac{1}{\ln(1.5)} \times [\ln(y)]_4^x$

$= \frac{1}{\ln(1.5)} \times (\ln(x) - \ln(4))$

for $4 \le x \le 6$

2.2.3 (a) Since

$\int_{-2}^0 \left(\frac{15}{64} + \frac{x}{64}\right) dx = \frac{7}{16}$

and

$\int_0^3 \left(\frac{3}{8} + cx\right) dx = \frac{9}{8} + \frac{9c}{2}$

it follows that

$\frac{7}{16} + \frac{9}{8} + \frac{9c}{2} = 1$

which gives $c = -\frac{1}{8}$.

(b) $P(-1 \le X \le 1) = \int_{-1}^0 \left(\frac{15}{64} + \frac{x}{64}\right) dx + \int_0^1 \left(\frac{3}{8} - \frac{x}{8}\right) dx$

$= \frac{69}{128}$

(c) $F(x) = \int_{-2}^{x} \left(\frac{15}{64} + \frac{y}{64}\right) dy$

$= \frac{x^2}{128} + \frac{15x}{64} + \frac{7}{16}$

for $-2 \leq x \leq 0$

$F(x) = \frac{7}{16} + \int_{0}^{x} \left(\frac{3}{8} - \frac{y}{8}\right) dy$

$= -\frac{x^2}{16} + \frac{3x}{8} + \frac{7}{16}$

for $0 \leq x \leq 3$

2.2.4 (b) $P(X \leq 2) = F(2) = \frac{1}{4}$

(c) $P(1 \leq X \leq 3) = F(3) - F(1)$

$= \frac{9}{16} - \frac{1}{16} = \frac{1}{2}$

(d) $f(x) = \frac{dF(x)}{dx} = \frac{x}{8}$

for $0 \leq x \leq 4$

2.2.5 (a) Since $F(\infty) = 1$ it follows that $A = 1$.
Then $F(0) = 0$ gives $1 + B = 0$ so that $B = -1$ and
$F(x) = 1 - e^{-x}$.

(b) $P(2 \leq X \leq 3) = F(3) - F(2)$
$= e^{-2} - e^{-3} = 0.0855$

(c) $f(x) = \frac{dF(x)}{dx} = e^{-x}$

for $x \geq 0$

2.2.6 (a) Since

$\int_{0.125}^{0.5} A\,(0.5 - (x - 0.25)^2)\; dx = 1$

it follows that $A = 5.5054$.

(b) $F(x) = \int_{0.125}^{x} f(y)\; dy$

$= 5.5054 \left(\frac{x}{2} - \frac{(x-0.25)^3}{3} - 0.06315\right)$

for $0.125 \leq x \leq 0.5$

(c) $F(0.2) = 0.203$

2.2.7 (a) Since

$F(0) = A + B\ln(2) = 0$

and

$F(10) = A + B\ln(32) = 1$

it follows that $A = -0.25$ and $B = \frac{1}{\ln(16)} = 0.361$.

(b) $P(X > 2) = 1 - F(2) = 0.5$

(c) $f(x) = \frac{dF(x)}{dx} = \frac{1.08}{3x+2}$

for $0 \leq x \leq 10$

2.2.8 (a) Since

$\int_0^{10} A\,(e^{10-\theta} - 1)\; d\theta = 1$

it follows that

$A = (e^{10} - 11)^{-1} = 4.54 \times 10^{-5}$.

(b) $F(\theta) = \int_0^{\theta} f(y)\; dy$

$= \frac{e^{10} - \theta - e^{10-\theta}}{e^{10} - 11}$

for $0 \leq \theta \leq 10$

(c) $1 - F(8) = 0.0002$

2.2.9 (a) Since $F(0) = 0$ and $F(50) = 1$
it follows that $A = 1.0007$ and $B = -125.09$.

(b) $P(X \leq 10) = F(10) = 0.964$

(c) $P(X \geq 30) = 1 - F(30) = 1 - 0.998 = 0.002$

(d) $f(r) = \frac{dF(r)}{dr} = \frac{375.3}{(r+5)^4}$

for $0 \leq r \leq 50$

2.2.10 (a) $F(200) = 0.1$

(b) $F(700) - F(400) = 0.65$

2.2.11 (a) Since

$\int_{10}^{11} Ax(130 - x^2)\ dx = 1$

it follows that

$A = \frac{4}{819}$.

(b) $F(x) = \int_{10}^{x} \frac{4y(130-y^2)}{819}\ dy$

$= \frac{4}{819}\left(65x^2 - \frac{x^4}{4} - 4000\right)$

for $10 \leq x \leq 11$

(c) $F(10.5) - F(10.25) = 0.623 - 0.340 = 0.283$

2.3 The Expectation of a Random Variable

2.3.1 $E(X) = (0 \times 0.08) + (1 \times 0.11) + (2 \times 0.27) + (3 \times 0.33) + (4 \times 0.21)$
$= 2.48$

2.3.2 $E(X) = \left(1 \times \frac{1}{36}\right) + \left(2 \times \frac{2}{36}\right) + \left(3 \times \frac{2}{36}\right) + \left(4 \times \frac{3}{36}\right) + \left(5 \times \frac{2}{36}\right) + \left(6 \times \frac{4}{36}\right)$
$+ \left(8 \times \frac{2}{36}\right) + \left(9 \times \frac{1}{36}\right) + \left(10 \times \frac{2}{36}\right) + \left(12 \times \frac{4}{36}\right) + \left(15 \times \frac{2}{36}\right) + \left(16 \times \frac{1}{36}\right)$
$+ \left(18 \times \frac{2}{36}\right) + \left(20 \times \frac{2}{36}\right) + \left(24 \times \frac{2}{36}\right) + \left(25 \times \frac{1}{36}\right) + \left(30 \times \frac{2}{36}\right) + \left(36 \times \frac{1}{36}\right)$
$= 12.25$

2.3.3 With replacement:
$E(X) = (0 \times 0.5625) + (1 \times 0.3750) + (2 \times 0.0625)$
$= 0.5$

Without replacement:
$E(X) = (0 \times 0.5588) + (1 \times 0.3824) + (2 \times 0.0588)$
$= 0.5$

2.3.4 $E(X) = \left(1 \times \frac{2}{5}\right) + \left(2 \times \frac{3}{10}\right) + \left(3 \times \frac{1}{5}\right) + \left(4 \times \frac{1}{10}\right)$
$= 2$

2.3.5

x_i	2	3	4	5	6	7	8	9	10	15
p_i	$\frac{1}{13}$	$\frac{1}{13}$	$\frac{1}{13}$	$\frac{1}{13}$	$\frac{1}{13}$	$\frac{1}{13}$	$\frac{1}{13}$	$\frac{1}{13}$	$\frac{1}{13}$	$\frac{4}{13}$

$E(X) = \left(2 \times \frac{1}{13}\right) + \left(3 \times \frac{1}{13}\right) + \left(4 \times \frac{1}{13}\right) + \left(5 \times \frac{1}{13}\right) + \left(6 \times \frac{1}{13}\right)$
$+ \left(7 \times \frac{1}{13}\right) + \left(8 \times \frac{1}{13}\right) + \left(9 \times \frac{1}{13}\right) + \left(10 \times \frac{1}{13}\right) + \left(15 \times \frac{4}{13}\right)$
$= \$8.77$

If \$9 is paid to play the game, the expected loss would be 23 cents.

2.3.6

x_i	1	2	3	4	5	6	7	8	9	10	11	12
p_i	$\frac{6}{72}$	$\frac{7}{72}$	$\frac{8}{72}$	$\frac{9}{72}$	$\frac{10}{72}$	$\frac{11}{72}$	$\frac{6}{72}$	$\frac{5}{72}$	$\frac{4}{72}$	$\frac{3}{72}$	$\frac{2}{72}$	$\frac{1}{72}$

$E(X) = \left(1 \times \frac{6}{72}\right) + \left(2 \times \frac{6}{72}\right) + \left(3 \times \frac{6}{72}\right) + \left(4 \times \frac{6}{72}\right) + \left(5 \times \frac{6}{72}\right) + \left(6 \times \frac{6}{72}\right)$
$+ \left(7 \times \frac{6}{72}\right) + \left(8 \times \frac{6}{72}\right) + \left(9 \times \frac{6}{72}\right) + \left(10 \times \frac{6}{72}\right) + \left(11 \times \frac{6}{72}\right) + \left(12 \times \frac{6}{72}\right)$
$= 5.25$

2.3.7 $P(\text{three sixes are rolled}) = \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6}$
$= \frac{1}{216}$
so that
$E(\text{net winnings}) = \left(-\$1 \times \frac{215}{216}\right) + \left(\$499 \times \frac{1}{216}\right)$
$= \$1.31.$

If you can play the game a large number of times then you should play the game as often as you can.

2.3.8 The expected net winnings will be negative.

2.3.9

x_i	0	1	2	3	4	5
p_i	0.1680	0.2816	0.2304	0.1664	0.1024	0.0512

$E(\text{payment}) = (0 \times 0.1680) + (1 \times 0.2816) + (2 \times 0.2304)$
$+ (3 \times 0.1664) + (4 \times 0.1024) + (5 \times 0.0512)$
$= 1.9072$

$E(\text{winnings}) = \$2 - \$1.91 = \$0.09$

The expected winnings increase to 9 cents per game.
Increasing the probability of scoring a three reduces the expected value of the difference in the scores of the two dice.

2.3.10 (a) $E(X) = \int_4^6 x \frac{1}{x\ln(1.5)} \, dx = 4.94$

(b) Solving $F(x) = 0.5$ gives $x = 4.90$.

2.3.11 (a) $E(X) = \int_0^4 x \frac{x}{8} \, dx = 2.67$

(b) Solving $F(x) = 0.5$ gives $x = \sqrt{8} = 2.83$.

2.3.12 $E(X) = \int_{0.125}^{0.5} x \; 5.5054 \; (0.5 - (x - 0.25)^2) \; dx = 0.3095$

Solving $F(x) = 0.5$ gives $x = 0.3081$.

2.3.13 $E(X) = \int_0^{10} \frac{\theta}{e^{10}-11} \; (e^{10-\theta} - 1) \; d\theta = 0.9977$

Solving $F(\theta) = 0.5$ gives $\theta = 0.6927$.

2.3.14 $E(X) = \int_0^{50} \frac{375.3\, r}{(r+5)^4} \; dr = 2.44$

Solving $F(r) = 0.5$ gives $r = 1.30$.

2.3.15 Let $f(x)$ be a probability density function that is symmetric about the point μ, so that $f(\mu + x) = f(\mu - x)$.
Then

$E(X) = \int_{-\infty}^{\infty} \; xf(x) \; dx$

which under the transformation $x = \mu + y$ gives

$E(X) = \int_{-\infty}^{\infty} \; (\mu + y)f(\mu + y) \; dy$

$= \mu \; \int_{-\infty}^{\infty} \; f(\mu + y) \; dy \; + \; \int_0^{\infty} \; y \; (f(\mu + y) - f(\mu - y)) \; dy$

$= (\mu \times 1) + 0 = \mu$.

2.3.16 $E(X) = (3 \times \frac{1}{20}) + (4 \times \frac{3}{20}) + (5 \times \frac{6}{20}) + (6 \times \frac{10}{20})$

$= \frac{105}{20} = 5.25$

2.3.17 (a) $E(X) = \int_{10}^{11} \frac{4x^2(130-x^2)}{819} \; dx$

$= 10.418234$

(b) Solving $F(x) = 0.5$ gives the median as 10.385.

2.3.18 (a) Since

$\int_2^3 A(x - 1.5)dx = 1$

it follows that

$A\left[x^2 - 1.5x\right]_2^3 = 1$

so that $A = 1$.

(b) Let the median be m.
Then

$\int_2^m (x - 1.5)dx = 0.5$

so that

$\left[x^2 - 1.5x\right]_2^m = 0.5$

which gives

$0.5m^2 - 1.5m + 1 = 0.5$.

Therefore,

$m^2 - 3m + 1 = 0$

so that

$m = \frac{3 \pm \sqrt{5}}{2}$.

Since $2 \leq m \leq 3$ it follows that $m = \frac{3+\sqrt{5}}{2} = 2.618$.

2.4 The Variance of a Random Variable

2.4.1 (a) $E(X) = \left(-2 \times \frac{1}{3}\right) + \left(1 \times \frac{1}{6}\right) + \left(4 \times \frac{1}{3}\right) + \left(6 \times \frac{1}{6}\right)$

$= \frac{11}{6}$

(b) $\mathrm{Var}(X) = \left(\frac{1}{3} \times \left(-2 - \frac{11}{6}\right)^2\right) + \left(\frac{1}{6} \times \left(1 - \frac{11}{6}\right)^2\right)$

$+ \left(\frac{1}{3} \times \left(4 - \frac{11}{6}\right)^2\right) + \left(\frac{1}{6} \times \left(6 - \frac{11}{6}\right)^2\right)$

$= \frac{341}{36}$

(c) $E(X^2) = \left(\frac{1}{3} \times (-2)^2\right) + \left(\frac{1}{6} \times 1^2\right) + \left(\frac{1}{3} \times 4^2\right) + \left(\frac{1}{6} \times 6^2\right)$

$= \frac{77}{6}$

$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \frac{77}{6} - \left(\frac{11}{6}\right)^2 = \frac{341}{36}$

2.4.2 $E(X^2) = (0^2 \times 0.08) + (1^2 \times 0.11) + (2^2 \times 0.27)$
$+ (3^2 \times 0.33) + (4^2 \times 0.21) = 7.52$

Then $E(X) = 2.48$ so that
$\mathrm{Var}(X) = 7.52 - (2.48)^2 = 1.37$
and $\sigma = 1.17$.

2.4.3 $E(X^2) = \left(1^2 \times \frac{2}{5}\right) + \left(2^2 \times \frac{3}{10}\right) + \left(3^2 \times \frac{1}{5}\right) + \left(4^2 \times \frac{1}{10}\right)$
$= 5$

Then $E(X) = 2$ so that
$\mathrm{Var}(X) = 5 - 2^2 = 1$
and $\sigma = 1$.

2.4.4 See Problem 2.3.9.
$E(X^2) = (0^2 \times 0.168) + (1^2 \times 0.2816) + (3^2 \times 0.1664)$
$+ (4^2 \times 0.1024) + (5^2 \times 0.0512)$
$= 5.6192$

Then $E(X) = 1.9072$ so that
$\mathrm{Var}(X) = 5.6192 - 1.9072^2 = 1.98$

and $\sigma = 1.41$.

A small variance is generally preferable if the expected winnings are positive.

2.4.5 (a) $E(X^2) = \int_4^6 x^2 \frac{1}{x\ln(1.5)}\, dx = 24.66$

Then $E(X) = 4.94$ so that
$\mathrm{Var}(X) = 24.66 - 4.94^2 = 0.25$.

(b) $\sigma = \sqrt{0.25} = 0.5$

(c) Solving $F(x) = 0.25$ gives $x = 4.43$.
Solving $F(x) = 0.75$ gives $x = 5.42$.

(d) The interquartile range is $5.42 - 4.43 = 0.99$.

2.4.6 (a) $E(X^2) = \int_0^4 x^2 \left(\frac{x}{8}\right) dx = 8$

Then $E(X) = \frac{8}{3}$ so that
$\mathrm{Var}(X) = 8 - \left(\frac{8}{3}\right)^2 = \frac{8}{9}$.

(b) $\sigma = \sqrt{\frac{8}{9}} = 0.94$

(c) Solving $F(x) = 0.25$ gives $x = 2$.
Solving $F(x) = 0.75$ gives $x = \sqrt{12} = 3.46$.

(d) The interquartile range is $3.46 - 2.00 = 1.46$.

2.4.7 (a) $E(X^2) = \int_{0.125}^{0.5} x^2\, 5.5054\, (0.5 - (x - 0.25)^2)\, dx = 0.1073$

Then $E(X) = 0.3095$ so that
$\mathrm{Var}(X) = 0.1073 - 0.3095^2 = 0.0115$.

(b) $\sigma = \sqrt{0.0115} = 0.107$

(c) Solving $F(x) = 0.25$ gives $x = 0.217$.
Solving $F(x) = 0.75$ gives $x = 0.401$.

(d) The interquartile range is $0.401 - 0.217 = 0.184$.

2.4.8 (a) $E(X^2) = \int_0^{10} \frac{\theta^2}{e^{10}-11}(e^{10-\theta} - 1)\ d\theta$

$= 1.9803$

Then $E(X) = 0.9977$ so that
$\text{Var}(X) = 1.9803 - 0.9977^2 = 0.985$.

(b) $\sigma = \sqrt{0.985} = 0.992$

(c) Solving $F(\theta) = 0.25$ gives $\theta = 0.288$.
Solving $F(\theta) = 0.75$ gives $\theta = 1.385$.

(d) The interquartile range is $1.385 - 0.288 = 1.097$.

2.4.9 (a) $E(X^2) = \int_0^{50} \frac{375.3\ r^2}{(r+5)^4}\ dr = 18.80$

Then $E(X) = 2.44$ so that
$\text{Var}(X) = 18.80 - 2.44^2 = 12.8$.

(b) $\sigma = \sqrt{12.8} = 3.58$

(c) Solving $F(r) = 0.25$ gives $r = 0.50$.
Solving $F(r) = 0.75$ gives $r = 2.93$.

(d) The interquartile range is $2.93 - 0.50 = 2.43$.

2.4.10 Adding and subtracting two standard deviations from the mean value gives:

$P(60.4 \leq X \leq 89.6) \geq 0.75$

Adding and subtracting three standard deviations from the mean value gives:

$P(53.1 \leq X \leq 96.9) \geq 0.89$

2.4.11 The interval $(109.55, 112.05)$ is $(\mu - 2.5c, \mu + 2.5c)$
so Chebyshev's inequality gives:
$P(109.55 \leq X \leq 112.05) \geq 1 - \frac{1}{2.5^2} = 0.84$

2.4.12 $E(X^2) = \left(3^2 \times \frac{1}{20}\right) + \left(4^2 \times \frac{3}{20}\right) + \left(5^2 \times \frac{6}{20}\right) + \left(6^2 \times \frac{10}{20}\right)$

$= \frac{567}{20}$

$\text{Var}(X) = E(X^2) - (E(X))^2$

$= \frac{567}{20} - \left(\frac{105}{20}\right)^2 = \frac{63}{80}$

The standard deviation is $\sqrt{63/80} = 0.887$.

2.4.13 (a) $E(X^2) = \int_{10}^{11} \frac{4x^3(130-x^2)}{819}\, dx$

$= 108.61538$

Therefore,
$\text{Var}(X) = E(X^2) - (E(X))^2 = 108.61538 - 10.418234^2 = 0.0758$
and the standard deviation is $\sqrt{0.0758} = 0.275$.

(b) Solving $F(x) = 0.8$ gives the 80th percentile of the resistance as 10.69, and solving $F(x) = 0.1$ gives the 10th percentile of the resistance as 10.07.

2.4.14 (a) Since

$1 = \int_2^3 Ax^{2.5}\, dx = \frac{A}{3.5} \times (3^{3.5} - 2^{3.5})$

it follows that $A = 0.0987$.

(b) $E(X) = \int_2^3 0.0987\, x^{3.5}\, dx$

$= \frac{0.0987}{4.5} \times (3^{4.5} - 2^{4.5}) = 2.58$

(c) $E(X^2) = \int_2^3 0.0987\, x^{4.5}\, dx$

$= \frac{0.0987}{5.5} \times (3^{5.5} - 2^{5.5}) = 6.741$

Therefore,
$\text{Var}(X) = 6.741 - 2.58^2 = 0.085$
and the standard deviation is $\sqrt{0.085} = 0.29$.

(d) Solving

$0.5 = \int_2^x 0.0987\, y^{2.5}\, dy$

$= \frac{0.0987}{3.5} \times (x^{3.5} - 2^{3.5})$

gives $x = 2.62$.

2.4.15 $E(X) = (-1 \times 0.25) + (1 \times 0.4) + (4 \times 0.35)$

$= \$1.55$

$E(X^2) = ((-1)^2 \times 0.25) + (1^2 \times 0.4) + (4^2 \times 0.35)$

$= 6.25$

Therefore, the variance is
$E(X^2) - (E(X))^2 = 6.25 - 1.55^2 = 3.8475$
and the standard deviation is $\sqrt{3.8475} = \$1.96$.

2.4.16 (a) Since

$1 = \int_3^4 \frac{A}{\sqrt{x}}\ dx = 2A(2 - \sqrt{3})$

it follows that

$A = 1.866$.

(b) $F(x) = \int_3^x \frac{1.866}{\sqrt{y}}\ dy$

$= 3.732 \times (\sqrt{x} - \sqrt{3})$

(c) $E(X) = \int_3^4 x\ \frac{1.866}{\sqrt{x}}\ dx$

$= \frac{2}{3} \times 1.866 \times (4^{1.5} - 3^{1.5}) = 3.488$

(d) $E(X^2) = \int_3^4 x^2\ \frac{1.866}{\sqrt{x}}\ dx$

$= \frac{2}{5} \times 1.866 \times (4^{2.5} - 3^{2.5}) = 12.250$

Therefore,
$\mathrm{Var}(X) = 12.250 - 3.488^2 = 0.0834$
and the standard deviation is $\sqrt{0.0834} = 0.289$.

(e) Solving
$F(x) = 3.732 \times (\sqrt{x} - \sqrt{3}) = 0.5$
gives $x = 3.48$.

(f) Solving
$F(x) = 3.732 \times (\sqrt{x} - \sqrt{3}) = 0.75$
gives $x = 3.74$.

2.4.17 (a) $E(X) = (2 \times 0.11) + (3 \times 0.19) + (4 \times 0.55) + (5 \times 0.15)$
$= 3.74$

(b) $E(X^2) = (2^2 \times 0.11) + (3^2 \times 0.19) + (4^2 \times 0.55) + (5^2 \times 0.15)$
$= 14.70$

Therefore,
$\mathrm{Var}(X) = 14.70 - 3.74^2 = 0.7124$
and the standard deviation is $\sqrt{0.7124} = 0.844$.

2.4.18 (a) $E(X) = \int_{-1}^{1} \frac{x(1-x)}{2}\ dx = -\frac{1}{3}$

(b) $E(X^2) = \int_{-1}^{1} \frac{x^2(1-x)}{2}\ dx = \frac{1}{3}$

Therefore,

$\text{Var}(X) = E(X^2) - (E(X))^2 = \frac{1}{3} - \frac{1}{9} = \frac{2}{9}$

and the standard deviation is $\frac{\sqrt{2}}{3} = 0.471$.

(c) Solving

$\int_{-1}^{y} \frac{(1-x)}{2}\ dx = 0.75$

gives $y = 0$.

2.5 Jointly Distributed Random Variables

2.5.1 (a) $P(0.8 \leq X \leq 1, 25 \leq Y \leq 30)$

$= \int_{x=0.8}^{1} \int_{y=25}^{30} \left(\frac{39}{400} - \frac{17(x-1)^2}{50} - \frac{(y-25)^2}{10000} \right) dx\ dy$

$= 0.092$

(b) $E(Y) = \int_{20}^{35} y \left(\frac{83}{1200} - \frac{(y-25)^2}{10000} \right) dy = 27.36$

$E(Y^2) = \int_{20}^{35} y^2 \left(\frac{83}{1200} - \frac{(y-25)^2}{10000} \right) dy = 766.84$

$\mathrm{Var}(Y) = E(Y^2) - E(Y)^2 = 766.84 - (27.36)^2 = 18.27$

$\sigma_Y = \sqrt{18.274} = 4.27$

(c) $E(Y|X = 0.55) = \int_{20}^{35} y \left(0.073 - \frac{(y-25)^2}{3922.5} \right) dy = 27.14$

$E(Y^2|X = 0.55) = \int_{20}^{35} y^2 \left(0.073 - \frac{(y-25)^2}{3922.5} \right) dy = 753.74$

$\mathrm{Var}(Y|X = 0.55) = E(Y^2|X = 0.55) - E(Y|X = 0.55)^2$

$= 753.74 - (27.14)^2 = 17.16$

$\sigma_{Y|X=0.55} = \sqrt{17.16} = 4.14$

2.5.2 (a) $p_{1|Y=1} = P(X = 1|Y = 1) = \frac{p_{11}}{p_{+1}} = \frac{0.12}{0.32} = 0.37500$

$p_{2|Y=1} = P(X = 2|Y = 1) = \frac{p_{21}}{p_{+1}} = \frac{0.08}{0.32} = 0.25000$

$p_{3|Y=1} = P(X = 3|Y = 1) = \frac{p_{31}}{p_{+1}} = \frac{0.07}{0.32} = 0.21875$

$p_{4|Y=1} = P(X = 4|Y = 1) = \frac{p_{41}}{p_{+1}} = \frac{0.05}{0.32} = 0.15625$

$E(X|Y = 1) = (1 \times 0.375) + (2 \times 0.25) + (3 \times 0.21875) + (4 \times 0.15625)$
$= 2.15625$

$E(X^2|Y = 1) = (1^2 \times 0.375) + (2^2 \times 0.25) + (3^2 \times 0.21875) + (4^2 \times 0.15625)$
$= 5.84375$

$\mathrm{Var}(X|Y = 1) = E(X^2|Y = 1) - E(X|Y = 1)^2$
$= 5.84375 - 2.15625^2 = 1.1943$

$\sigma_{X|Y=1} = \sqrt{1.1943} = 1.093$

(b) $p_{1|X=2} = P(Y = 1|X = 2) = \frac{p_{21}}{p_{2+}} = \frac{0.08}{0.24} = \frac{8}{24}$

$p_{2|X=2} = P(Y = 2|X = 2) = \frac{p_{22}}{p_{2+}} = \frac{0.15}{0.24} = \frac{15}{24}$

$p_{3|X=2} = P(Y = 3|X = 2) = \frac{p_{23}}{p_{2+}} = \frac{0.01}{0.24} = \frac{1}{24}$

$E(Y|X = 2) = \left(1 \times \frac{8}{24}\right) + \left(2 \times \frac{15}{24}\right) + \left(3 \times \frac{1}{24}\right)$
$= \frac{41}{24} = 1.7083$

$E(Y^2|X = 2) = \left(1^2 \times \frac{8}{24}\right) + \left(2^2 \times \frac{15}{24}\right) + \left(3^2 \times \frac{1}{24}\right)$
$= \frac{77}{24} = 3.2083$

$\text{Var}(Y|X = 2) = E(Y^2|X = 2) - E(Y|X = 2)^2$
$= 3.2083 - 1.7083^2 = 0.290$

$\sigma_{Y|X=2} = \sqrt{0.290} = 0.538$

2.5.3 (a) Since

$\int_{x=-2}^{3} \int_{y=4}^{6} A(x-3)y\ dx\ dy = 1$

it follows that $A = -\frac{1}{125}$.

(b) $P(0 \leq X \leq 1, 4 \leq Y \leq 5)$

$= \int_{x=0}^{1} \int_{y=4}^{5} \frac{(3-x)y}{125}\ dx\ dy$

$= \frac{9}{100}$

(c) $f_X(x) = \int_4^6 \frac{(3-x)y}{125}\ dy = \frac{2(3-x)}{25}$

for $-2 \leq x \leq 3$

$f_Y(y) = \int_{-2}^{3} \frac{(3-x)y}{125}\ dx = \frac{y}{10}$

for $4 \leq x \leq 6$

(d) The random variables X and Y are independent since
$f_X(x) \times f_Y(y) = f(x, y)$
and the ranges of the random variables are not related.

(e) Since the random variables are independent it follows that
$f_{X|Y=5}(x)$ is equal to $f_X(x)$.

2.5.4 (a)

X\Y	0	1	2	3	p_{i+}
0	$\frac{1}{16}$	$\frac{1}{16}$	0	0	$\frac{2}{16}$
1	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	0	$\frac{6}{16}$
2	0	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{6}{16}$
3	0	0	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{2}{16}$
p_{+j}	$\frac{2}{16}$	$\frac{6}{16}$	$\frac{6}{16}$	$\frac{2}{16}$	1

(b) See the table above.

(c) The random variables X and Y are not independent.
For example, notice that
$p_{0+} \times p_{+0} = \frac{2}{16} \times \frac{2}{16} = \frac{1}{4} \neq p_{00} = \frac{1}{16}$.

(d) $E(X) = \left(0 \times \frac{2}{16}\right) + \left(1 \times \frac{6}{16}\right) + \left(2 \times \frac{6}{16}\right) + \left(3 \times \frac{2}{16}\right) = \frac{3}{2}$

$E(X^2) = \left(0^2 \times \frac{2}{16}\right) + \left(1^2 \times \frac{6}{16}\right) + \left(2^2 \times \frac{6}{16}\right) + \left(3^2 \times \frac{2}{16}\right) = 3$

$\mathrm{Var}(X) = E(X^2) - E(X)^2 = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4}$

The random variable Y has the same mean and variance as X.

(e) $E(XY) = \left(1 \times 1 \times \frac{3}{16}\right) + \left(1 \times 2 \times \frac{2}{16}\right) + \left(2 \times 1 \times \frac{2}{16}\right)$

$+ \left(2 \times 2 \times \frac{3}{16}\right) + \left(2 \times 3 \times \frac{1}{16}\right) + \left(3 \times 2 \times \frac{1}{16}\right) + \left(3 \times 3 \times \frac{1}{16}\right)$

$= \frac{44}{16}$

$\mathrm{Cov}(X, Y) = E(XY) - (E(X) \times E(Y))$

$= \frac{44}{16} - \left(\frac{3}{2} \times \frac{3}{2}\right) = \frac{1}{2}$

(f) $P(X = 0|Y = 1) = \frac{p_{01}}{p_{+1}} = \frac{\left(\frac{1}{16}\right)}{\left(\frac{6}{16}\right)} = \frac{1}{6}$

$P(X = 1|Y = 1) = \frac{p_{11}}{p_{+1}} = \frac{\left(\frac{3}{16}\right)}{\left(\frac{6}{16}\right)} = \frac{1}{2}$

$P(X = 2|Y = 1) = \frac{p_{21}}{p_{+1}} = \frac{\left(\frac{2}{16}\right)}{\left(\frac{6}{16}\right)} = \frac{1}{3}$

$P(X = 3|Y = 1) = \frac{p_{31}}{p_{+1}} = \frac{0}{\left(\frac{6}{16}\right)} = 0$

$E(X|Y = 1) = \left(0 \times \frac{1}{6}\right) + \left(1 \times \frac{1}{2}\right) + \left(2 \times \frac{1}{3}\right) + (3 \times 0) = \frac{7}{6}$

$E(X^2|Y = 1) = \left(0^2 \times \frac{1}{6}\right) + \left(1^2 \times \frac{1}{2}\right) + \left(2^2 \times \frac{1}{3}\right) + (3^2 \times 0) = \frac{11}{6}$

$\mathrm{Var}(X|Y = 1) = E(X^2|Y = 1) - E(X|Y = 1)^2$

$= \frac{11}{6} - \left(\frac{7}{6}\right)^2 = \frac{17}{36}$

2.5.5 (a) Since

$\int_{x=1}^{2} \int_{y=0}^{3} A(e^{x+y} + e^{2x-y})\, dx\, dy = 1$

it follows that $A = 0.00896$.

(b) $P(1.5 \leq X \leq 2, 1 \leq Y \leq 2)$

$= \int_{x=1.5}^{2} \int_{y=1}^{2} 0.00896\, (e^{x+y} + e^{2x-y})\, dx\, dy$

$= 0.158$

(c) $f_X(x) = \int_0^3 0.00896\, (e^{x+y} + e^{2x-y})\, dy$

$= 0.00896\, (e^{x+3} - e^{2x-3} - e^x + e^{2x})$

for $1 \leq x \leq 2$

$f_Y(y) = \int_1^2 0.00896\, (e^{x+y} + e^{2x-y})\, dx$

$= 0.00896\, (e^{2+y} + 0.5e^{4-y} - e^{1+y} - 0.5e^{2-y})$

for $0 \leq y \leq 3$

(d) No, since $f_X(x) \times f_Y(y) \neq f(x, y)$.

(e) $f_{X|Y=0}(x) = \frac{f(x,0)}{f_Y(0)} = \frac{e^x + e^{2x}}{28.28}$

2.5.6 (a)

X\Y	0	1	2	p_{i+}
0	$\frac{25}{102}$	$\frac{26}{102}$	$\frac{6}{102}$	$\frac{57}{102}$
1	$\frac{26}{102}$	$\frac{13}{102}$	0	$\frac{39}{102}$
2	$\frac{6}{102}$	0	0	$\frac{6}{102}$
p_{+j}	$\frac{57}{102}$	$\frac{39}{102}$	$\frac{6}{102}$	1

(b) See the table above.

(c) No, the random variables X and Y are not independent.
For example,
$p_{22} \neq p_{2+} \times p_{+2}$.

(d) $E(X) = \left(0 \times \frac{57}{102}\right) + \left(1 \times \frac{39}{102}\right) + \left(2 \times \frac{6}{102}\right) = \frac{1}{2}$

$E(X^2) = \left(0^2 \times \frac{57}{102}\right) + \left(1^2 \times \frac{39}{102}\right) + \left(2^2 \times \frac{6}{102}\right) = \frac{21}{34}$

$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \frac{21}{34} - \left(\frac{1}{2}\right)^2 = \frac{25}{68}$

The random variable Y has the same mean and variance as X.

(e) $E(XY) = 1 \times 1 \times p_{11} = \frac{13}{102}$

$\mathrm{Cov}(X,Y) = E(XY) - (E(X) \times E(Y))$

$= \frac{13}{102} - \left(\frac{1}{2} \times \frac{1}{2}\right) = -\frac{25}{204}$

(f) $\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = -\frac{1}{3}$

(g) $P(Y = 0|X = 0) = \frac{p_{00}}{p_{0+}} = \frac{25}{57}$

$P(Y = 1|X = 0) = \frac{p_{01}}{p_{0+}} = \frac{26}{57}$

$P(Y = 2|X = 0) = \frac{p_{02}}{p_{0+}} = \frac{6}{57}$

$P(Y = 0|X = 1) = \frac{p_{10}}{p_{1+}} = \frac{2}{3}$

$P(Y = 1|X = 1) = \frac{p_{11}}{p_{1+}} = \frac{1}{3}$

$P(Y = 2|X = 1) = \frac{p_{12}}{p_{1+}} = 0$

2.5.7 (a)

X\Y	0	1	2	p_{i+}
0	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{1}{16}$	$\frac{9}{16}$
1	$\frac{4}{16}$	$\frac{2}{16}$	0	$\frac{6}{16}$
2	$\frac{1}{16}$	0	0	$\frac{1}{16}$
p_{+j}	$\frac{9}{16}$	$\frac{6}{16}$	$\frac{1}{16}$	1

(b) See the table above.

(c) No, the random variables X and Y are not independent.
For example,
$p_{22} \neq p_{2+} \times p_{+2}$.

(d) $E(X) = \left(0 \times \frac{9}{16}\right) + \left(1 \times \frac{6}{16}\right) + \left(2 \times \frac{1}{16}\right) = \frac{1}{2}$

$E(X^2) = \left(0^2 \times \frac{9}{16}\right) + \left(1^2 \times \frac{6}{16}\right) + \left(2^2 \times \frac{1}{16}\right) = \frac{5}{8}$

$\text{Var}(X) = E(X^2) - E(X)^2 = \frac{5}{8} - \left(\frac{1}{2}\right)^2 = \frac{3}{8} = 0.3676$

The random variable Y has the same mean and variance as X.

(e) $E(XY) = 1 \times 1 \times p_{11} = \frac{1}{8}$

$\text{Cov}(X, Y) = E(XY) - (E(X) \times E(Y))$

$= \frac{1}{8} - \left(\frac{1}{2} \times \frac{1}{2}\right) = -\frac{1}{8}$

(f) $\text{Corr}(X, Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = -\frac{1}{3}$

(g) $P(Y = 0|X = 0) = \frac{p_{00}}{p_{0+}} = \frac{4}{9}$

$P(Y = 1|X = 0) = \frac{p_{01}}{p_{0+}} = \frac{4}{9}$

$P(Y = 2|X = 0) = \frac{p_{02}}{p_{0+}} = \frac{1}{9}$

$P(Y = 0|X = 1) = \frac{p_{10}}{p_{1+}} = \frac{2}{3}$

$P(Y = 1|X = 1) = \frac{p_{11}}{p_{1+}} = \frac{1}{3}$

$P(Y = 2|X = 1) = \frac{p_{12}}{p_{1+}} = 0$

2.5.8 (a) Since

$\int_{x=0}^{5} \int_{y=0}^{5} A\ (20 - x - 2y)\ dx\ dy = 1$

it follows that $A = 0.0032$

(b) $P(1 \leq X \leq 2, 2 \leq Y \leq 3)$

$= \int_{x=1}^{2} \int_{y=2}^{3} 0.0032\ (20 - x - 2y)\ dx\ dy$

$= 0.0432$

(c) $f_X(x) = \int_{y=0}^{5} 0.0032\ (20 - x - 2y)\ dy = 0.016\ (15 - x)$

for $0 \leq x \leq 5$

$f_Y(y) = \int_{x=0}^{5} 0.0032\ (20 - x - 2y)\ dx = 0.008\ (35 - 4y)$

for $0 \leq y \leq 5$

(d) No, the random variables X and Y are not independent since $f(x, y) \neq f_X(x)f_Y(y)$.

(e) $E(X) = \int_{0}^{5} x\ 0.016\ (15 - x)\ dx = \frac{7}{3}$

$E(X^2) = \int_{0}^{5} x^2\ 0.016\ (15 - x)\ dx = \frac{15}{2}$

$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \frac{15}{2} - \left(\frac{7}{3}\right)^2 = \frac{37}{18}$

(f) $E(Y) = \int_{0}^{5} y\ 0.008\ (35 - 4y)\ dy = \frac{13}{6}$

$E(Y^2) = \int_{0}^{5} y^2\ 0.008\ (35 - 4y)\ dy = \frac{20}{3}$

$\mathrm{Var}(Y) = E(Y^2) - E(Y)^2 = \frac{20}{3} - \left(\frac{13}{6}\right)^2 = \frac{71}{36}$

(g) $f_{Y|X=3}(y) = \frac{f(3,y)}{f_X(3)} = \frac{17-2y}{60}$

for $0 \leq y \leq 5$

(h) $E(XY) = \int_{x=0}^{5} \int_{y=0}^{5} 0.0032\ xy\ (20 - x - 2y)\ dx\ dy = 5$

$\mathrm{Cov}(X, Y) = E(XY) - (E(X) \times (EY))$

$= 5 - \left(\frac{7}{3} \times \frac{13}{6}\right) = -\frac{1}{18}$

(i) $\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = -0.0276$

2.5.9 (a) $P(\text{same score}) = P(X = 1, Y = 1) + P(X = 2, Y = 2)$
$+ P(X = 3, Y = 3) + P(X = 4, Y = 4)$
$= 0.80$

(b) $P(X < Y) = P(X = 1, Y = 2) + P(X = 1, Y = 3) + P(X = 1, Y = 4)$
$+ P(X = 2, Y = 3) + P(X = 2, Y = 4) + P(X = 3, Y = 4)$
$= 0.07$

(c)

x_i	1	2	3	4
p_{i+}	0.12	0.20	0.30	0.38

$E(X) = (1 \times 0.12) + (2 \times 0.20) + (3 \times 0.30) + (4 \times 0.38) = 2.94$

$E(X^2) = (1^2 \times 0.12) + (2^2 \times 0.20) + (3^2 \times 0.30) + (4^2 \times 0.38) = 9.70$

$\text{Var}(X) = E(X^2) - E(X)^2 = 9.70 - (2.94)^2 = 1.0564$

(d)

y_j	1	2	3	4
p_{+j}	0.14	0.21	0.30	0.35

$E(Y) = (1 \times 0.14) + (2 \times 0.21) + (3 \times 0.30) + (4 \times 0.35) = 2.86$

$E(Y^2) = (1^2 \times 0.14) + (2^2 \times 0.21) + (3^2 \times 0.30) + (4^2 \times 0.35) = 9.28$

$\text{Var}(Y) = E(Y^2) - E(Y)^2 = 9.28 - (2.86)^2 = 1.1004$

(e) The scores are not independent.
For example, $p_{11} \neq p_{1+} \times p_{+1}$.
The scores would not be expected to be independent since they apply to the two inspectors' assessments of the same building. If they were independent it would suggest that one of the inspectors is randomly assigning a safety score without paying any attention to the actual state of the building.

(f) $P(Y = 1|X = 3) = \frac{p_{31}}{p_{3+}} = \frac{1}{30}$

$P(Y=2|X=3) = \frac{p_{32}}{p_{3+}} = \frac{3}{30}$

$P(Y=3|X=3) = \frac{p_{33}}{p_{3+}} = \frac{24}{30}$

$P(Y=4|X=3) = \frac{p_{34}}{p_{3+}} = \frac{2}{30}$

(g) $E(XY) = \sum_{i=1}^{4} \sum_{j=1}^{4} \; i\; j\; p_{ij} = 9.29$

$\text{Cov}(X,Y) = E(XY) - (E(X) \times E(Y))$
$= 9.29 - (2.94 \times 2.86) = 0.8816$

(h) $\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}X \; \text{Var}Y}} = \frac{0.8816}{\sqrt{1.0564 \times 1.1004}} = 0.82$

A high positive correlation indicates that the inspectors are consistent.
The closer the correlation is to one the more consistent the inspectors are.

2.5.10 (a) $\int_{x=0}^{2} \int_{y=0}^{2} \int_{z=0}^{2} \; \frac{3xyz^2}{32} \; dx \; dy \; dz = 1$

(b) $\int_{x=0}^{1} \int_{y=0.5}^{1.5} \int_{z=1}^{2} \; \frac{3xyz^2}{32} \; dx \; dy \; dz = \frac{7}{64}$

(c) $f_X(x) = \int_{x=0}^{2} \int_{y=0}^{2} \; \frac{3xyz^2}{32} \; dy \; dz = \frac{x}{2}$

for $0 \leq 2 \leq x$

2.6 Combinations and Functions of Random variables

2.6.1 (a) $E(3X + 7) = 3E(X) + 7 = 13$

$\text{Var}(3X + 7) = 3^2\text{Var}(X) = 36$

(b) $E(5X - 9) = 5E(X) - 9 = 1$

$\text{Var}(5X - 9) = 5^2\text{Var}(X) = 100$

(c) $E(2X + 6Y) = 2E(X) + 6E(Y) = -14$

$\text{Var}(2X + 6Y) = 2^2\text{Var}(X) + 6^2\text{Var}(Y) = 88$

(d) $E(4X - 3Y) = 4E(X) - 3E(Y) = 17$

$\text{Var}(4X - 3Y) = 4^2\text{Var}(X) + 3^2\text{Var}(Y) = 82$

(e) $E(5X - 9Z + 8) = 5E(X) - 9E(Z) + 8 = -54$

$\text{Var}(5X - 9Z + 8) = 5^2\text{Var}(X) + 9^2\text{Var}(Z) = 667$

(f) $E(-3Y - Z - 5) = -3E(Y) - E(Z) - 5 = -4$

$\text{Var}(-3Y - Z - 5) = (-3)^2\text{Var}(Y) + (-1)^2\text{Var}(Z) = 25$

(g) $E(X + 2Y + 3Z) = E(X) + 2E(Y) + 3E(Z) = 20$

$\text{Var}(X + 2Y + 3Z) = \text{Var}(X) + 2^2\text{Var}(Y) + 3^2\text{Var}(Z) = 75$

(h) $E(6X + 2Y - Z + 16) = 6E(X) + 2E(Y) - E(Z) + 16 = 14$

$\text{Var}(6X + 2Y - Z + 16) = 6^2\text{Var}(X) + 2^2\text{Var}(Y) + (-1)^2\text{Var}(Z) = 159$

2.6.2 $E(aX + b) = \int (ax + b)\ f(x)\ dx$

$= a \int x\ f(x)\ dx\ +\ b \int\ f(x)\ dx$

$= aE(X) + b$

$\text{Var}(aX + b) = E((aX + b - E(aX + b))^2)$

$= E((aX - aE(X))^2)$

$= a^2E((X - E(X))^2)$

$= a^2\text{Var}(X)$

2.6.3 $E(Y) = 3E(X_1) = 3\mu$

$\text{Var}(Y) = 3^2\text{Var}(X_1) = 9\sigma^2$

$E(Z) = E(X_1) + E(X_2) + E(X_3) = 3\mu$

$\text{Var}(Z) = \text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3) = 3\sigma^2$

The random variables Y and Z have the same mean but Z has a smaller variance than Y.

2.6.4 length $= A_1 + A_2 + B$

$E(\text{length}) = E(A_1) + E(A_2) + E(B) = 37 + 37 + 24 = 98$

$\text{Var}(\text{length}) = \text{Var}(A_1) + \text{Var}(A_2) + \text{Var}(B) = 0.7^2 + 0.7^2 + 0.3^2 = 1.07$

2.6.5 Let the random variable X_i be the winnings from the i^{th} game.
Then

$E(X_i) = \left(10 \times \frac{1}{8}\right) + \left((-1) \times \frac{7}{8}\right) = \frac{3}{8}$

and

$E(X_i^2) = \left(10^2 \times \frac{1}{8}\right) + \left((-1)^2 \times \frac{7}{8}\right) = \frac{107}{8}$

so that

$\text{Var}(X_i) = E(X_i^2) - (E(X_i))^2 = \frac{847}{64}$.

The total winnings from 50 (independent) games is

$Y = X_1 + \ldots + X_{50}$

and

$E(Y) = E(X_1) + \ldots + E(X_{50}) = 50 \times \frac{3}{8} = \frac{75}{4} = \$18.75$

with

$\text{Var}(Y) = \text{Var}(X_1) + \ldots + \text{Var}(X_{50}) = 50 \times \frac{847}{64} = 661.72$

so that $\sigma_Y = \sqrt{661.72} = \$25.72$.

2.6.6 (a) $E(\text{average weight}) = 1.12$ kg

$\text{Var}(\text{average weight}) = \frac{0.03^2}{25} = 3.6 \times 10^{-5}$

The standard deviation is $\frac{0.03}{\sqrt{25}} = 0.0012$ kg.

(b) It is required that $\frac{0.03}{\sqrt{n}} \leq 0.005$ which is satisfied for $n \geq 36$.

2.6.7 Let the random variable X_i be equal to 1 if an ace is drawn on the i^{th} drawing (which happens with a probability of $\frac{1}{13}$) and equal to 0 if an ace is not drawn on the i^{th} drawing (which happens with a probability of $\frac{12}{13}$).

Then the total number of aces drawn is $Y = X_1 + \ldots + X_{10}$.

Notice that $E(X_i) = \frac{1}{13}$ so that regardless of whether the drawing is performed with or without replacement it follows that

$E(Y) = E(X_1) + \ldots + E(X_{10}) = \frac{10}{13}$.

Also, notice that $E(X_i^2) = \frac{1}{13}$ so that

$\text{Var}(X_i) = \frac{1}{13} - \left(\frac{1}{13}\right)^2 = \frac{12}{169}$.

If the drawings are made *with replacement* then the random variables X_i are independent so that

$\text{Var}(Y) = \text{Var}(X_1) + \ldots + \text{Var}(X_{10}) = \frac{120}{169}$.

However, if the drawings are made *without replacement* then the random variables X_i are not independent.

2.6.8 $F_X(x) = P(X \leq x) = x^2$ for $0 \leq x \leq 1$

(a) $F_Y(y) = P(Y \leq y) = P(X^3 \leq y) = P(X \leq y^{1/3}) = F_X(y^{1/3}) = y^{2/3}$

and so

$f_Y(y) = \frac{2}{3}y^{-1/3}$

for $0 \leq y \leq 1$

$E(y) = \int_0^1 y\ f_Y(y)\ dy = 0.4$

(b) $F_Y(y) = P(Y \leq y) = P(\sqrt{X} \leq y) = P(X \leq y^2) = F_X(y^2) = y^4$

and so

$f_Y(y) = 4y^3$

for $0 \leq y \leq 1$

$E(y) = \int_0^1 y\ f_Y(y)\ dy = 0.8$

(c) $F_Y(y) = P(Y \leq y) = P(\frac{1}{1+X} \leq y) = P(X \geq \frac{1}{y} - 1)$

$= 1 - F_X(\frac{1}{y} - 1) = \frac{2}{y} - \frac{1}{y^2}$

and so

$f_Y(y) = -\frac{2}{y^2} + \frac{2}{y^3}$

for $\frac{1}{2} \leq y \leq 1$

$E(y) = \int_{0.5}^{1} \; y \; f_Y(y) \; dy = 0.614$

(d) $F_Y(y) = P(Y \leq y) = P(2^X \leq y) = P\left(X \leq \frac{\ln(y)}{\ln(2)}\right)$

$= F_X\left(\frac{\ln(y)}{\ln(2)}\right) = \left(\frac{\ln(y)}{\ln(2)}\right)^2$

and so

$f_Y(y) = \frac{2\;\ln(y)}{y\;(\ln(2))^2}$

for $1 \leq y \leq 2$

$E(y) = \int_{1}^{2} \; y \; f_Y(y) \; dy = 1.61$

2.6.9 (a) Since

$\int_0^2 \; A(1-(r-1)^2) \; dr = 1$

it follows that $A = \frac{3}{4}$.

This gives

$F_R(r) = \frac{3r^2}{4} - \frac{r^3}{4}$

for $0 \leq r \leq 2$.

(b) $V = \frac{4}{3}\pi r^3$

Since

$F_V(v) = P(V \leq v) = P\left(\frac{4}{3}\pi r^3 \;\; \leq \;\; v\right) = F_R\left(\left(\frac{3v}{4\pi}\right)^{1/3}\right)$

it follows that

$f_V(v) = \frac{1}{2}(\frac{3}{4\pi})^{2/3}v^{-1/3} \;\; - \;\; \frac{3}{16\pi}$

for $0 \leq v \leq \frac{32\pi}{3}$.

(c) $E(V) = \int_0^{\frac{32\pi}{3}} \; v \; f_V(v) \; dv = \frac{32\pi}{15}$

2.6.10 (a) Since

$\int_0^L Ax(L-x) \; dx = 1$

it follows that $A = \frac{6}{L^3}$.

Therefore,

$F_X(x) = \frac{x^2(3L-2x)}{L^3}$

for $0 \leq x \leq L$.

(b) The random variable corresponding to the difference between the lengths of the two pieces of rod is
$W = |L - 2X|$.

Therefore,

$F_W(w) = P\left(\frac{L}{2} - \frac{w}{2} \leq X \leq \frac{L}{2} + \frac{w}{2}\right) = F_X\left(\frac{L}{2} + \frac{w}{2}\right) - F_X\left(\frac{L}{2} - \frac{w}{2}\right)$

$= \frac{w(3L^2-w^2)}{2L^3}$

and

$f_W(w) = \frac{3(L^2-w^2)}{2L^3}$

for $0 \leq w \leq L$.

(c) $E(W) = \int_0^L \; w \; f_W(w) \; dw = \frac{3}{8}L$

2.6.11 (a) The return has an expectation of \$100, a standard deviation of \$20, and a variance of 400.

(b) The return has an expectation of \$100, a standard deviation of \$30, and a variance of 900.

(c) The return from fund A has an expectation of \$50, a standard deviation of \$10, and a variance of 100.

The return from fund B has an expectation of \$50, a standard deviation of \$15, and a variance of 225.

Therefore, the total return has an expectation of \$100 and a variance of 325, so that the standard deviation is \$18.03.

(d) The return from fund A has an expectation of \$0.1x,
a standard deviation of \$0.02x,
and a variance of $0.0004x^2$.

The return from fund B has an expectation of \$0.1$(1000 - x)$,
a standard deviation of \$0.03$(1000 - x)$,
and a variance of $0.0009(1000 - x)^2$.

Therefore, the total return has an expectation of \$100
and a variance of $0.0004x^2 + 0.0009(1000 - x)^2$.

This variance is minimized by taking $x = \$692$,
and the minimum value of the variance is 276.9
which corresponds to a standard deviation of \$16.64.

This problem illustrates that the variability of the return on an investment can be reduced by *diversifying* the investment, so that it is spread over several funds.

2.6.12 The expected value of the total resistance is

$5 \times E(X) = 5 \times 10.418234 = 52.09.$

The variance of the total resistance is

$5 \times \mathrm{Var}(X) = 5 \times 0.0758 = 0.379$

so that the standard deviation is $\sqrt{0.379} = 0.616$.

2.6.13 (a) The mean is

$$E(X) = \left(\tfrac{1}{3} \times E(X_1)\right) + \left(\tfrac{1}{3} \times E(X_2)\right) + \left(\tfrac{1}{3} \times E(X_3)\right)$$

$$= \left(\tfrac{1}{3} \times 59\right) + \left(\tfrac{1}{3} \times 67\right) + \left(\tfrac{1}{3} \times 72\right) = 66$$

The variance is

$$\mathrm{Var}(X) = \left(\left(\tfrac{1}{3}\right)^2 \times \mathrm{Var}(X_1)\right) + \left(\left(\tfrac{1}{3}\right)^2 \times \mathrm{Var}(X_2)\right) + \left(\left(\tfrac{1}{3}\right)^2 \times \mathrm{Var}(X_3)\right)$$

$$= \left(\left(\tfrac{1}{3}\right)^2 \times 10^2\right) + \left(\left(\tfrac{1}{3}\right)^2 \times 13^2\right) + \left(\left(\tfrac{1}{3}\right)^2 \times 4^2\right) = \tfrac{95}{3}$$

so that the standard deviation is $\sqrt{95/3} = 5.63$.

(b) The mean is

$$E(X) = (0.4 \times E(X_1)) + (0.4 \times E(X_2)) + (0.2 \times E(X_3))$$

$$= (0.4 \times 59) + (0.4 \times 67) + (0.2 \times 72) = 64.8.$$

The variance is

$$\mathrm{Var}(X) = (0.4^2 \times \mathrm{Var}(X_1)) + (0.4^2 \times \mathrm{Var}(X_2)) + (0.2^2 \times \mathrm{Var}(X_3))$$

$$= (0.4^2 \times 10^2) + (0.4^2 \times 13^2) + (0.2^2 \times 4^2) = 43.68$$

so that the standard deviation is $\sqrt{43.68} = 6.61$.

2.6.14 $1000 = E(Y) = a + bE(X) = a + (b \times 77)$

$10^2 = \mathrm{Var}(Y) = b^2\mathrm{Var}(X) = b^2 \times 9^2$

Solving these equations gives $a = 914.44$ and $b = 1.11$, or $a = 1085.56$ and $b = -1.11$.

2.6.15 (a) The mean is $\mu = 65.90$.
The standard deviation is $\frac{\sigma}{\sqrt{5}} = \frac{0.32}{\sqrt{5}} = 0.143$.

(b) The mean is $8\mu = 8 \times 65.90 = 527.2$.
The standard deviation is $\sqrt{8}\sigma = \sqrt{8} \times 0.32 = 0.905$.

2.6.16 (a) $E(A) = \frac{E(X_1)+E(X_2)}{2} = \frac{W+W}{2} = W$

$\text{Var}(A) = \frac{\text{Var}(X_1)+\text{Var}(X_2)}{4} = \frac{3^2+4^2}{4} = \frac{25}{4}$

The standard deviation is $\frac{5}{2} = 2.5$.

(b) $\text{Var}(B) = \delta^2\text{Var}(X_1) + (1-\delta)^2\text{Var}(X_2) = 9\delta^2 + 16(1-\delta)^2$

This is minimized when $\delta = \frac{16}{25}$ and the minimum value is $\frac{144}{25}$

so that the minimum standard deviation is $\frac{12}{5} = 2.4$.

2.6.17 When a die is rolled once the expectation is 3.5 and the standard deviation is 1.71 (see Games of Chance in section 2.4).
Therefore, the sum of eighty die rolls has an expectation of $80 \times 3.5 = 280$
and a standard deviation of $\sqrt{80} \times 1.71 = 15.3$.

2.6.18 (a) The expectation is $4 \times 33.2 = 132.8$ seconds.
The standard deviation is $\sqrt{4} \times 1.4 = 2.8$ seconds.

(b) $E(A_1 + A_2 + A_3 + A_4 - B_1 - B_2 - B_3 - B_4)$
$= E(A_1) + E(A_2) + E(A_3) + E(A_4) - E(B_1) - E(B_2) - E(B_3) - E(B_4)$
$= (4 \times 33.2) - (4 \times 33.0) = 0.8$

$\text{Var}(A_1 + A_2 + A_3 + A_4 - B_1 - B_2 - B_3 - B_4)$
$= \text{Var}(A_1) + \text{Var}(A_2) + \text{Var}(A_3) + \text{Var}(A_4)$
$+ \text{Var}(B_1) + \text{Var}(B_2) + \text{Var}(B_3) + \text{Var}(B_4)$
$= (4 \times 1.4^2) + (4 \times 1.3^2) = 14.6$

The standard deviation is $\sqrt{14.6} = 3.82$.

(c) $E\left(A_1 - \frac{A_2+A_3+A_4}{3}\right)$

$= E(A_1) - \frac{E(A_2)}{3} - \frac{E(A_3)}{3} - \frac{E(A_4)}{3} = 0$

$\text{Var}\left(A_1 - \frac{A_2+A_3+A_4}{3}\right)$

$= \text{Var}(A_1) + \frac{\text{Var}(A_2)}{9} + \frac{\text{Var}(A_3)}{9} + \frac{\text{Var}(A_4)}{9}$

$= \frac{4}{3} \times 1.4^2 = 2.613$

The standard deviation is $\sqrt{2.613} = 1.62$.

2.6.19 Let X be the temperature in Fahrenheit and let Y be the temperature in Centigrade.

$E(Y) = E\left(\frac{5(X-32)}{9}\right) = \left(\frac{5(E(X)-32)}{9}\right) = \left(\frac{5(110-32)}{9}\right) = 43.33$

$\text{Var}(Y) = \text{Var}\left(\frac{5(X-32)}{9}\right) = \left(\frac{5^2\text{Var}(X)}{9^2}\right) = \left(\frac{5^2\times 2.2^2}{9^2}\right) = 1.49$

The standard deviation is $\sqrt{1.49} = 1.22$.

2.6.20 $\text{Var}(0.5X_\alpha + 0.3X_\beta + 0.2X_\gamma)$

$= 0.5^2\text{Var}(X_\alpha) + 0.3^2\text{Var}(X_\beta) + 0.2^2\text{Var}(X_\gamma)$

$= (0.5^2 \times 1.2^2) + (0.3^2 \times 2.4^2) + (0.2^2 \times 3.1^2) = 1.26$

The standard deviation is $\sqrt{1.26} = 1.12$.

2.6.21 The inequality $\frac{56}{\sqrt{n}} \leq 10$ is satisfied for $n \geq 32$.

2.6.22 (a) $E(X_1 + X_2) = E(X_1) + E(X_2) = 7.74$
$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) = 0.0648$
The standard deviation is $\sqrt{0.0648} = 0.255$.

(b) $E(X_1 + X_2 + X_3) = E(X_1) + E(X_2) + E(X_3) = 11.61$
$\text{Var}(X_1 + X_2 + X_3) = \text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3) = 0.0972$
The standard deviation is $\sqrt{0.0972} = 0.312$.

(c) $E\left(\frac{X_1+X_2+X_3+X_4}{4}\right)$

$= \frac{E(X_1)+E(X_2)+E(X_3)+E(X_4)}{4}$

$= 3.87$

$\text{Var}\left(\frac{X_1+X_2+X_3+X_4}{4}\right)$

$= \frac{\text{Var}(X_1)+\text{Var}(X_2)+\text{Var}(X_3)+\text{Var}(X_4)}{16}$

$= 0.0081$

The standard deviation is $\sqrt{0.0081} = 0.09$.

(d) $E\left(X_3 - \frac{X_1+X_2}{2}\right) = E(X_3) - \frac{E(X_1)+E(X_2)}{2} = 0$

$\mathrm{Var}\left(X_3 - \frac{X_1+X_2}{2}\right) = \mathrm{Var}(X_3) + \frac{\mathrm{Var}(X_1)+\mathrm{Var}(X_2)}{4} = 0.0486$

The standard deviation is $\sqrt{0.0486} = 0.220$.

2.8 Supplementary Problems

2.8.1 (a)

x_i	2	3	4	5	6
p_i	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{4}{15}$	$\frac{5}{15}$

(b) $E(X) = \left(2 \times \frac{1}{15}\right) + \left(3 \times \frac{2}{15}\right) + \left(4 \times \frac{3}{15}\right) + \left(5 \times \frac{4}{15}\right) + \left(6 \times \frac{5}{15}\right) = \frac{14}{3}$

2.8.2 (a)

x_i	0	1	2	3	4	5	6
$F(x_i)$	0.21	0.60	0.78	0.94	0.97	0.99	1.00

(b) $E(X) = (0 \times 0.21) + (1 \times 0.39) + (2 \times 0.18) + (3 \times 0.16)$
$+ (4 \times 0.03) + (5 \times 0.02) + (6 \times 0.01)$
$= 1.51$

(c) $E(X^2) = (0^2 \times 0.21) + (1^2 \times 0.39) + (2^2 \times 0.18) + (3^2 \times 0.16)$
$+ (4^2 \times 0.03) + (5^2 \times 0.02) + (6^2 \times 0.01)$
$= 3.89$

$\text{Var}(X) = 3.89 - (1.51)^2 = 1.61$

(d) The expectation is $1.51 \times 60 = 90.6$
and the variance is $1.61 \times 60 = 96.6$.

2.8.3 (a)

x_i	2	3	4	5
p_i	$\frac{2}{30}$	$\frac{13}{30}$	$\frac{13}{30}$	$\frac{2}{30}$

(b) $E(X) = \left(2 \times \frac{2}{30}\right) + \left(3 \times \frac{13}{30}\right) + \left(4 \times \frac{13}{30}\right) + \left(5 \times \frac{2}{30}\right) = \frac{7}{2}$

$E(X^2) = \left(2^2 \times \frac{2}{30}\right) + \left(3^2 \times \frac{13}{30}\right) + \left(4^2 \times \frac{13}{30}\right) + \left(5^2 \times \frac{2}{30}\right) = \frac{383}{30}$

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \frac{383}{30} - \left(\frac{7}{2}\right)^2 = \frac{31}{60}$$

(c)

x_i	2	3	4	5
p_i	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{3}{10}$	$\frac{2}{10}$

$$E(X) = \left(2 \times \frac{2}{10}\right) + \left(3 \times \frac{3}{10}\right) + \left(4 \times \frac{3}{10}\right) + \left(5 \times \frac{2}{10}\right) = \frac{7}{2}$$

$$E(X^2) = \left(2^2 \times \frac{2}{10}\right) + \left(3^2 \times \frac{3}{10}\right) + \left(4^2 \times \frac{3}{10}\right) + \left(5^2 \times \frac{2}{10}\right) = \frac{133}{10}$$

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \frac{133}{10} - \left(\frac{7}{2}\right)^2 = \frac{21}{20}$$

2.8.4 Let X_i be the value of the i^{th} card dealt.

Then

$$E(X_i) = \left(2 \times \frac{1}{13}\right) + \left(3 \times \frac{1}{13}\right) + \left(4 \times \frac{1}{13}\right) + \left(5 \times \frac{1}{13}\right) + \left(6 \times \frac{1}{13}\right)$$

$$+ \left(7 \times \frac{1}{13}\right) + \left(8 \times \frac{1}{13}\right) + \left(9 \times \frac{1}{13}\right) + \left(10 \times \frac{1}{13}\right) + \left(15 \times \frac{4}{13}\right) = \frac{114}{13}$$

The total score of the hand is

$Y = X_1 + \ldots + X_{13}$

which has an expectation

$E(Y) = E(X_1) + \ldots + E(X_{13}) = 13 \times \frac{114}{13} = 114.$

2.8.5 (a) Since

$\int_1^{11} A \left(\frac{3}{2}\right)^x dx = 1$

it follows that $A = \frac{\ln(1.5)}{1.5^{11} - 1.5} = \frac{1}{209.6}$.

(b) $F(x) = \int_1^x \frac{1}{209.6} \left(\frac{3}{2}\right)^y dy$

$= 0.01177 \left(\frac{3}{2}\right)^x - 0.01765$

for $1 \leq x \leq 11$

(c) Solving $F(x) = 0.5$ gives $x = 9.332$.

(d) Solving $F(x) = 0.25$ gives $x = 7.706$.

Solving $F(x) = 0.75$ gives $x = 10.305$.

The interquartile range is $10.305 - 7.706 = 2.599$.

2.8.6 (a) $f_X(x) = \int_1^2 4x(2-y)\ dy = 2x$ for $0 \leq x \leq 1$

(b) $f_Y(y) = \int_0^1 4x(2-y)\ dx = 2(2-y)$ for $1 \leq y \leq 2$

Since $f(x,y) = f_X(x) \times f_Y(y)$ the random variables are independent.

(c) $\mathrm{Cov}(X,Y) = 0$ because the random variables are independent.

(d) $f_{X|Y=1.5}(x) = f_X(x)$ because the random variables are independent.

2.8.7 (a) Since

$\int_5^{10} A\left(x + \frac{2}{x}\right) dx = 1$

it follows that $A = 0.02572$.

(b) $F(x) = \int_5^x 0.02572\left(y + \frac{2}{y}\right) dy$

$= 0.0129x^2 + 0.0514\ln(x) - 0.404$

for $5 \leq x \leq 10$

(c) $E(X) = \int_5^{10} 0.02572\ x\ \left(x + \frac{2}{x}\right) dx = 7.759$

(d) $E(X^2) = \int_5^{10} 0.02572\ x^2\ \left(x + \frac{2}{x}\right) dx = 62.21$

$\mathrm{Var}(X) = E(X^2) - E(X)^2 = 62.21 - 7.759^2 = 2.01$

(e) Solving $F(x) = 0.5$ gives $x = 7.88$.

(f) Solving $F(x) = 0.25$ gives $x = 6.58$.
Solving $F(x) = 0.75$ gives $x = 9.00$.
The interquartile range is $9.00 - 6.58 = 2.42$.

(g) The expectation is $E(X) = 7.759$.

The variance is $\frac{\mathrm{Var}(X)}{10} = 0.0201$.

2.8.8 $\mathrm{Var}(a_1X_1 + a_2X_2 + \ldots + a_nX_n + b)$
$= \mathrm{Var}(a_1X_1) + \ldots + \mathrm{Var}(a_nX_n) + \mathrm{Var}(b)$
$= a_1^2\mathrm{Var}(X_1) + \ldots + a_n^2\mathrm{Var}(X_n) + 0$

2.8.9 $Y = \frac{5}{3}X - 25$

2.8.10 Notice that $E(Y) = aE(X) + b$ and $\text{Var}(Y) = a^2\text{Var}(X)$.

Also,

$\text{Cov}(X,Y) = E(\ (X - E(X))\ (Y - E(Y))\ )$

$= E(\ (X - E(X))\ a(X - E(X))\ )$

$= a\text{Var}(X)$.

Therefore,

$\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{a\text{Var}(X)}{\sqrt{\text{Var}(X)a^2\text{Var}(X)}} = \frac{a}{|a|}$

which is 1 if $a > 0$ and is -1 if $a < 0$.

2.8.11 The expected amount of a claim is

$E(X) = \int_0^{1800} x\ \frac{x(1800-x)}{972{,}000{,}000}\ dx = \$900$.

Consequently, the expected profit from each customer is

$\$100 - \$5 - (0.1 \times \$900) = \$5$.

The expected profit from 10,000 customers is therefore $10,000 \times \$5 = \$50,000$.

The profits may or may not be independent depending on the type of insurance and the pool of customers.

If large natural disasters affect the pool of customers all at once then the claims would not be independent.

2.8.12 (a) The expectation is $5 \times 320 = 1600$ seconds.
The variance is $5 \times 63^2 = 19845$
and the standard deviation is $\sqrt{19845} = 140.9$ seconds.

(b) The expectation is 320 seconds.
The variance is $\frac{63^2}{10} = 396.9$
and the standard deviation is $\sqrt{396.9} = 19.92$ seconds.

2.8.13 (a) The state space is the positive integers from 1 to n,
with each outcome having a probability value of $\frac{1}{n}$.

(b) $E(X) = \left(\frac{1}{n} \times 1\right) + \left(\frac{1}{n} \times 2\right) + \ldots + \left(\frac{1}{n} \times n\right) = \frac{n+1}{2}$

(c) $E(X^2) = \left(\frac{1}{n} \times 1^2\right) + \left(\frac{1}{n} \times 2^2\right) + \ldots + \left(\frac{1}{n} \times n^2\right) = \frac{(n+1)(2n+1)}{6}$

Therefore,
$\text{Var}(X) = E(X^2) - (E(X))^2 = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2-1}{12}$.

2.8.14 (a) Let X_T be the amount of time that Tom spends on the bus
and let X_N be the amount of time that Nancy spends on the bus.
Therefore, the sum of the times is $X = X_T + X_N$ and
$E(X) = E(X_T) + E(X_N) = 87 + 87 = 174$ minutes.

If Tom and Nancy ride on different buses then the random variables X_T and X_N are independent so that
$\text{Var}(X) = \text{Var}(X_T) + \text{Var}(X_N) = 3^2 + 3^2 = 18$
and the standard deviation is $\sqrt{18} = 4.24$ minutes.

(b) If Tom and Nancy ride together on the same bus then $X_T = X_N$ so that $X = 2 \times X_T$, twice the time of the ride.
In this case
$E(X) = 2 \times E(X_T) = 2 \times 87 = 174$ minutes
and
$\text{Var}(X) = 2^2 \times \text{Var}(X_1) = 2^2 \times 3^2 = 36$
so that the standard deviation is $\sqrt{36} = 6$ minutes.

2.8.15 (a) Two heads gives a total score of 20.
One head and one tail gives a total score of 30.
Two tails gives a total score of 40.
Therefore, the state space is $\{20, 30, 40\}$.

(b) $P(X = 20) = \frac{1}{4}$

$P(X = 30) = \frac{1}{2}$

$P(X = 40) = \frac{1}{4}$

(c) $P(X \leq 20) = \frac{1}{4}$

$P(X \leq 30) = \frac{3}{4}$

$P(X \leq 40) = 1$

(d) $E(X) = \left(20 \times \frac{1}{4}\right) + \left(30 \times \frac{1}{2}\right) + \left(40 \times \frac{1}{4}\right) = 30$

(e) $E(X^2) = \left(20^2 \times \frac{1}{4}\right) + \left(30^2 \times \frac{1}{2}\right) + \left(40^2 \times \frac{1}{4}\right) = 950$

$\text{Var}(X) = 950 - 30^2 = 50$

The standard deviation is $\sqrt{50} = 7.07$.

2.8.16 (a) Since

$\int_5^6 Ax\ dx = \frac{A}{2} \times (6^2 - 5^2) = 1$

it follows that $A = \frac{2}{11}$.

(b) $F(x) = \int_5^x \frac{2y}{11}\ dy = \frac{x^2-25}{11}$

(c) $E(X) = \int_5^6 \frac{2x^2}{11}\ dx = \frac{2\times(6^3-5^3)}{33} = \frac{182}{33} = 5.52$

(d) $E(X^2) = \int_5^6 \frac{2x^3}{11}\ dx = \frac{6^4-5^4}{22} = \frac{671}{22} = 30.5$

$\text{Var}(X) = 30.5 - \left(\frac{182}{33}\right)^2 = 0.083$

The standard deviation is $\sqrt{0.083} = 0.29$.

2.8.17 (a) The expectation is $3 \times 438 = 1314$.
The standard deviation is $\sqrt{3} \times 4 = 6.93$.

(b) The expectation is 438.
The standard deviation is $\frac{4}{\sqrt{8}} = 1.41$.

2.8.18 (a) If a 1 is obtained from the die the net winnings are $(3 \times \$1) - \$5 = -\$2$
If a 2 is obtained from the die the net winnings are $\$2 - \$5 = -\$3$
If a 3 is obtained from the die the net winnings are $(3 \times \$3) - \$5 = \$4$
If a 4 is obtained from the die the net winnings are $\$4 - \$5 = -\$1$
If a 5 is obtained from the die the net winnings are $(3 \times \$5) - \$5 = \$10$
If a 6 is obtained from the die the net winnings are $\$6 - \$5 = \$1$

Each of these values has a probability of $\frac{1}{6}$.

(b) $E(X) = \left(-3 \times \frac{1}{6}\right) + \left(-2 \times \frac{1}{6}\right) + \left(-1 \times \frac{1}{6}\right)$

$+ \ \left(1 \times \frac{1}{6}\right) + \left(4 \times \frac{1}{6}\right) + \left(10 \times \frac{1}{6}\right) = \frac{3}{2}$

$E(X^2) = \left((-3)^2 \times \frac{1}{6}\right) + \left((-2)^2 \times \frac{1}{6}\right) + \left((-1)^2 \times \frac{1}{6}\right)$

$+ \ \left(1^2 \times \frac{1}{6}\right) + \left(4^2 \times \frac{1}{6}\right) + \left(10^2 \times \frac{1}{6}\right) = \frac{131}{6}$

The variance is

$\frac{131}{6} - \left(\frac{3}{2}\right)^2 = \frac{235}{12}$

and the standard deviation is $\sqrt{\frac{235}{12}} = \$4.43$.

(c) The expectation is $10 \times \frac{3}{2} = \$15$.
The standard deviation is $\sqrt{10} \times 4.43 = \$13.99$.

2.8.19 (a) False

(b) True

(c) True

(d) True

(e) True

(f) False

2.8.20 $E(\text{total time}) = 5 \times \mu = 5 \times 22 = 110$ minutes
The standard deviation of the total time is $\sqrt{5}\sigma = \sqrt{5} \times 1.8 = 4.02$ minutes.

$E(\text{average time}) = \mu = 22$ minutes
The standard deviation of the average time is $\frac{\sigma}{\sqrt{5}} = \frac{1.8}{\sqrt{5}} = 0.80$ minutes.

2.8.21 (a) $E(X) = (0 \times 0.12) + (1 \times 0.43) + (2 \times 0.28) + (3 \times 0.17) = 1.50$

(b) $E(X^2) = (0^2 \times 0.12) + (1^2 \times 0.43) + (2^2 \times 0.28) + (3^2 \times 0.17) = 3.08$
The variance is $3.08 - 1.50^2 = 0.83$
and the standard deviation is $\sqrt{0.83} = 0.911$.

(c) $E(X_1 + X_2) = E(X_1) + E(X_2) = 1.50 + 1.50 = 3.00$
$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) = 0.83 + 0.83 = 1.66$
The standard deviation is $\sqrt{1.66} = 1.288$.

2.8.22 (a) $E(X) = (-22 \times 0.3) + (3 \times 0.2) + (19 \times 0.1) + (23 \times 0.4) = 5.1$

(b) $E(X^2) = ((-22)^2 \times 0.3) + (3^2 \times 0.2) + (19^2 \times 0.1) + (23^2 \times 0.4) = 394.7$
$\text{Var}(X) = 394.7 - 5.1^2 = 368.69$
The standard deviation is $\sqrt{368.69} = 19.2$.

2.8.23 (a) Since

$\int_2^4 f(x)\, dx = \int_2^4 \frac{A}{x^2}\, dx = \frac{A}{4} = 1$

it follows that $A = 4$.

(b) Since

$\frac{1}{4} = \int_2^y f(x)\ dx = \int_2^y \frac{4}{x^2}\ dx = \left(2 - \frac{4}{y}\right)$

it follows that $y = \frac{16}{7} = 2.29$.

2.8.24 (a) $100 = E(Y) = c + dE(X) = c + (d \times 250)$
$1 = \mathrm{Var}(Y) = d^2\mathrm{Var}(X) = d^2 \times 16$

Solving these equations gives $d = \frac{1}{4}$ and $c = \frac{75}{2}$
or $d = -\frac{1}{4}$ and $c = \frac{325}{2}$.

(b) The mean is $10 \times 250 = 1000$.
The standard deviation is $\sqrt{10} \times 4 = 12.65$.

2.8.25 Since

$E(c_1X_1 + c_2X_2) = c_1E(X_1) + c_2E(X_2) = (c_1 + c_2) \times 100 = 100$

it is necessary that $c_1 + c_2 = 1$.

Also,

$\mathrm{Var}(c_1X_1 + c_2X_2) = c_1^2\mathrm{Var}(X_1) + c_2^2\mathrm{Var}(X_2) = (c_1^2 \times 144) + (c_2^2 \times 169) = 100.$

Solving these two equations gives $c_1 = 0.807$ and $c_2 = 0.193$
or $c_1 = 0.273$ and $c_2 = 0.727$.

2.8.26 (a) The mean is $3\mu_A = 3 \times 134.9 = 404.7$.

The standard deviation is $\sqrt{3}\ \sigma_A = \sqrt{3} \times 0.7 = 1.21$.

(b) The mean is $2\mu_A + 2\mu_B = (2 \times 134.9) + (2 \times 138.2) = 546.2$.

The standard deviation is $\sqrt{0.7^2 + 0.7^2 + 1.1^2 + 1.1^2} = 1.84$.

(c) The mean is $\frac{4\mu_A + 3\mu_B}{7} = \frac{(4\times134.9)+(3\times138.2)}{7} = 136.3$.

The standard deviation is $\frac{\sqrt{0.7^2+0.7^2+0.7^2+0.7^2+1.1^2+1.1^2+1.1^2}}{7} = 0.34$.

Chapter 3

Discrete Probability Distributions

3.1 The Binomial Distribution

3.1.1 (a) $P(X = 3) = \binom{10}{3} \times 0.12^3 \times 0.88^7 = 0.0847$

(b) $P(X = 6) = \binom{10}{6} \times 0.12^6 \times 0.88^4 = 0.0004$

(c) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$
$= 0.2785 + 0.3798 + 0.2330$
$= 0.8913$

(d) $P(X \geq 7) = P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$
$= 3.085 \times 10^{-5}$

(e) $E(X) = 10 \times 0.12 = 1.2$

(f) $\text{Var}(X) = 10 \times 0.12 \times 0.88 = 1.056$

3.1.2 (a) $P(X = 4) = \binom{7}{4} \times 0.8^4 \times 0.2^3 = 0.1147$

(b) $P(X \neq 2) = 1 - P(X = 2)$

$= 1 - \binom{7}{2} \times 0.8^2 \times 0.2^5$

$= 0.9957$

(c) $P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 0.0334$

(d) $P(X \geq 6) = P(X = 6) + P(X = 7) = 0.5767$

(e) $E(X) = 7 \times 0.8 = 5.6$

(f) $\text{Var}(X) = 7 \times 0.8 \times 0.2 = 1.12$

3.1.3 $X \sim B(6, 0.5)$

x_i	0	1	2	3	4	5	6
p_i	0.0156	0.0937	0.2344	0.3125	0.2344	0.0937	0.0156

$E(X) = 6 \times 0.5 = 3$
$\text{Var}(X) = 6 \times 0.5 \times 0.5 = 1.5$
$\sigma = \sqrt{1.5} = 1.22$

$X \sim B(6, 0.7)$

x_i	0	1	2	3	4	5	6
p_i	0.0007	0.0102	0.0595	0.1852	0.3241	0.3025	0.1176

$E(X) = 6 \times 0.7 = 4.2$
$\text{Var}(X) = 6 \times 0.7 \times 0.3 = 1.26$
$\sigma = \sqrt{1.5} = 1.12$

3.1.4 $X \sim B(9, 0.09)$

(a) $P(X = 2) = 0.1507$

(b) $P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 0.1912$

$E(X) = 9 \times 0.09 = 0.81$

3.1.5 (a) $P\left(B\left(8, \frac{1}{2}\right) = 5\right) = 0.2187$

(b) $P\left(B\left(8, \frac{1}{6}\right) = 1\right) = 0.3721$

(c) $P\left(B\left(8, \frac{1}{6}\right) = 0\right) = 0.2326$

(d) $P\left(B\left(8, \frac{2}{3}\right) \geq 6\right) = 0.4682$

3.1.6 $P(B(10, 0.2) \geq 7) = 0.0009$

$P(B(10, 0.5) \geq 7) = 0.1719$

3.1.7 Let the random variable X be the number of employees taking sick leave.

Then $X \sim B(180, 0.35)$.

Therefore, the *proportion* of the workforce who need to take sick leave is

$Y = \frac{X}{180}$

so that

$E(Y) = \frac{E(X)}{180} = \frac{180 \times 0.35}{180} = 0.35$

and

$\text{Var}(Y) = \frac{\text{Var}(X)}{180^2} = \frac{180 \times 0.35 \times 0.65}{180^2} = 0.0013.$

In general, the variance is

$\text{Var}(Y) = \frac{\text{Var}(X)}{180^2} = \frac{180 \times p \times (1-p)}{180^2} = \frac{p \times (1-p)}{180}$

which is maximized when $p = 0.5$.

3.1.8 The random variable Y can be considered to be the number of successes out of $n_1 + n_2$ trials.

3.1.9 $X \sim B(18, 0.6)$

(a) $P(X = 8) + P(X = 9) + P(X = 10)$

$= \binom{18}{8} \times 0.6^8 \times 0.4^{10} + \binom{18}{9} \times 0.6^9 \times 0.4^9 + \binom{18}{10} \times 0.6^{10} \times 0.4^8$

$= 0.0771 + 0.1284 + 0.1734 = 0.3789$

(b) $P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4)$

$$= \binom{18}{0} \times 0.6^0 \times 0.4^{18} + \binom{18}{1} \times 0.6^1 \times 0.4^{17} + \binom{18}{2} \times 0.6^2 \times 0.4^{16}$$

$$+ \binom{18}{3} \times 0.6^3 \times 0.4^{15} + \binom{18}{4} \times 0.6^4 \times 0.4^{14}$$

$$= 0.0013$$

3.2 The Geometric and Negative Binomial Distributions

3.2.1 (a) $P(X = 4) = (1 - 0.7)^3 \times 0.7 = 0.0189$

(b) $P(X = 1) = (1 - 0.7)^0 \times 0.7 = 0.7$

(c) $P(X \leq 5) = 1 - (1 - 0.7)^5 = 0.9976$

(d) $P(X \geq 8) = 1 - P(X \leq 7) = (1 - 0.7)^7 = 0.0002$

3.2.2 (a) $P(X = 5) = \binom{4}{2} \times (1 - 0.6)^2 \times 0.6^3 = 0.2074$

(b) $P(X = 8) = \binom{7}{2} \times (1 - 0.6)^5 \times 0.6^3 = 0.0464$

(c) $P(X \leq 7) = P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7)$
$= 0.9037$

(d) $P(X \geq 7) = 1 - P(X = 3) - P(X = 4) - P(X = 5) - P(X = 6)$
$= 0.1792$

3.2.4 Notice that a negative binomial distribution with parameters p and r can be thought of as the number of trials up to and including the r^{th} success in a sequence of independent Bernoulli trials with a constant success probability p, which can be considered to be the number of trials up to and including the first success, plus the number of trials after the first success and up to and including the second success, plus the number of trials after the second success and up to and including the third success, and so on. Each of these r components has a geometric distribution with parameter p.

3.2.5 (a) Consider a geometric distribution with parameter $p = 0.09$.
$(1 - 0.09)^3 \times 0.09 = 0.0678$

(b) Consider a negative binomial distribution with parameters $p = 0.09$ and $r = 3$.
$\binom{9}{2} \times (1 - 0.09)^7 \times 0.09^3 = 0.0136$

(c) $\frac{1}{0.09} = 11.11$

(d) $\frac{3}{0.09} = 33.33$

3.2.6 (a) $\frac{1}{0.37} = 2.703$

(b) $\frac{3}{0.37} = 8.108$

(c) The required probability is
$P(X \leq 10) = 0.7794$
where the random variable X has a negative binomial distribution with parameters $p = 0.37$ and $r = 3$.

Alternatively, the required probability is
$P(Y \geq 3) = 0.7794$
where the random variable Y has a binomial distribution with parameters
$n = 10$ and $p = 0.37$.

(d) $P(X = 10) = \binom{9}{2} \times (1 - 0.37)^7 \times 0.37^3 = 0.0718$

3.2.7 (a) Consider a geometric distribution with parameter $p = 0.25$.
$(1 - 0.25)^2 \times 0.25 = 0.1406$

(b) Consider a negative binomial distribution with parameters $p = 0.25$ and $r = 4$.
$\binom{9}{3} \times (1 - 0.25)^6 \times 0.25^4 = 0.0584$

The expected number of cards drawn before the fourth heart is obtained is the expectation of a negative binomial distribution with parameters $p = 0.25$ and $r = 4$, which is $\frac{4}{0.25} = 16$.

If the first two cards are spades then the probability that the first heart card is obtained on the fifth drawing is the same as the probability in part (a).

3.2.8 (a) $\frac{1}{0.77} = 1.299$

(b) Consider a geometric distribution with parameter $p = 0.23$.
$(1 - 0.23)^4 \times 0.23 = 0.0809$

(c) Consider a negative binomial distribution with parameters $p = 0.77$ and $r = 3$.
$\binom{5}{2} \times (1 - 0.77)^3 \times 0.77^3 = 0.0555$

(d) $P(B(8, 0.77) \geq 3) = 0.9973$

3.2.9 (a) Consider a geometric distribution with parameter $p = 0.6$.
$P(X = 5) = (1 - 0.6)^4 \times 0.6 = 0.01536$

(b) Consider a negative binomial distribution with parameters $p = 0.6$ and $r = 4$.
$P(X = 8) = \binom{7}{3} \times 0.6^4 \times 0.4^4 = 0.116$

3.2.10 $E(X) = \frac{r}{p} = \frac{3}{1/6} = 18$

3.2.11 $P(X = 10) = \binom{9}{4} \times \left(\frac{1}{2}\right)^5 \times \left(\frac{1}{2}\right)^5 = 0.123$

3.3 The Hypergeometric Distribution

3.3.1 (a) $P(X=4)=\dfrac{\binom{6}{4}\times\binom{5}{3}}{\binom{11}{7}}=\frac{5}{11}$

(b) $P(X=5)=\dfrac{\binom{6}{5}\times\binom{5}{2}}{\binom{11}{7}}=\frac{2}{11}$

(c) $P(X\leq 3)=P(X=2)+P(X=3)=\frac{23}{66}$

3.3.2

x_i	0	1	2	3	4	5
p_i	$\frac{3}{429}$	$\frac{40}{429}$	$\frac{140}{429}$	$\frac{168}{429}$	$\frac{70}{429}$	$\frac{8}{429}$

3.3.3 (a) $\dfrac{\binom{10}{3}\times\binom{7}{2}}{\binom{17}{5}}=\frac{90}{221}$

(b) $\dfrac{\binom{10}{1}\times\binom{7}{4}}{\binom{17}{5}}=\frac{25}{442}$

(c) $P(\text{no red balls})+P(\text{one red ball})+P(\text{two red balls})=\frac{139}{442}$

3.3.4 $\dfrac{\binom{16}{5}\times\binom{18}{7}}{\binom{34}{12}}=0.2535$

$P\left(B\left(12,\frac{18}{34}\right)=7\right)=0.2131$

3.3.5 $$\frac{\binom{12}{3} \times \binom{40}{2}}{\binom{52}{5}} = \frac{55}{833}$$

The number of picture cards X in a hand of 13 cards has a hypergeometric distribution with $N = 52$, $n = 13$, and $r = 12$. The expected value is

$E(X) = \frac{13 \times 12}{52} = 3$

and the variance is

$\mathrm{Var}(X) = \left(\frac{52-13}{52-1}\right) \times 13 \times \frac{12}{52} \times \left(1 - \frac{12}{52}\right) = \frac{30}{17}$.

3.3.6 $$\frac{\binom{4}{1} \times \binom{5}{2} \times \binom{6}{2}}{\binom{15}{5}} = \frac{200}{1001}$$

3.3.7 (a) $$\frac{\binom{7}{3} \times \binom{4}{0}}{\binom{11}{3}} = \frac{7}{33}$$

(b) $$\frac{\binom{7}{1} \times \binom{4}{2}}{\binom{11}{3}} = \frac{7}{165}$$

3.3.8 $P(5 \leq X \leq 7) = P(X = 5) + P(X = 6) + P(X = 7)$

$$= \frac{\binom{9}{4} \times \binom{6}{1}}{\binom{15}{5}} + \frac{\binom{9}{3} \times \binom{6}{2}}{\binom{15}{5}} + \frac{\binom{9}{2} \times \binom{6}{3}}{\binom{15}{5}}$$

$= 0.911$

3.3.9 (a) $$\frac{\binom{8}{2} \times \binom{8}{2}}{\binom{16}{4}} = \frac{28}{65} = 0.431$$

(b) $P\left(B\left(4, \frac{1}{2}\right) = 2\right) = \binom{4}{2} \times \left(\frac{1}{2}\right)^2 \times \left(\frac{1}{2}\right)^2 = \frac{3}{8} = 0.375$

3.3.10 $$\frac{\binom{19}{4}\times\binom{6}{1}}{\binom{25}{5}} + \frac{\binom{19}{5}\times\binom{6}{0}}{\binom{25}{5}} = 0.4377 + 0.2189 = 0.6566$$

3.4 The Poisson Distribution

3.4.1 (a) $P(X = 1) = \frac{e^{-3.2} \times 3.2^1}{1!} = 0.1304$

(b) $P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 0.6025$

(c) $P(X \geq 6) = 1 - P(X \leq 5) = 0.1054$

(d) $P(X = 0|X \leq 3) = \frac{P(X=0)}{P(X \leq 3)} = \frac{0.0408}{0.6025} = 0.0677$

3.4.2 (a) $P(X = 0) = \frac{e^{-2.1} \times 2.1^0}{0!} = 0.1225$

(b) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0.6496$

(c) $P(X \geq 5) = 1 - P(X \leq 4) = 0.0621$

(d) $P(X = 1|X \leq 2) = \frac{P(X=1)}{P(X \leq 2)} = \frac{0.2572}{0.6496} = 0.3959$

3.4.4 $P(X = 0) = \frac{e^{-2.4} \times 2.4^0}{0!} = 0.0907$

$P(X \geq 4) = 1 - P(X \leq 3) = 0.2213$

3.4.5 It is best to use a Poisson distribution with $\lambda = \frac{25}{100} = 0.25$.

$P(X = 0) = \frac{e^{-0.25} \times 0.25^0}{0!} = 0.7788$

$P(X \leq 1) = P(X = 0) + P(X = 1) = 0.9735$

3.4.6 It is best to use a Poisson distribution with $\lambda = 4$.

(a) $P(X = 0) = \frac{e^{-4} \times 4^0}{0!} = 0.0183$

(b) $P(X \geq 6) = 1 - P(X \leq 5) = 0.2149$

3.4.7 A $B(500, 0.005)$ distribution can be approximated by a
Poisson distribution with $\lambda = 500 \times 0.005 = 2.5$.
Therefore,
$P(B(500, 0.005) \leq 3)$
$\simeq \frac{e^{-2.5} \times 2.5^0}{0!} + \frac{e^{-2.5} \times 2.5^1}{1!} + \frac{e^{-2.5} \times 2.5^2}{2!} + \frac{e^{-2.5} \times 2.5^3}{3!}$

$= 0.7576$

3.4.8 $X \sim P(9.2)$

(a) $P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$

$= \frac{e^{-9.2} \times 9.2^6}{6!} + \frac{e^{-9.2} \times 9.2^7}{7!} + \frac{e^{-9.2} \times 9.2^8}{8!} + \frac{e^{-9.2} \times 9.2^9}{9!} + \frac{e^{-9.2} \times 9.2^{10}}{10!}$

$= 0.0851 + 0.1118 + 0.1286 + 0.1315 + 0.1210$

$= 0.5780$

(b) $P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$

$= \frac{e^{-9.2} \times 9.2^0}{0!} + \frac{e^{-9.2} \times 9.2^1}{1!} + \frac{e^{-9.2} \times 9.2^2}{2!} + \frac{e^{-9.2} \times 9.2^3}{3!} + \frac{e^{-9.2} \times 9.2^4}{4!}$

$= 0.0001 + 0.0009 + 0.0043 + 0.0131 + 0.0302$

$= 0.0486$

3.5 The Multinomial Distribution

3.5.1 (a) $\frac{11!}{4!\times 5!\times 2!} \times 0.23^4 \times 0.48^5 \times 0.29^2 = 0.0416$

(b) $P(B(7, 0.23) < 3) = 0.7967$

3.5.2 (a) $\frac{15!}{3!\times 3!\times 9!} \times \left(\frac{1}{6}\right)^3 \times \left(\frac{1}{6}\right)^3 \times \left(\frac{2}{3}\right)^9 = 0.0558$

(b) $\frac{15!}{3!\times 3!\times 4!\times 5!} \times \left(\frac{1}{6}\right)^3 \times \left(\frac{1}{6}\right)^3 \times \left(\frac{1}{6}\right)^4 \times \left(\frac{1}{2}\right)^5 = 0.0065$

(c) $\frac{15!}{2!\times 13!} \times \left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^{13} = 0.2726$

The expected number of sixes is $\frac{15}{6} = 2.5$.

3.5.3 (a) $\frac{8!}{2!\times 5!\times 1!} \times 0.09^2 \times 0.79^5 \times 0.12^1 = 0.0502$

(b) $\frac{8!}{1!\times 5!\times 2!} \times 0.09^1 \times 0.79^5 \times 0.12^2 = 0.0670$

(c) $P(B(8, 0.09) \geq 2) = 0.1577$

The expected number of misses is $8 \times 0.12 = 0.96$.

3.5.4 The expected number of dead seedlings is $22 \times 0.08 = 1.76$
the expected number of slow growth seedlings is $22 \times 0.19 = 4.18$
the expected number of medium growth seedlings is $22 \times 0.42 = 9.24$
and the expected number of strong growth seedlings is $22 \times 0.31 = 6.82$.

(a) $\frac{22!}{3!\times 4!\times 6!\times 9!} \times 0.08^3 \times 0.19^4 \times 0.42^6 \times 0.31^9 = 0.0029$

(b) $\frac{22!}{5!\times 5!\times 5!\times 7!} \times 0.08^5 \times 0.19^5 \times 0.42^5 \times 0.31^7 = 0.00038$

(c) $P(B(22, 0.08) \leq 2) = 0.7442$

3.5.5 The probability that an order is received over the internet and it is large is $0.6 \times 0.3 = 0.18$.

The probability that an order is received over the internet and it is small is $0.6 \times 0.7 = 0.42$.

The probability that an order is not received over the internet and it is large is $0.4 \times 0.4 = 0.16$.

The probability that an order is not received over the internet and it is small is $0.4 \times 0.6 = 0.24$.

The answer is $\frac{8!}{2! \times 2! \times 2! \times 2!} \times 0.18^2 \times 0.42^2 \times 0.16^2 \times 0.24^2 = 0.0212$.

3.7 Supplementary Problems

3.7.1 (a) $P(B(18, 0.085) \geq 3) = 1 - P(B(18, 0.085) \leq 2) = 0.1931$

(b) $P(B(18, 0.085) \leq 1) = 0.5401$

(c) $18 \times 0.085 = 1.53$

3.7.2 $P(B(13, 0.4) \geq 3) = 1 - P(B(13, 0.4) \leq 2) = 0.9421$
The expected number of cells is $13 + (13 \times 0.4) = 18.2$.

3.7.3 (a) $\frac{8!}{2! \times 3! \times 3!} \times 0.40^2 \times 0.25^3 \times 0.35^3 = 0.0600$

(b) $\frac{8!}{3! \times 1! \times 4!} \times 0.40^3 \times 0.25^1 \times 0.35^4 = 0.0672$

(c) $P(B(8, 0.35) \leq 2) = 0.4278$

3.7.4 (a) $P(X = 0) = \frac{e^{-2/3} \times (2/3)^0}{0!} = 0.5134$

(b) $P(X = 1) = \frac{e^{-2/3} \times (2/3)^1}{1!} = 0.3423$

(c) $P(X \geq 3) = 1 - P(X \leq 2) = 0.0302$

3.7.5 $P(X = 2) = \frac{e^{-3.3} \times (3.3)^2}{2!} = 0.2008$

$P(X \geq 6) = 1 - P(X \leq 5) = 0.1171$

3.7.6 (a) Consider a negative binomial distribution with parameters $p = 0.55$ and $r = 4$.

(b) $P(X = 7) = \binom{6}{3} \times (1 - 0.55)^3 \times 0.55^4 = 0.1668$

(c) $P(X = 6) = \binom{5}{3} \times (1 - 0.55)^2 \times 0.55^4 = 0.1853$

(d) The probability that team A wins the series in game 5 is

$\binom{4}{3} \times (1 - 0.55)^1 \times 0.55^4 = 0.1647$.

The probability that team B wins the series in game 5 is

$$\binom{4}{3} \times (1-0.45)^1 \times 0.45^4 = 0.0902.$$

The probability that the series is over after game five is $0.1647+0.0902 = 0.2549$.

(e) The probability that team A wins the series in game 4 is $0.55^4 = 0.0915$.
The probability that team A wins the series is
$0.0915 + 0.1647 + 0.1853 + 0.1668 = 0.6083$.

3.7.7 (a) Consider a negative binomial distribution with parameters $p = 0.58$ and $r = 3$.

$$P(X = 9) = \binom{8}{2} \times (1-0.58)^6 \times 0.58^3 = 0.0300$$

(b) Consider a negative binomial distribution with parameters $p = 0.42$ and $r = 4$.

$$P(X \leq 7) = P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7) = 0.3294$$

3.7.8 $P(\text{two red balls} \mid \text{head}) = \dfrac{\binom{6}{2} \times \binom{5}{1}}{\binom{11}{3}} = \frac{5}{11}$

$P(\text{two red balls} \mid \text{tail}) = \dfrac{\binom{5}{2} \times \binom{6}{1}}{\binom{11}{3}} = \frac{4}{11}$

Therefore,

$P(\text{two red balls}) = (P(\text{head}) \times P(\text{two red balls} \mid \text{head}))$

$+\ (P(\text{tail}) \times P(\text{two red balls} \mid \text{tail}))$

$= \left(0.5 \times \frac{5}{11}\right) + \left(0.5 \times \frac{4}{11}\right) = \frac{9}{22}$

and

$P(\text{head} \mid \text{two red balls}) = \frac{P(\text{head and two red balls})}{P(\text{two red balls})}$

$= \frac{P(\text{head}) \times P(\text{two red balls} \mid \text{head})}{P(\text{two red balls})} = \frac{5}{9}.$

3.7.9 Using the hypergeometric distribution, the answer is

$$P(X=0)+P(X=1)=\frac{\binom{36}{5}\times\binom{4}{0}}{\binom{40}{5}}+\frac{\binom{36}{4}\times\binom{4}{1}}{\binom{40}{5}}=0.9310.$$

For a collection of 4,000,000 items of which 400,000 are defective a $B(5,0.1)$ distribution can be used.

$$P(X=0)+P(X=1)=\binom{5}{0}\times 0.1^0\times 0.9^5+\binom{5}{1}\times 0.1^1\times 0.9^4=0.9185$$

3.7.10 (a) $P\left(B\left(22,\frac{1}{6}\right)=3\right)=\binom{22}{3}\times\left(\frac{1}{6}\right)^3\times\left(\frac{5}{6}\right)^{19}=0.223$

(b) Using a negative binomial distribution with $p=\frac{1}{6}$ and $r=3$ the required probability is

$$P(X=10)=\binom{9}{2}\times\left(\frac{1}{6}\right)^3\times\left(\frac{5}{6}\right)^7=0.047$$

(c) $P(B(11,0.5)\leq 3)$

$$=\binom{11}{0}\times 0.5^0\times 0.5^{11}+\binom{11}{1}\times 0.5^1\times 0.5^{10}+\binom{11}{2}\times 0.5^2\times 0.5^9+\binom{11}{3}\times 0.5^3\times 0.5^8$$

$=0.113$

3.7.11 $\dfrac{\binom{11}{3}\times\binom{8}{3}}{\binom{19}{6}}=0.3406$

3.7.12 (a) True

(b) True

(c) True

(d) True

3.7.13 (a) $P\left(B\left(10,\frac{1}{6}\right)=3\right)=\binom{10}{3}\times\left(\frac{1}{6}\right)^3\times\left(\frac{5}{6}\right)^7=0.155$

(b) Using a negative binomial distribution with $p = \frac{1}{6}$ and $r = 4$ the required probability is

$$P(X = 20) = \binom{19}{3} \times \left(\frac{1}{6}\right)^4 \times \left(\frac{5}{6}\right)^{16} = 0.040$$

(c) Using the multinomial distribution the required probability is

$$\frac{9!}{5!\times 2!\times 2!} \times \left(\frac{2}{3}\right)^5 \times \left(\frac{1}{6}\right)^2 \times \left(\frac{1}{6}\right)^2 = 0.077$$

3.7.14 (a) $P(\text{top quality}) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$

$$= \left(e^{-8.3} \times \frac{8.3^0}{0!}\right) + \left(e^{-8.3} \times \frac{8.3^1}{1!}\right) + \left(e^{-8.3} \times \frac{8.3^2}{2!}\right) + \left(e^{-8.3} \times \frac{8.3^3}{3!}\right) + \left(e^{-8.3} \times \frac{8.3^4}{4!}\right)$$

$= 0.0837$

$P(\text{good quality}) = P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8)$

$$= \left(e^{-8.3} \times \frac{8.3^5}{5!}\right) + \left(e^{-8.3} \times \frac{8.3^6}{6!}\right) + \left(e^{-8.3} \times \frac{8.3^7}{7!}\right) + \left(e^{-8.3} \times \frac{8.3^8}{8!}\right)$$

$= 0.4671$

$P(\text{normal quality}) = P(X = 9) + P(X = 10) + P(X = 11) + P(X = 12)$

$$= \left(e^{-8.3} \times \frac{8.3^9}{9!}\right) + \left(e^{-8.3} \times \frac{8.3^{10}}{10!}\right) + \left(e^{-8.3} \times \frac{8.3^{11}}{11!}\right) + \left(e^{-8.3} \times \frac{8.3^{12}}{12!}\right)$$

$= 0.3699$

$P(\text{bad quality}) = 1 - 0.0837 - 0.4671 - 0.3699 = 0.0793$

Using the multinomial distribution the required probability is

$$\frac{7!}{2!\times 2!\times 2!\times 1!} \times 0.0837^2 \times 0.4671^2 \times 0.3699^2 \times 0.0793 = 0.0104$$

(b) The expectation is $10 \times 0.3699 = 3.699$.

The standard deviation is $\sqrt{10 \times 0.3699 \times (1 - 0.3699)} = 1.527$.

(c) The probability of being either top quality or good quality is $0.0837 + 0.4671 = 0.5508$.

$$P(B(8, 0.5508) \le 3) = \binom{8}{0} \times 0.5508^0 \times 0.4492^8 + \binom{8}{1} \times 0.5508^1 \times 0.4492^7$$

$$+ \binom{8}{2} \times 0.5508^2 \times 0.4492^6 + \binom{8}{3} \times 0.5508^3 \times 0.4492^5$$

$= 0.2589$

Chapter 4

Continuous Probability Distributions

4.1 The Uniform Distribution

4.1.1 (a) $E(X) = \frac{-3+8}{2} = 2.5$

(b) $\sigma = \frac{8-(-3)}{\sqrt{12}} = 3.175$

(c) The upper quartile is 5.25.

(d) $P(0 \leq X \leq 4) = \int_0^4 \frac{1}{11}\ dx = \frac{4}{11}$

4.1.2 (a) $E(X) = \frac{1.43+1.60}{2} = 1.515$

(b) $\sigma = \frac{1.60-1.43}{\sqrt{12}} = 0.0491$

(c) $F(x) = \frac{x-1.43}{1.60-1.43} = \frac{x-1.43}{0.17}$
for $1.43 \leq x \leq 1.60$

(d) $F(1.48) = \frac{1.48-1.43}{0.17} = \frac{0.05}{0.17} = 0.2941$

(e) $F(1.5) = \frac{1.5-1.43}{0.17} = \frac{0.07}{0.17} = 0.412$

The number of batteries with a voltage less than 1.5 Volts has a binomial distribution with parameters $n = 50$ and $p = 0.412$ so that the expected value is
$E(X) = n \times p = 50 \times 0.412 = 20.6$
and the variance is
$\text{Var}(X) = n \times p \times (1-p) = 50 \times 0.412 \times 0.588 = 12.11$.

4.1.3 (a) These four intervals have probabilities 0.30, 0.20, 0.25, and 0.25 respectively, and the expectations and variances are calculated from the binomial distribution.

The expectations are:
$20 \times 0.30 = 6$
$20 \times 0.20 = 4$
$20 \times 0.25 = 5$
$20 \times 0.25 = 5$

The variances are:
$20 \times 0.30 \times 0.70 = 4.2$
$20 \times 0.20 \times 0.80 = 3.2$
$20 \times 0.25 \times 0.75 = 3.75$
$20 \times 0.25 \times 0.75 = 3.75$

(b) Using the multinomial distribution the probability is

$\frac{20!}{5!\times 5!\times 5!\times 5!} \times 0.30^5 \times 0.20^5 \times 0.25^5 \times 0.25^5 = 0.0087.$

4.1.4 (a) $E(X) = \frac{0.0+2.5}{2} = 1.25$

$\mathrm{Var}(X) = \frac{(2.5-0.0)^2}{12} = 0.5208$

(b) The probability that a piece of scrap wood is longer than 1 meter is

$\frac{1.5}{2.5} = 0.6.$

The required probability is

$P(B(25, 0.6) \geq 20) = 0.0294.$

4.1.5 (a) The probability is $\frac{4.184-4.182}{4.185-4.182} = \frac{2}{3}$.

(b) $P(\text{difference} \leq 0.0005 \mid \text{fits in hole}) = \frac{P(4.1835 \leq \text{diameter} \leq 4.1840)}{P(\text{diameter} \leq 4.1840)}$

$= \frac{4.1840-4.1835}{4.1840-4.1820} = \frac{1}{4}$

4.1.6 (a) $P(X \leq 85) = \frac{85-60}{100-60} = \frac{5}{8}$

$$P\left(B\left(6, \tfrac{5}{8}\right) = 3\right) = \binom{6}{3} \times \left(\tfrac{5}{8}\right)^3 \times \left(1 - \tfrac{5}{8}\right)^3 = 0.257$$

(b) $P(X \leq 80) = \frac{80-60}{100-60} = \frac{1}{2}$

$P(80 \leq X \leq 90) = \frac{90-60}{100-60} - \frac{80-60}{100-60} = \frac{1}{4}$

$P(X \geq 90) = 1 - \frac{90-60}{100-60} = \frac{1}{4}$

Using the multinomial distribution the required probability is

$\frac{6!}{2!\times 2!\times 2!} \times \left(\frac{1}{4}\right)^2 \times \left(\frac{1}{2}\right)^2 \times \left(\frac{1}{4}\right)^2 = 0.088.$

(c) The number of employees that need to be tested before 3 are found with a score larger than 90 has a negative binomial distribution with $r = 3$ and $p = \frac{1}{4}$, which has an expectation of $\frac{r}{p} = 12$.

4.2 The Exponential Distribution

4.2.2 (a) $E(X) = \frac{1}{0.1} = 10$

(b) $P(X \geq 10) = 1 - F(10) = 1 - (1 - e^{-0.1\times 10}) = e^{-1} = 0.3679$

(c) $P(X \leq 5) = F(5) = 1 - e^{-0.1\times 5} = 0.3935$

(d) The *additional* waiting time also has an exponential distribution with parameter $\lambda = 0.1$.
The probability that the total waiting time is longer than 15 minutes is the probability that the *additional* waiting time is longer than 10 minutes, which is 0.3679 from part (b).

(e) $E(X) = \frac{0+20}{2} = 10$ as in the previous case.
However, in this case the *additional* waiting time has a $U(0, 15)$ distribution.

4.2.3 (a) $E(X) = \frac{1}{0.2} = 5$

(b) $\sigma = \frac{1}{0.2} = 5$

(c) The median is $\frac{0.693}{0.2} = 3.47$.

(d) $P(X \geq 7) = 1 - F(7) = 1 - (1 - e^{-0.2\times 7}) = e^{-1.4} = 0.2466$

(e) The memoryless property of the exponential distribution implies that the required probability is
$P(X \geq 2) = 1 - F(2) = 1 - (1 - e^{-0.2\times 2}) = e^{-0.4} = 0.6703$.

4.2.4 (a) $P(X \leq 5) = F(5) = 1 - e^{-0.31\times 5} = 0.7878$

(b) Consider a binomial distribution with parameters $n = 12$ and $p = 0.7878$.
The expected value is
$E(X) = n \times p = 12 \times 0.7878 = 9.45$
and the variance is
$\text{Var}(X) = n \times p \times (1 - p) = 12 \times 0.7878 \times 0.2122 = 2.01$.

(c) $P(B(12, 0.7878) \leq 9) = 0.4845$

4.2.5 $F(x) = \int_{-\infty}^{x} \frac{1}{2}\lambda e^{-\lambda(\theta-y)}\, dy = \frac{1}{2}e^{-\lambda(\theta-x)}$

for $-\infty \leq x \leq \theta$, and

$F(x) = \frac{1}{2} + \int_\theta^x \frac{1}{2}\lambda e^{-\lambda(y-\theta)}\, dy = 1 - \frac{1}{2}e^{-\lambda(x-\theta)}$

for $\theta \le x \le \infty$.

(a) $P(X \le 0) = F(0) = \frac{1}{2}e^{-3(2-0)} = 0.0012$

(b) $P(X \ge 1) = 1 - F(1) = 1 - \frac{1}{2}e^{-3(2-1)} = 0.9751$

4.2.6 (a) $E(X) = \frac{1}{2} = 0.5$

(b) $P(X \ge 1) = 1 - F(1) = 1 - (1 - e^{-2\times 1}) = e^{-2} = 0.1353$

(c) A Poisson distribution with parameter $2 \times 3 = 6$.

(d) $P(X \le 4) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$

$= \frac{e^{-6}\times 6^0}{0!} + \frac{e^{-6}\times 6^1}{1!} + \frac{e^{-6}\times 6^2}{2!} + \frac{e^{-6}\times 6^3}{3!} + \frac{e^{-6}\times 6^4}{4!} = 0.2851$

4.2.7 (a) $\lambda = 1.8$

(b) $E(X) = \frac{1}{1.8} = 0.5556$

(c) $P(X \ge 1) = 1 - F(1) = 1 - (1 - e^{-1.8\times 1}) = e^{-1.8} = 0.1653$

(d) A Poisson distribution with parameter $1.8 \times 4 = 7.2$.

(e) $P(X \ge 4) = 1 - P(X = 0) - P(X = 1) - P(X = 2) - P(X = 3)$

$= 1 - \frac{e^{-7.2}\times 7.2^0}{0!} - \frac{e^{-7.2}\times 7.2^1}{1!} - \frac{e^{-7.2}\times 7.2^2}{2!} - \frac{e^{-7.2}\times 7.2^3}{3!} = 0.9281$

4.2.8 (a) Solving
$F(5) = 1 - e^{-\lambda\times 5} = 0.90$
gives $\lambda = 0.4605$.

(b) $F(3) = 1 - e^{-0.4605\times 3} = 0.75$

4.2.9 (a) $P(X \ge 1.5) = e^{-0.8\times 1.5} = 0.301$

(b) The number of arrivals Y has a Poisson distribution with parameter
$0.8 \times 2 = 1.6$
so that the required probability is

$P(Y \ge 3) = 1 - P(Y = 0) - P(Y = 1) - P(Y = 2)$

$$= 1 - \left(e^{-1.6} \times \tfrac{1.6^0}{0!}\right) - \left(e^{-1.6} \times \tfrac{1.6^1}{1!}\right) - \left(e^{-1.6} \times \tfrac{1.6^2}{2!}\right) = 0.217$$

4.2.10 $P(X \leq 1) = 1 - e^{-0.3\times 1} = 0.259$
$P(X \leq 3) = 1 - e^{-0.3\times 3} = 0.593$

Using the multinomial distribution the required probability is

$\frac{10!}{2!\times 4!\times 4!} \times 0.259^2 \times (0.593 - 0.259)^4 \times (1 - 0.593)^4 = 0.072.$

4.2.11 (a) $P(X \leq 6) = 1 - e^{-0.2\times 6} = 0.699$

(b) The number of arrivals Y has a Poisson distribution with parameter
$0.2 \times 10 = 2$
so that the required probability is
$P(Y = 3) = e^{-2} \times \frac{2^3}{3!} = 0.180$

4.2.12 $P(X \geq 150) = e^{-0.0065\times 150} = 0.377$
The number of components Y in the box with lifetimes longer than 150 days has a $B(10, 0.377)$ distribution.

$P(Y \geq 8) = P(Y = 8) + P(Y = 9) + P(Y = 10)$

$= \binom{10}{8} \times 0.377^8 \times 0.623^2 + \binom{10}{9} \times 0.377^9 \times 0.623^1 + \binom{10}{10} \times 0.377^{10} \times 0.623^0$

$= 0.00713 + 0.00096 + 0.00006 = 0.00815$

4.2.13 The number of signals X in a 100 meter stretch has a Poisson distribution with mean $0.022 \times 100 = 2.2$.

$P(X \leq 1) = P(X = 0) + P(X = 1)$

$= \left(e^{-2.2} \times \frac{2.2^0}{0!}\right) + \left(e^{-2.2} \times \frac{2.2^1}{1!}\right)$

$= 0.111 + 0.244 = 0.355$

4.2.14 Since

$F(263) = \frac{50}{90} = 1 - e^{-263\lambda}$

it follows that $\lambda = 0.00308$.

Therefore,
$F(x) = \frac{80}{90} = 1 - e^{-0.00308x}$
gives $x = 732.4$.

4.3 The Gamma Distribution

4.3.1 $\Gamma(5.5) = 4.5 \times 3.5 \times 2.5 \times 1.5 \times 0.5 \times \sqrt{\pi} = 52.34$

4.3.3 (a) $f(3) = 0.2055$
$F(3) = 0.3823$
$F^{-1}(0.5) = 3.5919$

(b) $f(3) = 0.0227$
$F(3) = 0.9931$
$F^{-1}(0.5) = 1.3527$

(c) $f(3) = 0.2592$
$F(3) = 0.6046$
$F^{-1}(0.5) = 2.6229$

In this case

$f(3) = \frac{1.4^4 \times 3^{4-1} \times e^{-1.4 \times 3}}{3!} = 0.2592.$

4.3.4 (a) $E(X) = \frac{5}{0.9} = 5.556$

(b) $\sigma = \frac{\sqrt{5}}{0.9} = 2.485$

(c) From the computer the lower quartile is
$F^{-1}(0.25) = 3.743$
and the upper quartile is
$F^{-1}(0.75) = 6.972.$

(d) From the computer $P(X \geq 6) = 0.3733.$

4.3.5 (a) A gamma distribution with parameters $k = 4$ and $\lambda = 2$.

(b) $E(X) = \frac{4}{2} = 2$

(c) $\sigma = \frac{\sqrt{4}}{2} = 1$

(d) The probability can be calculated as
$P(X \geq 3) = 0.1512$
where the random variable X has a gamma distribution with parameters $k = 4$ and $\lambda = 2$.

The probability can also be calculated as
$P(Y \leq 3) = 0.1512$
where the random variable Y has a Poisson distribution with parameter
$2 \times 3 = 6$
which counts the number of imperfections in a 3 meter length of fiber.

4.3.6 (a) A gamma distribution with parameters $k = 3$ and $\lambda = 1.8$.

(b) $E(X) = \frac{3}{1.8} = 1.667$

(c) $\text{Var}(X) = \frac{3}{1.8^2} = 0.9259$

(d) The probability can be calculated as
$P(X \geq 3) = 0.0948$
where the random variable X has a gamma distribution with parameters
$k = 3$ and $\lambda = 1.8$.

The probability can also be calculated as
$P(Y \leq 2) = 0.0948$
where the random variable Y has a Poisson distribution with parameter
$1.8 \times 3 = 5.4$
which counts the number of arrivals in a 3 hour period.

4.3.7 (a) The expectation is $E(X) = \frac{44}{0.7} = 62.86$

the variance is $\text{Var}(X) = \frac{44}{0.7^2} = 89.80$

and the standard deviation is $\sqrt{89.80} = 9.48$.

(b) $F(60) = 0.3991$

4.4 The Weibull Distribution

4.4.2 (a) $\frac{(-\ln(1-0.5))^{1/4.9}}{0.22} = 4.218$

(b) $\frac{(-\ln(1-0.75))^{1/4.9}}{0.22} = 4.859$

$\frac{(-\ln(1-0.25))^{1/4.9}}{0.22} = 3.525$

(c) $F(x) = 1 - e^{-(0.22x)^{4.9}}$

$P(2 \leq X \leq 7) = F(7) - F(2) = 0.9820$

4.4.3 (a) $\frac{(-\ln(1-0.5))^{1/2.3}}{1.7} = 0.5016$

(b) $\frac{(-\ln(1-0.75))^{1/2.3}}{1.7} = 0.6780$

$\frac{(-\ln(1-0.25))^{1/2.3}}{1.7} = 0.3422$

(c) $F(x) = 1 - e^{-(1.7x)^{2.3}}$

$P(0.5 \leq X \leq 1.5) = F(1.5) - F(0.5) = 0.5023$

4.4.4 (a) $\frac{(-\ln(1-0.5))^{1/3}}{0.5} = 1.77$

(b) $\frac{(-\ln(1-0.01))^{1/3}}{0.5} = 0.43$

(c) $E(X) = \frac{1}{0.5}\, \Gamma\left(1 + \frac{1}{3}\right) = 1.79$

$\mathrm{Var}(X) = \frac{1}{0.5^2}\left\{\Gamma\left(1 + \frac{2}{3}\right) - \Gamma\left(1 + \frac{1}{3}\right)^2\right\} = 0.42$

(d) $P(X \leq 3) = F(3) = 1 - e^{-(0.5\times 3)^3} = 0.9658$

The probability that at least one circuit is working after three hours is $1 - 0.9688^4 = 0.13$.

4.4.5 (a) $\frac{(-\ln(1-0.5))^{1/0.4}}{0.5} = 0.8000$

(b) $\frac{(-\ln(1-0.75))^{1/0.4}}{0.5} = 4.5255$

$\frac{(-\ln(1-0.25))^{1/0.4}}{0.5} = 0.0888$

(c) $\frac{(-\ln(1-0.95))^{1/0.4}}{0.5} = 31.066$

$\frac{(-\ln(1-0.99))^{1/0.4}}{0.5} = 91.022$

(d) $F(x) = 1 - e^{-(0.5x)^{0.4}}$

$P(3 \leq X \leq 5) = F(5) - F(3) = 0.0722$

4.4.6 (a) $\frac{(-\ln(1-0.5))^{1/1.5}}{0.03} = 26.11$

$\frac{(-\ln(1-0.75))^{1/1.5}}{0.03} = 41.44$

$\frac{(-\ln(1-0.99))^{1/1.5}}{0.03} = 92.27$

(b) $F(x) = 1 - e^{-(0.03x)^{1.5}}$

$P(X \geq 30) = 1 - F(30) = 0.4258$

The number of components still operating after 30 minutes has a binomial distribution with parameters $n = 500$ and $p = 0.4258$.
The expected value is
$E(X) = n \times p = 500 \times 0.4258 = 212.9$
and the variance is
$\text{Var}(X) = n \times p \times (1-p) = 500 \times 0.4258 \times 0.5742 = 122.2.$

4.4.7 The probability that a culture has developed within four days is
$F(4) = 1 - e^{-(0.3\times 4)^{0.6}} = 0.672.$

Using the negative binomial distribution, the probability that exactly ten cultures are opened is

$\binom{9}{4} \times (1-0.672)^5 \times 0.672^5 = 0.0656.$

4.4.8 A Weibull distribution can be used with

$F(7) = 1 - e^{-(7\lambda)^a} = \frac{9}{82}$

and

$F(14) = 1 - e^{-(14\lambda)^a} = \frac{24}{82}.$

This gives $a = 1.577$ and $\lambda = 0.0364$ so that the median time is the solution to

$1 - e^{-(0.0364x)^{1.577}} = 0.5$

which is 21.7 days.

4.5 The Beta Distribution

4.5.1 (a) Since

$\int_0^1 A\, x^3(1-x)^2\, dx = 1$

it follows that $A = 60$.

(b) $E(X) = \int_0^1 60\, x^4(1-x)^2\, dx = \frac{4}{7}$

$E(X^2) = \int_0^1 60\, x^5(1-x)^2\, dx = \frac{5}{14}$

Therefore,

$\text{Var}(X) = \frac{5}{14} - \left(\frac{4}{7}\right)^2 = \frac{3}{98}$.

(c) This is a beta distribution with $a = 4$ and $b = 3$.

$E(X) = \frac{4}{4+3} = \frac{4}{7}$

$\text{Var}(X) = \frac{4\times 3}{(4+3)^2\times(4+3+1)} = \frac{3}{98}$

4.5.2 (a) This is a beta distribution with $a = 10$ and $b = 4$.

(b) $A = \frac{\Gamma(10+4)}{\Gamma(10)\times\Gamma(4)} = \frac{13!}{9!\times 3!} = 2860$

(c) $E(X) = \frac{10}{10+4} = \frac{5}{7}$

(d) $\text{Var}(X) = \frac{10\times 4}{(10+4)^2\times(10+4+1)} = \frac{2}{147}$

$\sigma = \sqrt{\frac{2}{147}} = 0.1166$

(e) $F(x) = \int_0^x 2860\, y^9\, (1-y)^3\, dy$

$= 2860\left(\frac{x^{10}}{10} - \frac{3x^{11}}{11} + \frac{x^{12}}{4} - \frac{x^{13}}{13}\right)$

for $0 \leq x \leq 1$

4.5.3 (a) $f(0.5) = 1.9418$
$F(0.5) = 0.6753$
$F^{-1}(0.5) = 0.5406$

(b) $f(0.5) = 0.7398$
$F(0.5) = 0.7823$
$F^{-1}(0.5) = 0.4579$

(c) $f(0.5) = 0.6563$
$F(0.5) = 0.9375$
$F^{-1}(0.5) = 0.3407$

In this case

$f(0.5) = \frac{\Gamma(2+6)}{\Gamma(2)\times\Gamma(6)} \times 0.5^{2-1} \times (1-0.5)^{6-1} = 0.65625.$

4.5.4 (a) $3 \leq y \leq 7$

(b) $E(X) = \frac{2.1}{2.1+2.1} = \frac{1}{2}$

Therefore, $E(Y) = 3 + (4 \times E(X)) = 5.$

$\text{Var}(X) = \frac{2.1\times2.1}{(2.1+2.1)^2\times(2.1+2.1+1)} = 0.0481$

Therefore, $\text{Var}(Y) = 4^2 \times \text{Var}(X) = 0.1923.$

(c) The random variable X has a symmetric beta distribution so
$P(Y \leq 5) = P(X \leq 0.5) = 0.5.$

4.5.5 (a) $E(X) = \frac{7.2}{7.2+2.3} = 0.7579$

$\text{Var}(X) = \frac{7.2\times2.3}{(7.2+2.3)^2\times(7.2+2.3+1)} = 0.0175$

(b) From the computer $P(X \geq 0.9) = 0.1368.$

4.5.6 (a) $E(X) = \frac{8.2}{8.2+11.7} = 0.4121$

(b) $\text{Var}(X) = \frac{8.2\times11.7}{(8.2+11.7)^2\times(8.2+11.7+1)} = 0.0116$

$\sigma = \sqrt{0.0116} = 0.1077$

(c) From the computer $F^{-1}(0.5) = 0.4091.$

4.7 Supplementary Problems

4.7.1 $F(0) = P(\text{winnings} = 0) = \frac{1}{4}$

$F(x) = P(\text{winnings} \leq x) = \frac{1}{4} + \frac{x}{720}$ for $0 \leq x \leq 360$

$F(x) = P(\text{winnings} \leq x) = \frac{\sqrt{x+72540}}{360}$ for $360 \leq x \leq 57060$

$F(x) = 1$ for $57060 \leq x$

4.7.2 (a) Solving

$\frac{0.693}{\lambda} = 1.5$

gives $\lambda = 0.462$.

(b) $P(X \geq 2) = 1 - F(2) = 1 - (1 - e^{-0.462\times 2}) = e^{-0.924} = 0.397$

$P(X \leq 1) = F(1) = 1 - e^{-0.462\times 1} = 0.370$

4.7.3 (a) $E(X) = \frac{1}{0.7} = 1.4286$

(b) $P(X \geq 3) = 1 - F(3) = 1 - (1 - e^{-0.7\times 3}) = e^{-2.1} = 0.1225$

(c) $\frac{0.693}{0.7} = 0.9902$

(d) A Poisson distribution with parameter $0.7 \times 10 = 7$.

(e) $P(X \geq 5) = 1 - P(X = 0) - P(X = 1) - P(X = 2) - P(X = 3) - P(X = 4)$
$= 0.8270$

(f) A gamma distribution with parameters $k = 10$ and $\lambda = 0.7$.

$E(X) = \frac{10}{0.7} = 14.286$

$\text{Var}(X) = \frac{10}{0.7^2} = 20.408$

4.7.4 (a) $E(X) = \frac{1}{5.2} = 0.1923$

(b) $P\left(X \leq \frac{1}{6}\right) = F\left(\frac{1}{6}\right) = 1 - e^{-5.2\times 1/6} = 0.5796$

(c) A gamma distribution with parameters $k = 10$ and $\lambda = 5.2$.

(d) $E(X) = \frac{10}{5.2} = 1.923$

(e) The probability is
$P(X > 5) = 0.4191$
where the random variable X has a Poisson distribution with parameter 5.2.

4.7.5 (a) The total area under the triangle is equal to 1
so the height at the midpoint is $\frac{2}{b-a}$.

(b) $P\left(X \leq \frac{a}{4} + \frac{3b}{4}\right) = P\left(X \leq a + \frac{3(b-a)}{4}\right) = \frac{7}{8}$

(c) $\text{Var}(X) = \frac{(b-a)^2}{24}$

(d) $F(x) = \frac{2(x-a)^2}{(b-a)^2}$

for $a \leq x \leq \frac{a+b}{2}$

and

$F(x) = 1 - \frac{2(b-x)^2}{(b-a)^2}$

for $\frac{a+b}{2} \leq x \leq b$

4.7.6 (a) $\frac{(-\ln(1-0.5))^{1/4}}{0.2} = 4.56$

$\frac{(-\ln(1-0.75))^{1/4}}{0.2} = 5.43$

$\frac{(-\ln(1-0.95))^{1/4}}{0.2} = 6.58$

(b) $E(X) = \frac{1}{0.2}\,\Gamma\left(1 + \frac{1}{4}\right) = 4.53$

$\text{Var}(X) = \frac{1}{0.2^2}\left\{\Gamma\left(1 + \frac{2}{4}\right) - \Gamma\left(1 + \frac{1}{4}\right)^2\right\} = 1.620$

(c) $F(x) = 1 - e^{-(0.2x)^4}$

$P(5 \leq X \leq 6) = F(6) - F(5) = 0.242$

4.7.7 (a) $E(X) = \frac{2.7}{2.7+2.9} = 0.4821$

(b) $\text{Var}(X) = \frac{2.7 \times 2.9}{(2.7+2.9)^2 \times (2.7+2.9+1)} = 0.0378$

$\sigma = \sqrt{0.0378} = 0.1945$

(c) From the computer $P(X \geq 0.5) = 0.4637$.

4.7.8 Let the random variable Y be the starting time of the class in minutes after 10 o'clock, so that $Y \sim U(0,5)$.

If $x \leq 0$, the expected penalty is
$A_1(|x| + E(Y)) = A_1(|x| + 2.5)$.

If $x \geq 5$, the expected penalty is
$A_2(x - E(Y)) = A_2(x - 2.5)$.

If $0 \leq x \leq 5$, the penalty is
$A_1(Y - x)$ for $Y \geq x$ and $A_2(x - Y)$ for $Y \leq x$.

The expected penalty is therefore

$\int_x^5 A_1(y-x)f(y)\ dy + \int_0^x A_2(x-y)f(y)\ dy$

$= \int_x^5 A_1(y-x)\frac{1}{5}\ dy + \int_0^x A_2(x-y)\frac{1}{5}\ dy$

$= \frac{A_1(5-x)^2}{10} + \frac{A_2x^2}{10}$.

The expected penalty is minimized by taking

$x = \frac{5A_1}{A_1+A_2}$.

4.7.9 (a) Solving simultaneously

$F(35) = 1 - e^{-(\lambda\times 35)^a} = 0.25$

and

$F(65) = 1 - e^{-(\lambda\times 65)^a} = 0.75$

gives $\lambda = 0.0175$ and $a = 2.54$.

(b) Solving

$F(x) = 1 - e^{-(0.0175\times x)^{2.54}} = 0.90$

gives x as about 79 days.

4.7.10 For this beta distribution $F(0.5) = 0.0925$ and $F(0.8) = 0.9851$
so that the probability of a solution being too weak is 0.0925
the probability of a solution being satisfactory is $0.9851 - 0.0925 = 0.8926$
and the probability of a solution being too strong is $1 - 0.9851 = 0.0149$.

Using the multinomial distribution, the required answer is

$\frac{10!}{1!\times 8!\times 1!} \times 0.0925 \times 0.8926^8 \times 0.0149 = 0.050$.

4.7.11 (a) The number of visits within a two hour interval has a Poisson distribution with parameter $2 \times 4 = 8$.

$P(X = 10) = e^{-8} \times \frac{8^{10}}{10!} = 0.099$

(b) A gamma distribution with $k = 10$ and $\lambda = 4$.

4.7.12 (a) $\frac{1}{\lambda} = \frac{1}{0.48} = 2.08$ cm

(b) $\frac{10}{\lambda} = \frac{10}{0.48} = 20.83$ cm

(c) $P(X \leq 0.5) = 1 - e^{-0.48 \times 0.5} = 0.213$

(d) $P(8 \leq X \leq 12) = \sum_{i=8}^{12} e^{-0.48 \times 20} \frac{(0.48 \times 20)^i}{i!} = 0.569$

4.7.13 (a) False

(b) True

(c) True

(d) True

4.7.14 Using the multinomial distribution the probability is

$\frac{5!}{2! \times 2! \times 1!} \times \left(\frac{2}{5}\right)^2 \times \left(\frac{2}{5}\right)^2 \times \left(\frac{1}{5}\right)^2 = \frac{96}{625} = 0.154.$

4.7.15 (a) The number of events in the interval has a Poisson distribution with parameter $8 \times 0.5 = 4$.

$P(X = 4) = e^{-4} \times \frac{4^4}{4!} = 0.195$

(b) The probability is obtained from an exponential distribution with $\lambda = 8$ and is

$1 - e^{-8 \times 0.2} = 0.798.$

4.7.16 $P(X \leq 8) = 1 - e^{-(0.09 \times 8)^{2.3}} = 0.375$

$P(8 \leq X \leq 12) = 1 - e^{-(0.09 \times 12)^{2.3}} - 0.375 = 0.322$

$P(X \geq 12) = 1 - 0.375 - 0.322 = 0.303$

Using the multinomial distribution the required probability is

$\frac{10!}{3! \times 4! \times 3!} \times 0.375^3 \times 0.322^4 \times 0.303^3 = 0.066.$

Chapter 5

The Normal Distribution

5.1 Probability Calculations using the Normal Distribution

5.1.1 (a) $\Phi(1.34) = 0.9099$

(b) $1 - \Phi(-0.22) = 0.5871$

(c) $\Phi(0.43) - \Phi(-2.19) = 0.6521$

(d) $\Phi(1.76) - \Phi(0.09) = 0.4249$

(e) $\Phi(0.38) - \Phi(-0.38) = 0.2960$

(f) Solving $\Phi(x) = 0.55$ gives $x = 0.1257$.

(g) Solving $1 - \Phi(x) = 0.72$ gives $x = -0.5828$.

(h) Solving $\Phi(x) - \Phi(-x) = (2 \times \Phi(x)) - 1 = 0.31$ gives $x = 0.3989$.

5.1.2 (a) $\Phi(-0.77) = 0.2206$

(b) $1 - \Phi(0.32) = 0.3745$

(c) $\Phi(-1.59) - \Phi(-3.09) = 0.0549$

(d) $\Phi(1.80) - \Phi(-0.82) = 0.7580$

(e) $1 - (\Phi(0.91) - \Phi(-0.91)) = 0.3628$

(f) Solving $\Phi(x) = 0.23$ gives $x = -0.7388$.

(g) Solving $1 - \Phi(x) = 0.51$ gives $x = -0.0251$.

(h) Solving $1 - (\Phi(x) - \Phi(-x)) = 2 - (2 \times \Phi(x)) = 0.42$ gives $x = 0.8064$.

5.1.3 (a) $P(X \leq 10.34) = \Phi\left(\frac{10.34-10}{\sqrt{2}}\right) = 0.5950$

(b) $P(X \geq 11.98) = 1 - \Phi\left(\frac{11.98-10}{\sqrt{2}}\right) = 0.0807$

(c) $P(7.67 \leq X \leq 9.90) = \Phi\left(\frac{9.90-10}{\sqrt{2}}\right) - \Phi\left(\frac{7.67-10}{\sqrt{2}}\right) = 0.4221$

(d) $P(10.88 \leq X \leq 13.22) = \Phi\left(\frac{13.22-10}{\sqrt{2}}\right) - \Phi\left(\frac{10.88-10}{\sqrt{2}}\right) = 0.2555$

(e) $P(|X - 10| \leq 3) = P(7 \leq X \leq 13)$

$= \Phi\left(\frac{13-10}{\sqrt{2}}\right) - \Phi\left(\frac{7-10}{\sqrt{2}}\right) = 0.9662$

(f) Solving $P(N(10,2) \leq x) = 0.81$ gives $x = 11.2415$.

(g) Solving $P(N(10,2) \geq x) = 0.04$ gives $x = 12.4758$.

(h) Solving $P(|N(10,2) - 10| \geq x) = 0.63$ gives $x = 0.6812$.

5.1.4 (a) $P(X \leq 0) = \Phi\left(\frac{0-(-7)}{\sqrt{14}}\right) = 0.9693$

(b) $P(X \geq -10) = 1 - \Phi\left(\frac{-10-(-7)}{\sqrt{14}}\right) = 0.7887$

(c) $P(-15 \leq X \leq -1) = \Phi\left(\frac{-1-(-7)}{\sqrt{14}}\right) - \Phi\left(\frac{-15-(-7)}{\sqrt{14}}\right) = 0.9293$

(d) $P(-5 \leq X \leq 2) = \Phi\left(\frac{2-(-7)}{\sqrt{14}}\right) - \Phi\left(\frac{-5-(-7)}{\sqrt{14}}\right) = 0.2884$

(e) $P(|X + 7| \geq 8) = 1 - P(-15 \leq X \leq 1)$

$= 1 - \left(\Phi\left(\frac{1-(-7)}{\sqrt{14}}\right) - \Phi\left(\frac{-15-(-7)}{\sqrt{14}}\right)\right)$

$= 0.0326$

(f) Solving $P(N(-7,14) \leq x) = 0.75$ gives $x = -4.4763$.

(g) Solving $P(N(-7,14) \geq x) = 0.27$ gives $x = -4.7071$.

(h) Solving $P(|N(-7,14)+7| \leq x) = 0.44$ gives $x = 2.18064$.

5.1.5 Solving

$P(X \leq 5) = \Phi\left(\frac{5-\mu}{\sigma}\right) = 0.8$

and

$P(X \geq 0) = 1 - \Phi\left(\frac{0-\mu}{\sigma}\right) = 0.6$

gives $\mu = 1.1569$ and $\sigma = 4.5663$.

5.1.6 Solving

$P(X \leq 10) = \Phi\left(\frac{10-\mu}{\sigma}\right) = 0.55$

and

$P(X \leq 0) = \Phi\left(\frac{0-\mu}{\sigma}\right) = 0.4$

gives $\mu = 6.6845$ and $\sigma = 26.3845$.

5.1.7 $P(X \leq \mu + \sigma z_\alpha) = \Phi\left(\frac{\mu+\sigma z_\alpha-\mu}{\sigma}\right)$

$= \Phi(z_\alpha) = 1 - \alpha$

$P(\mu - \sigma z_{\alpha/2} \leq X \leq \mu + \sigma z_{\alpha/2}) = \Phi\left(\frac{\mu+\sigma z_{\alpha/2}-\mu}{\sigma}\right) - \Phi\left(\frac{\mu-\sigma z_{\alpha/2}-\mu}{\sigma}\right)$

$= \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2})$

$= 1 - \alpha/2 - \alpha/2 = 1 - \alpha$

5.1.8 Solving $\Phi(x) = 0.75$ gives $x = 0.6745$.
Solving $\Phi(x) = 0.25$ gives $x = -0.6745$.
The interquartile range of a $N(0,1)$ distribution is therefore
$0.6745 - (-0.6745) = 1.3490$.

The interquartile range of a $N(\mu, \sigma^2)$ distribution is $1.3490 \times \sigma$.

5.1.9 (a) $P(N(3.00, 0.12^2) \geq 3.2) = 0.0478$

(b) $P(N(3.00, 0.12^2) \leq 2.7) = 0.0062$

(c) Solving
$P(3.00 - c \leq N(3.00, 0.12^2) \leq 3.00 + c) = 0.99$
gives
$c = 0.12 \times z_{0.005} = 0.12 \times 2.5758 = 0.3091$.

5.1.10 (a) $P(N(1.03, 0.014^2) \leq 1) = 0.0161$

(b) $P(N(1.05, 0.016^2) \leq 1) = 0.0009$
There is a decrease in the proportion of underweight packets.

(c) The expected excess weight is $\mu - 1$ which is 0.03 and 0.05.

5.1.11 (a) Solving $P(N(4.3, 0.12^2) \leq x) = 0.75$ gives $x = 4.3809$.
Solving $P(N(4.3, 0.12^2) \leq x) = 0.25$ gives $x = 4.2191$.

(b) Solving
$P(4.3 - c \leq N(4.3, 0.12^2) \leq 4.3 + c) = 0.80$
gives
$c = 0.12 \times z_{0.10} = 0.12 \times 1.2816 = 0.1538$.

5.1.12 (a) $P(N(0.0046, 9.6 \times 10^{-8}) \leq 0.005) = 0.9017$

(b) $P(0.004 \leq N(0.0046, 9.6 \times 10^{-8}) \leq 0.005) = 0.8753$

(c) Solving $P(N(0.0046, 9.6 \times 10^{-8}) \leq x) = 0.10$ gives $x = 0.0042$.

(d) Solving $P(N(0.0046, 9.6 \times 10^{-8}) \leq x) = 0.99$ gives $x = 0.0053$.

5.1.13 (a) $P(N(23.8, 1.28) \leq 23.0) = 0.2398$

(b) $P(N(23.8, 1.28) \geq 24.0) = 0.4298$

(c) $P(24.2 \leq N(23.8, 1.28) \leq 24.5) = 0.0937$

(d) Solving $P(N(23.8, 1.28) \leq x) = 0.75$ gives $x = 24.56$.

(e) Solving $P(N(23.8, 1.28) \leq x) = 0.95$ gives $x = 25.66$.

5.1.14 Solving

$P(N(\mu, 0.05^2) \leq 10) = 0.01$

gives

$\mu = 10 + (0.05 \times z_{0.01}) = 10 + (0.05 \times 2.3263) = 10.1163.$

5.1.15 (a) $P(2599 \leq X \leq 2601) = \Phi\left(\frac{2601-2600}{0.6}\right) - \Phi\left(\frac{2599-2600}{0.6}\right)$

$= 0.9522 - 0.0478 = 0.9044$

The probability of being outside the range is $1 - 0.9044 = 0.0956$.

(b) It is required that

$P(2599 \leq X \leq 2601) = \Phi\left(\frac{2601-2600}{\sigma}\right) - \left(\frac{2599-2600}{\sigma}\right)$

$= 1 - 0.001 = 0.999.$

Consequently,

$\Phi\left(\frac{1}{\sigma}\right) - \Phi\left(\frac{-1}{\sigma}\right)$

$= 2\Phi\left(\frac{1}{\sigma}\right) - 1 = 0.999$

so that

$\Phi\left(\frac{1}{\sigma}\right) = 0.9995$

which gives

$\frac{1}{\sigma} = \Phi^{-1}(0.9995) = 3.2905$

with

$\sigma = 0.304.$

5.1.16 $P(N(1320, 15^2) \leq 1300) = P\left(N(0,1) \leq \frac{1300-1320}{15}\right)$

$= \Phi(-1.333) = 0.0912$

$P(N(1320, 15^2) \leq 1330) = P\left(N(0,1) \leq \frac{1330-1320}{15}\right)$

$= \Phi(0.667) = 0.7475$

Using the multinomial distribution the required probability is

$\frac{10!}{3! \times 4! \times 3!} \times 0.0912^3 \times (0.7475 - 0.0912)^4 \times (1 - 0.7475)^3 = 0.0095.$

5.1.17 $0.95 = P(N(\mu, 4.2^2) \leq 100) = P\left(N(0,1) \leq \frac{100-\mu}{4.2}\right)$

Therefore,

$\frac{100-\mu}{4.2} = z_{0.05} = 1.645$

so that $\mu = 93.09$.

5.2 Linear Combinations of Normal Random Variables

5.2.1 (a) $P(N(3.2 + (-2.1), 6.5 + 3.5) \geq 0) = 0.6360$

(b) $P(N(3.2 + (-2.1) - (2 \times 12.0), 6.5 + 3.5 + (2^2 \times 7.5)) \leq -20) = 0.6767$

(c) $P(N((3 \times 3.2) + (5 \times (-2.1)), (3^2 \times 6.5) + (5^2 \times 3.5)) \geq 1) = 0.4375$

(d) The mean is $(4 \times 3.2) - (4 \times (-2.1)) + (2 \times 12.0) = 45.2$.
The variance is $(4^2 \times 6.5) + (4^2 \times 3.5) + (2^2 \times 7.5) = 190$.
$P(N(45.2, 190) \leq 25) = 0.0714$

(e) $P(\mid N(3.2 + (6 \times (-2.1)) + 12.0, 6.5 + (6^2 \times 3.5) + 7.5) \mid \geq 2) = 0.8689$

(f) $P(\mid N((2 \times 3.2) - (-2.1) - 6, (2^2 \times 6.5) + 3.5) \mid \leq 1) = 0.1315$

5.2.2 (a) $P(N(-1.9 - 3.3, 2.2 + 1.7) \geq -3) = 0.1326$

(b) The mean is $(2 \times (-1.9)) + (3 \times 3.3) + (4 \times 0.8) = 9.3$.
The variance is $(2^2 \times 2.2) + (3^2 \times 1.7) + (4^2 \times 0.2) = 27.3$.
$P(N(9.3, 27.3) \leq 10) = 0.5533$

(c) $P(N((3 \times 3.3) - 0.8, (3^2 \times 1.7) + 0.2) \leq 8) = 0.3900$

(d) The mean is $(2 \times (-1.9)) - (2 \times 3.3) + (3 \times 0.8) = -8.0$.
The variance is $(2^2 \times 2.2) + (2^2 \times 1.7) + (3^2 \times 0.2) = 17.4$.
$P(N(-8.0, 17.4) \leq -6) = 0.6842$

(e) $P(\mid N(-1.9 + 3.3 - 0.8, 2.2 + 1.7 + 0.2) \mid \geq 1.5) = 0.4781$

(f) $P(\mid N((4 \times (-1.9)) - 3.3 + 10, (4^2 \times 2.2) + 1.7) \mid \leq 0.5) = 0.0648$

5.2.3 (a) $\Phi(0.5) - \Phi(-0.5) = 0.3830$

(b) $P\left(\mid N\left(0, \frac{1}{8}\right) \mid \leq 0.5\right) = 0.8428$

(c) It is required that

$0.5\sqrt{n} \geq z_{0.005} = 2.5758$

which is satisfied for $n \geq 27$.

5.2.4 (a) $N(4.3 + 4.3, 0.12^2 + 0.12^2) = N(8.6, 0.0288)$

(b) $N\left(4.3, \frac{0.12^2}{12}\right) = N(4.3, 0.0012)$

(c) It is required that

$z_{0.0015} \times \frac{0.12}{\sqrt{n}} = 2.9677 \times \frac{0.12}{\sqrt{n}} \leq 0.05$

which is satisfied for $n \geq 51$.

5.2.5 $P(144 \leq N(37 + 37 + 24 + 24 + 24, 0.49 + 0.49 + 0.09 + 0.09 + 0.09) \leq 147) = 0.7777$

5.2.6 (a) $\text{Var}(Y) = (p^2 \times \sigma_1^2) + ((1-p)^2 \times \sigma_2^2)$

The minimum variance is

$\frac{1}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}} = \frac{\sigma_1^2\,\sigma_2^2}{\sigma_1^2+\sigma_2^2}$.

(b) In this case

$\text{Var}(Y) = \sum_{i=1}^n\ p_i^2\ \sigma_i^2$.

The variance is minimized with

$p_i = \frac{\frac{1}{\sigma_i^2}}{\frac{1}{\sigma_1^2}+\ldots+\frac{1}{\sigma_n^2}}$

and the minimum variance is

$\frac{1}{\frac{1}{\sigma_1^2}+\ldots+\frac{1}{\sigma_n^2}}$.

5.2.7 (a) $1.05y + 1.05(1000 - y) = \$1050$

(b) $0.0002y^2 + 0.0003(1000 - y)^2$

(c) The variance is minimized with $y = 600$ and the minimum variance is 120.

$P(N(1050, 120) \geq 1060) = 0.1807$

5.2.8 (a) $P(N(3.00 + 3.00 + 3.00, 0.12^2 + 0.12^2 + 0.12^2) \geq 9.50) = 0.0081$

(b) $P\left(N\left(3.00, \frac{0.12^2}{7}\right) \leq 3.10\right) = 0.9863$

5.2.9 (a) $N(22 \times 1.03, 22 \times 0.014^2) = N(22.66, 4.312 \times 10^{-3})$

(b) Solving $P(N(22.66, 4.312 \times 10^{-3}) \leq x) = 0.75$ gives $x = 22.704$.
Solving $P(N(22.66, 4.312 \times 10^{-3}) \leq x) = 0.25$ gives $x = 22.616$.

5.2.10 (a) Let the random variables X_i be the widths of the components.
Then

$P(X_1 + X_2 + X_3 + X_4 \leq 10402.5) = P(N(4 \times 2600, 4 \times 0.6^2) \leq 10402.5)$

$= \Phi\left(\frac{10402.5-10400}{1.2}\right) = \Phi(2.083) = 0.9814$.

(b) Let the random variable Y be the width of the slot.
Then

$P(X_1 + X_2 + X_3 + X_4 - Y \leq 0)$

$= P(N((4 \times 2600) - 10402.5, (4 \times 0.6^2) + 0.4^2) \leq 0)$

$= \Phi\left(\frac{2.5}{1.2649}\right) = \Phi(1.976) = 0.9759$.

5.2.11 (a) $P\left(4.2 \leq N\left(4.5, \frac{0.88}{15}\right) \leq 4.9\right)$

$= P\left(\frac{\sqrt{15}(4.2-4.5)}{\sqrt{0.88}} \leq N(0,1) \leq \frac{\sqrt{15}(4.9-4.5)}{\sqrt{0.88}}\right)$

$= \Phi(1.651) - \Phi(-1.239)$

$= 0.951 - 0.108 = 0.843$

(b) $0.99 = P\left(4.5 - c \leq N\left(4.5, \frac{0.88}{15}\right) \leq 4.5 + c\right)$

$= P\left(\frac{-c\sqrt{15}}{\sqrt{0.88}} \leq N(0,1) \leq \frac{c\sqrt{15}}{\sqrt{0.88}}\right)$

Therefore,

$\frac{c\sqrt{15}}{\sqrt{0.88}} = z_{0.005} = 2.576$

so that $c = 0.624$.

5.2.12 (a) $P(X_1 + X_2 + X_3 + X_4 + X_5 \geq 45)$

$= P(N(8 + 8 + 8 + 8 + 8, 2^2 + 2^2 + 2^2 + 2^2 + 2^2) \geq 45)$

$= P\left(N(0,1) \geq \frac{45-40}{\sqrt{20}}\right)$

$= 1 - \Phi(1.118) = 0.132$

(b) $P(N(28, 5^2) \geq N(8 + 8 + 8, 2^2 + 2^2 + 2^2))$

$= P(N(28 - 24, 25 + 12) \geq 0)$

$= P\left(N(0,1) \geq \frac{-4}{\sqrt{37}}\right)$

$= 1 - \Phi(-0.658) = 0.745$

5.2.13 The height of a stack of 4 components of type A has a normal distribution with mean $4 \times 190 = 760$ and a standard deviation $\sqrt{4} \times 10 = 20$.

The height of a stack of 5 components of type B has a normal distribution with mean $5 \times 150 = 750$ and a standard deviation $\sqrt{5} \times 8 = 17.89$.

$P(N(760, 20^2) > N(750, 17.89^2))$

$= P(N(760 - 750, 20^2 + 17.78^2) > 0)$

$= P\left(N(0,1) > \frac{-10}{\sqrt{720}}\right)$

$= 1 - \Phi(-0.373) = 0.645$

5.2.14 Let the random variables X_i be the times taken by worker 1 to perform a task and let the random variables Y_i be the times taken by worker 2 to perform a task.

$P(X_1 + X_2 + X_3 + X_4 - Y_1 - Y_2 - Y_3 \leq 0)$

$= P(N(13+13+13+13-17-17-17, 0.5^2+0.5^2+0.5^2+0.5^2+0.6^2+0.6^2+0.6^2) \leq 0)$

$= P(N(1, 2.08) \leq 0)$

$= P\left(N(0,1) \leq \frac{-1}{\sqrt{2.08}}\right)$

$= \Phi(-0.693) = 0.244$

5.2.15 It is required that

$P\left(N\left(110, \frac{4}{n}\right) \leq 111\right)$

$= P\left(N(0,1) \leq \frac{\sqrt{n}(111-110)}{2}\right) \geq 0.99.$

Therefore,

$\frac{\sqrt{n}(111-110)}{2} \geq z_{0.01} = 2.326$

which is satisfied for $n \geq 22$.

5.2.16 If X has mean of 7.2 m and a standard deviation of 0.11 m, then $\frac{X}{2}$ has a mean of $\frac{7.2}{2} = 3.6$ m and a standard deviation of $\frac{0.11}{2} = 0.055$ m.

5.2.17 (a) $E(X) = 20\mu = 20 \times 63400 = 1268000$
The standard deviation is $\sqrt{20}\ \sigma = \sqrt{20} \times 2500 = 11180$.

(b) $E(X) = \mu = 63400$
The standard deviation is $\frac{\sigma}{\sqrt{30}} = \frac{2500}{\sqrt{30}} = 456.4$.

5.2.18 (a) $P(X < 800) = \Phi\left(\frac{800-T}{47}\right) = 0.1$

so that

$\frac{800-T}{47} = -z_{0.1} = -1.282.$

This gives $T = 860.3$.

(b) The average Y is distributed as a $N\left(850, \frac{47^2}{10}\right)$ random variable.

Therefore,

$P(Y < 875) = \Phi\left(\frac{875-850}{47/\sqrt{10}}\right) = 0.954.$

5.3 Approximating Distributions with the Normal Distribution

5.3.1 (a) The exact probability is 0.3823.

The normal approximation is

$1 - \Phi\left(\frac{8-0.5-(10\times0.7)}{\sqrt{10\times0.7\times0.3}}\right) = 0.3650.$

(b) The exact probability is 0.9147.

The normal approximation is

$\Phi\left(\frac{7+0.5-(15\times0.3)}{\sqrt{15\times0.3\times0.7}}\right) - \Phi\left(\frac{1+0.5-(15\times0.3)}{\sqrt{15\times0.3\times0.7}}\right) = 0.9090.$

(c) The exact probability is 0.7334.

The normal approximation is

$\Phi\left(\frac{4+0.5-(9\times0.4)}{\sqrt{9\times0.4\times0.6}}\right) = 0.7299.$

(d) The exact probability is 0.6527.

The normal approximation is

$\Phi\left(\frac{11+0.5-(14\times0.6)}{\sqrt{14\times0.6\times0.4}}\right) - \Phi\left(\frac{7+0.5-(14\times0.6)}{\sqrt{14\times0.6\times0.4}}\right) = 0.6429.$

5.3.2 (a) The exact probability is 0.0106.

The normal approximation is

$1 - \Phi\left(\frac{7-0.5-(10\times0.3)}{\sqrt{10\times0.3\times0.7}}\right) = 0.0079.$

(b) The exact probability is 0.6160.

The normal approximation is

$\Phi\left(\frac{12+0.5-(21\times0.5)}{\sqrt{21\times0.5\times0.5}}\right) - \Phi\left(\frac{8+0.5-(21\times0.5)}{\sqrt{21\times0.5\times0.5}}\right) = 0.6172.$

(c) The exact probability is 0.9667.

The normal approximation is

$\Phi\left(\frac{3+0.5-(7\times0.2)}{\sqrt{7\times0.2\times0.8}}\right) = 0.9764.$

(d) The exact probability is 0.3410.

The normal approximation is

$\Phi\left(\frac{11+0.5-(12\times0.65)}{\sqrt{12\times0.65\times0.35}}\right) - \Phi\left(\frac{8+0.5-(12\times0.65)}{\sqrt{12\times0.65\times0.35}}\right) = 0.3233.$

5.3.3 The required probability is

$$\Phi\left(0.02\sqrt{n}+\frac{1}{\sqrt{n}}\right)-\Phi\left(-0.02\sqrt{n}-\frac{1}{\sqrt{n}}\right)$$

which is equal to

0.2358 for $n = 100$

0.2764 for $n = 200$

0.3772 for $n = 500$

0.4934 for $n = 1000$

and 0.6408 for $n = 2000$.

5.3.4 (a) $\Phi\left(\frac{180+0.5-(1{,}000\times 1/6)}{\sqrt{1{,}000\times 1/6\times 5/6}}\right)-\Phi\left(\frac{149+0.5-(1{,}000\times 1/6)}{\sqrt{1{,}000\times 1/6\times 5/6}}\right)=0.8072$

(b) It is required that

$$1-\Phi\left(\frac{50-0.5-n/6}{\sqrt{n\times 1/6\times 5/6}}\right)\geq 0.99$$

which is satisfied for $n \geq 402$.

5.3.5 (a) A normal distribution can be used with

$\mu = 500 \times 2.4 = 1200$

and

$\sigma^2 = 500 \times 2.4 = 1200$.

(b) $P(N(1200, 1200) \geq 1250) = 0.0745$

5.3.6 The normal approximation is

$$1-\Phi\left(\frac{135-0.5-(15{,}000\times 1/125)}{\sqrt{15{,}000\times 1/125\times 124/125}}\right)=0.0919.$$

5.3.7 The normal approximation is

$$\Phi\left(\frac{200+0.5-(250{,}000\times 0.0007)}{\sqrt{250{,}000\times 0.0007\times 0.9993}}\right)=0.9731.$$

5.3.8 (a) The normal approximation is

$$1-\Phi\left(\frac{30-0.5-(60\times 0.25)}{\sqrt{60\times 0.25\times 0.75}}\right)\simeq 0.$$

(b) It is required that $P(B(n, 0.25) \leq 0.35n) \geq 0.99$.

Using the normal approximation this can be written

$\Phi\left(\frac{0.35n+0.5-0.25n}{\sqrt{n\times 0.25\times 0.75}}\right) \geq 0.99$

which is satisfied for $n \geq 92$.

5.3.9 The yearly income can be approximated by a normal distribution with

$\mu = 365 \times \frac{5}{0.9} = 2027.8$

and

$\sigma^2 = 365 \times \frac{5}{0.9^2} = 2253.1$.

$P(N(2027.8, 2253.1) \geq 2000) = 0.7210$

5.3.10 The normal approximation is

$P(N(1500 \times 0.6, 1500 \times 0.6 \times 0.4) \geq 925 - 0.5)$

$= 1 - \Phi(1.291) = 0.0983$.

5.3.11 The expectation of the strength of a chemical solution is

$E(X) = \frac{18}{18+11} = 0.6207$

and the variance is

$\mathrm{Var}(X) = \frac{18\times 11}{(18+11)^2(18+11+1)} = 0.007848$.

Using the central limit theorem the required probability can be estimated as

$P\left(0.60 \leq N\left(0.6207, \frac{0.007848}{20}\right) \leq 0.65\right)$

$= \Phi(1.479) - \Phi(-1.045) = 0.7824$.

5.3.12 $P(B(1550, 0.135) \geq 241)$

$\simeq P(N(1550 \times 0.135, 1550 \times 0.135 \times 0.865) \geq 240.5)$

$= P\left(N(0,1) \geq \frac{240.5-209.25}{\sqrt{181.00}}\right)$

$= 1 - \Phi(2.323) = 0.010$

5.3.13 $P(60 \leq X \leq 100) = (1 - e^{-100/84}) - (1 - e^{-60/84}) = 0.1855$

$P(B(350, 0.1855) \geq 55)$

$\simeq P(N(350 \times 0.1855, 350 \times 0.1855 \times 0.8145) \geq 54.5)$

$= P\left(N(0,1) \geq \frac{54.5-64.925}{7.272}\right)$

$= 1 - \Phi(-1.434) = 0.92$

5.3.14 $P(X \geq 20) = e^{-(0.056 \times 20)^{2.5}} = 0.265$

$P(B(500, 0.265) \geq 125)$

$\simeq P(N(500 \times 0.265, 500 \times 0.265 \times 0.735) \geq 124.5)$

$= P\left(N(0,1) \geq \frac{124.5-132.57}{9.871}\right)$

$= 1 - \Phi(-0.818) = 0.79$

5.3.15 (a) $P(X \geq 891.2) = \frac{892-891.2}{892-890} = 0.4$

Using the negative binomial distribution the required probability is

$\binom{5}{2} \times 0.4^3 \times 0.6^3 = 0.138.$

(b) $P(X \geq 890.7) = \frac{892-890.7}{892-890} = 0.65$

$P(B(200, 0.65) \geq 100)$

$\simeq P(N(200 \times 0.65, 200 \times 0.65 \times 0.35) \geq 99.5)$

$= P\left(N(0,1) \geq \frac{99.5-130}{\sqrt{45.5}}\right)$

$= 1 - \Phi(-4.52) \simeq 1$

5.3.16 $P(\text{spoil within 10 days}) = 1 - e^{10/8} = 0.713$

The number of packets X with spoiled food has a binomial distribution with $n = 100$ and $p = 0.713$,
so that the expectation is $100 \times 0.713 = 71.3$
and the standard deviation is $\sqrt{100 \times 0.713 \times 0.287} = 4.52$.

$P(X \geq 75) \simeq P(N(71.3, 4.52^2) \geq 74.5)$

$= 1 - \Phi\left(\frac{74.5-71.3}{4.52}\right) = 1 - \Phi(0.71) = 0.24$

5.4 Distributions Related to the Normal Distribution

5.4.1 (a) $E(X) = e^{1.2+(1.5^2/2)} = 10.23$

(b) $\mathrm{Var}(X) = e^{(2\times 1.2)+1.5^2} \times (e^{1.5^2} - 1) = 887.69$

(c) Since $z_{0.25} = 0.6745$ the upper quartile is
$e^{1.2+(1.5\times 0.6745)} = 9.13.$

(d) The lower quartile is
$e^{1.2+(1.5\times(-0.6745))} = 1.21.$

(e) The interquartile range is $9.13 - 1.21 = 7.92$.

(f) $P(5 \leq X \leq 8) = \Phi\left(\frac{\ln(8)-1.2}{1.5}\right) - \Phi\left(\frac{\ln(5)-1.2}{1.5}\right) = 0.1136.$

5.4.2 (a) $E(X) = e^{-0.3+(1.1^2/2)} = 1.357$

(b) $\mathrm{Var}(X) = e^{(2\times(-0.3))+1.1^2} \times (e^{1.1^2} - 1) = 4.331$

(c) Since $z_{0.25} = 0.6745$ the upper quartile is
$e^{-0.3+(1.1\times 0.6745)} = 1.556.$

(d) The lower quartile is
$e^{-0.3+(1.1\times(-0.6745))} = 0.353.$

(e) The interquartile range is $1.556 - 0.353 = 1.203$.

(f) $P(0.1 \leq X \leq 7) = \Phi\left(\frac{\ln(7)-(-0.3)}{1.1}\right) - \Phi\left(\frac{\ln(0.1)-(-0.3)}{1.1}\right) = 0.9451.$

5.4.4 (a) $E(X) = e^{2.3+(0.2^2/2)} = 10.18$

(b) The median is $e^{2.3} = 9.974$.

(c) Since $z_{0.25} = 0.6745$ the upper quartile is
$e^{2.3+(0.2\times 0.6745)} = 11.41.$

(d) $P(X \geq 15) = 1 - \Phi\left(\frac{\ln(15)-2.3}{0.2}\right) = 0.0207$

(e) $P(X \leq 6) = \Phi\left(\frac{\ln(6)-2.3}{0.2}\right) = 0.0055$

5.4.5 (a) $\chi^2_{0.10,9} = 14.68$

(b) $\chi^2_{0.05,20} = 31.41$

(c) $\chi^2_{0.01,26} = 45.64$

(d) $\chi^2_{0.90,50} = 39.69$

(e) $\chi^2_{0.95,6} = 1.635$

5.4.6 (a) $\chi^2_{0.12,8} = 12.77$

(b) $\chi^2_{0.54,19} = 17.74$

(c) $\chi^2_{0.023,32} = 49.86$

(d) $P(X \leq 13.3) = 0.6524$

(e) $P(9.6 \leq X \leq 15.3) = 0.4256$

5.4.7 (a) $t_{0.10,7} = 1.415$

(b) $t_{0.05,19} = 1.729$

(c) $t_{0.01,12} = 2.681$

(d) $t_{0.025,30} = 2.042$

(e) $t_{0.005,4} = 4.604$

5.4.8 (a) $t_{0.27,14} = 0.6282$

(b) $t_{0.09,22} = 1.385$

(c) $t_{0.016,7} = 2.670$

(d) $P(X \leq 1.78) = 0.9556$

(e) $P(-0.65 \leq X \leq 2.98) = 0.7353$

(f) $P(|X| \geq 3.02) = 0.0062$

5.4.9 (a) $F_{0.10,9,10} = 2.347$

(b) $F_{0.05,6,20} = 2.599$

(c) $F_{0.01,15,30} = 2.700$

(d) $F_{0.05,4,8} = 3.838$

(e) $F_{0.01,20,13} = 3.665$

5.4.10 (a) $F_{0.04,7,37} = 2.393$

(b) $F_{0.87,17,43} = 0.6040$

(c) $F_{0.035,3,8} = 4.732$

(d) $P(X \geq 2.35) = 0.0625$

(e) $P(0.21 \leq X \leq 2.92) = 0.9286$

5.4.11 This follows from the definitions

$t_\nu \sim \frac{N(0,1)}{\sqrt{\chi^2_\nu/\nu}}$

and

$F_{1,\nu} \sim \frac{\chi^2_1}{\chi^2_\nu/\nu}$.

5.4.12 (a) $x = t_{0.05,23} = 1.714$

(b) $y = -t_{0.025,60} = -2.000$

(c) $\chi^2_{0.90,29} = 19.768$ and $\chi^2_{0.05,29} = 42.557$
so
$P(19.768 \leq \chi^2_{29} \leq 42.557) = 0.95 - 0.10 = 0.85$

5.4.13 $P(F_{5,20} \geq 4.00) = 0.011$

5.4.14 $P(t_{35} \geq 2.50) = 0.009$

5.4.15 (a) $P(F_{10,50} \geq 2.5) = 0.016$

(b) $P(\chi^2_{17} \leq 12) = 0.200$

(c) $P(t_{24} \geq 3) = 0.003$

(d) $P(t_{14} \geq -2) = 0.967$

5.4.16 (a) $P(t_{21} \leq 2.3) = 0.984$

(b) $P(\chi^2_6 \geq 13.0) = 0.043$

(c) $P(t_{10} \leq -1.9) = 0.043$

(d) $P(t_7 \geq -2.7) = 0.985$

5.4.17 (a) $P(t_{16} \leq 1.9) = 0.962$

(b) $P(\chi^2_{25} \geq 42.1) = 0.018$

(c) $P(F_{9,14} \leq 1.8) = 0.844$

(d) $P(-1.4 \leq t_{29} \leq 3.4) = 0.913$

5.6 Supplementary Problems

5.6.1 (a) $P(N(500, 50^2) \geq 625) = 0.0062$

(b) Solving $P(N(500, 50^2) \leq x) = 0.99$ gives $x = 616.3$.

(c) $P(N(500, 50^2) \geq 700) \simeq 0$
There is a strong suggestion that an eruption is imminent.

5.6.2 (a) $P(N(12500, 200000) \geq 13000) = 0.1318$

(b) $P(N(12500, 200000) \leq 11400) = 0.0070$

(c) $P(12200 \leq N(12500, 200000) \leq 14000) = 0.7484$

(d) Solving $P(N(12500, 200000) \leq x) = 0.95$ gives $x = 13200$.

5.6.3 (a) $P(N(70, 5.4^2) \geq 80) = 0.0320$

(b) $P(N(70, 5.4^2) \leq 55) = 0.0027$

(c) $P(65 \leq N(70, 5.4^2) \leq 78) = 0.7536$

(d) $c = \sigma \times z_{0.025} = 5.4 \times 1.9600 = 10.584$

5.6.4 (a) $P(X_1 - X_2 \geq 0) = P(N(0, 2 \times 5.4^2) \geq 0) = 0.5$

(b) $P(X_1 - X_2 \geq 10) = P(N(0, 2 \times 5.4^2) \geq 10) = 0.0952$

(c) $P\left(\frac{X_1+X_2}{2} - X_3 \geq 10\right) = P(N(0, 1.5 \times 5.4^2) \geq 10) = 0.0653$

5.6.5 $P(\mid X_1 - X_2 \mid \leq 3)$
$= P(\mid N(0, 2 \times 2^2) \mid \leq 3)$
$= P(-3 \leq N(0, 8) \leq 3) = 0.7112$

5.6.6 $E(X) = \frac{1.43+1.60}{2} = 1.515$

$\mathrm{Var}(X) = \frac{(1.60-1.43)^2}{12} = 0.002408$

Therefore, the required probability can be estimated as

$P(180 \leq N(120 \times 1.515, 120 \times 0.002408) \leq 182) = 0.6447.$

5.6.7 $E(X) = \frac{1}{0.31} = 3.2258$

$\text{Var}(X) = \frac{1}{0.31^2} = 10.406$

Therefore, the required probability can be estimated as

$P\left(3.10 \leq N\left(3.2258, \frac{10.406}{2000}\right) \leq 3.25\right) = 0.5908.$

5.6.8 The required probability is $P(B(350000, 0.06) \geq 20,800)$.

The normal approximation is

$1 - \Phi\left(\frac{20800-0.5-(350000\times0.06)}{\sqrt{350000\times0.06\times0.94}}\right) = 0.9232.$

5.6.9 (a) The median is $e^{5.5} = 244.7$.

Since $z_{0.25} = 0.6745$ the upper quartile is

$e^{5.5+(2.0\times0.6745)} = 942.9.$

The lower quartile is

$e^{5.5-(2.0\times0.6745)} = 63.50.$

(b) $P(X \geq 75000) = 1 - \Phi\left(\frac{\ln(75000)-5.5}{2.0}\right) = 0.0021$

(c) $P(X \leq 1000) = \Phi\left(\frac{\ln(1000)-5.5}{2.0}\right) = 0.7592$

5.6.10 Using the central limit theorem the required probability can be estimated as

$P(N(100 \times 9.2, 100 \times 9.2) \leq 1000) = \Phi(2.638) = 0.9958.$

5.6.11 If the variables are measured in minutes after 2pm, the probability of making the connection is

$P(X_1 + 30 - X_2 \leq 0)$

where $X_1 \sim N(47, 11^2)$ and $X_2 \sim N(95, 3^2)$.

This probability is

$P(N(47 + 30 - 95, 11^2 + 3^2) \leq 0) = \Phi(1.579) = 0.9428.$

5.6.12 The normal approximation is
$P(N(80 \times 0.25, 80 \times 0.25 \times 0.75) \geq 25 - 0.5)$
$= 1 - \Phi(1.162) = 0.1226.$

If physician D leaves the clinic, then the normal approximation is
$P(N(80 \times 0.3333, 80 \times 0.3333 \times 0.6667) \geq 25 - 0.5)$
$= 1 - \Phi(-0.514) = 0.6963.$

5.6.13 (a) $P(B(235, 0.9) \geq 221)$
$\simeq P(N(235 \times 0.9, 235 \times 0.9 \times 0.1) \geq 221 - 0.5)$
$= 1 - \Phi(1.957) = 0.025$

(b) If n passengers are booked on the flight, it is required that
$P(B(n, 0.9) \geq 221)$
$\simeq P(N(n \times 0.9, n \times 0.9 \times 0.1) \geq 221 - 0.5) \leq 0.25.$

This is satisfied at $n = 241$ but not at $n = 242$.
Therefore, the airline can book up to 241 passengers on the flight.

5.6.14 (a) $P(0.6 \leq N(0, 1) \leq 2.2)$
$= \Phi(2.2) - \Phi(0.6)$
$= 0.9861 - 0.7257 = 0.2604$

(b) $P(3.5 \leq N(4.1, 0.25^2) \leq 4.5)$

$= P\left(\frac{3.5-4.1}{0.25} \leq N(0, 1) \leq \frac{4.5-4.1}{0.25}\right)$

$= \Phi(1.6) - \Phi(-2.4)$

$= 0.9452 - 0.0082 = 0.9370$

(c) Since $\chi^2_{0.95,28} = 16.928$ and $\chi^2_{0.90,28} = 18.939$ the required probability is $0.95 - 0.90 = 0.05.$

(d) Since $t_{0.05,22} = 1.717$ and $t_{0.005,22} = 2.819$ the required probability is $(1 - 0.005) - 0.05 = 0.945.$

5.6.15 $P(X \geq 25) = 1 - \Phi\left(\frac{\ln(25)-3.1}{0.1}\right)$

$= 1 - \Phi(1.189) = 0.117$

$P(B(200, 0.117) \geq 30)$

$\simeq P(N(200 \times 0.117, 200 \times 0.117 \times 0.883) \geq 29.5)$

$= P\left(N(0,1) \geq \frac{29.5-23.4}{\sqrt{20.66}}\right)$

$= 1 - \Phi(1.342) = 0.090$

5.6.16 (a) True

(b) True

(c) True

(d) True

(e) True

5.6.17 $P(B(400, 0.2) \geq 90)$

$\simeq P(N(400 \times 0.2, 400 \times 0.2 \times 0.8) \geq 89.5)$

$= P\left(N(0,1) \geq \frac{89.5-80}{\sqrt{64}}\right)$

$= 1 - \Phi(1.1875) = 0.118$

5.6.18 (a) The probability that an expression is larger than 0.800 is

$P(N(0.768, 0.083^2) \geq 0.80) = P\left(N(0,1) \geq \frac{0.80-0.768}{0.083}\right)$

$= 1 - \Phi(0.386) = 0.350$

If Y measures the number of samples out of six that have an expression larger than 0.80, then Y has a binomial distribution with $n = 6$ and $p = 0.350$.

$P(Y \geq 3) = 1 - P(Y < 3)$

$= 1 - \left(\binom{6}{0} \times (0.35)^0 \times (0.65)^6 + \binom{6}{1} \times (0.35)^1 \times (0.65)^5\right.$

$\left.+ \binom{6}{2} \times (0.35)^2 \times (0.65)^4\right)$

$= 0.353$

(b) Let Y_1 be the number of samples that have an expression smaller than 0.70, let Y_2 be the number of samples that have an expression between 0.70 and 0.75, let Y_3 be the number of samples that have an expression between 0.75 and 0.78, and let Y_4 be the number of samples that have an expression larger than 0.78.

$P(X_i \leq 0.7) = \Phi(-0.819) = 0.206$
$P(0.7 \leq X_i \leq 0.75) = \Phi(-0.217) - \Phi(-0.819) = 0.414 - 0.206 = 0.208$
$P(0.75 \leq X_i \leq 0.78) = \Phi(0.145) - \Phi(-0.217) = 0.557 - 0.414 = 0.143$

$P(X_i \geq 0.78) = 1 - \Phi(0.145) = 1 - 0.557 = 0.443$

$P(Y_1 = 2, Y_2 = 2, Y_3 = 0, Y_4 = 2)$

$= \frac{6!}{2! \times 2! \times 2! \times 0!} \times 0.206^2 \times 0.208^2 \times 0.143^0 \times 0.443^2$

$= 0.032$

(c) A negative binomial distribution can be used with $r = 3$ and $p = P(X \leq 0.76) = \Phi(-0.096) = 0.462$.

The required probability is

$P(Y = 6) = \binom{5}{3} \times (1 - 0.462)^3 \times 0.462^3 = 0.154.$

(d) A geometric distribution can be used with $p = P(X \leq 0.68) = \Phi(-1.060) = 0.145$.

The required probability is
$P(Y = 5) = (1 - 0.145)^4 \times 0.145 = 0.077.$

(e) Using the hypergeometric distribution the required probability is

$$\frac{\binom{5}{3} \times \binom{5}{3}}{\binom{10}{6}} = 0.476.$$

5.6.19 (a) $P(X \leq 8000) = \Phi\left(\frac{8000-8200}{350}\right)$

$= \Phi(-0.571) = 0.284$

$P(8000 \leq X \leq 8300) = \Phi\left(\frac{8300-8200}{350}\right) - \Phi\left(\frac{8000-8200}{350}\right)$

$= \Phi(0.286) - \Phi(-0.571) = 0.330$

$P(X \geq 8300) = 1 - \Phi\left(\frac{8300-8200}{350}\right)$

$= 1 - \Phi(0.286) = 0.386$

Using the multinomial distribution the required probability is

$\frac{3!}{1! \times 1! \times 1!} \times 0.284^1 \times 0.330^1 \times 0.386^1 = 0.217.$

(b) $P(X \leq 7900) = \Phi\left(\frac{7900-8200}{350}\right) = \Phi(-0.857) = 0.195$

Using the negative binomial distribution the required probability is

$\binom{5}{1} \times (1-0.195)^4 \times 0.195^2 = 0.080.$

(c) $P(X \geq 8500) = 1 - \Phi\left(\frac{8500-8200}{350}\right) = 1 - \Phi(0.857) = 0.195$

Using the binomial distribution the required probability is

$\binom{7}{3} \times (0.195)^3 \times (1-0.195)^4 = 0.109.$

5.6.20 $0.90 = P(X_A \leq X_B)$

$= P(N(220, 11^2) \leq N(t+185, 9^2))$

$= P(N(220-t-185, 11^2+9^2) \leq 0)$

$= P\left(N(0,1) \leq \frac{t-35}{\sqrt{202}}\right)$

Therefore,

$\frac{t-35}{\sqrt{202}} = z_{0.10} = 1.282$

so that $t = 53.22$.

Consequently, operator B started working at 9:53 am.

5.6.21 (a) $P(X \leq 30) = 1 - e^{-(0.03\times 30)^{0.8}} = 0.601$

Using the binomial distribution the required probability is

$\binom{5}{2} \times 0.601^2 \times (1-0.601)^3 = 0.23.$

(b) $P(B(500, 0.399) \leq 210)$

$\simeq P(N(500 \times 0.399, 500 \times 0.399 \times 0.601) \leq 210.5)$

$= P\left(N(0,1) \leq \frac{210.5-199.5}{10.95}\right)$

$= \Phi(1.005) = 0.843$

5.6.22 $P(N(3 \times 45.3, 3 \times 0.02^2) \leq 135.975)$

$= P\left(N(0,1) \leq \frac{135.975-135.9}{\sqrt{3}\times 0.02}\right)$

$= \Phi(2.165) = 0.985$

5.6.23 $P(X_A - X_{B1} - X_{B2} \geq 0)$

$= P(N(67.2, 1.9^2) - N(33.2, 1.1^2) - N(33.2, 1.1^2) \geq 0)$

$= P(N(67.2 - 33.2 - 33.2, 1.9^2 + 1.1^2 + 1.1^2) \geq 0)$

$= P(N(0.8, 6.03) \geq 0)$

$= P\left(N(0,1) \geq \frac{-0.8}{\sqrt{6.03}}\right)$

$= 1 - \Phi(-0.326) = 0.628$

5.6.24 $P(X \geq 25) = e^{-25/32} = 0.458$

$P(B(240, 0.458) \geq 120)$

$\simeq P\left(N(240 \times 0.458, 240 \times 0.458 \times 0.542) \geq 119.5\right)$

$= P\left(N(0,1) \geq \frac{119.5 - 109.9}{\sqrt{59.57}}\right)$

$= 1 - \Phi(1.24) = 0.108$

5.6.25 (a) $P(N(55980, 10^2) \geq N(55985, 9^2))$

$= P(N(55980 - 55985, 10^2 + 9^2) \geq 0)$

$= P(N(-5, 181) \geq 0)$

$= P\left(N(0,1) \geq \frac{5}{\sqrt{181}}\right)$

$= 1 - \Phi(0.372) = 0.355$

(b) $P(N(55980, 10^2) \leq N(56000, 10^2))$

$= P(N(55980 - 56000, 10^2 + 10^2) \leq 0)$

$= P(N(-20, 200) \leq 0)$

$= P\left(N(0,1) \leq \frac{20}{\sqrt{200}}\right)$

$= \Phi(1.414) = 0.921$

(c) $P(N(56000, 10^2) \leq 55995) \times P(N(56005, 8^2) \leq 55995)$

$= P\left(N(0,1) \leq \frac{55995 - 56000}{10}\right) \times P\left(N(0,1) \leq \frac{55995 - 56005}{8}\right)$

$= \Phi(-0.5) \times \Phi(-1.25)$

$= 0.3085 \times 0.1056 = 0.033$

5.6.26 (a) $t_{0.10,40} = 1.303$ and $t_{0.025,40} = 2.021$ so that
$P(-1.303 \leq t_{40} \leq 2.021) = 0.975 - 0.10 = 0.875$

(b) $P(t_{17} \geq 2.7) = 0.008$

5.6.27 (a) $P(F_{16,20} \leq 2) = 0.928$

(b) $P(\chi^2_{28} \geq 47) = 0.014$

(c) $P(t_{29} \geq 1.5) = 0.072$

(d) $P(t_7 \leq -1.3) = 0.117$

(e) $P(t_{10} \geq -2) = 0.963$

5.6.28 (a) $P(\chi^2_{40} > 65.0) = 0.007$

(b) $P(t_{20} < -1.2) = 0.122$

(c) $P(t_{26} < 3.0) = 0.997$

(d) $P(F_{8,14} > 4.8) = 0.0053.$

5.6.29 Let the time be measured in minutes after 9:40am.
The doctor's consultation starts at time $X_1 \sim N(62, 4^2)$.
The length of the consultation is $X_2 \sim N(17, 5^2)$.
The time spent at the laboratory is $X_3 \sim N(11, 3^2)$.
The time spent at the pharmacy is $X_4 \sim N(15, 5^2)$.

Therefore,

$P(X_1 + X_2 + 1 + X_3 + 1 + X_4 \leq 120)$

$= P(N(62 + 17 + 1 + 11 + 1 + 15, 4^2 + 5^2 + 3^2 + 5^2) \leq 120)$

$= P(N(107, 75) \leq 120) = P\left(N(0,1) \leq \frac{120-107}{\sqrt{75}}\right)$

$= \Phi(1.50) = 0.933.$

Chapter 6

Descriptive Statistics

6.1 Experimentation

6.1.1 For this problem the population is the somewhat imaginary concept of "all possible die rolls."

The sample should be representative if the die is shaken properly.

6.1.2 The population may be all television sets that are ready to be shipped during a certain period of time, although the representativeness of the sample depends on whether the television sets that are ready to be shipped on that Friday morning are in any way different from television sets that are ready to be shipped at other times.

6.1.3 Is the population all students? - or the general public? - or perhaps it should just be computing students at that college?

You have to consider whether the eye colors of computing students are representative of the eye colors of all students or of all people.

Perhaps eye colors are affected by race and the racial make-up of the class may not reflect that of the student body or the general public as a whole.

6.1.4 The population is all service times under certain conditions.

The conditions depend upon how representative the period between 2:00 and 3:00 on that Saturday afternoon is of other serving periods.

The service times would be expected to depend on how busy the restaurant is and on the number of servers available.

6.1.5 The population is all peach boxes received by the supermarket within the time period.

The random sampling within each day's shipment and the recording of an observation every day should ensure that the sample is reasonably representative.

6.1.6 The population is the number of calls received in each minute of every day during the period of investigation.

The spacing of the sampling times should ensure that the sample is representative.

6.1.7 The population may be all bricks shipped by that company, or just the bricks in that particular delivery.

The random selection of the sample should ensure that it is representative of that particular delivery of bricks.

However, that specific delivery of bricks may not be representative of all of the deliveries from that company.

6.1.8 The population is all car panels spray painted by the machine.

The selection method of the sample should ensure that it is representative.

6.1.9 The population is all plastic panels made by the machine.

If the 80 sample panels are selected in some random manner then they should be representative of the entire population.

6.2 Data Presentation

6.2.3 The smallest observation 1.097 and the largest observation 1.303 both appear to be outliers.

6.2.4 The largest observation 66.00 can be considered to be an outlier.

In addition, the second largest observation 51 might also be considered to be an outlier.

6.2.5 There would appear to be no reason to doubt that the die is a fair one.

A test of the fairness of the die could be made using the methods presented in section 10.3.

6.2.6 It appears that worse grades are assigned less frequently than better grades.

6.2.7 The assignment "other" is employed considerably less frequently than blue, green, and brown, which are each about equally frequent.

6.2.8 The data set appears to be slightly positively skewed.

The observations 186, 177, 143, and 135 can all be considered to be outliers.

6.2.9 The observations 25 and 14 can be considered to be outliers.

6.2.10 The histogram is bimodal.

It may possibly be considered to be a mixture of two distributions corresponding to "busy" periods and "slow" periods.

6.2.11 The smallest observation 0.874 can be considered to be an outlier.

6.2.12 The largest observation 0.538 can be considered to be an outlier.

6.2.13 This is a negatively skewed data set.

The smallest observations 6.00 and 6.04 can be considered to be outliers, and possibly some of the other small observations may also be considered to be outliers.

6.2.14 A bar chart represents discrete or categorical data while a histogram represents continuous data.

6.3 Sample Statistics

Note: The sample statistics for the problems in this section depend upon whether any observations have been removed as outliers. To avoid confusion, the answers given here assume that **no** observations have been removed.

The trimmed means given here are those obtained by removing the largest 5% and the smallest 5% of the data observations.

6.3.1 The sample mean is $\bar{x} = 155.95$.

The sample median is 159.

The sample trimmed mean is 156.50.

The sample standard deviation is $s = 18.43$.

The upper sample quartile is 169.5.

The lower sample quartile is 143.25.

6.3.2 The sample mean is $\bar{x} = 1.2006$.

The sample median is 1.2010.

The sample trimmed mean is 1.2007.

The sample standard deviation is $s = 0.0291$.

The upper sample quartile is 1.2097.

The lower sample quartile is 1.1890.

6.3.3 The sample mean is $\bar{x} = 37.08$.

The sample median is 35.

The sample trimmed mean is 36.35.

The sample standard deviation is $s = 8.32$.

The upper sample quartile is 40.

The lower sample quartile is 33.5.

6.3.4 The sample mean is $\bar{x} = 3.567$.

The sample median is 3.5.

The sample trimmed mean is 3.574.

The sample standard deviation is $s = 1.767$.

The upper sample quartile is 5.

The lower sample quartile is 2.

6.3.5 The sample mean is $\bar{x} = 69.35$.

The sample median is 66.

The sample trimmed mean is 67.88.

The sample standard deviation is $s = 17.59$.

The upper sample quartile is 76.

The lower sample quartile is 61.

6.3.6 The sample mean is $\bar{x} = 3.291$.

The sample median is 2.

The sample trimmed mean is 2.755.

The sample standard deviation is $s = 3.794$.

The upper sample quartile is 4.

The lower sample quartile is 1.

6.3.7 The sample mean is $\bar{x} = 12.211$.

The sample median is 12.

The sample trimmed mean is 12.175.

The sample standard deviation is $s = 2.629$.

The upper sample quartile is 14.

The lower sample quartile is 10.

6.3.8 The sample mean is $\bar{x} = 1.1106$.

The sample median is 1.1102.

The sample trimmed mean is 1.1112.

The sample standard deviation is $s = 0.0530$.

The upper sample quartile is 1.1400.

The lower sample quartile is 1.0813.

6.3.9 The sample mean is $\bar{x} = 0.23181$.

The sample median is 0.220.

The sample trimmed mean is 0.22875.

The sample standard deviation is $s = 0.07016$.

The upper sample quartile is 0.280.

The lower sample quartile is 0.185.

6.3.10 The sample mean is $\bar{x} = 9.2294$.

The sample median is 9.435.

The sample trimmed mean is 9.3165.

The sample standard deviation is $s = 0.8423$.

The upper sample quartile is 9.81.

The lower sample quartile is 8.9825.

6.3.11 The sample mean is

$\frac{65+x}{6}$

and

$\sum_{i=1}^{6} x_i^2 = 1037 + x^2$.

Therefore,

$s^2 = \frac{1037+x^2-(65+x)^2/6}{5}$

which by differentiation can be shown to be minimized when $x = 13$ (which is the average of the other five data points).

6.6 Supplementary Problems

6.6.1 The population from which the sample is drawn would be all of the birds on the island.

However, the sample may not be representative if some species are more likely to be observed than others.

It appears that the grey markings are the most common, followed by the black markings.

6.6.2 There do not appear to be any seasonal effects, although there may possibly be a correlation from one month to the next.

The sample mean is $\bar{x} = 17.79$.

The sample median is 17.

The sample trimmed mean is 17.36.

The sample standard deviation is $s = 6.16$.

The upper sample quartile is 21.75.

The lower sample quartile is 14.

6.6.3 One question of interest in interpreting this data set is whether or not the month of sampling is representative of other months.

The sample is skewed.

The most frequent data value (the sample mode) is one error.

The sample mean is $\bar{x} = 1.633$.

The sample median is 1.5.

The sample trimmed mean is 1.615.

The sample standard deviation is $s = 0.999$.

The upper sample quartile is 2.

The lower sample quartile is 1.

6.6.4 The population would be all adult males who visit the clinic.

This could be representative of all adult males in the population unless there is something special about the clientele of this clinic.

The largest observation 75.9 looks like an outlier on a histogram but may be a valid observation.

The sample mean is $\bar{x} = 69.618$.

The sample median is 69.5.

The sample trimmed mean is 69.513.

The sample standard deviation is $s = 1.523$.

The upper sample quartile is 70.275.

The lower sample quartile is 68.6.

6.6.5 Two or three of the smallest observations and the largest observation may be considered to be outliers.

The sample mean is $\bar{x} = 32.042$.

The sample median is 32.55.

The sample trimmed mean is 32.592.

The sample standard deviation is $s = 5.817$.

The upper sample quartile is 35.5.

The lower sample quartile is 30.425.

6.6.6 The population of interest can be considered to be the soil throughout the construction site.

If the soil is of a fairly uniform type, and if the samples were taken representatively throughout the site, then they should provide accurate information on the soil throughout the entire construction site.

The sample mean is $\bar{x} = 25.318$.

The sample median is 25.301.

The sample trimmed mean is 25.319.

The sample standard deviation is $s = 0.226$.

The upper sample quartile is 25.501.

The lower sample quartile is 25.141.

6.6.7 (a) True

(b) False

(c) True

(d) False

Chapter 7

Statistical Estimation and Sampling Distributions

7.2 Properties of Point Estimates

7.2.1 (a) $\text{bias}(\hat{\mu}_1) = 0$
The point estimate $\hat{\mu}_1$ is unbiased.

$\text{bias}(\hat{\mu}_2) = 0$
The point estimate $\hat{\mu}_2$ is unbiased.

$\text{bias}(\hat{\mu}_3) = 9 - \frac{\mu}{2}$

(b) $\text{Var}(\hat{\mu}_1) = 6.2500$

$\text{Var}(\hat{\mu}_2) = 9.0625$

$\text{Var}(\hat{\mu}_3) = 1.9444$

The point estimate $\hat{\mu}_3$ has the smallest variance.

(c) $\text{MSE}(\hat{\mu}_1) = 6.2500$

$\text{MSE}(\hat{\mu}_2) = 9.0625$

$\text{MSE}(\hat{\mu}_3) = 1.9444 + (9 - \frac{\mu}{2})^2$
This is equal to 26.9444 when $\mu = 8$.

7.2.2 (a) $\text{bias}(\hat{\mu}_1) = 0$

$\text{bias}(\hat{\mu}_2) = -0.217\mu$

$\text{bias}(\hat{\mu}_3) = 2 - \frac{\mu}{4}$

The point estimate $\hat{\mu}_1$ is unbiased.

(b) $\text{Var}(\hat{\mu}_1) = 4.444$

$\text{Var}(\hat{\mu}_2) = 2.682$

$\text{Var}(\hat{\mu}_3) = 2.889$

The point estimate $\hat{\mu}_2$ has the smallest variance.

(c) $\text{MSE}(\hat{\mu}_1) = 4.444$

$\text{MSE}(\hat{\mu}_2) = 2.682 + 0.0469\mu^2$
This is equal to 3.104 when $\mu = 3$.

$\text{MSE}(\hat{\mu}_3) = 2.889 + (2 - \frac{\mu}{4})^2$
This is equal to 4.452 when $\mu = 3$.

7.2.3 (a) $\text{Var}(\hat{\mu}_1) = 2.5$

(b) The value $p = 0.6$ produces the smallest variance which is $\text{Var}(\hat{\mu}) = 2.4$.

(c) The relative efficiency is $\frac{2.4}{2.5} = 0.96$.

7.2.4 (a) $\text{Var}(\hat{\mu}_1) = 2$

(b) The value $p = 0.875$ produces the smallest variance which is $\text{Var}(\hat{\mu}) = 0.875$.

(c) The relative efficiency is $\frac{0.875}{2} = 0.4375$.

7.2.5 (a) $a_1 + \ldots + a_n = 1$

(b) $a_1 = \ldots = a_n = \frac{1}{n}$

7.2.6 $\text{MSE}(\hat{\theta}_1) = 0.02\,\theta^2 + (0.13\,\theta)^2 = 0.0369\,\theta^2$

$\text{MSE}(\hat{\theta}_2) = 0.07\,\theta^2 + (0.05\,\theta)^2 = 0.0725\,\theta^2$

$\text{MSE}(\hat{\theta}_3) = 0.005\,\theta^2 + (0.24\,\theta)^2 = 0.0626\,\theta^2$

The point estimate $\hat{\theta}_1$ has the smallest mean square error.

7.2.7 $\text{bias}(\hat{\mu}) = \frac{\mu_0 - \mu}{2}$

$\text{Var}(\hat{\mu}) = \frac{\sigma^2}{4}$

$\text{MSE}(\hat{\mu}) = \frac{\sigma^2}{4} + \frac{(\mu_0 - \mu)^2}{4}$

$\text{MSE}(X) = \sigma^2$

7.2.8 (a) $\text{bias}(\hat{p}) = -\frac{p}{11}$

(b) $\text{Var}(\hat{p}) = \frac{10\,p\,(1-p)}{121}$

(c) $\text{MSE}(\hat{p}) = \frac{10\,p\,(1-p)}{121} + \left(\frac{p}{11}\right)^2 = \frac{10p-9p^2}{121}$

(d) $\text{bias}\left(\frac{X}{10}\right) = 0$

$\text{Var}\left(\frac{X}{10}\right) = \frac{p(1-p)}{10}$

$\text{MSE}\left(\frac{X}{10}\right) = \frac{p(1-p)}{10}$

7.2.9 $\text{Var}\left(\frac{X_1+X_2}{2}\right)$

$= \frac{\text{Var}(X_1)+\text{Var}(X_2)}{4}$

$= \frac{5.39^2+9.43^2}{4}$

$= 29.49$

The standard deviation is $\sqrt{29.49} = 5.43$.

7.3 Sampling Distributions

7.3.1 $\text{Var}\left(\frac{X_1}{n_1}\right) = \frac{p(1-p)}{n_1}$

$\text{Var}\left(\frac{X_2}{n_2}\right) = \frac{p(1-p)}{n_2}$

The relative efficiency is the ratio of these two variances which is $\frac{n_1}{n_2}$.

7.3.2 (a) $P\left(\left|N\left(0, \frac{1}{10}\right)\right| \leq 0.3\right) = 0.6572$

(b) $P\left(\left|N\left(0, \frac{1}{30}\right)\right| \leq 0.3\right) = 0.8996$

7.3.3 (a) $P\left(\left|N\left(0, \frac{7}{15}\right)\right| \leq 0.4\right) = 0.4418$

(b) $P\left(\left|N\left(0, \frac{7}{50}\right)\right| \leq 0.4\right) = 0.7150$

7.3.4 (a) Solving

$P\left(5 \times \frac{\chi^2_{30}}{30} \leq c\right) = P(\chi^2_{30} \leq 6c) = 0.90$

gives $c = 6.709$.

(b) Solving

$P\left(5 \times \frac{\chi^2_{30}}{30} \leq c\right) = P(\chi^2_{30} \leq 6c) = 0.95$

gives $c = 7.296$.

7.3.5 (a) Solving

$P\left(32 \times \frac{\chi^2_{20}}{20} \leq c\right) = P\left(\chi^2_{20} \leq \frac{5c}{8}\right) = 0.90$

gives $c = 45.46$.

(b) Solving

$P\left(32 \times \frac{\chi^2_{20}}{20} \leq c\right) = P\left(\chi^2_{20} \leq \frac{5c}{8}\right) = 0.95$

gives $c = 50.26$.

7.3.6 (a) Solving

$P(|t_{15}| \leq c) = 0.95$

gives $c = t_{0.025,15} = 2.131$.

(b) Solving

$P(|t_{15}| \leq c) = 0.99$

gives $c = t_{0.005,15} = 2.947$.

7.3.7 (a) Solving

$P\left(\frac{|t_{20}|}{\sqrt{21}} \leq c\right) = 0.95$

gives $c = \frac{t_{0.025,20}}{\sqrt{21}} = 0.4552$.

(b) Solving

$P\left(\frac{|t_{20}|}{\sqrt{21}} \leq c\right) = 0.99$

gives $c = \frac{t_{0.005,20}}{\sqrt{21}} = 0.6209$.

7.3.8 $\hat{p} = \frac{234}{450} = 0.52$

$\text{s.e.}(\hat{p}) = \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} = \sqrt{\frac{0.52\times0.48}{450}} = 0.0236$

7.3.9 $\hat{\mu} = \bar{x} = 974.3$

$\text{s.e.}(\hat{\mu}) = \frac{s}{\sqrt{n}} = \sqrt{\frac{452.1}{35}} = 3.594$

7.3.10 $\hat{p} = \frac{24}{120} = 0.2$

$\text{s.e.}(\hat{p}) = \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} = \sqrt{\frac{0.2\times0.8}{120}} = 0.0365$

7.3.11 $\hat{p} = \frac{33}{150} = 0.22$

$\text{s.e.}(\hat{p}) = \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} = \sqrt{\frac{0.22\times0.78}{150}} = 0.0338$

7.3.12 $\hat{p} = \frac{26}{80} = 0.325$

$\text{s.e.}(\hat{p}) = \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} = \sqrt{\frac{0.325\times0.675}{80}} = 0.0524$

7.3.13 $\hat{\mu} = \bar{x} = 69.35$

s.e.$(\hat{\mu}) = \frac{s}{\sqrt{n}} = \frac{17.59}{\sqrt{200}} = 1.244$

7.3.14 $\hat{\mu} = \bar{x} = 3.291$

s.e.$(\hat{\mu}) = \frac{s}{\sqrt{n}} = \frac{3.794}{\sqrt{55}} = 0.512$

7.3.15 $\hat{\mu} = \bar{x} = 12.211$

s.e.$(\hat{\mu}) = \frac{s}{\sqrt{n}} = \frac{2.629}{\sqrt{90}} = 0.277$

7.3.16 $\hat{\mu} = \bar{x} = 1.1106$

s.e.$(\hat{\mu}) = \frac{s}{\sqrt{n}} = \frac{0.0530}{\sqrt{125}} = 0.00474$

7.3.17 $\hat{\mu} = \bar{x} = 0.23181$

s.e.$(\hat{\mu}) = \frac{s}{\sqrt{n}} = \frac{0.07016}{\sqrt{75}} = 0.00810$

7.3.18 $\hat{\mu} = \bar{x} = 9.2294$

s.e.$(\hat{\mu}) = \frac{s}{\sqrt{n}} = \frac{0.8423}{\sqrt{80}} = 0.0942$

7.3.19 If a sample of size $n = 100$ is used, then the probability is
$P(0.24 - 0.05 \leq \hat{p} \leq 0.24 + 0.05) = P(19 \leq B(100, 0.24) \leq 29)$.

Using a normal approximation this can be estimated as

$\Phi\left(\frac{29+0.5-100\times0.24}{\sqrt{100\times0.24\times0.76}}\right) - \Phi\left(\frac{19-0.5-100\times0.24}{\sqrt{100\times0.24\times0.76}}\right)$

$= \Phi(1.288) - \Phi(-1.288) = 0.8022.$

If a sample of size $n = 200$ is used, then the probability is
$P(38 \leq B(200, 0.24) \leq 58)$.

Using a normal approximation this can be estimated as

$\Phi\left(\frac{58+0.5-200\times0.24}{\sqrt{200\times0.24\times0.76}}\right) - \Phi\left(\frac{38-0.5-200\times0.24}{\sqrt{200\times0.24\times0.76}}\right)$

$= \Phi(1.738) - \Phi(-1.738) = 0.9178.$

7.3.20 $P(173 \leq \hat{\mu} \leq 175) = P(173 \leq \bar{X} \leq 175)$

where

$\bar{X} \sim N\left(174, \frac{2.8^2}{30}\right)$.

This is

$\Phi\left(\frac{175-174}{\sqrt{2.8^2/30}}\right) - \Phi\left(\frac{173-174}{\sqrt{2.8^2/30}}\right)$

$= \Phi(1.956) - \Phi(-1.956) = 0.9496.$

7.3.21 $P(0.62 \leq \hat{p} \leq 0.64)$

$= P(300 \times 0.62 \leq B(300, 0.63) \leq 300 \times 0.64)$

$\simeq P(185.5 \leq N(300 \times 0.63, 300 \times 0.63 \times 0.37) \leq 192.5)$

$= P\left(\frac{185.5-189}{\sqrt{69.93}} \leq N(0,1) \leq \frac{192.5-189}{\sqrt{69.93}}\right)$

$= \Phi(0.419) - \Phi(-0.419) = 0.324$

7.3.22 $P\left(109.9 \leq N\left(110.0, \frac{0.4^2}{22}\right) \leq 110.1\right)$

$= P\left(\frac{\sqrt{22}(109.9-110.0)}{0.4} \leq N(0,1) \leq \frac{\sqrt{22}(110.1-110.0)}{0.4}\right)$

$= \Phi(1.173) - \Phi(-1.173) = 0.759$

7.3.23 $\sqrt{\frac{0.126 \times 0.874}{360}} = 0.017$

7.3.24 $P\left(N\left(341, \frac{2^2}{20}\right) \leq 341.5\right)$

$= P\left(N(0.1) \leq \frac{\sqrt{20} \times (341.5-341)}{2}\right)$

$= \Phi(1.118) = 0.547$

7.3.25 $P\left(\mu - 2 \leq N\left(\mu, \frac{5.2^2}{18}\right) \leq \mu + 2\right)$

$= P\left(\frac{-\sqrt{18} \times 2}{5.2} \leq N(0.1) \leq \frac{\sqrt{18} \times 2}{5.2}\right)$

$= \Phi(1.632) - \Phi(-1.632) = 0.103$

7.3.26 The largest standard error is obtained when $\hat{p} = 0.5$ and is equal to

$\sqrt{\frac{0.5\times 0.5}{1400}} = 0.0134.$

7.3.27 $P(X \geq 60) = e^{-0.02\times 60} = 0.301$

Let Y be the number of components that last longer than one hour.

$P\left(0.301 - 0.05 \leq \frac{Y}{110} \leq 0.301 + 0.05\right)$

$= P(27.6 \leq Y \leq 38.6)$

$= P(28 \leq B(110, 0.301) \leq 38)$

$\simeq P\,(27.5 \leq N(110 \times 0.301, 110 \times 0.301 \times 0.699) \leq 38.5)$

$= P\left(\frac{27.5-33.11}{\sqrt{23.14}} \leq N(0,1) \leq \frac{38.5-33.11}{\sqrt{23.14}}\right)$

$= \Phi(1.120) - \Phi(-1.166)$

$= 0.869 - 0.122 = 0.747$

7.3.28 (a) $P(\mu - 0.5 \leq \bar{X} \leq \mu + 0.5)$

$= P\left(\mu - 0.5 \leq N\left(\mu, \frac{0.82^2}{5}\right) \leq \mu + 0.5\right)$

$= \Phi\left(\frac{0.5\sqrt{5}}{0.82}\right) - \Phi\left(\frac{-0.5\sqrt{5}}{0.82}\right) = 0.827$

(b) $P(\mu - 0.5 \leq \bar{X} \leq \mu + 0.5)$

$= P\left(\mu - 0.5 \leq N\left(\mu, \frac{0.82^2}{10}\right) \leq \mu + 0.5\right)$

$= \Phi\left(\frac{0.5\sqrt{10}}{0.82}\right) - \Phi\left(\frac{-0.5\sqrt{10}}{0.82}\right) = 0.946$

(c) In order for

$P\left(\mu - 0.5 \leq N\left(\mu, \frac{0.82^2}{n}\right) \leq \mu + 0.5\right)$

$= \Phi\left(\frac{0.5\sqrt{n}}{0.82}\right) - \Phi\left(\frac{-0.5\sqrt{n}}{0.82}\right) \geq 0.99$

it is necessary that

$\frac{0.5\sqrt{n}}{0.82} \geq z_{0.005} = 2.576$

which is satisfied for a sample size n of at least 18.

7.3.29 (a) $p = \frac{592}{3288} = 0.18$

$P(p - 0.1 \leq \hat{p} \leq p + 0.1)$

$= P\left(0.08 \leq \frac{X}{20} \leq 0.28\right)$

$= P(1.6 \leq X \leq 5.6)$

where $X \sim B(20, 0.18)$.

This probability is

$P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$

$$= \binom{20}{2} \times 0.18^2 \times 0.82^{18} + \binom{20}{3} \times 0.18^3 \times 0.82^{17}$$

$$+ \binom{20}{4} \times 0.18^4 \times 0.82^{16} + \binom{20}{5} \times 0.18^5 \times 0.82^{15}$$

$= 0.7626.$

(b) The probability that a sampled meter is operating outside the acceptable tolerance limits is now

$p^* = \frac{184}{2012} = 0.09.$

$P(p - 0.1 \leq \hat{p} \leq p + 0.1)$

$= P\left(0.08 \leq \frac{Y}{20} \leq 0.28\right)$

$= P(1.6 \leq Y \leq 5.6)$

where $Y \sim B(20, 0.09)$.

This probability is

$P(Y = 2) + P(Y = 3) + P(Y = 4) + P(Y = 5)$

$$= \binom{20}{2} \times 0.09^2 \times 0.91^{18} + \binom{20}{3} \times 0.09^3 \times 0.91^{17}$$

$$+ \binom{20}{4} \times 0.09^4 \times 0.91^{16} + \binom{20}{5} \times 0.09^5 \times 0.91^{15}$$

$= 0.5416.$

7.4 Constructing Parameter Estimates

7.4.1 $\hat{\lambda} = \bar{x} = 5.63$

s.e.$(\hat{\lambda}) = \sqrt{\frac{\hat{\lambda}}{n}} = \sqrt{\frac{5.63}{23}} = 0.495$

7.4.2 Using the method of moments the point estimates $\hat{a}$ and $\hat{b}$ are the solutions to the equations

$\frac{a}{a+b} = 0.782$

and

$\frac{ab}{(a+b)^2(a+b+1)} = 0.0083$

which are $\hat{a} = 15.28$ and $\hat{b} = 4.26$.

7.4.3 Using the method of moments

$E(X) = \frac{1}{\lambda} = \bar{x}$

which gives $\hat{\lambda} = \frac{1}{\bar{x}}$.

The likelihood is

$L(x_1, \ldots, x_n, \lambda) = \lambda^n \, e^{-\lambda(x_1+\ldots+x_n)}$

which is maximized at $\hat{\lambda} = \frac{1}{\bar{x}}$.

7.4.4 $\hat{p}_i = \frac{x_i}{n}$ for $1 \leq i \leq n$

7.4.5 Using the method of moments

$E(X) = \frac{5}{\lambda} = \bar{x}$

which gives $\hat{\lambda} = \frac{5}{\bar{x}}$.

The likelihood is

$L(x_1, \ldots, x_n, \lambda) = \left(\frac{1}{24}\right)^n \times \lambda^{5n} \times x_1^4 \times \ldots \times x_n^4 \times e^{-\lambda(x_1+\ldots+x_n)}$

which is maximimized at $\hat{\lambda} = \frac{5}{\bar{x}}$.

7.6 Supplementary Problems

7.6.1 $\text{bias}(\hat{\mu}_1) = 5 - \frac{\mu}{2}$

$\text{bias}(\hat{\mu}_2) = 0$

$\text{Var}(\hat{\mu}_1) = \frac{1}{8}$

$\text{Var}(\hat{\mu}_2) = \frac{1}{2}$

$\text{MSE}(\hat{\mu}_1) = \frac{1}{8} + (5 - \frac{\mu}{2})^2$

$\text{MSE}(\hat{\mu}_2) = \frac{1}{2}$

7.6.2 (a) $\text{bias}(\hat{p}) = -\frac{p}{7}$

(b) $\text{Var}(\hat{p}) = \frac{3p(1-p)}{49}$

(c) $\text{MSE}(\hat{p}) = \frac{3p(1-p)}{49} + (\frac{p}{7})^2 = \frac{3p-2p^2}{49}$

(d) $\text{MSE}\left(\frac{X}{12}\right) = \frac{p(1-p)}{12}$

7.6.3 (a) $F(t) = P(T \leq t) = P(X_1 \leq t) \times \ldots \times P(X_n \leq t)$

$= \frac{t}{\theta} \times \ldots \times \frac{t}{\theta} = (\frac{t}{\theta})^n$

for $0 \leq t \leq \theta$

(b) $f(t) = \frac{dF(t)}{dt} = n\ \frac{t^{n-1}}{\theta^n}$

for $0 \leq t \leq \theta$

(c) Notice that

$E(T) = \int_0^\theta\ t\ f(t)\ dt = \frac{n}{n+1}\theta$

so that $E(\hat{\theta}) = \theta$.

(d) Notice that

$E(T^2) = \int_0^\theta\ t^2\ f(t)\ dt = \frac{n}{n+2}\theta^2$

so that

$\text{Var}(T) = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2$

$= \frac{n\theta^2}{(n+2)(n+1)^2}$.

Consequently,

$\mathrm{Var}(\hat{\theta}) = \frac{(n+1)^2}{n^2}\mathrm{Var}(T) = \frac{\theta^2}{n(n+2)}$

and

s.e.$(\hat{\theta}) = \frac{\hat{\theta}}{\sqrt{n(n+2)}}$.

(e) $\hat{\theta} = \frac{11}{10} \times 7.3 = 8.03$

s.e.$(\hat{\theta}) = \frac{8.03}{\sqrt{10\times 12}} = 0.733$

7.6.4 Recall that $f(x_i, \theta) = \frac{1}{\theta}$ for $0 \leq x_i \leq \theta$
(and $f(x_i, \theta) = 0$ elsewhere)
so that the likelihood is $\frac{1}{\theta^n}$
as long as $x_i \leq \theta$ for $1 \leq i \leq n$
and is equal to zero otherwise.

bias$(\hat{\theta}) = -\frac{\theta}{n+1}$

7.6.5 Using the method of moments

$E(X) = \frac{1}{p} = \bar{x}$

which gives $\hat{p} = \frac{1}{\bar{x}}$.

The likelihood is

$L(x_1, \ldots, x_n, \lambda) = p^n(1-p)^{x_1+\ldots+x_n-n}$

which is maximimized at $\hat{p} = \frac{1}{\bar{x}}$.

7.6.6 $\hat{p} = \frac{35}{100} = 0.35$

s.e.$(\hat{p}) = \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} = \sqrt{\frac{0.35\times 0.65}{100}} = 0.0477$

7.6.7 $\hat{\mu} = \bar{x} = 17.79$

s.e.$(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{6.16}{\sqrt{24}} = 1.26$

7.6.8 $\hat{\mu} = \bar{x} = 1.633$

s.e.$(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{0.999}{\sqrt{30}} = 0.182$

7.6.9 $\hat{\mu} = \bar{x} = 69.618$

s.e.$(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{1.523}{\sqrt{60}} = 0.197$

7.6.10 $\hat{\mu} = \bar{x} = 32.042$

s.e.$(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{5.817}{\sqrt{40}} = 0.920$

7.6.11 $\text{Var}(s_1^2) = \text{Var}\left(\frac{\sigma^2 \chi^2_{n_1-1}}{n_1-1}\right)$

$= \left(\frac{\sigma^2}{n_1-1}\right)^2 \text{Var}(\chi^2_{n_1-1})$

$= \left(\frac{\sigma^2}{n_1-1}\right)^2 2(n_1-1) = \frac{2\sigma^4}{n_1-1}$

Similarly, $\text{Var}(s_2^2) = \frac{2\sigma^4}{n_2-1}$.

The ratio of these two variances is $\frac{n_1-1}{n_2-1}$.

7.6.12 The true proportion of "very satisfied" customers is

$p = \frac{11842}{24839} = 0.4768$.

The probability that the manager's estimate of the proportion of "very satisfied" customers is within 0.10 of $p = 0.4768$ is

$P(0.4768 - 0.10 \leq \hat{p} \leq 0.4768 + 0.10)$

$= P(0.3768 \times 80 \leq X \leq 0.5768 \times 80)$

$= P(30.144 \leq X \leq 46.144) = P(31 \leq X \leq 46)$

where $X \sim B(80, 0.4768)$.

This probability is 0.9264.

7.6.13 When a sample of size $n = 15$ is used

$P(62.8 - 0.5 \leq \hat{\mu} \leq 62.8 + 0.5)$

$= P(62.3 \leq \bar{X} \leq 63.3)$

where $\bar{X} \sim N(62.8, 3.9^2/15)$.

This probability is equal to

$\Phi\left(\frac{63.3-62.8}{\sqrt{3.9^2/15}}\right) - \Phi\left(\frac{62.3-62.8}{\sqrt{3.9^2/15}}\right)$

$= \Phi(0.4965) - \Phi(-0.4965) = 0.3804$.

When a sample of size $n = 40$ is used
$P(62.8 - 0.5 \leq \hat{\mu} \leq 62.8 + 0.5)$
$= P(62.3 \leq \bar{X} \leq 63.3)$
where $\bar{X} \sim N(62.8, 3.9^2/40)$.

This probability is equal to

$\Phi\left(\frac{63.3-62.8}{\sqrt{3.9^2/40}}\right) - \Phi\left(\frac{62.3-62.8}{\sqrt{3.9^2/40}}\right)$

$= \Phi(0.8108) - \Phi(-0.8108) = 0.5826.$

7.6.14 $\hat{\mu} = \bar{x} = 25.318$

s.e.$(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{0.226}{\sqrt{44}} = 0.0341$

The upper quartile of the distribution of soil compressibilities can be estimated by the upper sample quartile 25.50.

7.6.15 Probability theory

7.6.16 Probability theory

7.6.17 $\hat{p} = \frac{39}{220} = 0.177$

s.e.$(\hat{p}) = \sqrt{\frac{0.177 \times 0.823}{220}} = 0.026$

7.6.18 Let X be the number of cases where the treatment was effective.

$P\left(0.68 - 0.05 \leq \frac{X}{140} \leq 0.68 + 0.05\right)$

$= P(88.2 \leq X \leq 102.2)$

$= P(89 \leq B(140, 0.68) \leq 102)$

$\simeq P(88.5 \leq N(140 \times 0.68, 140 \times 0.68 \times 0.32) \leq 102.5)$

$= P\left(\frac{88.5-95.2}{5.519} \leq N(0,1) \leq \frac{102.5-95.2}{5.519}\right)$

$= \Phi(1.268) - \Phi(-1.268) = 0.80$

7.6.19 (a) $\hat{\mu} = \bar{x} = 70.58$

(b) $\frac{s}{\sqrt{n}} = \frac{12.81}{\sqrt{12}} = 3.70$

(c) $\frac{67+70}{2} = 68.5$

7.6.20 Statistical inference

7.6.21 Statistical inference

7.6.22 (a) True

(b) True

(c) True

(d) True

7.6.23 $P\left(722 \leq \bar{X} \leq 724\right)$

$= P\left(722 \leq N\left(723, \frac{3^2}{11}\right) \leq 724\right)$

$= P\left(\frac{-1\times\sqrt{11}}{3} \leq N(0,1) \leq \frac{1\times\sqrt{11}}{3}\right)$

$= \Phi(1.106) - \Phi(-1.106) = 0.73$

7.6.24 (a) $P\left(\mu - 20.0 \leq \bar{X} \leq \mu + 20.0\right)$

$= P\left(\mu - 20.0 \leq N\left(\mu, \frac{40.0^2}{10}\right) \leq \mu + 20.0\right)$

$= P\left(\frac{-20.0\times\sqrt{10}}{40.0} \leq N(0,1) \leq \frac{20.0\times\sqrt{10}}{40.0}\right)$

$= \Phi(1.58) - \Phi(-1.58) = 0.89$

(b) $P\left(\mu - 20.0 \leq \bar{X} \leq \mu + 20.0\right)$

$= P\left(\mu - 20.0 \leq N\left(\mu, \frac{40.0^2}{20}\right) \leq \mu + 20.0\right)$

$= P\left(\frac{-20.0\times\sqrt{20}}{40.0} \leq N(0,1) \leq \frac{20.0\times\sqrt{20}}{40.0}\right)$

$= \Phi(2.24) - \Phi(-2.24) = 0.97$

7.6.25 $\hat{p}_A = \frac{852}{1962} = 0.434$

s.e.$(\hat{p}_A) = \sqrt{\frac{0.434\times(1-0.434)}{1962}} = 0.011$

Chapter 8

Inferences on a Population Mean

8.1 Confidence Intervals

8.1.1 With $t_{0.025,30} = 2.042$ the confidence interval is

$$\left(53.42 - \frac{2.042\times3.05}{\sqrt{31}}, 53.42 + \frac{2.042\times3.05}{\sqrt{31}}\right) = (52.30, 54.54).$$

8.1.2 With $t_{0.005,40} = 2.704$ the confidence interval is

$$\left(3.04 - \frac{2.704\times0.124}{\sqrt{41}}, 3.04 + \frac{2.704\times0.124}{\sqrt{41}}\right) = (2.99, 3.09).$$

The confidence interval does not contain the value 2.90, and so 2.90 is not a plausible value for the mean glass thickness.

8.1.3 At 90% confidence the critical point is $t_{0.05,19} = 1.729$ and the confidence interval is

$$\left(436.5 - \frac{1.729\times11.90}{\sqrt{20}}, 436.5 + \frac{1.729\times11.90}{\sqrt{20}}\right) = (431.9, 441.1).$$

At 95% confidence the critical point is $t_{0.025,19} = 2.093$ and the confidence interval is

$$\left(436.5 - \frac{2.093\times11.90}{\sqrt{20}}, 436.5 + \frac{2.093\times11.90}{\sqrt{20}}\right) = (430.9, 442.1).$$

At 99% confidence the critical point is $t_{0.005,19} = 2.861$ and the confidence interval is

$$\left(436.5 - \frac{2.861\times11.90}{\sqrt{20}}, 436.5 + \frac{2.861\times11.90}{\sqrt{20}}\right) = (428.9, 444.1).$$

Even the 99% confidence level confidence interval does not contain the value 450.0, and so 450.0 is not a plausible value for the average breaking strength.

8.1.4 With $t_{0.005,15} = 2.947$ the confidence interval is

$\left(1.053 - \frac{2.947 \times 0.058}{\sqrt{16}}, 1.053 + \frac{2.947 \times 0.058}{\sqrt{16}}\right) = (1.010, 1.096)$.

The confidence interval contains the value 1.025, and so 1.025 is a plausible value for the average weight.

8.1.5 With $z_{0.025} = 1.960$ the confidence interval is

$\left(0.0328 - \frac{1.960 \times 0.015}{\sqrt{28}}, 0.0328 + \frac{1.960 \times 0.015}{\sqrt{28}}\right) = (0.0272, 0.0384)$.

8.1.6 At 90% confidence the critical point is $z_{0.05} = 1.645$ and the confidence interval is

$\left(19.50 - \frac{1.645 \times 1.0}{\sqrt{10}}, 19.50 + \frac{1.645 \times 1.0}{\sqrt{10}}\right) = (18.98, 20.02)$.

At 95% confidence the critical point is $z_{0.025} = 1.960$ and the confidence interval is

$\left(19.50 - \frac{1.960 \times 1.0}{\sqrt{10}}, 19.50 + \frac{1.960 \times 1.0}{\sqrt{10}}\right) = (18.88, 20.12)$.

At 99% confidence the critical point is $z_{0.005} = 2.576$ and the confidence interval is

$\left(19.50 - \frac{2.576 \times 1.0}{\sqrt{10}}, 19.50 + \frac{2.576 \times 1.0}{\sqrt{10}}\right) = (18.69, 20.31)$.

Even the 90% confidence level confidence interval contains the value 20.0, and so 20.0 is a plausible value for the average resilient modulus.

8.1.7 With $t_{0.025,n-1} \simeq 2.0$ a sufficient sample size can be estimated as

$n \geq 4 \times \left(\frac{t_{0.025,n-1}\, s}{L_0}\right)^2$

$= 4 \times \left(\frac{2.0 \times 10.0}{5}\right)^2 = 64$.

A sample size of about $n = 64$ should be sufficient.

8.1.8 With $t_{0.005,n-1} \simeq 3.0$ a sufficient sample size can be estimated as

$n \geq 4 \times \left(\frac{t_{0.005,n-1}\, s}{L_0}\right)^2$

$= 4 \times \left(\frac{3.0 \times 0.15}{0.2}\right)^2 = 20.25$.

A sample size slightly larger than 20 should be sufficient.

8.1.9 A total sample size of

$n \geq 4 \times \left(\frac{t_{0.025,n_1-1}\, s}{L_0}\right)^2$

$= 4 \times \left(\frac{2.042 \times 3.05}{2.0}\right)^2 = 38.8$

is required.

Therefore, an additional sample of at least $39 - 31 = 8$ observations should be sufficient.

8.1.10 A total sample size of

$n \geq 4 \times \left(\frac{t_{0.005,n_1-1}\, s}{L_0}\right)^2$

$= 4 \times \left(\frac{2.704 \times 0.124}{0.05}\right)^2 = 179.9$

is required.

Therefore, an additional sample of at least $180 - 41 = 139$ observations should be sufficient.

8.1.11 A total sample size of

$n \geq 4 \times \left(\frac{t_{0.005,n_1-1}\, s}{L_0}\right)^2$

$= 4 \times \left(\frac{2.861 \times 11.90}{10.0}\right)^2 = 46.4$

is required.

Therefore, an additional sample of at least $47 - 20 = 27$ observations should be sufficient.

8.1.12 With $t_{0.05,29} = 1.699$ the value of c is obtained as

$c = \bar{x} + \frac{t_{\alpha,n-1}\, s}{\sqrt{n}} = 14.62 + \frac{1.699 \times 2.98}{\sqrt{30}} = 15.54.$

The confidence interval does not contain the value 16.0, and so it is not plausible that $\mu \geq 16$.

8.1.13 With $t_{0.01,60} = 2.390$ the value of c is obtained as

$c = \bar{x} - \frac{t_{\alpha,n-1}\, s}{\sqrt{n}} = 0.768 - \frac{2.390 \times 0.0231}{\sqrt{61}} = 0.761.$

The confidence interval contains the value 0.765, and so it is plausible that the average solution density is less than 0.765.

8.1.14 With $z_{0.05} = 1.645$ the value of c is obtained as

$c = \bar{x} - \frac{z_\alpha \sigma}{\sqrt{n}} = 11.80 - \frac{1.645 \times 2.0}{\sqrt{19}} = 11.05.$

8.1.15 With $z_{0.01} = 2.326$ the value of c is obtained as

$c = \bar{x} + \frac{z_\alpha \sigma}{\sqrt{n}} = 415.7 + \frac{2.326 \times 10.0}{\sqrt{29}} = 420.0.$

The confidence interval contains the value 418.0, and so it is plausible that the mean radiation level is greater than 418.0.

8.1.16 The interval $(6.668, 7.054)$ is

$(6.861 - 0.193, 6.861 + 0.193)$

and

$0.193 = \frac{1.753 \times 0.440}{\sqrt{16}}.$

Since $1.753 = t_{0.05,15}$ it follows that the confidence level is
$1 - (2 \times 0.05) = 0.90.$

8.1.17 Using the critical point $t_{0.005,9} = 3.250$ the confidence interval is

$\left(2.752 - \frac{3.250 \times 0.280}{\sqrt{10}}, 2.752 + \frac{3.250 \times 0.280}{\sqrt{10}}\right) = (2.464, 3.040).$

The value 3.1 is outside this confidence interval, and so 3.1 is not a plausible value for the average corrosion rate.

Note: The sample statistics for the following problems in this section and the related problems in this chapter depend upon whether any observations have been removed as outliers. To avoid confusion, the answers given here assume that **no** observations have been removed. Notice that removing observations as outliers reduces the sample standard deviation s as well as affecting the sample mean $\bar{x}$.

8.1.18 At 95% confidence the critical point is $t_{0.025,199} = 1.972$ and the confidence interval is

$\left(69.35 - \frac{1.972 \times 17.59}{\sqrt{200}}, 69.35 + \frac{1.972 \times 17.59}{\sqrt{200}}\right) = (66.89, 71.80).$

8.1.19 At 95% confidence the critical point is $t_{0.025,89} = 1.987$ and the confidence interval is

$\left(12.211 - \frac{1.987 \times 2.629}{\sqrt{90}}, 12.211 + \frac{1.987 \times 2.629}{\sqrt{90}}\right) = (11.66, 12.76).$

8.1.20 At 95% confidence the critical point is $t_{0.025,124} = 1.979$ and the confidence interval is

$$\left(1.11059 - \frac{1.979 \times 0.05298}{\sqrt{125}}, 1.11059 + \frac{1.979 \times 0.05298}{\sqrt{125}}\right) = (1.101, 1.120).$$

8.1.21 At 95% confidence the critical point is $t_{0.025,74} = 1.9926$ and the confidence interval is

$$\left(0.23181 - \frac{1.9926 \times 0.07016}{\sqrt{75}}, 0.23181 + \frac{1.9926 \times 0.07016}{\sqrt{75}}\right) = (0.2157, 0.2480).$$

8.1.22 At 95% confidence the critical point is $t_{0.025,79} = 1.9905$ and the confidence interval is

$$\left(9.2294 - \frac{1.9905 \times 0.0942}{\sqrt{80}}, 9.2294 + \frac{1.9905 \times 0.0942}{\sqrt{80}}\right) = (9.0419, 9.4169).$$

8.1.23 Since

$$2.773 = 2.843 - \frac{t_{\alpha,8} \times 0.150}{\sqrt{9}}$$

it follows that $t_{\alpha,8} = 1.40$ so that $\alpha = 0.10$.

Therefore, the confidence level of the confidence interval is 90%.

8.1.24 (a) The sample median is 34.

(b) $\sum_{i=1}^{15} x_i = 532$

$\sum_{i=1}^{15} x_i^2 = 19336$

$\bar{x} = \frac{532}{15} = 35.47$

$s^2 = \frac{19336 - 532^2/15}{15-1} = 33.41$

Using the critical point $z_{0.005} = 2.576$ the confidence interval is

$35.47 \pm \frac{2.576 \times \sqrt{33.41}}{15} = (31.02, 39.91)$.

8.1.25 (a) Using the critical point $t_{0.025,13} = 2.160$ the confidence interval is

$\mu \in 5437.2 \pm \frac{2.160 \times 376.9}{\sqrt{14}} = (5219.6, 5654.8)$.

(b) With

$4 \times \left(\frac{2.160 \times 376.9}{300}\right)^2 = 29.5$

it can be estimated that an additional $30 - 14 = 16$ chemical solutions would need to be measured.

8.1.26 With $t_{0.025,n-1} \simeq 2$ the required sample size can be estimated to be about

$n = 4 \times \left(\frac{t\times\sigma}{L_0}\right)^2$

$= 4 \times \left(\frac{2\times 0.2031}{0.1}\right)^2$

$= 66.$

8.2 Hypothesis Testing

8.2.1 (a) The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{18}\times(57.74-55.0)}{11.2} = 1.04.$

The p-value is $2 \times P(t_{17} \geq 1.04) = 0.313$.

(b) The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{18}\times(57.74-65.0)}{11.2} = -2.75.$

The p-value is $P(t_{17} \leq -2.75) = 0.0068$.

8.2.2 (a) The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{39}\times(5532-5680)}{287.8} = -3.21.$

The p-value is $2 \times P(t_{38} \geq 3.21) = 0.003$.

(b) The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{39}\times(5{,}532-5{,}450)}{287.8} = 1.78.$

The p-value is $P(t_{38} \geq 1.78) = 0.042$.

8.2.3 (a) The test statistic is

$z = \frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma} = \frac{\sqrt{13}\times(2.879-3.0)}{0.325} = -1.34.$

The p-value is $2 \times \Phi(-1.34) = 0.180$.

(b) The test statistic is

$z = \frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma} = \frac{\sqrt{13}\times(2.879-3.1)}{0.325} = -2.45.$

The p-value is $\Phi(-2.45) = 0.007$.

8.2.4 (a) The test statistic is

$z = \frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma} = \frac{\sqrt{44}\times(87.90-90.0)}{5.90} = -2.36.$

The p-value is $2 \times \Phi(-2.36) = 0.018$.

(b) The test statistic is

$z = \frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma} = \frac{\sqrt{44}\times(87.90-86.0)}{5.90} = 2.14.$

The p-value is $1 - \Phi(2.14) = 0.016$.

8.2.5 (a) The critical point is $t_{0.05,40} = 1.684$
and the null hypothesis is accepted when $|t| \leq 1.684$.

(b) The critical point is $t_{0.005,40} = 2.704$
and the null hypothesis is rejected when $|t| > 2.704$.

(c) The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{41}\times(3.04-3.00)}{0.124} = 2.066.$

The null hypothesis is rejected at size $\alpha = 0.10$
and accepted at size $\alpha = 0.01$.

(d) The p-value is $2 \times P(t_{40} \geq 2.066) = 0.045$.

8.2.6 (a) The critical point is $t_{0.05,19} = 1.729$
and the null hypothesis is accepted when $|t| \leq 1.729$.

(b) The critical point is $t_{0.005,19} = 2.861$
and the null hypothesis is rejected when $|t| > 2.861$.

(c) The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{20}\times(436.5-430.0)}{11.90} = 2.443.$

The null hypothesis is rejected at size $\alpha = 0.10$
and accepted at size $\alpha = 0.01$.

(d) The p-value is $2 \times P(t_{19} \geq 2.443) = 0.025$.

8.2.7 (a) The critical point is $t_{0.05,15} = 1.753$
and the null hypothesis is accepted when $|t| \leq 1.753$.

(b) The critical point is $t_{0.005,15} = 2.947$
and the null hypothesis is rejected when $|t| > 2.947$.

(c) The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{16}\times(1.053-1.025)}{0.058} = 1.931.$

The null hypothesis is rejected at size $\alpha = 0.10$
and accepted at size $\alpha = 0.01$.

(d) The p-value is $2 \times P(t_{15} \geq 1.931) = 0.073$.

8.2.8 (a) The critical point is $z_{0.05} = 1.645$
and the null hypothesis is accepted when $|z| \leq 1.645$.

(b) The critical point is $z_{0.005} = 2.576$
and the null hypothesis is rejected when $|z| > 2.576$.

(c) The test statistic is

$z = \frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma} = \frac{\sqrt{10}\times(19.50-20.0)}{1.0} = -1.581.$

The null hypothesis is accepted at size $\alpha = 0.10$
and consequently also at size $\alpha = 0.01$.

(d) The p-value is $2 \times \Phi(-1.581) = 0.114$.

8.2.9 (a) The critical point is $t_{0.10,60} = 1.296$
and the null hypothesis is accepted when $t \leq 1.296$.

(b) The critical point is $t_{0.01,60} = 2.390$
and the null hypothesis is rejected when $t > 2.390$.

(c) The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{61}\times(0.0768-0.065)}{0.0231} = 3.990.$

The null hypothesis is rejected at size $\alpha = 0.01$
and consequently also at size $\alpha = 0.10$.

(d) The p-value is $P(t_{60} \geq 3.990) = 0.0001$.

8.2.10 (a) The critical point is $z_{0.10} = 1.282$
and the null hypothesis is accepted when $z \geq -1.282$.

(b) The critical point is $z_{0.01} = 2.326$
and the null hypothesis is rejected when $z < -2.326$.

(c) The test statistic is

$z = \frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma} = \frac{\sqrt{29}\times(415.7-420.0)}{10.0} = -2.316.$

The null hypothesis is rejected at size $\alpha = 0.10$
and accepted at size $\alpha = 0.01$.

(d) The p-value is $\Phi(-2.316) = 0.0103$.

8.2.11 Consider the hypotheses $H_0 : \mu = 44.350$ versus $H_A : \mu \neq 44.350$.

The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{24}\times(44.364-44.350)}{0.019} = 3.61.$

The p-value is $2 \times P(t_{23} \geq 3.61) = 0.0014$.

There is sufficient evidence to conclude that the machine is miscalibrated.

8.2.12 Consider the hypotheses $H_0 : \mu \leq 120$ versus $H_A : \mu > 120$.

The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{36}\times(122.5-120.0)}{13.4} = 1.12.$

The p-value is $P(t_{35} \geq 1.12) = 0.135$.

There is *not* sufficient evidence to conclude that the manufacturer's claim is incorrect.

8.2.13 Consider the hypotheses $H_0 : \mu \leq 12.50$ versus $H_A : \mu > 12.50$.

The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{15}\times(14.82-12.50)}{2.91} = 3.09.$

The p-value is $P(t_{14} \geq 3.09) = 0.004$.

There is sufficient evidence to conclude that the chemical plant is in violation of the working code.

8.2.14 Consider the hypotheses $H_0 : \mu \geq 0.25$ versus $H_A : \mu < 0.25$.

The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{23}\times(0.228-0.250)}{0.0872} = -1.21.$

The p-value is $P(t_{22} \leq -1.21) = 0.120$.

There is *not* sufficient evidence to conclude that the advertised claim is false.

8.2.15 Consider the hypotheses $H_0 : \mu \leq 2.5$ versus $H_A : \mu > 2.5$.

The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{10}\times(2.752-2.5)}{0.280} = 2.846.$

The p-value is $P(t_9 \geq 2.846) = 0.0096$.

There is sufficient evidence to conclude that the average corrosion rate of chilled cast iron of this type is larger than 2.5.

8.2.16 Consider the hypotheses $H_0 : \mu \leq 65$ versus $H_A : \mu > 65$.

The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{200}\times(69.35-65.00)}{17.59} = 3.50.$

The p-value is $P(t_{199} \geq 3.50) = 0.0003$.

There is sufficient evidence to conclude that the average service time is greater than 65 seconds and that the manager's claim is incorrect.

8.2.17 Consider the hypotheses $H_0 : \mu \geq 13$ versus $H_A : \mu < 13$.

The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{90}\times(12.211-13.000)}{2.629} = -2.85.$

The p-value is $P(t_{89} \leq -2.85) = 0.0027$.

There is sufficient evidence to conclude that the average number of calls taken per minute is less than 13 so that the manager's claim is false.

8.2.18 Consider the hypotheses $H_0 : \mu = 1.1$ versus $H_A : \mu \neq 1.1$.

The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{125}\times(1.11059-1.10000)}{0.05298} = 2.23.$

The p-value is $2 \times P(t_{124} \geq 2.23) = 0.028$.

There is some evidence that the manufacturing process needs adjusting but it is not overwhelming.

8.2.19 Consider the hypotheses $H_0 : \mu = 0.2$ versus $H_A : \mu \neq 0.2$.

The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{75}\times(0.23181-0.22500)}{0.07016} = 0.841.$

The p-value is $2 \times P(t_{74} \geq 0.841) = 0.40$.

There is not sufficient evidence to conclude that the spray painting machine is not performing properly.

8.2.20 Consider the hypotheses $H_0 : \mu \geq 9.5$ versus $H_A : \mu < 9.5$.

The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{80}\times(9.2294-9.5000)}{0.8423} = -2.87.$

The p-value is $P(t_{79} \leq -2.87) = 0.0026$.

There is sufficient evidence to conclude that the design criterion has not been met.

8.2.21 The hypotheses are $H_0 : \mu \leq 238.5$ versus $H_A : \mu > 238.5$

and the test statistic is

$t = \frac{\sqrt{16}(239.13-238.50)}{2.80} = 0.90.$

The p-value is $P(t_{15} > 0.90) = 0.191.$

There is not sufficient evidence to conclude that the average voltage of the batteries from the production line is at least 238.5.

8.2.22 (a) $0.002 \leq 2 \times P(t_{11} > 3.21) \leq 0.01$

(b) $0.05 \leq 2 \times P(t_{23} > 1.96) \leq 0.10$

(c) $2 \times P(t_{29} > 3.88) \leq 0.001$

8.2.23 The hypotheses are $H_0 : \mu = 82.50$ versus $H_A : \mu \neq 82.50$

and the test statistic is

$t = \frac{\sqrt{25}(82.40-82.50)}{0.14} = -3.571.$

The p-value is $2 \times P(t_{24} > 3.571) = 0.0015.$

There is sufficient evidence to conclude that the average length of the components is not 82.50.

8.2.24 The hypotheses are $H_0 : \mu \leq 70$ versus $H_A : \mu > 70$

and the test statistic is

$t = \frac{\sqrt{25}(71.97-70)}{7.44} = 1.324.$

The p-value is $P(t_{24} > 1.324) = 0.099.$

There is some evidence to conclude that the components have an average weight larger than 70, but the evidence is not overwhelming.

8.2.25 The hypotheses are $H_0 : \mu = 7.000$ versus $H_A : \mu \neq 7.000$

and the test statistic is

$t = \frac{\sqrt{28}(7.442-7.000)}{0.672} = 3.480.$

The p-value is $2 \times P(t_{27} > 3.480) = 0.002.$

There is sufficient evidence to conclude that the average breaking strength is not 7.000.

8.2.26 The hypotheses are $H_0 : \mu \leq 50$ versus $H_A : \mu > 50$

and the test statistic is

$t = \frac{\sqrt{25}(53.43-50)}{3.93} = 4.364.$

The p-value is $P(t_{24} > 4.364) = 0.0001.$

There is sufficient evidence to conclude that average failure time of this kind of component is at least 50 hours.

8.2.27 The hypotheses are $H_0 : \mu \geq 25$ versus $H_A : \mu < 25$.

8.2.28 (a) The t-statistic is

$t = \frac{\sqrt{20}(12.49-10)}{1.32} = 8.44$

and the p-value is $2 \times P(t_{19} > 8.44)$ which is less than 1%.

(b) The t-statistic is

$t = \frac{\sqrt{43}(3.03-3.2)}{0.11} = -10.13$

and the p-value is $P(t_{42} > -10.13)$ which is greater than 10%.

(c) The t-statistic is

$t = \frac{\sqrt{16}(73.43-85)}{16.44} = -2.815$

and the p-value is $P(t_{15} < -2.815)$ which is less than 1%.

8.2.29 (a) The sample mean is $\bar{x} = 11.975$

and the sample standard deviation is $s = 2.084$

so that the t-statistic is

$t = \frac{\sqrt{8}(11.975-11)}{2.084} = 1.32.$

The p-value is $P(t_7 > 1.32)$ which is greater than 10%.

Consequently, the experiment does not provide sufficient evidence to conclude that the average time to toxicity of salmon fillets under these storage conditions is more than 11 days.

(b) With $t_{0.005,7} = 3.499$ the confidence interval is

$11.975 \pm \frac{3.499 \times 2.084}{\sqrt{8}} = (9.40, 14.55).$

8.5 Supplementary Problems

8.5.1 (a) Consider the hypotheses $H_0 : \mu \leq 65$ versus $H_A : \mu > 65$.
The test statistic is
$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{15}\times(67.42-65.00)}{4.947} = 1.89$.
The p-value is $P(t_{14} \geq 1.89) = 0.040$.
There is some evidence that the average distance at which the target is detected is at least 65 miles although the evidence is not overwhelming.

(b) With $t_{0.01,14} = 2.624$ the confidence interval is
$\left(67.42 - \frac{2.624\times4.947}{\sqrt{15}}, \infty\right) = (64.07, \infty)$.

8.5.2 (a) Consider the hypotheses $H_0 : \mu \geq 10$ versus $H_A : \mu < 10$.
The test statistic is
$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{40}\times(9.39-10.00)}{1.041} = -3.71$.
The p-value is $P(t_{39} \leq -3.71) = 0.0003$.
The company can safely conclude that the telephone surveys will last on average less than ten minutes each.

(b) With $t_{0.01,39} = 2.426$ the confidence interval is
$\left(-\infty, 9.39 + \frac{2.426\times1.041}{\sqrt{40}}\right) = (-\infty, 9.79)$.

8.5.3 (a) Consider the hypotheses $H_0 : \mu = 75.0$ versus $H_A : \mu \neq 75.0$.
The test statistic is
$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{30}\times(74.63-75.00)}{2.095} = -0.1766$.
The p-value is $2 \times P(t_{29} \geq 0.1766) = 0.861$.
There is not sufficient evidence to conclude that the paper does not have an average weight of 75.0 g/m^2.

(b) With $t_{0.005,29} = 2.756$ the confidence interval is
$\left(74.63 - \frac{2.756\times2.095}{\sqrt{30}}, 74.63 + \frac{2.756\times2.095}{\sqrt{30}}\right) = (73.58, 75.68)$.

(c) A total sample size of
$n \geq 4 \times \left(\frac{t_{0.005,n_1-1}\, s}{L_0}\right)^2 = 4 \times \left(\frac{2.756\times2.095}{1.5}\right)^2 = 59.3$
is required.
Therefore, an additional sample of at least $60 - 30 = 30$ observations should be sufficient.

8.5.4 (a) Consider the hypotheses $H_0 : \mu \geq 0.50$ versus $H_A : \mu < 0.50$.
The test statistic is
$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{14}\times(0.497-0.500)}{0.0764} = -0.147$.
The p-value is $P(t_{13} \leq -0.147) = 0.443$.
There is not sufficient evidence to establish that the average deformity value of diseased arteries is less than 0.50.

(b) With $t_{0.005,13} = 3.012$ the confidence interval is
$\left(0.497 - \frac{3.012\times0.0764}{\sqrt{14}}, 0.497 + \frac{3.012\times0.0764}{\sqrt{14}}\right) = (0.435, 0.559)$.

(c) A total sample size of
$n \geq 4 \times \left(\frac{t_{0.005,n_1-1}\, s}{L_0}\right)^2 = 4 \times \left(\frac{3.012\times0.0764}{0.10}\right)^2 = 21.2$
is required.
Therefore, an additional sample of at least $22 - 14 = 8$ observations should be sufficient.

8.5.5 At a 90% confidence level the critical point is $t_{0.05,59} = 1.671$ and the confidence interval is

$\left(69.618 - \frac{1.671\times1.523}{\sqrt{60}}, 69.618 + \frac{1.671\times1.523}{\sqrt{60}}\right) = (69.29, 69.95)$.

At a 95% confidence level the critical point is $t_{0.025,59} = 2.001$ and the confidence interval is

$\left(69.618 - \frac{2.001\times1.523}{\sqrt{60}}, 69.618 + \frac{2.001\times1.523}{\sqrt{60}}\right) = (69.23, 70.01)$.

At a 99% confidence level the critical point is $t_{0.005,59} = 2.662$ and the confidence interval is

$\left(69.618 - \frac{2.662\times1.523}{\sqrt{60}}, 69.618 + \frac{2.662\times1.523}{\sqrt{60}}\right) = (69.10, 70.14)$.

There is not strong evidence that 70 inches is not a plausible value for the mean height because it is included in the 95% confidence level confidence interval.

8.5.6 At a 90% confidence level the critical point is $t_{0.05,39} = 1.685$ and the confidence interval is

$\left(32.042 - \frac{1.685\times5.817}{\sqrt{40}}, 32.042 + \frac{1.685\times5.817}{\sqrt{40}}\right) = (30.49, 33.59)$.

At a 95% confidence level the critical point is $t_{0.025,39} = 2.023$ and the confidence interval is

$\left(32.042 - \frac{2.023\times 5.817}{\sqrt{40}}, 32.042 + \frac{2.023\times 5.817}{\sqrt{40}}\right) = (30.18, 33.90)$.

At a 99% confidence level the critical point is $t_{0.005,39} = 2.708$ and the confidence interval is

$\left(32.042 - \frac{2.708\times 5.817}{\sqrt{40}}, 32.042 + \frac{2.708\times 5.817}{\sqrt{40}}\right) = (29.55, 34.53)$.

Since 35 and larger values are not contained within the 99% confidence level confidence interval they are not plausible values for the mean shoot height, and so these new results contradict the results of the previous study.

8.5.7 The interval $(472.56, 486.28)$ is

$(479.42 - 6.86, 479.42 + 6.86)$

and

$6.86 = \frac{2.787\times 12.55}{\sqrt{26}}$.

Since $2.787 = t_{0.005,25}$ it follows that the confidence level is

$1 - (2 \times 0.005) = 0.99$.

8.5.8 (a) Consider the hypotheses $H_0 : \mu \geq 0.36$ versus $H_A : \mu < 0.36$.
The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{18}\times(0.337-0.36)}{0.025} = -3.903$.

The p-value is $P(t_{17} \leq -3.903) = 0.0006$.
There is sufficient evidence to conclude that the average weight gain for composites of this kind is smaller than 0.36%.

(b) Using the critical point $t_{0.01,17} = 2.567$ the confidence interval is

$\left(-\infty, 0.337 + \frac{2.567\times 0.025}{\sqrt{18}}\right) = (-\infty, 0.352)$.

8.5.9 Using the critical point $t_{0.01,43} = 2.416$ the confidence interval is

$\left(-\infty, 25.318 + \frac{2.416\times 0.226}{\sqrt{44}}\right) = (-\infty, 25.400)$.

Consider the hypotheses $H_0 : \mu \geq 25.5$ versus $H_A : \mu < 25.5$.

The test statistic is

$t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{44}\times(25.318-25.5)}{0.226} = -5.342$.

The p-value is $P(t_{43} \leq -5.342) = 0.000$.

There is sufficient evidence to conclude that the average soil compressibility is no larger than 25.5.

8.5.11 At a 95% confidence level the critical points are

$\chi^2_{0.025,17} = 30.19$ and $\chi^2_{0.975,17} = 7.564$

so that the confidence interval is

$\left(\frac{(18-1)\times 6.48^2}{30.19}, \frac{(18-1)\times 6.48^2}{7.564}\right) = (23.6, 94.4).$

At a 99% confidence level the critical points are

$\chi^2_{0.005,17} = 35.72$ and $\chi^2_{0.995,17} = 5.697$

so that the confidence interval is

$\left(\frac{(18-1)\times 6.48^2}{35.72}, \frac{(18-1)\times 6.48^2}{5.697}\right) = (20.0, 125.3).$

8.5.12 At a 99% confidence level the critical points are

$\chi^2_{0.005,40} = 66.77$ and $\chi^2_{0.995,40} = 20.71$

so that the confidence interval is

$\left(\sqrt{\frac{(41-1)\times 0.124^2}{66.77}}, \sqrt{\frac{(41-1)\times 0.124^2}{20.71}}\right) = (0.095, 0.170).$

8.5.13 At a 95% confidence level the critical points are

$\chi^2_{0.025,19} = 32.85$ and $\chi^2_{0.975,19} = 8.907$

so that the confidence interval is

$\left(\frac{(20-1)\times 11.90^2}{32.85}, \frac{(20-1)\times 11.90^2}{8.907}\right) = (81.9, 302.1).$

8.5.14 At a 90% confidence level the critical points are

$\chi^2_{0.05,15} = 25.00$ and $\chi^2_{0.95,15} = 7.261$

so that the confidence interval is

$\left(\sqrt{\frac{(16-1)\times 0.058^2}{25.00}}, \sqrt{\frac{(16-1)\times 0.058^2}{7.261}}\right) = (0.045, 0.083).$

At a 95% confidence level the critical points are

$\chi^2_{0.025,15} = 27.49$ and $\chi^2_{0.975,15} = 6.262$

so that the confidence interval is

$\left(\sqrt{\frac{(16-1)\times 0.058^2}{27.49}}, \sqrt{\frac{(16-1)\times 0.058^2}{6.262}}\right) = (0.043, 0.090).$

At a 99% confidence level the critical points are

$\chi^2_{0.005,15} = 32.80$ and $\chi^2_{0.995,15} = 4.601$

so that the confidence interval is

$\left(\sqrt{\frac{(16-1)\times 0.058^2}{32.80}}, \sqrt{\frac{(16-1)\times 0.058^2}{4.601}}\right) = (0.039, 0.105).$

8.5.15 (a) The p-value is $2 \times P(t_7 > 1.31)$ which is more than 0.20.

(b) The p-value is $2 \times P(t_{29} > 2.82)$ which is between 0.002 and 0.01.

(c) The p-value is $2 \times P(t_{24} > 1.92)$ which is between 0.05 and 0.10.

8.5.16 The hypotheses are $H_0 : \mu \geq 81$ versus $H_A : \mu < 81$ and the test statistic is

$t = \frac{\sqrt{16}\times(76.99-81.00)}{5.37} = -2.987$

so that the p-value is $P(t_{15} \leq -2.987) = 0.005$.

There is sufficient evidence to conclude that the average clay compressibility at the location is less than 81.

8.5.17 The hypotheses are $H_0 : \mu \leq 260.0$ versus $H_A : \mu > 260.0$ and the test statistic is

$t = \frac{\sqrt{14}\times(266.5-260.0)}{18.6} = 1.308$

so that the p-value is $P(t_{13} \geq 1.308) = 0.107$.

There is not sufficient evidence to conclude that the average strength of fibers of this type is at least 260.0.

8.5.18 (a) $n = 18$

(b) $\frac{50+52}{2} = 51$

(c) $\bar{x} = 54.61$

(d) $s = 19.16$

(e) $s^2 = 367.07$

(f) $\frac{s}{\sqrt{n}} = 4.52$

(g) With $t_{0.01,17} = 2.567$ the confidence interval is

$\mu \in \left(54.61 - \frac{2.567\times 19.16}{\sqrt{18}}, \infty\right) = (43.02, \infty).$

(h) The test statistic is

$t = \frac{\sqrt{18}(54.61-50)}{19.16} = 1.021$

and the p-value is $2 \times P(t_{17} \geq 1.021)$.

The critical points in Table III imply that the p-value is larger than 0.20.

8.5.19 (a) True

(b) False

(c) True

(d) True

(e) True

(f) True

(g) True

8.5.20 The hypotheses are $H_0 : \mu = 200.0$ versus $H_A : \mu \neq 200.0$ and the test statistic is

$t = \frac{\sqrt{22}\times(193.7-200)}{11.2} = -2.639$

so that the p-value is $2 \times P(t_{21} \geq 2.639) = 0.015$.

There is some evidence that the average resistance of wires of this type is not 200.0 but the evidence is not overwhelming.

8.5.21 (a) Since

$L = 74.5 - 72.3 = 2.2 = 2 \times \frac{t_{0.005,9}s}{\sqrt{10}} = 2 \times \frac{3.250\times s}{\sqrt{10}}$

it follows that $s = 1.070$.

(b) Since

$4 \times \frac{3.250^2\times 1.070^2}{1^2} = 48.4$

it can be estimated that a further sample of size $49 - 10 = 39$ will be sufficient.

8.5.22 (a) The hypotheses are $H_0 : \mu = 600$ versus $H_A : \mu \neq 600$ and the test statistic is

$t = \frac{\sqrt{10}(614.5-600)}{42.9} = 1.069$

The p-value is $2 \times P(t_9 \geq 1.069) = 0.313$.

There is not sufficient evidence to establish that the population average is not 600.

(b) With $t_{0.01,9} = 3.250$ the confidence interval is

$\mu \in 614.5 \pm \frac{3.250\times 42.9}{\sqrt{10}} = (570.4, 658.6)$

(c) With

$4 \times \left(\frac{3.250\times 42.9}{30}\right)^2 = 86.4$

it can be estimated that about $87 - 10 = 77$ more items would need to be sampled.

8.5.23 (a) The hypotheses are $H_0 : \mu \geq 750$ versus $H_A : \mu < 750$ and the t-statistic is

$t = \frac{\sqrt{12}(732.9-750)}{12.5} = -4.74$

so that the p-value is $P(t_{11} < -4.74) = 0.0003$.
There is sufficient evidence to conclude that the flexibility of this kind of metal alloy is smaller than 750.

(b) With $t_{0.01,11} = 2.718$ the confidence interval is

$\left(-\infty, 732.9 + \frac{2.718\times 12.5}{\sqrt{12}}\right) = (-\infty, 742.7)$.

8.5.24 (a) $\bar{x} = \frac{\sum_{i=1}^{9} x_i}{9} = \frac{4047.4}{9} = 449.71$

(b) The ordered data are: 402.9 418.4 423.6 442.3 453.2 459 477.7 483 487.3
Therefore, the sample median is 453.2.

(c) $s^2 = \frac{(\sum_{i=1}^{9} x_i^2)-(\sum_{i=1}^{9} x_i)^2/9}{8} = 913.9111$

$s = 30.23$

(d) $\left(\bar{x} - \frac{t_{0.005,8}\times 30.23}{\sqrt{9}}, \bar{x} + \frac{t_{0.005,8}\times 30.23}{\sqrt{9}}\right)$

$= \left(449.71 - \frac{3.355\times 30.23}{3}, 449.71 + \frac{3.355\times 30.23}{3}\right)$

$= (415.9, 483.52)$

(e) $\left(-\infty, \bar{x} + \frac{t_{0.05,8}\times 30.23}{\sqrt{9}}\right)$

$= \left(-\infty, 449.71 + \frac{1.86\times 30.23}{3}\right)$

$= (-\infty, 468.45)$

(f) $n \geq 4 \times \left(\frac{t_{0.005,8}\times 30.23}{L_0}\right)^2$

$= 4 \times \left(\frac{3.355\times 30.23}{50}\right)^2$

$= 16.459 \simeq 17$
Therefore, $17 - 9 = 8$ additional samples should be sufficient.

(g) The hypotheses are $H_0 : \mu = 440$ versus $H_A : \mu \neq 440$.
The t-statistic is
$t = \frac{\sqrt{9}\times(449.71-440)}{30.23} = 0.9636$
so that the p-value is $2 \times P(t_8 \geq 0.9636) > 0.2$.
This large p-value indicates that H_0 should be accepted.

(h) The hypotheses are $H_0 : \mu \geq 480$ versus $H_A : \mu < 480$.
The t-statistic is
$t = \frac{\sqrt{9}\times(449.71-480)}{30.23} = -3.006$
so that the p-value is $P(t_8 \leq -3.006) < 0.01$.
This small p-value indicates that H_0 should be rejected.

8.5.25 (a) The sample mean is $\bar{x} = 3.669$
and the sample standard deviation is $s = 0.2531$.
The hypotheses are $H_0 : \mu \leq 3.50$ versus $H_A : \mu > 3.50$ and the t-statistic is
$t = \frac{\sqrt{8}(3.669-3.50)}{0.2531} = 1.89$
so that the p-value is $P(t_7 \geq 1.89) = 0.51$.
There is some evidence to establish that the average density of these kind of compounds is larger than 3.50, but the evidence is not overwhelming.

(b) With $t_{0.01,7} = 2.998$ the confidence interval is
$\mu \in \left(3.669 - \frac{0.2531\times2.998}{\sqrt{8}}, \infty\right) = (3.40, \infty)$.

8.5.26 (a) The hypotheses are $H_0 : \mu = 385$ versus $H_A : \mu \neq 385$ and the t-statistic is
$t = \frac{\sqrt{33}(382.97-385.00)}{3.81} = -3.06$
so that the p-value is $2 \times P(t_{32} \geq 3.06) = 0.004$.
There is sufficient evidence to establish that the population mean is not 385.

(b) With $t_{0.005,32} = 2.738$ the confidence interval is
$\mu \in \left(382.97 - \frac{3.81\times2.738}{\sqrt{33}}, 382.97 + \frac{3.81\times2.738}{\sqrt{33}}\right) = (381.1, 384.8)$.

8.5.27 (a) The t-statistic is
$t = \frac{\sqrt{24}\times(2.39-2.5)}{0.21} = -2.566$

so that the p-value is $2 \times P(t_{23} \geq |-2.566|)$.
The critical points in Table III imply that the p-value is between 0.01 and 0.02.

(b) The t-statistic is

$t = \frac{\sqrt{30}\times(0.538-0.54)}{0.026} = -0.421$

so that the p-value is $P(t_{29} \leq -0.421)$.
The critical points in Table III imply that the p-value is larger than 0.1.

(c) The t-statistic is

$t = \frac{\sqrt{10}\times(143.6-135)}{4.8} = 5.67$

so that the p-value is $P(t_9 \geq 5.67)$.
The critical points in Table III imply that the p-value is smaller than 0.0005.

Chapter 9

Comparing Two Population Means

9.2 Analysis of Paired Samples

9.2.1 The differences $z_i = x_i - y_i$ have a sample mean $\bar{z} = 7.12$ and a sample standard deviation $s = 34.12$.

Consider the hypotheses

$H_0 : \mu = \mu_A - \mu_B \leq 0$ versus $H_A : \mu = \mu_A - \mu_B > 0$

where the alternative hypothesis states that the new assembly method is quicker on average than the standard assembly method.

The test statistic is

$t = \frac{\sqrt{n}\,\bar{z}}{s} = \frac{\sqrt{35}\times 7.12}{34.12} = 1.23.$

The p-value is $P(t_{34} \geq 1.23) = 0.114$.

There is *not* sufficient evidence to conclude that the new assembly method is any quicker on average than the standard assembly method.

With $t_{0.05,34} = 1.691$ a one-sided 95% confidence level confidence interval for $\mu = \mu_A - \mu_B$ is

$\left(7.12 - \frac{1.691\times 34.12}{\sqrt{35}}, \infty\right) = (-2.63, \infty).$

9.2.2 The differences $z_i = x_i - y_i$ have a sample mean $\bar{z} = -1.36$ and a sample standard deviation $s = 6.08$.

Consider the hypotheses

$H_0 : \mu = \mu_A - \mu_B = 0$ versus $H_A : \mu = \mu_A - \mu_B \neq 0.$

The test statistic is

$t = \frac{\sqrt{n}\,\bar{z}}{s} = \frac{\sqrt{14}\times(-1.36)}{6.08} = -0.837.$

The p-value is $2 \times P(t_{13} \leq -0.837) = 0.418$.

There is *not* sufficient evidence to conclude that the different stimulation conditions affect the adhesion of the red blood cells.

With $t_{0.025,13} = 2.160$ a two-sided 95% confidence level confidence interval

for $\mu = \mu_A - \mu_B$ is

$\left(-1.36 - \frac{2.160\times6.08}{\sqrt{14}}, -1.36 + \frac{2.160\times6.08}{\sqrt{14}}\right) = (-4.87, 2.15)$.

9.2.3 The differences $z_i = x_i - y_i$ have a sample mean $\bar{z} = 0.570$ and a sample standard deviation $s = 0.813$.

Consider the hypotheses

$H_0 : \mu = \mu_A - \mu_B \leq 0$ versus $H_A : \mu = \mu_A - \mu_B > 0$

where the alternative hypothesis states that the new tires have a smaller average reduction in tread depth than the standard tires.

The test statistic is

$t = \frac{\sqrt{n}\,\bar{z}}{s} = \frac{\sqrt{20}\times0.570}{0.813} = 3.14.$

The p-value is $P(t_{19} \geq 3.14) = 0.003$.

There is sufficient evidence to conclude that the new tires are better than the standard tires in terms of the average reduction in tread depth.

With $t_{0.05,19} = 1.729$ a one-sided 95% confidence level confidence interval

for $\mu = \mu_A - \mu_B$ is

$\left(0.570 - \frac{1.729\times0.813}{\sqrt{20}}, \infty\right) = (0.256, \infty)$.

9.2.4 The differences $z_i = x_i - y_i$ have a sample mean $\bar{z} = -7.70$ and a sample standard deviation $s = 14.64$.

Consider the hypotheses

$H_0 : \mu = \mu_A - \mu_B \geq 0$ versus $H_A : \mu = \mu_A - \mu_B < 0$

where the alternative hypothesis states that the new teaching method produces higher scores on average than the standard teaching method.

The test statistic is

$t = \frac{\sqrt{n}\,\bar{z}}{s} = \frac{\sqrt{40}\times(-7.70)}{14.64} = -3.33.$

The p-value is $P(t_{39} \leq -3.33) = 0.001$.

There is sufficient evidence to conclude that the new teaching method is better since it produces higher scores on average than the standard teaching method.

With $t_{0.05,39} = 1.685$ a one-sided 95% confidence level confidence interval for $\mu = \mu_A - \mu_B$ is

$\left(-\infty, -7.70 + \frac{1.685 \times 14.64}{\sqrt{40}}\right) = (-\infty, -3.80)$.

9.2.5 The differences $z_i = x_i - y_i$ have a sample mean $\bar{z} = 2.20$ and a sample standard deviation $s = 147.8$.

Consider the hypotheses

$H_0 : \mu = \mu_A - \mu_B = 0$ versus $H_A : \mu = \mu_A - \mu_B \neq 0$.

The test statistic is

$t = \frac{\sqrt{n}\,\bar{z}}{s} = \frac{\sqrt{18} \times 2.20}{147.8} = 0.063$.

The p-value is $2 \times P(t_{17} \geq 0.063) = 0.95$.

There is *not* sufficient evidence to conclude that the two laboratories are any different in the datings that they provide.

With $t_{0.025,17} = 2.110$ a two-sided 95% confidence level confidence interval

for $\mu = \mu_A - \mu_B$ is

$\left(2.20 - \frac{2.110 \times 147.8}{\sqrt{18}}, 2.20 + \frac{2.110 \times 147.8}{\sqrt{18}}\right) = (-71.3, 75.7)$.

9.2.6 The differences $z_i = x_i - y_i$ have a sample mean $\bar{z} = -1.42$ and a sample standard deviation $s = 12.74$.

Consider the hypotheses

$H_0 : \mu = \mu_A - \mu_B \geq 0$ versus $H_A : \mu = \mu_A - \mu_B < 0$

where the alternative hypothesis states that the new golf balls travel further on average than the standard golf balls.

The test statistic is

$t = \frac{\sqrt{n}\,\bar{z}}{s} = \frac{\sqrt{24} \times (-1.42)}{12.74} = -0.546$.

The p-value is $P(t_{23} \leq -0.546) = 0.30$.

There is *not* sufficient evidence to conclude that the new golf balls travel further on average than the standard golf balls.

With $t_{0.05,23} = 1.714$ a one-sided 95% confidence level confidence interval

for $\mu = \mu_A - \mu_B$ is

$\left(-\infty, -1.42 + \frac{1.714 \times 12.74}{\sqrt{24}}\right) = (-\infty, 3.04)$.

9.2.7 The differences $z_i = x_i - y_i$ have a sample mean $\bar{z} = -2.800$ and a sample standard deviation $s = 6.215$.

The hypotheses are

$H_0 : \mu = \mu_A - \mu_B = 0$ versus $H_A : \mu = \mu_A - \mu_B \neq 0$

and the test statistic is

$t = \frac{\sqrt{10}\times(-2.800)}{6.215} = -1.425.$

The p-value is $2 \times P(t_9 \geq 1.425) = 0.188$.

There is not sufficient evidence to conclude that procedures A and B give different readings on average.

The reviewer's comments are plausible.

9.2.8 The differences $z_i = x_i - y_i$ have a sample mean $\bar{z} = 1.375$ and a sample standard deviation $s = 1.785$.

Consider the hypotheses

$H_0 : \mu = \mu_S - \mu_N \leq 0$ versus $H_A : \mu = \mu_S - \mu_N > 0$

where the alternative hypothesis states that the new antibiotic is quicker than the standard antibiotic.

The test statistic is

$t = \frac{\sqrt{n}\,\bar{z}}{s} = \frac{\sqrt{8}\times 1.375}{1.785} = 2.18.$

The p-value is $P(t_7 \geq 2.18) = 0.033$.

Consequently, there is some evidence that the new antibiotic is quicker than the standard antibiotic, but the evidence is not overwhelming.

9.2.9 The differences $z_i = x_i - y_i$ have a sample mean $\bar{z} = 0.85$ and a sample standard deviation $s = 4.283$.

Consider the hypotheses

$H_0 : \mu = \mu_A - \mu_B = 0$ versus $H_A : \mu = \mu_A - \mu_B \neq 0$

where the alternative hypothesis states that the addition of the surfactant has an effect on the amount of uranium-oxide removed from the water.

The test statistic is

$t = \frac{\sqrt{n}\,\bar{z}}{s} = \frac{\sqrt{6}\times 0.85}{4.283} = 0.486.$

The p-value is $2 \times P(t_5 \geq 0.486) = 0.65$.

Consequently, there is *not* sufficient evidence to conclude that the addition of the surfactant has an effect on the amount of uranium-oxide removed from the water.

9.3 Analysis of Independent Samples

9.3.2 (a) The pooled variance is

$s_p^2 = \frac{(n-1)s_x^2+(m-1)s_y^2}{n+m-2} = \frac{((14-1)\times 4.30^2)+((14-1)\times 5.23^2)}{14+14-2} = 22.92.$

With $t_{0.005,26} = 2.779$ a 99% two-sided confidence interval for $\mu_A - \mu_B$ is

$32.45 - 41.45 \pm 2.779 \times \sqrt{22.92} \times \sqrt{\frac{1}{14} + \frac{1}{14}}$

$= (-14.03, -3.97).$

(b) Since

$$\frac{\left(\frac{4.30^2}{14}+\frac{5.23^2}{14}\right)^2}{\frac{4.30^4}{14^2\times(14-1)}+\frac{5.23^4}{14^2\times(14-1)}} = 25.06$$

the degrees of freedom are $\nu = 25$.

Using a critical point $t_{0.005,25} = 2.787$
a 99% two-sided confidence interval for $\mu_A - \mu_B$ is

$32.45 - 41.45 \pm 2.787 \times \sqrt{\frac{4.30^2}{14} + \frac{5.23^2}{14}}$

$= (-14.04, -3.96).$

(c) The test statistic is

$t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}} = \frac{32.45-41.45}{\sqrt{\frac{4.30^2}{14}+\frac{5.23^2}{14}}} = 4.97.$

The null hypothesis is rejected since $|t| = 4.97$ is larger than the critical point $t_{0.005,26} = 2.779$.
The p-value is $2 \times P(t_{26} \geq 4.97) = 0.000$.

9.3.3 (a) The pooled variance is

$s_p^2 = \frac{(n-1)s_x^2+(m-1)s_y^2}{n+m-2} = \frac{((8-1)\times 44.76^2)+((17-1)\times 38.94^2)}{8+17-2} = 1664.6.$

With $t_{0.005,23} = 2.807$ a 99% two-sided confidence interval for $\mu_A - \mu_B$ is

$675.1 - 702.4 \pm 2.807 \times \sqrt{1664.6} \times \sqrt{\frac{1}{8} + \frac{1}{17}}$

$= (-76.4, 21.8).$

(b) Since

$$\frac{\left(\frac{44.76^2}{8}+\frac{38.94^2}{17}\right)^2}{\frac{44.76^4}{8^2\times(8-1)}+\frac{38.94^4}{17^2\times(17-1)}} = 12.2$$

the degrees of freedom are $\nu = 12$.

Using a critical point $t_{0.005,12} = 3.055$
a 99% two-sided confidence interval for $\mu_A - \mu_B$ is

$675.1 - 702.4 \pm 3.055 \times \sqrt{\frac{44.76^2}{8} + \frac{38.94^2}{17}}$

$= (-83.6, 29.0)$.

(c) The test statistic is

$t = \frac{\bar{x}-\bar{y}}{s_p\sqrt{\frac{1}{n}+\frac{1}{m}}} = \frac{675.1-702.4}{\sqrt{1664.6}\times\sqrt{\frac{1}{8}+\frac{1}{17}}} = -1.56.$

The null hypothesis is accepted since $|t| = 1.56$ is smaller than the critical point $t_{0.005,23} = 2.807$.
The p-value is $2 \times P(t_{23} \geq 1.56) = 0.132$.

9.3.4 (a) Since

$$\frac{\left(\frac{1.07^2}{10}+\frac{0.62^2}{9}\right)^2}{\frac{1.07^4}{10^2\times(10-1)}+\frac{0.62^4}{9^2\times(9-1)}} = 14.7$$

the degrees of freedom are $\nu = 14$.

Using a critical point $t_{0.01,14} = 2.624$
a 99% one-sided confidence interval for $\mu_A - \mu_B$ is

$\left(7.76 - 6.88 - 2.624 \times \sqrt{\frac{1.07^2}{10} + \frac{0.62^2}{9}}, \infty\right)$

$= (-0.16, \infty)$.

(b) The value of c increases with a confidence level of 95%.

(c) The test statistic is

$t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}} = \frac{7.76-6.88}{\sqrt{\frac{1.07^2}{10}+\frac{0.62^2}{9}}} = 2.22.$

The null hypothesis is accepted since
$t = 2.22 \leq t_{0.01,14} = 2.624$.
The p-value is $P(t_{14} \geq 2.22) = 0.022$.

9.3.5 (a) The pooled variance is

$s_p^2 = \frac{(n-1)s_x^2+(m-1)s_y^2}{n+m-2} = \frac{((13-1)\times 0.00128^2)+((15-1)\times 0.00096^2)}{13+15-2}$

$= 1.25 \times 10^{-6}$.

With $t_{0.05,26} = 1.706$ a 95% one-sided confidence interval for $\mu_A - \mu_B$ is

$\left(-\infty, 0.0548 - 0.0569 + 1.706 \times \sqrt{1.25 \times 10^{-6}} \times \sqrt{\frac{1}{13} + \frac{1}{15}}\right)$

$= (-\infty, -0.0014)$.

(b) The test statistic is

$t = \frac{\bar{x}-\bar{y}}{s_p\sqrt{\frac{1}{n}+\frac{1}{m}}} = \frac{0.0548-0.0569}{\sqrt{1.25\times 10^{-6}}\times\sqrt{\frac{1}{13}+\frac{1}{15}}} = -4.95.$

The null hypothesis is rejected at size $\alpha = 0.01$ since $t = -4.95 < -t_{0.01,26} = -2.479$.
The null hypothesis is consequently also rejected at size $\alpha = 0.05$.
The p-value is $P(t_{26} \leq -4.95) = 0.000$.

9.3.6 (a) The pooled variance is

$s_p^2 = \frac{(n-1)s_x^2+(m-1)s_y^2}{n+m-2} = \frac{((41-1)\times 0.124^2)+((41-1)\times 0.137^2)}{41+41-2} = 0.01707.$

The test statistic is

$t = \frac{\bar{x}-\bar{y}}{s_p\sqrt{\frac{1}{n}+\frac{1}{m}}} = \frac{3.04-3.12}{\sqrt{0.01707}\times\sqrt{\frac{1}{41}+\frac{1}{41}}} = -2.77.$

The null hypothesis is rejected at size $\alpha = 0.01$ since $|t| = 2.77$ is larger than $t_{0.005,80} = 2.639$.
The p-value is $2 \times P(t_{80} \leq -2.77) = 0.007$.

(b) With $t_{0.005,80} = 2.639$ a 99% two-sided confidence interval for $\mu_A - \mu_B$ is

$3.04 - 3.12 \pm 2.639 \times \sqrt{0.01707} \times \sqrt{\frac{1}{41} + \frac{1}{41}}$

$= (-0.156, -0.004)$.

(c) There is sufficient evidence to conclude that the average thicknesses of sheets produced by the two processes are different.

9.3.7 (a) Since

$$\frac{\left(\frac{11.90^2}{20}+\frac{4.61^2}{25}\right)^2}{\frac{11.90^4}{20^2\times(20-1)}+\frac{4.61^4}{25^2\times(25-1)}} = 23.6$$

the degrees of freedom are $\nu = 23$.

Consider the hypotheses
$H_0 : \mu = \mu_A - \mu_B \geq 0$ versus $H_A : \mu = \mu_A - \mu_B < 0$
where the alternative hypothesis states that the synthetic fiber bundles have an average breaking strength larger than the wool fiber bundles.

The test statistic is

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}} = \frac{436.5-452.8}{\sqrt{\frac{11.90^2}{20}+\frac{4.61^2}{25}}} = -5.788.$$

The null hypothesis is rejected at size $\alpha = 0.01$ since
$t = -5.788 < -t_{0.01,23} = -2.500$.
The p-value is $P(t_{23} \leq -5.788) = 0.000$.

(b) With a critical point $t_{0.01,23} = 2.500$
a 99% one-sided confidence interval for $\mu_A - \mu_B$ is

$$\left(-\infty, 436.5 - 452.8 + 2.500 \times \sqrt{\frac{11.90^2}{20} + \frac{4.61^2}{25}}\right)$$

$= (-\infty, -9.3)$.

(c) There is sufficient evidence to conclude that the synthetic fiber bundles have an average breaking strength larger than the wool fiber bundles.

9.3.8 Since

$$\frac{\left(\frac{0.058^2}{16}+\frac{0.062^2}{16}\right)^2}{\frac{0.058^4}{16^2\times(16-1)}+\frac{0.062^4}{16^2\times(16-1)}} = 29.9$$

the appropriate degrees of freedom for a general analysis without assuming equal population variances are $\nu = 29$.

Consider the hypotheses
$H_0 : \mu = \mu_A - \mu_B \geq 0$ versus $H_A : \mu = \mu_A - \mu_B < 0$
where the alternative hypothesis states that the brand B sugar packets weigh slightly more on average than brand A sugar packets.

The test statistic is

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}} = \frac{1.053-1.071}{\sqrt{\frac{0.058^2}{16}+\frac{0.062^2}{16}}} = -0.848$$

and the p-value is $P(t_{29} \leq -0.848) = 0.202$.

There is *not* sufficient evidence to conclude that the brand B sugar packets weigh slightly more on average than brand A sugar packets.

9.3.9 (a) The test statistic is

$$z = \frac{\bar{x}-\bar{y}-\delta}{\sqrt{\frac{\sigma_A^2}{n}+\frac{\sigma_B^2}{m}}} = \frac{100.85-89.32-3}{\sqrt{\frac{25^2}{47}+\frac{20^2}{62}}} = 1.92$$

and the p-value is $2 \times \Phi(-1.92) = 0.055$.

(b) With a critical point $z_{0.05} = 1.645$ a 90% two-sided confidence interval for $\mu_A - \mu_B$ is

$$100.85 - 89.32 \pm 1.645 \times \sqrt{\frac{25^2}{47} + \frac{20^2}{62}}$$

$$= (4.22, 18.84).$$

9.3.10 (a) The test statistic is

$$z = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_A^2}{n}+\frac{\sigma_B^2}{m}}} = \frac{5.782-6.443}{\sqrt{\frac{2.0^2}{38}+\frac{2.0^2}{40}}} = -1.459$$

and the p-value is $\Phi(-1.459) = 0.072$.

(b) With a critical point $z_{0.01} = 2.326$ a 99% one-sided confidence interval for $\mu_A - \mu_B$ is

$$\left(-\infty, 5.782 - 6.443 + 2.326 \times \sqrt{\frac{2.0^2}{38} + \frac{2.0^2}{40}}\right)$$

$$= (-\infty, 0.393).$$

9.3.11 (a) The test statistic is

$$z = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_A^2}{n}+\frac{\sigma_B^2}{m}}} = \frac{19.50-18.64}{\sqrt{\frac{1.0^2}{10}+\frac{1.0^2}{12}}} = 2.009$$

and the p-value is $2 \times \Phi(-2.009) = 0.045$.

(b) With a critical point $z_{0.05} = 1.645$ a 90% two-sided confidence interval for $\mu_A - \mu_B$ is

$$19.50 - 18.64 \pm 1.645 \times \sqrt{\frac{1.0^2}{10} + \frac{1.0^2}{12}}$$

$$= (0.16, 1.56).$$

With a critical point $z_{0.025} = 1.960$ a 95% two-sided confidence interval for $\mu_A - \mu_B$ is

$$19.50 - 18.64 \pm 1.960 \times \sqrt{\frac{1.0^2}{10} + \frac{1.0^2}{12}}$$

$$= (0.02, 1.70).$$

With a critical point $z_{0.005} = 2.576$ a 99% two-sided confidence interval for $\mu_A - \mu_B$ is

$19.50 - 18.64 \pm 2.576 \times \sqrt{\frac{1.0^2}{10} + \frac{1.0^2}{12}}$

$= (-0.24, 1.96).$

9.3.12 Using 2.6 as an upper bound for $t_{0.005,\nu}$ equal sample sizes of

$$n = m \geq \frac{4\, t^2_{\alpha/2,\nu}\, (\sigma_A^2+\sigma_B^2)}{L_0^2} = \frac{4\times 2.6^2 \times (10.0^2+15.0^2)}{10.0^2} = 87.88$$

should be sufficient.

Equal sample sizes of at least 88 can be recommended.

9.3.13 Using 2.0 as an upper bound for $t_{0.025,\nu}$ equal sample sizes of

$$n = m \geq \frac{4\, t^2_{\alpha/2,\nu}\, (\sigma_A^2+\sigma_B^2)}{L_0^2} = \frac{4\times 2.0^2 \times (1.2^2+1.2^2)}{1.0^2} = 46.08$$

should be sufficient.

Equal sample sizes of at least 47 can be recommended.

9.3.14 Using $t_{0.005,26} = 2.779$ equal total sample sizes of

$$n = m \geq \frac{4\, t^2_{\alpha/2,\nu}\, (s_x^2+s_y^2)}{L_0^2} = \frac{4\times 2.779^2 \times (4.30^2+5.23^2)}{5.0^2} = 56.6$$

should be sufficient.

Additional sample sizes of at least $57 - 14 = 43$ from each population can be recommended.

9.3.15 Using $t_{0.005,80} = 2.639$ equal total sample sizes of

$$n = m \geq \frac{4\, t^2_{\alpha/2,\nu}\, (s_x^2+s_y^2)}{L_0^2} = \frac{4\times 2.639^2 \times (0.124^2+0.137^2)}{0.1^2} = 95.1$$

should be sufficient.

Additional sample sizes of at least $96 - 41 = 55$ from each population can be recommended.

9.3.16 (a) The appropriate degrees of freedom are

$$\frac{\left(\frac{0.315^2}{12}+\frac{0.297^2}{13}\right)^2}{\frac{0.315^4}{12^2\times(12-1)}+\frac{0.297^4}{13^2\times(13-1)}} = 22.5$$

which should be rounded down to $\nu = 22$.

Consider the two-sided hypotheses
$H_0 : \mu_A = \mu_B$ versus $H_A : \mu_A \neq \mu_B$
for which the test statistic is

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}} = \frac{2.462-2.296}{\sqrt{\frac{0.315^2}{12}+\frac{0.297^2}{13}}} = 1.35$$

and the p-value is $2 \times P(t_{22} \geq 1.35) = 0.190$.
There is not sufficient evidence to conclude that the amount of chromium content has an effect on the average corrosion rate of chilled cast iron.

(b) With a critical point $t_{0.005,22} = 2.819$ a 99% two-sided confidence interval for the difference of the average corrosion rates of chilled cast iron at the two levels of chromium content is

$$2.462 - 2.296 \pm 2.819 \times \sqrt{\frac{0.315^2}{12} + \frac{0.297^2}{13}}$$

$$= (-0.180, 0.512).$$

9.3.17 There is sufficient evidence to conclude that the paving slabs from company A weigh more on average than the paving slabs from company B.

There is also more variability in the weights of the paving slabs from company A.

9.3.18 There is a fairly strong suggestion that the paint thicknesses from production line A are larger than those from production line B, although the evidence is not completely overwhelming (the p-value is 0.011).

9.3.19 There is sufficient evidence to conclude that the damped feature is effective in reducing the heel-strike force.

9.3.20 The high level of hydrogen peroxide seems to produce more variability in the whiteness measurements than the low level.

There is not sufficient evidence to conclude that the high level of hydrogen peroxide produces a larger average whiteness measurement than the low level of hydrogen peroxide.

9.3.21 There is not sufficient evidence to conclude that the average service times are any different at these two times of day.

9.3.22 The hypotheses are

$H_0 : \mu_N \leq \mu_S$ versus $H_A : \mu_N > \mu_S$

and

$$\frac{\left(\frac{6.30^2}{14}+\frac{7.15^2}{20}\right)^2}{\frac{6.30^4}{14^2\times(14-1)}+\frac{7.15^4}{20^2\times(20-1)}}=30.2$$

so that the degrees of freedom are $\nu = 30$.

The test statistic is

$$t=\frac{56.43-62.11}{\sqrt{\frac{6.30^2}{14}+\frac{7.15^2}{20}}}=-2.446$$

and the p-value is $P(t_{30} < -2.446) = 0.0103$.

Since the p-value is almost equal to 0.01, there is sufficient evidence to conclude that the new procedure has a larger breaking strength on average than the standard procedure.

9.3.23 $\bar{x}_A = 142.4$

$s_A = 9.24$

$n_A = 10$

$\bar{x}_B = 131.6$

$s_B = 7.97$

$n_B = 10$

The hypotheses are

$H_0 : \mu_A \leq \mu_B$ versus $H_A : \mu_A > \mu_B$

and

$$\frac{\left(\frac{9.24^2}{10}+\frac{7.97^2}{10}\right)^2}{\frac{9.24^4}{10^2\times(10-1)}+\frac{7.97^4}{10^2\times(10-1)}}=17.6$$

so that the degrees of freedom are $\nu = 17$.

The test statistic is

$$t=\frac{142.4-131.6}{\sqrt{\frac{9.24^2}{10}+\frac{7.97^2}{10}}}=2.799$$

and the p-value is $P(t_{17} > 2.799) = 0.006$.

There is sufficient evidence to conclude that on average medicine A provides a higher response than medicine B.

9.3.24 (a) $\bar{x}_M = 132.52$

$s_M = 1.31$

$n_M = 8$

$\bar{x}_A = 133.87$

$s_A = 1.72$
$n_A = 10$

The hypotheses are
$H_0 : \mu_M = \mu_A$ versus $H_A : \mu_M \neq \mu_A$
and

$$\frac{\left(\frac{1.31^2}{8}+\frac{1.72^2}{10}\right)^2}{\frac{1.31^4}{8^2\times(8-1)}+\frac{1.72^4}{10^2\times(10-1)}} = 15.98$$

so that the degrees of freedom are $\nu = 15$.

The test statistic is

$$t = \frac{132.52-133.87}{\sqrt{\frac{1.31^2}{8}+\frac{1.72^2}{10}}} = -1.89$$

and the p-value is $2 \times P(t_{15} > 1.89)$ which is between 5% and 10%.
There is some evidence to suggest that there is a difference between the running times in the morning and afternoon, but the evidence is not overwhelming.

(b) With $t_{0.005,15} = 2.947$ the confidence interval is

$$\mu_M - \mu_A \in 132.52 - 133.87 \pm 2.947 \times \sqrt{\frac{1.31^2}{8} + \frac{1.72^2}{10}} = (-3.46, 0.76).$$

9.3.25 $\bar{x}_A = 152.3$

$s_A = 1.83$

$n_A = 10$

$s_B = 1.94$

$n_B = 8$

The hypotheses are

$H_0 : \mu_A \leq \mu_B$ versus $H_A : \mu_A > \mu_B$

and

$$\frac{\left(\frac{1.83^2}{10}+\frac{1.94^2}{8}\right)^2}{\frac{1.83^4}{10^2\times(10-1)}+\frac{1.94^4}{8^2\times(8-1)}} = 14.7$$

so that the degrees of freedom are $\nu = 14$.

Since the p-value is $P(t_{14} > t) < 0.01$, it follows that

$$t = \frac{\bar{x}_A-\bar{x}_B}{\sqrt{\frac{s_A^2}{n_A}+\frac{s_B^2}{n_B}}} = \frac{152.3-\bar{x}_B}{0.8974} > t_{0.01,14} = 2.624$$

so that $\bar{x}_B < 149.9$.

9.6 Supplementary Problems

9.6.1 The differences $z_i = x_i - y_i$ have a sample mean $\bar{z} = 2.85$ and a sample standard deviation $s = 5.30$.

Consider the hypotheses

$H_0 : \mu = \mu_A - \mu_B \leq 0$ versus $H_A : \mu = \mu_A - \mu_B > 0$

where the alternative hypothesis states that the color displays are more effective than the black and white displays.

The test statistic is

$t = \frac{\sqrt{n}\,\bar{z}}{s} = \frac{\sqrt{22}\times 2.85}{5.30} = 2.52$

and the p-value is $P(t_{21} \geq 2.52) = 0.010$.

There is sufficient evidence to conclude that the color displays are more effective than the black and white displays.

With $t_{0.05,21} = 1.721$ a one-sided 95% confidence level confidence interval

for $\mu = \mu_A - \mu_B$ is

$\left(2.85 - \frac{1.721\times 5.30}{\sqrt{22}}, \infty\right)$

$= (0.91, \infty)$.

9.6.2 The differences $z_i = x_i - y_i$ have a sample mean $\bar{z} = 7.50$ and a sample standard deviation $s = 6.84$.

Consider the hypotheses

$H_0 : \mu = \mu_A - \mu_B = 0$ versus $H_A : \mu = \mu_A - \mu_B \neq 0$.

The test statistic is

$t = \frac{\sqrt{n}\,\bar{z}}{s} = \frac{\sqrt{14}\times 7.50}{6.84} = 4.10$

and the p-value is $2 \times P(t_{13} \geq 4.10) = 0.001$.

There is sufficient evidence to conclude that the water absorption properties of the fabric are different for the two different roller pressures.

With $t_{0.025,13} = 2.160$ a two-sided 95% confidence level confidence interval

for $\mu = \mu_A - \mu_B$ is

$\left(7.50 - \frac{2.160\times 6.84}{\sqrt{14}}, 7.50 + \frac{2.160\times 6.84}{\sqrt{14}}\right)$

$= (3.55, 11.45)$.

9.6.3 (a) Since

$$\frac{\left(\frac{5.20^2}{35}+\frac{3.06^2}{35}\right)^2}{\frac{5.20^4}{35^2\times(35-1)}+\frac{3.06^4}{35^2\times(35-1)}} = 55.03$$

the degrees of freedom are $\nu = 55$.

The test statistic is

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}} = \frac{22.73-12.66}{\sqrt{\frac{5.20^2}{35}+\frac{3.06^2}{35}}} = 9.87$$

and the p-value is $2 \times P(t_{55} \geq 9.87) = 0.000$.
It is not plausible that the average crystal size does not depend upon the pre-expansion temperature.

(b) With a critical point $t_{0.005,55} = 2.668$ a 99% two-sided confidence interval for $\mu_A - \mu_B$ is

$$22.73 - 12.66 \pm 2.668 \times \sqrt{\frac{5.20^2}{35} + \frac{3.06^2}{35}}$$

$$= (7.35, 12.79).$$

(c) Using $t_{0.005,55} = 2.668$ equal total sample sizes of

$$n = m \geq \frac{4\, t^2_{\alpha/2,\nu}\, (s_x^2+s_y^2)}{L_0^2} = \frac{4\times 2.668^2\times(5.20^2+3.06^2)}{4.0^2} = 64.8$$

should be sufficient.
Additional sample sizes of at least $65 - 35 = 30$ from each population can be recommended.

9.6.4 Since

$$\frac{\left(\frac{20.39^2}{48}+\frac{15.62^2}{10}\right)^2}{\frac{20.39^4}{48^2\times(48-1)}+\frac{15.62^4}{10^2\times(10-1)}} = 16.1$$

the appropriate degrees of freedom for a general analysis without assuming equal population variances are $\nu = 16$.

Consider the hypotheses

$H_0 : \mu = \mu_A - \mu_B \leq 0$ versus $H_A : \mu = \mu_A - \mu_B > 0$

where the alternative hypothesis states that the new driving route is quicker on average than the standard driving route.

The test statistic is

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}} = \frac{432.7-403.5}{\sqrt{\frac{20.39^2}{48}+\frac{15.62^2}{10}}} = 5.08$$

and the p-value is $P(t_{16} \geq 5.08) = 0.000$.

There is sufficient evidence to conclude that the new driving route is quicker on average than the standard driving route.

9.6.5 There is sufficient evidence to conclude that the additional sunlight results in larger heights on average.

9.6.6 There is *not* sufficient evidence to conclude that the reorganization has produced any improvement in the average waiting time.

However, the variability in the waiting times has been reduced following the reorganization.

9.6.7 This is a paired data set.

There is not any evidence of a difference in the average ocular motor measurements after reading a book and after reading a computer screen.

9.6.8 The variabilities in the viscosities appear to be about the same for the two engines, but there is sufficient evidence to conclude that the average viscosity is higher after having been used in engine 2 than after having been used in engine 1.

9.6.10 With $F_{0.05,17,20} = 2.1667$ and $F_{0.05,20,17} = 2.2304$

the confidence interval is

$$\left(\frac{6.48^2}{9.62^2 \times 2.1667}, \frac{6.48^2 \times 2.2304}{9.62^2}\right) = (0.21, 1.01).$$

9.6.11 With $F_{0.05,40,40} = 1.6928$

the confidence interval is

$$\left(\frac{0.124^2}{0.137^2 \times 1.6928}, \frac{0.124^2 \times 1.6928}{0.137^2}\right) = (0.484, 1.387).$$

9.6.12 With $F_{0.05,19,24} = 2.0399$ and $F_{0.05,24,19} = 2.1141$

the 90% confidence interval is

$$\left(\frac{11.90^2}{4.61^2 \times 2.0399}, \frac{11.90^2 \times 2.1141}{4.61^2}\right) = (3.27, 14.09).$$

With $F_{0.025,19,24} = 2.3452$ and $F_{0.025,24,19} = 2.4523$

the 95% confidence interval is

$$\left(\frac{11.90^2}{4.61^2 \times 2.3452}, \frac{11.90^2 \times 2.4523}{4.61^2}\right) = (2.84, 16.34).$$

With $F_{0.005,19,24} = 3.0920$ and $F_{0.005,24,19} = 3.3062$
the 99% confidence interval is

$$\left(\frac{11.90^2}{4.61^2 \times 3.0920}, \frac{11.90^2 \times 3.3062}{4.61^2}\right) = (2.16, 22.03).$$

9.6.13 $\bar{x}_A = 327433$
$s_A = 9832$
$n_A = 14$
$\bar{x}_B = 335537$
$s_B = 10463$
$n_B = 12$

The hypotheses are
$H_0 : \mu_A = \mu_B$ versus $H_A : \mu_A \neq \mu_B$
and since

$$\frac{\left(\frac{9832^2}{14} + \frac{10463^2}{12}\right)^2}{\frac{9832^4}{14^2 \times (14-1)} + \frac{10463^4}{12^2 \times (12-1)}} = 22.8$$

the degrees of freedom are $\nu = 22$.

The test statistic is

$$t = \frac{327433 - 335537}{\sqrt{\frac{9832^2}{14} + \frac{10463^2}{12}}} = -2.024$$

and the p-value is $2 \times P(t_{22} > 2.024)$ which is between 5% and 10%.
There is some evidence to suggest that there is a difference between the strengths of the two canvas types, but the evidence is not overwhelming.

9.6.14 Let x_i be the strength of the cement sample using procedure 1 and let y_i be the strength of the cement sample using procedure 2.
With $z_i = x_i - y_i$ it can be found that

$$\bar{z} = \frac{\sum_{i=1}^{9} z_i}{9} = -0.0222$$

and

$$s_z = \sqrt{\frac{\sum_{i=1}^{9}(z_i - \bar{z})^2}{8}} = 0.5911.$$

For the hypotheses
$H_0 : \mu_x = \mu_y$ versus $H_A : \mu_x \neq \mu_y$
the test statistic is

$t = \frac{\sqrt{9}(-0.0222-0)}{0.5911} = -0.113$

and the p-value is $2 \times P(t_8 \geq 0.113) = 0.91$.

Therefore, there is no evidence of any difference between the two procedures.

9.6.15 (a) False

(b) True

(c) True

(d) False

(e) False

(f) True

(g) True

(h) True

(i) True

9.6.16 Let x_i be the data obtained using therapy 1 and let y_i be the data obtained using therapy 2.

With $z_i = x_i - y_i$ it can be found that

$\bar{z} = \frac{\sum_{i=1}^{8} z_i}{8} = 1.000$

and

$s_z = \sqrt{\frac{\sum_{i=1}^{8}(z_i-\bar{z})^2}{7}} = 5.757.$

For the hypotheses

$H_0 : \mu_x = \mu_y$ versus $H_A : \mu_x \neq \mu_y$

the test statistic is

$t = \frac{\sqrt{8}(1.000-0)}{5.757} = 0.491$

and the p-value is $2 \times P(t_7 \geq 0.491) = 0.638$.

Therefore, there is not sufficient evidence to conclude that there is a difference between the two experimental drug therapies.

9.6.17 (a) The hypotheses are

$H_0 : \mu_A \geq \mu_B$ versus $H_A : \mu_A < \mu_B$

and the appropriate degrees of freedom are

$$\frac{\left(\frac{24.1^2}{20}+\frac{26.4^2}{24}\right)^2}{\frac{24.1^4}{20^2\times(20-1)}+\frac{26.4^4}{24^2\times(24-1)}} = 41.6$$

which should be rounded down to $\nu = 41$.

The test statistic is

$t = \frac{2376.3-2402.0}{\sqrt{\frac{24.1^2}{20}+\frac{26.4^2}{24}}} = -3.37$

and the p-value is $P(t_{41} \leq -3.37) = 0.0008$.

There is sufficient evidence to conclude that the items from manufacturer B provide larger measurements on average than the items from manufacturer A.

(b) With $t_{0.05,41} = 1.683$ the confidence interval is

$$\mu_B - \mu_A \in \left(-\infty, 2402.0 - 2376.3 + 1.683\sqrt{\frac{24.1^2}{20} + \frac{26.4^2}{24}}\right) = (-\infty, 38.5).$$

9.6.20 Let x_i be the mean error measurement for patient i using joystick design 1 and let y_i be the mean error measurement for patient i using joystick design 2.

With $z_i = x_i - y_i$ it can be found that

$\bar{z} = \frac{\sum_{i=1}^{9} z_i}{9} = 0.02067$

and

$s_z = \sqrt{\frac{\sum_{i=1}^{9}(z_i-\bar{z})^2}{8}} = 0.03201.$

For the hypotheses

$H_0 : \mu_x = \mu_y$ versus $H_A : \mu_x \neq \mu_y$

the test statistic is

$t = \frac{\sqrt{9}(0.02067-0)}{0.03201} = 1.937$

and the p-value is $2 \times P(t_8 \geq 1.937)$ which is between 5% and 10%.

Therefore, there is some evidence that the the two joystick designs result in different error rate measurements, but the evidence is not overwhelming.

With $t_{0.005,8} = 3.355$ a 99% confidence interval for the difference between the mean error measurements obtained from the two designs is

$0.02067 \pm \frac{3.355\times0.03201}{\sqrt{9}} = (-0.015, 0.056).$

Chapter 10

Discrete Data Analysis

10.1 Inferences on a Population Proportion

10.1.1 (a) With $z_{0.005} = 2.576$ the confidence interval is

$$\left(\frac{11}{32} - \frac{2.576}{32} \times \sqrt{\frac{11\times(32-11)}{32}}, \frac{11}{32} + \frac{2.576}{32} \times \sqrt{\frac{11\times(32-11)}{32}}\right)$$

$= (0.127, 0.560)$.

(b) With $z_{0.025} = 1.960$ the confidence interval is

$$\left(\frac{11}{32} - \frac{1.960}{32} \times \sqrt{\frac{11\times(32-11)}{32}}, \frac{11}{32} + \frac{1.960}{32} \times \sqrt{\frac{11\times(32-11)}{32}}\right)$$

$= (0.179, 0.508)$.

(c) With $z_{0.01} = 2.326$ the confidence interval is

$$\left(0, \frac{11}{32} + \frac{2.326}{32} \times \sqrt{\frac{11\times(32-11)}{32}}\right)$$

$= (0, 0.539)$.

(d) The exact p-value is $2 \times P(B(32, 0.5) \leq 11) = 0.110$.
The statistic for the normal approximation to the p-value is

$$z = \frac{x-np_0}{\sqrt{np_0(1-p_0)}} = \frac{11-(32\times 0.5)}{\sqrt{32\times 0.5\times(1-0.5)}} = -1.768$$

and the p-value is $2 \times \Phi(-1.768) = 0.077$.

10.1.2 (a) With $z_{0.005} = 2.576$ the confidence interval is

$$\left(\frac{21}{27} - \frac{2.576}{27} \times \sqrt{\frac{21\times(27-21)}{27}}, \frac{21}{27} + \frac{2.576}{27} \times \sqrt{\frac{21\times(27-21)}{27}}\right)$$

$= (0.572, 0.984)$.

(b) With $z_{0.025} = 1.960$ the confidence interval is

$$\left(\frac{21}{27} - \frac{1.960}{27} \times \sqrt{\frac{21\times(27-21)}{27}}, \frac{21}{27} + \frac{1.960}{27} \times \sqrt{\frac{21\times(27-21)}{27}}\right)$$

$= (0.621, 0.935)$.

(c) With $z_{0.05} = 1.645$ the confidence interval is

$$\left(\frac{21}{27} - \frac{1.645}{27} \times \sqrt{\frac{21\times(27-21)}{27}}, 1\right)$$

$= (0.646, 1)$.

(d) The exact p-value is $P(B(27, 0.6) \geq 21) = 0.042$.
The statistic for the normal approximation to the p-value is

$$z = \frac{x-np_0}{\sqrt{np_0(1-p_0)}} = \frac{21-(27\times0.6)}{\sqrt{27\times0.6\times(1-0.6)}} = 1.886$$

and the p-value is $1 - \Phi(1.886) = 0.030$.

10.1.3 (a) Let p be the probability that a value produced by the random number generator is a zero, and consider the hypotheses
$H_0 : p = 0.5$ versus $H_A : p \neq 0.5$
where the alternative hypothesis states that the random number generator is producing 0's and 1's with unequal probabilities.
The statistic for the normal approximation to the p-value is

$$z = \frac{x-np_0}{\sqrt{np_0(1-p_0)}} = \frac{25264-(50000\times0.5)}{\sqrt{50000}\times0.5\times(1-0.5)} = 2.361$$

and the p-value is $2 \times \Phi(-2.361) = 0.018$.
There is a fairly strong suggestion that the random number generator is producing 0's and 1's with unequal probabilities, although the evidence is not completely overwhelming.

(b) With $z_{0.005} = 2.576$ the confidence interval is

$$\left(\frac{25264}{50000} - \frac{2.576}{50000} \times \sqrt{\frac{25264\times(50000-25264)}{50000}}, \frac{25264}{50000} + \frac{2.576}{50000} \times \sqrt{\frac{25264\times(50000-25264)}{50000}}\right)$$

$= (0.4995, 0.5110)$.

(c) Using the worst case scenario
$\hat{p}(1-\hat{p}) = 0.25$
the total sample size required can be calculated as

$$n \geq \frac{4\, z_{\alpha/2}^2\, \hat{p}(1-\hat{p})}{L^2}$$

$$= \frac{4\times2.576^2\times0.25}{0.005^2} = 265431.04$$

so that an additional sample size of $265432 - 50000 \simeq 215500$ would be required.

10.1.4 With $z_{0.05} = 1.645$ the confidence interval is

$$\left(\frac{35}{44} - \frac{1.645}{44} \times \sqrt{\frac{35\times(44-35)}{44}}, 1\right)$$

$= (0.695, 1)$.

10.1.5 Let p be the probability that a six is scored on the die and consider the hypotheses

$H_0 : p \geq \frac{1}{6}$ versus $H_A : p < \frac{1}{6}$

where the alternative hypothesis states that the die has been weighted to reduce the chance of scoring a six.

In the first experiment the exact p-value is

$P\left(B\left(50, \frac{1}{6}\right) \leq 2\right) = 0.0066$

and in the second experiment the exact p-value is

$P\left(B\left(100, \frac{1}{6}\right) \leq 4\right) = 0.0001$

so that there is more support for foul play from the second experiment than from the first.

10.1.6 The exact p-value is

$2 \times P\left(B\left(100, \frac{1}{6}\right) \geq 21\right) = 0.304$

and the null hypothesis is accepted at size $\alpha = 0.05$.

10.1.7 Let p be the probability that a juror is selected from the county where the investigator lives, and consider the hypotheses

$H_0 : p = 0.14$ versus $H_A : p \neq 0.14$

where the alternative hypothesis implies that the jurors are not being randomly selected.

The statistic for the normal approximation to the p-value is

$$z = \frac{x-np_0}{\sqrt{np_0(1-p_0)}} = \frac{122-(1{,}386\times 0.14)}{\sqrt{1{,}386\times 0.14\times(1-0.14)}} = -5.577$$

and the p-value is $2 \times \Phi(-5.577) = 0.000$.

There is sufficient evidence to conclude that the jurors

are not being randomly selected.

10.1.8 The statistic for the normal approximation to the p-value is

$$z = \frac{x-np_0}{\sqrt{np_0(1-p_0)}} = \frac{23-(324\times 0.1)}{\sqrt{324\times 0.1\times(1-0.1)}} = -1.741$$

and the p-value is $\Phi(-1.741) = 0.041$.

With $z_{0.01} = 2.326$ the confidence interval is

$$\left(0, \frac{23}{324} + \frac{2.326}{324} \times \sqrt{\frac{23\times(324-23)}{324}}\right)$$

$= (0, 0.104)$.

It has not been conclusively shown that the screening test is acceptable.

10.1.9 With $z_{0.025} = 1.960$ and $L = 0.02$

the required sample size for the worst case scenario with

$\hat{p}(1-\hat{p}) = 0.25$

can be calculated as

$$n \geq \frac{4\, z_{\alpha/2}^2\, \hat{p}(1-\hat{p})}{L^2} = \frac{4\times 1.960^2\times 0.25}{0.02^2} = 9604.$$

If it can be assumed that

$\hat{p}(1-\hat{p}) \leq 0.75 \times 0.25 = 0.1875$

then the required sample size can be calculated as

$$n \geq \frac{4\, z_{\alpha/2}^2\, \hat{p}(1-\hat{p})}{L^2} = \frac{4\times 1.960^2\times 0.1875}{0.02^2} = 7203.$$

10.1.10 With $z_{0.005} = 2.576$ and $L = 0.04$

the required sample size for the worst case scenario with

$\hat{p}(1-\hat{p}) = 0.25$

can be calculated as

$$n \geq \frac{4\, z_{\alpha/2}^2\, \hat{p}(1-\hat{p})}{L^2} = \frac{4\times 2.576^2\times 0.25}{0.04^2} = 4148.$$

If it can be assumed that

$\hat{p}(1-\hat{p}) \leq 0.4 \times 0.6 = 0.24$

then the required sample size can be calculated as

$$n \geq \frac{4\, z_{\alpha/2}^2\, \hat{p}(1-\hat{p})}{L^2} = \frac{4\times 2.576^2\times 0.24}{0.04^2} = 3982.$$

10.1.11 With $z_{0.005} = 2.576$ the confidence interval is

$$\left(\frac{73}{120} - \frac{2.576}{120} \times \sqrt{\frac{73\times(120-73)}{120}}, \frac{73}{120} + \frac{2.576}{120} \times \sqrt{\frac{73\times(120-73)}{120}}\right)$$

$= (0.494, 0.723)$.

Using

$\hat{p}(1-\hat{p}) = \frac{73}{120} \times \left(1 - \frac{73}{120}\right) = 0.238$

the total sample size required can be calculated as

$n \geq \frac{4\, z_{\alpha/2}^2\, \hat{p}(1-\hat{p})}{L^2} = \frac{4\times 2.576^2\times 0.238}{0.1^2} = 631.7$

so that an additional sample size of $632 - 120 = 512$ would be required.

10.1.12 Let p be the proportion of defective chips in the shipment.

With $z_{0.05} = 1.645$ a 95% upper confidence bound on p is

$$\left(0, \frac{8}{200} + \frac{1.645}{200} \times \sqrt{\frac{8\times(200-8)}{200}}\right)$$

$= (0, 0.06279)$.

A 95% upper confidence bound on the total number of defective chips in the shipment can therefore be calculated as

$0.06279 \times 100000 = 6279$ chips.

10.1.13 With $z_{0.025} = 1.960$ the confidence interval is

$$\left(\frac{12}{20} - \frac{1.960}{20} \times \sqrt{\frac{12\times(20-12)}{20}}, \frac{12}{20} + \frac{1.960}{20} \times \sqrt{\frac{12\times(20-12)}{20}}\right)$$

$= (0.385, 0.815)$.

10.1.14 Let p be the proportion of the applications that contained errors.

With $z_{0.05} = 1.645$ a 95% lower confidence bound on p is

$$\left(\frac{17}{85} - \frac{1.645}{85} \times \sqrt{\frac{17\times(85-17)}{85}}, 1\right)$$

$= (0.1286, 1)$.

A 95% lower confidence bound on the total number of applications which contained errors can therefore be calculated as

$0.1286 \times 7607 = 978.5$ or 979 applications.

10.1.15 With $z_{0.025} = 1.960$ and $L = 0.10$

the required sample size for the worst case scenario with

$\hat{p}(1-\hat{p}) = 0.25$

can be calculated as

$n \geq \frac{4\, z_{\alpha/2}^2\, \hat{p}(1-\hat{p})}{L^2} = \frac{4 \times 1.960^2 \times 0.25}{0.10^2} = 384.2$

or 385 householders.

If it can be assumed that

$\hat{p}(1-\hat{p}) \leq 0.333 \times 0.667 = 0.222$

then the required sample size can be calculated as

$n \geq \frac{4\, z_{\alpha/2}^2\, \hat{p}(1-\hat{p})}{L^2} = \frac{4 \times 1.960^2 \times 0.222}{0.10^2} = 341.1$

or 342 householders.

10.1.16 With $z_{0.005} = 2.576$ the confidence interval is

$\left(\frac{22}{542} - \frac{2.576}{542} \times \sqrt{\frac{22 \times (542-22)}{542}}, \frac{22}{542} + \frac{2.576}{542} \times \sqrt{\frac{22 \times (542-22)}{542}} \right)$

$= (0.019, 0.062)$.

10.1.17 The standard confidence interval is $(0.161, 0.557)$.

The alternative confidence interval is $(0.195, 0.564)$.

10.1.18 (a) Let p be the probability that the dielectric breakdown strength is below the threshold level, and consider the hypotheses

$H_0 : p \leq 0.05$ versus $H_A : p > 0.05$

where the alternative hypothesis states that the probability of an insulator of this type having a dielectric breakdown strength below the specified threshold level is larger than 5%.

The statistic for the normal approximation to the p-value is

$z = \frac{x - np_0 - 0.5}{\sqrt{np_0(1-p_0)}} = \frac{13 - (62 \times 0.05) - 0.5}{\sqrt{62 \times 0.05 \times (1-0.05)}} = 5.48$

and the p-value is $1 - \Phi(5.48) = 0.000$.

There is sufficient evidence to conclude that the probability of an insulator of this type having a dielectric breakdown strength below the specified threshold level is larger than 5%.

(b) With $z_{0.05} = 1.645$ the confidence interval is

$$\left(\frac{13}{62} - \frac{1.645}{62}\sqrt{\frac{13(62-13)}{62}}, 1\right) = (0.125, 1).$$

10.1.19 $\hat{p} = \frac{31}{210} = 0.148$

With $z_{0.005} = 2.576$ the confidence interval is

$p \in 0.148 \pm \frac{2.576}{210}\sqrt{\frac{31\times(210-31)}{210}}$

$= (0.085, 0.211)$.

10.1.20 Let p be the probability of preferring cushion type A.

Then

$\hat{p} = \frac{28}{38} = 0.737$

and the hypotheses of interest are

$H_0 : p \leq \frac{2}{3}$ versus $H_A : p > \frac{2}{3}$.

The test statistic is

$z = \frac{28-(38\times 2/3)-0.5}{\sqrt{38\times 2/3\times 1/3}} = 0.75$

and the p-value is $1 - \Phi(0.75) = 0.227$.

The data set does not provide sufficient evidence to establish that cushion type A is at least twice as popular as cushion type B.

10.1.21 If $793 = \frac{z^2_{\alpha/2}}{(2\times 0.035)^2}$

then $z^2_{\alpha/2} = 1.97$ so that $\alpha \simeq 0.05$.

Therefore, the margin of error was calculated with 95% confidence under the worst case scenario where the estimated probability could be close to 0.5.

10.2 Comparing Two Population Proportions

10.2.1 (a) With $z_{0.005} = 2.576$ the confidence interval is

$\frac{14}{37} - \frac{7}{26} \pm 2.576 \times \sqrt{\frac{14\times(37-14)}{37^3} + \frac{7\times(26-7)}{26^3}}$

$= (-0.195, 0.413)$.

(b) With $z_{0.025} = 1.960$ the confidence interval is

$\frac{14}{37} - \frac{7}{26} \pm 1.960 \times \sqrt{\frac{14\times(37-14)}{37^3} + \frac{7\times(26-7)}{26^3}}$

$= (-0.122, 0.340)$.

(c) With $z_{0.01} = 2.326$ the confidence interval is

$\left(\frac{14}{37} - \frac{7}{26} - 2.326 \times \sqrt{\frac{14\times(37-14)}{37^3} + \frac{7\times(26-7)}{26^3}}, 1\right)$

$= (-0.165, 1)$.

(d) With the pooled probability estimate

$\hat{p} = \frac{x+y}{n+m} = \frac{14+7}{37+26} = 0.333$

the test statistic is

$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{\frac{14}{37}-\frac{7}{26}}{\sqrt{0.333\times(1-0.333)\times\left(\frac{1}{37}+\frac{1}{26}\right)}} = 0.905$

and the p-value is $2 \times \Phi(-0.905) = 0.365$.

10.2.2 (a) With $z_{0.005} = 2.576$ the confidence interval is

$\frac{261}{302} - \frac{401}{454} \pm 2.576 \times \sqrt{\frac{261\times(302-261)}{302^3} + \frac{401\times(454-401)}{454^3}}$

$= (-0.083, 0.045)$.

(b) With $z_{0.05} = 1.645$ the confidence interval is

$\frac{261}{302} - \frac{401}{454} \pm 1.645 \times \sqrt{\frac{261\times(302-261)}{302^3} + \frac{401\times(454-401)}{454^3}}$

$= (-0.060, 0.022)$.

(c) With $z_{0.05} = 1.645$ the confidence interval is

$\left(-1, \frac{261}{302} - \frac{401}{454} + 1.645 \times \sqrt{\frac{261\times(302-261)}{302^3} + \frac{401\times(454-401)}{454^3}}\right)$

$= (-1, 0.022)$.

(d) With the pooled probability estimate

$\hat{p} = \frac{x+y}{n+m} = \frac{261+401}{302+454} = 0.876$

the test statistic is

$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{\frac{261}{302} - \frac{401}{454}}{\sqrt{0.876\times(1-0.876)\times\left(\frac{1}{302}+\frac{1}{454}\right)}} = -0.776$

and the p-value is $2 \times \Phi(-0.776) = 0.438$.

10.2.3 (a) With $z_{0.005} = 2.576$ the confidence interval is

$\frac{35}{44} - \frac{36}{52} \pm 2.576 \times \sqrt{\frac{35\times(44-35)}{44^3} + \frac{36\times(52-36)}{52^3}}$

$= (-0.124, 0.331)$.

(b) With the pooled probability estimate

$\hat{p} = \frac{x+y}{n+m} = \frac{35+36}{44+52} = 0.740$

the test statistic is

$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{\frac{35}{44} - \frac{36}{52}}{\sqrt{0.740\times(1-0.740)\times\left(\frac{1}{44}+\frac{1}{52}\right)}} = 1.147$

and the p-value is $2 \times \Phi(-1.147) = 0.251$.

There is *not* sufficient evidence to conclude that one radar system is any better than the other radar system.

10.2.4 (a) With $z_{0.005} = 2.576$ the confidence interval is

$\frac{4}{50} - \frac{10}{50} \pm 2.576 \times \sqrt{\frac{4\times(50-4)}{50^3} + \frac{10\times(50-10)}{50^3}}$

$= (-0.296, 0.056)$.

(b) With the pooled probability estimate

$\hat{p} = \frac{x+y}{n+m} = \frac{4+10}{50+50} = 0.14$

the test statistic is

$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{\frac{4}{50} - \frac{10}{50}}{\sqrt{0.14\times(1-0.14)\times\left(\frac{1}{50}+\frac{1}{50}\right)}} = -1.729$

and the p-value is $2 \times \Phi(-1.729) = 0.084$.

(c) In this case the confidence interval is

$\frac{40}{500} - \frac{100}{500} \pm 2.576 \times \sqrt{\frac{40\times(500-40)}{500^3} + \frac{100\times(500-100)}{500^3}}$

$= (-0.176, -0.064)$.

With the pooled probability estimate

$\hat{p} = \frac{x+y}{n+m} = \frac{40+100}{500+500} = 0.14$

the test statistic is

$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{\frac{40}{500} - \frac{100}{500}}{\sqrt{0.14 \times (1-0.14) \times \left(\frac{1}{500}+\frac{1}{500}\right)}} = -5.468$

and the p-value is $2 \times \Phi(-5.468) = 0.000$.

10.2.5 Let p_A be the probability of crystallization within 24 hours *without* seed crystals and let p_B be the probability of crystallization within 24 hours *with* seed crystals.

With $z_{0.05} = 1.645$ a 95% upper confidence bound for $p_A - p_B$ is

$$\left(-1, \frac{27}{60} - \frac{36}{60} + 1.645 \times \sqrt{\frac{27\times(60-27)}{60^3} + \frac{36\times(60-36)}{60^3}}\right)$$

$= (-1, -0.002)$.

Consider the hypotheses

$H_0 : p_A \geq p_B$ versus $H_A : p_A < p_B$

where the alternative hypothesis states that the presence of seed crystals increases the probability of crystallization within 24 hours.

With the pooled probability estimate

$\hat{p} = \frac{x+y}{n+m} = \frac{27+36}{60+60} = 0.525$

the test statistic is

$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{\frac{27}{60} - \frac{36}{60}}{\sqrt{0.525 \times (1-0.525) \times \left(\frac{1}{60}+\frac{1}{60}\right)}} = -1.645$

and the p-value is $\Phi(-1.645) = 0.050$.

There is some evidence that the presence of seed crystals increases the probability of crystallization within 24 hours but it is not overwhelming.

10.2.6 Let p_A be the probability of an improved condition with the standard drug and let p_B be the probability of an improved condition with the new drug.

With $z_{0.05} = 1.645$ a 95% upper confidence bound for $p_A - p_B$ is

$$\left(-1, \frac{72}{100} - \frac{83}{100} + 1.645 \times \sqrt{\frac{72\times(100-72)}{100^3} + \frac{83\times(100-83)}{100^3}}\right)$$

$= (-1, -0.014)$.

Consider the hypotheses

$H_0 : p_A \geq p_B$ versus $H_A : p_A < p_B$

where the alternative hypothesis states that the new drug increases the probability of an improved condition.

With the pooled probability estimate

$\hat{p} = \frac{x+y}{n+m} = \frac{72+83}{100+100} = 0.775$

the test statistic is

$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{\frac{72}{100} - \frac{83}{100}}{\sqrt{0.775\times(1-0.775)\times\left(\frac{1}{100}+\frac{1}{100}\right)}} = -1.863$

and the p-value is $\Phi(-1.863) = 0.031$.

There is some evidence that the new drug increases the probability of an improved condition but it is not overwhelming.

10.2.7 Let p_A be the probability that a television set from production line A does not meet the quality standards and let p_B be the probability that a television set from production line B does not meet the quality standards.

With $z_{0.025} = 1.960$ a 95% two-sided confidence interval for $p_A - p_B$ is

$\frac{23}{1128} - \frac{24}{962} \pm 1.960 \times \sqrt{\frac{23\times(1128-23)}{1128^3} + \frac{24\times(962-24)}{962^3}}$

$= (-0.017, 0.008)$.

Consider the hypotheses

$H_0 : p_A = p_B$ versus $H_A : p_A \neq p_B$

where the alternative hypothesis states that there is a difference in the operating standards of the two production lines.

With the pooled probability estimate

$\hat{p} = \frac{x+y}{n+m} = \frac{23+24}{1128+962} = 0.022$

the test statistic is

$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{\frac{23}{1128} - \frac{24}{962}}{\sqrt{0.022\times(1-0.022)\times\left(\frac{1}{1128}+\frac{1}{962}\right)}} = -0.708$

and the p-value is $2 \times \Phi(-0.708) = 0.479$.

There is *not* sufficient evidence to conclude that there is a difference in the operating standards of the two production lines.

10.2.8 Let p_A be the probability of a successful outcome for the standard procedure and let p_B be the probability of a successful outcome for the new procedure.

With $z_{0.05} = 1.645$ a 95% upper confidence bound for $p_A - p_B$ is

$\left(-1, \frac{73}{120} - \frac{101}{120} + 1.645 \times \sqrt{\frac{73\times(120-73)}{120^3} + \frac{101\times(120-101)}{120^3}}\right)$

$= (-1, -0.142)$.

Consider the hypotheses

$H_0 : p_A \geq p_B$ versus $H_A : p_A < p_B$

where the alternative hypothesis states that the new procedure increases the probability of a successful outcome.

With the pooled probability estimate

$\hat{p} = \frac{x+y}{n+m} = \frac{73+101}{120+120} = 0.725$

the test statistic is

$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{\frac{73}{120} - \frac{101}{120}}{\sqrt{0.725 \times (1-0.725) \times \left(\frac{1}{120}+\frac{1}{120}\right)}} = -4.05$

and the p-value is $\Phi(-4.05) \simeq 0.0000$.

There is sufficient evidence to conclude that the new procedure increases the probability of a successful outcome.

10.2.9 Let p_A be the probability that a computer chip from supplier A is defective and let p_B be the probability that a computer chip from supplier B is defective.

With $z_{0.025} = 1.960$ a 95% two-sided confidence interval for $p_A - p_B$ is

$\frac{8}{200} - \frac{13}{250} \pm 1.960 \times \sqrt{\frac{8 \times (200-8)}{200^3} + \frac{13 \times (250-13)}{250^3}}$

$= (-0.051, 0.027)$.

Consider the hypotheses

$H_0 : p_A = p_B$ versus $H_A : p_A \neq p_B$

where the alternative hypothesis states that there is a difference in the quality of the computer chips from the two suppliers.

With the pooled probability estimate

$\hat{p} = \frac{x+y}{n+m} = \frac{8+13}{200+250} = 0.047$

the test statistic is

$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{\frac{8}{200} - \frac{13}{250}}{\sqrt{0.047 \times (1-0.047) \times \left(\frac{1}{200}+\frac{1}{250}\right)}} = -0.600$

and the p-value is $2 \times \Phi(-0.600) = 0.549$.

There is *not* sufficient evidence to conclude that there is a difference in the quality of the computer chips from the two suppliers.

10.2.10 Let p_A be the probability of an error in an application processed during the first two weeks and let p_B be the probability of an error in an application processed after the first two weeks.

With $z_{0.05} = 1.645$ a 95% lower confidence bound for $p_A - p_B$ is

$$\left(\frac{17}{85} - \frac{16}{132} - 1.645 \times \sqrt{\frac{17\times(85-17)}{85^3} + \frac{16\times(132-16)}{132^3}}, 1\right)$$

$= (-0.007, 1)$.

Consider the hypotheses

$H_0 : p_A \leq p_B$ versus $H_A : p_A > p_B$

where the alternative hypothesis states that the probability of an error in the processing of an application is larger during the first two weeks.

With the pooled probability estimate

$$\hat{p} = \frac{x+y}{n+m} = \frac{17+16}{85+132} = 0.152$$

the test statistic is

$$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{\frac{17}{85} - \frac{16}{132}}{\sqrt{0.152\times(1-0.152)\times\left(\frac{1}{85}+\frac{1}{132}\right)}} = 1.578$$

and the p-value is $1 - \Phi(1.578) = 0.057$.

There is some evidence that the probability of an error in the processing of an application is larger during the first two weeks but it is not overwhelming.

10.2.11 With the pooled probability estimate

$$\hat{p} = \frac{x+y}{n+m} = \frac{159+138}{185+185} = 0.803$$

the test statistic is

$$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{\frac{159}{185} - \frac{138}{185}}{\sqrt{0.803\times(1-0.803)\times\left(\frac{1}{185}+\frac{1}{185}\right)}} = 2.745$$

and the two-sided p-value is $2 \times \Phi(-2.745) = 0.006$.

The two-sided null hypothesis $H_0 : p_A = p_B$ is rejected and there is sufficient evidence to conclude that machine A is better than machine B.

10.2.12 Let p_A be the probability of a link being followed with the original design and let p_B be the probability of a link being followed with the modified design.

With $z_{0.05} = 1.645$ a 95% upper confidence bound for $p_A - p_B$ is

$$\left(-1, \frac{22}{542} - \frac{64}{601} + 1.645 \times \sqrt{\frac{22\times(542-22)}{542^3} + \frac{64\times(601-64)}{601^3}}\right)$$

$= (-1, -0.041)$.

Consider the hypotheses

$H_0 : p_A \geq p_B$ versus $H_A : p_A < p_B$

where the alternative hypothesis states that the probability of a link being followed is larger after the modifications.

With the pooled probability estimate

$\hat{p} = \frac{x+y}{n+m} = \frac{22+64}{542+601} = 0.0752$

the test statistic is

$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{\frac{22}{542}-\frac{64}{601}}{\sqrt{0.0752\times(1-0.0752)\times\left(\frac{1}{542}+\frac{1}{601}\right)}} = -4.22$

and the p-value is $\Phi(-4.22) \simeq 0.000$.

There is sufficient evidence to conclude that the probability of a link being followed has been increased by the modifications.

10.2.13 (a) Consider the hypotheses

$H_0 : p_{180} \geq p_{250}$ versus $H_A : p_{180} < p_{250}$

where the alternative hypothesis states that the probability of an insulator of this type having a dielectric breakdown strength below the specified threshold level is larger at 250 degrees Centigrade than it is at 180 degrees Centigrade.

With the pooled probability estimate

$\frac{x+y}{n+m} = \frac{13+20}{62+70} = 0.25$

the test statistic is

$z = \frac{\hat{p}_{180} - \hat{p}_{250}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{\frac{13}{62}-\frac{20}{70}}{\sqrt{0.25\times(1-0.25)\left(\frac{1}{62}+\frac{1}{70}\right)}} = -1.007$

and the p-value is $\Phi(-1.007) = 0.1570$.

There is not sufficient evidence to conclude that the probability of an insulator of this type having a dielectric breakdown strength below the specified threshold level is larger at 250 degrees Centigrade than it is at 180 degrees Centigrade.

(b) With $z_{0.005} = 2.576$ the confidence interval is

$\frac{13}{62} - \frac{20}{70} \pm 2.576 \times \sqrt{\frac{13\times(62-13)}{62^3} + \frac{20\times(70-20)}{70^3}}$

$= (-0.269, 0.117)$.

10.2.14 $\hat{p}_A = \frac{72}{125} = 0.576$

$\hat{p}_B = \frac{60}{125} = 0.480$

The pooled estimate is

$\hat{p} = \frac{72+60}{125+125} = 0.528$

and the hypotheses are

$H_0 : p_A = p_B$ versus $H_A : p_A \neq p_B$.

The test statistic is

$z = \frac{0.576-0.480}{\sqrt{0.528\times0.472\times\left(\frac{1}{125}+\frac{1}{125}\right)}} = 1.520$

and the p-value is $2 \times \Phi(-1.520) = 0.128$.

There is not sufficient evidence to conclude that there is a difference between the two treatments.

10.2.15 $\hat{p}_1 = \frac{76}{243} = 0.313$

$\hat{p}_2 = \frac{122}{320} = 0.381$

With $z_{0.005} = 2.576$ the confidence interval is

$p_1 - p_2 \in 0.313 - 0.381 \pm 2.576 \times \sqrt{\frac{76\times(243-76)}{243^3} + \frac{122\times(320-122)}{320^3}}$

$= (-0.172, 0.036)$

The confidence interval contains zero so there is not sufficient evidence to conclude that the failure rates due to operator misuse are different for the two products.

10.3 Goodness of Fit Tests for One-way Contingency Tables

10.3.1 (a) The expected cell frequencies are $e_i = \frac{500}{6} = 83.33$.

(b) The Pearson chi-square statistic is

$X^2 = \frac{(80-83.33)^2}{83.33} + \frac{(71-83.33)^2}{83.33} + \frac{(90-83.33)^2}{83.33} + \frac{(87-83.33)^2}{83.33}$

$+ \frac{(78-83.33)^2}{83.33} + \frac{(94-83.33)^2}{83.33} = 4.36.$

(c) The likelihood ratio chi-square statistic is

$G^2 = 2 \times \left(80 \ln\left(\frac{80}{83.33}\right) + 71 \ln\left(\frac{71}{83.33}\right) + 90 \ln\left(\frac{90}{83.33}\right)\right.$

$\left. + 87 \ln\left(\frac{87}{83.33}\right) + 78 \ln\left(\frac{78}{83.33}\right) + 94 \ln\left(\frac{94}{83.33}\right)\right) = 4.44.$

(d) The p-values are $P(\chi_5^2 \geq 4.36) = 0.499$ and $P(\chi_5^2 \geq 4.44) = 0.488$.

A size $\alpha = 0.01$ test of the null hypothesis that the die is fair is accepted.

(e) With $z_{0.05} = 1.645$ the confidence interval is

$\left(\frac{94}{500} - \frac{1.645}{500} \times \sqrt{\frac{94\times(500-94)}{500}}, \frac{94}{500} + \frac{1.645}{500} \times \sqrt{\frac{94\times(500-94)}{500}}\right)$

$= (0.159, 0.217).$

10.3.2 The expected cell frequencies are

1	2	3	4	5	6	7	8	9	≥10
50.00	41.67	34.72	28.94	24.11	20.09	16.74	13.95	11.62	58.16

The Pearson chi-square statistic is $X^2 = 10.33$.

The p-value is $P(\chi_9^2 \geq 10.33) = 0.324$.

The geometric distribution with $p = \frac{1}{6}$ is plausible.

10.3.3 (a) The expected cell frequencies are:

$e_1 = 221 \times \frac{4}{7} = 126.29$

$e_2 = 221 \times \frac{2}{7} = 63.14$

$e_3 = 221 \times \frac{1}{7} = 31.57$

The Pearson chi-square statistic is

$X^2 = \frac{(113-126.29)^2}{126.29} + \frac{(82-63.14)^2}{63.14} + \frac{(26-31.57)^2}{31.57} = 8.01.$

The p-value is $P(\chi_2^2 \geq 8.01) = 0.018$.

There is a fairly strong suggestion that the supposition is not plausible although the evidence is not completely overwhelming.

(b) With $z_{0.005} = 2.576$ the confidence interval is

$\left(\frac{113}{221} - \frac{2.576}{221} \times \sqrt{\frac{113\times(221-113)}{221}}, \frac{113}{221} + \frac{2.576}{221} \times \sqrt{\frac{113\times(221-113)}{221}}\right)$

$= (0.425, 0.598)$.

10.3.4 The expected cell frequencies are:

$e_1 = 964 \times 0.14 = 134.96$

$e_2 = 964 \times 0.22 = 212.08$

$e_3 = 964 \times 0.35 = 337.40$

$e_4 = 964 \times 0.16 = 154.24$

$e_5 = 964 \times 0.13 = 125.32$

The Pearson chi-square statistic is $X^2 = 14.6$.

The p-value is $P(\chi_4^2 \geq 14.6) = 0.006$.

There is sufficient evidence to conclude that the

jurors have not been selected randomly.

10.3.5 (a) The expected cell frequencies are:

$e_1 = 126 \times 0.5 = 63.0$

$e_2 = 126 \times 0.4 = 50.4$

$e_3 = 126 \times 0.1 = 12.6$

The likelihood ratio chi-square statistic is

$G^2 = 2 \times \left(56 \ln\left(\frac{56}{63.0}\right) + 51 \ln\left(\frac{51}{50.4}\right) + 19 \ln\left(\frac{19}{12.6}\right)\right) = 3.62.$

The p-value is $P(\chi_2^2 \geq 3.62) = 0.164$.

These probability values are plausible.

(b) With $z_{0.025} = 1.960$ the confidence interval is

$\left(\frac{56}{126} - \frac{1.960}{126} \times \sqrt{\frac{56\times(126-56)}{126}}, \frac{56}{126} + \frac{1.960}{126} \times \sqrt{\frac{56\times(126-56)}{126}}\right)$

$= (0.358, 0.531)$.

10.3.6 If the three soft drink formulations are equally likely
then the expected cell frequencies are

$e_i = 600 \times \frac{1}{3} = 200$.

The Pearson chi-square statistic is

$X^2 = \frac{(225-200)^2}{200} + \frac{(223-200)^2}{200} + \frac{(152-200)^2}{200} = 17.29$.

The p-value is $P(\chi_2^2 \geq 17.29) = 0.0002$.

It is not plausible that the three soft drink formulations are equally likely.

10.3.7 The first two cells should be pooled so that there are 13 cells altogether.

The Pearson chi-square statistic is $X^2 = 92.9$
and the p-value is $P(\chi_{12}^2 \geq 92.9) = 0.0000$.

It is not reasonable to model the number of arrivals with a
Poisson distribution with mean $\lambda = 7$.

10.3.8 A Poisson distribution with mean $\lambda = \bar{x} = 4.49$ can be considered.
The first two cells should be pooled and the last two cells should be pooled
so that there are 9 cells altogether.

The Pearson chi-square statistic is $X^2 = 8.3$
and the p-value is $P(\chi_7^2 \geq 8.3) = 0.307$.

It is reasonable to model the number of radioactive particles emitted
with a Poisson distribution.

10.3.9 If the pearl oyster diameters have a uniform distribution then the expected cell frequencies are:

$e_1 = 1490 \times 0.1 = 149$

$e_2 = 1490 \times 0.2 = 298$

$e_3 = 1490 \times 0.2 = 298$

$e_4 = 1490 \times 0.5 = 745$

The Pearson chi-square statistic is

$X^2 = \frac{(161-149)^2}{149} + \frac{(289-298)^2}{298} + \frac{(314-298)^2}{298} + \frac{(726-745)^2}{745} = 2.58.$

The p-value is $P(\chi_3^2 \geq 2.58) = 0.461.$

It is plausible that the pearl oyster diameters have a uniform distribution between 0 and 10 mm.

10.3.13 According to the genetic theory the probabilities are $\frac{9}{16}$, $\frac{3}{16}$, $\frac{3}{16}$ and $\frac{1}{16}$, so that the expected cell frequencies are:

$e_1 = \frac{9\times727}{16} = 408.9375$

$e_2 = \frac{3\times727}{16} = 136.3125$

$e_3 = \frac{3\times727}{16} = 136.3125$

$e_4 = \frac{1\times727}{16} = 45.4375$

The Pearson chi-square statistic is

$X^2 = \frac{(412-408.9375)^2}{408.9375} + \frac{(121-136.3125)^2}{136.3125}$

$+\frac{(148-136.3125)^2}{136.3125} + \frac{(46-45.4375)^2}{45.4375} = 2.75$

and the likelihood ratio chi-square statistic is

$G^2 = 2 \times \left(412 \ln\left(\frac{412}{408.9375}\right) + 121 \ln\left(\frac{121}{136.3125}\right)\right.$

$\left.+148 \ln\left(\frac{148}{136.3125}\right) + 46 \ln\left(\frac{46}{45.4375}\right)\right) = 2.79.$

The p-values are $P(\chi_3^2 \geq 2.75) = 0.432$ and $P(\chi_3^2 \geq 2.79) = 0.425$ so that the data set is consistent with the proposed genetic theory.

10.3.14 $e_1 = e_2 = e_3 = 205 \times \frac{1}{3} = 68.33$

The Pearson chi-square statistic is

$X^2 = \frac{(83-68.33)^2}{68.33} + \frac{(75-68.33)^2}{68.33} + \frac{(47-68.33)^2}{68.33} = 10.46$

so that the p-value is $P(X_2^2 \geq 10.46) = 0.005.$

There is sufficient evidence to conclude that the three products do not have equal probabilities of being chosen.

10.3.15 (a) $\hat{p}_3 = \frac{489}{630} = 0.776$

The hypotheses are
$H_0 : p_3 = 0.80$ versus $H_A : p_3 \neq 0.80$
and the test statistic is

$z = \frac{489-(630\times 0.8)}{\sqrt{630\times 0.8\times 0.2}} = -1.494.$

The p-value is $2 \times \Phi(-1.494) = 0.135$.

There is not sufficient evidence to conclude that the probability that a solution has normal acidity is not 0.80.

(b) $e_1 = 630 \times 0.04 = 25.2$
$e_2 = 630 \times 0.06 = 37.8$
$e_3 = 630 \times 0.80 = 504.0$
$e_4 = 630 \times 0.06 = 37.8$
$e_5 = 630 \times 0.04 = 25.2$

The Pearson chi-square statistic is

$X^2 = \frac{(34-25.2)^2}{25.2} + \frac{(41-37.8)^2}{37.8} + \frac{(489-504.0)^2}{504.0} + \frac{(52-37.8)^2}{37.8} + \frac{(14-25.2)^2}{25.2} = 14.1$

so that the p-value is $P(X_4^2 \geq 14.1) = 0.007$.

The data is not consistent with the claimed probabilities.

10.3.16 $P(X \leq 24) = 1 - e^{-(0.065\times 24)^{0.45}} = 0.705$

$P(X \leq 48) = 1 - e^{-(0.065\times 48)^{0.45}} = 0.812$

$P(X \leq 72) = 1 - e^{-(0.065\times 72)^{0.45}} = 0.865$

The observed cell frequencies are $x_1 = 12$, $x_2 = 53$, $x_3 = 39$, and $x_4 = 21$.

The expected cell frequencies are:

$e_1 = 125 \times 0.705 = 88.125$

$e_2 = 125 \times (0.812 - 0.705) = 13.375$

$e_3 = 125 \times (0.865 - 0.812) = 6.625$

$e_4 = 125 \times (1 - 0.865) = 16.875$

The Pearson chi-square statistic is

$X^2 = \frac{(12-88.125)^2}{88.125} + \frac{(53-13.375)^2}{13.375} + \frac{(39-6.625)^2}{6.625} + \frac{(21-16.875)^2}{16.875} = 342$

so that the p-value is $P(\chi_3^2 \geq 342) \simeq 0$.

It is not plausible that for these batteries under these storage conditions the time in hours until the charge drops below the threshold level has a Weibull distribution with parameters $\lambda = 0.065$ and $a = 0.45$.

10.3.17 The total sample size is $n = 76$.

Under the specified Poisson distribution the expected cell frequencies are:

$e_1 = 76 \times e^{-2.5} \times \frac{2.5^0}{0!} = 6.238$

$e_2 = 76 \times e^{-2.5} \times \frac{2.5^1}{1!} = 15.596$

$e_3 = 76 \times e^{-2.5} \times \frac{2.5^2}{2!} = 19.495$

$e_4 = 76 \times e^{-2.5} \times \frac{2.5^3}{3!} = 16.246$

$e_5 = 76 \times e^{-2.5} \times \frac{2.5^4}{4!} = 10.154$

$e_6 = 76 - e_1 - e_2 - e_3 - e_4 - e_5 = 8.270$

The Pearson chi-square statistic is

$X^2 = \frac{(3-6.238)^2}{6.238} + \frac{(12-15.596)^2}{15.596} + \frac{(23-19.495)^2}{19.495}$

$+ \frac{(18-16.246)^2}{16.246} + \frac{(13-10.154)^2}{10.154} + \frac{(7-8.270)^2}{8.270} = 4.32$

so that the p-value is $P(\chi^2_5 \geq 4.32) = 0.50$.

It is plausible that the number of shark attacks per year follows a Poisson distribution with mean 2.5.

10.4 Testing for Independence in Two-way Contingency Tables

10.4.1 (a) The expected cell frequencies are

	Acceptable	Defective
Supplier A	186.25	13.75
Supplier B	186.25	13.75
Supplier C	186.25	13.75
Supplier D	186.25	13.75

(b) The Pearson chi-square statistic is $X^2 = 7.087$.

(c) The likelihood ratio chi-square statistic is $G^2 = 6.889$.

(d) The p-values are $P(\chi_3^2 \geq 7.087) = 0.069$ and $P(\chi_3^2 \geq 6.889) = 0.076$ where the degrees of freedom of the chi-square random variable are calculated as $(4-1) \times (2-1) = 3$.

(e) The null hypothesis that the defective rates are identical for the four suppliers is accepted at size $\alpha = 0.05$.

(f) With $z_{0.025} = 1.960$ the confidence interval is

$\frac{10}{200} \pm \frac{1.960}{200} \times \sqrt{\frac{10\times(200-10)}{200}}$

$= (0.020, 0.080)$.

(g) With $z_{0.025} = 1.960$ the confidence interval is

$\frac{15}{200} - \frac{21}{200} \pm 1.960 \times \sqrt{\frac{15\times(200-15)}{200^3} + \frac{21\times(200-21)}{200^3}}$

$= (-0.086, 0.026)$.

10.4.2 The expected cell frequencies are

	Dead	Slow growth	Medium growth	Strong growth
No fertilizer	57.89	93.84	172.09	163.18
Fertilizer I	61.22	99.23	181.98	172.56
Fertilizer II	62.89	101.93	186.93	177.25

The Pearson chi-square statistic is $X^2 = 13.66$.

The p-value is $P(\chi_6^2 \geq 13.66) = 0.034$ where the degrees of freedom of the chi-square random variable are $(3-1) \times (4-1) = 6$.

There is a fairly strong suggestion that the seedlings growth pattern is different for the different growing conditions, although the evidence is not overwhelming.

10.4.3 The expected cell frequencies are

	Formulation I	Formulation II	Formulation III
10-25	75.00	74.33	50.67
26-50	75.00	74.33	50.67
≥ 51	75.00	74.33	50.67

The Pearson chi-square statistic is $X^2 = 6.11$.

The p-value is $P(\chi_4^2 \geq 6.11) = 0.191$ where the degrees of freedom of the chi-square random variable are calculated as $(3-1) \times (3-1) = 4$.

There is *not* sufficient evidence to conclude that the preferences for the different formulations change with age.

10.4.4 (a) The expected cell frequencies are

	Pass	Fail
Line 1	166.2	13.8
Line 2	166.2	13.8
Line 3	166.2	13.8
Line 4	166.2	13.8
Line 5	166.2	13.8

The Pearson chi-square statistic is $X^2 = 13.72$.

The p-value is $P(\chi_4^2 \geq 13.72) = 0.008$ where the degrees of freedom of the chi-square random variable are calculated as $(5-1) \times (2-1) = 4$.

There is sufficient evidence to conclude that the pass rates are different for the five production lines.

(b) With $z_{0.025} = 1.960$ the confidence interval is

$\frac{11}{180} - \frac{15}{180} \pm 1.960 \times \sqrt{\frac{11\times(180-11)}{180^3} + \frac{15\times(180-15)}{180^3}}$

$= (-0.076, 0.031)$.

10.4.5 The expected cell frequencies are

	Completely satisfied	Somewhat satisfied	Not satisfied
Technician 1	71.50	22.36	4.14
Technician 2	83.90	26.24	4.86
Technician 3	45.96	14.37	2.66
Technician 4	57.64	18.03	3.34

The Pearson chi-square statistic is $X^2 = 32.11$.

The p-value is $P(\chi_6^2 \geq 32.11) = 0.000$ where the degrees of freedom of the chi-square random variable are calculated as $(4-1) \times (3-1) = 6$.

There is sufficient evidence to conclude that some technicians are better than others in satisfying their customers.

Note: In this analysis 4 of the cells have expected values less than 5 and it may be preferable to pool together the categories "somewhat satisfied" and "not satisfied". In this case the Pearson chi-square statistic is $X^2 = 31.07$ and comparison with a chi-square distribution with 3 degrees of freedom again gives a p-value of 0.000. The conclusion remains the same.

10.4.7 (a) The expected cell frequencies are

	Less than one week	More than one week
Standard drug	88.63	64.37
New drug	79.37	57.63

The Pearson chi-square statistic is $X^2 = 15.71$.

The p-value is $P(\chi_1^2 \geq 15.71) = 0.0000$ where the degrees of freedom of the chi-square random variable are calculated as $(2-1) \times (2-1) = 1$.

There is sufficient evidence to conclude that $p_s \neq p_n$.

(b) With $z_{0.005} = 2.576$ the confidence interval is

$\frac{72}{153} - \frac{96}{137} \pm 2.576 \times \sqrt{\frac{72\times(153-72)}{153^3} + \frac{96\times(137-96)}{137^3}}$

$= (-0.375, -0.085)$.

10.4.8 The Pearson chi-square statistic is

$X^2 = \frac{1986\times(1078\times111-253\times544)^2}{1331\times655\times1622\times364} = 1.247$

which gives a p-value of $P(\chi_1^2 \geq 1.247) = 0.264$ where the degrees of freedom of the chi-square random variable are calculated as $(2-1) \times (2-1) = 1$.

It is plausible that the completeness of the structure and the etch depth are independent factors.

10.4.9 The expected cell frequencies are

Type	Warranty purchased	Warranty not purchased
A	34.84	54.16
B	58.71	91.29
C	43.45	67.55

The Pearson chi-square statistic is $X^2 = 2.347$.

The p-value is $P(\chi_2^2 \geq 2.347) = 0.309$.

The null hypothesis of independence is plausible and there is not sufficient evidence to conclude that the probability of a customer purchasing the extended warranty is different for the three product types.

10.4.10 The expected cell frequencies are

Type	Minor cracking	Medium cracking	Severe cracking
A	35.77	13.09	8.14
B	30.75	11.25	7.00
C	56.48	20.66	12.86

The Pearson chi-square statistic is $X^2 = 5.024$.

The p-value is $P(\chi_4^2 \geq 5.024) = 0.285$.

The null hypothesis of independence is plausible and there is not sufficient evidence to conclude that the three types of asphalt are different with respect to cracking.

10.6 Supplementary Problems

10.6.1 With $z_{0.025} = 1.960$ the confidence interval is

$$\left(\frac{27}{60} - \frac{1.960}{60} \times \sqrt{\frac{27\times(60-27)}{60}}, \frac{27}{60} + \frac{1.960}{60} \times \sqrt{\frac{27\times(60-27)}{60}}\right)$$

$= (0.324, 0.576)$.

10.6.2 Let p be the probability that a bag of flour is underweight and consider the hypotheses

$H_0 : p \leq \frac{1}{40} = 0.025$ versus $H_A : p > \frac{1}{40} = 0.025$

where the alternative hypothesis states that the consumer watchdog organization can take legal action.

The statistic for the normal approximation to the p-value is

$$z = \frac{x-np_0}{\sqrt{np_0(1-p_0)}} = \frac{18-(500\times 0.025)}{\sqrt{500\times 0.025\times(1-0.025)}} = 1.575$$

and the p-value is $1 - \Phi(1.575) = 0.058$.

There is a fairly strong suggestion that the proportion of underweight bags is more than 1 in 40 although the evidence is not overwhelming.

10.6.3 Let p be the proportion of customers who request the credit card.

With $z_{0.005} = 2.576$ a 99% two-sided confidence interval for p is

$$\left(\frac{384}{5000} - \frac{2.576}{5000} \times \sqrt{\frac{384\times(5000-384)}{5000}}, \frac{384}{5000} + \frac{2.576}{5000} \times \sqrt{\frac{384\times(5000-384)}{5000}}\right)$$

$= (0.0671, 0.0865)$.

The number of customers out of 1,000,000 who request the credit card can be estimated as being between 67,100 and 86,500.

10.6.4 Let p_A be the probability that an operation performed in the morning is a total success and let p_B be the probability that an operation performed in the afternoon is a total success.

With $z_{0.05} = 1.645$ a 95% lower confidence bound for $p_A - p_B$ is

$$\left(\frac{443}{564} - \frac{388}{545} - 1.645 \times \sqrt{\frac{443\times(564-443)}{564^3} + \frac{388\times(545-388)}{545^3}}, 1\right)$$

$= (0.031, 1)$.

Consider the hypotheses

$H_0 : p_A \leq p_B$ versus $H_A : p_A > p_B$

where the alternative hypothesis states that the probability that an operation is a total success is smaller in the afternoon than in the morning.

With the pooled probability estimate

$\hat{p} = \frac{x+y}{n+m} = \frac{443+388}{564+545} = 0.749$

the test statistic is

$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{\frac{443}{564} - \frac{388}{545}}{\sqrt{0.749 \times (1-0.749) \times \left(\frac{1}{564}+\frac{1}{545}\right)}} = 2.822$

and the p-value is $1 - \Phi(2.822) = 0.002$.

There is sufficient evidence to conclude that the probability that an operation is a total success is smaller in the afternoon than in the morning.

10.6.5 Let p_A be the probability that a householder with an income above \$60,000 supports the tax increase and let p_B be the probability that a householder with an income below \$60,000 supports the tax increase.

With $z_{0.025} = 1.960$ a 95% two-sided confidence interval for $p_A - p_B$ is

$\frac{32}{106} - \frac{106}{221} \pm 1.960 \times \sqrt{\frac{32 \times (106-32)}{106^3} + \frac{106 \times (221-106)}{221^3}}$

$= (-0.287, -0.068)$.

Consider the hypotheses

$H_0 : p_A = p_B$ versus $H_A : p_A \neq p_B$

where the alternative hypothesis states that the support for the tax increase does depend upon the householder's income.

With the pooled probability estimate

$\hat{p} = \frac{x+y}{n+m} = \frac{32+106}{106+221} = 0.422$

the test statistic is

$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{\frac{32}{106} - \frac{106}{221}}{\sqrt{0.422 \times (1-0.422) \times \left(\frac{1}{106}+\frac{1}{221}\right)}} = -3.05$

and the p-value is $2 \times \Phi(-3.05) = 0.002$.

There is sufficient evidence to conclude that the support for the tax increase does depend upon the householder's income.

10.6.6 The expected cell frequencies are:

$e_1 = 619 \times 0.1 = 61.9$

$e_2 = 619 \times 0.8 = 495.2$

$e_3 = 619 \times 0.1 = 61.9$

The Pearson chi-square statistic is

$X^2 = \frac{(61-61.9)^2}{61.9} + \frac{(486-495.2)^2}{495.2} + \frac{(72-61.9)^2}{61.9} = 1.83$

so that the p-value is $P(\chi_2^2 \geq 3.62) = 0.400$.

These probability values are plausible.

10.6.7 A Poisson distribution with mean $\lambda = \bar{x} = 2.95$ can be considered.

The last two cells can be pooled so that there are 8 cells altogether.

The Pearson chi-square statistic is $X^2 = 13.1$ and the p-value is
$P(\chi_6^2 \geq 13.1) = 0.041$.

There is some evidence that a Poisson distribution is not appropriate although the evidence is not overwhelming.

10.6.8 If the random numbers have a uniform distribution then the expected cell frequencies are $e_i = 1000$.

The Pearson chi-square statistic is $X^2 = 9.07$ and the
p-value is $P(\chi_9^2 \geq 9.07) = 0.431$.

There is no evidence that the random number generator is not operating correctly.

10.6.10 The expected cell frequencies are

	A	B	C
This year	112.58	78.18	30.23
Last year	211.42	146.82	56.77

The Pearson chi-square statistic is $X^2 = 1.20$.

The p-value is $P(\chi_2^2 \geq 1.20) = 0.549$ where the degrees of freedom of the chi-square random variable are calculated as $(2-1) \times (3-1) = 2$.

There is *not* sufficient evidence to conclude that there has been a change in preferences for the three types of tire between the two years.

10.6.11 The expected cell frequencies are

	Completely healed	Partially healed	No change
Treatment 1	19.56	17.81	6.63
Treatment 2	22.22	20.24	7.54
Treatment 3	14.22	12.95	4.83

The Pearson chi-square statistic is $X^2 = 5.66$.

The p-value is $P(\chi_4^2 \geq 5.66) = 0.226$ where the degrees of freedom of the chi-square random variable are calculated as $(3-1) \times (3-1) = 4$.

There is *not* sufficient evidence to conclude that the three medications are not equally effective.

10.6.12 The expected cell frequencies are

	Computers	Library
Engineering	72.09	70.91
Arts & Sciences	49.91	49.09

The Pearson chi-square statistic is $X^2 = 4.28$.

The p-value is $P(\chi_1^2 \geq 4.28) = 0.039$ where the degrees of freedom of the chi-square random variable are calculated as $(2-1) \times (2-1) = 1$.

There is a fairly strong suggestion that the opinions differ between the two colleges but the evidence is not overwhelming.

10.6.13 (a) Let p be the probability that a part has a length outside the specified tolerance range, and consider the hypotheses
$H_0 : p \leq 0.10$ versus $H_A : p > 0.10$
where the alternative hypothesis states that the probability that a part has a length outside the specified tolerance range is larger than 10%.

The statistic for the normal approximation to the p-value is

$z = \frac{x - np_0 - 0.5}{\sqrt{np_0(1-p_0)}} = \frac{445 - (3877 \times 0.10) - 0.5}{\sqrt{3877 \times 0.10 \times (1-0.10)}} = 3.041$

and the p-value is $1 - \Phi(3.041) = 0.0012$.

There is sufficient evidence to conclude that the probability that a part has a length outside the specified tolerance range is larger than 10%.

(b) With $z_{0.01} = 2.326$ the confidence interval is

$\left(\frac{445}{3877} - \frac{2.326}{3877}\sqrt{\frac{445(3877-445)}{3877}}, 1\right)$

$= (0.103, 1)$.

(c) The Pearson chi-square statistic is

$X^2 = \frac{3877 \times (161 \times 420 - 3271 \times 25)^2}{186 \times 3691 \times 3432 \times 445} = 0.741$

which gives a p-value of $P(\chi_1^2 \geq 0.741) = 0.389$ where the degrees of freedom of the chi-square random variable are calculated as $(2-1) \times (2-1) = 1$.

It is plausible that the acceptability of the length and the acceptability of the width of the parts are independent of each other.

10.6.14 (a) The expected cell frequencies are:

$800 \times 0.80 = 640$

$800 \times 0.15 = 120$

$800 \times 0.05 = 40$

The Pearson chi-square statistic is

$X^2 = \frac{(619-640)^2}{640} + \frac{(124-120)^2}{120} + \frac{(57-40)^2}{40} = 8.047$

and the likelihood ratio chi-square statistic is

$G^2 = 2 \times \left(619 \ln\left(\frac{619}{640}\right) + 124 \ln\left(\frac{124}{120}\right) + 57 \ln\left(\frac{57}{40}\right)\right) = 7.204.$

The p-values are $P(\chi_2^2 \geq 8.047) = 0.018$ and $P(\chi_2^2 \geq 7.204) = 0.027$.

There is some evidence that the claims made by the research report are incorrect, although the evidence is not overwhelming.

(b) With $z_{0.01} = 2.326$ the confidence interval is

$\left(0, \frac{57}{800} + \frac{2.326}{800}\sqrt{\frac{57(800-57)}{800}}\right)$

$= (0, 0.092)$.

10.6.15 The expected cell frequencies are

	Weak	Satisfactory	Strong
Preparation method 1	13.25	42.65	15.10
Preparation method 2	23.51	75.69	26.80
Preparation method 3	13.25	42.65	15.10

The Pearson chi-square statistic is $X^2 = 16.797$ and the p-value is

$P(\chi_4^2 \geq 16.797) = 0.002$

where the degrees of freedom of the chi-square random variable are calculated as $(3-1) \times (3-1) = 4$.

There is sufficient evidence to conclude that the three preparation methods are not equivalent in terms of the quality of chemical solutions which they produce.

10.6.16 (a) The expected cell frequencies are

	No damage	Slight damage	Medium damage	Severe damage
Type I	87.33	31.33	52.00	49.33
Type II	87.33	31.33	52.00	49.33
Type III	87.33	31.33	52.00	49.33

The Pearson chi-square statistic is $X^2 = 50.08$ so that the p-value is

$P(\chi_6^2 \geq 50.08) = 0.000$

where the degrees of freedom of the chi-square random variable are calculated as $(4-1) \times (3-1) = 6$.

Consequently, there is sufficient evidence to conclude that the three types of metal alloy are not all the same in terms of the damage that they suffer.

(b) $\hat{p}_{Se1} = \frac{42}{220} = 0.1911$

$\hat{p}_{Se3} = \frac{32}{220} = 0.1455$

The pooled estimate is

$\hat{p} = \frac{42+32}{220+220} = 0.1682.$

The test statistic is

$z = \frac{0.1911-0.1455}{\sqrt{0.1682\times0.8318\times\left(\frac{1}{220}+\frac{1}{220}\right)}} = 1.27$

and the p-value is $2 \times \Phi(-1.27) = 0.20$.

There is not sufficient evidence to conclude that the probability of suffering severe damage is different for alloys of type I and type III.

(c) $\hat{p}_{N2} = \frac{52}{220} = 0.236$

With $z_{0.005} = 2.576$ the confidence interval is

$0.236 \pm \frac{2.576}{220}\sqrt{\frac{52\times(220-52)}{220}}$

$= (0.163, 0.310)$.

10.6.17 (a) The expected cell frequencies are:

$e_1 = 655 \times 0.25 = 163.75$

$e_2 = 655 \times 0.10 = 65.50$

$e_3 = 655 \times 0.40 = 262.00$

$e_4 = 655 \times 0.25 = 163.75$

The Pearson chi-square statistic is

$X^2 = \frac{(119-163.75)^2}{163.75} + \frac{(54-65.50)^2}{65.50} + \frac{(367-262.00)^2}{262.00} + \frac{(115-163.75)^2}{163.75} = 70.8$

so that the p-value is $P(\chi^2_3 \geq 70.8) = 0.000$.

The data is not consistent with the claimed probabilities.

(b) With $z_{0.005} = 2.576$ the confidence interval is

$p_C \in \frac{367}{655} \pm \frac{2.576}{655}\sqrt{\frac{367\times(655-367)}{655}}$

$= (0.510, 0.610)$.

10.6.18 $P(N(120, 4^2) \leq 115) = P\left(N(0,1) \leq \frac{115-120}{4}\right) = \Phi(-1.25) = 0.1056$

$P(N(120, 4^2) \leq 120) = P\left(N(0,1) \leq \frac{120-120}{4}\right) = \Phi(0) = 0.5000$

$P(N(120, 4^2) \leq 125) = P\left(N(0,1) \leq \frac{125-120}{4}\right) = \Phi(1.25) = 0.8944$

The observed cell frequencies are $x_1 = 17$, $x_2 = 32$, $x_3 = 21$, and $x_4 = 14$.

The expected cell frequencies are:

$e_1 = 84 \times 0.1056 = 8.87$

$e_2 = 84 \times (0.5000 - 0.1056) = 33.13$

$e_3 = 84 \times (0.8944 - 0.5000) = 33.13$

$e_4 = 84 \times (1 - 0.8944) = 8.87$

The Pearson chi-square statistic is

$X^2 = \frac{(17-8.87)^2}{8.87} + \frac{(32-33.13)^2}{33.13} + \frac{(21-33.13)^2}{33.13} + \frac{(14-8.87)^2}{8.87} = 14.88$

so that the p-value is $P(\chi_3^2 \geq 14.88) = 0.002$.

There is sufficient evidence to conclude that the breaking strength of concrete of this type is not normally distributed with a mean of 120 and a standard deviation of 4.

10.6.19 (a) $\hat{p}_M = \frac{28}{64} = 0.438$

$\hat{p}_F = \frac{31}{85} = 0.365$

The hypotheses are
$H_0 : p_M = p_F$ versus $H_A : p_M \neq p_F$
and the pooled estimate is

$\hat{p} = \frac{28+31}{64+85} = 0.396$.

The test statistic is

$z = \frac{0.438-0.365}{\sqrt{0.396 \times 0.604 \times \left(\frac{1}{64}+\frac{1}{85}\right)}} = 0.90$

and the p-value is $2 \times \Phi(-0.90) = 0.37$.

There is not sufficient evidence to conclude that the support for the proposal is different for men and women.

(b) With $z_{0.005} = 2.576$ the confidence interval is

$p_M - p_F \in 0.438 - 0.365 \pm 2.576\sqrt{\frac{28\times 36}{64^3} + \frac{31\times 54}{85^3}}$

$= (-0.14, 0.28)$.

10.6.20 (a) $\hat{p}_A = \frac{56}{94} = 0.596$

$\hat{p}_B = \frac{64}{153} = 0.418$

The hypotheses are
$H_0 : p_A \leq 0.5$ versus $H_A : p_A > 0.5$
and the test statistic is

$z = \frac{56-(94\times0.5)-0.5}{\sqrt{94\times0.5\times0.5}} = 1.753$

so that the p-value is $1 - \Phi(1.753) = 0.040$.

There is some evidence that the chance of success for patients with Condition A is better than 50%, but the evidence is not overwhelming.

(b) With $z_{0.005} = 2.576$ the confidence interval is

$p_A - p_B \in 0.596 - 0.418 \pm 2.576\sqrt{\frac{56\times38}{94^3} + \frac{64\times89}{153^3}}$

$= (0.012, 0.344)$.

(c) The Pearson chi-square statistic is

$X^2 = \frac{n(x_{11}x_{22}-x_{12}x_{21})^2}{x_{1.}x_{.1}x_{2.}x_{.2}} = \frac{247\times(56\times89-38\times64)^2}{94\times120\times153\times127} = 7.34$

and the p-value is $P(\chi_1^2 \geq 7.34) = 0.007$.

There is sufficient evidence to conclude that the success probabilities are different for patients with Condition A and with Condition B.

10.6.21 (a) True

(b) True

(c) False

(d) False

(e) True

(f) True

(g) True

(h) True

(i) True

(j) False

10.6.22 (a) $\hat{p} = \frac{485}{635} = 0.764$

With $z_{0.025} = 1.960$ the confidence interval is

$0.764 \pm \frac{1.960}{635} \times \sqrt{\frac{485\times(635-485)}{635}}$

$= 0.764 \pm 0.033$

$= (0.731, 0.797)$.

(b) The hypotheses are
$H_0 : p \leq 0.75$ versus $H_A : p > 0.75$
and the test statistic is

$z = \frac{485-(635\times0.75)-0.5}{\sqrt{635\times0.75\times0.25}} = 0.756$

so that the p-value is $1 - \Phi(0.756) = 0.225$.

There is not sufficient evidence to establish that at least 75% of the customers are satisfied.

10.6.23 (a) The expected cell frequencies are

	Hospital 1	Hospital 2	Hospital 3	Hospital 4	Hospital 5
Admitted	38.07	49.60	21.71	74.87	50.75
Returned home	324.93	423.40	185.29	639.13	433.25

The Pearson chi-square statistic is $X^2 = 26.844$
so that the p-value is $P(\chi_4^2 \geq 26.844) \simeq 0$.

Consequently, there is sufficient evidence to support the claim that the hospital admission rates differ between the five hospitals.

(b) $\hat{p}_3 = \frac{42}{207} = 0.203$

$\hat{p}_4 = \frac{57}{714} = 0.080$

With $z_{0.25} = 1.960$ the confidence interval is

$p_3 - p_4 \in 0.203 - 0.080 \pm 1.960\sqrt{\frac{0.203\times0.797}{207} + \frac{0.080\times0.920}{714}}$

$= 0.123 \pm 0.058$

$= (0.065, 0.181)$.

(c) $\hat{p}_1 = \frac{39}{363} = 0.107$

The hypotheses are
$H_0 : p_1 \leq 0.1$ versus $H_A : p_1 > 0.1$
and the test statistic is

$z = \frac{39-(363\times0.1)-0.5}{\sqrt{363\times0.1\times0.9}} = 0.385$

so that the p-value is $1 - \Phi(0.385) = 0.35$.

There is not sufficient evidence to conclude that the admission rate for hospital 1 is larger than 10%.

10.6.24 (a) The expected cell frequencies are

	Minimal scour depth	Substantial scour depth	Severe scour depth
Pier design 1	7.86	13.82	7.32
Pier design 2	8.13	14.30	7.57
Pier design 3	13.01	22.88	12.11

The Pearson chi-square statistic is

$X^2 = \sum_{i=1}^{3} \sum_{j=1}^{3} \frac{(x_{ij}-e_{ij})^2}{e_{ij}} = 17.41$

so that the p-value is $P(\chi^2_4 \geq 17.41) = 0.002$.

Consequently, there is sufficient evidence to conclude that the pier design has an effect on the amount of scouring.

The likelihood ratio chi-square statistic is

$G^2 = 2\sum_{i=1}^{3} \sum_{j=1}^{3} x_{ij} \ln\left(\frac{x_{ij}}{e_{ij}}\right) = 20.47$

which provides a similar conclusion.

(b) The expected cell frequencies are

$e_1 = e_2 = e_3 = \frac{29}{3}$

and the Pearson chi-square statistic is

$X^2 = \frac{(12-\frac{29}{3})^2}{\frac{29}{3}} + \frac{(15-\frac{29}{3})^2}{\frac{29}{3}} + \frac{(2-\frac{29}{3})^2}{\frac{29}{3}} = 9.59$

so that the p-value is $P(\chi^2_2 \geq 9.59) = 0.008$.

Consequently, the hypothesis of homogeneity is not plausible and the data set provides sufficient evidence to conclude that for pier design 1 the three levels of scouring are not equally likely.

(c) Let p_{3m} be the probability of minimal scour depth with pier design 3, so that the hypotheses of interest are
$H_0 : p_{3m} = 0.25$ versus $H_A : p_{3m} \neq 0.25$.

Since

$\hat{p}_{3m} = \frac{x}{n} = \frac{15}{48} = 0.3125 > 0.25$

the exact p-value is $2 \times P(B(48, 0.25) \geq 15)$.

With a test statistic

$z = \frac{x - np_0}{\sqrt{np_0(1-p_0)}} = \frac{15 - 48(0.25)}{\sqrt{48(0.25)(1-0.25)}} = \frac{3}{3} = 1$

the normal approximation to the p-value is

$2\Phi(-|z|) = 2\Phi(-1) = 0.3174$.

Consequently, the null hypothesis is not rejected and it is plausible that the probability of minimal scour for pier design 3 is 25%.

(d) $\hat{p}_{1s} = \frac{2}{29} = 0.0690$

$\hat{p}_{2s} = \frac{8}{30} = 0.2667$

With $z_{0.005} = 2.576$ the confidence interval is

$p_{1s} - p_{2s} \in 0.0690 - 0.2667 \pm 2.576\sqrt{\frac{0.0690(1-0.0690)}{29} + \frac{0.2667(1-0.2667)}{30}}$

$= -0.1977 \pm 0.2407$

$= (-0.4384, 0.0430)$.

Chapter 11

The Analysis of Variance

11.1 One Factor Analysis of Variance

11.1.1 (a) $P(X \geq 4.2) = 0.0177$

(b) $P(X \geq 2.3) = 0.0530$

(c) $P(X \geq 31.7) \leq 0.0001$

(d) $P(X \geq 9.3) = 0.0019$

(e) $P(X \geq 0.9) = 0.5010$

11.1.2

Source	df	SS	MS	F	p-value
Treatments	5	557.0	111.4	5.547	0.0017
Error	23	461.9	20.08		
Total	28	1018.9			

11.1.3

Source	df	SS	MS	F	p-value
Treatments	7	126.95	18.136	5.01	0.0016
Error	22	79.64	3.62		
Total	29	206.59			

11.1.4

Source	df	SS	MS	F	p-value
Treatments	6	7.66	1.28	0.78	0.59
Error	77	125.51	1.63		
Total	83	133.18			

11.1.5

Source	df	SS	MS	F	p-value
Treatments	3	162.19	54.06	6.69	0.001
Error	40	323.34	8.08		
Total	43	485.53			

11.1.6

Source	df	SS	MS	F	p-value
Treatments	2	46.8	23.4	2.7	0.08
Error	52	451.2	8.7		
Total	54	498.0			

11.1.7

Source	df	SS	MS	F	p-value
Treatments	3	0.0079	0.0026	1.65	0.189
Error	52	0.0829	0.0016		
Total	55	0.0908			

11.1.8 (a) $\mu_1 - \mu_2 \in \left(48.05 - 44.74 - \frac{\sqrt{4.96}\times 3.49}{\sqrt{11}}, 48.05 - 44.74 + \frac{\sqrt{4.96}\times 3.49}{\sqrt{11}}\right)$

$= (0.97, 5.65)$

$\mu_1 - \mu_3 \in \left(48.05 - 49.11 - \frac{\sqrt{4.96}\times 3.49}{\sqrt{11}}, 48.05 - 49.11 + \frac{\sqrt{4.96}\times 3.49}{\sqrt{11}}\right)$

$= (-3.40, 1.28)$

$\mu_2 - \mu_3 \in \left(44.74 - 49.11 - \frac{\sqrt{4.96}\times 3.49}{\sqrt{11}}, 44.74 - 49.11 + \frac{\sqrt{4.96}\times 3.49}{\sqrt{11}}\right)$

$= (-6.71, -2.03)$

(c) The total sample size required from each factor level can be estimated as

$n \geq \frac{4\ s^2\ q^2_{\alpha,k,\nu}}{L^2} = \frac{4\times 4.96\times 3.49^2}{2.0^2} = 60.4$

so that an additional sample size of $61 - 11 = 50$ observations from each factor level can be recommended.

11.1.9 (a) $\mu_1 - \mu_2 \in \left(136.3 - 152.1 - \frac{\sqrt{15.95}\times 4.30}{\sqrt{6}}, 136.3 - 152.1 + \frac{\sqrt{15.95}\times 4.30}{\sqrt{6}}\right)$

$= (-22.8, -8.8)$

$\mu_1 - \mu_3 \in (3.6, 17.6)$

$\mu_1 - \mu_4 \in (-0.9, 13.1)$

$\mu_1 - \mu_5 \in (-13.0, 1.0)$

$\mu_1 - \mu_6 \in (1.3, 15.3)$

$\mu_2 - \mu_3 \in (19.4, 33.4)$

$\mu_2 - \mu_4 \in (14.9, 28.9)$

$\mu_2 - \mu_5 \in (2.8, 16.8)$

$\mu_2 - \mu_6 \in (17.1, 31.1)$

$\mu_3 - \mu_4 \in (-11.5, 2.5)$

$\mu_3 - \mu_5 \in (-23.6, -9.6)$

$\mu_3 - \mu_6 \in (-9.3, 4.7)$

$\mu_4 - \mu_5 \in (-19.1, -5.1)$

$\mu_4 - \mu_6 \in (-4.8, 9.2)$

$\mu_5 - \mu_6 \in (7.3, 21.3)$

(c) The total sample size required from each factor level can be estimated as

$n \geq \frac{4\, s^2\, q^2_{\alpha,k,\nu}}{L^2} = \frac{4 \times 15.95 \times 4.30^2}{10.0^2} = 11.8$

so that an additional sample size of $12 - 6 = 6$ observations from each factor level can be recommended.

11.1.10 The p-value remains unchanged.

11.1.11 (a) $\bar{x}_{1.} = 5.633$
$\bar{x}_{2.} = 5.567$
$\bar{x}_{3.} = 4.778$

(b) $\bar{x}_{..} = 5.326$

(c) $SSTR = 4.076$

(d) $\sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}^2 = 791.30$

(e) $SST = 25.432$

(f) $SSE = 21.356$

(g)

Source	df	SS	MS	F	p-value
Treatments	2	4.076	2.038	2.29	0.123
Error	24	21.356	0.890		
Total	26	25.432			

(h) $\mu_1 - \mu_2 \in \left(5.633 - 5.567 - \frac{\sqrt{0.890} \times 3.53}{\sqrt{9}}, 5.633 - 5.567 + \frac{\sqrt{0.890} \times 3.53}{\sqrt{9}}\right)$

$= (-1.04, 1.18)$

$\mu_1 - \mu_3 \in \left(5.633 - 4.778 - \frac{\sqrt{0.890} \times 3.53}{\sqrt{9}}, 5.633 - 4.778 + \frac{\sqrt{0.890} \times 3.53}{\sqrt{9}}\right)$

$= (-0.25, 1.97)$

$\mu_2 - \mu_3 \in \left(5.567 - 4.778 - \frac{\sqrt{0.890} \times 3.53}{\sqrt{9}}, 5.567 - 4.778 + \frac{\sqrt{0.890} \times 3.53}{\sqrt{9}}\right)$

$= (-0.32, 1.90)$

(j) The total sample size required from each factor level can be estimated as

$n \geq \frac{4\, s^2\, q^2_{\alpha,k,\nu}}{L^2} = \frac{4 \times 0.890 \times 3.53^2}{1.0^2} = 44.4$

so that an additional sample size of $45 - 9 = 36$ observations from each factor level can be recommended.

11.1.12 (a) $\bar{x}_{1.} = 10.560$

$\bar{x}_{2.} = 15.150$

$\bar{x}_{3.} = 17.700$

$\bar{x}_{4.} = 11.567$

(b) $\bar{x}_{..} = 14.127$

(c) $SSTR = 364.75$

(d) $\sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}^2 = 9346.74$

(e) $SST = 565.23$

(f) $SSE = 200.47$

(g)

Source	df	SS	MS	F	p-value
Treatments	3	364.75	121.58	24.26	0.000
Error	40	200.47	5.01		
Total	43	565.23			

(h) $\mu_1 - \mu_2 \in (-7.16, -2.02)$

$\mu_1 - \mu_3 \in (-9.66, -4.62)$

$\mu_1 - \mu_4 \in (-3.76, 1.75)$

$\mu_2 - \mu_3 \in (-4.95, -0.15)$

$\mu_2 - \mu_4 \in (0.94, 6.23)$

$\mu_3 - \mu_4 \in (3.53, 8.74)$

Note: In the remainder of this section the confidence intervals for the pairwise differences of the factor level means are provided with an overall confidence level of 95%.

11.1.13

Source	df	SS	MS	F	p-value
Treatments	2	0.0085	0.0042	0.24	0.787
Error	87	1.5299	0.0176		
Total	89	1.5384			

$\mu_1 - \mu_2 \in (-0.08, 0.08)$

$\mu_1 - \mu_3 \in (-0.06, 0.10)$

$\mu_2 - \mu_3 \in (-0.06, 0.10)$

There is *not* sufficient evidence to conclude that there is a difference between the three production lines.

11.1.14

Source	df	SS	MS	F	p-value
Treatments	2	278.0	139.0	85.4	0.000
Error	50	81.3	1.63		
Total	52	359.3			

$\mu_1 - \mu_2 \in (3.06, 5.16)$

$\mu_1 - \mu_3 \in (4.11, 6.11)$

$\mu_2 - \mu_3 \in (-0.08, 2.08)$

There is sufficient evidence to conclude that Monday is slower than the other two days.

11.1.15

Source	df	SS	MS	F	p-value
Treatments	2	0.0278	0.0139	1.26	0.299
Error	30	0.3318	0.0111		
Total	32	0.3596			

$\mu_1 - \mu_2 \in (-0.15, 0.07)$

$\mu_1 - \mu_3 \in (-0.08, 0.14)$

$\mu_2 - \mu_3 \in (-0.04, 0.18)$

There is *not* sufficient evidence to conclude that the radiation readings are affected by the background radiation levels.

11.1.16

Source	df	SS	MS	F	p-value
Treatments	2	121.24	60.62	52.84	0.000
Error	30	34.42	1.15		
Total	32	155.66			

$\mu_1 - \mu_2 \in (-5.12, -2.85)$

$\mu_1 - \mu_3 \in (-0.74, 1.47)$

$\mu_2 - \mu_3 \in (3.19, 5.50)$

There is sufficient evidence to conclude that layout 2 is slower than the other two layouts.

11.1.17

Source	df	SS	MS	F	p-value
Treatments	2	0.4836	0.2418	7.13	0.001
Error	93	3.1536	0.0339		
Total	95	3.6372			

$\mu_1 - \mu_2 \in (-0.01, 0.22)$

$\mu_1 - \mu_3 \in (0.07, 0.29)$

$\mu_2 - \mu_3 \in (-0.03, 0.18)$

There is sufficient evidence to conclude that the average particle diameter is larger at the low amount of stabilizer than at the high amount of stabilizer.

11.1.18

Source	df	SS	MS	F	p-value
Treatments	2	135.15	67.58	19.44	0.000
Error	87	302.50	3.48		
Total	89	437.66			

$\mu_1 - \mu_2 \in (-1.25, 1.04)$

$\mu_1 - \mu_3 \in (1.40, 3.69)$

$\mu_2 - \mu_3 \in (1.50, 3.80)$

There is sufficient evidence to conclude that method 3 is quicker than the other two methods.

11.1.19 $\bar{x}_{\cdot\cdot} = \frac{(8\times 42.91)+(11\times 44.03)+(10\times 43.72)}{8+11+10} = \frac{1264.81}{29} = 43.61$

$SSTr = (8 \times 42.91^2) + (11 \times 44.03^2) + (10 \times 43.72^2) - \frac{1264.81^2}{29} = 5.981$

$SSE = (7 \times 5.33^2) + (10 \times 4.01^2) + (9 \times 5.10^2) = 593.753$

Source	df	SS	MS	F	p-value
Treatments	2	5.981	2.990	0.131	0.878
Error	26	593.753	22.837		
Total	28	599.734			

There is not sufficient evidence to conclude that there is a difference between the catalysts in terms of the strength of the compound.

11.1.20 (a) $\bar{x}_{1.} = 33.6$

$\bar{x}_{2.} = 40.0$

$\bar{x}_{3.} = 20.4$

$\bar{x}_{4.} = 31.0$

$\bar{x}_{5.} = 26.5$

Source	df	SS	MS	F	p-value
Treatments	4	1102.7	275.7	18.51	0.000
Error	20	297.9	14.9		
Total	24	1400.6			

(b) $q_{0.05,5,20} = 4.23$

$s = \sqrt{MSE} = \sqrt{14.9} = 3.86$

The pairwise comparisons which contain zero are:
treatment 1 and treatment 2
treatment 1 and treatment 4
treatment 3 and treatment 5
treatment 4 and treatment 5

The treatment with the largest average quality score is either treatment 1 or treatment 2.

The treatment with the smallest average quality score is either treatment 3 or treatment 5.

11.1.21 $q_{0.05,5,43} = 4.04$

With a 95% confidence level the pairwise confidence intervals that contain zero are:

$\mu_1 - \mu_2$

$\mu_2 - \mu_5$

$\mu_3 - \mu_4$

It can be inferred that the largest mean is either μ_3 or μ_4

and that the smallest mean is either μ_2 or μ_5.

11.1.22 (a) $\bar{x}_{..} = \frac{(8\times10.50)+(8\times9.22)+(9\times6.32)+(6\times11.39)}{31}$

$= 9.1284$

$SSTr = (8 \times 10.50^2) + (8 \times 9.22^2) + (9 \times 6.32^2) + (6 \times 11.39^2) - (31 \times 9.1284^2)$

$= 116.79$

$SSE = (7 \times 1.02^2) + (7 \times 0.86^2) + (8 \times 1.13^2) + (5 \times 0.98^2)$

$= 27.48$

Source	df	SS	MS	F	p-value
Alloy	3	116.79	38.93	38.3	0.000
Error	27	27.48	1.02		
Total	30	144.27			

There is sufficient evidence to conclude that the average strengths of the four metal alloys are not all the same.

(b) $q_{0.05,4,27} = 3.88$

$\mu_1 - \mu_2 \in 10.50 - 9.22 \pm \frac{\sqrt{1.02}\times3.88}{\sqrt{2}}\sqrt{\frac{1}{8}+\frac{1}{8}} = (-0.68, 3.24)$

$\mu_1 - \mu_3 \in 10.50 - 6.32 \pm \frac{\sqrt{1.02}\times3.88}{\sqrt{2}}\sqrt{\frac{1}{8}+\frac{1}{9}} = (2.28, 6.08)$

$\mu_1 - \mu_4 \in 10.50 - 11.39 \pm \frac{\sqrt{1.02}\times3.88}{\sqrt{2}}\sqrt{\frac{1}{8}+\frac{1}{6}} = (-3.00, 1.22)$

$\mu_2 - \mu_3 \in 9.22 - 6.32 \pm \frac{\sqrt{1.02}\times3.88}{\sqrt{2}}\sqrt{\frac{1}{8}+\frac{1}{9}} = (1.00, 4.80)$

$\mu_2 - \mu_4 \in 9.22 - 11.39 \pm \frac{\sqrt{1.02}\times3.88}{\sqrt{2}}\sqrt{\frac{1}{8}+\frac{1}{6}} = (-4.28, -0.06)$

$\mu_3 - \mu_4 \in 6.32 - 11.39 \pm \frac{\sqrt{1.02}\times3.88}{\sqrt{2}}\sqrt{\frac{1}{9}+\frac{1}{6}} = (-7.13, -3.01)$

The strongest metal alloy is either type A or type D.

The weakest metal alloy is type C.

11.1.23 $\bar{x}_{1.} = 40.80$

$\bar{x}_{2.} = 32.80$

$\bar{x}_{3.} = 25.60$

$\bar{x}_{4.} = 50.60$

$\bar{x}_{5.} = 41.80$

$\bar{x}_{6.} = 31.80$

Source	df	SS	MS	F	p-value
Physician	5	1983.8	396.8	15.32	0.000
Error	24	621.6	25.9		
Total	29	2605.4			

The p-value of 0.000 implies that there is sufficient evidence to conclude that the times taken by the physicians for the investigatory surgical procedures are different.

Since

$\frac{s \times q_{0.05,6,24}}{\sqrt{5}} = \frac{\sqrt{25.9} \times 4.37}{\sqrt{5}} = 9.95$

it follows that two physicians cannot be concluded to be different if their sample averages have a difference of less than 9.95.

The slowest physician is either physician 1, physician 4, or physician 5.

The quickest physician is either physician 2, physician 3, or physician 6.

11.1.24 $\bar{x}_{1.} = 29.00$

$\bar{x}_{2.} = 28.75$

$\bar{x}_{3.} = 28.75$

$\bar{x}_{4.} = 37.00$

$\bar{x}_{5.} = 42.00$

Source	df	SS	MS	F	p-value
Treatments	4	596.3	149.08	24.44	0.000
Error	15	91.50	6.10		
Total	19	687.80			

The small p-value in the analysis of variance table implies that there is sufficient evidence to conclude that the E. Coli pollution levels are not the same at all five locations.

Since

$\frac{s \times q_{0.05,5,15}}{\sqrt{n}} = \frac{\sqrt{6.10} \times 4.37}{\sqrt{4}} = 5.40$

the pairwise comparisons reveal that the pollution levels at both locations 4 and 5 are larger than the pollution levels at the other three locations.

The highest E. Coli pollution level is at either location 4 or 5, and the smallest E. Coli pollution level is at either location 1, 2 or 3.

11.1.25 (a) $\bar{x}_{1.} = 46.83$

$\bar{x}_{2.} = 47.66$

$\bar{x}_{3.} = 48.14$

$\bar{x}_{4.} = 48.82$

$\bar{x}_{..} = 47.82$

$SSTr = \sum_{i=1}^{4} n_i(\bar{x}_{i.} - \bar{x}_{..})^2 = 13.77$

Since the p-value is 0.01, the F-statistic in the analysis of variance table must be $F_{0.01,3,24} = 4.72$ so that the complete analysis of variance table is

Source	df	SS	MS	F	p-value
Treatments	3	13.77	4.59	4.72	0.01
Error	24	23.34	0.97		
Total	27	37.11			

(b) With $s = \sqrt{MSE} = 0.986$ and $q_{0.05,4,24} = 3.90$ the pairwise confidence intervals for the treatment means are:

$\mu_1 - \mu_2 \in (-2.11, 0.44)$

$\mu_1 - \mu_3 \in (-2.63, 0.01)$

$\mu_1 - \mu_4 \in (-3.36, -0.62)$

$\mu_2 - \mu_3 \in (-1.75, 0.79)$

$\mu_2 - \mu_4 \in (-2.48, 0.18)$

$\mu_3 - \mu_4 \in (-2.04, 0.70)$

There is sufficient evidence to establish that μ_4 is larger than μ_1.

11.2 Randomized Block Designs

11.2.1

Source	df	SS	MS	F	p-value
Treatments	3	10.15	3.38	3.02	0.047
Blocks	9	24.53	2.73	2.43	0.036
Error	27	30.24	1.12		
Total	39	64.92			

11.2.2

Source	df	SS	MS	F	p-value
Treatments	7	26.39	3.77	3.56	0.0036
Blocks	7	44.16	6.31	5.95	0.0000
Error	49	51.92	1.06		
Total	63	122.47			

11.2.3

Source	df	SS	MS	F	p-value
Treatments	3	58.72	19.57	0.63	0.602
Blocks	9	2839.97	315.55	10.17	0.0000
Error	27	837.96	31.04		
Total	39	3736.64			

11.2.4

Source	df	SS	MS	F	p-value
Treatments	4	240.03	60.01	18.59	0.0000
Blocks	14	1527.12	109.08	33.80	0.0000
Error	56	180.74	3.228		
Total	74	1947.89			

11.2.5 (a)

Source	df	SS	MS	F	p-value
Treatments	2	8.17	4.085	8.96	0.0031
Blocks	7	50.19	7.17	15.72	0.0000
Error	14	6.39	0.456		
Total	23	64.75			

(b) $\mu_1 - \mu_2 \in \left(5.93 - 4.62 - \frac{\sqrt{0.456}\times 3.70}{\sqrt{8}}, 5.93 - 4.62 + \frac{\sqrt{0.456}\times 3.70}{\sqrt{8}}\right)$

$= (0.43, 2.19)$

$\mu_1 - \mu_3 \in \left(5.93 - 4.78 - \frac{\sqrt{0.456}\times 3.70}{\sqrt{8}}, 5.93 - 4.78 + \frac{\sqrt{0.456}\times 3.70}{\sqrt{8}}\right)$

$= (0.27, 2.03)$

$\mu_2 - \mu_3 \in \left(4.62 - 4.78 - \frac{\sqrt{0.456}\times 3.70}{\sqrt{8}}, 4.62 - 4.78 + \frac{\sqrt{0.456}\times 3.70}{\sqrt{8}}\right)$

$= (-1.04, 0.72)$

11.2.6 The numbers in the "Blocks" row change (except for the degrees of freedom) and the total sum of squares changes.

11.2.7 (a) $\bar{x}_{1.} = 6.0617$

$\bar{x}_{2.} = 7.1967$

$\bar{x}_{3.} = 5.7767$

(b) $\bar{x}_{.1} = 7.4667$

$\bar{x}_{.2} = 5.2667$

$\bar{x}_{.3} = 5.1133$

$\bar{x}_{.4} = 7.3300$

$\bar{x}_{.5} = 6.2267$

$\bar{x}_{.6} = 6.6667$

(c) $\bar{x}_{..} = 6.345$

(d) $SSTr = 6.7717$

(e) $SSBl = 15.0769$

(f) $\sum_{i=1}^{3}\sum_{j=1}^{6} x_{ij}^2 = 752.1929$

(g) $SST = 27.5304$

(h) $SSE = 5.6818$

(i)

Source	df	SS	MS	F	p-value
Treatments	2	6.7717	3.3859	5.96	0.020
Blocks	5	15.0769	3.0154	5.31	0.012
Error	10	5.6818	0.5682		
Total	17	27.5304			

(j) $\mu_1 - \mu_2 \in \left(6.06 - 7.20 - \frac{\sqrt{0.5682}\times 3.88}{\sqrt{6}}, 6.06 - 7.20 + \frac{\sqrt{0.5682}\times 3.88}{\sqrt{6}}\right)$

$= (-2.33, 0.05)$

$\mu_1 - \mu_3 \in \left(6.06 - 5.78 - \frac{\sqrt{0.5682}\times 3.88}{\sqrt{6}}, 6.06 - 5.78 + \frac{\sqrt{0.5682}\times 3.88}{\sqrt{6}}\right)$

$= (-0.91, 1.47)$

$\mu_2 - \mu_3 \in \left(7.20 - 5.78 - \frac{\sqrt{0.5682}\times 3.88}{\sqrt{6}}, 7.20 - 5.78 + \frac{\sqrt{0.5682}\times 3.88}{\sqrt{6}}\right)$

$= (0.22, 2.61)$

(l) The total sample size required from each factor level (number of blocks) can be estimated as

$n \geq \frac{4\ s^2\ q^2_{\alpha,k,\nu}}{L^2} = \frac{4\times 0.5682\times 3.88^2}{2.0^2} = 8.6$

so that an additional $9 - 6 = 3$ blocks can be recommended.

11.2.8

Source	df	SS	MS	F	p-value
Treatments	3	67.980	22.660	5.90	0.004
Blocks	7	187.023	26.718	6.96	0.000
Error	21	80.660	3.841		
Total	31	335.662			

$\mu_1 - \mu_2 \in (-2.01, 3.46)$

$\mu_1 - \mu_3 \in (-5.86, -0.39)$

$\mu_1 - \mu_4 \in (-3.95, 1.52)$

$\mu_2 - \mu_3 \in (-6.59, -1.11)$

$\mu_2 - \mu_4 \in (-4.68, 0.79)$

$\mu_3 - \mu_4 \in (-0.83, 4.64)$

The total sample size required from each factor level (number of blocks) can be estimated as

$n \geq \frac{4\ s^2\ q^2_{\alpha,k,\nu}}{L^2} = \frac{4\times 3.841\times 3.95^2}{4.0^2} = 14.98$

so that an additional $15 - 8 = 7$ blocks can be recommended.

Note: In the remainder of this section the confidence intervals for the pairwise differences of the factor level means are provided with an overall confidence level of 95%.

11.2.9

Source	df	SS	MS	F	p-value
Treatments	2	17.607	8.803	2.56	0.119
Blocks	6	96.598	16.100	4.68	0.011
Error	12	41.273	3.439		
Total	20	155.478			

$\mu_1 - \mu_2 \in (-1.11, 4.17)$

$\mu_1 - \mu_3 \in (-0.46, 4.83)$

$\mu_2 - \mu_3 \in (-1.99, 3.30)$

There is *not* sufficient evidence to conclude that the calciners are operating at different efficiencies.

11.2.10

Source	df	SS	MS	F	p-value
Treatments	2	133.02	66.51	19.12	0.000
Blocks	7	1346.76	192.39	55.30	0.000
Error	14	48.70	3.48		
Total	23	1528.49			

$\mu_1 - \mu_2 \in (-8.09, -3.21)$

$\mu_1 - \mu_3 \in (-4.26, 0.62)$

$\mu_2 - \mu_3 \in (1.39, 6.27)$

There is sufficient evidence to conclude that radar system 2 is better than the other two radar systems.

11.2.11

Source	df	SS	MS	F	p-value
Treatments	3	3231.2	1,077.1	4.66	0.011
Blocks	8	29256.1	3,657.0	15.83	0.000
Error	24	5545.1	231.0		
Total	35	38032.3			

$\mu_1 - \mu_2 \in (-8.20, 31.32)$

$\mu_1 - \mu_3 \in (-16.53, 22.99)$

$\mu_1 - \mu_4 \in (-34.42, 5.10)$

$\mu_2 - \mu_3 \in (-28.09, 11.43)$

$\mu_2 - \mu_4 \in (-45.98, -6.46)$

$\mu_3 - \mu_4 \in (-37.65, 1.87)$

There is sufficient evidence to conclude that driver 4 is better than driver 2.

11.2.12

Source	df	SS	MS	F	p-value
Treatments	2	7.47	3.73	0.34	0.718
Blocks	9	313.50	34.83	3.15	0.018
Error	18	199.20	11.07		
Total	29	520.17			

$\mu_1 - \mu_2 \in (-3.00, 4.60)$

$\mu_1 - \mu_3 \in (-2.60, 5.00)$

$\mu_2 - \mu_3 \in (-3.40, 4.20)$

There is *not* sufficient evidence to conclude that there is any difference between the assembly methods.

11.2.13

Source	df	SS	MS	F	p-value
Treatments	4	8.462×10^8	2.116×10^8	66.55	0.000
Blocks	11	19.889×10^8	1.808×10^8	56.88	0.000
Error	44	1.399×10^8	3.179×10^6		
Total	59	29.750×10^8			

$\mu_1 - \mu_2 \in (4372, 8510)$

$\mu_1 - \mu_3 \in (4781, 8919)$

$\mu_1 - \mu_4 \in (5438, 9577)$

$\mu_1 - \mu_5 \in (-3378, 760)$

$\mu_2 - \mu_3 \in (-1660, 2478)$

$\mu_2 - \mu_4 \in (-1002, 3136)$

$\mu_2 - \mu_5 \in (-9819, -5681)$

$\mu_3 - \mu_4 \in (-1411, 2727)$

$\mu_3 - \mu_5 \in (-10228, -6090)$

$\mu_4 - \mu_5 \in (-10886, -6748)$

There is sufficient evidence to conclude that either agent 1 or agent 5 is the best agent.

The worst agent is either agent 2, 3 or 4.

11.2.14

Source	df	SS	MS	F	p-value
Treatments	3	10.637	3.546	2.01	0.123
Blocks	19	169.526	8.922	5.05	0.000
Error	57	100.641	1.766		
Total	79	280.805			

$\mu_1 - \mu_2 \in (-1.01, 1.21)$

$\mu_1 - \mu_3 \in (-1.89, 0.34)$

$\mu_1 - \mu_4 \in (-1.02, 1.20)$

$\mu_2 - \mu_3 \in (-1.98, 0.24)$

$\mu_2 - \mu_4 \in (-1.12, 1.11)$

$\mu_3 - \mu_4 \in (-0.24, 1.98)$

There is *not* sufficient evidence to conclude that there is any difference between the four formulations.

11.2.15 (a)

Source	df	SS	MS	F	p-value
Treatments	3	0.151	0.0503	5.36	0.008
Blocks	6	0.324	0.054	5.75	0.002
Error	18	0.169	0.00939		
Total	27	0.644			

(b) With $q_{0.05,4,18} = 4.00$ and

$\frac{\sqrt{MSE} \times q_{0.05,4,18}}{\sqrt{b}} = \frac{\sqrt{0.00939} \times 4.00}{\sqrt{7}} = 0.146$

the pairwise confidence intervals are:

$\mu_2 - \mu_1 \in 0.630 - 0.810 \pm 0.146 = (-0.326, -0.034)$

$\mu_2 - \mu_3 \in 0.630 - 0.797 \pm 0.146 = (-0.313, -0.021)$

$\mu_2 - \mu_4 \in 0.630 - 0.789 \pm 0.146 = (-0.305, -0.013)$

None of these confidence intervals contains zero so there is sufficient evidence to conclude that treatment 2 has a smaller mean value than each of the other treatments.

11.2.16 $\bar{x}_{..} = \frac{107.68+109.86+111.63}{3} = \frac{329.17}{3} = 109.72$

$SSTR = 4 \times (107.68^2 + 109.86^2 + 111.63^2) - 12 \times \left(\frac{329.17}{3}\right)^2 = 31.317$

$MSE = \hat{\sigma}^2 = 1.445^2 = 2.088$

Source	df	SS	MS	F	p-value
Treatments	2	31.317	15.659	7.50	0.023
Blocks	3	159.720	53.240	25.50	0.001
Error	6	12.528	2.088		
Total	11	203.565			

11.2.17 The new analysis of variance table is

Source	df	SS	MS	F	p-value
Treatments	same	a^2 SSTr	a^2 MSTr	same	same
Blocks	same	a^2 SSBl	a^2 MSBl	same	same
Error	same	a^2 SSE	a^2 MSE		
Total	same	a^2 SST			

11.2.18 $\bar{x}_{..} = \frac{\bar{x}_{1.}+\bar{x}_{2.}+\bar{x}_{3.}+\bar{x}_{4.}}{4} = \frac{3107.3}{4} = 776.825$

$SSTr = 7 \times (763.9^2 + 843.9^2 + 711.3^2 + 788.2^2) - 4 \times 7 \times 776.825^2 = 63623.2$

Source	df	SS	MS	F	p-value
Treatments	3	63623.2	21207.7	54.13	0.000
Blocks	6	13492.3	2248.7	5.74	0.002
Error	18	7052.8	391.8		
Total	27	84168.3			

There is sufficient evidence to conclude that the treatments are not all the same.

Since

$\frac{s \times q_{0.05,4,18}}{\sqrt{b}} = \frac{\sqrt{391.8} \times 4.00}{\sqrt{7}} = 29.9$

it follows that treatments are only known to be different if their sample averages are more than 29.9 apart.

It is known that treatment 2 has the largest mean, and that treatment 3 has the smallest mean.

Treatments 1 and 4 are indistinguishable.

11.2.19 $\bar{x}_{1.} = 23.18$

$\bar{x}_{2.} = 23.58$

$\bar{x}_{3.} = 23.54$

$\bar{x}_{4.} = 22.48$

Source	df	SS	MS	F	p-value
Locations	3	3.893	1.298	0.49	0.695
Time	4	472.647	118.162	44.69	0.000
Error	12	31.729	2.644		
Total	19	508.270			

The p-value of 0.695 implies that there is not sufficient evidence to conclude that the pollution levels are different at the four locations.

The confidence intervals for all of the pairwise comparisons contain zero, so the graphical representation has one line joining all four sample means.

11.4 Supplementary Problems

11.4.1

Source	df	SS	MS	F	p-value
Treatments	3	1.9234	0.6411	22.72	0.000
Error	16	0.4515	0.0282		
Total	19	2.3749			

$\mu_1 - \mu_2 \in (-0.35, 0.26)$

$\mu_1 - \mu_3 \in (0.38, 0.99)$

$\mu_1 - \mu_4 \in (-0.36, 0.25)$

$\mu_2 - \mu_3 \in (0.42, 1.03)$

$\mu_2 - \mu_4 \in (-0.31, 0.30)$

$\mu_3 - \mu_4 \in (-1.04, -0.43)$

There is sufficient evidence to conclude that type 3 has a lower average Young's modulus.

11.4.2

Source	df	SS	MS	F	p-value
Treatments	3	5.77	1.92	0.49	0.690
Error	156	613.56	3.93		
Total	159	619.33			

$\mu_1 - \mu_2 \in (-1.27, 1.03)$

$\mu_1 - \mu_3 \in (-0.82, 1.61)$

$\mu_1 - \mu_4 \in (-1.16, 1.17)$

$\mu_2 - \mu_3 \in (-0.64, 1.67)$

$\mu_2 - \mu_4 \in (-0.97, 1.22)$

$\mu_3 - \mu_4 \in (-1.55, 0.77)$

There is *not* sufficient evidence to conclude that any of the cars is getting better gas mileage than the others.

11.4.3

Source	df	SS	MS	F	p-value
Treatments	4	2,716.8	679.2	3.57	0.024
Blocks	5	4,648.2	929.6	4.89	0.004
Error	20	3,806.0	190.3		
Total	29	11,171.0			

There is not conclusive evidence that the different temperature levels have an effect on the cement strength.

11.4.4

Source	df	SS	MS	F	p-value
Treatments	4	10,381.4	2,595.3	25.70	0.000
Blocks	9	6,732.7	748.1	7.41	0.000
Error	36	3,635.8	101.0		
Total	49	20,749.9			

There is sufficient evidence to conclude that either fertilizer type 4 or type 5 provides the highest yield.

11.4.5

Source	df	SS	MS	F	p-value
Treatments	3	115.17	38.39	4.77	0.007
Blocks	11	4,972.67	452.06	56.12	0.000
Error	33	265.83	8.06		
Total	47	5,353.67			

There is sufficient evidence to conclude that clinic 3 is different from clinics 2 and 4.

11.4.6

Source	df	SS	MS	F	p-value
Treatments	2	142.89	71.44	16.74	0.000
Error	24	102.42	4.27		
Total	26	245.31			

$\mu_h - \mu_a \in (-5.13, -0.27)$

$\mu_h - \mu_b \in (0.50, 5.36)$

$\mu_a - \mu_b \in (3.20, 8.06)$

There is sufficient evidence to conclude that each of the three positions produce different average insertion gains.

11.4.7

Source	df	SS	MS	F	p-value
Treatments	3	1175.3	391.8	8.11	0.000
Error	33	1595.1	48.3		
Total	36	2770.4			

$\mu_1 - \mu_2 \in (3.45, 21.32)$

$\mu_1 - \mu_3 \in (3.29, 20.13)$

$\mu_1 - \mu_4 \in (-6.94, 10.36)$

$\mu_2 - \mu_3 \in (-9.61, 8.26)$

$\mu_2 - \mu_4 \in (-19.82, -1.53)$

$\mu_3 - \mu_4 \in (-18.65, -1.35)$

The drags for designs 1 and 4 are larger than the drags for designs 2 and 3.

11.4.8

Source	df	SS	MS	F	p-value
Treatments	3	0.150814	0.050271	5.39	0.008
Blocks	6	0.325043	0.054174	5.80	0.002
Error	18	0.167986	0.009333		
Total	27	0.643843			

There is sufficient evidence to conclude that the shrinkage from preparation method 2 is smaller than from the other preparation methods.

11.4.9 (a) True

(b) False

(c) True

(d) True

(e) True

(f) True

(g) False

(h) False

11.4.10 (a) $\bar{x}_{1.} = 16.667$

$\bar{x}_{2.} = 19.225$

$\bar{x}_{3.} = 14.329$

Source	df	SS	MS	F	p-value
Alloys	2	89.83	44.91	13.84	0.000
Error	18	58.40	3.24		
Total	20	148.23			

There is sufficient evidence to establish that the alloys are not all the same with respect to their hardness measurements.

(b) With $q_{0.05,3,18} = 3.61$ the pairwise confidence intervals are:

$\mu_1 - \mu_2 \in 16.667 - 19.225 \pm \frac{3.61\times\sqrt{3.24}}{\sqrt{2}}\sqrt{\frac{1}{6}+\frac{1}{8}} = (-5.042, -0.075)$

$\mu_1 - \mu_3 \in 16.667 - 14.329 \pm \frac{3.61\times\sqrt{3.24}}{\sqrt{2}}\sqrt{\frac{1}{6}+\frac{1}{7}} = (-0.220, 4.896)$

$\mu_2 - \mu_3 \in 19.225 - 14.329 \pm \frac{3.61\times\sqrt{3.24}}{\sqrt{2}}\sqrt{\frac{1}{8}+\frac{1}{7}} = (2.517, 7.276)$

These confidence intervals show that alloy 2 has larger hardness measurements than both alloys 1 and 3, which are indistinguishable.

Alloy 2 has the largest mean.

Either alloy 1 or alloy 3 has the smallest mean.

11.4.11 $\bar{x}_{..} = \frac{\bar{x}_{1.}+\bar{x}_{2.}+\bar{x}_{3.}+\bar{x}_{4.}}{4} = \frac{50.1}{4} = 12.525$

$SSTr = 9 \times (11.43^2 + 12.03^2 + 14.88^2 + 11.76^2) - 4 \times 9 \times 12.525^2 = 68.18$

Source	df	SS	MS	F	p-value
Treatments	3	68.18	22.73	38.63	0.000
Blocks	8	53.28	6.66	11.32	0.000
Error	24	14.12	0.588		
Total	35	135.58			

There is sufficient evidence to conclude that the treatments are not all the same.

Since

$\frac{s\times q_{0.05,4,24}}{\sqrt{b}} = \frac{\sqrt{0.588}\times 3.90}{\sqrt{9}} = 0.997$

it follows that two treatments are only known to be different if their sample averages are more than 0.997 apart.

Therefore, treatment 3 is known to have a larger mean than treatments 1, 2, and 4, which are indistinguishable.

11.4.12 $\bar{x}_{1.} = 310.83$

$\bar{x}_{2.} = 310.17$

$\bar{x}_{3.} = 315.33$

$\bar{x}_{4.} = 340.33$

$\bar{x}_{5.} = 300.00$

Source	df	SS	MS	F	p-value
Rivers	4	5442.3	1360.6	20.71	0.000
Error	25	1642.3	65.7		
Total	29	7084.7			

There is sufficient evidence to conclude that the average radon levels in the five rivers are different.

Since

$\frac{s \times q_{0.05,5,25}}{\sqrt{n}} = \frac{\sqrt{65.7} \times 4.165}{\sqrt{6}} = 13.7$

it follows that rivers are only known to be different if their sample averages are more than 13.7 apart.

River 4 can be determined to be the river with the highest radon level.

11.4.13 $\mu_1 - \mu_2 \in (3.23, 11.57)$

$\mu_1 - \mu_3 \in (4.32, 11.68)$

$\mu_1 - \mu_4 \in (-5.85, 1.65)$

$\mu_2 - \mu_3 \in (-3.44, 4.64)$

$\mu_2 - \mu_4 \in (-13.60, -5.40)$

$\mu_3 - \mu_4 \in (-13.70, -6.50)$

Chapter 12

Simple Linear Regression and Correlation

12.1 The Simple Linear Regression Model

12.1.1 (a) $4.2 + (1.7 \times 10) = 21.2$

(b) $3 \times 1.7 = 5.1$

(c) $P(N(4.2 + (1.7 \times 5), 3.2^2) \geq 12) = 0.587$

(d) $P(N(4.2 + (1.7 \times 8), 3.2^2) \leq 17) = 0.401$

(e) $P(N(4.2 + (1.7 \times 6), 3.2^2) \geq N(4.2 + (1.7 \times 7), 3.2^2)) = 0.354$

12.1.2 (a) $123.0 + (-2.16 \times 20) = 79.8$

(b) $-2.16 \times 10 = -21.6$

(c) $P(N(123.0 + (-2.16 \times 25), 4.1^2) \leq 60) = 0.014$

(d) $P(30 \leq N(123.0 + (-2.16 \times 40), 4.1^2) \leq 40) = 0.743$

(e) $P(N(123.0 + (-2.16 \times 30), 4.1^2) \leq N(123.0 + (-2.16 \times 27.5), 4.1^2)) = 0.824$

12.1.3 (a) $y = 5 + (0.9 \times 20) = 23.0$

(b) The expected value of the porosity decreases by $5 \times 0.9 = 4.5$.

(c) $P(N(5 + (0.9 \times 25), 1.4^2) \leq 30) = 0.963$

(d) $P\left(17 \leq N\left(5+(0.9\times 15), \frac{1.4^2}{4}\right) \leq 20\right) = 0.968$

12.1.4 Since the model minimizes the sum of squares of residuals (vertical differences) the model will change if the x and y variables are exchanged.

If there is a reason to consider that one variable can be naturally thought of as being "determined" by the choice of the other variable, then that indicates the appropriate choice of the x and y variables (the y variable should be the "determined" variable).

In addition, if the model is to be used to predict one variable for given values of the other variable, then that also indicates the appropriate choice of the x and y variables (the y variable should be the variable that is being predicted).

12.1.5 $P(N(675.30-(5.87\times 80), 7.32^2) \leq 220)$

$= P\left(N(0,1) \leq \frac{220-205.7}{7.32}\right)$

$= \Phi(1.954) = 0.975$

12.2 Fitting the Regression Line

12.2.2 $n = 20$

$\sum_{i=1}^{20} x_i = 8.552$

$\sum_{i=1}^{20} y_i = 398.2$

$\sum_{i=1}^{20} x_i^2 = 5.196$

$\sum_{i=1}^{20} y_i^2 = 9356$

$\sum_{i=1}^{20} x_i y_i = 216.6$

$\bar{x} = \frac{8.552}{20} = 0.4276$

$\bar{y} = \frac{398.2}{20} = 19.91$

$S_{XX} = 5.196 - (20 \times 0.4276^2) = 1.539$

$S_{XY} = 216.6 - (20 \times 0.4276 \times 19.91) = 46.330$

Using these values

$\hat{\beta}_1 = \frac{46.330}{1.539} = 30.101$

$\hat{\beta}_0 = 19.91 - (30.10 \times 0.4276) = 7.039$

and

$SSE = 9356 - (7.039 \times 398.2) - (30.101 \times 216.6) = 33.291$

so that

$\hat{\sigma}^2 = \frac{33.291}{20-2} = 1.85.$

When $x = 0.5$ the fitted value is
$7.039 + (30.101 \times 0.5) = 22.09.$

12.2.3 $\hat{\beta}_0 = 39.5$

$\hat{\beta}_1 = -2.04$

$\hat{\sigma}^2 = 17.3$

$39.5 + (-2.04 \times (-2.0)) = 43.6$

12.2.4 (a) $\hat{\beta}_0 = -2,277$

$\hat{\beta}_1 = 1.003$

(b) $1.003 \times 1000 = \$1003$

(c) $-2277 + (1.003 \times 10000) = 7753$

The predicted cost is \$7,753,000.

(d) $\hat{\sigma}^2 = 774211$

(e) If the model is used then it would be extrapolation,
so the prediction may be inaccurate.

12.2.5 (a) $\hat{\beta}_0 = 36.19$

$\hat{\beta}_1 = 0.2659$

(b) $\hat{\sigma}^2 = 70.33$

(c) Yes, since $\hat{\beta}_1 > 0$.

(d) $36.19 + (0.2659 \times 72) = 55.33$

12.2.6 (a) $\hat{\beta}_0 = 54.218$

$\hat{\beta}_1 = -0.3377$

(b) No, $\hat{\beta}_1 < 0$ suggests that aerobic fitness deteriorates with age.

The predicted change in VO2-max for an additional 5 years of age is
$-0.3377 \times 5 = -1.6885$.

(c) $54.218 + (-0.3377 \times 50) = 37.33$

(d) If the model is used then it would be extrapolation,
so the prediction may be inaccurate.

(e) $\hat{\sigma}^2 = 57.30$

12.2.7 (a) $\hat{\beta}_0 = -29.59$

$\hat{\beta}_1 = 0.07794$

(b) $-29.59 + (0.07794 \times 2,600) = 173.1$

(c) $0.07794 \times 100 = 7.794$

(d) $\hat{\sigma}^2 = 286$

12.2.8 (a) $\hat{\beta}_0 = -1.911$

$\hat{\beta}_1 = 1.6191$

(b) $1.6191 \times 1 = 1.6191$

The expert is underestimating the times.

$-1.911 + (1.6191 \times 7) = 9.42$

(c) If the model is used then it would be extrapolation, so the prediction may be inaccurate.

(d) $\hat{\sigma}^2 = 12.56$

12.2.9 (a) $\hat{\beta}_0 = 12.864$

$\hat{\beta}_1 = 0.8051$

(b) $12.864 + (0.8051 \times 69) = 68.42$

(c) $0.8051 \times 5 = 4.03$

(d) $\hat{\sigma}^2 = 3.98$

12.3 Inferences on the Slope Parameter $\hat{\beta}_1$

12.3.1 (a) $(0.522 - (2.921 \times 0.142), 0.522 + (2.921 \times 0.142)) = (0.107, 0.937)$

(b) The t-statistic is

$\frac{0.522}{0.142} = 3.68$

and the p-value is 0.002.

12.3.2 (a) $(56.33 - (2.086 \times 3.78), 56.33 + (2.086 \times 3.78)) = (48.44, 64.22)$

(b) The t-statistic is

$\frac{56.33-50.0}{3.78} = 1.67$

and the p-value is 0.110.

12.3.3 (a) $s.e.(\hat{\beta}_1) = 0.08532$

(b) $(1.003 - (2.145 \times 0.08532), 1.003 + (2.145 \times 0.08532)) = (0.820, 1.186)$

(c) The t-statistic is

$\frac{1.003}{0.08532} = 11.76$

and the p-value is 0.000.

12.3.4 (a) $s.e.(\hat{\beta}_1) = 0.2383$

(b) $(0.2659 - (1.812 \times 0.2383), 0.2659 + (1.812 \times 0.2383)) = (-0.166, 0.698)$

(c) The t-statistic is

$\frac{0.2659}{0.2383} = 1.12$

and the p-value is 0.291.

(d) There is *not* sufficient evidence to conclude that on average trucks take longer to unload when the temperature is higher.

12.3.5 (a) $s.e.(\hat{\beta}_1) = 0.1282$

(b) $(-\infty, -0.3377 + (1.734 \times 0.1282)) = (-\infty, -0.115)$

(c) The t-statistic is

$\frac{-0.3377}{0.1282} = -2.63$

and the (two-sided) p-value is 0.017.

12.3.6 (a) $s.e.(\hat{\beta}_1) = 0.00437$

(b) $(0.0779 - (3.012 \times 0.00437), 0.0779 + (3.012 \times 0.00437)) = (0.0647, 0.0911)$

(c) The t-statistic is

$\frac{0.0779}{0.00437} = 17.83$

and the p-value is 0.000.

There is sufficient evidence to conclude that the house price depends upon the size of the house.

12.3.7 (a) $s.e.(\hat{\beta}_1) = 0.2829$

(b) $(1.619 - (2.042 \times 0.2829), 1.619 + (2.042 \times 0.2829)) = (1.041, 2.197)$

(c) If $\beta_1 = 1$ then the actual times are equal to the estimated times except for a constant difference of β_0.

The t-statistic is

$\frac{1.619-1.000}{0.2829} = 2.19$

and the p-value is 0.036.

12.3.8 (a) $s.e.(\hat{\beta}_1) = 0.06427$

(b) $(0.8051 - (2.819 \times 0.06427), 0.8051 + (2.819 \times 0.06427)) = (0.624, 0.986)$

(c) The t-statistic is

$\frac{0.8051}{0.06427} = 12.53$

and the p-value is 0.000.

There is sufficient evidence to conclude that resistance
increases with temperature.

12.3.9 For the hypotheses

$H_0 : \beta_1 = 0$ versus $H_A : \beta_1 \neq 0$

the t-statistic is

$t = \frac{54.87}{21.20} = 2.588$

so that the p-value is $2 \times P(t_{18} \geq 2.588) = 0.019$.

12.3.10 The model is $y = \beta_0 + \beta_1 x$,

where y is the density of the ceramic and x is the baking time.

$n = 10$

$\sum_{i=1}^{10} x_i = 2600$

$\sum_{i=1}^{10} y_i = 31.98$

$\sum_{i=1}^{10} x_i^2 = 679850$

$\sum_{i=1}^{10} y_i^2 = 102.3284$

$\sum_{i=1}^{10} x_i y_i = 8321.15$

$\bar{x} = \frac{2600}{10} = 260$

$\bar{y} = \frac{31.98}{10} = 3.198$

$S_{XX} = 679850 - (10 \times 260^2) = 3850$

$S_{YY} = 102.3284 - (10 \times 3.198^2) = 0.05636$

$S_{XY} = 8321.15 - (10 \times 260 \times 3.198) = 6.35$

Using these values

$\hat{\beta}_1 = \frac{6.35}{3850} = 0.00165$

$\hat{\beta}_0 = 3.198 - (0.00165 \times 260) = 2.769$

and

$SSE = 102.3284 - (2.769 \times 31.98) - (0.00165 \times 8321.15) = 0.04589$

so that

$\hat{\sigma}^2 = \frac{0.04589}{10-2} = 0.00574.$

For the hypotheses

$H_0 : \beta_1 = 0$ versus $H_A : \beta_1 \neq 0$

the t-statistic is

$t = \frac{0.00165}{\sqrt{0.00574/3850}} = 1.35$

so that the p-value is $2 \times P(t_9 \geq 1.35) = 0.214$.

Therefore, the regression is not significant and there is not sufficient evidence to establish that the baking time has an effect on the density of the ceramic.

12.4 Inferences on the Regression Line

12.4.2 $(1392, 1400)$

12.4.3 $(21.9, 23.2)$

12.4.4 $(6754, 7755)$

12.4.5 $(33.65, 41.02)$

12.4.6 $(201.4, 238.2)$

12.4.7 $(-\infty, 10.63)$

12.4.8 $(68.07, 70.37)$

12.4.9 $\sum_{i=1}^{8} x_i = 122.6$

$\sum_{i=1}^{8} x_i^2 = 1939.24$

$\bar{x} = \frac{122.6}{8} = 15.325$

$S_{XX} = 1939.24 - \frac{122.6^2}{8} = 60.395$

With $t_{0.025,6} = 2.447$ the confidence interval is

$\beta_0 + (\beta_1 \times 15) \in 75.32 + (0.0674 \times 15) \pm 2.447 \times 0.0543 \times \sqrt{\frac{1}{8} + \frac{(15-15.325)^2}{60.395}}$

which is $(76.284, 76.378)$.

12.4.10 $n = 7$

$\sum_{i=1}^{7} x_i = 240.8$

$\sum_{i=1}^{7} y_i = 501.8$

$\sum_{i=1}^{7} x_i^2 = 8310.44$

$\sum_{i=1}^{7} y_i^2 = 36097.88$

$\sum_{i=1}^{7} x_i y_i = 17204.87$

$\bar{x} = \frac{240.8}{7} = 34.400$

$\bar{y} = \frac{501.8}{7} = 71.686$

$S_{XX} = 8310.44 - \frac{240.8^2}{7} = 26.920$

$S_{YY} = 36097.88 - \frac{501.8^2}{7} = 125.989$

$S_{XY} = 17204.87 - \frac{240.8 \times 501.8}{7} = -57.050$

Using these values

$\hat{\beta}_1 = \frac{-57.050}{26.920} = -2.119$

$\hat{\beta}_0 = 71.686 - (-2.119 \times 34.400) = 144.588$

and

$SSE = 36097.88 - (144.588 \times 501.8) - (-2.119 \times 17204.87) = 5.086$

so that

$\hat{\sigma}^2 = \frac{5.086}{7-2} = 1.017.$

With $t_{0.005,5} = 4.032$ the confidence interval is

$\beta_0 + (\beta_1 \times 35) \in 144.588 + (-2.119 \times 35) \pm 4.032 \times \sqrt{1.017} \times \sqrt{\frac{1}{7} + \frac{(35-34.400)^2}{26.920}}$

which is

$70.414 \pm 1.608 = (68.807, 72.022).$

12.5 Prediction Intervals for Future Response Values

12.5.1 $(1386, 1406)$

12.5.2 $(19.7, 25.4)$

12.5.3 $(5302, 9207)$

12.5.4 $(21.01, 53.66)$

12.5.5 $(165.7, 274.0)$

12.5.6 $(-\infty, 15.59)$

12.5.7 $(63.48, 74.96)$

12.5.8 $\bar{x} = \frac{603.36}{30} = 20.112$

$S_{XX} = 12578.22 - \frac{603.36^2}{30} = 443.44$

$\hat{\sigma}^2 = \frac{329.77}{30-2} = 11.778$

With $t_{0.025,28} = 2.048$ the prediction interval is

$51.98 + (3.44 \times 22) \pm 2.048 \times \sqrt{11.778} \times \sqrt{\frac{31}{30} + \frac{(22-20.112)^2}{443.44}}$

$= 127.66 \pm 7.17 = (120.49, 134.83)$.

12.5.9 $n = 7$

$\sum_{i=1}^{7} x_i = 142.8$

$\sum_{i=1}^{7} y_i = 361.5$

$\sum_{i=1}^{7} x_i^2 = 2942.32$

$\sum_{i=1}^{7} y_i^2 = 18771.5$

$\sum_{i=1}^{7} x_i y_i = 7428.66$

$\bar{x} = \frac{142.8}{7} = 20.400$

$\bar{y} = \frac{361.5}{7} = 51.643$

$S_{XX} = 2942.32 - \frac{142.8^2}{7} = 29.200$

$S_{YY} = 18771.5 - \frac{361.5^2}{7} = 102.607$

$S_{XY} = 7428.66 - \frac{142.8 \times 361.5}{7} = 54.060$

Using these values

$\hat{\beta}_1 = \frac{54.060}{29.200} = 1.851$

$\hat{\beta}_0 = 51.643 - (1.851 \times 20.400) = 13.875$

and

$SSE = 18771.5 - (13.875 \times 361.5) - (1.851 \times 7428.66) = 2.472$

so that

$\hat{\sigma}^2 = \frac{2.472}{7-2} = 0.494.$

With $t_{0.005,5} = 4.032$ the prediction interval is

$13.875 + (1.851 \times 20) \pm 4.032 \times \sqrt{0.494} \times \sqrt{\frac{8}{7} + \frac{(20-20.400)^2}{29.200}}$

which is

$50.902 \pm 3.039 = (47.864, 53.941).$

12.6 The Analysis of Variance Table

12.6.1

Source	df	SS	MS	F	p-value
Regression	1	40.53	40.53	2.32	0.137
Error	33	576.51	17.47		
Total	34	617.04			

$R^2 = \frac{40.53}{617.04} = 0.066$

12.6.2

Source	df	SS	MS	F	p-value
Regression	1	120.61	120.61	6.47	0.020
Error	19	354.19	18.64		
Total	20	474.80			

$R^2 = \frac{120.61}{474.80} = 0.254$

12.6.3

Source	df	SS	MS	F	p-value
Regression	1	870.43	870.43	889.92	0.000
Error	8	7.82	0.9781		
Total	9	878.26			

$R^2 = \frac{870.43}{878.26} = 0.991$

12.6.4

Source	df	SS	MS	F	p-value
Regression	1	6.82×10^6	6.82×10^6	1.64	0.213
Error	23	95.77×10^6	4.16×10^6		
Total	24	102.59×10^6			

$R^2 = \frac{6.82\times10^6}{102.59\times10^6} = 0.06$

12.6.5

Source	df	SS	MS	F	p-value
Regression	1	10.71×10^7	10.71×10^7	138.29	0.000
Error	14	1.08×10^7	774,211		
Total	15	11.79×10^7			

$R^2 = \frac{10.71\times10^7}{11.79\times10^7} = 0.908$

12.6.6

Source	df	SS	MS	F	p-value
Regression	1	87.59	87.59	1.25	0.291
Error	10	703.33	70.33		
Total	11	790.92			

$R^2 = \frac{87.59}{790.92} = 0.111$

The large p-value implies that there is not sufficient evidence to conclude that on average trucks take longer to unload when the temperature is higher.

12.6.7

Source	df	SS	MS	F	p-value
Regression	1	397.58	397.58	6.94	0.017
Error	18	1031.37	57.30		
Total	19	1428.95			

$R^2 = \frac{397.58}{1428.95} = 0.278$

The R^2 value implies that about 28% of the variability in VO2-max can be accounted for by changes in age.

12.6.8

Source	df	SS	MS	F	p-value
Regression	1	90907	90907	318.05	0.000
Error	13	3716	286		
Total	14	94622			

$R^2 = \frac{90907}{94622} = 0.961.$

The high R^2 value indicates that there is almost a perfect linear relationship between appraisal value and house size.

12.6.9

Source	df	SS	MS	F	p-value
Regression	1	411.26	411.26	32.75	0.000
Error	30	376.74	12.56		
Total	31	788.00			

$R^2 = \frac{411.26}{788.00} = 0.522$

The p-value is not very meaningful because it tests the null hypothesis that the actual times are unrelated to the estimated times.

12.6.10

Source	df	SS	MS	F	p-value
Regression	1	624.70	624.70	156.91	0.000
Error	22	87.59	3.98		
Total	23	712.29			

$R^2 = \frac{624.70}{712.29} = 0.877$

12.7 Residual Analysis

12.7.1 There is no suggestion that the fitted regression model is not appropriate.

12.7.2 There is no suggestion that the fitted regression model is not appropriate.

12.7.3 There is a possible suggestion of a slight reduction in the variability of the VO2-max values as age increases.

12.7.4 The observation with an area of 1,390 square feet appears to be an outlier.

There is no suggestion that the fitted regression model is not appropriate.

12.7.5 The variability of the actual times increases as the estimated time increases.

12.7.6 There is a possible suggestion of a slight increase in the variability of the resistances at higher temperatures.

12.8 Variable Transformations

12.8.1 The model

$y = \gamma_0 \, e^{\gamma_1 x}$

is appropriate.

A linear regression can be performed with $\ln(y)$ as the dependent variable and with x as the input variable.

$\hat{\gamma}_0 = 9.12$

$\hat{\gamma}_1 = 0.28$

$\hat{\gamma}_0 \, e^{\hat{\gamma}_1 \times 2.0} = 16.0$

12.8.2 The model

$y = \frac{x}{\gamma_0 + \gamma_1 x}$

is appropriate.

A linear regression can be performed with $\frac{1}{y}$ as the dependent variable and with $\frac{1}{x}$ as the input variable.

$\hat{\gamma}_0 = 1.067$

$\hat{\gamma}_1 = 0.974$

$\frac{2.0}{\hat{\gamma}_0 + (\hat{\gamma}_1 \times 2.0)} = 0.66$

12.8.3 $\hat{\gamma}_0 = 8.81$

$\hat{\gamma}_1 = 0.523$

$\gamma_0 \in (6.84, 11.35)$

$\gamma_1 \in (0.473, 0.573)$

12.8.4 (b) $\hat{\gamma}_0 = 89.7$

$\hat{\gamma}_1 = 4.99$

(c) $\gamma_0 \in (68.4, 117.7)$

$\gamma_1 \in (4.33, 5.65)$

12.8.5 $\hat{\gamma}_0 = e^{\hat{\beta}_0} = e^{2.628} = 13.85$

$\hat{\gamma}_1 = \hat{\beta}_1 = 0.341$

With $t_{0.025,23} = 2.069$ the confidence interval for γ_1 (and β_1) is
$0.341 \pm (2.069 \times 0.025) = (0.289, 0.393)$.

12.8.6 The model can be rewritten

$y = \gamma_0 \ln(\gamma_1) - 2\gamma_0 \ln(x)$.

If a simple linear regression is performed with $\ln(x)$ as the input variable and y as the output variable, then

$\hat{\beta}_0 = \hat{\gamma}_0 \ln(\hat{\gamma}_1)$

and

$\hat{\beta}_1 = -2\hat{\gamma}_0$.

Therefore,

$\hat{\gamma}_0 = \frac{-\hat{\beta}_1}{2}$

and

$\hat{\gamma}_1 = e^{-2\hat{\beta}_0/\hat{\beta}_1}$.

12.8.7 $\hat{\gamma}_0 = 12.775$

$\hat{\gamma}_1 = -0.5279$

When the crack length is 2.1 the expected breaking load is

$12.775 \times e^{-0.5279 \times 2.1} = 4.22$.

12.9 Correlation Analysis

12.9.3 The sample correlation coefficient is $r = 0.95$.

12.9.4 The sample correlation coefficient is $r = 0.33$.

12.9.5 The sample correlation coefficient is $r = -0.53$.

12.9.6 The sample correlation coefficient is $r = 0.98$.

12.9.7 The sample correlation coefficient is $r = 0.72$.

12.9.8 The sample correlation coefficient is $r = 0.94$.

12.9.9 The sample correlation coefficient is $r = 0.431$.

12.9.10 It is known that $\hat{\beta}_1 > 0$ but nothing is known about the p-value.

12.9.11 The variables A and B may both be related to a third surrogate variable C. It is possible that the variables A and C have a causal relationship, and that the variables B and C have a causal relationship, without there being a causal relationship between the variables A and B.

12.11 Supplementary Problems

12.11.1 (a) $\hat{\beta}_0 = 95.77$

$\hat{\beta}_1 = -0.1003$

$\hat{\sigma}^2 = 67.41$

(b) The sample correlation coefficient is $r = -0.69$.

(c) $(-0.1003 - (2.179 \times 0.0300), -0.1003 + (2.179 \times 0.0300))$

$= (-0.1657, -0.0349)$

(d) The t-statistic is

$\frac{-0.1003}{0.0300} = -3.34$

and the p-value is 0.006.

There is sufficient evidence to conclude that the time taken to finish the test depends upon the SAT score.

(e) $-0.1003 \times 10 = -1.003$

(f) $95.77 + (-0.1003 \times 550) = 40.6$

The confidence interval is $(35.81, 45.43)$.

The prediction interval is $(22.09, 59.15)$.

(g) There is no suggestion that the fitted regression model is not appropriate.

12.11.2 (a) $\hat{\beta}_0 = 18.35$

$\hat{\beta}_1 = 6.72$

$\hat{\sigma}^2 = 93.95$

(b) The sample correlation coefficient is $r = 0.84$.

(c) The t-statistic is 23.91 and the p-value is 0.000.

There is sufficient evidence to conclude that the amount of scrap material depends upon the number of passes.

(d) $(6.72 - (1.960 \times 0.2811), 6.72 + (1.960 \times 0.2811)) = (6.17, 7.27)$

(e) It increases by $6.72 \times 1 = 6.72$.

(f) $18.35 + (6.72 \times 7) = 65.4$

The prediction interval is $(46.1, 84.7)$.

(g) Observations $x = 2$, $y = 67.71$ and $x = 9$, $y = 48.17$ have standardized residuals with absolute values larger than three.

The linear model is reasonable, but a non-linear model with a decreasing slope may be more appropriate.

12.11.3 $\hat{\beta}_0 = 29.97$

$\hat{\beta}_1 = 0.0923$

$\hat{\sigma} = 0.09124$

The t-statistic for the null hypothesis $H_0 : \beta_1 = 0$ is

$\frac{0.09234}{0.01026} = 9.00$

and the p-value is 0.000.

There is a significant association between power loss and bearing diameter.

The sample correlation coefficient is $r = 0.878$.

The fitted value for the power loss of a new engine with a bearing diameter of 25.0 is 32.28 and a 95% prediction interval is $(32.09, 32.47)$.

There are no data points with values $\frac{e_i}{\hat{\sigma}}$ larger than three in absolute value.

12.11.4 $\hat{\beta}_0 = 182.61$

$\hat{\beta}_1 = 0.8623$

$\hat{\sigma} = 32.08$

The sample correlation coefficient is $r = 0.976$.

When the energy lost by the hot liquid is 500, the fitted value for the energy gained by the cold liquid is 613.8 and a 95% prediction interval is $(547.1, 680.3)$.

12.11.5 $\hat{\beta}_0 = 3.252$

$\hat{\beta}_1 = 0.01249$

$\hat{\sigma} = 2.997$

The t-statistic for the null hypothesis $H_0 : \beta_1 = 0$ is

$\frac{0.01249}{0.003088} = 4.04$

and the p-value is 0.001.

There is a significant association between the pulse time and the capacitance value.

The sample correlation coefficient is $r = 0.690$.

For a capacitance of 1700 microfarads, the fitted value for the pulse time is 24.48 milliseconds and a 95% prediction interval is $(17.98, 30.99)$.

The data point with a pulse time of 28.52 milliseconds for a capacitance of 1400 microfarads has a residual $e_i = 7.784$ so that

$\frac{e_i}{\sigma} = \frac{7.784}{2.997} = 2.60.$

12.11.6 (b) $\hat{\gamma}_0 = 0.199$

$\hat{\gamma}_1 = 0.537$

(c) $\gamma_0 \in (0.179, 0.221)$

$\gamma_1 \in (0.490, 0.584)$

(d) $0.199 + \frac{0.537}{10.0} = 0.253$

12.11.7 (a) The model is $y = \beta_0 + \beta_1 x$ where y is the strength of the chemical solution and x is the amount of the catalyst.

$n = 8$

$\sum_{i=1}^{8} x_i = 197$

$\sum_{i=1}^{8} y_i = 225$

$\sum_{i=1}^{8} x_i^2 = 4951$

$\sum_{i=1}^{8} y_i^2 = 7443$

$\sum_{i=1}^{8} x_i y_i = 5210$

$\bar{x} = \frac{197}{8} = 24.625$

$\bar{y} = \frac{225}{8} = 28.125$

$S_{XX} = 4951 - (8 \times 24.625^2) = 99.874$

$S_{YY} = 7443 - (8 \times 28.125^2) = 1114.875$

$S_{XY} = 5210 - (8 \times 24.625 \times 28.125) = -330.625$

Using these values

$\hat{\beta}_1 = \frac{-330.625}{99.874} = -3.310$

$\hat{\beta}_0 = 28.125 - (-3.310 \times 24.625) = 109.643$

and

$SSE = 7443 - (109.643 \times 225) - (-3.310 \times 5210) = 20.378$

so that

$\hat{\sigma}^2 = \frac{20.378}{8-2} = 3.396.$

(b) The t-statistic is

$\frac{-3.310}{\sqrt{3.396/99.874}} = -17.95$

so that the p-value is $2 \times P(t_6 > 17.95) = 0.000$.

Therefore, the null hypothesis $H_0 : \beta_1 = 0$ can be rejected and the regression is significant.

There is sufficient evidence to establish that the amount of the catalyst does effect the strength of the chemical solution.

(c) With $t_{0.025,6} = 2.447$ the prediction interval is

$109.643 + (-3.310 \times 21.0) \pm 2.447 \times \sqrt{3.396} \times \sqrt{1 + \frac{1}{8} + \frac{(21.0-24.625)^2}{99.874}}$

$= 40.125 \pm 5.055$

$= (35.070, 45.180)$.

(d) $e_2 = 17 - (109.643 + (-3.310 \times 28)) = 0.05$

12.11.8 (a) $\bar{x} = \frac{856}{20} = 42.8$

$S_{XX} = 37636 - (20 \times 42.8^2) = 999.2$

With $t_{0.025,18} = 2.101$ the prediction interval is

$123.57 - (3.90 \times 40) \pm 2.101 \times 11.52 \times \sqrt{\frac{21}{20} + \frac{(40-42.8)^2}{999.2}}$

$= (-57.32, -7.54)$

(b) $SST = 55230 - \frac{(-869)^2}{20} = 17472$

Source	df	SS	MS	F	p-value
Regression	1	15083	15083	114	0.000
Error	18	2389	133		
Total	19	17472			

$R^2 = \frac{15083}{17472} = 86\%$

12.11.9 (a) The model is $y = \beta_0 + \beta_1 x$ where y is the bacteria yield and x is the temperature.

$n = 8$

$\sum_{i=1}^{8} x_i = 197$

$\sum_{i=1}^{8} y_i = 429$

$\sum_{i=1}^{8} x_i^2 = 4943$

$\sum_{i=1}^{8} y_i^2 = 23805$

$\sum_{i=1}^{8} x_i y_i = 10660$

$\bar{x} = \frac{197}{8} = 24.625$

$\bar{y} = \frac{429}{8} = 53.625$

$S_{XX} = 4943 - (8 \times 24.625^2) = 91.875$

$S_{YY} = 23805 - (8 \times 53.625^2) = 799.875$

$S_{XY} = 10660 - (8 \times 24.625 \times 53.625) = 95.875$

Using these values

$\hat{\beta}_1 = \frac{95.875}{91.875} = 1.044$

$\hat{\beta}_0 = 53.625 - (1.044 \times 24.625) = 27.93$

and

$SSE = 23805 - (27.93 \times 429) - (1.044 \times 10660) = 699.8$

so that

$\hat{\sigma}^2 = \frac{699.8}{8-2} = 116.6.$

(b) The t-statistic is

$\frac{1.044}{\sqrt{116.6/91.875}} = 0.93$

so that the p-value is $2 \times P(t_6 > 0.93) = 0.390$.

Therefore, the null hypothesis $H_0 : \beta_1 = 0$ cannot be rejected and the regression is not significant.
There is not sufficient evidence to establish that the bacteria yield does depend on temperature.

(c) $e_1 = 54 - (27.93 + 1.044 \times 22) = 3.1$

12.11.10 The F-statistic from the analysis of variance table is

$F = \frac{MSR}{MSE} = \frac{(n-2)SSR}{SSE} = \frac{(n-2)R^2}{1-R^2} = \frac{18 \times 0.853}{1-0.853} = 104.4$

The p-value is $P(F_{1,18} \geq 104.4) = 0.000$.

12.11.11 (a) The model is $y = \beta_0 + \beta_1 x$ where y is the amount of gold obtained and x is the amount of ore processed.

$n = 7$

$\sum_{i=1}^{7} x_i = 85.8$

$\sum_{i=1}^{7} y_i = 87.9$

$\sum_{i=1}^{7} x_i^2 = 1144.40$

$\sum_{i=1}^{7} y_i^2 = 1158.91$

$\sum_{i=1}^{7} x_i y_i = 1146.97$

$\bar{x} = \frac{85.8}{7} = 12.257$

$\bar{y} = \frac{87.9}{7} = 12.557$

$S_{XX} = 1144.40 - (7 \times 12.257^2) = 92.737$

$S_{YY} = 1158.91 - (7 \times 12.557^2) = 55.137$

$S_{XY} = 1146.97 - (7 \times 12.257 \times 12.557) = 69.567$

Using these values

$\hat{\beta}_1 = \frac{69.567}{92.737} = 0.750$

and

$\hat{\beta}_0 = 12.557 - (0.750 \times 12.257) = 3.362.$

(b) Since

$SSE = 1158.91 - (3.362 \times 87.9) - (0.750 \times 69.567) = 2.9511$

it follows that

$\hat{\sigma}^2 = \frac{2.9511}{7-2} = 0.5902.$

The t-statistic is

$\frac{0.750}{\sqrt{0.5902/92.737}} = 9.40$

so that the p-value is $2 \times P(t_5 > 9.40) \simeq 0.$

Therefore, the null hypothesis $H_0 : \beta_1 = 0$ can be rejected and it can be concluded that the regression is significant.

(c) $r = \frac{69.567}{\sqrt{92.737}\sqrt{55.137}} = 0.973$

(d) $R^2 = r^2 = 0.973^2 = 0.946$

(e) With $t_{0.025,5} = 2.571$ the prediction interval is

$3.362 + (0.750 \times 15) \pm 2.571 \times \sqrt{0.5902} \times \sqrt{1 + \frac{1}{7} + \frac{(15-12.257)^2}{92.737}}$

which is

$14.615 \pm 2.185 = (12.430, 16.800)$.

(f) $e_1 = 8.9 - (3.362 + 0.750 \times 7.3) = 0.06$
$e_2 = 11.3 - (3.362 + 0.750 \times 9.1) = 1.11$
$e_3 = 10.6 - (3.362 + 0.750 \times 10.2) = -0.41$
$e_4 = 11.6 - (3.362 + 0.750 \times 11.5) = -0.38$
$e_5 = 12.2 - (3.362 + 0.750 \times 13.2) = -1.06$
$e_6 = 15.7 - (3.362 + 0.750 \times 16.1) = 0.26$
$e_7 = 17.6 - (3.362 + 0.750 \times 18.4) = 0.43$

12.11.12 (a) False

(b) True

(c) True

(d) False

(e) True

(f) False

(g) True

(h) False

(i) True

(j) True

(k) False

(l) True

(m) False

12.11.13 (a) The model is $y = \beta_0 + \beta_1 x$ where y is the downloading time and x is the file size.

$n = 9$

$\sum_{i=1}^{9} x_i = 50.06$

$\sum_{i=1}^{9} y_i = 1156$

$\sum_{i=1}^{9} x_i^2 = 319.3822$

$\sum_{i=1}^{9} y_i^2 = 154520$

$\sum_{i=1}^{9} x_i y_i = 6894.34$

$\bar{x} = \frac{50.06}{9} = 5.562$

$\bar{y} = \frac{1156}{9} = 128.444$

$S_{XX} = 319.3822 - (9 \times 5.562^2) = 40.9374$

$S_{YY} = 154520 - (9 \times 128.444^2) = 6038.2223$

$S_{XY} = 6894.34 - (9 \times 5.562 \times 128.444) = 464.4111$

Using these values

$\hat{\beta}_1 = \frac{464.4111}{40.9374} = 11.344$

$\hat{\beta}_0 = 128.444 - (11.344 \times 5.562) = 65.344$

and

$SSE = 154520 - (65.344 \times 1156) - (11.344 \times 6894.34) = 769.737$

so that

$\hat{\sigma}^2 = \frac{769.737}{9-2} = 109.962.$

(b) The t-statistic is

$\frac{11.344}{\sqrt{109.962/40.9374}} = 6.92$

so that the p-value is $2 \times P(t_7 > 6.92) \simeq 0$.

Therefore, the null hypothesis $H_0 : \beta_1 = 0$ can be rejected and it can be concluded that the regression is significant.

(c) $65.344 + (11.344 \times 6) = 133.41$

(d) Since

$SSR = SST - SSE = 6038.2223 - 769.737 = 5268.485$

it follows that

$R^2 = \frac{5268.485}{6038.2223} = 87.25\%.$

(e) With $t_{0.025,7} = 2.365$ the prediction interval is

$65.344 + (11.344 \times 6) \pm 2.365 \times \sqrt{109.962} \times \sqrt{1 + \frac{1}{9} + \frac{(6-5.562)^2}{40.9374}}$

which is

$133.41 \pm 26.19 = (107.22, 159.60).$

(f) $103 - (65.344 + (4.56 \times 6)) = -14.07$

(g) $r = \sqrt{R^2} = \sqrt{0.8725} = 0.934$

(h) It may be quite unreliable to extrapolate the model to predict the downloading time of a file of size 0.40.

12.11.14 (a) The model is $y = \beta_0 + \beta_1 x$ where y is the speed and x is the depth.

$n = 18$

$\sum_{i=1}^{18} x_i = 56.988$

$\sum_{i=1}^{18} y_i = 27343.03$

$\sum_{i=1}^{18} x_i^2 = 234.255$

$\sum_{i=1}^{18} y_i^2 = 41535625$

$\sum_{i=1}^{18} x_i y_i = 86560.46$

$\bar{x} = \frac{56.988}{18} = 3.166$

$\bar{y} = \frac{27343.03}{18} = 1519.06$

$S_{XX} = 234.255 - (18 \times 3.166^2) = 53.8307$

$S_{YY} = 41535625 - (18 \times 1519.06^2) = 5.2843$

$S_{XY} = 86560.46 - (18 \times 3.166 \times 1519.06) = -16.666$

Using these values

$\hat{\beta}_1 = \frac{-16.666}{53.8307} = -0.3096$

and

$\hat{\beta}_0 = 1519.06 - (-0.31 \times 3.16) = 1520.04.$

(b) With

$SSE = 41535625 - (1520.04 \times 27343.03) - (-0.3096 \times 86560.46) = 0.1232$

it follows that

$\hat{\sigma}^2 = \frac{0.1232}{18-2} = 0.00770.$

(c) The t-statistic is

$\frac{-0.3096}{\sqrt{0.00770/53.8307}} = -25.89$

so that the p-value is $2 \times P(t_{16} > 25.84) \simeq 0.$

Therefore, the null hypothesis $H_0 : \beta_1 = 0$ can be rejected and it can be concluded that the regression is significant.

(d) With $t_{0.025,16} = 2.120$ the confidence interval is

$\beta_0 + (\beta_1 \times 4) \in 1520.04 + (-0.3096 \times 4) \pm 2.120 \times \sqrt{0.00770} \times \sqrt{\frac{1}{18} + \frac{(4-3.166)^2}{53.8307}}$

which is

$1518.80 \pm 0.05 = (1518.75, 1518.85)$.

(e) Since

$SSR = SST - SSE = 5.2843 - 0.1232 = 5.1611$

it follows that

$R^2 = \frac{5.1611}{5.2843} = 97.7\%$.

Chapter 13

Multiple Linear Regression and Nonlinear Regression

13.1 Introduction to Multiple Linear Regression

13.1.1 (a) $R^2 = 0.89$

(b)

Source	df	SS	MS	F	p-value
Regression	3	96.5	32.17	67.4	0.000
Error	26	12.4	0.477		
Total	29	108.9			

(c) $\hat{\sigma}^2 = 0.477$

(d) The p-value is 0.000.

(e) $(16.5 - (2.056 \times 2.6), 16.5 + (2.056 \times 2.6)) = (11.2, 21.8)$

13.1.2 (a) $R^2 = 0.23$

(b)

Source	df	SS	MS	F	p-value
Regression	6	2.67	0.445	1.89	0.108
Error	38	8.95	0.2355		
Total	44	11.62			

(c) $\hat{\sigma}^2 = 0.2355$

(d) The p-value is 0.108.

(e) $(1.05 - (2.024 \times 0.91), 1.05 + (2.024 \times 0.91)) = (-0.79, 2.89)$

13.1.3 (a) $(132.4 - (2.365 \times 27.6), 132.4 + (2.365 \times 27.6)) = (67.1, 197.7)$

(b) The test statistic is $t = 4.80$ and the p-value is 0.002.

13.1.4 (a) $(0.954 - (2.201 \times 0.616), 0.954 + (2.201 \times 0.616)) = (-0.402, 2.310)$

(b) The test statistic is $t = 1.55$ and the p-value is 0.149.

13.1.5 The test statistic for $H_0 : \beta_1 = 0$ is $t = 11.30$ and the p-value is 0.000.

The test statistic for $H_0 : \beta_2 = 0$ is $t = 5.83$ and the p-value is 0.000.

The test statistic for $H_0 : \beta_3 = 0$ is $t = 1.15$ and the p-value is 0.257.

Variable x_3 should be removed from the model.

13.1.6 The test statistic is $F = 1.56$ and the p-value is 0.233.

13.1.7 The test statistic is $F = 5.29$ and the p-value is 0.013.

13.1.8 (b) $\hat{y} = 7.280 - 0.313 - 0.1861 = 6.7809$

13.1.9 (a) $\hat{y} = 104.9 + (12.76 \times 10) + (409.6 \times 0.3) = 355.38$

(b) $(355.38 - (2.110 \times 17.6), 355.38 + (2.110 \times 17.6)) = (318.24, 392.52)$

13.1.10 (a) $\hat{y} = 65.98 + (23.65 \times 1.5) + (82.04 \times 1.5) + (17.04 \times 2.0) = 258.6$

(b) $(258.6 - (2.201 \times 2.55), 258.6 + (2.201 \times 2.55)) = (253.0, 264.2)$.

13.1.11 $MSE = 4.33^2 = 18.749$

$SST = 694.09 - \frac{(-5.68)^2}{20} = 692.477$

Source	df	SS	MS	F	p-value
Regression	3	392.495	130.832	6.978	0.003
Error	16	299.982	18.749		
Total	19	692.477			

The p-value in the analysis of variance table is for the null hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$.

The proportion of the variability of the y variable that is explained by the model is

$R^2 = \frac{392.495}{692.477} = 56.7\%$.

13.1.12 (a) $R^2 = \frac{SSR}{SST} = \frac{45.76-23.98}{45.76} = 47.6\%$

(b)

Source	df	SS	MS	F	p-value
Regression	4	21.78	5.445	3.860	0.021
Error	17	23.98	1.411		
Total	21	45.76			

(c) $\hat{\sigma}^2 = MSE = 1.411$

(d) From the analysis of variance table the p-value is 0.021.

(e) With $t_{0.25,17} = 2.110$ the confidence interval is

$\beta_2 \in 183.2 \pm (2.110 \times 154.3) = (-142.4, 508.8)$

13.2 Examples of Multiple Linear Regression

13.2.1 (b) The variable competitor's price has a p-value of 0.216 and is not needed in the model.

The sample correlation coefficient between the competitor's price and the sales is $r = -0.91$.

The sample correlation coefficient between the competitor's price and the company's price is $r = 0.88$.

(c) The sample correlation coefficient between the company's price and the sales is $r = -0.96$.

Using the model

sales $= 107.4 - (3.67 \times$ company's price)

the predicted sales are $107.4 - (3.67 \times 10.0) = 70.7$.

13.2.2 $\hat{\beta}_0 = 20.011$

$\hat{\beta}_1 = -0.633$

$\hat{\beta}_2 = -1.467$

$\hat{\beta}_3 = 2.083$

$\hat{\beta}_4 = -1.717$

$\hat{\beta}_5 = 0.925$

All terms should be kept in the model.

It can be estimated that the fiber strength is maximized at

$x_1 = -0.027$ and $x_2 = 0.600$.

13.2.3 (a) $\hat{\beta}_0 = -3,238.6$

$\hat{\beta}_1 = 0.9615$

$\hat{\beta}_2 = 0.732$

$\hat{\beta}_3 = 2.889$

$\hat{\beta}_4 = 389.9$

(b) The variable geology has a p-value of 0.737 and is not needed in the model.

The sample correlation coefficient between the cost and geology is $r = 0.89$.

The sample correlation coefficient between the depth and geology is $r = 0.92$.

The variable geology is not needed in the model because it is highly correlated with the variable depth which is in the model.

(c) The variable rig-index can also be removed from the model.

A final model

cost $= -3011 + (1.04 \times \text{depth}) + (2.67 \times \text{downtime})$

can be recommended.

13.2.4 A final model

VO2-max $= 88.8 - (0.343 \times \text{heart rate}) - (0.195 \times \text{age}) - (0.901 \times \text{bodyfat})$

can be recommended.

13.2.5 Two indicator variables x_1 and x_2 are needed.
One way is to have $(x_1, x_2) = (0, 0)$ at level 1,
$(x_1, x_2) = (0, 1)$ at level 2,
and $(x_1, x_2) = (1, 0)$ at level 3.

13.2.6 No bounds can be put on the p-value for x_2 in the simple linear regression.
It can take any value.

13.3 Matrix Algebra Formulation of Multiple Linear Regression

13.3.1 (a)

$$\mathbf{Y} = \begin{pmatrix} 2 \\ -2 \\ 4 \\ -2 \\ 2 \\ -4 \\ 1 \\ 3 \\ 1 \\ -5 \end{pmatrix}$$

(b)

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 1 & 1 & 4 \\ 1 & 1 & -4 \\ 1 & -1 & 2 \\ 1 & -1 & -2 \\ 1 & 2 & 0 \\ 1 & 2 & 0 \\ 1 & -2 & 3 \\ 1 & -2 & -3 \end{pmatrix}$$

(c)

$$\mathbf{X'X} = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 20 & 0 \\ 0 & 0 & 60 \end{pmatrix}$$

(d)

$$(\mathbf{X'X})^{-\mathbf{1}} = \begin{pmatrix} 0.1000 & 0 & 0 \\ 0 & 0.0500 & 0 \\ 0 & 0 & 0.0167 \end{pmatrix}$$

(e)

$$\mathbf{X'Y} = \begin{pmatrix} 0 \\ 20 \\ 58 \end{pmatrix}$$

(g)

$$\hat{\mathbf{Y}} = \begin{pmatrix} 0.967 \\ -0.967 \\ 4.867 \\ -2.867 \\ 0.933 \\ -2.933 \\ 2.000 \\ 2.000 \\ 0.900 \\ -4.900 \end{pmatrix}$$

(h)

$$\mathbf{e} = \begin{pmatrix} 1.033 \\ -1.033 \\ -0.867 \\ 0.867 \\ 1.067 \\ -1.067 \\ -1.000 \\ 1.000 \\ 0.100 \\ -0.100 \end{pmatrix}$$

(i) $SSE = 7.933$

(k) $s.e.(\hat{\beta}_1) = 0.238$

$s.e.(\hat{\beta}_2) = 0.137$

Both input variables should be kept in the model.

(l) The fitted value is

$0 + (1 \times 1) + \left(\frac{29}{30} \times 2\right) = 2.933.$

The standard error is 0.496.

The confidence interval is $(1.76, 4.11)$.

(m) The prediction interval is $(0.16, 5.71)$.

13.3.2 (a)

$$\mathbf{Y} = \begin{pmatrix} 3 \\ -5 \\ 2 \\ 4 \\ 4 \\ 6 \\ 3 \\ 15 \end{pmatrix}$$

(b)

$$\mathbf{X} = \begin{pmatrix} 1 & -3 & 0.5 \\ 1 & -2 & -3.0 \\ 1 & -1 & 0.5 \\ 1 & 0 & -1.0 \\ 1 & 0 & -1.0 \\ 1 & 1 & 1.5 \\ 1 & 2 & -1.0 \\ 1 & 3 & 3.5 \end{pmatrix}$$

(c)

$$\mathbf{X'X} = \begin{pmatrix} 8 & 0 & 0 \\ 0 & 28 & 14 \\ 0 & 14 & 27 \end{pmatrix}$$

(d)

$$(\mathbf{X'X})^{-\mathbf{1}} = \begin{pmatrix} 0.125 & 0 & 0 \\ 0 & 0.048 & -0.025 \\ 0 & -0.025 & 0.050 \end{pmatrix}$$

(e)

$$\mathbf{X'Y} = \begin{pmatrix} 32 \\ 56 \\ 68 \end{pmatrix}$$

(f)

$$\hat{\beta} = \begin{pmatrix} 4 \\ 1 \\ 2 \end{pmatrix}$$

(g)

$$\hat{\mathbf{Y}} = \begin{pmatrix} 2 \\ -4 \\ 4 \\ 2 \\ 2 \\ 8 \\ 4 \\ 14 \end{pmatrix}$$

(h)

$$\mathbf{e} = \begin{pmatrix} 1 \\ -1 \\ -2 \\ 2 \\ 2 \\ -2 \\ -1 \\ 1 \end{pmatrix}$$

(j) $\hat{\sigma}^2 = 4$

(k) $s.e.(\hat{\beta}_1) = 0.439$

$s.e.(\hat{\beta}_2) = 0.447$

Perhaps the variable x_1 could be dropped from the model (the p-value is 0.072).

(l) The fitted value is

$4 + (1 \times 1) + (2 \times 1) = 7.$

The standard error is 0.832.

The confidence interval is $(4.86, 9.14)$.

(m) The prediction interval is $(1.43, 12.57)$.

13.3.3

$$\mathbf{Y} = \begin{pmatrix} 10 \\ 0 \\ -5 \\ 2 \\ 3 \\ -6 \end{pmatrix}$$

$$\mathbf{X} = \begin{pmatrix} 1 & -3 & 1 & 3 \\ 1 & -2 & 1 & 0 \\ 1 & -1 & 1 & -5 \\ 1 & 1 & -6 & 1 \\ 1 & 2 & -3 & 0 \\ 1 & 3 & 6 & 1 \end{pmatrix}$$

$$\mathbf{X'X} = \begin{pmatrix} 6 & 0 & 0 & 0 \\ 0 & 28 & 0 & 0 \\ 0 & 0 & 84 & -2 \\ 0 & 0 & -2 & 36 \end{pmatrix}$$

$$(\mathbf{X'X})^{-\mathbf{1}} = \begin{pmatrix} 0.16667 & 0 & 0 & 0 \\ 0 & 0.03571 & 0 & 0 \\ 0 & 0 & 0.01192 & 0.00066 \\ 0 & 0 & 0.00066 & 0.02781 \end{pmatrix}$$

$$\mathbf{X'Y} = \begin{pmatrix} 4 \\ -35 \\ -52 \\ 51 \end{pmatrix}$$

$$\hat{\beta} = \begin{pmatrix} 0.6667 \\ -1.2500 \\ -0.5861 \\ 1.3841 \end{pmatrix}$$

13.4 Evaluating Model Accuracy

13.4.1 (a) There is a slight suggestion of a greater variability in the yields at higher temperatures.

(b) There are no unusually large standardized residuals.

(c) The points $(90, 85)$ and $(200, 702)$ have leverage values $h_{ii} = 0.547$.

13.4.2 (a) The residual plots do not indicate any problems.

(b) If it were beneficial to add the variable geology to the model, then there would be some pattern in this residual plot.

(d) The observation with a cost of 8089.5 has a standardized residual of 2.01.

13.4.3 (a) The residual plots do not indicate any problems.

(b) If it were beneficial to add the variable weight to the model, then there would be some pattern in this residual plot.

(d) The observation with VO2-max = 23 has a standardized residual of -2.15.

13.4.4 The leverage values only depend upon the design matrix $\mathbf{X}$ and will not change if any of the values y_i are altered.

13.6 Supplementary Problems

13.6.1 (b)

Source	df	SS	MS	F	p-value
Regression	2	2,224.8	1,112.4	228.26	0.000
Error	8	39.0	4.875		
Total	10	2,263.7			

(c) The test statistic is $t = 5.85$ and the p-value is 0.000.

(d) The fitted value is

$$18.18 - (44.90 \times 1) + (44.08 \times 1^2) = 17.36.$$

The confidence interval is $(15.04, 19.68)$.

13.6.2 (a)

$$\mathbf{Y} = \begin{pmatrix} 24 \\ 8 \\ 14 \\ 6 \\ 0 \\ 2 \\ -8 \\ -8 \\ -12 \\ -16 \end{pmatrix}$$

$$\mathbf{X} = \begin{pmatrix} 1 & -4 & 5 \\ 1 & -4 & -5 \\ 1 & -2 & 2 \\ 1 & -2 & -2 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 4 & 2 \\ 1 & 4 & -2 \\ 1 & 6 & 5 \\ 1 & 6 & -5 \end{pmatrix}$$

(b)

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 10 & 10 & 0 \\ 10 & 146 & 0 \\ 0 & 0 & 116 \end{pmatrix}$$

(e)

$$\hat{\mathbf{Y}} = \begin{pmatrix} 21 \\ 11 \\ 12 \\ 8 \\ 1 \\ 1 \\ -6 \\ -10 \\ -9 \\ -19 \end{pmatrix}$$

$$\mathbf{e} = \begin{pmatrix} 3 \\ -3 \\ 2 \\ -2 \\ -1 \\ 1 \\ -2 \\ 2 \\ -3 \\ 3 \end{pmatrix}$$

(f) $SSE = 54$

(h) $s.e.(\hat{\beta}_1) = 0.238$

$s.e.(\hat{\beta}_2) = 0.258$

Both input variables should be kept in the model.

(i) The fitted value is

$4 - (3 \times 2) + (1 \times (-2)) = -4.$

The standard error is 1.046.

(j) The prediction interval is $(-11.02, 3.02)$.

13.6.3 (b) The fitted model is

powerloss $= 30.2 + (0.0933 \times \text{diameter}) - (4.081 \times \text{clearance})$

with $\hat{\sigma} = 0.0764$.

When the bearing diameter is 25 and the bearing clearance is 0.07, the fitted value is 32.29 and a 95% prediction interval is $(32.13, 32.45)$.

The data point with a bearing diameter of 28.2, a bearing clearance of 0.086, and a powerloss of 32.35 has a standardized residual of -2.46.

13.6.4 (a) The values of the additive levels and the temperature levels used in the experiment have been chosen according to a grid pattern.

(b) The data point obtained with an additive level of 2.3 and a temperature of 160 has a standardized residual of -3.01.

(c) The fitted model is maximized with an additive level of 2.80 and a temperature of 155.4.

13.6.5 $e_1 = 288.9 - (-67.5 + (34.5 \times 12.3) - (0.44 \times 143.4) + (108.6 \times (-7.2)) + (55.8 \times 14.4))$

$= -26.454$

Since

$e_1^* = \frac{e_1}{\hat{\sigma}\sqrt{1-h_{11}}}$

it follows that

$-1.98 = \frac{-26.454}{\hat{\sigma}\sqrt{1-0.0887}}$

so that $\hat{\sigma} = 13.996$.

Therefore,

$SSE = MSE \times (44 - 4 - 1) = 13.996^2 \times 39 = 7639$

so that

$R^2 = \frac{SST-SSE}{SST} = \frac{20554-7639}{20554} = 62.8\%$.

13.6.6 The sample correlation coefficient between y and x_3 could be either negative, zero, or positive.

13.6.7 (a) True

(b) False

13.6.8 (a) $\frac{\hat{\beta}_1}{s.e.(\hat{\beta}_1)} = \frac{-45.2}{39.5} = -1.14$

$\frac{\hat{\beta}_2}{s.e.(\hat{\beta}_2)} = \frac{3.55}{5.92} = 0.60$

The p-value for variable 1 is $2 \times P(t_{27} \geq 1.14) = 0.26$.

The p-value for variable 2 is $2 \times P(t_{27} \geq 0.60) = 0.55$.

The variables should be removed sequentially. Variable 2 should be removed first since it has the largest p-value. When variable 2 has been removed and a simple linear regression is performed with variable 1, the p-value of variable 1 may change. Therefore, it is not clear whether variable 1 should also be removed from the model. It is not clear that both variables should be removed from the model. False.

(b) $\frac{\hat{\beta}_1}{s.e.(\hat{\beta}_1)} = \frac{-45.2}{8.6} = -5.26$

$\frac{\hat{\beta}_2}{s.e.(\hat{\beta}_2)} = \frac{3.55}{0.63} = 5.63$

The p-value for variable 1 is $2 \times P(t_{27} \geq -5.26) = 0.000$.

The p-value for variable 2 is $2 \times P(t_{27} \geq 5.63) = 0.000$.

Neither variable should be removed from the model. True.

Chapter 14

Multifactor Experimental Design and Analysis

14.1 Experiments with Two Factors

14.1.1

Source	df	SS	MS	F	p-value
Fuel	1	96.33	96.33	3.97	0.081
Car	1	75.00	75.00	3.09	0.117
Fuel*Car	1	341.33	341.33	14.08	0.006
Error	8	194.00	24.25		
Total	11	706.66			

14.1.2 (a)

Source	df	SS	MS	F	p-value
Type	3	160.61	53.54	9.63	0.002
Temp	2	580.52	290.26	52.22	0.000
Type*Temp	6	58.01	9.67	1.74	0.195
Error	12	66.71	5.56		
Total	23	865.85			

(c) With a confidence level 95% the pairwise comparisons are:

$\alpha_1 - \alpha_2 \in (0.26, 8.34)$

$\alpha_1 - \alpha_3 \in (-2.96, 5.12)$

$\alpha_1 - \alpha_4 \in (-6.97, 1.11)$

$\alpha_2 - \alpha_3 \in (-7.26, 0.82)$

$\alpha_2 - \alpha_4 \in (-11.27, -3.19)$

$\alpha_3 - \alpha_4 \in (-8.05, 0.03)$

(d) With a confidence level 95% the pairwise comparisons are:

$\beta_1 - \beta_2 \in (4.61, 10.89)$

$\beta_1 - \beta_3 \in (8.72, 15.00)$

$\beta_2 - \beta_3 \in (0.97, 7.25)$

14.1.3 (a)

Source	df	SS	MS	F	p-value
Tip	2	0.1242	0.0621	1.86	0.175
Material	2	14.1975	7.0988	212.31	0.000
Tip*Material	4	0.0478	0.0120	0.36	0.837
Error	27	0.9028	0.0334		
Total	35	15.2723			

(c) Apart from one large negative residual there appears to be less variability in the measurements from the third tip.

14.1.4 (a)

Source	df	SS	MS	F	p-value
Material	3	51.7	17.2	0.11	0.957
Magnification	3	13493.7	4,497.9	27.47	0.000
Material*Magnification	9	542.1	60.2	0.37	0.947
Error	80	13098.8	163.7		
Total	95	27186.3			

(c) Material type 3 has the least amount of variability.

14.1.5

Source	df	SS	MS	F	p-value
Glass	2	3.134	1.567	0.32	0.732
Acidity	1	18.201	18.201	3.72	0.078
Glass*Acidity	2	83.421	41.711	8.52	0.005
Error	12	58.740	4.895		
Total	17	163.496			

14.1.6

Source	df	SS	MS	F	p-value
A	2	230.11	115.06	11.02	0.004
B	2	7.44	3.72	0.36	0.710
A*B	4	26.89	6.72	0.64	0.645
Error	9	94.00	10.44		
Total	17	358.44			

The low level of ingredient B has the smallest amount of variability in the percentage improvements.

14.1.7

Source	df	SS	MS	F	p-value
Design	2	3.896×10^3	1.948×10^3	0.46	0.685
Material	1	0.120×10^3	0.120×10^3	0.03	0.882
Error	2	8.470×10^3	4.235×10^3		
Total	5	12.487×10^3			

14.2 Experiments with Three or More Factors

14.2.1 (d)

Source	df	SS	MS	F	p-value
Drink	2	90.65	45.32	5.39	0.007
Gender	1	6.45	6.45	0.77	0.384
Age	2	23.44	11.72	1.39	0.255
Drink*Gender	2	17.82	8.91	1.06	0.352
Drink*Age	4	24.09	6.02	0.72	0.583
Gender*Age	2	24.64	12.32	1.47	0.238
Drink*Gender*Age	4	27.87	6.97	0.83	0.511
Error	72	605.40	8.41		
Total	89	820.36			

14.2.2 (a)

Source	df	SS	MS	F	p-value
Rice	2	527.0	263.5	1.72	0.193
Fert	1	2394.2	2394.2	15.62	0.000
Sun	1	540.0	540.0	3.52	0.069
Rice*Fert	2	311.6	155.8	1.02	0.372
Rice*Sun	2	2076.5	1038.3	6.78	0.003
Fert*Sun	1	77.5	77.5	0.51	0.481
Rice*Fert*Sun	2	333.3	166.6	1.09	0.348
Error	36	5516.3	153.2		
Total	47	11776.5			

(b) Yes

(c) No

(d) Yes

14.2.3 (a)

Source	df	SS	MS	F	p-value
Add-A	2	324.11	162.06	8.29	0.003
Add-B	2	5.18	2.59	0.13	0.877
Conditions	1	199.28	199.28	10.19	0.005
Add-A*Add-B	4	87.36	21.84	1.12	0.379
Add-A*Conditions	2	31.33	15.67	0.80	0.464
Add-B*Conditions	2	2.87	1.44	0.07	0.930
Add-A*Add-B*Conditions	4	21.03	5.26	0.27	0.894
Error	18	352.05	19.56		
Total	35	1023.21			

The amount of additive B does not effect the expected value of the gas mileage although the variability of the gas mileage increases as more of additive B is used.

14.2.4

Source	df	SS	MS	F	p-value
Radar	3	40.480	13.493	5.38	0.009
Aircraft	1	2.750	2.750	1.10	0.311
Period	1	0.235	0.235	0.09	0.764
Radar*Aircraft	3	142.532	47.511	18.94	0.000
Radar*Period	3	8.205	2.735	1.09	0.382
Aircraft*Period	1	5.152	5.152	2.05	0.171
Radar*Aircraft*Period	3	5.882	1.961	0.78	0.521
Error	16	40.127	2.508		
Total	31	245.362			

14.2.5 (d)

Source	df	SS	MS	F	p-value
Machine	1	387.1	387.1	3.15	0.095
Temp	1	29.5	29.5	0.24	0.631
Position	1	1271.3	1271.3	10.35	0.005
Angle	1	6865.0	6685.0	55.91	0.000
Machine*Temp	1	43.0	43.0	0.35	0.562
Machine*Position	1	54.9	54.9	0.45	0.513
Machine*Angle	1	1013.6	1013.6	8.25	0.011
Temp*Position	1	67.6	67.6	0.55	0.469
Temp*Angle	1	8.3	8.3	0.07	0.798
Position*Angle	1	61.3	61.3	0.50	0.490
Machine*Temp*Position	1	21.0	21.0	0.17	0.685
Machine*Temp*Angle	1	31.4	31.4	0.26	0.620
Machine*Position*Angle	1	13.7	13.7	0.11	0.743
Temp*Position*Angle	1	17.6	17.6	0.14	0.710
Machine*Temp*Position*Angle	1	87.5	87.5	0.71	0.411
Error	16	1964.7	122.8		
Total	31	11937.3			

14.2.6

Source	df	SS	MS	F	p-value
Player	1	72.2	72.2	0.21	0.649
Club	1	289.0	289.0	0.84	0.365
Ball	1	225.0	225.0	0.65	0.423
Weather	1	2626.6	2626.6	7.61	0.008
Player*Club	1	72.2	72.2	0.21	0.649
Player*Ball	1	169.0	169.0	0.49	0.488
Player*Weather	1	826.6	826.6	2.39	0.128
Club*Ball	1	5700.3	5700.3	16.51	0.000
Club*Weather	1	10.6	10.6	0.03	0.862
Ball*Weather	1	115.6	115.6	0.33	0.566
Player*Club*Ball	1	22500.0	22500.0	65.17	0.000
Player*Club*Weather	1	297.6	297.6	0.86	0.358
Player*Ball*Weather	1	115.6	115.6	0.33	0.566
Club*Ball*Weather	1	14.1	14.1	0.04	0.841
Player*Club*Ball*Weather	1	0.6	0.6	0.00	0.968
Error	48	16571.0	345.2		
Total	63	49605.8			

14.2.7 A redundant

B redundant

C redundant

D redundant

A*B not significant

A*C redundant

A*D redundant

B*C significant

B*D significant

C*D redundant

A*B*C not significant

A*B*D not significant

A*C*D significant

B*C*D not significant

A*B*C*D not significant

14.2.8 A plot should be made of the data averaged over the levels of factor B.

	A low	A middle	A high
C low	45	42.5	51
C high	45	45.5	68

14.3 Supplementary Problems

14.3.1 (a)

Source	df	SS	MS	F	p-value
Material	2	106.334	53.167	34.35	0.000
Pressure	2	294.167	147.084	95.03	0.000
Material*Pressure	4	2.468	0.617	0.40	0.808
Error	27	41.788	1.548		
Total	35	444.756			

(c) With a confidence level 95% the pairwise comparisons are:

$\alpha_1 - \alpha_2 \in (2.61, 5.13)$

$\alpha_1 - \alpha_3 \in (2.11, 4.64)$

$\alpha_2 - \alpha_3 \in (-1.75, 0.77)$

(d) With a confidence level 95% the pairwise comparisons are:

$\beta_1 - \beta_2 \in (-0.96, 1.56)$

$\beta_1 - \beta_3 \in (-7.17, -4.65)$

$\beta_2 - \beta_3 \in (-7.47, -4.95)$

14.3.2

Source	df	SS	MS	F	p-value
Location	1	34.13	34.13	2.29	0.144
Coating	2	937.87	468.93	31.40	0.000
Location*Coating	2	43.47	21.73	1.46	0.253
Error	24	358.40	14.93		
Total	29	1373.87			

14.3.3

Source	df	SS	MS	F	p-value
Drug	3	593.19	197.73	62.03	0.000
Severity	1	115.56	115.56	36.25	0.000
Drug*Severity	3	86.69	28.90	9.07	0.006
Error	8	25.50	3.19		
Total	15	820.94			

14.3.4 (c)

Source	df	SS	MS	F	p-value
Furnace	1	570.38	570.38	27.77	0.000
Layer	2	18.08	9.04	0.44	0.654
Position	1	495.04	495.04	24.10	0.000
Furnace*Layer	2	23.25	11.63	0.57	0.582
Furnace*Position	1	18.38	18.38	0.89	0.363
Layer*Position	2	380.08	190.04	9.25	0.004
Furnace*Layer*Position	2	84.25	42.13	2.05	0.171
Error	12	246.50	20.54		
Total	23	1835.96			

14.3.5

Source	df	SS	MS	F	p-value
Monomer	1	19.220	19.220	15.40	0.001
Stab	1	13.781	13.781	11.04	0.004
Cat	1	36.125	36.125	28.94	0.000
Water	1	4.061	4.061	3.25	0.090
Monomer*Stab	1	0.000	0.000	0.00	1.000
Monomer*Cat	1	22.781	22.781	18.25	0.001
Monomer*Water	1	11.520	11.520	9.23	0.008
Stab*Cat	1	0.405	0.405	0.32	0.577
Stab*Water	1	0.011	0.011	0.01	0.926
Cat*Water	1	0.845	0.845	0.68	0.423
Monomer*Stab*Cat	1	2.101	2.101	1.68	0.213
Monomer*Stab*Water	1	2.000	2.000	1.60	0.224
Monomer*Cat*Water	1	0.281	0.281	0.23	0.641
Stab*Cat*Water	1	1.445	1.445	1.16	0.298
Monomer*Stab*Cat*Water	1	0.101	0.101	0.08	0.779
Error	16	19.970	1.248		
Total	31	134.649			

14.3.6

Source	df	SS	MS	F	p-value
Lathe	1	1144.7	1144.7	8.92	0.006
Operator	2	2325.7	1162.9	9.06	0.001
Lathe*Operator	2	485.7	242.9	1.89	0.168
Error	30	3849.5	128.3		
Total	35	7805.6			

There is sufficient evidence to conclude that lathe 2 is more efficient than lathe 1.

There is no evidence of an interaction effect, so there is no evidence that the difference between the lathes is not the same for each of the operators.

14.3.7

Source	df	SS	MS	F	p-value
Speed	2	400.509	200.254	171.65	0.000
Cooler	3	215.884	71.961	61.68	0.000
Position	1	0.101	0.101	0.09	0.771
Speed*Cooler	6	59.550	9.925	8.51	0.000
Speed*Position	2	41.595	20.798	17.83	0.000
Cooler*Position	3	55.034	18.345	15.72	0.000
Speed*Cooler*Position	6	47.280	7.880	6.75	0.000
Error	24	28.000	1.167		
Total	47	847.953			

Chapter 15

Nonparametric Statistical Analysis

15.1 The Analysis of a Single Population

15.1.1 (c) It is not plausible.

(d) It is not plausible.

(e) $S(65) = 84$
The p-value is 0.064.

(f) The p-value is 0.001.

(g) The confidence interval from the sign test is $(65.0, 69.0)$.
The confidence interval from the signed rank test is $(66.0, 69.5)$.

15.1.2 (c) A $N(1.1, 0.05^2)$ distribution is plausible while
a $N(1.0, 0.05^2)$ distribution is not plausible.

(d) $S(1.1) = 51$
The p-value is 0.049.

(e) The p-values are 0.014 for the signed rank test and 0.027 for the t-test.

(f) The confidence interval from the sign test is $(1.102, 1.120)$.
The confidence interval from the signed rank test is $(1.102, 1.120)$.
The confidence interval from the t-test is $(1.101, 1.120)$.

15.1.3 The p-values for the hypotheses
$H_0 : \mu = 0.2$ versus $H_A : \mu \neq 0.2$
are 0.004 for the sign test,

0.000 for the signed rank test,

and 0.000 for the t-test.

Confidence intervals for μ with a confidence level of at least 95% are

$(0.207, 0.244)$ for the sign test,

$(0.214, 0.244)$ for the signed rank test,

and $(0.216, 0.248)$ for the t-test.

There is sufficient evidence to conclude that the median paint thickness is larger than 0.2 mm.

15.1.4 The p-values for the hypotheses

$H_0 : \mu \geq 9.5$ versus $H_A : \mu < 9.5$

are 0.288 for the sign test,

0.046 for the signed rank test,

and 0.003 for the t-test.

A histogram of the data shows a skewed distribution, so that the assumptions of symmetry and normality required by the signed rank test and the t-test respectively appear to be invalid.

The sign test does not provide support for the statement that the median is less than 9.5.

15.1.5 (a) $S(18.0) = 14$

(b) The exact p-value is
$2 \times P(B(20, 0.5) \geq 14) = 0.115$.

(c) $2 \times \Phi(-1.57) = 0.116$

(d) $S_+(18.0) = 37$

(e) $2 \times \Phi(-2.52) = 0.012$

15.1.6 (a) $S(40) = 7$

(b) The exact p-value is
$2 \times P(B(25, 0.5) \leq 7) = 0.064$.

(c) $2 \times \Phi(-2.00) = 0.046$

(d) $S_+(40) = 241$

(e) $2 \times \Phi(-2.10) = 0.036$

15.1.7 It is reasonable to assume that the differences of the data have a symmetric distribution in which case the signed rank test can be used.

The p-values for the hypotheses

$H_0 : \mu_A - \mu_B = 0$ versus $H_A : \mu_A - \mu_B \neq 0$

are 0.296 for the sign test and 0.300 for the signed rank test.

Confidence intervals for $\mu_A - \mu_B$ with a confidence level of at least 95% are

$(-1.0, 16.0)$ for the sign test and

$(-6.0, 17.0)$ for the signed rank test.

There is not sufficient evidence to conclude that there is a difference between the two assembly methods.

15.1.8 The p-values for the hypotheses

$H_0 : \mu_A - \mu_B = 0$ versus $H_A : \mu_A - \mu_B \neq 0$

are 0.774 for the sign test and 0.480 for the signed rank test.

Confidence intervals for $\mu_A - \mu_B$ with a confidence level of at least 95% are

$(-6.0, 4.0)$ for the sign test and

$(-4.0, 2.0)$ for the signed rank test.

There is not sufficient evidence to conclude that there is a difference between the two stimulation conditions.

15.1.9 The p-values for the hypotheses

$H_0 : \mu_A - \mu_B = 0$ versus $H_A : \mu_A - \mu_B \neq 0$

are 0.003 for the sign test and 0.002 for the signed rank test.

Confidence intervals for $\mu_A - \mu_B$ with a confidence level of at least 95% are

$(-13.0, -1.0)$ for the sign test and

$(-12.0, -3.5)$ for the signed rank test.

The signed rank test shows that the new teaching method is better by at least 3.5 points on average.

15.1.10 The p-values for the hypotheses

$H_0 : \mu_A - \mu_B = 0$ versus $H_A : \mu_A - \mu_B \neq 0$

are 0.815 for the sign test and 0.879 for the signed rank test.

Confidence intervals for $\mu_A - \mu_B$ with a confidence level of at least 95% are

$(-70.0, 80.0)$ for the sign test and

$(-65.0, 65.0)$ for the signed rank test.

There is not sufficient evidence to conclude that there is a difference between the two dating methods.

15.1.11 The p-values for the hypotheses

$H_0 : \mu_A - \mu_B = 0$ versus $H_A : \mu_A - \mu_B \neq 0$

are 0.541 for the sign test and 0.721 for the signed rank test.

Confidence intervals for $\mu_A - \mu_B$ with a confidence level of at least 95% are

$(-13.6, 7.3)$ for the sign test and

$(-6.6, 6.3)$ for the signed rank test.

There is not sufficient evidence to conclude that there is a difference between the two ball types.

15.2 Comparing Two Populations

15.2.1 (c) The Kolmogorov-Smirnov statistic is $M = 0.2006$, which is larger than

$d_{0.01}\sqrt{\frac{1}{200} + \frac{1}{180}} = 0.167.$

There is sufficient evidence to conclude that the two distribution functions are different.

15.2.2 (c) The Kolmogorov-Smirnov statistic is $M = 0.376$, which is larger than

$d_{0.01}\sqrt{\frac{1}{125} + \frac{1}{125}} = 0.206.$

There is sufficient evidence to conclude that the two distribution functions are different.

15.2.3 The Kolmogorov-Smirnov statistic is $M = 0.40$, which is larger than

$d_{0.01}\sqrt{\frac{1}{50} + \frac{1}{50}} = 0.326.$

There is sufficient evidence to conclude that the two distribution functions are different.

15.2.4 (b) $S_A = 75.5$

(c) $U_A = 75.5 - \frac{8\times(8+1)}{2} = 39.5$

(d) Since

$U_A = 39.5 < \frac{mn}{2} = \frac{8\times 13}{2} = 52$

the value of U_A is consistent with the observations from population A being smaller than the observations from population B.

(e) The p-value is 0.385.
There is not sufficient evidence to conclude that there is a difference between the two distribution functions.

15.2.5 (b) $S_A = 245$

(c) $U_A = 245 - \frac{14\times(14+1)}{2} = 140$

(d) Since

$U_A = 140 > \frac{mn}{2} = \frac{14\times 12}{2} = 84$

the value of U_A is consistent with the observations from population A being larger than the observations from population B.

(e) The p-value is 0.004.
There is sufficient evidence to conclude that the observations from population A tend to be larger than the observations from population B.

15.2.6 (b) $S_A = 215.5$

(c) $U_A = 215.5 - \frac{15\times(15+1)}{2} = 95.5$

(d) Since

$U_A = 95.5 < \frac{mn}{2} = \frac{15\times15}{2} = 112.5$

the value of U_A is consistent with the hypothesis that the observations from the standard treatment are smaller than the observations from the new treatment.

(e) The one-sided p-value is 0.247.
There is not sufficient evidence to conclude that there is a difference between the new and the standard treatments.

15.2.7 (c) The Kolmogorov-Smirnov statistic is $M = 0.218$, which is approximately equal to

$d_{0.05}\sqrt{\frac{1}{75} + \frac{1}{82}} = 0.217.$

There is some evidence that the two distribution functions are different, although the evidence is not overwhelming.

(d) $S_A = 6555.5$

$U_A = 6555.5 - \frac{75\times(75+1)}{2} = 3705.5$

Since

$U_A = 3705.5 > \frac{mn}{2} = \frac{75\times82}{2} = 3075.0$

the value of U_A is consistent with the observations from production line A being larger than the observations from production line B.

The two-sided p-value is 0.027.

A 95% confidence interval for the difference in the population medians is $(0.003, 0.052)$.

The rank sum test is based on the assumption that the two distribution functions are identical except for a location difference, and the plots of the empirical cumulative distribution functions in (a) suggest that this assumption is not unreasonable.

15.2.8 The rank sum test has a two-sided p-value of 0.24 and there is not sufficient evidence to conclude that there is a difference between the low and high levels of hydrogen peroxide.

15.3 Comparing Three or More Populations

15.3.1 (b) $\bar{r}_{1.} = 16.6$

$\bar{r}_{2.} = 15.5$

$\bar{r}_{3.} = 9.9$

(c) $H = 3.60$

(d) The p-value is $P(\chi^2_2 > 3.60) = 0.165$.

15.3.2 (a) $\bar{r}_{1.} = 10.4$

$\bar{r}_{2.} = 26.1$

$\bar{r}_{3.} = 35.4$

$\bar{r}_{4.} = 12.5$

(b) $H = 28.52$

(c) The p-value is $P(\chi^2_3 > 28.52) = 0.000$.

15.3.3 (a) $\bar{r}_{1.} = 17.0$

$\bar{r}_{2.} = 19.8$

$\bar{r}_{3.} = 14.2$

$H = 1.84$

The p-value is $P(\chi^2_2 > 1.84) = 0.399$.

There is not sufficient evidence to conclude that the radiation readings are affected by the background radiation level.

(b) See Problem 11.1.15.

15.3.4 $\bar{r}_{1.} = 13.0$

$\bar{r}_{2.} = 28.5$

$\bar{r}_{3.} = 10.9$

$H = 20.59$

The p-value is $P(\chi^2_2 > 20.59) = 0.000$.

There is sufficient evidence to conclude that the different layouts affect the time taken to perform a task.

15.3.5 $\bar{r}_{1.} = 55.1$

$\bar{r}_{2.} = 55.7$

$\bar{r}_{3.} = 25.7$

$H = 25.86$

The p-value is $P(\chi_2^2 > 25.86) = 0.000$.

There is sufficient evidence to conclude that the computer assembly times are affected by the different assembly methods.

15.3.6 (b) $\bar{r}_{1.} = 1.50$

$\bar{r}_{2.} = 2.83$

$\bar{r}_{3.} = 1.67$

(c) $S = 6.33$

(d) The p-value is $P(\chi_2^2 > 6.33) = 0.043$.

15.3.7 (a) $\bar{r}_{1.} = 2.250$

$\bar{r}_{2.} = 1.625$

$\bar{r}_{3.} = 3.500$

$\bar{r}_{4.} = 2.625$

(b) $S = 8.85$

(c) The p-value is $P(\chi_3^2 > 8.85) = 0.032$.

15.3.8 (a) $\bar{r}_{1.} = 2.429$

$\bar{r}_{2.} = 2.000$

$\bar{r}_{3.} = 1.571$

$S = 2.57$

The p-value is $P(\chi_2^2 > 2.57) = 0.277$.

There is not sufficient evidence to conclude that the calciners are operating at different efficiencies.

(b) See Problem 11.2.9.

15.3.9 $\bar{r}_{1.} = 1.125$

$\bar{r}_{2.} = 2.875$

$\bar{r}_{3.} = 2.000$

$S = 12.25$

The p-value is $P(\chi_2^2 > 12.25) = 0.002$.

There is sufficient evidence to conclude that there is a difference between the radar systems.

15.3.10 $\bar{r}_{1.} = 2.4$

$\bar{r}_{2.} = 1.7$

$\bar{r}_{3.} = 1.9$

$S = 2.60$

The p-value is $P(\chi_2^2 > 2.60) = 0.273$.

There is not sufficient evidence to conclude that there is any difference between the assembly methods.

15.3.11 $\bar{r}_{1.} = 4.42$

$\bar{r}_{2.} = 2.50$

$\bar{r}_{3.} = 1.79$

$\bar{r}_{4.} = 1.71$

$\bar{r}_{5.} = 4.58$

$S = 37.88$

The p-value is $P(\chi_4^2 > 37.88) = 0.000$.

There is sufficient evidence to conclude that there is a difference in the performances of the agents.

15.3.12 $\bar{r}_{1.} = 2.375$

$\bar{r}_{2.} = 2.225$

$\bar{r}_{3.} = 3.100$

$\bar{r}_{4.} = 2.300$

$S = 5.89$

The p-value is $P(\chi_3^2 > 5.89) = 0.118$.

There is not sufficient evidence to conclude that there is any difference between the detergent formulations.

15.4 Supplementary Problems

15.4.1 (c) The distribution is not plausible.

(d) The distribution is not plausible.

(e) $S(70) = 38$
The p-value is 0.011.

(f) The p-value is 0.006.

(g) Confidence intervals for μ with a confidence level of at least 95% are
$(69.00, 70.00)$ for the sign test,
$(69.15, 69.85)$ for the signed rank test,
and $(69.23, 70.01)$ for the t-test.

15.4.2 The p-values for the hypotheses
$H_0 : \mu \geq 35$ versus $H_A : \mu < 35$
are 0.005 for the sign test,
0.000 for the signed rank test,
and 0.001 for the t-test.

Confidence intervals for μ with a confidence level of at least 95% are
$(30.9, 33.8)$ for the sign test,
$(31.3, 34.0)$ for the signed rank test,
and $(30.2, 33.9)$ for the t-test.

15.4.3 The p-values for the hypotheses
$H_0 : \mu_A - \mu_B = 0$ versus $H_A : \mu_A - \mu_B \neq 0$
are 0.115 for the sign test,
0.012 for the signed rank test,
and 0.006 for the t-test.

Confidence intervals for $\mu_A - \mu_B$ with a confidence level of at least 95% are
$(-1.20, 0.10)$ for the sign test,
$(-1.05, -0.20)$ for the signed rank test,
and $(-0.95, -0.19)$ for the t-test.

15.4.4 The p-values for the hypotheses

$H_0 : \mu_A - \mu_B = 0$ versus $H_A : \mu_A - \mu_B \neq 0$

are 0.134 for the sign test,

0.036 for the signed rank test,

and 0.020 for the t-test.

Confidence intervals for $\mu_A - \mu_B$ with a confidence level of at least 95% are

$(-1.00, 6.90)$ for the sign test,

$(0.20, 5.15)$ for the signed rank test,

and $(0.49, 5.20)$ for the t-test.

15.4.5 (c) The Kolmogorov-Smirnov statistic is $M = 0.20$, which is smaller than

$d_{0.20}\sqrt{\frac{1}{40} + \frac{1}{40}} = 0.239.$

This does not provide any evidence of a difference between the distributions of the waiting times before and after the reorganization.

15.4.6 (c) The Kolmogorov-Smirnov statistic is $M = 0.525$, which is larger than

$d_{0.01}\sqrt{\frac{1}{40} + \frac{1}{40}} = 0.364.$

There is sufficient evidence to conclude that there is a difference between the two distribution functions.

(d) $S_A = 1143$

$U_A = 1143 - \frac{40\times(40+1)}{2} = 323$

Since

$U_A = 323 < \frac{mn}{2} = \frac{40\times 40}{2} = 800$

and the p-value is 0.000, there is sufficient evidence to conclude that the heights under growing conditions A tend to be smaller than the heights under growing conditions B.

A 95% confidence interval for the difference between the median bamboo shoot heights for the two growing conditions is $(-8.30, -3.50)$.

15.4.7 (b) $S_A = 292$

(c) $U_A = 292 - \frac{20\times(20+1)}{2} = 82$

(d) The value

$U_A = 82 < \frac{mn}{2} = \frac{20\times 25}{2} = 250$

is consistent with the observations being smaller without anthraquinone than with anthraquinone.

(e) The one-sided p-value is 0.000.

15.4.8 (b) $\bar{r}_{1.} = 12.4$

$\bar{r}_{2.} = 12.6$

$\bar{r}_{3.} = 3.0$

$\bar{r}_{4.} = 14.0$

(c) $H = 10.93$

(d) The p-value is $P(\chi^2_3 > 10.93) = 0.012$.

The p-value is about equal to the boundary value of 1%.

15.4.9 (a) $\bar{r}_{1.} = 80.6$

$\bar{r}_{2.} = 84.2$

$\bar{r}_{3.} = 75.3$

$\bar{r}_{4.} = 80.9$

$H = 0.75$

The p-value is $P(\chi^2_3 > 0.75) = 0.861$.

There is not sufficient evidence to conclude that any of the cars is getting a better gas mileage than the others.

(b) See Problem 11.3.3.

15.4.10 (b) $\bar{r}_{1.} = 3.500$

$\bar{r}_{2.} = 2.500$

$\bar{r}_{3.} = 1.583$

$\bar{r}_{4.} = 4.000$

$\bar{r}_{5.} = 3.417$

(c) $S = 8.83$

(d) The p-value is $P(\chi_4^2 > 8.83) = 0.066$.

There is some evidence that the different temperature levels have an effect on the cement strength, but the evidence is not overwhelming.

15.4.11 $\bar{r}_{1.} = 1.7$

$\bar{r}_{2.} = 1.5$

$\bar{r}_{3.} = 3.5$

$\bar{r}_{4.} = 4.2$

$\bar{r}_{5.} = 4.1$

$S = 27.36$

The p-value is $P(\chi_4^2 > 27.36) = 0.000$.

There is sufficient evidence to conclude that there is a difference between the fertilizers.

15.4.12 (a) $\bar{r}_{1.} = 2.292$

$\bar{r}_{2.} = 2.000$

$\bar{r}_{3.} = 3.708$

$\bar{r}_{4.} = 2.000$

$S = 14.43$

The p-value is $P(\chi_3^2 > 14.43) = 0.002$.

There is sufficient evidence to conclude that there is a difference between the clinics.

(b) See Problem 11.3.6.

15.4.13 For the hypotheses

$H_0 : \mu \geq 25.5$ versus $H_A : \mu < 25.5$

the sign test has a p-value of 0.0006

and the signed rank test has a p-value of 0.0000.

There is sufficient evidence to conclude that the average soil compressibility is no larger than 25.5.

15.4.14 This is a paired data set.

The p-values for the hypotheses

$H_0 : \mu_A = \mu_B$ versus $H_A : \mu_A \neq \mu_B$

are 0.754 for the sign test

and 0.610 for the signed rank test.

There is not sufficient evidence to conclude that there is a difference in the average ocular motor measurements after reading a book and after reading a computer screen.

15.4.15 The rank sum test has a two-sided p-value of 0.002 and there is sufficient evidence to conclude that the average viscosity is higher after having being used in engine 2 than after having being used in engine 1.

15.4.16 The Kruskal-Wallis test gives a p-value of 0.000 and there is sufficient evidence to conclude that there is a difference between the three positions.

15.4.17 The Kruskal-Wallis test gives a p-value of 0.001 and there is sufficient evidence to conclude that there is a difference between the four different vehicle designs.

15.4.18 The Friedman test gives a p-value of 0.04.

This provides some evidence of a difference between the four different preparation methods, although the evidence is not overwhelming.

Chapter 16

Quality Control Methods

16.2 Statistical Process Control

16.2.1 (a) The center line is 10.0 and the control limits are 9.7 and 10.3.

(b) The process is declared to be out of control at $\bar{x} = 9.5$ but not at $\bar{x} = 10.25$.

(c) $P\left(9.7 \leq N\left(10.15, \frac{0.2^2}{4}\right) \leq 10.3\right) = 0.9332$

The probability that an observation lies outside the control limits is therefore $1 - 0.9332 = 0.0668$.

The average run length for detecting the change is $\frac{1}{0.0668} = 15.0$.

16.2.2 (a) The center line is 0.650 and the control limits are 0.605 and 0.695.

(b) There is no evidence that the process is out of control at either $\bar{x} = 0.662$ or at $\bar{x} = 0.610$.

(c) $P(0.605 \leq N(0.630, 0.015^2) \leq 0.695) = 0.9522$.

The probability that an observation lies outside the control limits is therefore $1 - 0.9522 = 0.0478$.

The average run length for detecting the change is $\frac{1}{0.0478} = 20.9$.

16.2.3 (a) $P(\mu - 2\sigma \leq N(\mu, \sigma^2) \leq \mu + 2\sigma) = 0.9544$

The probability that an observation lies outside the control limits is therefore $1 - 0.9544 = 0.0456$.

(b) $P(\mu - 2\sigma \leq N(\mu + \sigma, \sigma^2) \leq \mu + 2\sigma) = 0.8400$

The probability that an observation lies outside the control limits is therefore $1 - 0.8400 = 0.1600$.

The average run length for detecting the change is $\frac{1}{0.1600} = 6.25$.

16.2.4 The average run length is about $\frac{1}{1-0.9974} = 380$.

16.2.5 The probability that a point is above the center line and within the upper control limit is

$P(\mu \leq N(\mu, \sigma^2) \leq \mu + 3\sigma) = 0.4987$.

The probability that all eight points lie above the center line and within the upper control limit is therefore $0.4987^8 = 0.0038$.

Similarly, the probability that all eight points lie below the center line and within the lower control limit is $0.4987^8 = 0.0038$.

Consequently, the probability that all eight points lie on the same side of the center line and within the control limits is $2 \times 0.0038 = 0.0076$.

Since this probability is very small, if all eight points lie on the same side of the centerline this suggests that the process has moved out of control, even though the points may all lie within the control limits.

16.3 Variable Control Charts

16.3.1 (a) The $\bar{X}$-chart has a center line at 91.33 and control limits at 87.42 and 95.24.

The R-chart has a center line at 5.365 and control limits at 0 and 12.24.

(b) No

(c) $\bar{x} = 92.6$

$r = 13.1$

The process can be declared to be out of control due to an increase in the variability.

(d) $\bar{x} = 84.6$

$r = 13.5$

The process can be declared to be out of control due to an increase in the variability and a decrease in the mean value.

(e) $\bar{x} = 91.8$

$r = 5.7$

There is no evidence that the process is out of control.

(f) $\bar{x} = 95.8$

$r = 5.4$

The process can be declared to be out of control due to an increase in the mean value.

16.3.2 (a) The $\bar{X}$-chart has a center line at 12.02 and control limits at 11.27 and 12.78.

The R-chart has a center line at 1.314 and control limits at 0 and 2.779.

(b) Sample 8 lies above the upper control limits.

(c) If sample 8 is removed then the following modified control charts can be employed.

The $\bar{X}$-chart has a center line at 11.99 and control limits at 11.28 and 12.70.

The R-chart has a center line at 1.231 and control limits at 0 and 2.602.

16.3.3 (a) The $\bar{X}$-chart has a center line at 2.993 and control limits at 2.801 and 3.186.

The R-chart has a center line at 0.2642 and control limits at 0 and 0.6029.

(b) $\bar{x} = 2.97$

$r = 0.24$

There is no evidence that the process is out of control.

16.4 Attribute Control Charts

16.4.1 The p-chart has a center line at 0.0500 and control limits at 0.0000 and 0.1154.

(a) No

(b) In order for

$\frac{x}{100} \geq 0.1154$

it is necessary that $x \geq 12$.

16.4.2 (a) Samples 8 and 22 are above the upper control limit on the p-chart.

(b) If samples 8 and 22 are removed from the data set then a p-chart with a center line at 0.1400 and control limits at 0.0880 and 0.1920 is obtained.

(c) In order for

$\frac{x}{400} \geq 0.1920$

it is necessary that $x \geq 77$.

16.4.3 The c-chart has a center line at 12.42 and control limits at 1.85 and 22.99.

(a) No

(b) At least 23.

16.4.4 (a) The c-chart has a center line at 2.727 and control limits at 0 and 7.682. Samples 16 and 17 lie above the upper control limit.

(b) If samples 16 and 17 are removed then a c-chart with a center line at 2.150 and control limits at 0 and 6.549 is obtained.

(c) At least 7.

16.5 Acceptance Sampling

16.5.1 (a) With $p_0 = 0.06$ there would be 3 defective items in the batch of $N = 50$ items. The producer's risk is 0.0005.

(b) With $p_1 = 0.20$ there would be 10 defective items in the batch of $N = 50$ items. The consumer's risk is 0.952.

Using a binomial approximation these probabilities are estimated to be 0.002 and 0.942.

16.5.2 (a) With $p_0 = 0.10$ there would be 2 defective items in the batch of $N = 20$ items. The producer's risk is 0.016.

(b) With $p_1 = 0.20$ there would be 4 defective items in the batch of $N = 20$ items. The consumer's risk is 0.912.

Using a binomial approximation these probabilities are estimated to be 0.028 and 0.896.

16.5.3 (a) The producer's risk is 0.000.

(b) The consumer's risk is 0.300.

16.5.4 (a) The producer's risk is 0.000.

(b) The consumer's risk is 0.991.

(c) If $c = 9$ then the producer's risk is 0.000 and the consumer's risk is 0.976.

16.5.5 The smallest value of c for which

$P(B(30, 0.10) > c) \leq 0.05$

is $c = 6$.

16.6 Supplementary Problems

16.6.1 (a) The center line is 1250 and the control limits are 1214 and 1286.

(b) Yes
Yes

(c) $P(1214 \leq N(1240, 12^2) \leq 1286) = 0.9848$

The probability that an observation lies outside the control limits is therefore $1 - 0.9848 = 0.0152$.

The average run length for detecting the change is $\frac{1}{0.0152} = 66$.

16.6.2 (a) Sample 3 appears to have been out of control.

(b) If sample 3 is removed then the following modified control charts can be employed.

The $\bar{X}$-chart has a center line at 74.99 and control limits at 72.25 and 77.73.

The R-chart has a center line at 2.680 and control limits at 0 and 6.897.

(c) $\bar{x} = 74.01$

$r = 3.4$

There is no evidence that the process is out of control.

(d) $\bar{x} = 77.56$

$r = 3.21$

There is no evidence that the process is out of control.

16.6.3 (a) No

(b) The p-chart has a center line at 0.0205 and control limits at 0 and 0.0474.

(c) In order for

$\frac{x}{250} \geq 0.0474$

it is necessary that $x \geq 12$.

16.6.4 (a) Sample 13 lies above the center line of a c-chart.
If sample 13 is removed then a c-chart with a center line at 2.333 and control limits at 0 and 6.916 is obtained.

(b) At least seven flaws.

16.6.5 The smallest value of c for which

$P(B(50, 0.06) > c) \leq 0.025$

is $c = 7$.

The consumer's risk is 0.007.

Chapter 17

Reliability Analysis and Life Testing

17.1 System Reliability

17.1.1 $r = 0.9985$

17.1.2 $r = 0.9886$

17.1.3 (a) If r_1 is the individual reliability then in order for

$r_1^4 \geq 0.95$

it is necessary that $r_1 \geq 0.9873$.

(b) If r_1 is the individual reliability then in order for

$1 - (1 - r_1)^4 \geq 0.95$

it is necessary that $r_1 \geq 0.5271$.

(c) Suppose that n components with individual reliabilities r_1 are used, then an overall reliability of r is achieved as long as

$r_1 \geq r^{1/n}$

when the components are placed in series, and as long as

$r_1 \geq 1 - (1 - r)^{1/n}$

when the components are placed in parallel.

17.1.4 (a) The fourth component should be placed in parallel with the first component.

(b) In general, the fourth component (regardless of the value of r_4) should be placed in parallel with the component with the smallest reliability.

17.1.5 $r = 0.9017$

17.1.6 $r = 0.9507$

17.2 Modeling Failure Rates

17.2.1 The parameter is $\lambda = \frac{1}{225}$.

(a) $P(T \geq 250) = e^{-250/225} = 0.329$

(b) $P(T \leq 150) = 1 - e^{-150/225} = 0.487$

(c) $P(T \geq 100) = e^{-100/225} = 0.641$

If three components are placed in series then the system reliability is $0.641^3 = 0.264$.

17.2.2 The parameter is $\lambda = \frac{1}{35}$.

(a) $P(T \geq 35) = e^{-35/35} = 0.368$

(b) $P(T \leq 40) = 1 - e^{-40/35} = 0.681$

(c) $P(T \geq 5) = e^{-5/35} = 0.867$

If six components are placed in series then the system reliability is $0.867^6 = 0.424$.

17.2.3 $\frac{1}{\frac{1}{125}+\frac{1}{60}+\frac{1}{150}+\frac{1}{100}} = 24.2$ minutes

17.2.4 The failure time distribution is exponential with parameter $\lambda = 0.2$.

(a) $P(T \geq 4) = e^{-0.2\times 4} = 0.449$

(b) $P(T \leq 6) = 1 - e^{-0.2\times 6} = 0.699$

17.2.5 (a) $P(T \geq 40) = 1 - \Phi\left(\frac{\ln(40)-2.5}{1.5}\right) = 0.214$

(b) $P(T \leq 10) = \Phi\left(\frac{\ln(10)-2.5}{1.5}\right) = 0.448$

(c) $e^{2.5+1.5^2/2} = 37.5$

(d) Solving

$\Phi\left(\frac{\ln(t)-2.5}{1.5}\right) = 0.5$

gives $t = e^{2.5} = 12.2$.

17.2.6 (a) $P(T \geq 50) = 1 - \Phi\left(\frac{\ln(50)-3.0}{0.5}\right) = 0.034$

(b) $P(T \leq 40) = \Phi\left(\frac{\ln(40)-3.0}{0.5}\right) = 0.916$

(c) $e^{3.0+0.5^2/2} = 22.8$

(d) Solving

$\Phi\left(\frac{\ln(t)-3.0}{0.5}\right) = 0.5$

gives $t = e^{3.0} = 20.1$.

17.2.7 (a) $P(T \geq 5) = e^{-(0.25\times5)^{3.0}} = 0.142$

(b) $P(T \leq 3) = 1 - e^{-(0.25\times3)^{3.0}} = 0.344$

(c) Solving

$1 - e^{-(0.25\times t)^{3.0}} = 0.5$

gives $t = 3.54$.

(d) The hazard rate is

$h(t) = 3.0 \times 0.25^{3.0} \times t^{3.0-1} = 0.0469 \times t^2$.

(e) $\frac{h(5)}{h(3)} = 2.78$

17.2.8 (a) $P(T \geq 12) = e^{-(0.1\times12)^{4.5}} = 0.103$

(b) $P(T \leq 8) = 1 - e^{-(0.1\times8)^{4.5}} = 0.307$

(c) Solving

$1 - e^{-(0.1\times t)^{4.5}} = 0.5$

gives $t = 9.22$.

(d) The hazard rate is

$h(t) = 4.5 \times 0.1^{4.5} \times t^{4.5-1} = 0.0001423 \times t^{3.5}$.

(e) $\frac{h(12)}{h(8)} = 4.13$

17.3 Life Testing

17.3.1 (a) With $\chi^2_{60,0.005} = 91.952$ and $\chi^2_{60,0.995} = 35.534$ the confidence interval is

$\left(\frac{2\times30\times132.4}{91.952}, \frac{2\times30\times132.4}{35.534}\right) = (86.4, 223.6)$.

(b) The value 150 is within the confidence interval, so the claim is plausible.

17.3.2 (a) With $\chi^2_{40,0.025} = 59.342$ and $\chi^2_{40,0.975} = 24.433$, and with $\bar{t} = 12.145$, the confidence interval is

$\left(\frac{2\times20\times12.145}{59.342}, \frac{2\times20\times12.145}{24.433}\right) = (8.19, 19.88)$.

(b) The value 14 is within the confidence interval so it is a plausible value.

17.3.3 (a) With $\chi^2_{60,0.005} = 91.952$ and $\chi^2_{60,0.995} = 35.534$, and with $\bar{t} = 176.5/30 = 5.883$, the confidence interval is

$\left(\frac{2\times30\times5.883}{91.952}, \frac{2\times30\times5.883}{35.534}\right) = (3.84, 9.93)$.

(b) The value 10 is not included within the confidence interval, and so it is not plausible that the mean time to failure is 10 hours.

17.3.4 (a) The natural logarithms of the data values have a sample mean $\hat{\mu} = 2.007$ and a sample standard deviation $\hat{\sigma} = 0.3536$.

(b) $P(T \geq 10) = 1 - \Phi\left(\frac{\ln(10)-2.007}{0.3536}\right) = 0.202$

17.3.5 (a)

$$
\begin{aligned}
0 < t \leq 67 \quad &\Rightarrow \quad \hat{r}(t) = 1\\
67 < t \leq 72 \quad &\Rightarrow \quad \hat{r}(t) = 1 \times (27-1)/27 = 0.963\\
72 < t \leq 79 \quad &\Rightarrow \quad \hat{r}(t) = 0.963 \times (26-2)/26 = 0.889\\
79 < t \leq 81 \quad &\Rightarrow \quad \hat{r}(t) = 0.889 \times (24-1)/24 = 0.852\\
81 < t \leq 82 \quad &\Rightarrow \quad \hat{r}(t) = 0.852 \times (22-1)/22 = 0.813\\
82 < t \leq 89 \quad &\Rightarrow \quad \hat{r}(t) = 0.813 \times (21-1)/21 = 0.774\\
89 < t \leq 93 \quad &\Rightarrow \quad \hat{r}(t) = 0.774 \times (18-1)/18 = 0.731\\
93 < t \leq 95 \quad &\Rightarrow \quad \hat{r}(t) = 0.731 \times (17-1)/17 = 0.688\\
95 < t \leq 101 \quad &\Rightarrow \quad \hat{r}(t) = 0.688 \times (16-1)/16 = 0.645\\
101 < t \leq 104 \quad &\Rightarrow \quad \hat{r}(t) = 0.645 \times (15-1)/15 = 0.602\\
104 < t \leq 105 \quad &\Rightarrow \quad \hat{r}(t) = 0.602 \times (14-1)/14 = 0.559\\
105 < t \leq 109 \quad &\Rightarrow \quad \hat{r}(t) = 0.559 \times (13-1)/13 = 0.516\\
109 < t \leq 114 \quad &\Rightarrow \quad \hat{r}(t) = 0.516 \times (11-1)/11 = 0.469\\
114 < t \leq 122 \quad &\Rightarrow \quad \hat{r}(t) = 0.469 \times (9-2)/9 = 0.365\\
122 < t \leq 126 \quad &\Rightarrow \quad \hat{r}(t) = 0.365 \times (7-1)/7 = 0.313\\
126 < t \leq 135 \quad &\Rightarrow \quad \hat{r}(t) = 0.313 \times (6-2)/6 = 0.209\\
135 < t \leq 138 \quad &\Rightarrow \quad \hat{r}(t) = 0.209 \times (3-1)/3 = 0.139\\
138 < t \quad &\Rightarrow \quad \hat{r}(t) = 0.139 \times (2-2)/2 = 0.000
\end{aligned}
$$

(b) $\mathrm{Var}(\hat{r}(100)) = 0.645^2 \times$
$\left(\frac{1}{27(27-1)} + \frac{2}{26(26-2)} + \frac{1}{24(24-1)} + \frac{1}{22(22-1)}\right.$
$\left. + \frac{1}{21(21-1)} + \frac{1}{18(18-1)} + \frac{1}{17(17-1)} + \frac{1}{16(16-1)}\right)$
$= 0.0091931$

The confidence interval is

$(0.645 - 1.960 \times \sqrt{0.0091931}, 0.645 + 1.960 \times \sqrt{0.0091931}) = (0.457, 0.833).$

17.4 Supplementary Problems

17.4.1 (a) In order for

$1 - (1 - 0.90)^n \geq 0.995$

it is necessary that $n \geq 3$.

(b) In order for

$1 - (1 - r_i)^n \geq r$

it is necessary that $n \geq \frac{\ln(1-r)}{\ln(1-r_i)}$.

17.4.2 $r = 0.9890$

17.4.3 The failure time distribution is exponential with parameter $\lambda = 0.31$.

(a) $P(T \geq 6) = e^{-0.31\times 6} = 0.156$

(b) $P(T \leq 2) = 1 - e^{-0.31\times 2} = 0.462$

17.4.4 (a) $P(T \geq 120) = e^{-(0.01\times 120)^{2.5}} = 0.207$

(b) $P(T \leq 50) = 1 - e^{-(0.01\times 50)^{2.5}} = 0.162$

(c) Solving

$1 - e^{-(0.01\times t)^{2.5}} = 0.5$

gives $t = 86.4$ days.

(d) The hazard rate is

$h(t) = 2.5 \times 0.01^{2.5} \times t^{2.5-1} = 2.5 \times 10^{-5} \times t^{1.5}$.

(e) $\frac{h(120)}{h(100)} = 1.31$

17.4.5 (a) With $\chi^2_{50,0.025} = 71.420$ and $\chi^2_{50,0.975} = 32.357$, and with $\bar{t} = 141.2$, the confidence interval is

$\left(\frac{2\times 25\times 141.2}{71.420}, \frac{2\times 25\times 141.2}{32.357}\right) = (98.85, 218.19)$.

(b) The value $7 \times 24 = 168$ is within the confidence interval so it is a plausible value.

17.4.6 (a) The natural logarithms of the data values have a sample mean $\hat{\mu} = 2.5486$ and a sample standard deviation $\hat{\sigma} = 0.2133$.

(b) $P(T \geq 15) = 1 - \Phi\left(\frac{\ln(15) - 2.5486}{0.2133}\right) = 0.227$

17.4.7 (a)

$$
\begin{array}{rcl}
0 < t \leq 99 & \Rightarrow & \hat{r}(t) = 1 \\
99 < t \leq 123 & \Rightarrow & \hat{r}(t) = 1 \times (39-1)/39 = 0.974 \\
123 < t \leq 133 & \Rightarrow & \hat{r}(t) = 0.974 \times (38-1)/38 = 0.949 \\
133 < t \leq 142 & \Rightarrow & \hat{r}(t) = 0.949 \times (37-2)/37 = 0.897 \\
142 < t \leq 149 & \Rightarrow & \hat{r}(t) = 0.897 \times (35-1)/35 = 0.872 \\
149 < t \leq 154 & \Rightarrow & \hat{r}(t) = 0.872 \times (32-2)/32 = 0.817 \\
154 < t \leq 155 & \Rightarrow & \hat{r}(t) = 0.817 \times (30-1)/30 = 0.790 \\
155 < t \leq 168 & \Rightarrow & \hat{r}(t) = 0.790 \times (29-1)/29 = 0.763 \\
168 < t \leq 172 & \Rightarrow & \hat{r}(t) = 0.763 \times (27-2)/27 = 0.706 \\
172 < t \leq 176 & \Rightarrow & \hat{r}(t) = 0.706 \times (24-2)/24 = 0.647 \\
176 < t \leq 179 & \Rightarrow & \hat{r}(t) = 0.647 \times (21-4)/21 = 0.524 \\
179 < t \leq 181 & \Rightarrow & \hat{r}(t) = 0.524 \times (16-1)/16 = 0.491 \\
181 < t \leq 182 & \Rightarrow & \hat{r}(t) = 0.491 \times (15-1)/15 = 0.459 \\
182 < t \leq 184 & \Rightarrow & \hat{r}(t) = 0.459 \times (14-1)/14 = 0.426 \\
184 < t \leq 185 & \Rightarrow & \hat{r}(t) = 0.426 \times (13-2)/13 = 0.360 \\
185 < t \leq 186 & \Rightarrow & \hat{r}(t) = 0.360 \times (11-1)/11 = 0.328 \\
186 < t \leq 191 & \Rightarrow & \hat{r}(t) = 0.328 \times (10-1)/10 = 0.295 \\
191 < t \leq 193 & \Rightarrow & \hat{r}(t) = 0.295 \times (9-1)/9 = 0.262 \\
193 < t \leq 199 & \Rightarrow & \hat{r}(t) = 0.262 \times (8-1)/8 = 0.229 \\
199 < t \leq 207 & \Rightarrow & \hat{r}(t) = 0.229 \times (6-1)/6 = 0.191 \\
207 < t \leq 211 & \Rightarrow & \hat{r}(t) = 0.191 \times (4-1)/4 = 0.143 \\
211 < t \leq 214 & \Rightarrow & \hat{r}(t) = 0.143 \times (3-1)/3 = 0.096 \\
214 < t \leq 231 & \Rightarrow & \hat{r}(t) = 0.096 \times (2-1)/2 = 0.048 \\
231 < t & \Rightarrow & \hat{r}(t) = 0.048 \times (1-1)/1 = 0.000
\end{array}
$$

(b) $\mathrm{Var}(\hat{r}(200)) = 0.191^2 \times \left(\frac{1}{39(39-1)} + \frac{1}{38(38-1)} + \frac{2}{37(37-2)} + \frac{1}{35(35-1)} + \frac{2}{32(32-2)}\right.$
$+ \frac{1}{30(30-1)} + \frac{1}{29(29-1)} + \frac{2}{27(27-2)} + \frac{2}{24(24-2)} + \frac{4}{21(21-4)} + \frac{1}{16(16-1)} + \frac{1}{15(15-1)}$
$\left. + \frac{1}{14(14-1)} + \frac{2}{13(13-2)} + \frac{1}{11(11-1)} + \frac{1}{10(10-1)} + \frac{1}{9(9-1)} + \frac{1}{8(8-1)} + \frac{1}{6(6-1)}\right)$
$= 0.005103$

The confidence interval is

$(0.191 - 1.960 \times \sqrt{0.005103}, 0.191 + 1.960 \times \sqrt{0.005103}) = (0.051, 0.331)$.