INSTRUCTOR'S
SOLUTION MANUAL

KEYING YE AND SHARON MYERS

for

PROBABILITY & STATISTICS
FOR ENGINEERS & SCIENTISTS

EIGHTH EDITION

WALPOLE, MYERS, MYERS, YE

Contents

Chapter 1

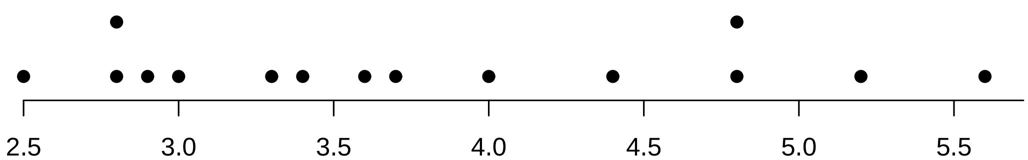
Introduction to Statistics and Data Analysis

1.1 (a) 15.

(b) $\bar{x} = \frac{1}{15}(3.4 + 2.5 + 4.8 + \cdots + 4.8) = 3.787.$

(c) Sample median is the 8th value, after the data is sorted from smallest to largest: 3.6.

(d) A dot plot is shown below.



(e) After trimming total 40% of the data (20% highest and 20% lowest), the data becomes:

2.9	3.0	3.3	3.4	3.6
3.7	4.0	4.4	4.8	

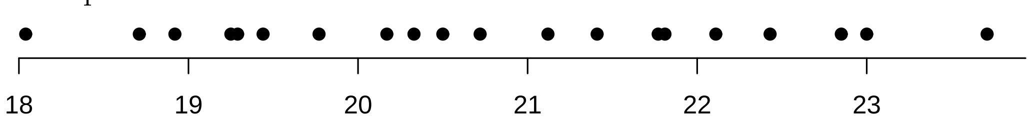
So. the trimmed mean is

$$\bar{x}_{\text{tr20}} = \frac{1}{9}(2.9 + 3.0 + \cdots + 4.8) = 3.678.$$

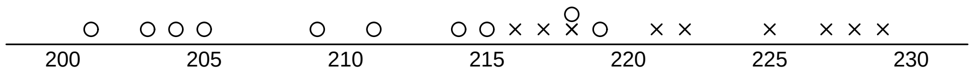
1.2 (a) Mean=20.768 and Median=20.610.

(b) $\bar{x}_{\text{tr10}} = 20.743.$

(c) A dot plot is shown below.



1.3 (a) A dot plot is shown below.



In the figure, "×" represents the "No aging" group and "∘" represents the "Aging" group.

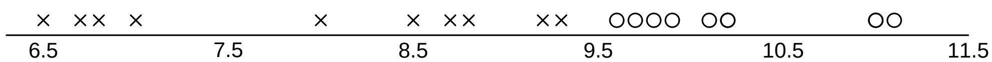
(b) Yes; tensile strength is greatly reduced due to the aging process.

(c) $\text{Mean}_{\text{Aging}} = 209.90$, and $\text{Mean}_{\text{No aging}} = 222.10$.

(d) $\text{Median}_{\text{Aging}} = 210.00$, and $\text{Median}_{\text{No aging}} = 221.50$. The means and medians for each group are similar to each other.

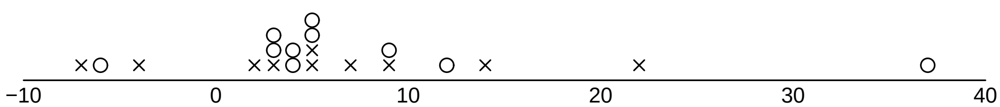
1.4 (a) $\bar{X}_A = 7.950$ and $\tilde{X}_A = 8.250$;
$\bar{X}_B = 10.260$ and $\tilde{X}_B = 10.150$.

(b) A dot plot is shown below.



In the figure, "×" represents company A and "∘" represents company B. The steel rods made by company B show more flexibility.

1.5 (a) A dot plot is shown below.

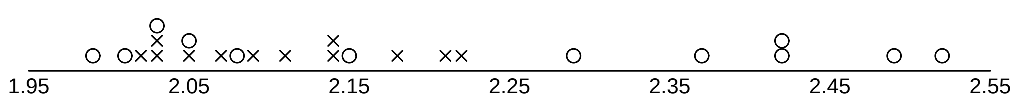


In the figure, "×" represents the control group and "∘" represents the treatment group.

(b) $\bar{X}_{\text{Control}} = 5.60$, $\tilde{X}_{\text{Control}} = 5.00$, and $\bar{X}_{\text{tr}(10);\text{Control}} = 5.13$;
$\bar{X}_{\text{Treatment}} = 7.60$, $\tilde{X}_{\text{Treatment}} = 4.50$, and $\bar{X}_{\text{tr}(10);\text{Treatment}} = 5.63$.

(c) The difference of the means is 2.0 and the differences of the medians and the trimmed means are 0.5, which are much smaller. The possible cause of this might be due to the extreme values (outliers) in the samples, especially the value of 37.

1.6 (a) A dot plot is shown below.



In the figure, "×" represents the 20°C group and "∘" represents the 45°C group.

(b) $\bar{X}_{20^\circ\text{C}} = 2.1075$, and $\bar{X}_{45^\circ\text{C}} = 2.2350$.

(c) Based on the plot, it seems that high temperature yields more high values of tensile strength, along with a few low values of tensile strength. Overall, the temperature does have an influence on the tensile strength.

(d) It also seems that the variation of the tensile strength gets larger when the cure temperature is increased.

1.7 $s^2 = \frac{1}{15-1}[(3.4-3.787)^2+(2.5-3.787)^2+(4.8-3.787)^2+\cdots+(4.8-3.787)^2] = 0.94284$;
$s = \sqrt{s^2} = \sqrt{0.9428} = 0.971$.

1.8 $s^2 = \frac{1}{20-1}[(18.71-20.768)^2 + (21.41-20.768)^2 + \cdots + (21.12-20.768)^2] = 2.5345$;
$s = \sqrt{2.5345} = 1.592$.

1.9 $s^2_{\text{No Aging}} = \frac{1}{10-1}[(227-222.10)^2 + (222-222.10)^2 + \cdots + (221-222.10)^2] = 42.12$;
$s_{\text{No Aging}} = \sqrt{42.12} = 6.49$.
$s^2_{\text{Aging}} = \frac{1}{10-1}[(219-209.90)^2 + (214-209.90)^2 + \cdots + (205-209.90)^2] = 23.62$;
$s_{\text{Aging}} = \sqrt{23.62} = 4.86$.

1.10 For company A: $s^2_A = 1.2078$ and $s_A = \sqrt{1.2078} = 1.099$.
For company B: $s^2_B = 0.3249$ and $s_B = \sqrt{0.3249} = 0.570$.

1.11 For the control group: $s^2_{\text{Control}} = 69.39$ and $s_{\text{Control}} = 8.33$.
For the treatment group: $s^2_{\text{Treatment}} = 128.14$ and $s_{\text{Treatment}} = 11.32$.

1.12 For the cure temperature at 20°C: $s^2_{20^\circ\text{C}} = 0.005$ and $s_{20^\circ\text{C}} = 0.071$.
For the cure temperature at 45°C: $s^2_{45^\circ\text{C}} = 0.0413$ and $s_{45^\circ\text{C}} = 0.2032$.
The variation of the tensile strength is influenced by the increase of cure temperature.

1.13 (a) Mean $= \bar{X} = 124.3$ and median $= \tilde{X} = 120$;

(b) 175 is an extreme observation.

1.14 (a) Mean $= \bar{X} = 570.5$ and median $= \tilde{X} = 571$;

(b) Variance $= s^2 = 10$; standard deviation$= s = 3.162$; range$=10$;

(c) Variation of the diameters seems too big.

1.15 Yes. The value 0.03125 is actually a P-value and a small value of this quantity means that the outcome (i.e., $HHHHH$) is very unlikely to happen with a fair coin.

1.16 The term on the left side can be manipulated to

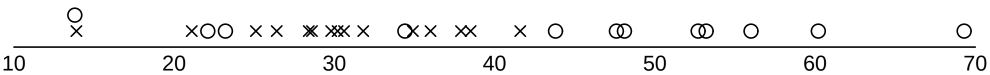
$$\sum_{i=1}^{n} x_i - n\bar{x} = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0,$$

which is the term on the right side.

1.17 (a) $\bar{X}_{\text{smokers}} = 43.70$ and $\bar{X}_{\text{nonsmokers}} = 30.32$;

(b) $s_{\text{smokers}} = 16.93$ and $s_{\text{nonsmokers}} = 7.13$;

(c) A dot plot is shown below.



In the figure, "×" represents the nonsmoker group and "○" represents the smoker group.

(d) Smokers appear to take longer time to fall asleep and the time to fall asleep for smoker group is more variable.

1.18 (a) A stem-and-leaf plot is shown below.

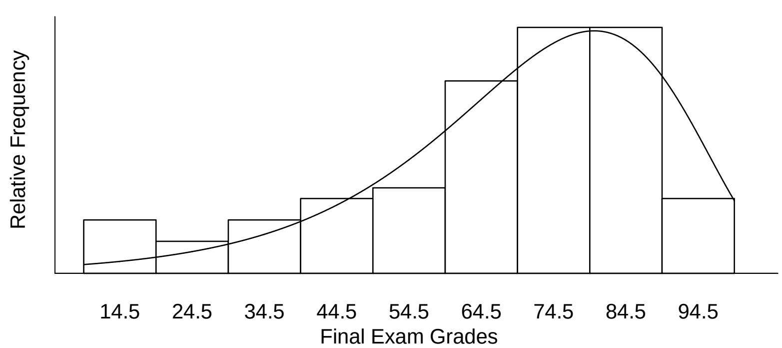
Stem	Leaf	Frequency
1	057	3
2	35	2
3	246	3
4	1138	4
5	22457	5
6	00123445779	11
7	01244456678899	14
8	00011223445589	14
9	0258	4

(b) The following is the relative frequency distribution table.

Relative Frequency Distribution of Grades

Class Interval	Class Midpoint	Frequency, f	Relative Frequency
10 − 19	14.5	3	0.05
20 − 29	24.5	2	0.03
30 − 39	34.5	3	0.05
40 − 49	44.5	4	0.07
50 − 59	54.5	5	0.08
60 − 69	64.5	11	0.18
70 − 79	74.5	14	0.23
80 − 89	84.5	14	0.23
90 − 99	94.5	4	0.07

(c) A histogram plot is given below.



The distribution skews to the left.

(d) $\bar{X} = 65.48$, $\tilde{X} = 71.50$ and $s = 21.13$.

1.19 (a) A stem-and-leaf plot is shown below.

Stem	**Leaf**	**Frequency**
0	22233457	8
1	023558	6
2	035	3
3	03	2
4	057	3
5	0569	4
6	0005	4

(b) The following is the relative frequency distribution table.

Relative Frequency Distribution of Years

Class Interval	**Class Midpoint**	**Frequency, f**	**Relative Frequency**
$0.0 - 0.9$	0.45	8	0.267
$1.0 - 1.9$	1.45	6	0.200
$2.0 - 2.9$	2.45	3	0.100
$3.0 - 3.9$	3.45	2	0.067
$4.0 - 4.9$	4.45	3	0.100
$5.0 - 5.9$	5.45	4	0.133
$6.0 - 6.9$	6.45	4	0.133

(c) $\bar{X} = 2.797$, $s = 2.227$ and Sample range is $6.5 - 0.2 = 6.3$.

1.20 (a) A stem-and-leaf plot is shown next.

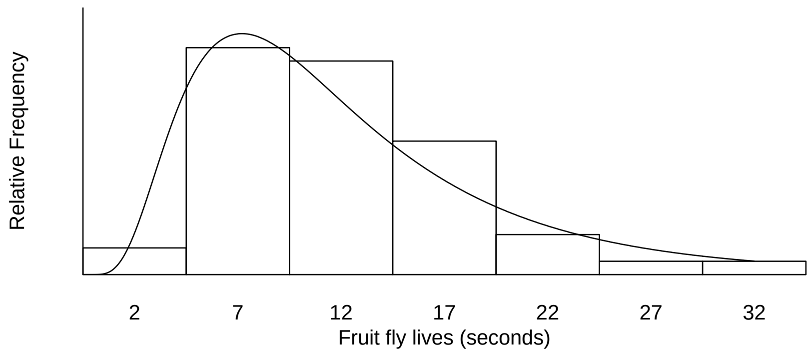
Stem	**Leaf**	**Frequency**
0*	34	2
0	56667777777889999	17
1*	0000001223333344	16
1	5566788899	10
2*	034	3
2	7	1
3*	2	1

(b) The relative frequency distribution table is shown next.

Relative Frequency Distribution of Fruit Fly Lives

Class Interval	**Class Midpoint**	**Frequency, f**	**Relative Frequency**
$0-4$	2	2	0.04
$5-9$	7	17	0.34
$10-14$	12	16	0.32
$15-19$	17	10	0.20
$20-24$	22	3	0.06
$25-29$	27	1	0.02
$30-34$	32	1	0.02

(c) A histogram plot is shown next.



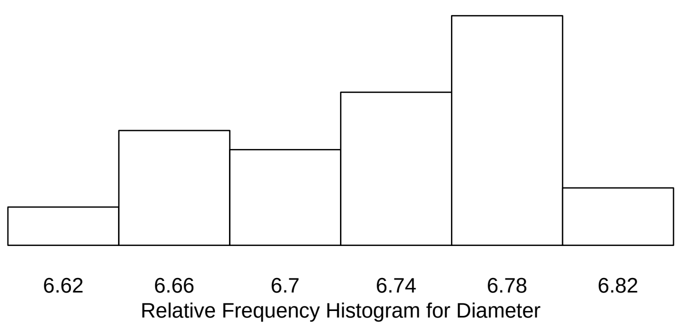
(d) $\tilde{X} = 10.50$.

1.21 (a) $\bar{X} = 1.7743$ and $\tilde{X} = 1.7700$;

(b) $s = 0.3905$.

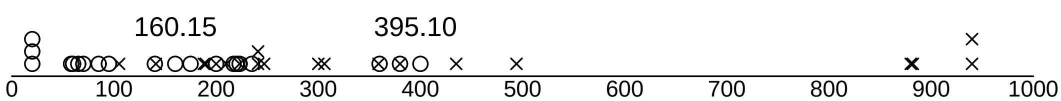
1.22 (a) $\bar{X} = 6.7261$ and $\tilde{X} = 0.0536$.

(b) A histogram plot is shown next.



(c) The data appear to be skewed to the left.

1.23 (a) A dot plot is shown next.

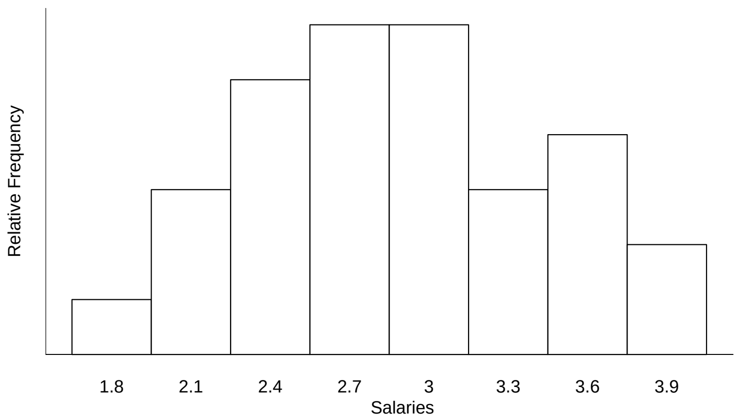


(b) $\bar{X}_{1980} = 395.1$ and $\bar{X}_{1990} = 160.2$.

(c) The sample mean for 1980 is over twice as large as that of 1990. The variability for 1990 decreased also as seen by looking at the picture in (a). The gap represents an increase of over 400 ppm. It appears from the data that hydrocarbon emissions decreased considerably between 1980 and 1990 and that the extreme large emission (over 500 ppm) were no longer in evidence.

1.24 (a) $\bar{X} = 2.8973$ and $s = 0.5415$.

(b) A histogram plot is shown next.



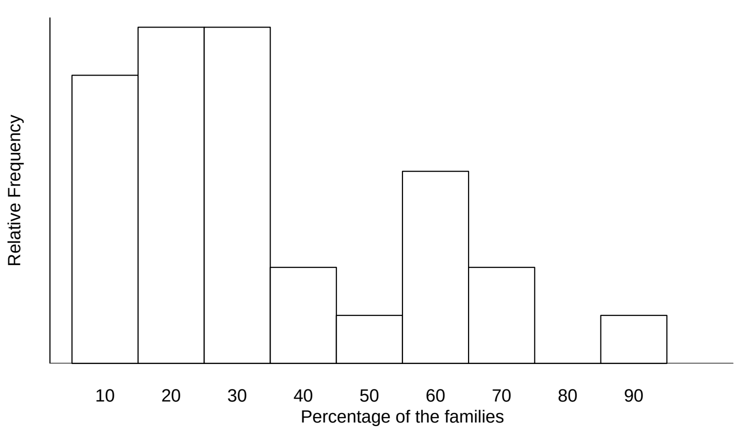
(c) Use the double-stem-and-leaf plot, we have the following.

Stem	**Leaf**	**Frequency**
1	(84)	1
2*	(05)(10)(14)(37)(44)(45)	6
2	(52)(52)(67)(68)(71)(75)(77)(83)(89)(91)(99)	11
3*	(10)(13)(14)(22)(36)(37)	6
3	(51)(54)(57)(71)(79)(85)	6

1.25 (a) $\bar{X} = 33.31$;

(b) $\tilde{X} = 26.35$;

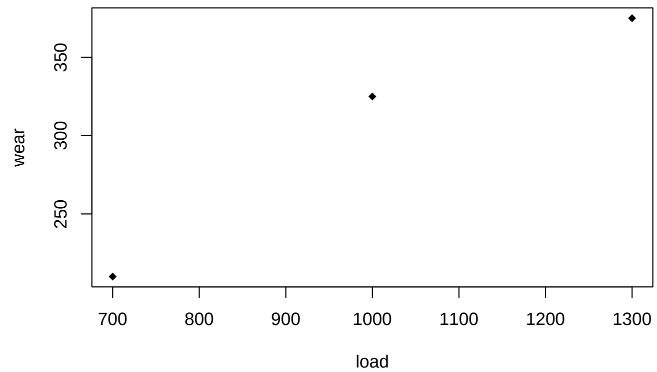
(c) A histogram plot is shown next.



(d) $\bar{X}_{\text{tr}(10)} = 30.97$. This trimmed mean is in the middle of the mean and median using the full amount of data. Due to the skewness of the data to the right (see plot in (c)), it is common to use trimmed data to have a more robust result.

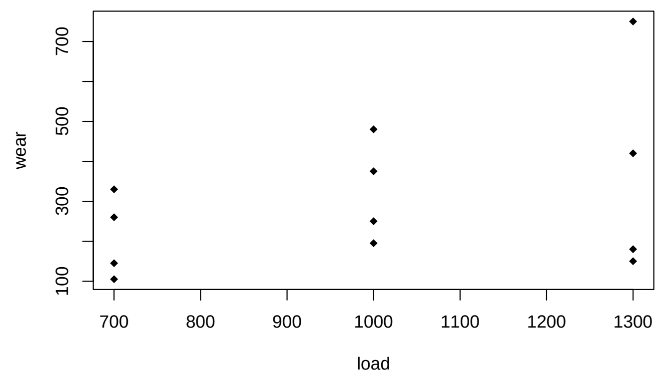
1.26 If a model using the function of percent of families to predict staff salaries, it is likely that the model would be wrong due to several extreme values of the data. Actually if a scatter plot of these two data sets is made, it is easy to see that some outlier would influence the trend.

1.27 (a) The averages of the wear are plotted here.



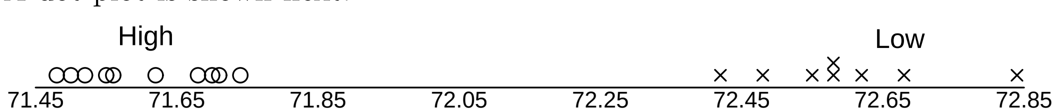
(b) When the load value increases, the wear value also increases. It does show certain relationship.

(c) A plot of wears is shown next.



(d) The relationship between load and wear in (c) is not as strong as the case in (a), especially for the load at 1300. One reason is that there is an extreme value (750) which influence the mean value at the load 1300.

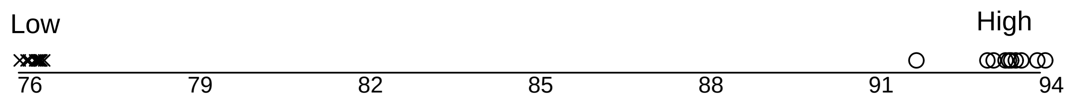
1.28 (a) A dot plot is shown next.



In the figure, "×" represents the low-injection-velocity group and "∘" represents the high-injection-velocity group.

(b) It appears that shrinkage values for the low-injection-velocity group is higher than those for the high-injection-velocity group. Also, the variation of the shrinkage is a little larger for the low injection velocity than that for the high injection velocity.

1.29 (a) A dot plot is shown next.

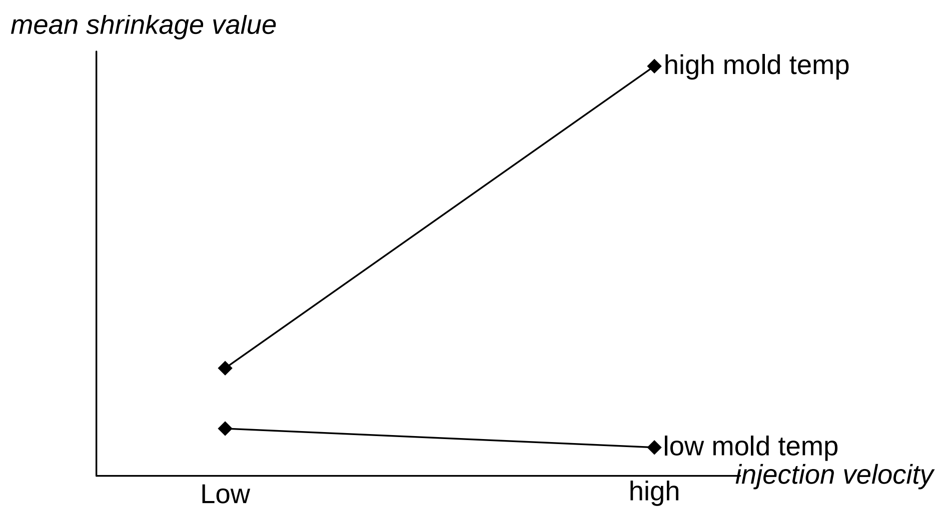


In the figure, "$\times$" represents the low-injection-velocity group and "$\circ$" represents the high-injection-velocity group.

(b) In this time, the shrinkage values are much higher for the high-injection-velocity group than those for the low-injection-velocity group. Also, the variation for the former group is much higher as well.

(c) Since the shrinkage effects change in different direction between low mode temperature and high mold temperature, the apparent interactions between the mold temperature and injection velocity are significant.

1.30 An interaction plot is shown next.



It is quite obvious to find the interaction between the two variables. Since in this experimental data, those two variables can be controlled each at two levels, the interaction can be investigated. However, if the data are from an observational studies, in which the variable values cannot be controlled, it would be difficult to study the interactions among these variables.

Chapter 2

Probability

2.1 (a) $S = \{8, 16, 24, 32, 40, 48\}$.

(b) For $x^2 + 4x - 5 = (x + 5)(x - 1) = 0$, the only solutions are $x = -5$ and $x = 1$. $S = \{-5, 1\}$.

(c) $S = \{T, HT, HHT, HHH\}$.

(d) $S = \{\text{N. America}, \text{S. America}, \text{Europe}, \text{Asia}, \text{Africa}, \text{Australia}, \text{Antarctica}\}$.

(e) Solving $2x - 4 \geq 0$ gives $x \geq 2$. Since we must also have $x < 1$, it follows that $S = \phi$.

2.2 $S = \{(x, y) \mid x^2 + y^2 < 9;\ x \geq 0,\ y \geq 0\}$.

2.3 (a) $A = \{1, 3\}$.

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

(c) $C = \{x \mid x^2 - 4x + 3 = 0\} = \{x \mid (x - 1)(x - 3) = 0\} = \{1, 3\}$.

(d) $D = \{0, 1, 2, 3, 4, 5, 6\}$. Clearly, $A = C$.

2.4 (a) $S = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}$.

(b) $S = \{(x, y) \mid 1 \leq x, y \leq 6\}$.

2.5 $S = \{1HH, 1HT, 1TH, 1TT, 2H, 2T, 3HH, 3HT, 3TH, 3TT, 4H, 4T, 5HH, 5HT, 5TH, 5TT, 6H, 6T\}$.

2.6 $S = \{A_1A_2, A_1A_3, A_1A_4, A_2A_3, A_2A_4, A_3A_4\}$.

2.7 $S_1 = \{MMMM, MMMF, MMFM, MFMM, FMMM, MMFF, MFMF, MFFM, FMFM, FFMM, FMMF, MFFF, FMFF, FFMF, FFFM, FFFF\}$.
$S_2 = \{0, 1, 2, 3, 4\}$.

2.8 (a) $A = \{(3, 6), (4, 5), (4, 6), (5, 4), (5, 5), (5, 6), (6, 3), (6, 4), (6, 5), (6, 6)\}$.

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

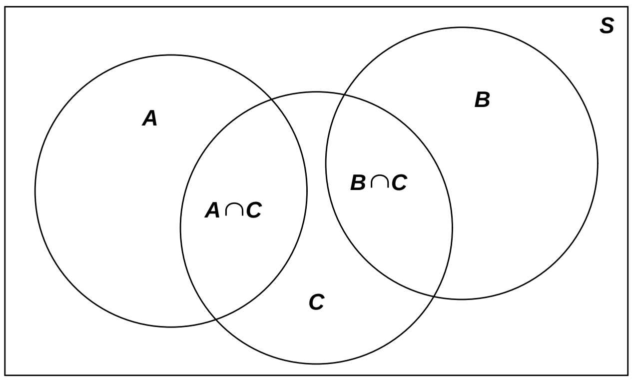
(c) $C = \{(5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}$.

(d) $A \cap C = \{(5, 4), (5, 5), (5, 6), (6, 3), (6, 4), (6, 5), (6, 6)\}$.

(e) $A \cap B = \phi$.

(f) $B \cap C = \{(5, 2), (6, 2)\}$.

(g) A Venn diagram is shown next.



2.9 (a) $A = \{1HH, 1HT, 1TH, 1TT, 2H, 2T\}$.

(b) $B = \{1TT, 3TT, 5TT\}$.

(c) $A' = \{3HH, 3HT, 3TH, 3TT, 4H, 4T, 5HH, 5HT, 5TH, 5TT, 6H, 6T\}$.

(d) $A' \cap B = \{3TT, 5TT\}$.

(e) $A \cup B = \{1HH, 1HT, 1TH, 1TT, 2H, 2T, 3TT, 5TT\}$.

2.10 (a) $S = \{FFF, FFN, FNF, NFF, FNN, NFN, NNF, NNN\}$.

(b) $E = \{FFF, FFN, FNF, NFF\}$.

(c) The second river was safe for fishing.

2.11 (a) $S = \{M_1M_2, M_1F_1, M_1F_2, M_2M_1, M_2F_1, M_2F_2, F_1M_1, F_1M_2, F_1F_2, F_2M_1, F_2M_2, F_2F_1\}$.

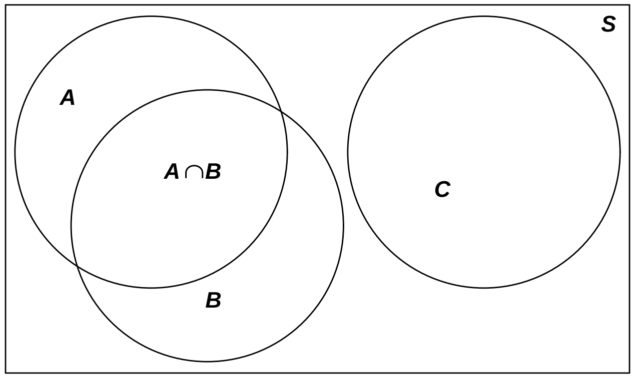
(b) $A = \{M_1M_2, M_1F_1, M_1F_2, M_2M_1, M_2F_1, M_2F_2\}$.

(c) $B = \{M_1F_1, M_1F_2, M_2F_1, M_2F_2, F_1M_1, F_1M_2, F_2M_1, F_2M_2\}$.

(d) $C = \{F_1F_2, F_2F_1\}$.

(e) $A \cap B = \{M_1F_1, M_1F_2, M_2F_1, M_2F_2\}$.

(f) $A \cup C = \{M_1M_2, M_1F_1, M_1F_2, M_2M_1, M_2F_1, M_2F_2, F_1F_2, F_2F_1\}$.

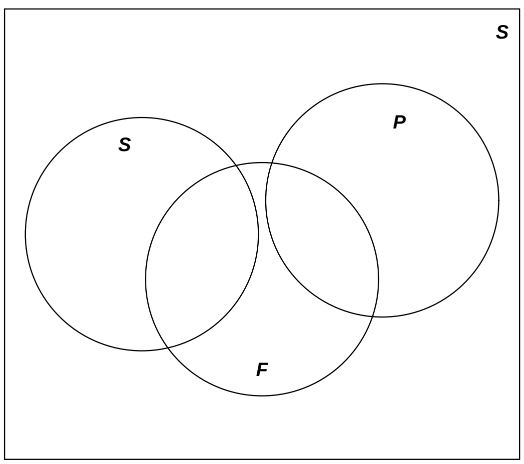
(g) 

2.12 (a) $S = \{ZYF, ZNF, WYF, WNF, SYF, SNF, ZYM\}$.

(b) $A \cup B = \{ZYF, ZNF, WYF, WNF, SYF, SNF\} = A$.

(c) $A \cap B = \{WYF, SYF\}$.

2.13 A Venn diagram is shown next.



2.14 (a) $A \cup C = \{0, 2, 3, 4, 5, 6, 8\}$.

(b) $A \cap B = \phi$.

(c) $C' = \{0, 1, 6, 7, 8, 9\}$.

(d) $C' \cap D = \{1, 6, 7\}$, so $(C' \cap D) \cup B = \{1, 3, 5, 6, 7, 9\}$.

(e) $(S \cap C)' = C' = \{0, 1, 6, 7, 8, 9\}$.

(f) $A \cap C = \{2, 4\}$, so $A \cap C \cap D' = \{2, 4\}$.

2.15 (a) $A' = \{\text{nitrogen}, \text{potassium}, \text{uranium}, \text{oxygen}\}$.

(b) $A \cup C = \{\text{copper}, \text{sodium}, \text{zinc}, \text{oxygen}\}$.

(c) $A \cap B' = \{\text{copper}, \text{zinc}\}$ and
$C' = \{\text{copper}, \text{sodium}, \text{nitrogen}, \text{potassium}, \text{uranium}, \text{zinc}\}$;
so $(A \cap B') \cup C' = \{\text{copper}, \text{sodium}, \text{nitrogen}, \text{potassium}, \text{uranium}, \text{zinc}\}$.

(d) $B' \cap C' = \{\text{copper}, \text{uranium}, \text{zinc}\}$.

(e) $A \cap B \cap C = \phi$.

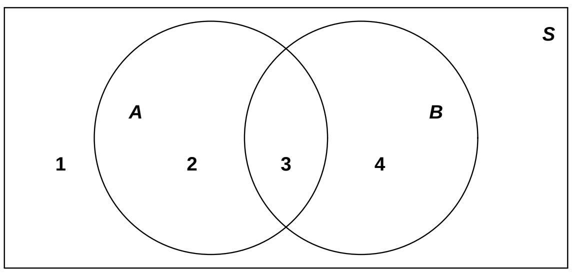
(f) $A' \cup B' = \{\text{copper}, \text{nitrogen}, \text{potassium}, \text{uranium}, \text{oxygen}, \text{zinc}\}$ and $A' \cap C = \{\text{oxygen}\}$; so, $(A' \cup B') \cap (A' \cap C) = \{\text{oxygen}\}$.

2.16 (a) $M \cup N = \{x \mid 0 < x < 9\}$.

(b) $M \cap N = \{x \mid 1 < x < 5\}$.

(c) $M' \cap N' = \{x \mid 9 < x < 12\}$.

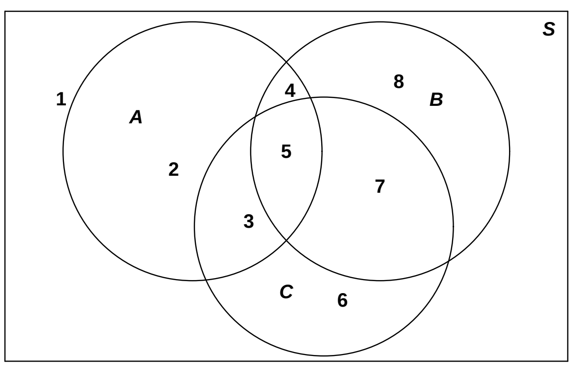
2.17 A Venn diagram is shown next.



(a) From the above Venn diagram, $(A \cap B)'$ contains the regions of 1, 2 and 4.

(b) $(A \cup B)'$ contains region 1.

(c) A Venn diagram is shown next.



$(A \cap C) \cup B$ contains the regions of 3, 4, 5, 7 and 8.

2.18 (a) Not mutually exclusive.

(b) Mutually exclusive.

(c) Not mutually exclusive.

(d) Mutually exclusive.

2.19 (a) The family will experience mechanical problems but will receive no ticket for traffic violation and will not arrive at a campsite that has no vacancies.

(b) The family will receive a traffic ticket and arrive at a campsite that has no vacancies but will not experience mechanical problems.

(c) The family will experience mechanical problems and will arrive at a campsite that has no vacancies.

(d) The family will receive a traffic ticket but will not arrive at a campsite that has no vacancies.

(e) The family will not experience mechanical problems.

2.20 (a) 6;

(b) 2;

(c) 2, 5, 6;

(d) 4, 5, 6, 8.

2.21 With $n_1 = 6$ sightseeing tours each available on $n_2 = 3$ different days, the multiplication rule gives $n_1 n_2 = (6)(3) = 18$ ways for a person to arrange a tour.

2.22 With $n_1 = 8$ blood types and $n_2 = 3$ classifications of blood pressure, the multiplication rule gives $n_1 n_2 = (8)(3) = 24$ classifications.

2.23 Since the die can land in $n_1 = 6$ ways and a letter can be selected in $n_2 = 26$ ways, the multiplication rule gives $n_1 n_2 = (6)(26) = 156$ points in S.

2.24 Since a student may be classified according to $n_1 = 4$ class standing and $n_2 = 2$ gender classifications, the multiplication rule gives $n_1 n_2 = (4)(2) = 8$ possible classifications for the students.

2.25 With $n_1 = 5$ different shoe styles in $n_2 = 4$ different colors, the multiplication rule gives $n_1 n_2 = (5)(4) = 20$ different pairs of shoes.

2.26 Using Theorem 2.8, we obtain the followings.

(a) There are $\binom{7}{5} = 21$ ways.

(b) There are $\binom{5}{3} = 10$ ways.

2.27 Using the generalized multiplication rule, there are $n_1 \times n_2 \times n_3 \times n_4 = (4)(3)(2)(2) = 48$ different house plans available.

2.28 With $n_1 = 5$ different manufacturers, $n_2 = 3$ different preparations, and $n_3 = 2$ different strengths, the generalized multiplication rule yields $n_1 n_2 n_3 = (5)(3)(2) = 30$ different ways to prescribe a drug for asthma.

2.29 With $n_1 = 3$ race cars, $n_2 = 5$ brands of gasoline, $n_3 = 7$ test sites, and $n_4 = 2$ drivers, the generalized multiplication rule yields $(3)(5)(7)(2) = 210$ test runs.

2.30 With $n_1 = 2$ choices for the first question, $n_2 = 2$ choices for the second question, and so forth, the generalized multiplication rule yields $n_1 n_2 \cdots n_9 = 2^9 = 512$ ways to answer the test.

2.31 (a) With $n_1 = 4$ possible answers for the first question, $n_2 = 4$ possible answers for the second question, and so forth, the generalized multiplication rule yields $4^5 = 1024$ ways to answer the test.

(b) With $n_1 = 3$ wrong answers for the first question, $n_2 = 3$ wrong answers for the second question, and so forth, the generalized multiplication rule yields

$$n_1 n_2 n_3 n_4 n_5 = (3)(3)(3)(3)(3) = 3^5 = 243$$

ways to answer the test and get all questions wrong.

2.32 (a) By Theorem 2.3, $7! = 5040$.

(b) Since the first letter must be m, the remaining 6 letters can be arranged in $6! = 720$ ways.

2.33 Since the first digit is a 5, there are $n_1 = 9$ possibilities for the second digit and then $n_2 = 8$ possibilities for the third digit. Therefore, by the multiplication rule there are $n_1 n_2 = (9)(8) = 72$ registrations to be checked.

2.34 (a) By Theorem 2.3, there are $6! = 720$ ways.

(b) A certain 3 persons can follow each other in a line of 6 people in a specified order is 4 ways or in $(4)(3!) = 24$ ways with regard to order. The other 3 persons can then be placed in line in $3! = 6$ ways. By Theorem 2.1, there are total $(24)(6) = 144$ ways to line up 6 people with a certain 3 following each other.

(c) Similar as in (b), the number of ways that a specified 2 persons can follow each other in a line of 6 people is $(5)(2!)(4!) = 240$ ways. Therefore, there are $720 - 240 = 480$ ways if a certain 2 persons refuse to follow each other.

2.35 The first house can be placed on any of the $n_1 = 9$ lots, the second house on any of the remaining $n_2 = 8$ lots, and so forth. Therefore, there are $9! = 362,880$ ways to place the 9 homes on the 9 lots.

2.36 (a) Any of the 6 nonzero digits can be chosen for the hundreds position, and of the remaining 6 digits for the tens position, leaving 5 digits for the units position. So, there are $(6)(5)(5) = 150$ three digit numbers.

(b) The units position can be filled using any of the 3 odd digits. Any of the remaining 5 nonzero digits can be chosen for the hundreds position, leaving a choice of 5 digits for the tens position. By Theorem 2.2, there are $(3)(5)(5) = 75$ three digit odd numbers.

(c) If a 4, 5, or 6 is used in the hundreds position there remain 6 and 5 choices, respectively, for the tens and units positions. This gives $(3)(6)(5) = 90$ three digit numbers beginning with a 4, 5, or 6. If a 3 is used in the hundreds position, then a 4, 5, or 6 must be used in the tens position leaving 5 choices for the units position. In this case, there are $(1)(3)(5) = 15$ three digit number begin with a 3. So, the total number of three digit numbers that are greater than 330 is $90 + 15 = 105$.

2.37 The first seat must be filled by any of 5 girls and the second seat by any of 4 boys. Continuing in this manner, the total number of ways to seat the 5 girls and 4 boys is $(5)(4)(4)(3)(3)(2)(2)(1)(1) = 2880$.

2.38 (a) $8! = 40320$.

(b) There are 4! ways to seat 4 couples and then each member of a couple can be interchanged resulting in $2^4(4!) = 384$ ways.

(c) By Theorem 2.3, the members of each gender can be seated in 4! ways. Then using Theorem 2.1, both men and women can be seated in $(4!)(4!) = 576$ ways.

2.39 (a) Any of the $n_1 = 8$ finalists may come in first, and of the $n_2 = 7$ remaining finalists can then come in second, and so forth. By Theorem 2.3, there $8! = 40320$ possible orders in which 8 finalists may finish the spelling bee.

(b) The possible orders for the first three positions are ${}_8P_3 = \frac{8!}{5!} = 336$.

2.40 By Theorem 2.4, ${}_8P_5 = \frac{8!}{3!} = 6720$.

2.41 By Theorem 2.4, ${}_6P_4 = \frac{6!}{2!} = 360$.

2.42 By Theorem 2.4, ${}_{40}P_3 = \frac{40!}{37!} = 59,280$.

2.43 By Theorem 2.5, there are $4! = 24$ ways.

2.44 By Theorem 2.5, there are $7! = 5040$ arrangements.

2.45 By Theorem 2.6, there are $\frac{8!}{3!2!} = 3360$.

2.46 By Theorem 2.6, there are $\frac{9!}{3!4!2!} = 1260$ ways.

2.47 By Theorem 2.7, there are $\binom{12}{7,3,2} = 7920$ ways.

2.48 $\binom{9}{1,4,4} + \binom{9}{2,4,3} + \binom{9}{1,3,5} + \binom{9}{2,3,4} + \binom{9}{2,2,5} = 4410$.

2.49 By Theorem 2.8, there are $\binom{8}{3} = 56$ ways.

2.50 Assume February 29th as March 1st for the leap year. There are total 365 days in a year. The number of ways that all these 60 students will have different birth dates (i.e, arranging 60 from 365) is ${}_{365}P_{60}$. This is a very large number.

2.51 (a) Sum of the probabilities exceeds 1.

(b) Sum of the probabilities is less than 1.

(c) A negative probability.

(d) Probability of both a heart and a black card is zero.

2.52 Assuming equal weights

(a) $P(A) = \frac{5}{18}$;

(b) $P(C) = \frac{1}{3}$;

(c) $P(A \cap C) = \frac{7}{36}$.

2.53 $S = \{\$10, \$25, \$100\}$ with weights $275/500 = 11/20$, $150/500 = 3/10$, and $75/500 = 3/20$, respectively. The probability that the first envelope purchased contains less than \$100 is equal to $11/20 + 3/10 = 17/20$.

2.54 (a) $P(S \cap D') = 88/500 = 22/125$.

(b) $P(E \cap D \cap S') = 31/500$.

(c) $P(S' \cap E') = 171/500$.

2.55 Consider the events
S: industry will locate in Shanghai,
B: industry will locate in Beijing.

(a) $P(S \cap B) = P(S) + P(B) - P(S \cup B) = 0.7 + 0.4 - 0.8 = 0.3$.

(b) $P(S' \cap B') = 1 - P(S \cup B) = 1 - 0.8 = 0.2$.

2.56 Consider the events
B: customer invests in tax-free bonds,
M: customer invests in mutual funds.

(a) $P(B \cup M) = P(B) + P(M) - P(B \cap M) = 0.6 + 0.3 - 0.15 = 0.75$.

(b) $P(B' \cap M') = 1 - P(B \cup M) = 1 - 0.75 = 0.25$.

2.57 (a) Since 5 of the 26 letters are vowels, we get a probability of 5/26.

(b) Since 9 of the 26 letters precede j, we get a probability of 9/26.

(c) Since 19 of the 26 letters follow g, we get a probability of 19/26.

2.58 (a) Let A = Defect in brake system; B = Defect in fuel system; $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.25 + 0.17 - 0.15 = 0.27$.

(b) $P(\text{No defect}) = 1 - P(A \cup B) = 1 - 0.27 = 0.73$.

2.59 By Theorem 2.2, there are $N = (26)(25)(24)(9)(8)(7)(6) = 47,174,400$ possible ways to code the items of which $n = (5)(25)(24)(8)(7)(6)(4) = 4,032,000$ begin with a vowel and end with an even digit. Therefore, $\frac{n}{N} = \frac{10}{117}$.

2.60 (a) Of the $(6)(6) = 36$ elements in the sample space, only 5 elements (2,6), (3,5), (4,4), (5,3), and (6,2) add to 8. Hence the probability of obtaining a total of 8 is then 5/36.

(b) Ten of the 36 elements total at most 5. Hence the probability of obtaining a total of at most is 10/36=5/18.

2.61 Since there are 20 cards greater than 2 and less than 8, the probability of selecting two of these in succession is

$$\left(\frac{20}{52}\right)\left(\frac{19}{51}\right) = \frac{95}{663}.$$

2.62 (a) $\frac{\binom{1}{1}\binom{8}{2}}{\binom{9}{3}} = \frac{1}{3}$.

(b) $\frac{\binom{5}{2}\binom{3}{1}}{\binom{9}{3}} = \frac{5}{14}$.

2.63 (a) $\frac{\binom{4}{3}\binom{48}{2}}{\binom{52}{5}} = \frac{94}{54145}$.

(b) $\frac{\binom{13}{4}\binom{13}{1}}{\binom{52}{5}} = \frac{143}{39984}$.

2.64 Any four of a kind, say four 2's and one 5 occur in $\binom{5}{1} = 5$ ways each with probability $(1/6)(1/6)(1/6)(1/6)(1/6) = (1/6)^5$. Since there are ${}_6P_2 = 30$ ways to choose various pairs of numbers to constitute four of one kind and one of the other (we use permutation instead of combination is because that four 2's and one 5, and four 5's and one 2 are two different ways), the probability is $(5)(30)(1/6)^5 = 25/1296$.

2.65 (a) $P(M \cup H) = 88/100 = 22/25$;

(b) $P(M' \cap H') = 12/100 = 3/25$;

(c) $P(H \cap M') = 34/100 = 17/50$.

2.66 (a) 9;

(b) 1/9.

2.67 (a) 0.32;

(b) 0.68;

(c) office or den.

2.68 (a) $1 - 0.42 = 0.58$;

(b) $1 - 0.04 = 0.96$.

2.69 $P(A) = 0.2$ and $P(B) = 0.35$

(a) $P(A') = 1 - 0.2 = 0.8$;

(b) $P(A' \cap B') = 1 - P(A \cup B) = 1 - 0.2 - 0.35 = 0.45$;

(c) $P(A \cup B) = 0.2 + 0.35 = 0.55$.

2.70 (a) $0.02 + 0.30 = 0.32 = 32\%$;

(b) $0.32 + 0.25 + 0.30 = 0.87 = 87\%$;

(c) $0.05 + 0.06 + 0.02 = 0.13 = 13\%$;

(d) $1 - 0.05 - 0.32 = 0.63 = 63\%$.

2.71 (a) $0.12 + 0.19 = 0.31$;

(b) $1 - 0.07 = 0.93$;

(c) $0.12 + 0.19 = 0.31$.

2.72 (a) $1 - 0.40 = 0.60$.

(b) The probability that all six purchasing the electric oven or all six purchasing the gas oven is $0.007 + 0.104 = 0.111$. So the probability that at least one of each type is purchased is $1 - 0.111 = 0.889$.

2.73 (a) $P(C) = 1 - P(A) - P(B) = 1 - 0.990 - 0.001 = 0.009$;

(b) $P(B') = 1 - P(B) = 1 - 0.001 = 0.999$;

(c) $P(B) + P(C) = 0.01$.

2.74 (a) $(\$4.50 - \$4.00) \times 50,000 = \$25,000$;

(b) Since the probability of underfilling is 0.001, we would expect $50,000 \times 0.001 = 50$ boxes to be underfilled. So, instead of having $(\$4.50 - \$4.00) \times 50 = \$25$ profit for those 50 boxes, there are a loss of $\$4.00 \times 50 = \$200$ due to the cost. So, the loss in profit expected due to underfilling is $\$25 + \$200 = \$250$.

2.75 (a) $1 - 0.95 - 0.002 = 0.048$;

(b) $(\$25.00 - \$20.00) \times 10,000 = \$50,000$;

(c) $(0.05)(10,000) \times \$5.00 + (0.05)(10,000) \times \$20 = \$12,500$.

2.76 $P(A' \cap B') = 1 - P(A \cup B) = 1 - (P(A) + P(B) - P(A \cap B) = 1 + P(A \cap B) - P(A) - P(B)$.

2.77 (a) The probability that a convict who pushed dope, also committed armed robbery.

(b) The probability that a convict who committed armed robbery, did not push dope.

(c) The probability that a convict who did not push dope also did not commit armed robbery.

2.78 $P(S \mid A) = 10/18 = 5/9$.

2.79 Consider the events:
M: a person is a male;
S: a person has a secondary education;
C: a person has a college degree.

(a) $P(M \mid S) = 28/78 = 14/39$;

(b) $P(C' \mid M') = 95/112$.

2.80 Consider the events:
A: a person is experiencing hypertension,
B: a person is a heavy smoker,
C: a person is a nonsmoker.

(a) $P(A \mid B) = 30/49$;

(b) $P(C \mid A') = 48/93 = 16/31$.

2.81 (a) $P(M \cap P \cap H) = \frac{10}{68} = \frac{5}{34}$;

(b) $P(H \cap M \mid P') = \frac{P(H \cap M \cap P')}{P(P')} = \frac{22-10}{100-68} = \frac{12}{32} = \frac{3}{8}$.

2.82 (a) $(0.90)(0.08) = 0.072$;

(b) $(0.90)(0.92)(0.12) = 0.099$.

2.83 (a) 0.018;

(b) $0.22 + 0.002 + 0.160 + 0.102 + 0.046 + 0.084 = 0.614$;

(c) $0.102/0.614 = 0.166$;

(d) $\frac{0.102+0.046}{0.175+0.134} = 0.479$.

2.84 Consider the events:
C: an oil change is needed,
F: an oil filter is needed.

(a) $P(F \mid C) = \frac{P(F \cap C)}{P(C)} = \frac{0.14}{0.25} = 0.56$.

(b) $P(C \mid F) = \frac{P(C \cap F)}{P(F)} = \frac{0.14}{0.40} = 0.35$.

2.85 Consider the events:
H: husband watches a certain show,
W: wife watches the same show.

(a) $P(W \cap H) = P(W)P(H \mid W) = (0.5)(0.7) = 0.35$.

(b) $P(W \mid H) = \frac{P(W \cap H)}{P(H)} = \frac{0.35}{0.4} = 0.875$.

(c) $P(W \cup H) = P(W) + P(H) - P(W \cap H) = 0.5 + 0.4 - 0.35 = 0.55$.

2.86 Consider the events:
H: the husband will vote on the bond referendum,
W: the wife will vote on the bond referendum.
Then $P(H) = 0.21$, $P(W) = 0.28$, and $P(H \cap W) = 0.15$.

(a) $P(H \cup W) = P(H) + P(W) - P(H \cap W) = 0.21 + 0.28 - 0.15 = 0.34$.

(b) $P(W \mid H) = \frac{P(H \cap W)}{P(H)} = \frac{0.15}{0.21} = \frac{5}{7}$.

(c) $P(H \mid W') = \frac{P(H \cap W')}{P(W')} = \frac{0.06}{0.72} = \frac{1}{12}$.

2.87 Consider the events:
A: the vehicle is a camper,
B: the vehicle has Canadian license plates.

(a) $P(B \mid A) = \frac{P(A\cap B)}{P(A)} = \frac{0.09}{0.28} = \frac{9}{28}$.

(b) $P(A \mid B) = \frac{P(A\cap B)}{P(B)} = \frac{0.09}{0.12} = \frac{3}{4}$.

(c) $P(B' \cup A') = 1 - P(A \cap B) = 1 - 0.09 = 0.91$.

2.88 Define
H: head of household is home,
C: a change is made in long distance carriers.
$P(H \cap C) = P(H)P(C \mid H) = (0.4)(0.3) = 0.12$.

2.89 Consider the events:
A: the doctor makes a correct diagnosis,
B: the patient sues.
$P(A' \cap B) = P(A')P(B \mid A') = (0.3)(0.9) = 0.27$.

2.90 (a) 0.43;

(b) $(0.53)(0.22) = 0.12$;

(c) $1 - (0.47)(0.22) = 0.90$.

2.91 Consider the events:
A: the house is open,
B: the correct key is selected.
$P(A) = 0.4$, $P(A') = 0.6$, and $P(B) = \frac{\binom{1}{1}\binom{7}{2}}{\binom{8}{3}} = \frac{3}{8} = 0.375$.
So, $P[A \cup (A' \cap B)] = P(A) + P(A')P(B) = 0.4 + (0.6)(0.375) = 0.625$.

2.92 Consider the events:
F: failed the test,
P: passed the test.

(a) $P(\text{failed at least one tests}) = 1 - P(P_1P_2P_3P_4) = 1 - (0.99)(0.97)(0.98)(0.99) = 1 - 0.93 = 0.07$,

(b) $P(\text{failed 2 or 3}) = P(P_1)P(P_4)(1 - P(P_2P_3)) = (0.99)(0.99)(1 - (0.97)(0.98)) = 0.0484$.

(c) $100 \times 0.07 = 7$.

(d) 0.25.

2.93 Let A and B represent the availability of each fire engine.

(a) $P(A' \cap B') = P(A')P(B') = (0.04)(0.04) = 0.0016$.

(b) $P(A \cup B) = 1 - P(A' \cap B') = 1 - 0.0016 = 0.9984$.

2.94 $P(T' \cap N') = P(T')P(N') = (1 - P(T))(1 - P(N)) = (0.3)(0.1) = 0.03$.

2.95 Consider the events:
A_1: aspirin tablets are selected from the overnight case,
A_2: aspirin tablets are selected from the tote bag,
L_2: laxative tablets are selected from the tote bag,
T_1: thyroid tablets are selected from the overnight case,
T_2: thyroid tablets are selected from the tote bag.

(a) $P(T_1 \cap T_2) = P(T_1)P(T_2) = (3/5)(2/6) = 1/5$.

(b) $P(T_1' \cap T_2') = P(T_1')P(T_2') = (2/5)(4/6) = 4/15$.

(c) $1 - P(A_1 \cap A_2) - P(T_1 \cap T_2) = 1 - P(A_1)P(A_2) - P(T_1)P(T_2) = 1 - (2/5)(3/6) - (3/5)(2/6) = 3/5$.

2.96 Consider the events:
X: a person has an X-ray,
C: a cavity is filled,
T: a tooth is extracted.
$P(X \cap C \cap T) = P(X)P(C \mid X)P(T \mid X \cap C) = (0.6)(0.3)(0.1) = 0.018$.

2.97 (a) $P(Q_1 \cap Q_2 \cap Q_3 \cap Q_4) = P(Q_1)P(Q_2 \mid Q_1)P(Q_3 \mid Q_1 \cap Q_2)P(Q_4 \mid Q_1 \cap Q_2 \cap Q_3) = (15/20)(14/19)(13/18)(12/17) = 91/323$.

(b) Let A be the event that 4 good quarts of milk are selected. Then

$$P(A) = \frac{\binom{15}{4}}{\binom{20}{4}} = \frac{91}{323}.$$

2.98 $P = (0.95)[1 - (1 - 0.7)(1 - 0.8)](0.9) = 0.8037$.

2.99 This is a parallel system of two series subsystems.

(a) $P = 1 - [1 - (0.7)(0.7)][1 - (0.8)(0.8)(0.8)] = 0.75112$.

(b) $P = \frac{P(A' \cap C \cap D \cap E)}{P\text{system works}} = \frac{(0.3)(0.8)(0.8)(0.8)}{0.75112} = 0.2045$.

2.100 Define S: the system works.
$P(A' \mid S') = \frac{P(A' \cap S')}{P(S')} = \frac{P(A')(1 - P(C \cap D \cap E))}{1 - P(S)} = \frac{(0.3)[1-(0.8)(0.8)(0.8)]}{1-0.75112} = 0.588$.

2.101 Consider the events:
C: an adult selected has cancer,
D: the adult is diagnosed as having cancer.
$P(C) = 0.05$, $P(D \mid C) = 0.78$, $P(C') = 0.95$ and $P(D \mid C') = 0.06$. So, $P(D) = P(C \cap D) + P(C' \cap D) = (0.05)(0.78) + (0.95)(0.06) = 0.096$.

2.102 Let S_1, S_2, S_3, and S_4 represent the events that a person is speeding as he passes through the respective locations and let R represent the event that the radar traps is operating resulting in a speeding ticket. Then the probability that he receives a speeding ticket:
$P(R) = \sum_{i=1}^{4} P(R \mid S_i)P(S_i) = (0.4)(0.2) + (0.3)(0.1) + (0.2)(0.5) + (0.3)(0.2) = 0.27.$

2.103 $P(C \mid D) = \frac{P(C \cap D)}{P(D)} = \frac{0.039}{0.096} = 0.40625.$

2.104 $P(S_2 \mid R) = \frac{P(R \cap S_2)}{P(R)} = \frac{0.03}{0.27} = 1/9.$

2.105 Consider the events:
A: no expiration date,
B_1: John is the inspector, $P(B_1) = 0.20$ and $P(A \mid B_1) = 0.005$,
B_2: Tom is the inspector, $P(B_2) = 0.60$ and $P(A \mid B_2) = 0.010$,
B_3: Jeff is the inspector, $P(B_3) = 0.15$ and $P(A \mid B_3) = 0.011$,
B_4: Pat is the inspector, $P(B_4) = 0.05$ and $P(A \mid B_4) = 0.005$,
$P(B_1 \mid A) = \frac{(0.005)(0.20)}{(0.005)(0.20)+(0.010)(0.60)+(0.011)(0.15)+(0.005)(0.05)} = 0.1124.$

2.106 Consider the events
E: a malfunction by other human errors,
A: station A, B: station B, and C: station C.
$P(C \mid E) = \frac{P(E \mid C)P(C)}{P(E \mid A)P(A)+P(E \mid B)P(B)+P(E \mid C)P(C)} = \frac{(5/10)(10/43)}{(7/18)(18/43)+(7/15)(15/43)+(5/10)(10/43)} = \frac{0.1163}{0.4419} = 0.2632.$

2.107 (a) $P(A \cap B \cap C) = P(C \mid A \cap B)P(B \mid A)P(A) = (0.20)(0.75)(0.3) = 0.045.$

(b) $P(B' \cap C) = P(A \cap B' \cap C) + P(A' \cap B' \cap C) = P(C \mid A \cap B')P(B' \mid A)P(A) + P(C \mid A' \cap B')P(B' \mid A')P(A') = (0.80)(1-0.75)(0.3) + (0.90)(1-0.20)(1-0.3) = 0.564.$

(c) Use similar argument as in (a) and (b), $P(C) = P(A \cap B \cap C) + P(A \cap B' \cap C) + P(A' \cap B \cap C) + P(A' \cap B' \cap C) = 0.045 + 0.060 + 0.021 + 0.504 = 0.630.$

(d) $P(A \mid B' \cap C) = P(A \cap B' \cap C)/P(B' \cap C) = (0.06)(0.564) = 0.1064.$

2.108 Consider the events:
A: a customer purchases latex paint,
A': a customer purchases semigloss paint,
B: a customer purchases rollers.
$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A)+P(B \mid A')P(A')} = \frac{(0.60)(0.75)}{(0.60)(0.75)+(0.25)(0.30)} = 0.857.$

2.109 Consider the events:
G: guilty of committing a crime,
I: innocent of the crime,
i: judged innocent of the crime,
g: judged guilty of the crime.
$P(I \mid g) = \frac{P(g \mid I)P(I)}{P(g \mid G)P(G)+P(g \mid I)P(I)} = \frac{(0.01)(0.95)}{(0.05)(0.90)+(0.01)(0.95)} = 0.1743.$

2.110 Let A_i be the event that the ith patient is allergic to some type of week.

(a) $P(A_1 \cap A_2 \cap A_3 \cap A_4') + P(A_1 \cap A_2 \cap A_3' \cap A_4) + P(A_1 \cap A_2' \cap A_3 \cap A_4) + P(A_1' \cap A_2 \cap A_3 \cap A_4) = P(A_1)P(A_2)P(A_3)P(A_4') + P(A_1)P(A_2)P(A_3')P(A_4) + P(A_1)P(A_2')P(A_3)P(A_4) + P(A_1')P(A_2)P(A_3)P(A_4) = (4)(1/2)^4 = 1/4.$

(b) $P(A_1' \cap A_2' \cap A_3' \cap A_4') = P(A_1')P(A_2')P(A_3')P(A_4') = (1/2)^4 = 1/16.$

2.111 No solution necessary.

2.112 (a) $0.28 + 0.10 + 0.17 = 0.55.$

(b) $1 - 0.17 = 0.83.$

(c) $0.10 + 0.17 = 0.27.$

2.113 $P = \frac{\binom{13}{4}\binom{13}{6}\binom{13}{1}\binom{13}{2}}{\binom{52}{13}}.$

2.114 (a) $P(M_1 \cap M_2 \cap M_3 \cap M_4) = (0.1)^4 = 0.0001$, where M_i represents that ith person make a mistake.

(b) $P(J \cap C \cap R' \cap W') = (0.1)(0.1)(0.9)(0.9) = 0.0081.$

2.115 Let R, S, and L represent the events that a client is assigned a room at the Ramada Inn, Sheraton, and Lakeview Motor Lodge, respectively, and let F represents the event that the plumbing is faulty.

(a) $P(F) = P(F \mid R)P(R) + P(F \mid S)P(S) + P(F \mid L)P(L) = (0.05)(0.2) + (0.04)(0.4) + (0.08)(0.3) = 0.054.$

(b) $P(L \mid F) = \frac{(0.08)(0.3)}{0.054} = \frac{4}{9}.$

2.116 (a) There are $\binom{9}{3} = 84$ possible committees.

(b) There are $\binom{4}{1}\binom{5}{2} = 40$ possible committees.

(c) There are $\binom{3}{1}\binom{1}{1}\binom{5}{1} = 15$ possible committees.

2.117 Denote by R the event that a patient survives. Then $P(R) = 0.8$.

(a) $P(R_1 \cap R_2 \cap R_3') + P(R_1 \cap R_2' \cap R_3)P(R_1' \cap R_2 \cap R_3) = P(R_1)P(R_2)P(R_3') + P(R_1)P(R_2')P(R_3) + P(R_1')P(R_2)P(R_3) = (3)(0.8)(0.8)(0.2) = 0.384.$

(b) $P(R_1 \cap R_2 \cap R_3) = P(R_1)P(R_2)P(R_3) = (0.8)^3 = 0.512.$

2.118 Consider events
M: an inmate is a male,
N: an inmate is under 25 years of age.
$P(M' \cap N') = P(M') + P(N') - P(M' \cup N') = 2/5 + 1/3 - 5/8 = 13/120.$

2.119 There are $\binom{4}{3}\binom{5}{3}\binom{6}{3} = 800$ possible selections.

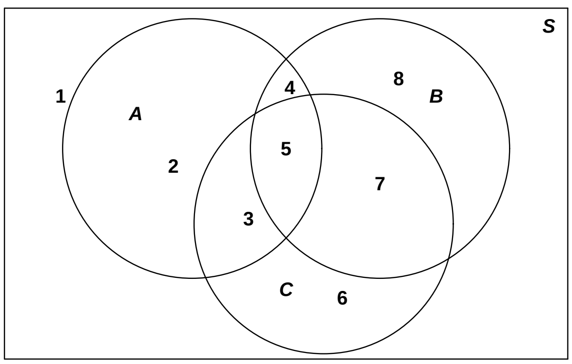
2.120 Consider the events:
B_i: a black ball is drawn on the ith drawl,
G_i: a green ball is drawn on the ith drawl.

(a) $P(B_1 \cap B_2 \cap B_3) + P(G_1 \cap G_2 \cap G_3) = (6/10)(6/10)(6/10) + (4/10)(4/10)(4/10) = 7/25$.

(b) The probability that each color is represented is $1 - 7/25 = 18/25$.

2.121 The total number of ways to receive 2 or 3 defective sets among 5 that are purchased is $\binom{3}{2}\binom{9}{3} + \binom{3}{3}\binom{9}{2} = 288$.

2.122 A Venn diagram is shown next.



(a) $(A \cap B)'$: 1, 2, 3, 6, 7, 8.

(b) $(A \cup B)'$: 1, 6.

(c) $(A \cap C) \cup B$: 3, 4, 5, 7, 8.

2.123 Consider the events:
O: overrun,
A: consulting firm A,
B: consulting firm B,
C: consulting firm C.

(a) $P(C \mid O) = \frac{P(O \mid C)P(C)}{P(O \mid A)P(A)+P(O \mid B)P(B)+P(O \mid C)P(C)} = \frac{(0.15)(0.25)}{(0.05)(0.40)+(0.03)(0.35)+(0.15)(0.25)} = \frac{0.0375}{0.0680} = 0.5515$.

(b) $P(A \mid O) = \frac{(0.05)(0.40)}{0.0680} = 0.2941$.

2.124 (a) 36;

(b) 12;

(c) order is not important.

2.125 (a) $\frac{1}{\binom{36}{2}} = 0.0016$;

(b) $\frac{\binom{12}{1}\binom{24}{1}}{\binom{36}{2}} = \frac{288}{630} = 0.4571$.

2.126 Consider the events:
C: a woman over 60 has the cancer,
P: the test gives a positive result.
So, $P(C) = 0.07$, $P(P' \mid C) = 0.1$ and $P(P \mid C') = 0.05$.
$P(C \mid P') = \frac{P(P' \mid C)P(C)}{P(P' \mid C)P(C)+P(P' \mid C')P(C')} = \frac{(0.1)(0.07)}{(0.1)(0.07)+(1-0.05)(1-0.07)} = \frac{0.007}{0.8905} = 0.00786.$

2.127 Consider the events:
A: two nondefective components are selected,
N: a lot does not contain defective components, $P(N) = 0.6$, $P(A \mid N) = 1$,
O: a lot contains one defective component, $P(O) = 0.3$, $P(A \mid O) = \frac{\binom{19}{2}}{\binom{20}{2}} = \frac{9}{10}$,
T: a lot contains two defective components,$P(T) = 0.1$, $P(A \mid T) = \frac{\binom{18}{2}}{\binom{20}{2}} = \frac{153}{190}$.

(a) $P(N \mid A) = \frac{P(A \mid N)P(N)}{P(A \mid N)P(N)+P(A \mid O)P(O)+P(A \mid T)P(T)} = \frac{(1)(0.6)}{(1)(0.6)+(9/10)(0.3)+(153/190)(0.1)}$
$= \frac{0.6}{0.9505} = 0.6312$;

(b) $P(O \mid A) = \frac{(9/10)(0.3)}{0.9505} = 0.2841$;

(c) $P(T \mid A) = 1 - 0.6312 - 0.2841 = 0.0847$.

2.128 Consider events:
D: a person has the rare disease, $P(D) = 1/500$,
P: the test shows a positive result, $P(P \mid D) = 0.95$ and $P(P \mid D') = 0.01$.
$P(D \mid P) = \frac{P(P \mid D)P(D)}{P(P \mid D)P(D)+P(P \mid D')P(D')} = \frac{(0.95)(1/500)}{(0.95)(1/500)+(0.01)(1-1/500)} = 0.1599.$

2.129 Consider the events:
1: engineer 1, $P(1) = 0.7$, and 2: engineer 2, $P(2) = 0.3$,
E: an error has occurred in estimating cost, $P(E \mid 1) = 0.02$ and $P(E \mid 2) = 0.04$.
$P(1 \mid E) = \frac{P(E \mid 1)P(1)}{P(E \mid 1)P(1)+P(E \mid 2)P(2)} = \frac{(0.02)(0.7)}{(0.02)(0.7)+(0.04)(0.3)} = 0.5385$, and
$P(2 \mid E) = 1 - 0.5385 = 0.4615$. So, more likely engineer 1 did the job.

2.130 Consider the events: D: an item is defective

(a) $P(D_1D_2D_3) = P(D_1)P(D_2)P(D_3) = (0.2)^3 = 0.008$.

(b) $P(\text{three out of four are defectives}) = \binom{4}{3}(0.2)^3(1 - 0.2) = 0.0256$.

2.131 Let A be the event that an injured worker is admitted to the hospital and N be the event that an injured worker is back to work the next day. $P(A) = 0.10$, $P(N) = 0.15$ and $P(A \cap N) = 0.02$. So, $P(A \cup N) = P(A) + P(N) - P(A \cap N) = 0.1 + 0.15 - 0.02 = 0.23$.

2.132 Consider the events:
T: an operator is trained, $P(T) = 0.5$,
M an operator meets quota, $P(M \mid T) = 0.9$ and $P(M \mid T') = 0.65$.
$P(T \mid M) = \frac{P(M \mid T)P(T)}{P(M \mid T)P(T)+P(M \mid T')P(T')} = \frac{(0.9)(0.5)}{(0.9)(0.5)+(0.65)(0.5)} = 0.5807.$

2.133 Consider the events:

A: purchased from vendor A,

D: a customer is dissatisfied.

Then $P(A) = 0.2$, $P(A \mid D) = 0.5$, and $P(D) = 0.1$.

So, $P(D \mid A) = \frac{P(A \mid D)P(D)}{P(A)} = \frac{(0.5)(0.1)}{0.2} = 0.25$.

2.134 (a) $P(\text{Union member} \mid \text{New company (same field)}) = \frac{13}{13+10} = \frac{13}{23} = 0.5652$.

(b) $P(\text{Unemployed} \mid \text{Union member}) = \frac{2}{40+13+4+2} = \frac{2}{59} = 0.034$.

2.135 Consider the events:

C: the queen is a carrier, $P(C) = 0.5$,

D: a prince has the disease, $P(D \mid C) = 0.5$.

$P(C \mid D_1'D_2'D_3') = \frac{P(D_1'D_2'D_3' \mid C)P(C)}{P(D_1'D_2'D_3' \mid C)P(C)+P(D_1'D_2'D_3' \mid C')P(C')} = \frac{(0.5)^3(0.5)}{(0.5)^3(0.5)+1(0.5)} = \frac{1}{9}$.

2.136 Using the solutions to Exercise 2.50, we know that there are total ${}_{365}P_{60}$ ways that no two students have the same birth date. Since the total number of ways of the birth dates that 60 students can have is 365^{60}, the probability that at least two students will have the same birth date in a class of 60 is $P = 1 - \frac{{}_{365}P_{60}}{365^{60}}$. To compute this number, regular calculator may not be able to handle it. Using approximation (such as Stirling's approximation formula), we obtain $P = 0.9941$, which is quite high.

Chapter 3

Random Variables and Probability Distributions

3.1 Discrete; continuous; continuous; discrete; discrete; continuous.

3.2 A table of sample space and assigned values of the random variable is shown next.

Sample Space	x
NNN	0
NNB	1
NBN	1
BNN	1
NBB	2
BNB	2
BBN	2
BBB	3

3.3 A table of sample space and assigned values of the random variable is shown next.

Sample Space	w
HHH	3
HHT	1
HTH	1
THH	1
HTT	-1
THT	-1
TTH	-1
TTT	-3

3.4 $S = \{HHH, THHH, HTHHH, TTHHH, TTTHHH, HTTHHH, THTHHH, HHTHHH, \dots\}$; The sample space is discrete containing as many elements as there are positive integers.

3.5 (a) $c = 1/30$ since $1 = \sum_{x=0}^{3} c(x^2+4) = 30c$.

(b) $c = 1/10$ since

$$1 = \sum_{x=0}^{2} c\binom{2}{x}\binom{3}{3-x} = c\left[\binom{2}{0}\binom{3}{3} + \binom{2}{1}\binom{3}{2} + \binom{2}{2}\binom{3}{1}\right] = 10c.$$

3.6 (a) $P(X > 200) = \int_{200}^{\infty} \frac{20000}{(x+100)^3}\, dx = -\left.\frac{10000}{(x+100)^2}\right|_{200}^{\infty} = \frac{1}{9}$.

(b) $P(80 < X < 200) = \int_{80}^{120} \frac{20000}{(x+100)^3}\, dx = -\left.\frac{10000}{(x+100)^2}\right|_{80}^{120} = \frac{1000}{9801} = 0.1020$.

3.7 (a) $P(X < 1.2) = \int_0^1 x\, dx + \int_1^{1.2} (2-x)\, dx = \left.\frac{x^2}{2}\right|_0^1 + \left.\left(2x - \frac{x^2}{2}\right)\right|_1^{1.2} = 0.68$.

(b) $P(0.5 < X < 1) = \int_{0.5}^{1} x\, dx = \left.\frac{x^2}{2}\right|_{0.5}^{1} = 0.375$.

3.8 Referring to the sample space in Exercise 3.3 and making use of the fact that $P(H) = 2/3$ and $P(T) = 1/3$, we have
$P(W = -3) = P(TTT) = (1/3)^3 = 1/27$;
$P(W = -1) = P(HTT) + P(THT) + P(TTH) = 3(2/3)(1/3)^2 = 2/9$;
$P(W = 1) = P(HHT) + P(HTH) + P(THH) = 3(2/3)^2(1/3) = 2/9$;
$P(W = 3) = P(HHH) = (2/3)^3 = 8/27$;
The probability distribution for W is then

w	-3	-1	1	3
$P(W = w)$	$1/27$	$2/9$	$2/9$	$8/27$

3.9 (a) $P(0 < X < 1) = \int_0^1 \frac{2(x+2)}{5}\, dx = \left.\frac{(x+2)^2}{5}\right|_0^1 = 1$.

(b) $P(1/4 < X < 1/2) = \int_{1/4}^{1/2} \frac{2(x+2)}{5}\, dx = \left.\frac{(x+2)^2}{5}\right|_{1/4}^{1/2} = 19/80$.

3.10 The die can land in 6 different ways each with probability 1/6. Therefore, $f(x) = \frac{1}{6}$, for $x = 1, 2, \ldots, 6$.

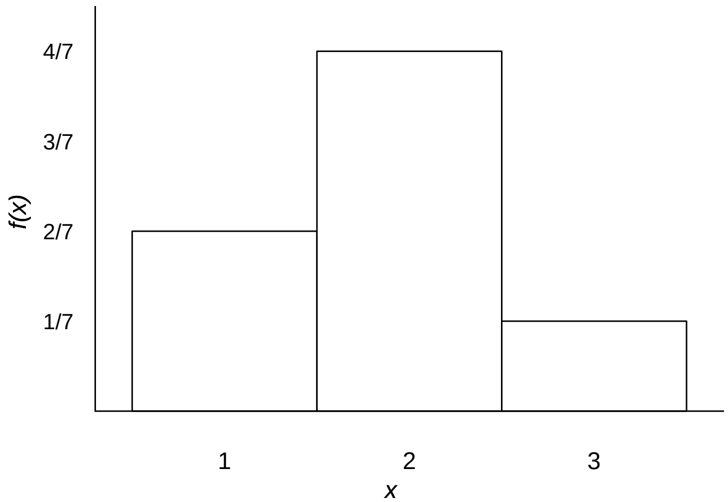
3.11 We can select x defective sets from 2, and $3 - x$ good sets from 5 in $\binom{2}{x}\binom{5}{3-x}$ ways. A random selection of 3 from 7 sets can be made in $\binom{7}{3}$ ways. Therefore,

$$f(x) = \frac{\binom{2}{x}\binom{5}{3-x}}{\binom{7}{3}}, \qquad x = 0, 1, 2.$$

In tabular form

x	0	1	2
$f(x)$	$2/7$	$4/7$	$1/7$

The following is a probability histogram:



3.12 (a) $P(T = 5) = F(5) - F(4) = 3/4 - 1/2 = 1/4$.

(b) $P(T > 3) = 1 - F(3) = 1 - 1/2 = 1/2$.

(c) $P(1.4 < T < 6) = F(6) - F(1.4) = 3/4 - 1/4 = 1/2$.

3.13 The c.d.f. of X is

$$F(x) = \begin{cases} 0, & \text{for } x < 0, \\ 0.41, & \text{for } 0 \le x < 1, \\ 0.78, & \text{for } 1 \le x < 2, \\ 0.94, & \text{for } 2 \le x < 3, \\ 0.99, & \text{for } 3 \le x < 4, \\ 1, & \text{for } x \ge 4. \end{cases}$$

3.14 (a) $P(X < 0.2) = F(0.2) = 1 - e^{-1.6} = 0.7981$;

(b) $f(x) = F'(x) = 8e^{-8x}$. Therefore, $P(X < 0.2) = 8\int_0^{0.2} e^{-8x}\, dx = -e^{-8x}\big|_0^{0.2} = 0.7981$.

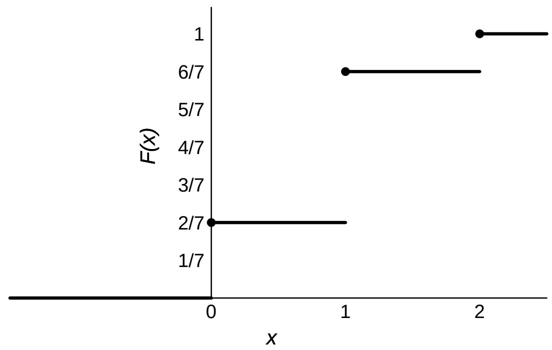
3.15 The c.d.f. of X is

$$F(x) = \begin{cases} 0, & \text{for } x < 0, \\ 2/7, & \text{for } 0 \le x < 1, \\ 6/7, & \text{for } 1 \le x < 2, \\ 1, & \text{for } x \ge 2. \end{cases}$$

(a) $P(X = 1) = P(X \le 1) - P(X \le 0) = 6/7 - 2/7 = 4/7$;

(b) $P(0 < X \le 2) = P(X \le 2) - P(X \le 0) = 1 - 2/7 = 5/7$.

3.16 A graph of the c.d.f. is shown next.



3.17 (a) Area $= \int_1^3 (1/2)\ dx = \frac{x}{2}\Big|_1^3 = 1$.

(b) $P(2 < X < 2.5) \int_2^{2.5} (1/2)\ dx = \frac{x}{2}\Big|_2^{2.5} = \frac{1}{4}$.

(c) $P(X \le 1.6) = \int_1^{1.6} (1/2)\ dx = \frac{x}{2}\Big|_1^{1.6} = 0.3$.

3.18 (a) $P(X < 4) = \int_2^4 \frac{2(1+x)}{27}\ dx = \frac{(1+x)^2}{27}\Big|_2^4 = 16/27$.

(b) $P(3 \le X < 4) = \int_3^4 \frac{2(1+x)}{27}\ dx = \frac{(1+x)^2}{27}\Big|_3^4 = 1/3$.

3.19 $F(x) = \int_1^x (1/2)\ dt = \frac{x-1}{2}$,
$P(2 < X < 2.5) = F(2.5) - F(2) = \frac{1.5}{2} - \frac{1}{2} = \frac{1}{4}$.

3.20 $F(x) = \frac{2}{27} \int_2^x (1+t)\ dt = \frac{2}{27}\left(t + \frac{t^2}{2}\right)\Big|_2^x = \frac{(x+4)(x-2)}{27}$,
$P(3 \le X < 4) = F(4) - F(3) = \frac{(8)(2)}{27} - \frac{(7)(1)}{27} = \frac{1}{3}$.

3.21 (a) $1 = k\int_0^1 \sqrt{x}\ dx = \frac{2k}{3} x^{3/2}\Big|_0^1 = \frac{2k}{3}$. Therefore, $k = \frac{3}{2}$.

(b) $F(x) = \frac{3}{2}\int_0^x \sqrt{t}\ dt = t^{3/2}\Big|_0^x = x^{3/2}$.
$P(0.3 < X < 0.6) = F(0.6) - F(0.3) = (0.6)^{3/2} - (0.3)^{3/2} = 0.3004$.

3.22 Denote by X the number of spades int he three draws. Let S and N stand for a spade and not a spade, respectively. Then
$P(X = 0) = P(NNN) = (39/52)(38/51)(37/50) = 703/1700$,
$P(X = 1) = P(SNN) + P(NSN) + P(NNS) = 3(13/52)(39/51)(38/50) = 741/1700$,
$P(X = 3) = P(SSS) = (13/52)(12/51)(11/50) = 11/850$, and
$P(X = 2) = 1 - 703/1700 - 741/1700 - 11/850 = 117/850$.
The probability mass function for X is then

x	0	1	2	3
$f(x)$	703/1700	741/1700	117/850	11/850

3.23 The c.d.f. of X is

$$F(x) = \begin{cases} 0, & \text{for } w < -3, \\ 1/27, & \text{for } -3 \le w < -1, \\ 7/27, & \text{for } -1 \le w < 1, \\ 19/27, & \text{for } 1 \le w < 3, \\ 1, & \text{for } w \ge 3, \end{cases}$$

(a) $P(W > 0 = 1 - P(W \le 0) = 1 - 7/27 = 20/27.$

(b) $P(-1 \le W < 3) = F(2) - F(-3) = 19/27 - 1/27 = 2/3.$

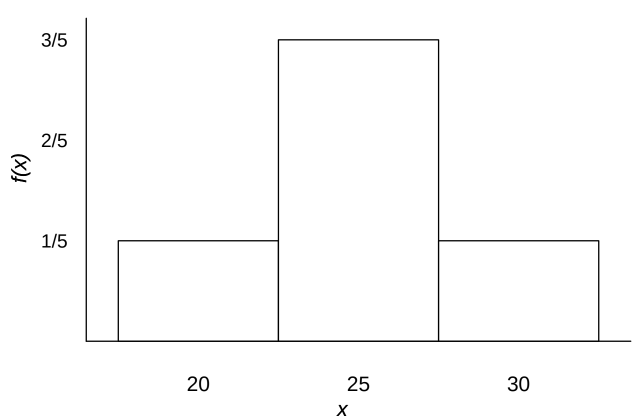
3.24 There are $\binom{10}{4}$ ways of selecting any 4 CDs from 10. We can select x jazz CDs from 5 and $4 - x$ from the remaining CDs in $\binom{5}{x}\binom{5}{4-x}$ ways. Hence

$$f(x) = \frac{\binom{5}{x}\binom{5}{4-x}}{\binom{10}{4}}, \qquad x = 0, 1, 2, 3, 4.$$

3.25 Let T be the total value of the three coins. Let D and N stand for a dime and nickel, respectively. Since we are selecting without replacement, the sample space containing elements for which $t = 20, 25$, and 30 cents corresponding to the selecting of 2 nickels and 1 dime, 1 nickel and 2 dimes, and 3 dimes. Therefore, $P(T = 20) = \frac{\binom{2}{2}\binom{4}{1}}{\binom{6}{3}} = \frac{1}{5}$,
$P(T = 25) = \frac{\binom{2}{1}\binom{4}{2}}{\binom{6}{3}} = \frac{3}{5}$,
$P(T = 30) = \frac{\binom{4}{3}}{\binom{6}{3}} = \frac{1}{5}$,
and the probability distribution in tabular form is

t	20	25	30
$P(T = t)$	1/5	3/5	1/5

As a probability histogram



3.26 Denote by X the number of green balls in the three draws. Let G and B stand for the colors of green and black, respectively.

Simple Event	x	$P(X = x)$
BBB	0	$(2/3)^3 = 8/27$
GBB	1	$(1/3)(2/3)^2 = 4/27$
BGB	1	$(1/3)(2/3)^2 = 4/27$
BBG	1	$(1/3)(2/3)^2 = 4/27$
BGG	2	$(1/3)^2(2/3) = 2/27$
GBG	2	$(1/3)^2(2/3) = 2/27$
GGB	2	$(1/3)^2(2/3) = 2/27$
GGG	3	$(1/3)^3 = 1/27$

The probability mass function for X is then

x	0	1	2	3
$P(X = x)$	8/27	4/9	2/9	1/27

3.27 (a) For $x \geq 0$, $F(x) = \int_0^x \frac{1}{2000} \exp(-t/2000)\, dt = -\exp(-t/2000)\big|_0^x$ $= 1 - \exp(-x/2000)$. So

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - \exp(-x/2000), & x \geq 0. \end{cases}$$

(b) $P(X > 1000) = 1 - F(1000) = 1 - [1 - \exp(-1000/2000)] = 0.6065$.

(c) $P(X < 2000) = F(2000) = 1 - \exp(-2000/2000) = 0.6321$.

3.28 (a) $f(x) \geq 0$ and $\int_{23.75}^{26.25} \frac{2}{5}\, dx = \frac{2}{5}t\big|_{23.75}^{26.25} = \frac{2.5}{2.5} = 1$.

(b) $P(X < 24) = \int_{23.75}^{24} \frac{2}{5}\, dx = \frac{2}{5}(24 - 23.75) = 0.1$.

(c) $P(X > 26) = \int_{26}^{26.25} \frac{2}{5}\, dx = \frac{2}{5}(26.25 - 26) = 0.1$. It is not extremely rare.

3.29 (a) $f(x) \geq 0$ and $\int_1^\infty 3x^{-4}\, dx = -3\, \frac{x^{-3}}{3}\Big|_1^\infty = 1$. So, this is a density function.

(b) For $x \geq 1$, $F(x) = \int_1^x 3t^{-4}\, dt = 1 - x^{-3}$. So,

$$F(x) = \begin{cases} 0, & x < 1, \\ 1 - x^{-3}, & x \geq 1. \end{cases}$$

(c) $P(X > 4) = 1 - F(4) = 4^{-3} = 0.0156$.

3.30 (a) $1 = k \int_{-1}^1 (3 - x^2)\, dx = k \left(3x - \frac{x^3}{3}\right)\Big|_{-1}^1 = \frac{16}{3}k$. So, $k = \frac{3}{16}$.

(b) For $-1 \le x < 1$, $F(x) = \frac{3}{16}\int_{-1}^{x}(3-t^2)\,dt = \left(3t - \frac{1}{3}t^3\right)\Big|_{-1}^{x} = \frac{1}{2} + \frac{9}{16}x - \frac{x^3}{16}$.
So, $P\left(X < \frac{1}{2}\right) = \frac{1}{2} - \left(\frac{9}{16}\right)\left(\frac{1}{2}\right) - \frac{1}{16}\left(\frac{1}{2}\right)^3 = \frac{99}{128}$.

(c) $P(|X| < 0.8) = P(X < -0.8) + P(X > 0.8) = F(-0.8) + 1 - F(0.8)$
$= 1 + \left(\frac{1}{2} - \frac{9}{16}0.8 + \frac{1}{16}0.8^3\right) - \left(\frac{1}{2} + \frac{9}{16}0.8 - \frac{1}{16}0.8^3\right) = 0.164$.

3.31 (a) For $y \ge 0$, $F(y) = \frac{1}{4}\int_0^y e^{-t/4}\,dy = 1 - e^{y/4}$. So, $P(Y > 6) = e^{-6/4} = 0.2231$. This probability certainly cannot be considered as "unlikely."

(b) $P(Y \le 1) = 1 - e^{-1/4} = 0.2212$, which is not so small either.

3.32 (a) $f(y) \ge 0$ and $\int_0^1 5(1-y)^4\,dy = -\left.(1-y)^5\right|_0^1 = 1$. So, this is a density function.

(b) $P(Y < 0.1) = -\left.(1-y)^5\right|_0^{0.1} = 1 - (1-0.1)^5 = 0.4095$.

(c) $P(Y > 0.5) = (1-0.5)^5 = 0.03125$.

3.33 (a) Using integral by parts and setting $1 = k\int_0^1 y^4(1-y)^3\,dy$, we obtain $k = 280$.

(b) For $0 \le y < 1$, $F(y) = 56y^5(1-Y)^3 + 28y^6(1-y)^2 + 8y^7(1-y) + y^8$. So, $P(Y \le 0.5) = 0.3633$.

(c) Using the cdf in (b), $P(Y > 0.8) = 0.0563$.

3.34 (a) The event $Y = y$ means that among 5 selected, exactly y tubes meet the specification (M) and $5-y$ (M') does not. The probability for one combination of such a situation is $(0.99)^y(1-0.99)^{5-y}$ if we assume independence among the tubes. Since there are $\frac{5!}{y!(5-y)!}$ permutations of getting y Ms and $5-y$ M's, the probability of this event ($Y = y$) would be what it is specified in the problem.

(b) Three out of 5 is outside of specification means that $Y = 2$. $P(Y = 2) = 9.8\times 10^{-6}$ which is extremely small. So, the conjecture is false.

3.35 (a) $P(X > 8) = 1 - P(X \le 8) = \sum_{x=0}^{8} e^{-6}\frac{6^x}{x!} = e^{-6}\left(\frac{6^0}{0!} + \frac{6^1}{1!} + \cdots + \frac{6^8}{8!}\right) = 0.1528$.

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

3.36 For $0 < x < 1$, $F(x) = 2\int_0^x (1-t)\,dt = -\left.(1-t)^2\right|_0^x = 1 - (1-x)^2$.

(a) $P(X \le 1/3) = 1 - (1-1/3)^2 = 5/9$.

(b) $P(X > 0.5) = (1-1/2)^2 = 1/4$.

(c) $P(X < 0.75 \mid X \ge 0.5) = \frac{P(0.5 \le X < 0.75)}{P(X \ge 0.5)} = \frac{(1-0.5)^2 - (1-0.75)^2}{(1-0.5)^2} = \frac{3}{4}$.

3.37 (a) $\sum_{x=0}^{3}\sum_{y=0}^{3} f(x,y) = c\sum_{x=0}^{3}\sum_{y=0}^{3} xy = 36c = 1$. Hence $c = 1/36$.

(b) $\sum_x\sum_y f(x,y) = c\sum_x\sum_y |x-y| = 15c = 1$. Hence $c = 1/15$.

3.38 The joint probability distribution of (X, Y) is

$f(x,y)$		x = 0	1	2	3
y	0	0	1/30	2/30	3/30
	1	1/30	2/30	3/30	4/30
	2	2/30	3/30	4/30	5/30

(a) $P(X \le 2, Y = 1) = f(0,1) + f(1,1) + f(2,1) = 1/30 + 2/30 + 3/30 = 1/5.$

(b) $P(X > 2, Y \le 1) = f(3,0) + f(3,1) = 3/30 + 4/30 = 7/30.$

(c) $P(X > Y) = f(1,0) + f(2,0) + f(3,0) + f(2,1) + f(3,1) + f(3,2)$
$= 1/30 + 2/30 + 3/30 + 3/30 + 4/30 + 5/30 = 3/5.$

(d) $P(X + Y = 4) = f(2,2) + f(3,1) = 4/30 + 4/30 = 4/15.$

3.39 (a) We can select x oranges from 3, y apples from 2, and $4 - x - y$ bananas from 3 in $\binom{3}{x}\binom{2}{y}\binom{3}{4-x-y}$ ways. A random selection of 4 pieces of fruit can be made in $\binom{8}{4}$ ways. Therefore,

$$f(x,y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{4-x-y}}{\binom{8}{4}}, \qquad x = 0,1,2,3; \quad y = 0,1,2; \quad 1 \le x + y \le 4.$$

(b) $P[(X,Y) \in A] = P(X + Y \le 2) = f(1,0) + f(2,0) + f(0,1) + f(1,1) + f(0,2)$
$= 3/70 + 9/70 + 2/70 + 18/70 + 3/70 = 1/2.$

3.40 (a) $g(x) = \frac{2}{3}\int_0^1 (x + 2y)\ dy = \frac{2}{3}(x + 1)$, for $0 \le x \le 1$.

(b) $h(y) = \frac{2}{3}\int_0^1 (x + 2y)\ dy = \frac{1}{3}(1 + 4y)$, for $0 \le y \le 1$.

(c) $P(X < 1/2) = \frac{2}{3}\int_0^{1/2} (x + 1)\ dx = \frac{5}{12}.$

3.41 (a) $P(X + Y \le 1/2) = \int_0^{1/2}\int_0^{1/2-y} 24xy\ dx\ dy = 12\int_0^{1/2}\left(\frac{1}{2} - y\right)^2 y\ dy = \frac{1}{16}.$

(b) $g(x) = \int_0^{1-x} 24xy\ dy = 12x(1 - x)^2$, for $0 \le x < 1$.

(c) $f(y|x) = \frac{24xy}{12x(1-x)^2} = \frac{2y}{(1-x)^2}$, for $0 \le y \le 1 - x$.
Therefore, $P(Y < 1/8 \mid X = 3/4) = 32\int_0^{1/8} y\ dy = 1/4.$

3.42 Since $h(y) = e^{-y}\int_0^{\infty} e^{-x}\ dx = e^{-y}$, for $y > 0$, then $f(x|y) = f(x,y)/h(y) = e^{-x}$, for $x > 0$. So, $P(0 < X < 1 \mid Y = 2) = \int_0^1 e^{-x}\ dx = 0.6321.$

3.43 (a) $P(0 \le X \le 1/2,\ 1/4 \le Y \le 1/2) = \int_0^{1/2}\int_{1/4}^{1/2} 4xy\ dy\ dx = 3/8\int_0^{1/2} x\ dx = 3/64.$

(b) $P(X < Y) = \int_0^1\int_0^y 4xy\ dx\ dy = 2\int_0^1 y^3\ dy = 1/2.$

3.44 (a) $1 = k\int_{30}^{50}\int_{30}^{50} (x^2 + y^2)\ dx\ dy = k(50 - 30)\left(\int_{30}^{50} x^2\ dx + \int_{30}^{50} y^2\ dy\right) = \frac{392k}{3} \cdot 10^4.$
So, $k = \frac{3}{392} \cdot 10^{-4}.$

(b) $P(30 \le X \le 40,\ 40 \le Y \le 50) = \frac{3}{392} \cdot 10^{-4} \int_{30}^{40} \int_{40}^{50} (x^2 + y^2)\ dy\ dx$
$= \frac{3}{392} \cdot 10^{-3} (\int_{30}^{40} x^2\ dx + \int_{40}^{50} y^2\ dy) = \frac{3}{392} \cdot 10^{-3} \left(\frac{40^3 - 30^3}{3} + \frac{50^3 - 40^3}{3} \right) = \frac{49}{196}$.

(c) $P(30 \le X \le 40,\ 30 \le Y \le 40) = \frac{3}{392} \cdot 10^{-4} \int_{30}^{40} \int_{30}^{40} (x^2 + y^2)\ dx\ dy$
$= 2\frac{3}{392} \cdot 10^{-4} (40 - 30) \int_{30}^{40} x^2\ dx = \frac{3}{196} \cdot 10^{-3} \frac{40^3 - 30^3}{3} = \frac{37}{196}$.

3.45 $P(X + Y > 1/2) = 1 - P(X + Y < 1/2) = 1 - \int_0^{1/4} \int_x^{1/2 - x} \frac{1}{y}\ dy\ dx$
$= 1 - \int_0^{1/4} \left[\ln\left(\frac{1}{2} - x\right) - \ln x \right]\ dx = 1 + \left[\left(\frac{1}{2} - x\right) \ln\left(\frac{1}{2} - x\right) - x \ln x \right] \Big|_0^{1/4}$
$= 1 + \frac{1}{4} \ln\left(\frac{1}{4}\right) = 0.6534$.

3.46 (a) From the column totals of Exercise 3.38, we have

x	0	1	2	3
$g(x)$	1/10	1/5	3/10	2/5

(b) From the row totals of Exercise 3.38, we have

y	0	1	2
$h(y)$	1/5	1/3	7/15

3.47 (a) $g(x) = 2 \int_x^1\ dy = 2(1 - x)$ for $0 < x < 1$;
$h(y) = 2 \int_0^y\ dx = 2y$, for $0 < y < 1$.
Since $f(x, y) \ne g(x)h(y)$, X and Y are not independent.

(b) $f(x|y) = f(x, y)/h(y) = 1/y$, for $0 < x < y$.
Therefore, $P(1/4 < X < 1/2 \mid Y = 3/4) = \frac{4}{3} \int_{1/4}^{1/2} dx = \frac{1}{3}$.

3.48 (a) $g(2) = \sum_{y=0}^{2} f(2, y) = f(2, 0) + f(2, 1) + f(2, 2) = 9/70 + 18/70 + 3/70 = 3/7$. So,
$f(y|2) = f(2, y)/g(2) = (7/3)f(2, y)$.
$f(0|2) = (7/3)f(2, 0) = (7/3)(9/70) = 3/10$, $f(1|2) = 3/5$ and $f(2|2) = 1/10$. In tabular form,

y	0	1	2
$f(y\|2)$	3/10	3/5	1/10

(b) $P(Y = 0 \mid X = 2) = f(0|2) = 3/10$.

3.49 (a)

x	1	2	3
$g(x)$	0.10	0.35	0.55

(b)

y	1	2	3
$h(y)$	0.20	0.50	0.30

(c) $P(Y = 3 \mid X = 2) = \frac{0.2}{0.05 + 0.10 + 0.20} = 0.5714$.

3.50

	$f(x,y)$	$x=2$	$x=4$	$h(y)$
y	1	0.10	0.15	0.25
	3	0.20	0.30	0.50
	5	0.10	0.15	0.25
	$g(x)$	0.40	0.60	

(a)

x	2	4
$g(x)$	0.40	0.60

(b)

y	1	3	5
$h(y)$	0.25	0.50	0.25

3.51 (a) Let X be the number of 4's and Y be the number of 5's. The sample space consists of 36 elements each with probability 1/36 of the form (m, n) where m is the outcome of the first roll of the die and n is the value obtained on the second roll. The joint probability distribution $f(x, y)$ is defined for $x = 0, 1, 2$ and $y = 0, 1, 2$ with $0 \leq x + y \leq 2$. To find $f(0, 1)$, for example, consider the event A of obtaining zero 4's and one 5 in the 2 rolls. Then $A = \{(1,5),(2,5),(3,5),(6,5),(5,1),(5,2),(5,3),(5,6)\}$, so $f(0,1) = 8/36 = 2/9$. In a like manner we find $f(0,0) = 16/36 = 4/9$, $f(0,2) = 1/36$, $f(1,0) = 2/9$, $f(2,0) = 1/36$, and $f(1,1) = 1/18$.

(b) $P[(X,Y) \in A] = P(2X + Y < 3) = f(0,0) + f(0,1) + f(0,2) + f(1,0) = 4/9 + 1/9 + 1/36 + 2/9 = 11/12$.

3.52 A tabular form of the experiment can be established as,

Sample Space	$\boldsymbol{x}$	$\boldsymbol{y}$
HHH	3	3
HHT	2	1
HTH	2	1
THH	2	1
HTT	1	-1
THT	1	-1
TTH	1	-1
TTT	0	-3

So, the joint probability distribution is,

	$f(x,y)$	$x=0$	$x=1$	$x=2$	$x=3$
y	-3	1/8			
	-1		3/8		
	1			3/8	
	3				1/8

3.53 (a) If (x, y) represents the selection of x kings and y jacks in 3 draws, we must have $x = 0, 1, 2, 3$; $y = 0, 1, 2, 3$; and $0 \leq x + y \leq 3$. Therefore, $(1, 2)$ represents the selection of 1 king and 2 jacks which will occur with probability

$$f(1,2) = \frac{\binom{4}{1}\binom{4}{2}}{\binom{12}{3}} = \frac{6}{55}.$$

Proceeding in a similar fashion for the other possibilities, we arrive at the following joint probability distribution:

	$f(x,y)$	x = 0	1	2	3
	0	1/55	6/55	6/55	1/55
y	1	6/55	16/55	6/55	
	2	6/55	6/55		
	3	1/55			

(b) $P[(X, Y) \in A] = P(X + Y \geq 2) = 1 - P(X + Y < 2) = 1 - 1/55 - 6/55 - 6/55 = 42/55$.

3.54 (a) $P(H) = 0.4$, $P(T) = 0.6$, and $S = \{HH, HT, TH, TT\}$. Let (W, Z) represent a typical outcome of the experiment. The particular outcome $(1, 0)$ indicating a total of 1 head and no heads on the first toss corresponds to the event TH. Therefore, $f(1, 0) = P(W = 1, Z = 0) = P(TH) = P(T)P(H) = (0.6)(0.4) = 0.24$. Similar calculations for the outcomes $(0, 0)$, $(1, 1)$, and $(2, 1)$ lead to the following joint probability distribution:

	$f(w,z)$	w = 0	1	2
z	0	0.36	0.24	
	1		0.24	0.16

(b) Summing the columns, the marginal distribution of W is

w	0	1	2
$g(w)$	0.36	0.48	0.16

(c) Summing the rows, the marginal distribution of Z is

z	0	1
$h(z)$	0.60	0.40

(d) $P(W \geq 1) = f(1, 0) + f(1, 1) + f(2, 1) = 0.24 + 0.24 + 0.16 = 0.64$.

3.55 $g(x) = \frac{1}{8}\int_2^4 (6-x-y)\ dy = \frac{3-x}{4}$, for $0 < x < 2$.
So, $f(y|x) = \frac{f(x,y)}{g(x)} = \frac{6-x-y}{2(3-x)}$, for $2 < y < 4$,
and $P(1 < Y < 3 \mid X = 1) = \frac{1}{4}\int_2^3 (5-y)\ dy = \frac{5}{8}$.

3.56 Since $f(1,1) \neq g(1)h(1)$, the variables are not independent.

3.57 X and Y are independent since $f(x,y) = g(x)h(y)$ for all (x,y).

3.58 (a) $h(y) = 6\int_0^{1-y} x\ dx = 3(1-y)^2$, for $0 < y < 1$. Since $f(x|y) = \frac{f(x,y)}{h(y)} = \frac{2x}{(1-y)^2}$, for $0 < x < 1-y$, involves the variable y, X and Y are not independent.

(b) $P(X > 0.3 \mid Y = 0.5) = 8\int_{0.3}^{0.5} x\ dx = 0.64$.

3.59 (a) $1 = k\int_0^1\int_0^1\int_0^2 xy^2z\ dx\ dy\ dz = 2k\int_0^1\int_0^1 y^2z\ dy\ dz = \frac{2k}{3}\int_0^1 z\ dz = \frac{k}{3}$. So, $k = 3$.

(b) $P\left(X < \frac{1}{4}, Y > \frac{1}{2}, 1 < Z < 2\right) = 3\int_0^{1/4}\int_{1/2}^1\int_1^2 xy^2z\ dx\ dy\ dz = \frac{9}{2}\int_0^{1/4}\int_{1/2}^1 y^2z\ dy\ dz$ $= \frac{21}{16}\int_0^{1/4} z\ dz = \frac{21}{512}$.

3.60 $g(x) = 4\int_0^1 xy\ dy = 2x$, for $0 < x < 1$; $h(y) = 4\int_0^1 xy\ dx = 2y$, for $0 < y < 1$. Since $f(x,y) = g(x)h(y)$ for all (x,y), X and Y are independent.

3.61 $g(x) = k\int_{30}^{50}(x^2+y^2)\ dy = k\left(x^2y + \frac{y^3}{3}\right)\Big|_{30}^{50} = k\left(20x^2 + \frac{98{,}000}{3}\right)$, and
$h(y) = k\left(20y^2 + \frac{98{,}000}{3}\right)$.
Since $f(x,y) \neq g(x)h(y)$, X and Y are not independent.

3.62 (a) $g(y,z) = \frac{4}{9}\int_0^1 xyz^2\ dx = \frac{2}{9}yz^2$, for $0 < y < 1$ and $0 < z < 3$.

(b) $h(y) = \frac{2}{9}\int_0^3 yz^2\ dz = 2y$, for $0 < y < 1$.

(c) $P\left(\frac{1}{4} < X < \frac{1}{2}, Y > \frac{1}{3}, Z < 2\right) = \frac{4}{9}\int_1^2\int_{1/3}^1\int_{1/4}^{1/2} xyz^2\ dx\ dy\ dz = \frac{7}{162}$.

(d) Since $f(x|y,z) = \frac{f(x,y,z)}{g(y,z)} = 2x$, for $0 < x < 1$, $P\left(0 < X < \frac{1}{2} \mid Y = \frac{1}{4}, Z = 2\right) = 2\int_0^{1/2} x\ dx = \frac{1}{4}$.

3.63 $g(x) = 24\int_0^{1-x} xy\ dy = 12x(1-x)^2$, for $0 < x < 1$.

(a) $P(X \geq 0.5) = 12\int_{0.5}^1 x(1-x)^2\ dx = \int_{0.5}^1 (12x - 24x^2 + 12x^3)\ dx = \frac{5}{16} = 0.3125$.

(b) $h(y) = 24\int_0^{1-y} xy\ dx = 12y(1-y)^2$, for $0 < y < 1$.

(c) $f(x|y) = \frac{f(x,y)}{h(y)} = \frac{24xy}{12y(1-y)^2} = \frac{2x}{(1-y)^2}$, for $0 < x < 1-y$.
So, $P\left(X < \frac{1}{8} \mid Y = \frac{3}{4}\right) = \int_0^{1/8} \frac{2x}{1/16}\ dx = 32\int_0^{1/8} = 0.25$.

3.64 (a)

x	1	3	5	7
$f(x)$	0.4	0.2	0.2	0.2

(b) $P(4 < X \leq 7) = P(X \leq 7) - P(X \leq 4) = F(7) - F(4) = 1 - 0.6 = 0.4$.

3.65 (a) $g(x) = \int_0^\infty ye^{-y(1+x)}\ dy = -\frac{1}{1+x} ye^{-y(1+x)}\Big|_0^\infty + \frac{1}{1+x}\int_0^\infty e^{-y(1+x)}\ dy$
$= -\frac{1}{(1+x)^2}\ e^{-y(1+x)}\Big|_0^\infty$
$= \frac{1}{(1+x)^2}$, for $x > 0$.
$h(y) = ye^{-y}\int_0^\infty e^{-yx}\ dx = -e^{-y}\ e^{-yx}|_0^\infty = e^{-y}$, for $y > 0$.

(b) $P(X \geq 2, Y \geq 2) = \int_2^\infty \int_2^\infty ye^{-y(1+x)}\ dx\ dy = -\int_2^\infty e^{-y}\ e^{-yx}|_2^\infty\ dy = \int_2^\infty e^{-3y} dy$
$= -\frac{1}{3} e^{-3y}\Big|_2^\infty = \frac{1}{3e^6}$.

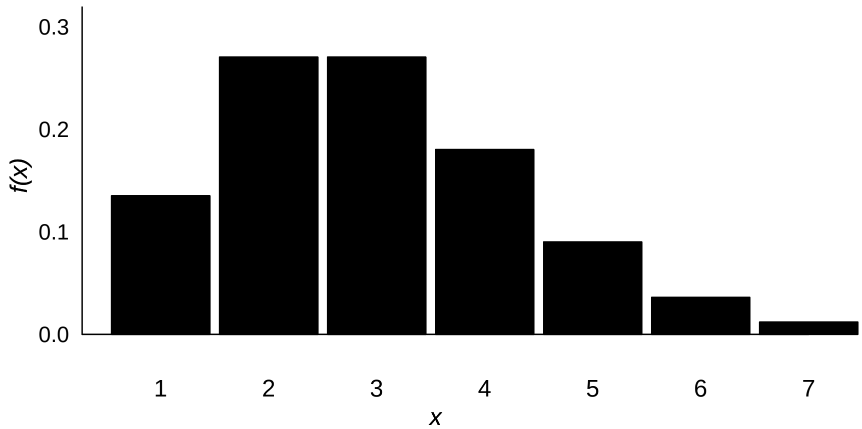
3.66 (a) $P\left(X \leq \frac{1}{2}, Y \leq \frac{1}{2}\right) = \frac{3}{2}\int_0^{1/2}\int_0^{1/2}(x^2+y^2)\ dxdy = \frac{3}{2}\int_0^{1/2}\left(x^2 y + \frac{y^3}{3}\right)\Big|_0^{1/2}\ dx$
$= \frac{3}{4}\int_0^{1/2}\left(x^2 + \frac{1}{12}\right)\ dx = \frac{1}{16}$.

(b) $P\left(X \geq \frac{3}{4}\right) = \frac{3}{2}\int_{3/4}^1\left(x^2 + \frac{1}{3}\right)\ dx = \frac{53}{128}$.

3.67 (a)

x	0	1	2	3	4	5	6
$f(x)$	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361	0.0120

(b) A histogram is shown next.



(c)

x	0	1	2	3	4	5	6
$F(x)$	0.1353	0.4060	0.6767	0.8571	0.9473	0.9834	0.9954

3.68 (a) $g(x) = \int_0^1 (x+y)\ dy = x + \frac{1}{2}$, for $0 < x < 1$, and $h(y) = y + \frac{1}{2}$ for $0 < y < 1$.

(b) $P(X > 0.5, Y > 0.5) = \int_{0.5}^1\int_{0.5}^1 (x+y)\ dx\ dy = \int_{0.5}^1\left(\frac{x^2}{2} + xy\right)\Big|_{0.5}^1\ dy$
$= \int_{0.5}^1\left[\left(\frac{1}{2}+y\right) - \left(\frac{1}{8}+\frac{y}{2}\right)\right]\ dy = \frac{3}{8}$.

3.69 $f(x) = \binom{5}{x}(0.1)^x(1-0.1)^{5-x}$, for $x = 0, 1, 2, 3, 4, 5$.

3.70 (a) $g(x) = \int_1^2\left(\frac{3x-y}{9}\right)\ dy = \frac{3xy - y^2/2}{9}\Big|_1^2 = \frac{x}{3} - \frac{1}{6}$, for $1 < x < 3$, and
$h(y) = \int_1^3\left(\frac{3x-y}{9}\right)\ dx = \frac{4}{3} - \frac{2}{9}y$, for $1 < y < 2$.

(b) No, since $g(x)h(y) \neq f(x,y)$.

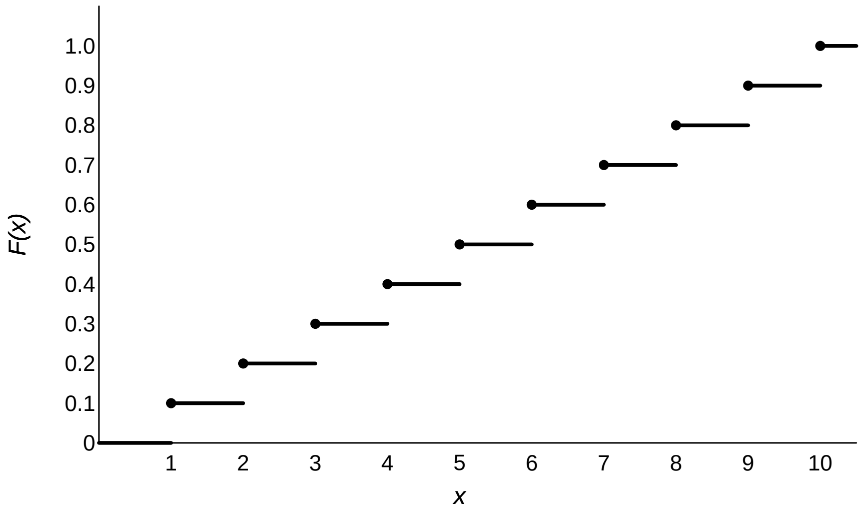
(c) $P(X > 2) = \int_2^3\left(\frac{x}{3} - \frac{1}{6}\right)\ dx = \left(\frac{x^2}{6} - \frac{x}{6}\right)\Big|_2^3 = \frac{2}{3}$.

3.71 (a) $f(x) = \frac{d}{dx}F(x) = \frac{1}{50}e^{-x/50}$, for $x > 0$.

(b) $P(X > 70) = 1 - P(X \leq 70) = 1 - F(70) = 1 - (1 - e^{-70/50}) = 0.2466$.

3.72 (a) $f(x) = \frac{1}{10}$, for $x = 1, 2, \ldots, 10$.

(b) A c.d.f. plot is shown next.



3.73 $P(X \geq 3) = \frac{1}{2}\int_3^\infty e^{-y/2} = e^{-3/2} = 0.2231$.

3.74 (a) $f(x) \geq 0$ and $\int_0^{10} \frac{1}{10}\ dx = 1$. This is a continuous uniform distribution.

(b) $P(X \leq 7) = \frac{1}{10}\int_0^7\ dx = 0.7$.

3.75 (a) $f(y) \geq 0$ and $\int_0^1 f(y)\ dy = 10\int_0^1 (1-y)^9\ dy = -\ \frac{10}{10}(1-y)^{10}\Big|_0^1 = 1$.

(b) $P(Y > 0.6) = \int_{0.6}^1 f(y)\ dy = -\ (1-y)^{10}\Big|_{0.6}^1 = (1-0.6)^{10} = 0.0001$.

3.76 (a) $P(Z > 20) = \frac{1}{10}\int_{20}^\infty e^{-z/10}\ dz = -\ e^{-z/10}\Big|_{20}^\infty = e^{-20/10} = 0.1353$.

(b) $P(Z \leq 10) = -\ e^{-z/10}\Big|_0^{10} = 1 - e^{-10/10} = 0.6321$.

3.77 (a) $g(x_1) = \int_{x_1}^1 2\ dx_2 = 2(1 - x_1)$, for $0 < x_1 < 1$.

(b) $h(x_2) = \int_0^{x_2} 2\ dx_1 = 2x_2$, for $0 < x_2 < 1$.

(c) $P(X_1 < 0.2, X_2 > 0, 5) = \int_{0.5}^1 \int_0^{0.2} 2\ dx_1\ dx_2 = 2(1 - 0.5)(0.2 - 0) = 0.2$.

(d) $f_{X_1|X_2}(x_1|x_2) = \frac{f(x_1,x_2)}{h(x_2)} = \frac{2}{2x_2} = \frac{1}{x_2}$, for $0 < x_1 < x_2$.

3.78 (a) $f_{X_1}(x_1) = \int_0^{x_1} 6x_2\ dx_2 = 3x_1^2$, for $0 < x_1 < 1$. Apparently, $f_{X_1}(x_1) \geq 0$ and $\int_0^1 f_{X_1}(x_1)\ dx_1 = \int_0^1 3x_1^2\ dx_1 = 1$. So, $f_{X_1}(x_1)$ is a density function.

(b) $f_{X_2|X_1}(x_2|x_1) = \frac{f(x_1,x_2)}{f_{X_1}(x_1)} = \frac{6x_2}{3x_1^2} = 2\frac{x_2}{x_1^2}$, for $0 < x_2 < x_1$.
So, $P(X_2 < 0.5 \mid X_1 = 0.7) = \frac{2}{0.7^2}\int_0^{0.5} x_2\ dx_2 = \frac{25}{49}$.

3.79 (a) $g(x) = \frac{9}{(16)4^y} \sum_{x=0}^{\infty} \frac{1}{4^x} = \frac{9}{(16)4^y} \frac{1}{1-1/4} = \frac{3}{4} \cdot \frac{1}{4^x}$, for $x = 0, 1, 2, \ldots$; similarly, $h(y) = \frac{3}{4} \cdot \frac{1}{4^y}$, for $y = 0, 1, 2, \ldots$. Since $f(x, y) = g(x)h(y)$, X and Y are independent.

(b) $P(X + Y < 4) = f(0, 0) + f(0, 1) + f(0, 2) + f(0, 3) + f(1, 0) + f(1, 1) + f(1, 2) + f(2, 0) + f(2, 1) + f(3, 0) = \frac{9}{16}\left(1 + \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \frac{1}{r^2} + \frac{1}{4^3} + \frac{1}{4^3}\right) = \frac{9}{16}\left(1 + \frac{2}{4} + \frac{3}{4^2} + \frac{4}{4^3}\right) = \frac{63}{64}$.

3.80 P(the system works) $= P$(all components work) $= (0.95)(0.99)(0.92) = 0.86526$.

3.81 P(the system does not fail) $= P$(at least one of the components works) $= 1 - P$(all components fail) $= 1 - (1 - 0.95)(1 - 0.94)(1 - 0.90)(1 - 0.97) = 0.999991$.

3.82 Denote by X the number of components (out of 5) work.
Then, P(the system is operational) $= P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5) = \binom{5}{3}(0.92)^3(1 - 0.92)^2 + \binom{5}{4}(0.92)^4(1 - 0.92) + \binom{5}{5}(0.92)^5 = 0.9955$.

Chapter 4

Mathematical Expectation

4.1 $E(X) = \frac{1}{\pi a^2} \int_{-a}^{a} \int_{-\sqrt{a^2-y^2}}^{\sqrt{a^2-y^2}} x \; dx \; dy = \frac{1}{\pi a^2} \left[\left(\frac{a^2-y^2}{2} \right) - \left(\frac{a^2-y^2}{2} \right) \right] \; dy = 0.$

4.2 $E(X) = \sum_{x=0}^{3} x \; f(x) = (0)(27/64) + (1)(27/64) + (2)(9/64) + (3)(1/64) = 3/4.$

4.3 $\mu = E(X) = (20)(1/5) + (25)(3/5) + (30)(1/5) = 25$ cents.

4.4 Assigning wrights of $3w$ and w for a head and tail, respectively. We obtain $P(H) = 3/4$ and $P(T) = 1/4$. The sample space for the experiment is $S = \{HH, HT, TH, TT\}$. Now if X represents the number of tails that occur in two tosses of the coin, we have

$$\begin{aligned} P(X = 0) &= P(HH) = (3/4)(3/4) = 9/16, \\ P(X = 1) &= P(HT) + P(TH) = (2)(3/4)(1/4) = 3/8, \\ P(X = 2) &= P(TT) = (1/4)(1/4) = 1/16. \end{aligned}$$

The probability distribution for X is then

x	0	1	2
$f(x)$	9/16	3/8	1/16

from which we get $\mu = E(X) = (0)(9/16) + (1)(3/8) + (2)(1/16) = 1/2$.

4.5 $\mu = E(X) = (0)(0.41) + (1)(0.37) + (2)(0.16) + (3)(0.05) + (4)(0.01) = 0.88.$

4.6 $\mu = E(X) = (\$7)(1/12) + (\$9)(1/12) + (\$11)(1/4) + (\$13)(1/4) + (\$15)(1/6) + (\$17)(1/6) = \$12.67.$

4.7 Expected gain $= E(X) = (4000)(0.3) + (-1000)(0.7) = \$500$.

4.8 Let $X =$ profit. Then

$$\mu = E(X) = (250)(0.22) + (150)(0.36) + (0)(0.28) + (-150)(0.14) = \$88.$$

4.9 Let $c =$ amount to play the game and $Y =$ amount won.

y	$5-c$	$3-c$	$-c$
$f(y)$	2/13	2/13	9/13

$E(Y) = (5-c)(2/13) + (3-c)(2/13) + (-c)(9/13) = 0$. So, $13c = 16$ which implies $c = \$1.23$.

4.10 $\mu_X = \sum xg(x) = (1)(0.17) + (2)(0.5) + (3)(0.33) = 2.16$,
$\mu_Y = \sum yh(y) = (1)(0.23) + (2)(0.5) + (3)(0.27) = 2.04$.

4.11 For the insurance of \$200,000 pilot, the distribution of the claim the insurance company would have is as follows:

Claim Amount	\$200,000	\$100,000	\$50,000	0
$f(x)$	0.002	0.01	0.1	0.888

So, the expected claim would be

$$(\$200,000)(0.002) + (\$100,000)(0.01) + (\$50,000)(0.1) + (\$0)(0.888) = \$6,400.$$

Hence the insurance company should charge a premium of $\$6,400 + \$500 = \$6,900$.

4.12 $E(X) = \int_0^1 2x(1-x)\ dx = 1/3$. So, $(1/3)(\$5,000) = \$1,667.67$.

4.13 $E(X) = \frac{4}{\pi}\int_0^1 \frac{x}{1+x^2}\ dx = \frac{\ln 4}{\pi}$.

4.14 $E(X) = \int_0^1 \frac{2x(x+2)}{5}\ dx = \frac{8}{15}$.

4.15 $E(X) = \int_0^1 x^2\ dx + \int_1^2 x(2-x)\ dx = 1$. Therefore, the average number of hours per year is $(1)(100) = 100$ hours.

4.16 $P(X_1 + X_2 = 1) = P(X_1 = 1, X_2 = 0) + P(X_1 = 0, X_2 = 1)$
$= \frac{\binom{980}{1}\binom{20}{1}}{\binom{1000}{2}} + \frac{\binom{980}{1}\binom{20}{1}}{\binom{1000}{2}} = (2)(0.0392) = 0.0784.$

4.17 The probability density function is,

x	-3	6	9
$f(x)$	1/6	1/2	1/3
$g(x)$	25	169	361

$\mu_{g(X)} = E[(2X+1)^2] = (25)(1/6) + (169)(1/2) + (361)(1/3) = 209.$

4.18 $E(X^2) = (0)(27/64) + (1)(27/64) + (4)(9/64) + (9)(1/64) = 9/8$.

4.19 Let $Y = 1200X - 50X^2$ be the amount spent.

x	0	1	2	3
$f(x)$	1/10	3/10	2/5	1/5
$y = g(x)$	0	1150	2200	3150

$\mu_Y = E(1200X - 50X^2) = (0)(1/10) + (1150)(3/10) + (2200)(2/5) + (3150)(1/5)$
$= \$1,855.$

4.20 $E[g(X)] = E(e^{2X/3}) = \int_0^\infty e^{2x/3}e^{-x}\ dx = \int_0^\infty e^{-x/3}\ dx = 3.$

4.21 $E(X^2) = \int_0^1 2x^2(1-x)\ dx = \frac{1}{6}$. Therefore, the average profit per new automobile is $(1/6)(\$5000.00) = \$833.33$.

4.22 $E(Y) = E(X+4) = \int_0^\infty 32(x+4)\frac{1}{(x+4)^3}\ dx = 8$ days.

4.23 (a) $E[g(X,Y)] = E(XY^2) = \sum_x \sum_y xy^2 f(x,y)$
$= (2)(1)^2(0.10) + (2)(3)^2(0.20) + (2)(5)^2(0.10) + (4)(1)^2(0.15) + (4)(3)^2(0.30)$
$+ (4)(5)^2(0.15) = 35.2.$

(b) $\mu_X = E(X) = (2)(0.40) + (4)(0.60) = 3.20,$
$\mu_Y = E(Y) = (1)(0.25) + (3)(0.50) + (5)(0.25) = 3.00.$

4.24 (a) $E(X^2Y - 2XY) = \sum_{x=0}^{3}\sum_{y=0}^{2}(x^2y - 2xy)f(x,y) = (1-2)(18/70) + (4-4)(18/70) +$
$\cdots + (8-8)(3/70) = -3/7.$

(b)

x	0	1	2	3
$g(x)$	5/70	30/70	30/70	5/70

y	0	1	2
$h(y)$	15/70	40/70	15/70

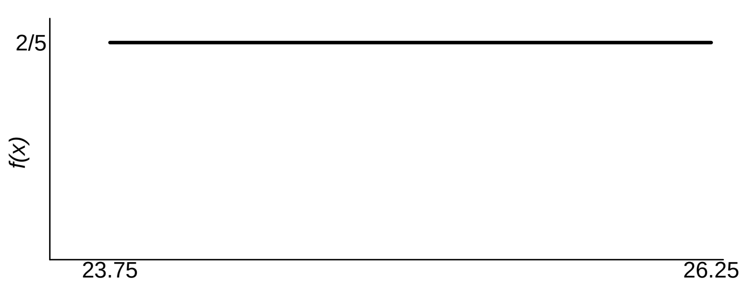
$\mu_X = E(X) = (0)(5/70) + (1)(30/70) + (2)(30/70) + (3)(5/70) = 3/2,$
$\mu_Y = E(Y) = (0)(15/70) + (1)(40/70) + (2)(15/70) = 1.$

4.25 $\mu_{X+Y} = E(X+Y) = \sum_{x=0}^{3}\sum_{y=0}^{3}(x+y)f(x,y) = (0+0)(1/55) + (1+0)(6/55) + \cdots + (0+3)(1/55) = 2.$

4.26 $E(Z) = E(\sqrt{X^2+Y^2}) = \int_0^1\int_0^1 4xy\sqrt{x^2+y^2}\ dx\ dy = \frac{4}{3}\int_0^1 [y(1+y^2)^{3/2} - y^4]\ dy$
$= 8(2^{3/2} - 1)/15 = 0.9752.$

4.27 $E(X) = \frac{1}{2000}\int_0^\infty x\exp(-x/2000)\ dx = 2000\int_0^\infty y\exp(-y)\ dy = 2000.$

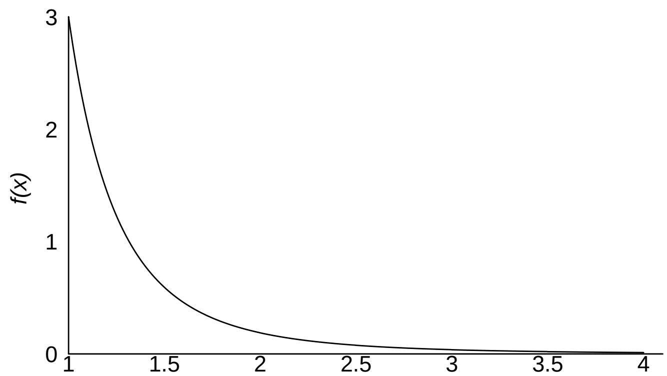
4.28 (a) The density function is shown next.



(b) $E(X) = \frac{2}{5}\int_{23.75}^{26.25} x\ dx = \frac{1}{5}(26.25^2 - 23.75^2) = 25.$

(c) The mean is exactly in the middle of the interval. This should not be surprised due to the symmetry of the density at 25.

4.29 (a) The density function is shown next



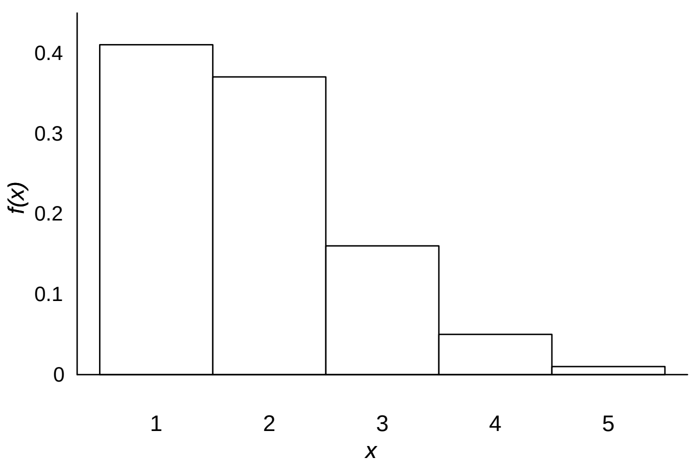
(b) $\mu = E(X) = \int_1^\infty 3x^{-3}\ dx = \frac{3}{2}.$

4.30 $E(Y) = \frac{1}{4}\int_0^\infty ye^{-y/4}\ dy = 4.$

4.31 (a) $\mu = E(Y) = 5\int_0^1 y(1-y)^4\ dy = -\int_0^1 y\ d(1-y)^5 = \int_0^\infty (1-y)^5\ dy = \frac{1}{6}.$

(b) $P(Y > 1/6) = \int_{1/6}^1 5(1-y)^4\ dy = -\left.(1-y)^5\right|_{1/6}^1 = (1-1/6)^5 = 0.4019.$

4.32 (a) A histogram is shown next.



(b) $\mu = (0)(0.41) + (1)(0.37) + (2)(0.16) + (3)(0.05) + (4)(0.01) = 0.88.$

(c) $E(X^2) = (0)^2(0.41) + (1)^2(0.37) + (2)^2(0.16) + (3)^2(0.05) + (4)^2(0.01) = 1.62.$

(d) $Var(X) = 1.62 - 0.88^2 = 0.8456.$

4.33 $\mu = \$500$. So, $\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x) = (-1500)^2(0.7) + (3500)^2(0.3) = \$5,250,000.$

4.34 $\mu = (-2)(0.3) + (3)(0.2) + (5)(0.5) = 2.5$ and
$E(X^2) = (-2)^2(0.3) + (3)^2(0.2) + (5)^2(0.5) = 15.5$.
So, $\sigma^2 = E(X^2) - \mu^2 = 9.25$ and $\sigma = 3.041$.

4.35 $\mu = (2)(0.01) + (3)(0.25) + (4)(0.4) + (5)(0.3) + (6)(0.04) = 4.11$,
$E(X^2) = (2)^2(0.01) + (3)^2(0.25) + (4)^2(0.4) + (5)^2(0.3) + (6)^2(0.04) = 17.63$.
So, $\sigma^2 = 17.63 - 4.11^2 = 0.74$.

4.36 $\mu = (0)(0.4) + (1)(0.3) + (2)(0.2) + (3)(0.1) = 1.0$,
and $E(X^2) = (0)^2(0.4) + (1)^2(0.3) + (2)^2(0.2) + (3)^2(0.1) = 2.0$.
So, $\sigma^2 = 2.0 - 1.0^2 = 1.0$.

4.37 It is know $\mu = 1/3$.
So, $E(X^2) = \int_0^1 2x^2(1-x)\ dx = 1/6$ and $\sigma^2 = 1/6 - (1/3)^2 = 1/18$. So, in the actual profit, the variance is $\frac{1}{18}(5000)^2$.

4.38 It is known $\mu = 8/15$.
Since $E(X^2) = \int_0^1 \frac{2}{5}x^2(x+2)\ dx = \frac{11}{30}$, then $\sigma^2 = 11/30 - (8/15)^2 = 37/450$.

4.39 It is known $\mu = 1$.
Since $E(X^2) = \int_0^1 x^2\ dx + \int_1^2 x^2(2-x)\ dx = 7/6$, then $\sigma^2 = 7/6 - (1)^2 = 1/6$.

4.40 $\mu_{g(X)} = E[g(X)] = \int_0^1 (3x^2+4)\left(\frac{2x+4}{5}\right)\ dx = \frac{1}{5}\int_0^1 (6x^3 + 12x^2 + 8x + 16)\ dx = 5.1$.
So, $\sigma^2 = E[g(X) - \mu]^2 = \int_0^1 (3x^2 + 4 - 5.1)^2 \left(\frac{2x+4}{5}\right)\ dx$
$= \int_0^1 (9x^4 - 6.6x^2 + 1.21)\left(\frac{2x+4}{5}\right)\ dx = 0.83$.

4.41 It is known $\mu_{g(X)} = E[(2X+1)^2] = 209$. Hence
$\sigma^2_{g(X)} = \sum_x [(2X+1)^2 - 209]^2 g(x)$
$= (25 - 209)^2(1/6) + (169 - 209)^2(1/2) + (361 - 209)^2(1/3) = 14,144$.
So, $\sigma_{g(X)} = \sqrt{14,144} = 118.9$.

4.42 It is known $\mu_{g(X)} = E(X^2) = 1/6$. Hence
$\sigma^2_{g(X)} = \int_0^1 2\left(x^2 - \frac{1}{6}\right)^2 (1-x)\ dx = 7/180$.

4.43 $\mu_Y = E(3X - 2) = \frac{1}{4}\int_0^\infty (3x-2)e^{-x/4}\ dx = 10$. So
$\sigma^2_Y = E\{[(3X-2) - 10]^2\} = \frac{9}{4}\int_0^\infty (x-4)^2 e^{-x/4}\ dx = 144$.

4.44 $E(XY) = \sum_x \sum_y xyf(x,y) = (1)(1)(18/70) + (2)(1)(18/70)$
$+ (3)(1)(2/70) + (1)(2)(9/70) + (2)(2)(3/70) = 9/7$;
$\mu_X = \sum_x \sum_y xf(x,y) = (0)f(0,1) + (0)f(0,2) + (1)f(1,0) + \cdots + (3)f(3,1) = 3/2$,
and $\mu_Y = 1$.
So, $\sigma_{XY} = E(XY) - \mu_X\mu_Y = 9/7 - (3/2)(1) = -3/14$.

4.45 $\mu_X = \sum_x xg(x) = 2.45$, $\mu_Y = \sum_y yh(y) = 2.10$, and
$E(XY) = \sum_x \sum_x xyf(x,y) = (1)(0.05) + (2)(0.05) + (3)(0.10) + (2)(0.05)$
$+ (4)(0.10) + (6)(0.35) + (3)(0) + (6)(0.20) + (9)(0.10) = 5.15$.
So, $\sigma_{XY} = 5.15 - (2.45)(2.10) = 0.005$.

4.46 From previous exercise, $k = \left(\frac{3}{392}\right)10^{-4}$, and $g(x) = k\left(20x^2 + \frac{98000}{3}\right)$, with
$\mu_X = E(X) = \int_{30}^{50} xg(x)\ dx = k\int_{30}^{50}\left(20x^3 + \frac{98000}{3}x\right)\ dx = 40.8163$.
Similarly, $\mu_Y = 40.8163$. On the other hand,
$E(XY) = k\int_{30}^{50}\int_{30}^{50} xy(x^2+y^2)\ dy\ dx = 1665.3061$.
Hence, $\sigma_{XY} = E(XY) - \mu_X\mu_Y = 1665.3061 - (40.8163)^2 = -0.6642$.

4.47 $g(x) = \frac{2}{3}\int_0^1 (x+2y)\ dy = \frac{2}{3}(x+1$, for $0 < x < 1$, so $\mu_X = \frac{2}{3}\int_0^1 x(x+1)\ dx = \frac{5}{9}$;
$h(y) = \frac{2}{3}\int_0^1 (x+2y)\ dx = \frac{2}{3}\left(\frac{1}{2}+2y\right)$, so $\mu_Y = \frac{2}{3}\int_0^1 y\left(\frac{1}{2}+2y\right)\ dy = \frac{11}{18}$; and
$E(XY) = \frac{2}{3}\int_0^1\int_0^1 xy(x+2y)\ dy\ dx = \frac{1}{3}$.
So, $\sigma_{XY} = E(XY) - \mu_X\mu_Y = \frac{1}{3} - \left(\frac{5}{9}\right)\left(\frac{11}{18}\right) = -0.0062$.

4.48 Since $\sigma_{XY} = Cov(a+bX, X) = b\sigma_X^2$ and $\sigma_Y^2 = b^2\sigma_X^2$, then
$\rho = \frac{\sigma_{XY}}{\sigma_X}\sigma_Y = \frac{b\sigma_X^2}{\sqrt{\sigma_X^2 b^2\sigma_X^2}} = \frac{b}{|b|} =$ sign of b.
Hence $\rho = 1$ if $b > 0$ and $\rho = -1$ if $b < 0$.

4.49 $E(X) = (0)(0.41) + (1)(0.37) + (2)(0.16) + (3)(0.05) + (4)(0.01) = 0.88$
and $E(X^2) = (0)^2(0.41) + (1)^2(0.37) + (2)^2(0.16) + (3)^2(0.05) + (4)^2(0.01) = 1.62$.
So, $Var(X) = 1.62 - 0.88^2 = 0.8456$ and $\sigma = \sqrt{0.8456} = 0.9196$.

4.50 $E(X) = 2\int_0^1 x(1-x)\ dx = 2\left(\frac{x^2}{2} - \frac{x^3}{3}\right)\Big|_0^1 = \frac{1}{3}$ and
$E(X^2) = 2\int_0^1 x^2(1-x)\ dx = 2\left(\frac{x^3}{3} - \frac{x^4}{4}\right)\Big|_0^1 = \frac{1}{6}$. Hence,
$Var(X) = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}$, and $\sigma = \sqrt{1/18} = 0.2357$.

4.51 Previously we found $\mu = 4.11$ and $\sigma^2 = 0.74$, Therefore,
$\mu_{g(X)} = E(3X-2) = 3\mu - 2 = (3)(4.11) - 2 = 10.33$ and $\sigma_{g(X)} = 9\sigma^2 = 6.66$.

4.52 Previously we found $\mu = 1$ and $\sigma^2 = 1$. Therefore,
$\mu_{g(X)} = E(5X+3) = 5\mu + 3 = (5)(1) + 3 = 8$ and $\sigma_{g(X)} = 25\sigma^2 = 25$.

4.53 Let X = number of cartons sold and Y = profit.
We can write $Y = 1.65X + (0.90)(5-X) - 6 = 0.75X - 1.50$. Now
$E(X) = (0)(1/15)+(1)(2/15)+(2)(2/15)+(3)(3/15)+(4)(4/15)+(5)(3/15) = 46/15$,
and $E(Y) = (0.75)E(X) - 1.50 = (0.75)(46/15) - 1.50 = \$0.80$.

4.54 $\mu_X = E(X) = \frac{1}{4}\int_0^\infty xe^{-x/4}\ dx = 4$.
Therefore, $\mu_Y = E(3X-2) = 3E(X) - 2 = (3)(4) - 2 = 10$.
Since $E(X^2) = \frac{1}{4}\int_0^\infty x^2e^{-x/4}\ dx = 32$, therefore, $\sigma_X^2 = E(X^2) - \mu_X^2 = 32 - 16 = 16$.
Hence $\sigma_Y^2 = 9\sigma_X^2 = (9)(16) = 144$.

4.55 $E(X) = (-3)(1/6) + (6)(1/2) + (9)(1/3) = 11/2$,
$E(X^2) = (-3)^2(1/6) + (6)^2(1/2) + (9)^2(1/3) = 93/2$. So,
$E[(2X+1)^2] = 4E(X^2) + 4E(X) + 1 = (4)(93/2) + (4)(11/2) + 1 = 209$.

4.56 Since $E(X) = \int_0^1 x^2\, dx + \int_1^2 x(2-x)\, dx = 1$, and
$E(X^2) = \int_0^1 x^3 2\, dx + \int_1^2 x^2(2-x)\, dx = 7/6$,then
$E(Y) = 60E(X^2) + 39E(X) = (60)(7/6) + (39)(1) = 109$ kilowatt hours.

4.57 The equations $E[(X-1)^2] = 10$ and $E[(X-2)^2] = 6$ may be written in the form:

$$E(X^2) - 2E(X) = 9, \qquad E(X^2) - 4E(X) = 2.$$

Solving these two equations simultaneously we obtain

$$E(X) = 7/2, \quad \text{and} \quad E(X^2) = 16.$$

Hence $\mu = 7/2$ and $\sigma^2 = 16 - (7/2)^2 = 15/4$.

4.58 $E(X) = (2)(0.40) + (4)(0.60) = 3.20$, and
$E(Y) = (1)(0.25) + (3)(0.50) + (5)(0.25) = 3$. So,

(a) $E(2X - 3Y) = 2E(X) - 3E(Y) = (2)(3.20) - (3)(3.00) = -2.60$.

(b) $E(XY) = E(X)E(Y) = (3.20)(3.00) = 9.60$.

4.59 $E(2XY^2 - X^2Y) = 2E(XY^2) - E(X^2Y)$. Now,
$E(XY^2) = \sum_{x=0}^{2}\sum_{y=0}^{2} xy^2 f(x,y) = (1)(1)^2(3/14) = 3.14$, and
$E(X^2Y) = \sum_{x=0}^{2}\sum_{y=0}^{2} x^2 y f(x,y) = (1)^2(1)(3/14) = 3.14$.
Therefore, $E(2XY^2 - X^2Y) = (2)(3/14) - (3/14) = 3/14$.

4.60 Using $\mu = 60$ and $\sigma = 6$ and Chebyshev's theorem

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2},$$

since from $\mu + k\sigma = 84$ we obtain $k = 4$.
So, $P(X < 84) \geq P(36 < X < 84) \geq 1 - \frac{1}{4^2} = 0.9375$. Therefore,

$$P(X \geq 84) \leq 1 - 0.9375 = 0.0625.$$

Since $1000(0.0625) = 62.5$, we claim that at most 63 applicants would have a score as 84 or higher. Since there will be 70 positions, the applicant will have the job.

4.61 $\mu = 900$ hours and $\sigma = 50$ hours. Solving $\mu - k\sigma = 700$ we obtain $k = 4$.
So, using Chebyshev's theorem with $P(\mu - 4\sigma < X < \mu + 4\sigma) \geq 1 - 1/4^2 = 0.9375$,
we obtain $P(700 < X < 1100) \geq 0.9375$. Therefore, $P(X \leq 700) \leq 0.03125$.

4.62 $\mu = 52$ and $\sigma = 6.5$. Solving $\mu + k\sigma = 71.5$ we obtain $k = 3$. So,

$$P(\mu - 3\sigma < X < \mu + 3\sigma) \geq 1 - \frac{1}{3^2} = 0.8889,$$

which is

$$P(32.5 < X < 71.5) \geq 0.8889.$$

we obtain $P(X > 71.5) < \frac{1-0.8889}{2} = 0.0556$ using the symmetry.

4.63 $n = 500$, $\mu = 4.5$ and $\sigma = 2.8733$. Solving $\mu + k(\sigma/\sqrt{500}) = 5$ we obtain

$$k = \frac{5 - 4.5}{2.87333/\sqrt{500}} = \frac{0.5}{0.1284} = 3.8924.$$

So, $P(4 \leq \bar{X} \leq 5) \geq 1 - \frac{1}{k^2} = 0.9340$.

4.64 $\sigma_Z^2 = \sigma_{-2X+4Y-3}^2 = 4\sigma_X^2 + 16\sigma_Y^2 = (4)(5) + (16)(3) = 68$.

4.65 $\sigma_Z^2 = \sigma_{-2X+4Y-3}^2 = 4\sigma_X^2 + 16\sigma_Y^2 - 16\sigma_{XY} = (4)(5) + (16)(3) - (16)(1) = 52$.

4.66 (a) $P(6 < X < 18) = P[12 - (2)(3) < X < 12 + (2)(3)] \geq 1 - \frac{1}{2^2} = \frac{3}{4}$.

(b) $P(3 < X < 21) = P[12 - (3)(3) < X < 12 + (3)(3)] \geq 1 - \frac{1}{3^2} = \frac{8}{9}$.

4.67 (a) $P(|X - 10| \geq 3) = 1 - P(|X - 10| < 3)$
$= 1 - P[10 - (3/2)(2) < X < 10 + (3/2)(2)] \leq 1 - \left[1 - \frac{1}{(3/2)^2}\right] = \frac{4}{9}$.

(b) $P(|X - 10| < 3) = 1 - P(|X - 10| \geq 3) \geq 1 - \frac{4}{9} = \frac{5}{9}$.

(c) $P(5 < X < 15) = P[10 - (5/2)(2) < X < 10 + (5/2)(2)] \geq 1 - \frac{1}{(5/2)^2} = \frac{21}{25}$.

(d) $P(|X - 10| \geq c) \leq 0.04$ implies that $P(|X - 10| < c) \geq 1 - 0.04 = 0.96$.
Solving $0.96 = 1 - \frac{1}{k^2}$ we obtain $k = 5$. So, $c = k\sigma = (5)(2) = 10$.

4.68 $\mu = E(X) = 6\int_0^1 x^2(1 - x)\ dx = 0.5$, $E(X^2) = 6\int_0^1 x^3(1 - x)\ dx = 0.3$, which imply $\sigma^2 = 0.3 - (0.5)^2 = 0.05$ and $\sigma = 0.2236$. Hence,

$$\begin{aligned} P(\mu - 2\sigma < X < \mu + 2\sigma) &= P(0.5 - 0.4472 < X < 0.5 + 0.4472) \\ &= P(0.0528 < X < 0.9472) = 6\int_{0.0528}^{0.9472} x(1 - x)\ dx = 0.9839, \end{aligned}$$

compared to a probability of at least 0.75 given by Chebyshev's theorem.

4.69 It is easy to see that the expectations of X and Y are both 3.5. So,

(a) $E(X + Y) = E(X) + E(Y) = 3.5 + 3.5 = 7.0$.

(b) $E(X - Y) = E(X) - E(Y) = 0$.

(c) $E(XY) = E(X)E(Y) = (3.5)(3.5) = 12.25$.

4.70 $E(Z) = E(XY) = E(X)E(Y) = \int_0^1 \int_2^\infty 16xy(y/x^3)\ dx\ dy = 8/3$.

4.71 $E[g(X,Y)] = E(X/Y^3 + X^2Y) = E(X/Y^3) + E(X^2Y)$.
$E(X/Y^3) = \int_1^2 \int_0^1 \frac{2x(x+2y)}{7y^3}\ dx\ dy = \frac{2}{7}\int_1^2 \left(\frac{1}{3y^3} + \frac{1}{y^2}\right)\ dy = \frac{15}{84}$;
$E(X^2Y) = \int_1^2 \int_0^1 \frac{2x^2y(x+2y)}{7}\ dx\ dy = \frac{2}{7}\int_1^2 y\left(\frac{1}{4} + \frac{2y}{3}\right)\ dy = \frac{139}{252}$.
Hence, $E[g(X,Y)] = \frac{15}{84} + \frac{139}{252} = \frac{46}{63}$.

4.72 $\mu_X = \mu_Y = 3.5$. $\sigma_X^2 = \sigma_Y^2 = [(1)^2 + (2)^2 + \cdots + (6)^2](1/6) - (3.5)^2 = \frac{35}{12}$.

(a) $\sigma_{2X-Y} = 4\sigma_X^2 + \sigma_Y^2 = \frac{175}{12}$;

(b) $\sigma_{X+3Y-5} = \sigma_X^2 + 9\sigma_Y^2 = \frac{175}{6}$.

4.73 (a) $\mu = \frac{1}{5}\int_0^5 x\ dx = 2.5$, $\sigma^2 = E(X^2) - \mu^2 = \frac{1}{5}x^2 \int_0^5 x^2\ dx - 2.5^2 = 2.08$.
So, $\sigma = \sqrt{\sigma^2} = 1.44$.

(b) By Chebyshev's theorem,

$$P[2.5 - (2)(1.44) < X < 2.5 + (2)(1.44)] = P(-0.38 < X < 5.38) \geq 0.75.$$

Using integration, $P(-0.38 < X < 5.38) = 1 \geq 0.75$;

$$P[2.5 - (3)(1.44) < X < 2.5 + (3)(1.44)] = P(-1.82 < X < 6.82) \geq 0.89.$$

Using integration, $P(-1.82 < X < 6.82) = 1 \geq 0.89$.

4.74 $P = I^2R$ with $R = 50$, $\mu_I = E(I) = 15$ and $\sigma_I^2 = Var(I) = 0.03$.
$E(P) = E(I^2R) = 50E(I^2) = 50[Var(I) + \mu_I^2] = 50(0.03 + 15^2) = 11251.5$. If we use the approximation formula, with $g(I) = I^2$, $g'(I) = 2I$ and $g''(I) = 2$, we obtain,

$$E(P) \approx 50\left[g(\mu_I) + 2\frac{\sigma_I^2}{2}\right] = 50(15^2 + 0.03) = 11251.5.$$

Since $Var[g(I)] \approx \left[\frac{\partial g(i)}{\partial i}\right]_{i=\mu_I}^2 \sigma_I^2$, we obtain

$$Var(P) = 50^2 Var(I^2) = 50^2(2\mu_I)^2\sigma_I^2 = 50^2(30)^2(0.03) = 67500.$$

4.75 For $0 < a < 1$, since $g(a) = \sum\limits_{x=0}^{\infty} a^x = \frac{1}{1-a}$, $g'(a) = \sum\limits_{x=1}^{\infty} xa^{x-1} = \frac{1}{(1-a)^2}$ and
$g''(a) = \sum\limits_{x=2}^{\infty} x(x-1)a^{x-2} = \frac{2}{(1-a)^3}$.

(a) $E(X) = (3/4)\sum_{x=1}^{\infty} x(1/4)^x = (3/4)(1/4)\sum_{x=1}^{\infty} x(1/4)^{x-1} = (3/16)[1/(1-1/4)^2]$
$= 1/3$, and $E(Y) = E(X) = 1/3$.
$E(X^2) - E(X) = E[X(X-1)] = (3/4)\sum_{x=2}^{\infty} x(x-1)(1/4)^x$
$= (3/4)(1/4)^2 \sum_{x=2}^{\infty} x(x-1)(1/4)^{x-2} = (3/4^3)[2/(1-1/4)^3] = 2/9$.
So, $Var(X) = E(X^2) - [E(X)]^2 = [E(X^2) - E(X)] + E(X) - [E(X)]^2$
$2/9 + 1/3 - (1/3)^2 = 4/9$, and $Var(Y) = 4/9$.

(b) $E(Z) = E(X) + E(Y) = (1/3) + (1/3) = 2/3$, and
$Var(Z) = Var(X+Y) = Var(X) + Var(Y) = (4/9) + (4/9) = 8/9$, since X and Y are independent (from Exercise 3.79).

4.76 (a) $g(x) = \frac{3}{2}\int_0^1 (x^2 + y^2)\, dy = \frac{1}{2}(3x^2 + 1)$ for $0 < x < 1$ and
$h(y) = \frac{1}{2}(3y^2 + 1)$ for $0 < y < 1$.
Since $f(x, y) \neq g(x)h(y)$, X and Y are not independent.

(b) $E(X+Y) = E(X) + E(Y) = 2E(X) = \int_0^1 x(3x^2 + 1)\, dx = 3/4 + 1/2 = 5/4$.
$E(XY) = \frac{3}{2}\int_0^1\int_0^1 xy(x^2+y^2)\, dx\, dy = \frac{3}{2}\int_0^1 y\left(\frac{1}{4} + \frac{y^2}{2}\right)\, dy$
$= \frac{3}{2}\left[\left(\frac{1}{4}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{4}\right)\right] = \frac{3}{8}$.

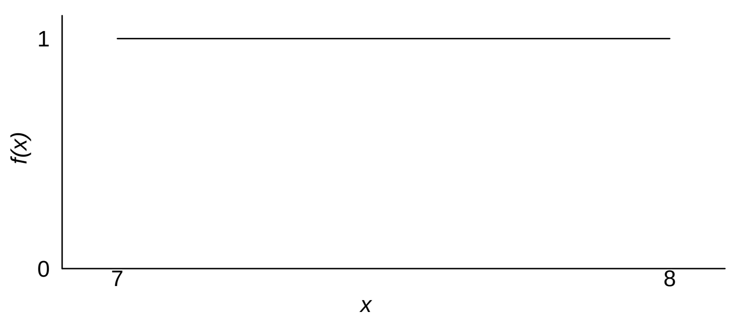
(c) $Var(X) = E(X^2) - [E(X)]^2 = \frac{1}{2}\int_0^1 x^2(3x^2+1)\, dx - \left(\frac{5}{8}\right)^2 = \frac{7}{15} - \frac{25}{64} = \frac{73}{960}$, and
$Var(Y) = \frac{73}{960}$. Also, $Cov(X, Y) = E(XY) - E(X)E(Y) = \frac{3}{8} - \left(\frac{5}{8}\right)^2 = -\frac{1}{64}$.

(d) $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X, Y) = 2\frac{73}{960} - 2\frac{1}{64} = \frac{29}{240}$.

4.77 (a) $E(Y) = \int_0^\infty ye^{-y/4}\, dy = 4$.

(b) $E(Y^2) = \int_0^\infty y^2 e^{-y/4}\, dy = 32$ and $Var(Y) = 32 - 4^2 = 16$.

4.78 (a) The density function is shown next.



(b) $E(Y) = \int_7^8 y\, dy = \frac{1}{2}[8^2 - 7^2] = \frac{15}{2} = 7.5$,
$E(Y^2) = \int_7^8 y^2\, dy = \frac{1}{3}[8^3 - 7^3] = \frac{169}{3}$, and $Var(Y) = \frac{169}{3} - \left(\frac{15}{2}\right)^2 = \frac{1}{12}$.

4.79 Using the exact formula, we have

$$E(e^Y) = \int_7^8 e^y\, dy = e^y|_7^8 = 1884.32.$$

Using the approximation, since $g(y) = e^y$, so $g''(y) = e^y$. Hence, using the approximation formula,

$$E(e^Y) \approx e^{\mu_Y} + e^{\mu_Y}\frac{\sigma_Y^2}{2} = \left(1 + \frac{1}{24}\right)e^{7.5} = 1883.38.$$

The approximation is very close to the true value.

4.80 Using the exact formula, $E(Z^2) = \int_7^8 e^{2y}\, dy = \frac{1}{2}\, e^{2y}|_7^8 = 3841753.12$. Hence,

$$Var(Z) = E(Z^2) - [E(Z)]^2 = 291091.3.$$

Using the approximation formula, we have

$$Var(e^Y) = (e^{\mu_Y})^2 Var(Y) = \frac{e^{(2)(7.5)}}{12} = 272418.11.$$

The approximation is not so close to each other. One reason is that the first order approximation may not always be good enough.

4.81 Define $I_1 = \{x_i |\ |x_i - \mu| < k\sigma\}$ and $I_2 = \{x_i |\ |x_i - \mu| \geq k\sigma\}$. Then

$$\begin{aligned}\sigma^2 = E[(X-\mu)^2] &= \sum_x (x-\mu)^2 f(x) = \sum_{x_i \in I_1}(x_i-\mu)^2 f(x_i) + \sum_{x_i \in I_2}(x_i-\mu)^2 f(x_i)\\ &\geq \sum_{x_i \in I_2}(x_i-\mu)^2 f(x_i) \geq k^2\sigma^2 \sum_{x_i \in I_2} f(x_i) = k^2\sigma^2 P(|X-\mu| \geq k\sigma),\end{aligned}$$

which implies

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Hence, $P(|X-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}$.

4.82 $E(XY) = \int_0^1 \int_0^1 xy(x+y)\, dx\, dy = \frac{1}{3}$, $E(X) = \int_0^1 \int_0^1 x(x+y)\, dx\, dy = \frac{7}{12}$ and $E(Y) = \frac{7}{12}$. Therefore, $\sigma_{XY} = E(XY) - \mu_X\mu_Y = \frac{1}{3} - \left(\frac{7}{12}\right)^2 = -\frac{1}{144}$.

4.83 $E(Y-X) = \int_0^1 \int_0^y 2(y-x)\, dx\, dy = \int_0^1 y^2\, dy = \frac{1}{3}$. Therefore, the average amount of kerosene left in the tank at the end of each day is $(1/3)(1000) = 333$ liters.

4.84 (a) $E(X) = \int_0^\infty \frac{x}{5} e^{-x/5}\, dx = 5$.

(b) $E(X^2) = \int_0^\infty \frac{x^2}{5} e^{-x/5}\, dx = 50$, so $Var(X) = 50 - 5^2 = 25$, and $\sigma = 5$.

(c) $E[(X+5)^2] = E\{[(X-5)+10]^2\} = E[(X-5)^2] + 10^2 + 20E(X-5)$ $= Var(X) + 100 = 125$.

4.85 $E(XY) = 24\int_0^1 \int_0^{1-y} x^2y^2\, dx\, dy = 8\int_0^1 y^2(1-y)^3\, dy = \frac{2}{15}$, $\mu_X = 24\int_0^1 \int_0^{1-y} x^2 y\, dx\, dy = \frac{2}{5}$ and $\mu_Y = 24\int_0^1 \int_0^{1-y} xy^2\, dx\, dy = \frac{2}{5}$. Therefore, $\sigma_{XY} = E(XY) - \mu_X\mu_Y = \frac{2}{15} - \left(\frac{2}{5}\right)^2 = -\frac{2}{75}$.

4.86 $E(X+Y)=\int_0^1\int_0^{1-y}24(x+y)xy\ dx\ dy=\frac{4}{5}$.

4.87 (a) $E(X)=\int_0^\infty \frac{x}{900}e^{-x/900}\ dx=900$ hours.

(b) $E(X^2)=\int_0^\infty \frac{x^2}{900}e^{-x/900}\ dx=1620000$ hours^2.

(c) $Var(X)=E(X^2)-[E(X)]^2=810000$ hours^2 and $\sigma=900$ hours.

4.88 It is known $g(x)=\frac{2}{3}(x+1)$, for $0<x<1$, and $h(y)=\frac{1}{3}(1+4y)$, for $0<y<1$.

(a) $\mu_X=\int_0^1 \frac{2}{3}x(x+1)\ dx=\frac{5}{9}$ and $\mu_Y=\int_0^1\frac{1}{3}y(1+4y)\ dy=\frac{11}{18}$.

(b) $E[(X+Y)/2]=\frac{1}{2}[E(X)+E(Y)]=\frac{7}{12}$.

4.89 $Cov(aX,bY)=E[(aX-a\mu_X)(bY-b\mu_Y)]=abE[(X-\mu_X)(Y-\mu_Y)]=abCov(X,Y)$.

4.90 It is known $\mu=900$ and $\sigma=900$. For $k=2$,

$$P(\mu-2\sigma<X<\mu+2\sigma)=P(-900<X<2700)\geq 0.75$$

using Chebyshev's theorem. On the other hand,

$$P(\mu-2\sigma<X<\mu+2\sigma)=P(-900<X<2700)=1-e^{-3}=0.9502.$$

For $k=3$, Chebyshev's theorem yields

$$P(\mu-3\sigma<X<\mu+3\sigma)=P(-1800<X<3600)\geq 0.8889,$$

while $P(-1800<X<3600)=1-e^{-4}=0.9817$.

4.91 $g(x)=\int_0^1\frac{16y}{x^3}\ dy=\frac{8}{x^3}$, for $x>2$, with $\mu_X=\int_2^\infty\frac{8}{x^2}\ dx=-\left.\frac{8}{x}\right|_2^\infty=4$,
$h(y)=\int_2^\infty\frac{16y}{x^3}\ dx=-\left.\frac{8y}{x^2}\right|_2^\infty=2y$, for $0<y<1$, with $\mu_Y=\int_0^1 2y^2=\frac{2}{3}$, and
$E(XY)=\int_2^\infty\int_0^1\frac{16y^2}{x^2}\ dy\ dx=\frac{16}{3}\int_2^\infty\frac{1}{x^2}\ dx=\frac{8}{3}$. Hence,
$\sigma_{XY}=E(XY)-\mu_X\mu_Y=\frac{8}{3}-(4)\left(\frac{2}{3}\right)=0$.

4.92 Since $\sigma_{XY}=1$, $\sigma_X^2=5$ and $\sigma_Y^2=3$, we have $\rho=\frac{\sigma_{XY}}{\sigma_X\sigma_Y}=\frac{1}{\sqrt{(5)(3)}}=0.2582$.

4.93 (a) From Exercise 4.37, we have $\sigma^2=1/18$, so $\sigma=0.2357$.

(b) Also, $\mu_X=1/3$ from Exercise 4.12. So,

$$\begin{aligned}P(\mu-2\sigma<X<\mu+2\sigma)&=P[1/3-(2)(0.2357)<X<1/3+(2)(0.2357)]\\&=P(0<X<0.8047)=\int_0^{0.8047}2(1-x)\ dx=0.9619.\end{aligned}$$

Using Chebyshev's theorem, the probability of this event should be larger than 0.75, which is true.

(c) $P(\text{profit}>\$500)=P(X>0.1)=\int_{0.1}^1 2(1-x)=0.81$.

4.94 Since $g(0)h(0) = (0.17)(0.23) \neq 0.10 = f(0,0)$, X and Y are not independent.

4.95 $E(X) = (-5000)(0.2) + (10000)(0.5) + (30000)(0.3) = \$13,000$.

4.96 (a) $f(x) = \binom{3}{x}(0.15)^x(0.85)^{3-x}$, for $x = 0, 1, 2, 3$.

x	0	1	2	3
$f(x)$	0.614125	0.325125	0.057375	0.003375

(b) $E(X) = 0.45$.

(c) $E(X^2) = 0.585$, so $Var(X) = 0.585 - 0.45^2 = 0.3825$.

(d) $P(X \leq 2) = 1 - P(X = 3) = 1 - 0.003375 = 0.996625$.

(e) 0.003375.

(f) Yes.

4.97 (a) $E(X) = (-\$15\text{k})(0.05)+(\$15\text{k})(0.15)+(\$25\text{k})(0.30)+(\$40\text{k})(0.15)+(\$50\text{k})(0.10)+(\$100\text{k})(0.05) + (\$150\text{k})(0.03) + (\$200\text{k})(0.02) = \$33.5\text{k}$.

(b) $E(X^2) = 2,697,500,000$ dollars^2. So, $\sigma = \sqrt{E(X^2) - [E(X)]^2} = \$39.689\text{k}$.

4.98 (a) $E(X) = \frac{3}{4\times 50^3}\int_{-50}^{50} x(50^2 - x^2)\ dx = 0$.

(b) $E(X^2) = \frac{3}{4\times 50^3}\int_{-50}^{50} x^2(50^2 - x^2)\ dx = 500$.

(c) $\sigma = \sqrt{E(X^2) - [E(X)]^2} = \sqrt{500 - 0} = 22.36$.

4.99 (a) The marginal density of X is

x_1	0	1	2	3	4
$f_{X_1}(x_1)$	0.13	0.21	0.31	0.23	0.12

(b) The marginal density of Y is

x_2	0	1	2	3	4
$f_{X_2}(x_2)$	0.10	0.30	0.39	0.15	0.06

(c) Given $X_2 = 3$, the conditional density function of X_1 is $f(x_1, 3)/0.15$. So

x_1	0	1	2	3	4
$f_{X_2}(x_2)$	$\frac{7}{15}$	$\frac{1}{5}$	$\frac{1}{15}$	$\frac{1}{5}$	$\frac{1}{15}$

(d) $E(X_1) = (0)(0.13) + (1)(0.21) + (2)(0.31) + (3)(0.23) + (4)(0.12) = 2$.

(e) $E(X_2) = (0)(0.10) + (1)(0.30) + (2)(0.39) + (3)(0.15) + (4)(0.06) = 1.77$.

(f) $E(X_1|X_2 = 3) = (0)\left(\frac{7}{15}\right) + (1)\left(\frac{1}{5}\right) + (2)\left(\frac{1}{15}\right) + (3)\left(\frac{1}{5}\right) + (4)\left(\frac{1}{15}\right) = \frac{18}{15} = \frac{6}{5} = 1.2$.

(g) $E(X_1^2) = (0)^2(0.13) + (1)^2(0.21) + (2)^2(0.31) + (3)^2(0.23) + (4)^2(0.12) = 5.44$. So, $\sigma_{X_1} = \sqrt{E(X_1^2) - [E(X_1)]^2} = \sqrt{5.44 - 2^2} = \sqrt{1.44} = 1.2$.

4.100 (a) The marginal densities of X and Y are, respectively,

x	0	1	2
$g(x)$	0.2	0.32	0.48

y	0	1	2
$h(y)$	0.26	0.35	0.39

The conditional density of X given $Y = 2$ is

x	0	1	2
$f_{X\|Y=2}(x\|2)$	$\frac{4}{39}$	$\frac{5}{39}$	$\frac{30}{39}$

(b) $E(X) = (0)(0.2) + (1)(0.32) + (2)(0.48) = 1.28$,
$E(X^2) = (0)^2(0.2) + (1)^2(0.32) + (2)^2(0.48) = 2.24$, and
$Var(X) = 2.24 - 1.28^2 = 0.6016$.

(c) $E(X|Y = 2) = (1)\frac{5}{39} + (2)\frac{30}{39} = \frac{65}{39}$ and $E(X^2|Y = 2) = (1)^2\frac{5}{39} + (2)^2\frac{30}{39} = \frac{125}{39}$. So, $Var(X) = \frac{125}{39} - \left(\frac{65}{39}\right)^2 = \frac{650}{1521} = \frac{50}{117}$.

4.101 The profit is $8X + 3Y - 10$ for each trip. So, we need to calculate the average of this quantity. The marginal densities of X and Y are, respectively,

x	0	1	2
$g(x)$	0.34	0.32	0.34

y	0	1	2	3	4	5
$h(y)$	0.05	0.18	0.15	0.27	0.19	0.16

So, $E(8X+3Y-10) = (8)[(1)(0.32)+(2)(0.34)]+(3)[(1)(0.18)+(2)(0.15)+(3)(0.27)+(4)(0.19) + (5)(0.16)] - 10 = \$6.55$.

4.102 Using the approximation formula, $Var(Y) \approx \sum_{i=1}^{k} \left[\frac{\partial h(x_1,x_2,\ldots,x_k)}{\partial x_i}\right]^2\Big|_{x_i=\mu_i,\ 1\le i\le k} \sigma_i^2$, we have

$$Var(\hat{Y}) \approx \sum_{i=0}^{2} \left(\frac{\partial e^{b_0+b_1k_1+b_2k_2}}{\partial b_i}\right)^2\Bigg|_{b_i=\beta_i,\ 0\le i\le 2} \sigma_{b_i}^2 = e^{2(\beta_0+k_1\beta_1+k_2\beta_2)}(\sigma_0^2 + k_1^2\sigma_1^2 + k_2^2\sigma_2^2).$$

4.103 (a) $E(Y) = 10\int_0^1 y(1-y)^9\, dy = -\, y(1-y)^{10}|_0^1 + \int_0^1 (1-y)^{10}\, dy = \frac{1}{11}$.

(b) $E(1-Y) = 1 - E(Y) = \frac{10}{11}$.

(c) $Var(Z) = Var(1-Y) = Var(Y) = E(Y^2) - [E(Y)]^2 = \frac{10}{11^2\times 12} = 0.006887$.

Chapter 5

Some Discrete Probability Distributions

5.1 This is a uniform distribution: $f(x) = \frac{1}{10}$, for $x = 1, 2, \ldots, 10$.
Therefore $P(X < 4) = \sum_{x=1}^{3} f(x) = \frac{3}{10}$.

5.2 Binomial distribution with $n = 12$ and $p = 0.5$. Hence
$P(X = 3) = P(X \leq 3) - P(X \leq 2) = 0.0730 - 0.0193 = 0.0537$.

5.3 $\mu = \sum_{x=1}^{10} \frac{x}{10} = 5.5$, and $\sigma^2 = \sum_{x=1}^{10} \frac{(x-5.5)^2}{10} = 8.25$.

5.4 For $n = 5$ and $p = 3/4$, we have

(a) $P(X = 2) = \binom{5}{2}(3/4)^2(1/4)^3 = 0.0879$,

(b) $P(X \leq 3) = \sum_{x=0}^{3} b(x; 5, 3/4) = 1 - P(X = 4) - P(X = 5)$
$= 1 - \binom{5}{4}(3/4)^4(1/4)^1 - \binom{5}{5}(3/4)^5(1/4)^0 = 0.3672$.

5.5 We are considering a $b(x; 20, 0.3)$.

(a) $P(X \geq 10) = 1 - P(X \leq 9) = 1 - 0.9520 = 0.0480$.

(b) $P(X \leq 4) = 0.2375$.

(c) $P(X = 5) = 0.1789$. This probability is not very small so this is not a rare event. Therefore, $P = 0.30$ is reasonable.

5.6 For $n = 6$ and $p = 1/2$.

(a) $P(2 \leq X \leq 5) = P(X \leq 5) - P(X \leq 1) = 0.9844 - 0.1094$.

(b) $P(X < 3) = P(X \leq 2) = 0.3438$.

5.7 $p = 0.7$.

(a) For $n = 10$, $P(X < 5) = P(X \leq 4) = 0.0474$.

(b) For $n = 20$, $P(X < 10) = P(X \leq 9) = 0.0171$.

5.8 For $n = 8$ and $p = 0.6$, we have

(a) $P(X = 3) = b(3; 8, 0.6) = P(X \leq 3) - P(X \leq 2) = 0.1737 - 0.0498 = 0.1239$.

(b) $P(X \geq 5) = 1 - P(X \leq 4) = 1 - 0.4059 = 0.5941$.

5.9 For $n = 15$ and $p = 0.25$, we have

(a) $P(3 \leq X \leq 6) = P(X \leq 6) - P(X \leq 2) = 0.9434 - 0.2361 = 0.7073$.

(b) $P(X < 4) = P(X \leq 3) = 0.4613$.

(c) $P(X > 5) = 1 - P(X \leq 5) = 1 - 0.8516 = 0.1484$.

5.10 From Table A.1 with $n = 12$ and $p = 0.7$, we have

(a) $P(7 \leq X \leq 9) = P(X \leq 9) - P(X \leq 6) = 0.7472 - 0.1178 = 0.6294$.

(b) $P(X \leq 5) = 0.0386$.

(c) $P(X \geq 8) = 1 - P(X \leq 7) = 1 - 0.2763 = 0.7237$.

5.11 From Table A.1 with $n = 7$ and $p = 0.9$, we have
$P(X = 5) = P(X \leq 5) - P(X \leq 4) = 0.1497 - 0.0257 = 0.1240$.

5.12 From Table A.1 with $n = 9$ and $p = 0.25$, we have $P(X < 4) = 0.8343$.

5.13 From Table A.1 with $n = 5$ and $p = 0.7$, we have
$P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0.1631 = 0.8369$.

5.14 (a) $n = 4$, $P(X = 4) = 1 - 0.3439 = 0.6561$.

(b) Assuming the series went to the seventh game, the probability that the Bulls won 3 of the first 6 games and then the seventh game is given by

$$\left[\binom{6}{3}(0.9)^3(0.1)^3\right](0.9) = 0.0131.$$

(c) The probability that the Bulls win is always 0.9.

5.15 $p = 0.4$ and $n = 5$.

(a) $P(X = 0) = 0.0778$.

(b) $P(X < 2) = P(X \leq 1) = 0.3370$.

(c) $P(X > 3) = 1 - P(X \leq 3) = 1 - 0.9130 = 0.0870$.

5.16 Probability of 2 or more of 4 engines operating when $p = 0.6$ is

$$P(X \geq 2) = 1 - P(X \leq 1) = 0.8208,$$

and the probability of 1 or more of 2 engines operating when $p = 0.6$ is

$$P(X \geq 1) = 1 - P(X = 0) = 0.8400.$$

The 2-engine plane has a slightly higher probability for a successful flight when $p = 0.6$.

5.17 Since $\mu = np = (5)(0.7) = 3.5$ and $\sigma^2 = npq = (5)(0.7)(0.3) = 1.05$ with $\sigma = 1.025$. Then $\mu \pm 2\sigma = 3.5 \pm (2)(1.025) = 3.5 \pm 2.050$ or from 1.45 to 5.55. Therefore, at least 3/4 of the time when 5 people are selected at random, anywhere from 2 to 5 are of the opinion that tranquilizers do not cure but only cover up the real problem.

5.18 (a) $\mu = np = (15)(0.25) = 3.75$.

(b) With $k = 2$ and $\sigma = \sqrt{npq} = \sqrt{(15)(0.25)(0.75)} = 1.677$, $\mu \pm 2\sigma = 3.75 \pm 3.354$ or from 0.396 to 7.104.

5.19 Let X_1 = number of times encountered green light with $P(\text{Green}) = 0.35$, X_2 = number of times encountered yellow light with $P(\text{Yellow}) = 0.05$, and X_3 = number of times encountered red light with $P(\text{Red}) = 0.60$. Then

$$f(x_1, x_2, x_3) = \binom{n}{x_1, x_2, x_3}(0.35)^{x_1}(0.05)^{x_2}(0.60)^{x_3}.$$

5.20 (a) $\binom{10}{2,5,3}(0.225)^2(0.544)^5(0.231)^3 = 0.0749$.

(b) $\binom{10}{10}(0.544)^{10}(0.456)^0 = 0.0023$.

(c) $\binom{10}{0}(0.225)^0(0.775)^{10} = 0.0782$.

5.21 Using the multinomial distribution with required probability is

$$\binom{7}{0, 0, 1, 4, 2}(0.02)(0.82)^4(0.1)^2 = 0.0095.$$

5.22 Using the multinomial distribution, we have $\binom{8}{5,2,1}(1/2)^5(1/4)^2(1/4) = 21/256$.

5.23 Using the multinomial distribution, we have

$$\binom{9}{3, 3, 1, 2}(0.4)^3(0.2)^3(0.3)(0.2)^2 = 0.0077.$$

5.24 $p = 0.40$ and $n = 6$, so $P(X = 4) = P(X \leq 4) - P(X \leq 3) = 0.9590 - 0.8208 = 0.1382$.

5.25 $n = 20$ and the probability of a defective is $p = 0.10$. So, $P(X \leq 3) = 0.8670$.

5.26 $n = 8$ and $p = 0.60$;

(a) $P(X = 6) = \binom{8}{6}(0.6)^6(0.4)^2 = 0.2090$.

(b) $P(X = 6) = P(X \le 6) - P(X \le 5) = 0.8936 - 0.6846 = 0.2090$.

5.27 $n = 20$ and $p = 0.90$;

(a) $P(X = 18) = P(X \le 18) - P(X \le 17) = 0.6083 - 0.3231 = 0.2852$.

(b) $P(X \ge 15) = 1 - P(X \le 14) = 1 - 0.0113 = 0.9887$.

(c) $P(X \le 18) = 0.6083$.

5.28 $n = 20$;

(a) $p = 0.20$, $P(X \ge x) \le 0.5$ and $P(X < x) > 0.5$ yields $x = 4$.

(b) $p = 0.80$, $P(Y \ge y) \ge 0.8$ and $P(Y < y) < 0.2$ yields $y = 14$.

5.29 Using the hypergeometric distribution, we get

(a) $\frac{\binom{12}{2}\binom{40}{5}}{\binom{52}{7}} = 0.3246$.

(b) $1 - \frac{\binom{48}{7}}{\binom{52}{7}} = 0.4496$.

5.30 $P(X \ge 1) = 1 - P(X = 0) = 1 - h(0; 15, 3, 6) = 1 - \frac{\binom{6}{0}\binom{9}{3}}{\binom{15}{3}} = \frac{53}{65}$.

5.31 Using the hypergeometric distribution, we get $h(2; 9, 6, 4) = \frac{\binom{4}{2}\binom{5}{4}}{\binom{9}{6}} = \frac{5}{14}$.

5.32 (a) Probability that all 4 fire $= h(4; 10, 4, 7) = \frac{1}{6}$.

(b) Probability that at most 2 will not fire $= \sum_{x=0}^{2} h(x; 10, 4, 3) = \frac{29}{30}$.

5.33 $h(x; 6, 3, 4) = \frac{\binom{4}{x}\binom{2}{3-x}}{\binom{6}{3}}$, for $x = 1, 2, 3$.
$P(2 \le X \le 3) = h(2; 6, 3, 4) + h(3; 6, 3, 4) = \frac{4}{5}$.

5.34 $h(2; 9, 5, 4) = \frac{\binom{4}{2}\binom{5}{3}}{\binom{9}{5}} = \frac{10}{21}$.

5.35 $P(X \le 2) = \sum_{x=0}^{2} h(x; 50, 5, 10) = 0.9517$.

5.36 (a) $P(X = 0) = h(0; 25, 3, 3) = \frac{77}{115}$.

(b) $P(X = 1) = h(1; 25, 3, 1) = \frac{3}{25}$.

5.37 (a) $P(X = 0) = b(0; 3, 3/25) = 0.6815$.

(b) $P(1 \le X \le 3) = \sum_{x=1}^{3} b(x; 3, 1/25) = 0.1153.$

5.38 Since $\mu = (4)(3/10) = 1.2$ and $\sigma^2 = (4)(3/10)(7/10)(6/9) = 504/900$ with $\sigma = 0.7483$, at least 3/4 of the time the number of defectives will fall in the interval

$$\mu \pm 2\sigma = 1.2 \pm (2)(0.7483), \text{ or from } -0.297 \text{ to } 2.697,$$

and at least 8/9 of the time the number of defectives will fall in the interval

$$\mu \pm 3\sigma = 1.2 \pm (3)(0.7483) \text{ or from } -1.045 \text{ to } 3.445.$$

5.39 Since $\mu = (13)(13/52) = 3.25$ and $\sigma^2 = (13)(1/4)(3/4)(39/51) = 1.864$ with $\sigma = 1.365$, at least 75% of the time the number of hearts lay between

$$\mu \pm 2\sigma = 3.25 \pm (2)(1.365) \text{ or from } 0.52 \text{ to } 5.98.$$

5.40 The binomial approximation of the hypergeometric with $p = 1 - 4000/10000 = 0.6$ gives a probability of $\sum_{x=0}^{7} b(x; 15, 0.6) = 0.2131$.

5.41 Using the binomial approximation of the hypergeometric with $p = 0.5$, the probability is $1 - \sum_{x=0}^{2} b(x; 10, 0.5) = 0.9453$.

5.42 Using the binomial approximation of the hypergeometric distribution with $p = 30/150 = 0.2$, the probability is $1 - \sum_{x=0}^{2} b(x; 10, 0.2) = 0.3222$.

5.43 Using the binomial approximation of the hypergeometric distribution with 0.7, the probability is $1 - \sum_{x=10}^{13} b(x; 18, 0.7) = 0.6077$.

5.44 Using the extension of the hypergeometric distribution the probability is

$$\frac{\binom{13}{5}\binom{13}{2}\binom{13}{3}\binom{13}{3}}{\binom{52}{13}} = 0.0129.$$

5.45 (a) The extension of the hypergeometric distribution gives a probability

$$\frac{\binom{2}{1}\binom{3}{1}\binom{5}{1}\binom{2}{1}}{\binom{12}{4}} = \frac{4}{33}.$$

(b) Using the extension of the hypergeometric distribution, we have

$$\frac{\binom{2}{1}\binom{3}{1}\binom{2}{2}}{\binom{12}{4}} + \frac{\binom{2}{2}\binom{3}{1}\binom{2}{1}}{\binom{12}{4}} + \frac{\binom{2}{1}\binom{3}{2}\binom{2}{1}}{\binom{12}{4}} = \frac{8}{165}.$$

5.46 Using the extension of the hypergeometric distribution the probability is

$$\frac{\binom{2}{2}\binom{4}{1}\binom{3}{2}}{\binom{9}{5}} + \frac{\binom{2}{2}\binom{4}{2}\binom{3}{1}}{\binom{9}{5}} + \frac{\binom{2}{2}\binom{4}{3}\binom{3}{0}}{\binom{9}{5}} = \frac{17}{63}.$$

5.47 $h(5; 25, 15, 10) = \frac{\binom{10}{5}\binom{15}{10}}{\binom{25}{15}} = 0.2315.$

5.48 (a) $\frac{\binom{2}{1}\binom{13}{4}}{\binom{15}{5}} = 0.4762.$

(b) $\frac{\binom{2}{2}\binom{13}{3}}{\binom{15}{5}} = 0.0952.$

5.49 (a) $\frac{\binom{3}{0}\binom{17}{5}}{\binom{20}{5}} = 0.3991.$

(b) $\frac{\binom{3}{2}\binom{17}{3}}{\binom{20}{5}} = 0.1316.$

5.50 $N = 10000$, $n = 30$ and $k = 300$. Using binomial approximation to the hypergeometric distribution with $p = 300/10000 = 0.03$, the probability of $\{X \geq 1\}$ can be determined by

$$1 - b(0; 30, 0.03) = 1 - (0.97)^{30} = 0.599.$$

5.51 Using the negative binomial distribution, the required probability is

$$b^*(10; 5, 0.3) = \binom{9}{4}(0.3)^5(0.7)^5 = 0.0515.$$

5.52 From the negative binomial distribution, we obtain

$$b^*(8; 2, 1/6) = \binom{7}{1}(1/6)^2(5/6)^6 = 0.0651.$$

5.53 (a) $P(X > 5) = \sum_{x=6}^{\infty} p(x; 5) = 1 - \sum_{x=0}^{5} p(x; 5) = 0.3840.$

(b) $P(X = 0) = p(0; 5) = 0.0067.$

5.54 (a) Using the negative binomial distribution, we get

$$b^*(7; 3, 1/2) = \binom{6}{2}(1/2)^7 = 0.1172.$$

(b) From the geometric distribution, we have $g(4; 1/2) = (1/2)(1/2)^3 = 1/16.$

5.55 The probability that all coins turn up the same is 1/4. Using the geometric distribution with $p = 3/4$ and $q = 1/4$, we have

$$P(X < 4) = \sum_{x=1}^{3} g(x; 3/4) = \sum_{x=1}^{3} (3/4)(1/4)^{x-1} = \frac{63}{64}.$$

5.56 (a) Using the geometric distribution, we have $g(5; 2/3) = (2/3)(1/3)^4 = 2/243$.

(b) Using the negative binomial distribution, we have

$$b^*(5; 3, 2/3) = \binom{4}{2}(2/3)^3(1/3)^2 = \frac{16}{81}.$$

5.57 Using the geometric distribution

(a) $P(X = 3) = g(3; 0.7) = (0.7)(0.3)^2 = 0.0630$.

(b) $P(X < 4) = \sum_{x=1}^{3} g(x; 0.7) = \sum_{x=1}^{3} (0.7)(0.3)^{x-1} = 0.9730$.

5.58 (a) Using the Poisson distribution with $x = 5$ and $\mu = 3$, we find from Table A.2 that

$$p(5; 3) = \sum_{x=0}^{5} p(x; 3) - \sum_{x=0}^{4} p(x; 3) = 0.1008.$$

(b) $P(X < 3) = P(X \le 2) = 0.4232$.

(c) $P(X \ge 2) = 1 - P(X \le 1) = 0.8009$.

5.59 (a) $P(X \ge 4) = 1 - P(X \le 3) = 0.1429$.

(b) $P(X = 0) = p(0; 2) = 0.1353$.

5.60 (a) $P(X < 4) = P(X \le 3) = 0.1512$.

(b) $P(6 \le X \le 8) = P(X \le 8) - P(X \le 5) = 0.4015$.

5.61 (a) Using the negative binomial distribution, we obtain $b^*(6; 4, 0.8) = \binom{5}{3}(0.8)^4(0.2)^2 = 0.1638$.

(b) From the geometric distribution, we have $g(3; 0.8) = (0.8)(0.2)^2 = 0.032$.

5.62 (a) Using the Poisson distribution with $\mu = 12$, we find from Table A.2 that $P(X < 7) = P(X \le 6) = 0.0458$.

(b) Using the binomial distribution with $p = 0.0458$, we get

$$b(2; 3, 0.0458) = \binom{3}{2}(0.0458)^2(0.9542) = 0.0060.$$

5.63 (a) Using the Poisson distribution with $\mu = 5$, we find

$$P(X > 5) = 1 - P(X \leq 5) = 1 - 0.6160 = 0.3840.$$

(b) Using the binomial distribution with $p = 0.3840$, we get

$$b(3; 4, 0.384) = \binom{4}{3}(0.3840)^3(0.6160) = 0.1395.$$

(c) Using the geometric distribution with $p = 0.3840$, we have

$$g(5; 0.384) = (0.394)(0.616)^4 = 0.0553.$$

5.64 $\mu = np = (2000)(0.002) = 4$, so $P(X < 5) = P(X \leq 4) \approx \sum\limits_{x=0}^{4} p(x; 4) = 0.6288.$

5.65 $\mu = np = (10000)(0.001) = 10$, so

$$P(6 \leq X \leq 8) = P(X \leq 8) - P(X \leq 5) \approx \sum_{x=0}^{8} p(x; 10) - \sum_{x=0}^{5} p(x; 10) = 0.2657.$$

5.66 (a) $\mu = np = (1875)(0.004) = 7.5$, so $P(X < 5) = P(X \leq 4) \approx 0.1321.$

(b) $P(8 \leq X \leq 10) = P(X \leq 10) - P(X \leq 7) \approx 0.8622 - 0.5246 = 0.3376.$

5.67 (a) $\mu = (2000)(0.002) = 4$ and $\sigma^2 = 4$.

(b) For $k = 2$, we have $\mu \pm 2\sigma = 4 \pm 4$ or from 0 to 8.

5.68 (a) $\mu = (10000)(0.001) = 10$ and $\sigma^2 = 10$.

(b) For $k = 3$, we have $\mu \pm 3\sigma = 10 \pm 3\sqrt{10}$ or from 0.51 to 19.49.

5.69 (a) $P(X \leq 3|\lambda t = 5) = 0.2650.$

(b) $P(X > 1|\lambda t = 5) = 1 - 0.0404 = 0.9596.$

5.70 (a) $P(X = 4|\lambda t = 6) = 0.2851 - 0.1512 = 0.1339.$

(b) $P(X \geq 4|\lambda t = 6) = 1 - 0.1512 = 0.8488.$

(c) $P(X \geq 75|\lambda t = 72) = 1 - \sum\limits_{x=0}^{74} p(x; 74) = 0.3773.$

5.71 (a) $P(X > 10|\lambda t = 14) = 1 - 0.1757 = 0.8243.$

(b) $\lambda t = 14$.

5.72 $\mu = np = (1875)(0.004) = 7.5.$

5.73 $\mu = (4000)(0.001) = 4.$

5.74 $\mu = 1$ and $\sigma^2 = 0.99$.

5.75 $\mu = \lambda t = (1.5)(5) = 7.5$ and $P(X = 0|\lambda t = 7.5) = e^{-7.5} = 5.53 \times 10^{-4}$.

5.76 (a) $P(X \leq 1|\lambda t = 2) = 0.4060$.

(b) $\mu = \lambda t = (2)(5) = 10$ and $P(X \leq 4|\lambda t = 10) = 0.0293$.

5.77 (a) $P(X > 10|\lambda t = 5) = 1 - P(X \leq 10|\lambda t = 5) = 1 - 0.9863 = 0.0137$.

(b) $\mu = \lambda t = (5)(3) = 15$, so $P(X > 20|\lambda t = 15) = 1 - P(X \leq 20|\lambda = 15) = 1 - 0.9170 = 0.0830$.

5.78 $p = 0.03$ with a $g(x; 0.03)$. So, $P(X = 16) = (0.03)(1 - 0.03)^{15} = 0.0190$ and $\mu = \frac{1}{0.03} - 1 = 32.33$.

5.79 So, Let Y = number of shifts until it fails. Then Y follows a geometric distribution with $p = 0.10$. So,

$$\begin{aligned} P(Y \leq 6) &= g(1; 0.1) + g(2; 0.1) + \cdots + g(6; 0.1) \\ &= (0.1)[1 + (0.9) + (0.9)^2 + \cdots + (0.9)^5] = 0.4686. \end{aligned}$$

5.80 (a) The number of people interviewed before the first refusal follows a geometric distribution with $p = 0.2$. So.

$$P(X \geq 51) = \sum_{x=51}^{\infty} (0.2)(1 - 0.2)^x = (0.2)\frac{(1 - 0.2)^{50}}{1 - (1 - 0.2)} = 0.00001,$$

which is a very rare event.

(b) $\mu = \frac{1}{0.2} - 1 = 4$.

5.81 $n = 15$ and $p = 0.05$.

(a) $P(X \geq 2) = 1 - P(X \leq 1) = 1 - \sum_{x=0}^{1} \binom{15}{x}(0.05)^x(1 - 0.05)^{15-x} = 1 - 0.8290 = 0.1710$.

(b) $p = 0.07$. So, $P(X \leq 1) = \sum_{x=0}^{1} \binom{15}{x}(0.07)^x(1 - 0.07)^{15-x} = 1 - 0.7168 = 0.2832$.

5.82 $n = 100$ and $p = 0.01$.

(a) $P(X > 3) = 1 - P(X \leq 3) = 1 - \sum_{x=0}^{3} \binom{100}{x}(0.01)^x(1 - 0.01)^{100-x} = 1 - 0.9816 = 0.0184$.

(b) For $p = 0.05$, $P(X \leq 3) = \sum_{x=0}^{3} \binom{100}{x}(0.05)^x(1 - 0.05)^{100-x} = 0.2578$.

5.83 Using the extension of the hypergeometric distribution, the probability is

$$\frac{\binom{5}{2}\binom{7}{3}\binom{4}{1}\binom{3}{1}\binom{4}{2}}{\binom{5}{2}} = 0.0308.$$

5.84 $\lambda = 2.7$ call/min.

(a) $P(X \le 4) = \sum_{x=0}^{4} \frac{e^{-2.7}(2.7)^x}{x!} = 0.8629.$

(b) $P(X \le 1) = \sum_{x=0}^{1} \frac{e^{-2.7}(2.7)^x}{x!} = 0.2487.$

(c) $\lambda t = 13.5$. So,
$P(X > 10) = 1 - P(X \le 10) = 1 - \sum_{x=0}^{10} \frac{e^{-13.5}(13.5)^x}{x!} = 1 - 0.2971 = 0.7129.$

5.85 $n = 15$ and $p = 0.05$.

(a) $P(X = 5) = \binom{15}{5}(0.05)^5(1-0.05)^{10} = 0.000562.$

(b) I would not believe the claim of 5% defective.

5.86 $\lambda = 0.2$, so $\lambda t = (0.2)(5) = 1$.

(a) $P(X \le 1) = \sum_{x=0}^{1} \frac{e^{-1}(1)^x}{x!} = 2e^{-1} = 0.7358$. Hence, $P(X > 1) = 1 - 0.7358 = 0.2642$.

(b) $\lambda = 0.25$, so $\lambda t = 1.25$. $P(X|le1) = \sum_{x=0}^{1} \frac{e^{-1.25}(1.25)^x}{x!} = 0.6446.$

5.87 (a) $1 - P(X \le 1) = 1 - \sum_{x=0}^{1} \binom{100}{x}(0.01)^x(0.99)^{100-x} = 1 - 0.7358 = 0.2642.$

(b) $P(X \le 1) = \sum_{x=0}^{1} \binom{100}{x}(0.05)^x(0.95)^{100-x} = 0.0371.$

5.88 (a) 100 visits/60 minutes with $\lambda t = 5$ visits/3 minutes. $P(X = 0) = \frac{e^{-5}5^0}{0!} = 0.0067.$

(b) $P(X > 5) = 1 - \sum_{x=0}^{5} \frac{e^{5}5^x}{x!} = 1 - 0.6160 = 0.3840.$

5.89 (a) $P(X \ge 1) = 1 - P(X = 0) = 1 - \binom{4}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^4 = 1 - 0.4822 = 0.5177.$

(b) $P(X \ge 1) = 1 - P(X = 0) = 1 - \binom{24}{0}\left(\frac{1}{36}\right)^0\left(\frac{35}{36}\right)^{24} = 1 - 0.5086 = 0.4914.$

5.90 $n = 5$ and $p = 0.4$; $P(X \ge 3) = 1 - P(X \le 2) = 1 - 0.6826 = 0.3174.$

5.91 (a) $\mu = bp = (200)(0.03) = 6.$

(b) $\sigma^2 = npq = 5.82.$

(c) $P(X = 0) = \frac{e^6 (6)^0}{0!} = 0.0025$ (using the Poisson approximation).
$P(X = 0) = (0.97)^{200} = 0.0023$ (using the binomial distribution).

5.92 (a) $p^{10} q^0 = (0.99)^{10} = 0.9044$.

(b) $p^{10} q^{12-10} = (0.99)^{10} (0.01)^2 = (0.9044)(0.0001) = 0.00009$.

5.93 $n = 75$ with $p = 0.999$.

(a) $X =$ the number of trials, and $P(X = 75) = (0.999)^{75}(0.001)^0 = 0.9277$.

(b) $Y =$ the number of trials before the first failure (geometric distribution), and $P(Y = 20) = (0.001)(0.999)^{19} = 0.000981$.

(c) $1 - P(\text{no failures}) = 1 - (0.001)^0 (0.999)^{10} = 0.01$.

5.94 (a) $\binom{10}{1} pq^9 = (10)(0.25)(0.75)^9 = 0.1877$.

(b) Let X be the number of drills until the first success. X follows a geometric distribution with $p = 0.25$. So, the probability of having the first 10 drills being failure is $q^{10} = (0.75)^{10} = 0.056$. So, there is a small prospects for bankruptcy. Also, the probability that the first success appears in the 11th drill is $pq^{10} = 0.014$ which is even smaller.

5.95 It is a negative binomial distribution. $\binom{x-1}{k-1} p^k q^{x-k} = \binom{6-1}{2-1} (0.25)^2 (0.75)^4 = 0.0989$.

5.96 It is a negative binomial distribution. $\binom{x-1}{k-1} p^k q^{x-k} = \binom{4-1}{2-1} (0.5)^2 (0.5)^2 = 0.1875$.

5.97 $n = 1000$ and $p = 0.01$, with $\mu = (1000)(0.01) = 10$. $P(X < 7) = P(X \leq 6) = 0.1301$.

5.98 $n = 500$;

(a) If $p = 0.01$,

$$P(X \geq 15) = 1 - P(X \leq 14) = 1 - \sum_{x=0}^{14} \binom{500}{x} (0.01)^x (0.99)^{500-x} = 0.00021.$$

This is a very rare probability and thus the original claim that $p = 0.01$ is questionable.

(b) $P(X = 3) = \binom{500}{3} (0.01)^3 (0.99)^{497} = 0.1402$.

(c) For (a), if $p = 0.01$, $\mu = (500)(0.01) = 5$. So

$$P(X \geq 15) = 1 - P(X \leq 14) = 1 - 0.9998 = 0.0002.$$

For (b),

$$P(X = 3) = 0.2650 - 0.1247 = 0.1403.$$

5.99 $N = 50$ and $n = 10$.

(a) $k = 2$; $P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{2}{0}\binom{48}{10}}{\binom{50}{10}} = 1 - 0.6367 = 0.3633$.

(b) Even though the lot contains 2 defectives, the probability of reject the lot is not very high. Perhaps more items should be sampled.

(c) $\mu = (10)(2/50) = 0.4$.

5.100 Suppose n items need to be sampled. $P(X \geq 1) = 1 - \frac{\binom{2}{0}\binom{48}{n}}{\binom{50}{n}} = 1 - \frac{(50-n)(49-n)}{(50)(49)} \geq 0.9$. The solution is $n = 34$.

5.101 Define X = number of screens will detect. Then $X \sim b(x; 3, 0.8)$.

(a) $P(X = 0) = (1 - 0.8)^3 = 0.008$.

(b) $P(X = 1) = (3)(0.2)^2(0.8) = 0.096$.

(c) $P(X \geq 2) = P(X = 2) + P(X = 3) = (3)(0.8)^2(0.2) + (0.8)^3 = 0.896$.

5.102 (a) $P(X = 0) = (1 - 0.8)^n \leq 0.0001$ implies that $n \geq 6$.

(b) $(1 - p)^3 \leq 0.0001$ implies $p \geq 0.9536$.

5.103 $n = 10$ and $p = \frac{2}{50} = 0.04$.

$$P(X \geq 1) = 1 - P(X = 0) \approx 1 - \binom{10}{0}(0.04)^0(1 - 0.04)^{10} = 1 - 0.6648 = 0.3351.$$

The approximation is not that good due to $\frac{n}{N} = 0.2$ is too large.

5.104 (a) $P = \frac{\binom{2}{2}\binom{3}{0}}{\binom{5}{2}} = 0.1$.

(b) $P = \frac{\binom{2}{1}\binom{1}{1}}{\binom{5}{2}} = 0.2$.

5.105 $n = 200$ with $p = 0.00001$.

(a) $P(X \geq 5) = 1 - P(X \leq 4) = 1 - \sum_{x=0}^{4} \binom{200}{x}(0.00001)^x(1 - 0.00001)^{200-x} \approx 0$. This is a rare event. Therefore, the claim does not seem right.

(b) $\mu = np = (200)(0.00001) = 0.02$. Using Poisson approximation,

$$P(X \geq 5) = 1 - P(X \leq 4) \approx 1 - \sum_{x=0}^{4} e^{-0.02}\frac{(0.02)^x}{x!} = 0.$$

Chapter 6

Some Continuous Probability Distributions

6.1 (a) Area=0.9236.

(b) Area=$1 - 0.1867 = 0.8133$.

(c) Area=$0.2578 - 0.0154 = 0.2424$.

(d) Area=0.0823.

(e) Area=$1 - 0.9750 = 0.0250$.

(f) Area=$0.9591 - 0.3156 = 0.6435$.

6.2 (a) The area to the left of z is $1 - 0.3622 = 0.6378$ which is closer to the tabled value 0.6368 than to 0.6406. Therefore, we choose $z = 0.35$.

(b) From Table A.3, $z = -1.21$.

(c) The total area to the left of z is 0.5000+0.4838=0.9838. Therefore, from Table A.3, $z = 2.14$.

(d) The distribution contains an area of 0.025 to the left of $-z$ and therefore a total area of 0.025+0.95=0.975 to the left of z. From Table A.3, $z = 1.96$.

6.3 (a) From Table A.3, $k = -1.72$.

(b) Since $P(Z > k) = 0.2946$, then $P(Z < k) = 0.7054$/ From Table A.3, we find $k = 0.54$.

(c) The area to the left of $z = -0.93$ is found from Table A.3 to be 0.1762. Therefore, the total area to the left of k is 0.1762+0.7235=0.8997, and hence $k = 1.28$.

6.4 (a) $z = (17 - 30)/6 = -2.17$. Area=$1 - 0.0150 = 0.9850$.

(b) $z = (22 - 30)/6 = -1.33$. Area=0.0918.

(c) $z_1 = (32-3)/6 = 0.33$, $z_2 = (41-30)/6 = 1.83$. Area $= 0.9664-0.6293 = 0.3371$.

(d) $z = 0.84$. Therefore, $x = 30 + (6)(0.84) = 35.04$.

(e) $z_1 = -1.15$, $z_2 = 1.15$. Therefore, $x_1 = 30 + (6)(-1.15) = 23.1$ and $x_2 = 30 + (6)(1.15) = 36.9$.

6.5 (a) $z = (15 - 18)/2.5 = -1.2$; $P(X < 15) = P(Z < -1.2) = 0.1151$.

(b) $z = -0.76$, $k = (2.5)(-0.76) + 18 = 16.1$.

(c) $z = 0.91$, $k = (2.5)(0.91) + 18 = 20.275$.

(d) $z_1 = (17 - 18)/2.5 = -0.4$, $z_2 = (21 - 18)/2.5 = 1.2$;
$P(17 < X < 21) = P(-0.4 < Z < 1.2) = 0.8849 - 0.3446 = 0.5403$.

6.6 $z_1 = [(\mu - 3\sigma) - \mu]/\sigma = -3$, $z_2 = [(\mu + 3\sigma) - \mu]/\sigma = 3$;
$P(\mu - 3\sigma < Z < \mu + 3\sigma) = P(-3 < Z < 3) = 0.9987 - 0.0013 = 0.9974$.

6.7 (a) $z = (32 - 40)/6.3 = -1.27$; $P(X > 32) = P(Z > -1.27) = 1 - 0.1020 = 0.8980$.

(b) $z = (28 - 40)/6.3 = -1.90$, $P(X < 28) = P(Z < -1.90) = 0.0287$.

(c) $z_1 = (37 - 40)/6.3 = -0.48$, $z_2 = (49 - 40)/6.3 = 1.43$;
So, $P(37 < X < 49) = P(-0.48 < Z < 1.43) = 0.9236 - 0.3156 = 0.6080$.

6.8 (a) $z = (31.7 - 30)/2 = 0.85$; $P(X > 31.7) = P(Z > 0.85) = 0.1977$.
Therefore, 19.77% of the loaves are longer than 31.7 centimeters.

(b) $z_1 = (29.3 - 30)/2 = -0.35$, $z_2 = (33.5 - 30)/2 = 1.75$;
$P(29.3 < X < 33.5) = P(-0.35 < Z < 1.75) = 0.9599 - 0.3632 = 0.5967$.
Therefore, 59.67% of the loaves are between 29.3 and 33.5 centimeters in length.

(c) $z = (25.5 - 30)/2 = -2.25$; $P(X < 25.5) = P(Z < -2.25) = 0.0122$.
Therefore, 1.22% of the loaves are shorter than 25.5 centimeters in length.

6.9 (a) $z = (224 - 200)/15 = 1.6$. Fraction of the cups containing more than 224 millimeters is $P(Z > 1.6) = 0.0548$.

(b) $z_1 = (191 - 200)/15 = -0.6$, $Z_2 = (209 - 200)/15 = 0.6$;
$P(191 < X < 209) = P(-0.6 < Z < 0.6) = 0.7257 - 0.2743 = 0.4514$.

(c) $z = (230 - 200)/15 = 2.0$; $P(X > 230) = P(Z > 2.0) = 0.0228$. Therefore, $(1000)(0.0228) = 22.8$ or approximately 23 cups will overflow.

(d) $z = -0.67$, $x = (15)(-0.67) + 200 = 189.95$ millimeters.

6.10 (a) $z = (10.075 - 10.000)/0.03 = 2.5$; $P(X > 10.075) = P(Z > 2.5) = 0.0062$.
Therefore, 0.62% of the rings have inside diameters exceeding 10.075 cm.

(b) $z_1 = (9.97 - 10)/0.03 = -1.0$, $z_2 = (10.03 - 10)/0.03 = 1.0$;
$P(9.97 < X < 10.03) = P(-1.0 < Z < 1.0) = 0.8413 - 0.1587 = 0.6826$.

(c) $z = -1.04$, $x = 10 + (0.03)(-1.04) = 9.969$ cm.

6.11 (a) $z = (30 - 24)/3.8 = 1.58$; $P(X > 30) = P(Z > 1.58) = 0.0571$.

(b) $z = (15 - 24)/3.8 = -2.37$; $P(X > 15) = P(Z > -2.37) = 0.9911$. He is late 99.11% of the time.

(c) $z = (25 - 24)/3.8 = 0.26$; $P(X > 25) = P(Z > 0.26) = 0.3974$.

(d) $z = 1.04$, $x = (3.8)(1.04) + 24 = 27.952$ minutes.

(e) Using the binomial distribution with $p = 0.0571$, we get
$b(2; 3, 0.0571) = \binom{3}{2}(0.0571)^2(0.9429) = 0.0092$.

6.12 $\mu = 99.61$ and $\sigma = 0.08$.

(a) $P(99.5 < X < 99.7) = P(-1.375 < Z < 1.125) = 0.8697 - 0.08455 = 0.7852$.

(b) $P(Z > 1.645) = 0.05$; $x = (1.645)(0.08) + 99.61 = 99.74$.

6.13 $z = -1.88$, $x = (2)(-1.88) + 10 = 6.24$ years.

6.14 (a) $z = (159.75 - 174.5)/6.9 = -2.14$; $P(X < 159.75) = P(Z < -2.14) = 0.0162$. Therefore, $(1000)(0.0162) = 16$ students.

(b) $z_1 = (171.25 - 174.5)/6.9 = -0.47$, $z_2 = (182.25 - 174.5)/6.9 = 1.12$.
$P(171.25 < X < 182.25) = P(-0.47 < Z < 1.12) = 0.8686 - 0.3192 = 0.5494$.
Therefore, $(1000)(0.5494) = 549$ students.

(c) $z_1 = (174.75 - 174.5)/6.9 = 0.04$, $z_2 = (175.25 - 174.5)/6.9 = 0.11$.
$P(174.75 < X < 175.25) = P(0.04 < Z < 0.11) = 0.5438 - 0.5160 = 0.0278$.
Therefore, $(1000)(0.0278)$=28 students.

(d) $z = (187.75 - 174.5)/6.9 = 1.92$; $P(X > 187.75) = P(Z > 1.92) = 0.0274$.
Therefore, $(1000)(0.0274) = 27$ students.

6.15 $\mu = \$15.90$ and $\sigma = \$1.50$.

(a) 51%, since $P(13.75 < X < 16.22) = P\left(\frac{13.745-15.9}{1.5} < Z < \frac{16.225-15.9}{1.5}\right)$
$= P(-1.437 < Z < 0.217) = 0.5871 - 0.0749 = 0.5122$.

(b) \$18.36, since $P(Z > 1.645) = 0.05$; $x = (1.645)(1.50) + 15.90 + 0.005 = 18.37$.

6.16 (a) $z = (9.55 - 8)/0.9 = 1.72$. Fraction of poodles weighing over 9.5 kilograms $= P(X > 9.55) = P(Z > 1.72) = 0.0427$.

(b) $z = (8.65 - 8)/0.9 = 0.72$. Fraction of poodles weighing at most 8.6 kilograms $= P(X < 8.65) = P(Z < 0.72) = 0.7642$.

(c) $z_1 = (7.25 - 8)/0.9 = -0.83$ and $z_2 = (9.15 - 8)/0.9 = 1.28$.
Fraction of poodles weighing between 7.3 and 9.1 kilograms inclusive
$= P(7.25 < X < 9.15) = P(-0.83 < Z < 1.28) = 0.8997 - 0.2033 = 0.6964$.

6.17 (a) $z = (10,175 - 10,000)/100 = 1.75$. Proportion of components exceeding 10.150 kilograms in tensile strength$= P(X > 10,175) = P(Z > 1.75) = 0.0401$.

(b) $z_1 = (9,775 - 10,000)/100 = -2.25$ and $z_2 = (10,225 - 10,000)/100 = 2.25$. Proportion of components scrapped$= P(X < 9,775) + P(X > 10,225) = P(Z < -2.25) + P(Z > 2.25) = 2P(Z < -2.25) = 0.0244$.

6.18 (a) $x_1 = \mu + 1.3\sigma$ and $x_2 = \mu - 1.3\sigma$. Then $z_1 = 1.3$ and $z_2 = -1.3$. $P(X > \mu+1.3\sigma)+P(X < 1.3\sigma) = P(Z > 1.3)+P(Z < -1.3) = 2P(Z < -1.3) = 0.1936$. Therefore, 19.36%.

(b) $x_1 = \mu+0.52\sigma$ and $x_2 = \mu-0.52\sigma$. Then $z_1 = 0.52$ and $z_2 = -0.52$. $P(\mu-0.52\sigma < X < \mu + 0.52\sigma) = P(-0.52 < Z < 0.52) = 0.6985 - 0.3015 = 0.3970$. Therefore, 39.70%.

6.19 $z = (94.5 - 115)/12 = -1.71$; $P(X < 94.5) = P(Z < -1.71) = 0.0436$. Therefore, $(0.0436)(600) = 26$ students will be rejected.

6.20 $f(x) = \frac{1}{B-A}$ for $A \le x \le B$.

(a) $\mu = \int_A^B \frac{x}{B-A}\, dx = \frac{B^2-A^2}{2(B-A)} = \frac{A+B}{2}$.

(b) $E(X^2) = \int_A^B \frac{x^2}{B-A}\, dx = \frac{B^3-A^3}{3(B-A)}$.
So, $\sigma^2 = \frac{B^3-A^3}{3(B-A)} - \left(\frac{A+B}{2}\right)^2 = \frac{4(B^2+AB+A^2)-3(B^2+2AB+A^2)}{12} = \frac{B^2-2AB+A^2}{12} = \frac{(B-A)^2}{12}$.

6.21 $A = 7$ and $B = 10$.

(a) $P(X \le 8.8) = \frac{8.8-7}{3} = 0.60$.

(b) $P(7.4 < X < 9.5) = \frac{9.5-7.4}{3} = 0.70$.

(c) $P(X \ge 8.5) = \frac{10-8.5}{3} = 0.50$.

6.22 (a) $P(X > 7) = \frac{10-7}{10} = 0.3$.

(b) $P(2 < X < 7) = \frac{7-2}{10} = 0.5$.

6.23 (a) From Table A.1 with $n = 15$ and $p = 0.2$ we have
$P(1 \le X \le 4) = \sum_{x=0}^{4} b(x; 15, 0.2) - b(0; 15, 0.2) = 0.8358 - 0.0352 = 0.8006$.

(b) By the normal-curve approximation we first find
$\mu = np = 3$ and then $\sigma^2 = npq = (15)(0.2)(0.8) = 2.4$. Then $\sigma = 1.549$.
Now, $z_1 = (0.5 - 3)/1.549 = -1.61$ and $z_2 = (4.5 - 3)/1.549 = 0.97$.
Therefore, $P(1 \le X \le 4) = P(-1.61 \le Z \le 0.97) = 0.8340 - 0.0537 = 0.7803$.

6.24 $\mu = np = (400)(1/2) = 200$, $\sigma = \sqrt{npq} = \sqrt{(400)(1/2)(1/2)} = 10$.

(a) $z_1 = (184.5 - 200)/10 = -1.55$ and $z_2 = (210.5 - 200)/10 = 1.05$.
$P(184.5 < X < 210.5) = P(-1.55 < Z < 1.05) = 0.8531 - 0.0606 = 0.7925$.

(b) $z_1 = (204.5 - 200)/10 = 0.45$ and $z_2 = (205.5 - 200)/10 = 0.55$.
$P(204.5 < X < 205.5) = P(0.45 < Z < 0.55) = 0.7088 - 0.6736 = 0.0352$.

(c) $z_1 = (175.5 - 200)/10 = -2.45$ and $z_2 = (227.5 - 200)/10 = 2.75$.
$P(X < 175.5) + P(X > 227.5) = P(Z < -2.45) + P(Z > 2.75)$
$= P(Z < -2.45) + 1 - P(Z < 2.75) = 0.0071 + 1 - 0.9970 = 0.0101$.

6.25 $n = 100$.

(a) $p = 0.01$ with $\mu = (100)(0.01) = 1$ and $\sigma = \sqrt{(100)(0.01)(0.99)} = 0.995$.
So, $z = (0.5 - 1)/0.995 = -0.503$. $P(X \leq 0) \approx P(Z \leq -0.503) = 0.3085$.

(b) $p = 0.05$ with $\mu = (100)(0.05) = 5$ and $\sigma = \sqrt{(100)(0.05)(0.95)} = 2.1794$.
So, $z = (0.5 - 5)/2.1794 = -2.06$. $P(X \leq 0) \approx P(X \leq -2.06) = 0.0197$.

6.26 $\mu = np = (100)(0.1) = 10$ and $\sigma = \sqrt{(100)(0.1)(0.9)} = 3$.

(a) $z = (13.5 - 10)/3 = 1.17$; $P(X > 13.5) = P(Z > 1.17) = 0.1210$.

(b) $z = (7.5 - 10)/3 = -0.83$; $P(X < 7.5) = P(Z < -0.83) = 0.2033$.

6.27 $\mu = (100)(0.9) = 90$ and $\sigma = \sqrt{(100)(0.9)(0.1)} = 3$.

(a) $z_1 = (83.5 - 90)/3 = -2.17$ and $z_2 = (95.5 - 90)/3 = 1.83$.
$P(83.5 < X < 95.5) = P(-2.17 < Z < 1.83) = 0.9664 - 0.0150 = 0.9514$.

(b) $z = (85.5 - 90)/3 = -1.50$; $P(X < 85.5) = P(Z < -1.50) = 0.0668$.

6.28 $\mu = (80)(3/4) = 60$ and $\sigma = \sqrt{(80)(3/4)(1/4)} = 3.873$.

(a) $z = (49.5 - 60)/3.873 = -2.71$; $P(X > 49.5) = P(Z > -2.71) = 1 - 0.0034 = 0.9966$.

(b) $z = (56.5 - 60)/3.873 = -0.90$; $P(X < 56.5) = P(Z < -0.90) = 0.1841$.

6.29 $\mu = (1000)(0.2) = 200$ and $\sigma = \sqrt{(1000)(0.2)(0.8)} = 12.649$.

(a) $z_1 = (169.5 - 200)/12.649 = -2.41$ and $z_2 = (185.5 - 200)/12.649 = -1.15$.
$P(169.5 < X < 185.5) = P(-2.41 < Z < -1.15) = 0.1251 - 0.0080 = 0.1171$.

(b) $z_1 = (209.5 - 200)/12.649 = 0.75$ and $z_2 = (225.5 - 200)/12.649 = 2.02$.
$P(209.5 < X < 225.5) = P(0.75 < Z < 2.02) = 0.9783 - 0.7734 = 0.2049$.

6.30 (a) $\mu = (100)(0.8) = 80$ and $\sigma = \sqrt{(100)(0.8)(0.2)} = 4$ with $z = (74.5 - 80)/4 = -1.38$.
$P(\text{Claim is rejected when } p = 0.8) = P(Z < -1.38) = 0.0838$.

(b) $\mu = (100)(0.7) = 70$ and $\sigma = \sqrt{(100)(0.7)(0.3)} = 4.583$ with $z = (74.5 - 70)/4.583 = 0.98$.
$P(\text{Claim is accepted when } p = 0.7) = P(Z > 0.98) = 1 - 0.8365 = 0.1635$.

6.31 $\mu = (180)(1/6) = 30$ and $\sigma = \sqrt{(180)(1/6)(5/6)} = 5$ with $z = (35.5 - 30)/5 = 1.1$.
$P(X > 35.5) = P(Z > 1.1) = 1 - 0.8643 = 0.1357$.

6.32 $\mu = (200)(0.05) = 10$ and $\sigma = \sqrt{(200)(0.05)(0.95)} = 3.082$ with
$z = (9.5 - 10)/3.082 = -0.16$. $P(X < 10) = P(Z < -0.16) = 0.4364$.

6.33 $\mu = (400)(1/10) = 40$ and $\sigma = \sqrt{(400)(1/10)(9/10)} = 6$.

(a) $z = (31.5 - 40)/6 = -1.42$; $P(X < 31.5) = P(Z < -1.42) = 0.0778$.

(b) $z = (49.5 - 40)/6 = 1.58$; $P(X > 49.5) = P(Z > 1.58) = 1 - 0.9429 = 0.0571$.

(c) $z_1 = (34.5 - 40)/6 = -0.92$ and $z_2 = (46.5 - 40)/6 = 1.08$;
$P(34.5 < X < 46.5) = P(-0.92 < Z < 1.08) = 0.8599 - 0.1788 = 0.6811$.

6.34 $\mu = (180)(1/6) = 30$ and $\sigma = \sqrt{(180)(1/6)(5/6)} = 5$.

(a) $z = (24.5 - 30)/5 = -1.1$; $P(X > 24.5) = P(Z > -1.1) = 1 - 0.1357 = 0.8643$.

(b) $z_1 = (32.5 - 30)/5 = 0.5$ and $z_2 = (41.5 - 30)/5 = 2.3$.
$P(32.5 < X < 41.5) = P(0.5 < Z < 2.3) = 0.9893 - 0.6915 = 0.2978$.

(c) $z_1 = (29.5 - 30)/5 = -0.1$ and $z_2 = (30.5 - 30)/5 = 0.1$.
$P(29.5 < X < 30.5) = P(-0.1 < Z < 0.1) = 0.5398 - 0.4602 = 0.0796$.

6.35 (a) $p = 0.05$, $n = 100$ with $\mu = 5$ and $\sigma = \sqrt{(100)(0.05)(0.95)} = 2.1794$.
So, $z = (2.5 - 5)/2.1794 = -1.147$; $P(X \geq 2) \approx P(Z \geq -1.147) = 0.8749$.

(b) $z = (10.5 - 5)/2.1794 = 2.524$; $P(X \geq 10) \approx P(Z > 2.52) = 0.0059$.

6.36 $n = 200$; $X =$ The number of no shows with $p = 0.02$. $z = \frac{3-0.5-4}{\sqrt{(200)(0.02)(0.98)}} = -0.76$.
Therefore, $P(\text{airline overbooks the flight}) = 1 - P(X \geq 3) \approx 1 - P(Z > -0.76) = 0.2236$.

6.37 (a) $P(X \geq 230) = P\left(Z > \frac{230-170}{30}\right) = 0.0228$.

(b) Denote by Y the number of students whose serum cholesterol level exceed 230 among the 300. Then $Y \sim b(y; 300, 0.0228$ with $\mu = (300)(0.0228) = 6.84$ and $\sigma = \sqrt{(300)(0.0228)(1 - 0.0228)} = 2.5854$. So, $z = \frac{8-0.5-6.84}{2.5854} = 0.26$ and
$P(X \geq 8) \approx P(Z > 0.26) = 0.3974$.

6.38 (a) Denote by X the number of failures among the 20. $X \sim b(x; 20, 0.01)$ and $P(X > 1) = 1 - b(0; 20, 0.01) - b(1; 20, 0.01) = 1 - \binom{20}{0}(0.01)^0(0.99)^{20} - \binom{20}{1}(0.01)(0.99)^{19} = 0.01686$.

(b) $n = 500$ and $p = 0.01$ with $\mu = (500)(0.01) = 5$ and $\sigma = \sqrt{(500)(0.01)(0.99)} = 2.2249$. So, $P(\text{more than 8 failures}) \approx P(Z > (8.5 - 5)/2.2249) = P(Z > 1.57) = 1 - 0.9418 = 0.0582$.

6.39 $P(1.8 < X < 2.4) = \int_{1.8}^{2.4} xe^{-x}\, dx = [-xe^{-x} - e^{-x}]|_{1.8}^{2.4} = 2.8e^{-1.8} - 3.4e^{-2.4} = 0.1545$.

6.40 $P(X > 9) = \frac{1}{9}\int_9^\infty x^{-x/3}\, dx = \left[-\frac{x}{3}e^{-x/3} - e^{-x/3}\right]\Big|_9^\infty = 4e^{-3} = 0.1992$.

6.41 Setting $\alpha = 1/2$ in the gamma distribution and integrating, we have

$$\frac{1}{\sqrt{\beta}\Gamma(1/2)} \int_0^\infty x^{-1/2} e^{-x/\beta}\, dx = 1.$$

Substitute $x = y^2/2$, $dx = y\ dy$, to give

$$\Gamma(1/2) = \frac{\sqrt{2}}{\sqrt{\beta}} \int_0^\infty e^{-y^2/2\beta}\ dy = 2\sqrt{\pi}\left(\frac{1}{\sqrt{2\pi}\sqrt{\beta}} \int_0^\infty e^{-y^2/2\beta}\ dy\right) = \sqrt{\pi},$$

since the quantity in parentheses represents one-half of the area under the normal curve $n(y; 0, \sqrt{\beta})$.

6.42 (a) $P(X < 1) = 4\int_0^1 xe^{-2x}\ dx = [-2xe^{-2x} - e^{-2x}]|_0^1 = 1 - 3e^{-2} = 0.5940$.

(b) $P(X > 2) = 4\int_0^\infty xe^{-2x}\ dx = [-2xe^{-2x} - e^{-2x}]|_2^\infty = 5e^{-4} = 0.0916$.

6.43 (a) $\mu = \alpha\beta = (2)(3) = 6$ million liters; $\sigma^2 = \alpha\beta^2 = (2)(9) = 18$.

(b) Water consumption on any given day has a probability of at least 3/4 of falling in the interval $\mu \pm 2\sigma = 6 \pm 2\sqrt{18}$ or from -2.485 to 14.485. That is from 0 to 14.485 million liters.

6.44 (a) $\mu = \alpha\beta = 6$ and $\sigma^2 = \alpha\beta^2 = 12$. Substituting $\alpha = 6/\beta$ into the variance formula we find $6\beta = 12$ or $\beta = 2$ and then $\alpha = 3$.

(b) $P(X > 12) = \frac{1}{16}\int_{12}^\infty x^2 e^{-x/2}\ dx$. Integrating by parts twice gives

$$P(X > 12) = \frac{1}{16}\left[-2x^2e^{-x/2} - 8xe^{-x/2} - 16e^{-x/2}\right]\Big|_{12}^\infty = 25e^{-6} = 0.0620.$$

6.45 $P(X < 3) = \frac{1}{4}\int_0^3 e^{-x/4}\ dx = -e^{-x/4}\Big|_0^3 = 1 - e^{-3/4} = 0.5276$.
Let Y be the number of days a person is served in less than 3 minutes. Then
$P(Y \geq 4) = \sum_{x=4}^{6} b(y; 6, 1 - e^{-3/4}) = \binom{6}{4}(0.5276)^4(0.4724)^2 + \binom{6}{5}(0.5276)^5(0.4724)$
$+ \binom{6}{6}(0.5276)^6 = 0.3968$.

6.46 $P(X < 1) = \frac{1}{2}\int_0^1 e^{-x/2}\ dx = -e^{-x/2}\Big|_0^1 = 1 - e^{-1/2} = 0.3935$. Let Y be the number of switches that fail during the first year. Using the normal approximation we find $\mu = (100)(0.3935) = 39.35$, $\sigma = \sqrt{(100)(0.3935)(0.6065)} = 4.885$, and $z = (30.5 - 39.35)/4.885 = -1.81$. Therefore, $P(Y \leq 30) = P(Z < -1.81) = 0.0352$.

6.47 (a) $E(X) = \int_0^\infty x^2 e^{-x^2/2}\ dx = -xe^{-x^2/2}\Big|_0^\infty + \int_0^\infty e^{-x^2/2}\ dx$
$= 0 + \sqrt{2\pi} \cdot \frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-x^2/2}\ dx = \frac{\sqrt{2\pi}}{2} = \sqrt{\frac{\pi}{2}} = 1.2533$.

(b) $P(X > 2) = \int_2^\infty xe^{-x^2/2}\ dx = -e^{-x^2/2}\Big|_2^\infty = e^{-2} = 0.1353$.

6.48 $\mu = E(T) = \alpha\beta\int_0^\infty t^\beta e^{-\alpha t^\beta}\ dt$. Let $y = \alpha t^\beta$, then $dy = \alpha\beta t^{\beta-1}\ dt$ and $t = (y/\alpha)^{1/\beta}$. Then

$$\mu = \int_0^\infty (y/\alpha)^{1/\beta} e^{-y}\ dy = \alpha^{-1/\beta}\int_0^\infty y^{(1+1/\beta)-1}e^{-y}\ dy = \alpha^{-1/\beta}\Gamma(1 + 1/\beta).$$

$$E(T^2) = \alpha\beta \int_0^\infty t^{\beta+1} e^{-\alpha t^\beta}\ dt = \int_0^\infty (y/\alpha)^{2/\beta} e^{-y}\ dy = \alpha^{-2/\beta} \int_0^\infty y^{(1+2/\beta)-1} e^{-y}\ dy$$
$$= \alpha^{-2/\beta}\Gamma(1+2/\beta).$$

So, $\sigma^2 = E(T^2) - \mu^2 = \alpha^{-2/\beta}\{\Gamma(1+2/\beta) - [\Gamma(1+1/\beta)]^2\}$.

6.49 $R(t) = ce^{-\int 1/\sqrt{t}\ dt} = ce^{-2\sqrt{t}}$. However, $R(0) = 1$ and hence $c = 1$. Now

$$f(t) = Z(t)R(t) = e^{-2\sqrt{t}}/\sqrt{t}, \qquad t > 0,$$

and

$$P(T > 4) = \int_4^\infty e^{-2\sqrt{t}}/\sqrt{t}\ dt = -e^{-2\sqrt{t}}\Big|_4^\infty = e^{-4} = 0.0183.$$

6.50 $f(x) = 12x^2(1-x)$, $0 < x < 1$. Therefore,

$$P(X > 0.8) = 12 \int_{0.8}^1 x^2(1-x)\ dx = 0.1808.$$

6.51 $\alpha = 5$; $\beta = 10$;

(a) $\alpha\beta = 50$.

(b) $\sigma^2 = \alpha\beta^2 = 500$; so $\sigma = \sqrt{500} = 22.36$.

(c) $P(X > 30) = \frac{1}{\beta^\alpha\Gamma(\alpha)} \int_{30}^\infty x^{\alpha-1}e^{-x/\beta}\ dx$. Using the incomplete gamma with $y = x/\beta$, then

$$1 - P(X \le 30) = 1 - P(Y \le 3) = 1 - \int_0^3 \frac{y^4 e^{-y}}{\Gamma(5)}\ dy = 1 - 0.185 = 0.815.$$

6.52 $\alpha\beta = 10$; $\sigma = \sqrt{\alpha\beta^2}\sqrt{50} = 7.07$.

(a) Using integration by parts,

$$P(X \le 50) = \frac{1}{\beta^\alpha\Gamma(\alpha)} \int_0^{50} x^{\alpha-1}e^{-x/\beta}\ dx = \frac{1}{25}\int_0^{50} xe^{-x/5}\ dx = 0.9995.$$

(b) $P(X < 10) = \frac{1}{\beta^\alpha\Gamma(\alpha)} \int_0^{10} x^{\alpha-1}e^{-x/\beta}\ dx$. Using the incomplete gamma with $y = x/\beta$, we have

$$P(X < 10) = P(Y < 2) = \int_0^2 ye^{-y}\ dy = 0.5940.$$

6.53 $\mu = 3$ seconds with $f(x) = \frac{1}{3}e^{-x/3}$ for $x > 0$.

(a) $P(X > 5) = \int_5^\infty \frac{1}{3} e^{-x/3}\, dx = \frac{1}{3}\left[-3e^{-x/3}\right]\Big|_5^\infty = e^{-5/3} = 0.1889$.

(b) $P(X > 10) = e^{-10/3} = 0.0357$.

6.54 $P(X > 270) = 1 - \Phi\left(\frac{\ln 270 - 4}{2}\right) = 1 - \Phi(0.7992) = 0.2119$.

6.55 $\mu = E(X) = e^{4+4/2} = e^6$; $\sigma^2 = e^{8+4}(e^4 - 1) = e^{12}(e^4 - 1)$.

6.56 $\beta = 1/5$ and $\alpha = 10$.

(a) $P(X > 10) = 1 - P(X \le 10) = 1 - 0.9863 = 0.0137$.

(b) $P(X > 2)$ before 10 cars arrive.

$$P(X \le 2) = \int_0^2 \frac{1}{\beta^\alpha} \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)}\, dx.$$

Given $y = x/\beta$, then

$$P(X \le 2) = P(Y \le 10) = \int_0^{10} \frac{y^{\alpha-1} e^{-y}}{\Gamma(\alpha)}\, dy = \int_0^{10} \frac{y^{10-1} e^{-y}}{\Gamma(10)}\, dy = 0.542,$$

with $P(X > 2) = 1 - P(X \le 2) = 1 - 0.542 = 0.458$.

6.57 (a) $P(X > 1) = 1 - P(X \le 1) = 1 - 10\int_0^1 e^{-10x}\, dx = e^{-10} = 0.000045$.

(b) $\mu = \beta = 1/10 = 0.1$.

6.58 Assume that $Z(t) = \alpha\beta t^{\beta-1}$, for $t > 0$. Then we can write $f(t) = Z(t)R(t)$, where $R(t) = ce^{-\int Z(t)\, dt} = ce^{-\int \alpha\beta t^{\beta-1} dt} = ce^{-\alpha t^\beta}$. From the condition that $R(0) = 1$, we find that $c = 1$. Hence $R(t) = e^{\alpha t^\beta}$ and $f(t) = \alpha\beta t^{\beta-1} e^{-\alpha t^\beta}$, for $t > 0$. Since

$$Z(t) = \frac{f(t)}{R(t)},$$

where

$$R(t) = 1 - F(t) = 1 - \int_0^t \alpha\beta x^{\beta-1} e^{-\alpha x^\beta} dx = 1 + \int_0^t de^{-\alpha x^\beta} = e^{-\alpha t^\beta},$$

then

$$Z(t) = \frac{\alpha\beta t^{\beta-1} e^{-\alpha t^\beta}}{e^{-\alpha t^\beta}} = \alpha\beta t^{\beta-1}, \quad t > 0.$$

6.59 $\mu = np = (1000)(0.49) = 490$, $\sigma = \sqrt{npq} = \sqrt{(1000)(0.49)(0.51)} = 15.808$.

$$z_1 = \frac{481.5 - 490}{15.808} = -0.54, \qquad z_2 = \frac{510.5 - 490}{15.808} = 1.3.$$

$P(481.5 < X < 510.5) = P(-0.54 < Z < 1.3) = 0.9032 - 0.2946 = 0.6086$.

6.60 $P(X > 1/4) = \int_{1/4}^{\infty} 6e^{-6x}\, dx = -e^{-6x}|_{1/4}^{\infty} = e^{-1.5} = 0.223$.

6.61 $P(X < 1/2) = 108 \int_0^{1/2} x^2 e^{-6x}\, dx$. Letting $y = 6x$ and using Table A.24 we have

$$P(X < 1/2) = P(Y < 3) = \int_0^3 y^2 e^{-y}\, dy = 0.577.$$

6.62 Manufacturer A:

$$P(X \geq 10000) = P\left(Z \geq \frac{100000 - 14000}{2000}\right) = P(Z \geq -2) = 0.9772.$$

Manufacturer B:

$$P(X \geq 10000) = P\left(Z \geq \frac{10000 - 13000}{1000}\right) = P(Z \geq -3) = 0.9987.$$

Manufacturer B will produce the fewest number of defective rivets.

6.63 Using the normal approximation to the binomial with $\mu = np = 650$ and $\sigma = \sqrt{npq} = 15.0831$. So,

$$P(590 \leq X \leq 625) = P(-10.64 < Z < -8.92) \approx 0.$$

6.64 (a) $\mu = \beta = 100$ hours.

(b) $P(X \geq 200) = 0.01 \int_{200}^{\infty} e^{-0.01x}\, dx = e^{-2} = 0.1353$.

6.65 (a) $\mu = 85$ and $\sigma = 4$. So, $P(X < 80) = P(Z < -1.25) = 0.1056$.

(b) $\mu = 79$ and $\sigma = 4$. So, $P(X \geq 80) = P(Z > 0.25) = 0.4013$.

6.66 $1/\beta = 1/5$ hours with $\alpha = 2$ failures and $\beta = 5$ hours.

(a) $\alpha\beta = (2)(5) = 10$.

(b) $P(X \geq 12) = \int_{12}^{\infty} \frac{1}{5^2\Gamma(2)} x e^{-x/5}\, dx = \frac{1}{25}\int_{12}^{\infty} x e^{-x/5}\, dx = \left[-\frac{x}{5}e^{-x/5} - e^{-x/5}\right]_{12}^{\infty}$ $= 0.3084$.

6.67 Denote by X the elongation. We have $\mu = 0.05$ and $\sigma = 0.01$.

(a) $P(X \geq 0.1) = P\left(Z \geq \frac{0.1-0.05}{0.01}\right) = P(Z \geq 5) \approx 0$.

(b) $P(X \leq 0.04) = P\left(Z \leq \frac{0.04-0.05}{0.01}\right) = P(Z \leq -1) = 0.1587$.

(c) $P(0.025 \leq X \leq 0.065) = P(-2.5 \leq Z \leq 1.5) = 0.9332 - 0.0062 = 0.9270$.

6.68 Let X be the error. $X \sim n(x; 0, 4)$. So,

$$P(\text{fails}) = 1 - P(-10 < X < 10) = 1 - P(-2.25 < Z < 2.25) = 2(0.0122) = 0.0244.$$

6.69 Let X be the time to bombing with $\mu = 3$ and $\sigma = 0.5$. Then

$$P(1 \le X \le 4) = P\left(\frac{1-3}{0.5} \le Z \le \frac{4-3}{0.5}\right) = P(-4 \le Z \le 2) = 0.9772.$$

P(of an undesirable product) is $1 - 0.9772 = 0.0228$. Hence a product is undesirable is 2.28% of the time.

6.70 $\alpha = 2$ and $\beta = 100$. $P(X \le 200) = \frac{1}{\beta^2}\int_0^{200} xe^{-x/\beta}\, dx$. Using the incomplete gamma table and let $y = x/\beta$, $\int_0^2 ye^{-y}\, dy = 0.594$.

6.71 $\mu = \alpha\beta = 200$ hours and $\sigma^2 = \alpha\beta^2 = 20,000$ hours.

6.72 X follows a lognormal distribution.

$$P(X \ge 50,000) = 1 - \Phi\left(\frac{\ln 50,000 - 5}{2}\right) = 1 - \Phi(2.9099) = 1 - 0.9982 = 0.0018.$$

6.73 The mean of X, which follows a lognormal distribution is $\mu = E(X) = e^{\mu+\sigma^2/2} = e^7$.

6.74 $\mu = 10$ and $\sigma = \sqrt{50}$.

(a) $P(X \le 50) = P(Z \le 5.66) \approx 1$.

(b) $P(X \le 10) = 0.5$.

(c) The results are very similar.

6.75 (a) Since $f(y) \ge 0$ and $\int_0^1 10(1-y)^9\, dy = -\left.(1-y)^{10}\right|_0^1 = 1$, it is a density function.

(b) $P(Y > 0.6) = -\left.(1-y)^{10}\right|_{0.6}^1 = (0.4)^{10} = 0.0001$.

(c) $\alpha = 1$ and $\beta = 10$.

(d) $\mu = \frac{\alpha}{\alpha+\beta} = \frac{1}{11} = 0.0909$.

(e) $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{(1)(10)}{(1+10)^2(1+10+1)} = 0.006887$.

6.76 (a) $\mu = \frac{1}{10}\int_0^\infty ze^{-z/10}\, dz = -\left.ze^{-z/10}\right|_0^\infty + \int_0^\infty e^{-z/10}\, dz = 10$.

(b) Using integral by parts twice, we get

$$E(Z^2) = \frac{1}{10}\int_0^\infty z^2 e^{-z/10}\, dz = 200.$$

So, $\sigma^2 = E(Z^2) - \mu^2 = 200 - (10)^2 = 100$.

(c) $P(Z > 10) = -\left.e^{z/10}\right|_{10}^\infty = e^{-1} = 0.3679$.

6.77 This is an exponential distribution with $\beta = 10$.

(a) $\mu = \beta = 10$.

(b) $\sigma^2 = \beta^2 = 100$.

6.78 $\mu = 0.5$ seconds and $\sigma = 0.4$ seconds.

(a) $P(X > 0.3) = P\left(Z > \frac{0.3-0.5}{0.4}\right) = P(Z > -0.5) = 0.6915$.

(b) $P(Z > -1.645) = 0.95$. So, $-1.645 = \frac{x-0.5}{0.4}$ yields $x = -0.158$ seconds. The negative number in reaction time is not reasonable. So, it means that the normal model may not be accurate enough.

6.79 (a) For an exponential distribution with parameter β,

$$P(X > a + b \mid X > a) = \frac{P(X > a+b)}{P(X > a)} = \frac{e^{-a-b}}{e^{-a}} = e^{-b} = P(X > b).$$

So, $P(\text{it will breakdown in the next 21 days} \mid \text{it just broke down}) = P(X > 21) = e^{-21/15} = e^{-1.4} = 0.2466$.

(b) $P(X > 30) = e^{-30/15} = e^{-2} = 0.1353$.

6.80 $\alpha = 2$ and $\beta = 50$. So,

$$P(X \leq 10) = 100 \int_0^{10} x^{49} e^{-2x^{50}}\ dx.$$

Let $y = 2x^{50}$ with $x = (y/2)^{1/50}$ and $dx = \frac{1}{2^{1/50}(50)} y^{-49/50}\ dy$.

$$P(X \leq 10) = \frac{100}{2^{1/50}(50)} \int_0^{(2)10^{50}} \left(\frac{y}{2}\right)^{49/50} y^{-49/50} e^{-y}\ dy = \int_0^{(2)10^{50}} e^{-y}\ dy \approx 1.$$

6.81 The density function of a Weibull distribution is

$$f(y) = \alpha\beta y^{\beta-1} e^{-\alpha y^{\beta}}, \qquad y > 0.$$

So, for any $y \geq 0$,

$$F(y) = \int_0^y f(t)\ dt = \alpha\beta \int_0^y t^{\beta-1} e^{-\alpha t^{\beta}}\ dt.$$

Let $z = t^{\beta}$ which yields $t = z^{1/\beta}$ and $dt = \frac{1}{\beta} z^{1/\beta-1}\ dz$. Hence,

$$F(y) = \alpha\beta \int_0^{y^{\beta}} z^{1-1/\beta} \frac{1}{\beta} z^{1/\beta-1} e^{-\alpha z}\ dz = \alpha \int_0^{y^{\beta}} e^{-\alpha z}\ dz = 1 - e^{-\alpha y^{\beta}}.$$

On the other hand, since $de^{-\alpha y^{\beta}} = -\alpha\beta y^{\beta-1} e^{-\alpha y^{\beta}}$, the above result follows immediately.

6.82 One of the basic assumptions for the exponential distribution centers around the "lack-of-memory" property for the associated Poisson distribution. Thus the drill bit of problem 6.80 is assumed to have no punishment through wear if the exponential distribution applies. A drill bit is a mechanical part that certainly will have significant wear over time. Hence the exponential distribution would not apply.

6.83 The chi-squared distribution is a special case of the gamma distribution when $\alpha = v/2$ and $\beta = 2$, where v is the degrees of the freedom of the chi-squared distribution. So, the mean of the chi-squared distribution, using the property from the gamma distribution, is $\mu = \alpha\beta = (v/2)(2) = v$, and the variance of the chi-squared distribution is $\sigma^2 = \alpha\beta^2 = (v/2)(2)^2 = 2v$.

6.84 Let X be the length of time in seconds. Then $Y = \ln(X)$ follows a normal distribution with $\mu = 1.8$ and $\sigma = 2$.

(a) $P(X > 20) = P(Y > \ln 20) = P(Z > (\ln 20 - 1.8)/2) = P(Z > 0.60) = 0.2743$.
$P(X > 60) = P(Y > \ln 60) = P(Z > (\ln 60 - 1.8)/2) = P(Z > 1.15) = 0.1251$.

(b) The mean of the underlying normal distribution is $e^{1.8+4/2} = 44.70$ seconds. So, $P(X < 44.70) = P(Z < (\ln 44.70 - 1.8)/2) = P(Z < 1) = 0.8413$.

Chapter 7

Functions of Random Variables

7.1 From $y = 2x - 1$ we obtain $x = (y+1)/2$, and given $x = 1, 2$, and 3, then

$$g(y) = f[(y+1)/2] = 1/3, \qquad \text{for } y = 1, 3, 5.$$

7.2 From $y = x^2$, $x = 0, 1, 2, 3$, we obtain $x = \sqrt{y}$,

$$g(y) = f(\sqrt{y}) = \binom{3}{\sqrt{y}} \left(\frac{2}{5}\right)^{\sqrt{y}} \left(\frac{3}{5}\right)^{3-\sqrt{y}}, \qquad \text{for} y = 0, 1, 4, 9.$$

7.3 The inverse functions of $y_1 = x_1 + x_2$ and $y_2 = x_1 - x_2$ are $x_1 = (y_1 + y_2)/2$ and $x_2 = (y_1 - y_2)/2$. Therefore,

$$g(y_1, y_2) = \binom{2}{\frac{y_1+y_2}{2}, \frac{y_1-y_2}{2}, 2-y_1} \left(\frac{1}{4}\right)^{(y_1+y_2)/2} \left(\frac{1}{3}\right)^{(y_1-y_2)/2} \left(\frac{5}{12}\right)^{2-y_1},$$

where $y_1 = 0, 1, 2$, $y_2 = -2, -1, 0, 1, 2$, $y_2 \le y_1$ and $y_1 + y_2 = 0, 2, 4$.

7.4 Let $W = X_2$. The inverse functions of $y = x_1 x_2$ and $w = x_2$ are $x_1 = y/w$ and $x_2 = w$, where $y/w = 1, 2$. Then

$$g(y, w) = (y/w)(w/18) = y/18, \quad y = 1, 2, 3, 4, 6;\ w = 1, 2, 3,\ y/w = 1, 2.$$

In tabular form the joint distribution $g(y, w)$ and marginal $h(y)$ are given by

				y		
$g(y,w)$		1	2	3	4	6
	1	1/18	2/18			
w	2		2/18		4/18	
	3			3/18		6/18
$h(y)$		1/18	2/9	1/6	2/9	1/3

The alternate solutions are:

$P(Y=1) = f(1,1) = 1/18,$

$P(Y=2) = f(1,2) + f(2,1) = 2/18 + 2/18 = 2/9,$

$P(Y=3) = f(1,3) = 3/18 = 1/6,$

$P(Y=4) = f(2,2) = 4/18 = 2/9,$

$P(Y=6) = f(2,3) = 6/18 = 1/3.$

7.5 The inverse function of $y = -2\ln x$ is given by $x = e^{-y/2}$ from which we obtain $|J| = |-e^{-y/2}/2| = e^{-y/2}/2$. Now,

$$g(y) = f(e^{y/2})|J| = e^{-y/2}/2, \qquad y > 0,$$

which is a chi-squared distribution with 2 degrees of freedom.

7.6 The inverse function of $y = 8x^3$ is $x = y^{1/3}/2$, for $0 < y < 8$ from which we obtain $|J| = y^{-2/3}/6$. Therefore,

$$g(y) = f(y^{1/3}/2)|J| = 2(y^{1/3}/2)(y^{-2/3}/6) = \frac{1}{6}y^{-1/3}, \quad 0 < y < 8.$$

7.7 To find k we solve the equation $k\int_0^\infty v^2 e^{-bv^2}\,dv = 1$. Let $x = bv^2$, then $dx = 2bv\,dv$ and $dv = \frac{x^{-1/2}}{2\sqrt{b}}\,dx$. Then the equation becomes

$$\frac{k}{2b^{3/2}}\int_0^\infty x^{3/2-1}e^{-x}\,dx = 1, \quad \text{or} \quad \frac{k\Gamma(3/2)}{2b^{3/2}} = 1.$$

Hence $k = \frac{4b^{3/2}}{\Gamma(1/2)}$.

Now the inverse function of $w = mv^2/2$ is $v = \sqrt{2w/m}$, for $w > 0$, from which we obtain $|J| = 1/\sqrt{2mw}$. It follows that

$$g(w) = f(\sqrt{2w/m})|J| = \frac{4b^{3/2}}{\Gamma(1/2)}(2w/m)e^{-2bw/m} = \frac{1}{(m/2b)^{3/2}\Gamma(3/2)}w^{3/2-1}e^{-(2b/m)w},$$

for $w > 0$, which is a gamma distribution with $\alpha = 3/2$ and $\beta = m/2b$.

7.8 (a) The inverse of $y = x^2$ is $x = \sqrt{y}$, for $0 < y < 1$, from which we obtain $|J| = 1/2\sqrt{y}$. Therefore,

$$g(y) = f(\sqrt{y})|J| = 2(1-\sqrt{y})/2\sqrt{y} = y^{-1/2} - 1, \quad 0 < y < 1.$$

(b) $P(Y < 1) = \int_0^1 (y^{-1/2} - 1)\,dy = (2y^{1/2} - y)\Big|_0^1 = 0.5324.$

7.9 (a) The inverse of $y = x + 4$ is $x = y - 4$, for $y > 4$, from which we obtain $|J| = 1$. Therefore,

$$g(y) = f(y-4)|J| = 32/y^3, \quad y > 4.$$

(b) $P(Y > 8) = 32\int_8^\infty y^{-3}\,dy = -\,16y^{-2}\big|_8^\infty = \frac{1}{4}$.

7.10 (a) Let $W = X$. The inverse functions of $z = x + y$ and $w = x$ are $x = w$ and $y = z - w$, $0 < w < z$, $0 < z < 1$, from which we obtain

$$J = \begin{vmatrix} \frac{\partial x}{\partial w} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial w} & \frac{\partial y}{\partial z} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1.$$

Then $g(w, z) = f(w, z - w)|J| = 24w(z - w)$, for $0 < w < z$ and $0 < z < 1$. The marginal distribution of Z is

$$f_1(z) = \int_0^1 24(z - w)w\,dw = 4z^3, \qquad 0 < z < 1.$$

(b) $P(1/2 < Z < 3/4) = 4\int_{1/2}^{3/4} z^3\,dz = 65/256$.

7.11 The amount of kerosene left at the end of the day is $Z = Y - X$. Let $W = Y$. The inverse functions of $z = y - x$ and $w = y$ are $x = w - z$ and $y = w$, for $0 < z < w$ and $0 < w < 1$, from which we obtain

$$J = \begin{vmatrix} \frac{\partial x}{\partial w} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial w} & \frac{\partial y}{\partial z} \end{vmatrix} = \begin{vmatrix} 1 & -1 \\ 1 & 0 \end{vmatrix} = 1.$$

Now,

$$g(w, z) = g(w - z, w) = 2, \quad 0 < z < w,\ 0 < w < 1,$$

and the marginal distribution of Z is

$$h(z) = 2\int_z^1 dw = 2(1 - z), \quad 0 < z < 1.$$

7.12 Since X_1 and X_2 are independent, the joint probability distribution is

$$f(x_1, x_2) = f(x_1)f(x_2) = e^{-(x_1+x_2)}, \quad x_1 > 0,\ x_2 > 0.$$

The inverse functions of $y_1 = x_1 + x_2$ and $y_2 = x_1/(x_1 + x_2)$ are $x_1 = y_1y_2$ and $x_2 = y_1(1 - y_2)$, for $y_1 > 0$ and $0 < y_2 < 1$, so that

$$J = \begin{vmatrix} \partial x_1/\partial y_1 & \partial x_1/\partial y_2 \\ \partial x_2/\partial y_1 & \partial x_2/\partial y_2 \end{vmatrix} = \begin{vmatrix} y_2 & y_1 \\ 1 - y_2 & -y_1 \end{vmatrix} = -y_1.$$

Then, $g(y_1, y_2) = f(y_1y_2, y_1(1 - y_2))|J| = y_1e^{-y_1}$, for $y_1 > 0$ and $0 < y_2 < 1$. Therefore,

$$g(y_1) = \int_0^1 y_1e^{-y_1}\,dy_2 = y_1e^{-y_1}, \quad y_1 > 0,$$

and

$$g(y_2) = \int_0^\infty y_1e^{-y_1}\,dy_1 = \Gamma(2) = 1, \quad 0 < y_2 < 1.$$

Since $g(y_1, y_2) = g(y_1)g(y_2)$, the random variables Y_1 and Y_2 are independent.

7.13 Since I and R are independent, the joint probability distribution is

$$f(i, r) = 12ri(1 - i), \quad 0 < i < 1,\ 0 < r < 1.$$

Let $V = R$. The inverse functions of $w = i^2 r$ and $v = r$ are $i = \sqrt{w/v}$ and $r = v$, for $w < v < 1$ and $0 < w < 1$, from which we obtain

$$J = \begin{vmatrix} \partial i/\partial w & \partial i/\partial v \\ \partial r/\partial w & \partial r/\partial v \end{vmatrix} = \frac{1}{2\sqrt{vw}}.$$

Then,

$$g(w, v) = f(\sqrt{w/v}, v)|J| = 12v\sqrt{w/v}(1 - \sqrt{w/v})\frac{1}{2\sqrt{vw}} = 6(1 - \sqrt{w/v}),$$

for $w < v < 1$ and $0 < w < 1$, and the marginal distribution of W is

$$h(w) = 6\int_w^1 (1 - \sqrt{w/v})\ dv = 6\ (v - 2\sqrt{wv})\Big|_{v=w}^{v=1} = 6 + 6w - 12\sqrt{w}, \quad 0 < w < 1.$$

7.14 The inverse functions of $y = x^2$ are given by $x_1 = \sqrt{y}$ and $x_2 = -\sqrt{y}$ from which we obtain $J_1 = 1/2\sqrt{y}$ and $J_2 = 1/2\sqrt{y}$. Therefore,

$$g(y) = f(\sqrt{y})|J_1| + f(-\sqrt{y})|J_2| = \frac{1 + \sqrt{y}}{2} \cdot \frac{1}{2\sqrt{y}} + \frac{1 - \sqrt{y}}{2} \cdot \frac{1}{2\sqrt{y}} = 1/2\sqrt{y},$$

for $0 < y < 1$.

7.15 The inverse functions of $y = x^2$ are $x_1 = \sqrt{y}$, $x_2 = -\sqrt{y}$ for $0 < y < 1$ and $x_1 = \sqrt{y}$ for $0 < y < 4$. Now $|J_1| = |J_2| = |J_3| = 1/2\sqrt{y}$, from which we get

$$g(y) = f(\sqrt{y})|J_1| + f(-\sqrt{y})|J_2| = \frac{2(\sqrt{y} + 1)}{9} \cdot \frac{1}{2\sqrt{y}} + \frac{2(-\sqrt{y} + 1)}{9} \cdot \frac{1}{2\sqrt{y}} = \frac{2}{9\sqrt{y}},$$

for $0 < y < 1$ and

$$g(y) = f(\sqrt{y})|J_3| = \frac{2(\sqrt{y} + 1)}{9} \cdot \frac{1}{2\sqrt{y}} = \frac{\sqrt{y} + 1}{9\sqrt{y}}, \quad \text{for } 1 < y < 4.$$

7.16 Using the formula we obtain

$$\begin{aligned} \mu_r' = E(X^r) &= \int_0^\infty x^r \cdot \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}\ dx = \frac{\beta^{\alpha+r}\Gamma(\alpha + r)}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty \frac{x^{\alpha+r-1}e^{-x/\beta}}{\beta^{\alpha+r}\Gamma(\alpha + r)}\ dx \\ &= \frac{\beta^r \Gamma(\alpha + r)}{\Gamma(\alpha)}, \end{aligned}$$

since the second integrand is a gamma density with parameters$\alpha + r$ and β.

7.17 The moment-generating function of X is

$$M_X(t) = E(e^{tX}) = \frac{1}{k}\sum_{x=1}^{k} e^{tx} = \frac{e^t(1-e^{kt})}{k(1-e^t)},$$

by summing the geometric series of k terms.

7.18 The moment-generating function of X is

$$M_X(t) = E(e^{tX}) = p\sum_{x=1}^{\infty} e^{tx}q^{x-1} = \frac{p}{q}\sum_{x=1}^{\infty}(e^t q)^x = \frac{pe^t}{1-qe^t},$$

by summing an infinite geometric series. To find out the moments, we use

$$\mu = M_X^{'}(0) = \left.\frac{(1-qe^t)pe^t + pqe^{2t}}{(1-qe^t)^2}\right|_{t=0} = \frac{(1-q)p+pq}{(1-q)^2} = \frac{1}{p},$$

and

$$\mu_2^{'} = M_X^{''}(0) = \left.\frac{(1-qe^t)^2pe^t + 2pqe^{2t}(1-qe^t)}{(1-qe^t)^4}\right|_{t=0} = \frac{2-p}{p^2}.$$

So, $\sigma^2 = \mu_2^{'} - \mu^2 = \frac{q}{p^2}$.

7.19 The moment-generating function of a Poisson random variable is

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{\infty}\frac{e^{tx}e^{-\mu}\mu^x}{x!} = e^{-\mu}\sum_{x=0}^{\infty}\frac{(\mu e^t)^x}{x!} = e^{-\mu}e^{\mu e^t} = e^{\mu(e^t-1)}.$$

So,

$$\mu = M_X^{'}(0) = \mu\, e^{\mu(e^t-1)+t}\Big|_{t=0} = \mu,$$
$$\mu_2^{'} = M_X^{''}(0) = \mu e^{\mu(e^t-1)+t}(\mu e^t+1)\Big|_{t=0} = \mu(\mu+1),$$

and

$$\sigma^2 = \mu_2^{'} - \mu^2 = \mu(\mu+1) - \mu^2 = \mu.$$

7.20 From $M_X(t) = e^{4(e^t-1)}$ we obtain $\mu = 6$, $\sigma^2 = 4$, and $\sigma = 2$. Therefore,

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = P(0 < X < 8) = \sum_{x=1}^{7} p(x;4) = 0.9489 - 0.0183 = 0.9306.$$

7.21 Using the moment-generating function of the chi-squared distribution, we obtain

$$\mu = M_X'(0) = v(1-2t)^{-v/2-1}\Big|_{t=0} = v,$$
$$\mu_2' = M_X''(0) = v(v+2)\,(1-2t)^{-v/2-2}\Big|_{t=0} = v(v+2).$$

So, $\sigma^2 = \mu_2' - \mu^2 = v(v+2) - v^2 = 2v$.

7.22
$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} f(x)\ dx = \int_{-\infty}^{\infty}\left(1 + tx + \frac{t^2x^2}{2!} + \cdots + \frac{t^r x^r}{r!} + \cdots\right) f(x)\ dx \\ &= \int_{-\infty}^{\infty} f(x)\ dx + t\int_{-\infty}^{\infty} x f(x)\ dx + \frac{t^2}{2}\int_{-\infty}^{\infty} x^2 f(x)\ dx \\ &\quad + \cdots + \frac{t^r}{r!}\int_{-\infty}^{\infty} x^r f(x)\ dx + \cdots = 1 + \mu t + \mu_1' \frac{t^2}{2!} + \cdots + \mu_r' \frac{t^r}{r!} + \cdots . \end{aligned}$$

7.23 The joint distribution of X and Y is $f_{X,Y}(x,y) = e^{-x-y}$ for $x > 0$ and $y > 0$. The inverse functions of $u = x + y$ and $v = x/(x+y)$ are $x = uv$ and $y = u(1-v)$ with $J = \begin{vmatrix} v & u \\ 1-v & -u \end{vmatrix} = u$ for $u > 0$ and $0 < v < 1$. So, the joint distribution of U and V is

$$f_{U,V}(u,v) = ue^{-uv} \cdot e^{-u(1-v)} = ue^{-u},$$

for $u > 0$ and $0 < v < 1$.

(a) $f_U(u) = \int_0^1 ue^{-u}\ dv = ue^{-u}$ for $u > 0$, which is a gamma distribution with parameters 2 and 1.

(b) $f_V(v) = \int_0^\infty ue^{-u}\ du = 1$ for $0 < v < 1$. This is a uniform (0,1) distribution.

Chapter 8

Fundamental Sampling Distributions and Data Descriptions

8.1 (a) Responses of all people in Richmond who have telephones.

(b) Outcomes for a large or infinite number of tosses of a coin.

(c) Length of life of such tennis shoes when worn on the professional tour.

(d) All possible time intervals for this lawyer to drive from her home to her office.

8.2 (a) Number of tickets issued by all state troopers in Montgomery County during the Memorial holiday weekend.

(b) Number of tickets issued by all state troopers in South Carolina during the Memorial holiday weekend.

8.3 (a) $\bar{x} = 2.4$.

(b) $\bar{x} = 2$.

(c) $m = 3$.

8.4 (a) $\bar{x} = 8.6$ minutes.

(b) $\bar{x} = 9.5$ minutes.

(c) Mode are 5 and 10 minutes.

8.5 (a) $\bar{x} = 3.2$ seconds.

(b) $\bar{x} = 3.1$ seconds.

8.6 (a) $\bar{x} = 35.7$ grams.

(b) $\bar{x} = 32.5$ grams.

(c) Mode=29 grams.

8.7 (a) $\bar{x} = 53.75$.

(b) Modes are 75 and 100.

8.8 $\bar{x} = 22.2$ days, $\tilde{x} = 14$ days and $m = 8$ days. $\tilde{x}$ is the best measure of the center of the data. The mean should not be used on account of the extreme value 95, and the mode is not desirable because the sample size is too small.

8.9 (a) Range $= 15 - 5 = 10$.

(b) $s^2 = \frac{n\sum\limits_{i=1}^{n} x_i^2 - (\sum\limits_{i=1}^{n} x_i)^2}{n(n-1)} = \frac{(10)(838)-86^2}{(10)(9)} = 10.933$. Taking the square root, we have $s = 3.307$.

8.10 (a) Range $= 4.3 - 2.3 = 2.0$.

(b) $s^2 = \frac{n\sum\limits_{i=1}^{n} x_i^2 - (\sum\limits_{i=1}^{n} x_i)^2}{n(n-1)} = \frac{(9)(96.14)-28.8^2}{(9)(8)} = 0.498$.

8.11 (a) $s^2 = \frac{1}{n-1}\sum\limits_{x=1}^{n}(x_i - \bar{x})^2 = \frac{1}{14}[(2-2.4)^2 + (1-2.4)^2 + \cdots + (2-2.4)^2] = 2.971$.

(b) $s^2 = \frac{n\sum\limits_{i=1}^{n} x_i^2 - (\sum\limits_{i=1}^{n} x_i)^2}{n(n-1)} = \frac{(15)(128)-36^2}{(15)(14)} = 2.971$.

8.12 (a) $\bar{x} = 11.69$ milligrams.

(b) $s^2 = \frac{n\sum\limits_{i=1}^{n} x_i^2 - (\sum\limits_{i=1}^{n} x_i)^2}{n(n-1)} = \frac{(8)(1168.21)-93.5^2}{(8)(7)} = 10.776$.

8.13 $s^2 = \frac{n\sum\limits_{i=1}^{n} x_i^2 - (\sum\limits_{i=1}^{n} x_i)^2}{n(n-1)} = \frac{(2)(148.55)-53.3^2}{(20)(19)} = 0.342$ and hence $s = 0.585$.

8.14 (a) Replace X_i in S^2 by $X_i + c$ for $i = 1, 2, \ldots, n$. Then $\bar{X}$ becomes $\bar{X} + c$ and

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}[(X_i + c) - (\bar{X} + c)]^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

(b) Replace X_i by cX_i in S^2 for $i = 1, 2, \ldots, n$. Then $\bar{X}$ becomes $c\bar{X}$ and

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(cX_i - c\bar{X})^2 = \frac{c^2}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

8.15 $s^2 = \frac{n\sum\limits_{i=1}^{n} x_i^2 - (\sum\limits_{i=1}^{n} x_i)^2}{n(n-1)} = \frac{(6)(207)-33^2}{(6)(5)} = 5.1$.

(a) Multiplying each observation by 3 gives $s^2 = (9)(5.1) = 45.9$.

(b) Adding 5 to each observation does not change the variance. Hence $s^2 = 5.1$.

8.16 Denote by D the difference in scores.

(a) $\bar{D} = 25.15$.

(b) $\tilde{D} = 31.00$.

8.17 $z_1 = -1.9$, $z_2 = -0.4$. Hence,

$$P(\mu_{\bar{X}} - 1.9\sigma_{\bar{X}} < \bar{X} < \mu_{\bar{X}} - 0.4\sigma_{\bar{X}}) = P(-1.9 < Z < -0.4) = 0.3446 - 0.0287 = 0.3159.$$

8.18 $n = 54$, $\mu_{\bar{X}} = 4$, $\sigma^2_{\bar{X}} = \sigma^2/n = (8/3)/54 = 4/81$ with $\sigma_{\bar{X}} = 2/9$. So,

$$z_1 = (4.15 - 4)/(2/9) = 0.68, \quad \text{and} \quad z_2 = (4.35 - 4)/(2/9) = 1.58,$$

and

$$P(4.15 < \bar{X} < 4.35) = P(0.68 < Z < 1.58) = 0.9429 - 0.7517 = 0.1912.$$

8.19 (a) For $n = 64$, $\sigma_{\bar{X}} = 5.6/8 = 0.7$, whereas for $n = 196$, $\sigma_{\bar{X}} = 5.6/14 = 0.4$. Therefore, the variance of the sample mean is reduced from 0.49 to 0.16 when the sample size is increased from 64 to 196.

(b) For $n = 784$, $\sigma_{\bar{X}} = 5.6/28 = 0.2$, whereas for $n = 49$, $\sigma_{\bar{X}} = 5.6/7 = 0.8$. Therefore, the variance of the sample mean is increased from 0.04 to 0.64 when the sample size is decreased from 784 to 49.

8.20 $n = 36$, $\sigma_{\bar{X}} = 2$. Hence $\sigma = \sqrt{n}\sigma_{\bar{X}} = (6)(2) = 12$. If $\sigma_{\bar{X}} = 1.2$, then $1.2 = 12/\sqrt{n}$ and $n = 100$.

8.21 $\mu_{\bar{X}} = \mu = 240$, $\sigma_{\bar{X}} = 15/\sqrt{40} = 2.372$. Therefore, $\mu_{\bar{X}} \pm 2\sigma_{\bar{X}} = 240 \pm (2)(2.372)$ or from 235.257 to 244.743, which indicates that a value of $x = 236$ milliliters is reasonable and hence the machine needs not be adjusted.

8.22 (a) $\mu_{\bar{X}} = \mu = 174.5$, $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 6.9/5 = 1.38$.

(b) $z_1 = (172.45 - 174.5)/1.38 = -1.49$, $z_2 = (175.85 - 174.5)/1.38 = 0.98$. So,

$$P(172.45 < \bar{X} < 175.85) = P(-1.49 < Z < 0.98) = 0.8365 - 0.0681 = 0.7684.$$

Therefore, the number of sample means between 172.5 and 175.8 inclusive is $(200)(0.7684) = 154$.

(c) $z = (171.95 - 174.5)/1.38 = -1.85$. So,

$$P(\bar{X} < 171.95) = P(Z < -1.85) = 0.0322.$$

Therefore, about $(200)(0.0322) = 6$ sample means fall below 172.0 centimeters.

8.23 (a) $\mu = \sum xf(x) = (4)(0.2) + (5)(0.4) + (6)(0.3) + (7)(0.1) = 5.3$, and $\sigma^2 = \sum(x - \mu)^2 f(x) = (4 - 5.3)^2(0.2) + (5 - 5.3)^2(0.4) + (6 - 5.3)^2(0.3) + (7 - 5.3)^2(0.1) = 0.81$.

(b) With $n = 36$, $\mu_{\bar{X}} = \mu = 5.3$ and $\sigma_{\bar{X}} = \sigma^2/n = 0.81/36 = 0.0225$.

(c) $n = 36$, $\mu_{\bar{X}} = 5.3$, $\sigma_{\bar{X}} = 0.9/6 = 0.15$, and $z = (5.5 - 5.3)/0.15 = 1.33$. So,

$$P(\bar{X} < 5.5) = P(Z < 1.33) = 0.9082.$$

8.24 $n = 36$, $\mu_{\bar{X}} = 40$, $\sigma_{\bar{X}} = 2/6 = 1/3$ and $z = (40.5 - 40)/(1/3) = 1.5$. So,

$$P\left(\sum_{i=1}^{36} X_i > 1458\right) = P(\bar{X} > 40.5) = P(Z > 1.5) = 1 - 0.9332 = 0.0668.$$

8.25 (a) $P(6.4 < \bar{X} < 7.2) = P(-1.8 < Z < 0.6) = 0.6898$.

(b) $z = 1.04$, $\bar{x} = z(\sigma/\sqrt{n}) + \mu = (1.04)(1/3) + 7 = 7.35$.

8.26 $n = 64$, $\mu_{\bar{X}} = 3.2$, $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 1.6/8 = 0.2$.

(a) $z = (2.7 - 3.2)/0.2 = -2.5$, $P(\bar{X} < 2.7) = P(Z < -2.5) = 0.0062$.

(b) $z = (3.5 - 3.2)/0.2 = 1.5$, $P(\bar{X} > 3.5) = P(Z > 1.5) = 1 - 0.9332 = 0.0668$.

(c) $z_1 = (3.2 - 3.2)/0.2 = 0$, $z_2 = (3.4 - 3.2)/0.2 = 1.0$,
$P(3.2 < \bar{X} < 3.4) = P(0 < Z < 1.0) = 0.9413 - 0.5000 = 0.3413$.

8.27 $n = 50$, $\bar{x} = 0.23$ and $\sigma = 0.1$. Now, $z = (0.23 - 0.2)/(0.1/\sqrt{50}) = 2.12$; so

$$P(\bar{X} \geq 0.23) = P(Z \geq 2.12) = 0.0170.$$

Hence the probability of having such observations, given the mean $\mu = 0.20$, is small. Therefore, the mean amount to be 0.20 is not likely to be true.

8.28 $\mu_1 - \mu_2 = 80 - 75 = 5$, $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{25/25 + 9/36} = 1.118$, $z_1 = (3.35 - 5)/1.118 = -1.48$ and $z_2 = (5.85 - 5)/1.118 = 0.76$. So,

$$P(3.35 < \bar{X}_1 - \bar{X}_2 < 5.85) = P(-1.48 < Z < 0.76) = 0.7764 - 0.0694 = 0.7070.$$

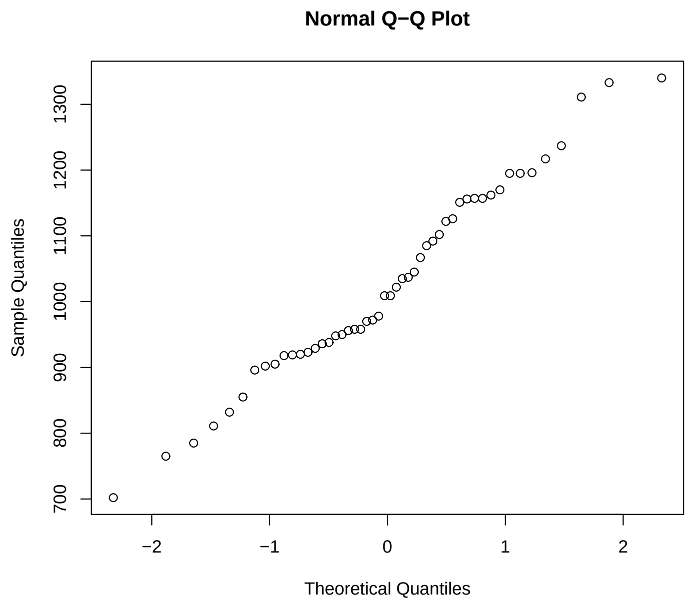
8.29 $\mu_{\bar{X}_1 - \bar{X}_2} = 72 - 28 = 44$, $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{100/64 + 25/100} = 1.346$ and $z = (44.2 - 44)/1.346 = 0.15$. So, $P(\bar{X}_1 - \bar{X}_2 < 44.2) = P(Z < 0.15) = 0.5596$.

8.30 $\mu_1 - \mu_2 = 0$, $\sigma_{\bar{X}_1 - \bar{X}_2} = 50\sqrt{1/32 + 1/50} = 11.319$.

(a) $z_1 = -20/11.319 = -1.77$, $z_2 = 20/11.319 = 1.77$, so
$P(|\bar{X}_1 - \bar{X}_2| > 20) = 2P(Z < -1.77) = (2)(0.0384) = 0.0768$.

(b) $z_1 = 5/11.319 = 0.44$ and $z_2 = 10/11.319 = 0.88$. So,
$P(-10 < \bar{X}_1 - \bar{X}_2 < -5) + P(5 < \bar{X}_1 - \bar{X}_2 < 10) = 2P(5 < \bar{X}_1 - \bar{X}_2 < 10) = 2P(0.44 < Z < 0.88) = 2(0.8106 - 0.6700) = 0.2812$.

8.31 The normal quantile-quantile plot is shown as



8.32 (a) If the two population mean drying times are truly equal, the probability that the difference of the two sample means is 1.0 is 0.0013, which is very small. This means that the assumption of the equality of the population means are not reasonable.

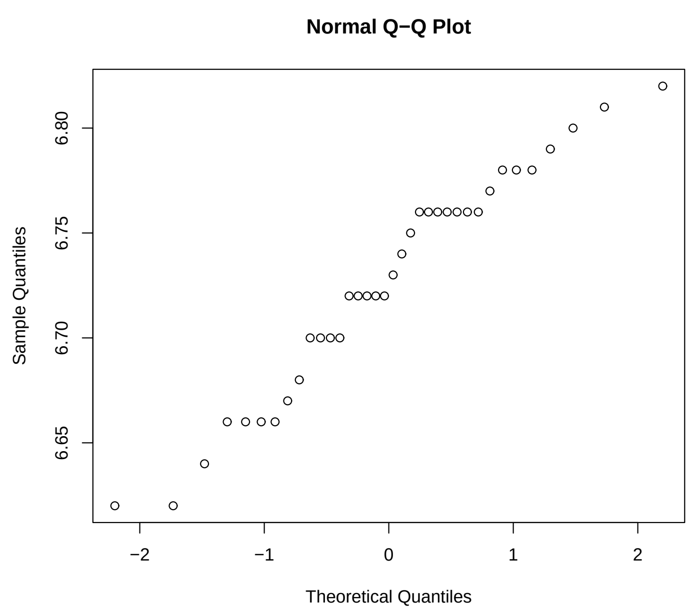
(b) If the experiment was run 10,000 times, there would be $(10000)(0.0013) = 13$ experiments where $\bar{X}_A - \bar{X}_B$ would be at least 1.0.

8.33 (a) $n_1 = n_2 = 36$ and $z = 0.2/\sqrt{1/36 + 1/36} = 0.85$. So,

$$P(\bar{X}_B - \bar{X}_A \geq 0.2) = P(Z \geq 0.85) = 0.1977.$$

(b) Since the probability in (a) is not negligible, the conjecture is not true.

8.34 The normal quantile-quantile plot is shown as



8.35 (a) When the population equals the limit, the probability of a sample mean exceeding the limit would be 1/2 due the symmetry of the approximated normal distribution.

(b) $P(\bar{X} \geq 7960 \mid \mu = 7950) = P(Z \geq (7960 - 7950)/(100/\sqrt{25})) = P(Z \geq 0.5) = 0.3085$. No, this is not very strong evidence that the population mean of the process exceeds the government limit.

8.36 (a) $\sigma_{\bar{X}_A - \bar{X}_B} = \sqrt{\frac{5^2}{30} + \frac{5^2}{30}} = 1.29$ and $z = \frac{4-0}{1.29} = 3.10$. So,

$$P(\bar{X}_A - \bar{X}_B > 4 \mid \mu_A = \mu_B) = P(Z > 3.10) = 0.0010.$$

Such a small probability means that the difference of 4 is not likely if the two population means are equal.

(b) Yes, the data strongly support alloy A.

8.37 Since the probability that $\bar{X} \leq 775$ is 0.0062, given that $\mu = 800$ is true, it suggests that this event is very rare and it is very likely that the claim of $\mu = 800$ is not true. On the other hand, if μ is truly, say, 760, the probability

$$P(\bar{X} \leq 775 \mid \mu = 760) = P(Z \leq (775 - 760)/(40/\sqrt{16})) = P(Z \leq 1.5) = 0.9332,$$

which is very high.

8.38 Define $W_i = \ln X_i$ for $i = 1, 2, \ldots$. Using the central limit theorem, $Z = (\bar{W} - \mu_{W_1})/(\sigma_{W_1}/\sqrt{n}) \sim n(z; 0, 1)$. Hence $\bar{W}$ follows, approximately, a normal distribution when n is large. Since

$$\bar{W} = \frac{1}{n}\sum_{i=1}^{n} \ln(X_i) = \frac{1}{n} \ln\left(\prod_{i=1}^{n} X_i\right) = \frac{1}{n}\ln(Y),$$

then it is easily seen that Y follows, approximately, a lognormal distribution.

8.39 (a) 27.488.

(b) 18.475.

(c) 36.415.

8.40 (a) 16.750.

(b) 30.144.

(c) 26.217.

8.41 (a) $\chi^2_\alpha = \chi^2_{0.99} = 0.297$.

(b) $\chi^2_\alpha = \chi^2_{0.025} = 32.852$.

(c) $\chi^2_{0.05} = 37.652$. Therefore, $\alpha = 0.05 - 0.045 = 0.005$. Hence, $\chi^2_\alpha = \chi^2_{0.005} = 46.928$.

8.42 (a) $\chi^2_\alpha = \chi^2_{0.01} = 38.932$.

(b) $\chi^2_\alpha = \chi^2_{0.05} = 12.592$.

(c) $\chi^2_{0.01} = 23.209$ and $\chi^2_{0.025} = 20.483$ with $\alpha = 0.01 + 0.015 = 0.025$.

8.43 (a) $P(S^2 > 9.1) = P\left(\frac{(n-1)S^2}{\sigma^2} > \frac{(24)(9.1)}{6}\right) = P(\chi^2 > 36.4) = 0.05$.

(b) $P(3.462 < S^2 < 10.745) = P\left(\frac{(24)(3.462)}{6} < \frac{(n-1)S^2}{\sigma^2} < \frac{(24)(10.745)}{6}\right)$
$= P(13.848 < \chi^2 < 42.980) = 0.95 - 0.01 = 0.94$.

8.44 $\chi^2 = \frac{(19)(20)}{8} = 47.5$ while $\chi^2_{0.01} = 36.191$. Conclusion values are not valid.

8.45 Since $\frac{(n-1)S^2}{\sigma^2}$ is a chi-square statistic, it follows that

$$\sigma^2_{(n-1)S^2/\sigma^2} = \frac{(n-1)^2}{\sigma^4}\sigma^2_{S^2} = 2(n-1).$$

Hence, $\sigma^2_{S^2} = \frac{2\sigma^4}{n-1}$, which decreases as n increases.

8.46 (a) 2.145.

(b) -1.372.

(c) -3.499.

8.47 (a) $P(T < 2.365) = 1 - 0.025 = 0.975$.

(b) $P(T > 1.318) = 0.10$.

(c) $P(T < 2.179) = 1 - 0.025 = 0.975$, $P(T < -1.356) = P(T > 1.356) = 0.10$. Therefore, $P(-1.356 < T < 2.179) = 0.975 - 0.010 = 0.875$.

(d) $P(T > -2.567) = 1 - P(T > 2.567) = 1 - 0.01 = 0.99$.

8.48 (a) Since $t_{0.01}$ leaves an area of 0.01 to the right, and $-t_{0.005}$ an area of 0.005 to the left, we find the total area to be $1 - 0.01 - 0.005 = 0.985$ between $-t_{0.005}$ and $t_{0.01}$. Hence, $P(-t_{0.005} < T < t_{0.01}) = 0.985$.

(b) Since $-t_{0.025}$ leaves an area of 0.025 to the left, the desired area is $1 - 0.025 = 0.975$. That is, $P(T > -t_{0.025}) = 0.975$.

8.49 (a) From Table A.4 we note that 2.069 corresponds to $t_{0.025}$ when $v = 23$. Therefore, $-t_{0.025} = -2.069$ which means that the total area under the curve to the left of $t = k$ is $0.025 + 0.965 = 0.990$. Hence, $k = t_{0.01} = 2.500$.

(b) From Table A.4 we note that 2.807 corresponds to $t_{0.005}$ when $v = 23$. Therefore the total area under the curve to the right of $t = k$ is $0.095 + 0.005 = 0.10$. Hence, $k = t_{0.10} = 1.319$.

(c) $t_{0.05} = 1.714$ for 23 degrees of freedom.

8.50 From Table A.4 we find $t_{0.025} = 2.131$ for $v = 15$ degrees of freedom. Since the value

$$t = \frac{27.5 - 30}{5/4} = -2.00$$

falls between -2.131 and 2.131, the claim is valid.

8.51 $t = (24 - 20)/(4.1/3) = 2.927$, $t_{0.01} = 2.896$ with 8 degrees of freedom. Conclusion: no, $\mu > 20$.

8.52 $\bar{x} = 0.475$, $s^2 = 0.0336$ and $t = (0.475 - 0.5)/0.0648 = -0.39$. Hence

$$P(\bar{X} < 0.475) = P(T < -0.39) \approx 0.35.$$

So, the result is inconclusive.

8.53 (a) 2.71.

(b) 3.51.

(c) 2.92.

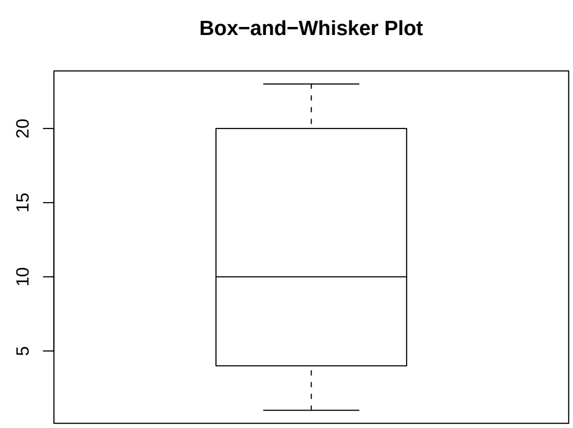
(d) $1/2.11 = 0.47$.

(e) $1/2.90 = 0.34$.

8.54 $s_1^2 = 10.441$ and $s_2^2 = 1.846$ which gives $f = 5.66$. Since, from Table A.6, $f_{0.05}(9, 7) = 3.68$ and $f_{0.01}(9, 7) = 6.72$, the probability of $P(F > 5.66)$ should be between 0.01 and 0.05, which is quite small. Hence the variances may not be equal. Furthermore, if a computer software can be used, the exact probability of $F > 5.66$ can be found 0.0162, or if two sides are considered, $P(F < 1/5.66) + P(F > 5.66) = 0.026$.

8.55 $s_1^2 = 15750$ and $s_2^2 = 10920$ which gives $f = 1.44$. Since, from Table A.6, $f_{0.05}(4, 5) = 5.19$, the probability of $F > 1.44$ is much bigger than 0.05, which means that the two variances may be considered equal. The actual probability of $F > 1.44$ is 0.3436 and $P(F < 1/1.44) + P(F > 1.44) = 0.7158$.

8.56 The box-and-whisker plot is shown below.



The sample mean = 12.32 and the sample standard deviation = 6.08.

8.57 The moment-generating function for the gamma distribution is given by

$$\begin{aligned}M_X(t) = E(e^{tX}) &= \frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^\infty e^{tx}x^{\alpha-1}e^{-x/\beta}\,dx\\ &= \frac{1}{\beta^\alpha(1/\beta-t)^\alpha}\frac{1}{(1/\beta-t)^{-\alpha}\Gamma(\alpha)}\int_0^\infty x^{\alpha-1}e^{-x\left(\frac{1}{\beta}-t\right)}\,dx\\ &= \frac{1}{(1-\beta t)^\alpha}\int_0^\infty \frac{x^{\alpha-1}e^{-x/(1/\beta-t)^{-1}}}{[(1/\beta-t)^{-1}]^\alpha\Gamma(\alpha)}\,dx = \frac{1}{(1-\beta t)^\alpha},\end{aligned}$$

for $t < 1/\beta$, since the last integral is one due to the integrand being a gamma density function. Therefore, the moment-generating function of an exponential distribution, by substituting α to 1, is given by $M_X(t) = (1-\theta t)^{-1}$. Hence, the moment-generating function of Y can be expressed as

$$M_Y(t) = M_{X_1}(t)M_{X_2}(t)\cdots M_{X_n}(t) = \prod_{i=1}^n (1-\theta t)^{-1} = (1-\theta t)^{-n},$$

which is seen to be the moment-generating function of a gamma distribution with $\alpha = n$ and $\beta = \theta$.

8.58 The variance of the carbon monoxide contents is the same as the variance of the coded measurements. That is, $s^2 = \frac{(15)(199.94)-39^2}{(15)(14)} = 7.039$, which results in $s = 2.653$.

8.59 $P\left(\frac{S_1^2}{S_2^2} < 4.89\right) = P\left(\frac{S_1^2/\sigma^2}{S_2^2/\sigma^2} < 4.89\right) = P(F < 4.89) = 0.99$, where F has 7 and 11 degrees of freedom.

8.60 $s^2 = 114,700,000$.

8.61 Let X_1 and X_2 be Poisson variables with parameters $\lambda_1 = 6$ and $\lambda_2 = 6$ representing the number of hurricanes during the first and second years, respectively. Then $Y = X_1 + X_2$ has a Poisson distribution with parameter $\lambda = \lambda_1 + \lambda_2 = 12$.

(a) $P(Y = 15) = \frac{e^{-12}12^{15}}{15!} = 0.0724$.

(b) $P(Y \le 9) = \sum_{y=0}^{9} \frac{e^{-12}12^y}{y!} = 0.2424$.

8.62 Dividing each observation by 1000 and then subtracting 55 yields the following data: -7, -2, -10, 6, 4, 1, 8, -6, -2, and -1. The variance of this coded data is

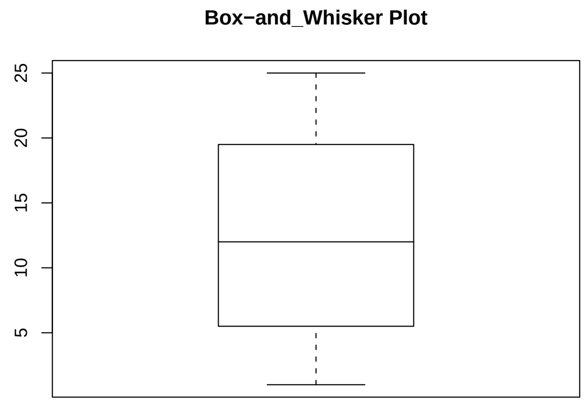
$$\frac{(10)(311)-(-9)^2}{(10)(9)} = 33.656.$$

Hence, with $c = 1000$, we have

$$s^2 = (1000)^2(33.656) = 33.656 \times 10^6,$$

and then $s = 5801$ kilometers.

8.63 The box-and-whisker plot is shown below.



The sample mean is 2.7967 and the sample standard deviation is 2.2273.

8.64 $P\left(\frac{S_1^2}{S_2^2} > 1.26\right) = P\left(\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} > \frac{(15)(1.26)}{10}\right) = P(F > 1.89) \approx 0.05$, where F has 24 and 30 degrees of freedom.

8.65 No outliers.

8.66 The value 32 is a possible outlier.

8.67 $\mu = 5{,}000$ psi, $\sigma = 400$ psi, and $n = 36$.

(a) Using approximate normal distribution (by CLT),

$$\begin{aligned} P(4800 < \bar{X} < 5200) &= P\left(\frac{4800 - 5000}{400/\sqrt{36}} < Z < \frac{5200 - 5000}{400/\sqrt{36}}\right) \\ &= P(-3 < Z < 3) = 0.9974. \end{aligned}$$

(b) To find a z such that $P(-z < Z < z) = 0.99$, we have $P(Z < z) = 0.995$, which results in $z = 2.575$. Hence, by solving $2.575 = \frac{5100-5000}{400/\sqrt{n}}$ we have $n \geq 107$. Note that the value n can be affected by the z values picked (2.57 or 2.58).

8.68 $\bar{x} = 54{,}100$ and $s = 5801.34$. Hence

$$t = \frac{54100 - 53000}{5801.34/\sqrt{10}} = 0.60.$$

So, $P(\bar{X} \geq 54,100) = P(T \geq 0.60)$ is a value between 0.20 and 0.30, which is not a rare event.

8.69 $n_A = n_B = 20$, $\bar{x}_A = 20.50$, $\bar{x}_B = 24.50$, and $\sigma_A = \sigma_B = 5$.

(a) $P(\bar{X}_A - \bar{X}_B \geq 4.0 \mid \mu_A = \mu_B) = P(Z > (24.5 - 20.5)/\sqrt{5^2/20 + 5^2/20})$
$= P(Z > 4.5/(5/\sqrt{10})) = P(Z > 2.85) = 0.0022.$

(b) It is extremely unlikely that $\mu_A = \mu_B$.

8.70 (a) $n_A = 30$, $\bar{x}_A = 64.5\%$ and $\sigma_A = 5\%$. Hence,

$$P(\bar{X}_A \leq 64.5 \mid \mu_A = 65) = P(Z < (64.5 - 65)/(5/\sqrt{30})) = P(Z < -0.55) = 0.2912.$$

There is no evidence that the μ_A is less than 65%.

(b) $n_B = 30$, $\bar{x}_B = 70\%$ and $\sigma_B = 5\%$. It turns out $\sigma_{\bar{X}_B - \bar{X}_A} = \sqrt{\frac{5^2}{30} + \frac{5^2}{30}} = 1.29\%$. Hence,

$$P(\bar{X}_B - \bar{X}_A \geq 5.5 \mid \mu_A = \mu_B) = P\left(Z \geq \frac{5.5}{1.29}\right) = P(Z \geq 4.26) \approx 0.$$

It does strongly support that μ_B is greater than μ_A.

(c) i) Since $\sigma_{\bar{X}_B} = \frac{5}{\sqrt{30}} = 0.9129$, $\bar{X}_B \sim n(x; 65\%, 0.9129\%)$.
ii) $\bar{X}_A - \bar{X}_B \sim n(x; 0, 1.29\%)$.
iii) $\frac{\bar{X}_A - \bar{X}_B}{\sigma\sqrt{2/30}} \sim n(z; 0, 1)$.

8.71 $P(\bar{X}_B \geq 70) = P\left(Z \geq \frac{70-65}{0.9129}\right) = P(Z \geq 5.48) \approx 0$.

8.72 It is known, from Table A.3, that $P(-1.96 < Z < 1.96) = 0.95$. Given $\mu = 20$ and $\sigma = \sqrt{9} = 3$, we equate $1.96 = \frac{20.1-20}{3/\sqrt{n}}$ to obtain $n = \left(\frac{(3)(1.96)}{0.1}\right)^2 = 3457.44 \approx 3458$.

8.73 It is known that $P(-2.575 < Z < 2.575) = 0.99$. Hence, by equating $2.575 = \frac{0.05}{1/\sqrt{n}}$, we obtain $n = \left(\frac{2.575}{0.05}\right)^2 = 2652.25 \approx 2653$.

8.74 $\mu = 9$ and $\sigma = 1$. Then

$$P(9 - 1.5 < X < 9 + 1.5) = P(7.5 < X < 10.5) = P(-1.5 < Z < 1.5) = 0.9322 - 0.0668 = 0.8654.$$

Thus the proportion of defective is $1 - 0.8654 = 0.1346$. To meet the specifications 99% of the time, we need to equate $2.575 = \frac{1.5}{\sigma}$, since $P(-2.575 < Z < 2.575) = 0.99$. Therefore, $\sigma = \frac{1.5}{2.575} = 0.5825$.

8.75 With the 39 degrees of freedom,

$$P(S^2 \leq 0.188 \mid \sigma^2 = 1.0) = P(\chi^2 \leq (39)(0.188)) = P(\chi^2 \leq 7.332) \approx 0,$$

which means that it is impossible to observe $s^2 = 0.188$ with $n = 40$ for $\sigma^2 = 1$.

Note that Table A.5 does not provide any values for the degrees of freedom to be larger than 30. However, one can deduce the conclusion based on the values in the last line of the table. Also, computer software gives the value of 0.

Chapter 9

One- and Two-Sample Estimation Problems

9.1 From Example 9.1 on page 271, we know that $E(S^2) = \sigma^2$. Therefore,

$$E(S'^2) = E\left[\frac{n-1}{n}S^2\right] = \frac{n-1}{n}E(S^2) = \frac{n-1}{n}\sigma^2.$$

9.2 (a) $E(X) = np$; $E(\hat{P}) = E(X/n) = E(X)/n = np/n = p$.

(b) $E(P') = \frac{E(X)+\sqrt{n}/2}{n+\sqrt{n}} = \frac{np+\sqrt{n}/2}{n+\sqrt{n}} \neq p$.

9.3 $\lim\limits_{n\to\infty} \frac{np+\sqrt{n}/2}{n+\sqrt{n}} = \lim\limits_{n\to\infty} \frac{p+1/2\sqrt{n}}{1+1/\sqrt{n}} = p$.

9.4 $n = 30, \bar{x} = 780$, and $\sigma = 40$. Also, $z_{0.02} = 2.054$. So, a 96% confidence interval for the population mean can be calculated as

$$780 - (2.054)(40/\sqrt{30}) < \mu < 780 + (2.054)(40/\sqrt{30}),$$

or $765 < \mu < 795$.

9.5 $n = 75, \bar{x} = 0.310, \sigma = 0.0015$, and $z_{0.025} = 1.96$. A 95% confidence interval for the population mean is

$$0.310 - (1.96)(0.0015/\sqrt{75}) < \mu < 0.310 + (1.96)(0.0015/\sqrt{75}),$$

or $0.3097 < \mu < 0.3103$.

9.6 $n = 50, \bar{x} = 174.5, \sigma = 6.9$, and $z_{0.01} = 2.33$.

(a) A 98% confidence interval for the population mean is
$174.5 - (2.33)(6.9/\sqrt{50}) < \mu < 174.5 + (2.33)(6.9/\sqrt{50})$, or $172.23 < \mu < 176.77$.

(b) $e < (2.33)(6.9)/\sqrt{50} = 2.27$.

9.7 $n = 100, \bar{x} = 23,500, \sigma = 3900$, and $z_{0.005} = 2.575$.

(a) A 99% confidence interval for the population mean is
$23,500 - (2.575)(3900/10) < \mu < 23,500 + (2.575)(3900/10)$, or
$22,496 < \mu < 24,504$.

(b) $e < (2.575)(3900/10) = 1004$.

9.8 $n = [(2.05)(40)/10]^2 = 68$ when rounded up.

9.9 $n = [(1.96)(0.0015)/0.0005]^2 = 35$ when rounded up.

9.10 $n = [(1.96)(40)/15]^2 = 28$ when rounded up.

9.11 $n = [(2.575)(5.8)/2]^2 = 56$ when rounded up.

9.12 $n = 20, \bar{x} = 11.3, s = 2.45$, and $t_{0.025} = 2.093$ with 19 degrees of freedom. A 95% confidence interval for the population mean is

$$11.3 - (2.093)(2.45/\sqrt{20}) < \mu < 11.3 + (2.093)(2.45/\sqrt{20}),$$

or $10.15 < \mu < 12.45$.

9.13 $n = 9, \bar{x} = 1.0056, s = 0.0245$, and $t_{0.005} = 3.355$ with 8 degrees of freedom. A 99% confidence interval for the population mean is

$$1.0056 - (3.355)(0.0245/3) < \mu < 1.0056 + (3.355)(0.0245/3),$$

or $0.978 < \mu < 1.033$.

9.14 $n = 10, \bar{x} = 230, s = 15$, and $t_{0.005} = 3.25$ with 9 degrees of freedom. A 99% confidence interval for the population mean is

$$230 - (3.25)(15/\sqrt{10}) < \mu < 230 + (3.25)(15/\sqrt{10}),$$

or $214.58 < \mu < 245.42$.

9.15 $n = 12, \bar{x} = 48.50, s = 1.5$, and $t_{0.05} = 1.796$ with 11 degrees of freedom. A 90% confidence interval for the population mean is

$$48.50 - (1.796)(1.5/\sqrt{12}) < \mu < 48.50 + (1.796)(1.5/\sqrt{12}),$$

or $47.722 < \mu < 49.278$.

9.16 $n = 12, \bar{x} = 79.3, s = 7.8$, and $t_{0.025} = 2.201$ with 11 degrees of freedom. A 95% confidence interval for the population mean is

$$79.3 - (2.201)(7.8/\sqrt{12}) < \mu < 79.3 + (2.201)(7.8/\sqrt{12}),$$

or $74.34 < \mu < 84.26$.

9.17 $n = 25, \bar{x} = 325.05, s = 0.5, \gamma = 5\%$, and $1 - \alpha = 90\%$, with $k = 2.208$. So, $325.05 \pm (2.208)(0.5)$ yields $(323.946, 326.151)$. Thus, we are 95% confident that this tolerance interval will contain 90% of the aspirin contents for this brand of buffered aspirin.

9.18 $n = 15, \bar{x} = 3.7867, s = 0.9709, \gamma = 1\%$, and $1 - \alpha = 95\%$, with $k = 3.507$. So, by calculating $3.7867 \pm (3.507)(0.9709)$ we obtain $(0.382, 7.192)$ which is a 99% tolerance interval that will contain 95% of the drying times.

9.19 $n = 100, \bar{x} = 23{,}500, s = 3,900, 1 - \alpha = 0.99$, and $\gamma = 0.01$, with $k = 3.096$. The tolerance interval is $23{,}500 \pm (3.096)(3{,}900)$ which yields $11{,}425 < \mu < 35{,}574$.

9.20 $n = 12, \bar{x} = 48.50, s = 1.5, 1 - \alpha = 0.90$, and $\gamma = 0.05$, with $k = 2.655$. The tolerance interval is $48.50 \pm (2.655)(1.5)$ which yields $(44.52, 52.48)$.

9.21 By definition, $MSE = E(\hat{\Theta} - \theta)^2$ which can be expressed as

$$\begin{aligned} MSE &= E[\hat{\Theta} - E(\hat{\Theta}) + E(\hat{\Theta}) - \theta]^2 \\ &= E[\hat{\Theta} - E(\hat{\Theta})]^2 + E[E(\hat{\Theta}) - \theta]^2 + 2E[\hat{\Theta} - E(\hat{\Theta})]E[E(\hat{\Theta}) - \theta]. \end{aligned}$$

The third term on the right hand side is zero since $E[\hat{\Theta} - E(\hat{\Theta})] = E[\hat{\Theta}] - E(\hat{\Theta}) = 0$. Hence the claim is valid.

9.22 (a) The bias is $E(S'^2) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = \frac{\sigma^2}{n}$.

(b) $\lim\limits_{n\to\infty} \text{Bias} = \lim\limits_{n\to\infty} \frac{\sigma^2}{n} = 0$.

9.23 Using Theorem 8.4, we know that $X^2 = \frac{(n-1)S^2}{\sigma^2}$ follows a chi-squared distribution with $n - 1$ degrees of freedom, whose variance is $2(n - 1)$. So, $Var(S^2) = Var\left(\frac{\sigma^2}{n-1}X^2\right) = \frac{2}{n-1}\sigma^4$, and $Var(S'^2) = Var\left(\frac{n-1}{n}S^2\right) = \left(\frac{n-1}{n}\right)^2 Var(S^2) = \frac{2(n-1)\sigma^4}{n^2}$. Therefore, the variance of S'^2 is smaller.

9.24 Using Exercises 9.21 and 9.23,

$$\begin{aligned} \frac{MSE(S^2)}{MSE(S'^2)} &= \frac{Var(S^2) + [\text{Bias}(S^2)]^2}{Var(S'^2) + [\text{Bias}(S'^2)]^2} = \frac{2\sigma^4/(n-1)}{2(n-1)\sigma^4/n^2 + \sigma^4/n^2} \\ &= 1 + \frac{3n - 1}{2n^2 - 3n + 1}, \end{aligned}$$

which is always larger than 1 when n is larger than 1. Hence the MSE of S'^2 is usually smaller.

9.25 $n = 20, \bar{x} = 11.3, s = 2.45$, and $t_{0.025} = 2.093$ with 19 degrees of freedom. A 95% prediction interval for a future observation is

$$11.3 \pm (2.093)(2.45)\sqrt{1 + 1/20} = 11.3 \pm 5.25,$$

which yields $(6.05, 16.55)$.

9.26 $n = 12, \bar{x} = 79.3, s = 7.8$, and $t_{0.025} = 2.201$ with 11 degrees of freedom. A 95% prediction interval for a future observation is

$$79.3 \pm (2.201)(7.8)\sqrt{1 + 1/12} = 79.3 \pm 17.87,$$

which yields $(61.43, 97.17)$.

9.27 $n = 15, \bar{x} = 3.7867, s = 0.9709$, and $t_{0.025} = 2.145$ with 14 degrees of freedom. A 95% prediction interval for a new observation is

$$3.7867 \pm (2.145)(0.9709)\sqrt{1 + 1/15} = 3.7867 \pm 2.1509,$$

which yields $(1.6358, 5.9376)$.

9.28 $n = 9, \bar{x} = 1.0056, s = 0.0245, 1 - \alpha = 0.95$, and $\gamma = 0.05$, with $k = 3.532$. The tolerance interval is $1.0056 \pm (3.532)(0.0245)$ which yields $(0.919, 1.092)$.

9.29 $n = 15, \bar{x} = 3.84$, and $s = 3.07$. To calculate an upper 95% prediction limit, we obtain $t_{0.05} = 1.761$ with 14 degrees of freedom. So, the upper limit is $3.84 + (1.761)(3.07)\sqrt{1 + 1/15} = 3.84 + 5.58 = 9.42$. This means that a new observation will have a chance of 95% to fall into the interval $(-\infty, 9.42)$. To obtain an upper 95% tolerance limit, using $1 - \alpha = 0.95$ and $\gamma = 0.05$, with $k = 2.566$, we get $3.84 + (2.566)(3.07) = 11.72$. Hence, we are 95% confident that $(-\infty, 11.72)$ will contain 95% of the orthophosphorous measurements in the river.

9.30 $n = 50, \bar{x} = 78.3$, and $s = 5.6$. Since $t_{0.05} = 1.677$ with 49 degrees of freedom, the bound of a lower 95% prediction interval for a single new observation is $78.3 - (1.677)(5.6)\sqrt{1 + 1/50} = 68.91$. So, the interval is $(68.91, \infty)$. On the other hand, with $1 - \alpha = 95\%$ and $\gamma = 0.01$, the k value for a one-sided tolerance limit is 2.269 and the bound is $78.3 - (2.269)(5.6) = 65.59$. So, the tolerance interval is $(65.59, \infty)$.

9.31 Since the manufacturer would be more interested in the mean tensile strength for future products, it is conceivable that prediction interval and tolerance interval may be more interesting than just a confidence interval.

9.32 This time $1 - \alpha = 0.99$ and $\gamma = 0.05$ with $k = 3.126$. So, the tolerance limit is $78.3 - (3.126)(5.6) = 60.79$. Since 62 exceeds the lower bound of the interval, yes, this is a cause of concern.

9.33 In Exercise 9.27, a 95% prediction interval for a new observation is calculated as $(1.6358, 5.9377)$. Since 6.9 is in the outside range of the prediction interval, this new observation is likely to be an outlier.

9.34 $n = 12, \bar{x} = 48.50, s = 1.5, 1 - \alpha = 0.95$, and $\gamma = 0.05$, with $k = 2.815$. The lower bound of the one-sided tolerance interval is $48.50 - (2.815)(1.5) = 44.275$. Their claim is not necessarily correct.

9.35 $n_1 = 25, n_2 = 36, \bar{x}_1 = 80, \bar{x}_2 = 75, \sigma_1 = 5, \sigma_2 = 3$, and $z_{0.03} = 1.88$. So, a 94% confidence interval for $\mu_1 - \mu_2$ is

$$(80-75) - (1.88)\sqrt{25/25 + 9/36} < \mu_1 - \mu_2 < (80-75) + (1.88)\sqrt{25/25 + 9/36},$$

which yields $2.9 < \mu_1 - \mu_2 < 7.1$.

9.36 $n_A = 50, n_B = 50, \bar{x}_A = 78.3, \bar{x}_B = 87.2, \sigma_A = 5.6$, and $\sigma_B = 6.3$. It is known that $z_{0.025} = 1.96$. So, a 95% confidence interval for the difference of the population means is

$$(87.2 - 78.3) \pm 1.96\sqrt{5.6^2/50 + 6.3^2/50} = 8.9 \pm 2.34,$$

or $6.56 < \mu_A - \mu_B < 11.24$.

9.37 $n_1 = 100, n_2 = 200, \bar{x}_1 = 12.2, \bar{x}_2 = 9.1, s_1 = 1.1$, and $s_2 = 0.9$. It is known that $z_{0.01} = 2.327$. So

$$(12.2 - 9.1) \pm 2.327\sqrt{1.1^2/100 + 0.9^2/200} = 3.1 \pm 0.30,$$

or $2.80 < \mu_1 - \mu_2 < 3.40$. The treatment appears to reduce the mean amount of metal removed.

9.38 $n_1 = 12, n_2 = 10, \bar{x}_1 = 85, \bar{x}_2 = 81, s_1 = 4, s_2 = 5$, and $s_p = 4.478$ with $t_{0.05} = 1.725$ with 20 degrees of freedom. So

$$(85 - 81) \pm (1.725)(4.478)\sqrt{1/12 + 1/10} = 4 \pm 3.31,$$

which yields $0.69 < \mu_1 - \mu_2 < 7.31$.

9.39 $n_1 = 12, n_2 = 18, \bar{x}_1 = 84, \bar{x}_2 = 77, s_1 = 4, s_2 = 6$, and $s_p = 5.305$ with $t_{0.005} = 2.763$ with 28 degrees of freedom. So,

$$(84 - 77) \pm (2.763)(5.305)\sqrt{1/12 + 1/18} = 7 \pm 5.46,$$

which yields $1.54 < \mu_1 - \mu_2 < 12.46$.

9.40 $n_1 = 10, n_2 = 10, \bar{x}_1 = 0.399, \bar{x}_2 = 0.565, s_1 = 0.07279, s_2 = 0.18674$, and $s_p = 0.14172$ with $t_{0.025} = 2.101$ with 18 degrees of freedom. So,

$$(0.565 - 0.399) \pm (2.101)(0.14172)\sqrt{1/10 + 1/10} = 0.166 \pm 0.133,$$

which yields $0.033 < \mu_1 - \mu_2 < 0.299$.

9.41 $n_1 = 14, n_2 = 16, \bar{x}_1 = 17, \bar{x}_2 = 19, s_1^2 = 1.5, s_2^2 = 1.8$, and $s_p = 1.289$ with $t_{0.005} = 2.763$ with 28 degrees of freedom. So,

$$(19 - 17) \pm (2.763)(1.289)\sqrt{1/16 + 1/14} = 2 \pm 1.30,$$

which yields $0.70 < \mu_1 - \mu_2 < 3.30$.

9.42 $n_1 = 12, n_2 = 10, \bar{x}_1 = 16, \bar{x}_2 = 11, s_1 = 1.0, s_2 = 0.8$, and $s_p = 0.915$ with $t_{0.05} = 1.725$ with 20 degrees of freedom. So,

$$(16 - 11) \pm (1.725)(0.915)\sqrt{1/12 + 1/10} = 5 \pm 0.68,$$

which yields $4.3 < \mu_1 - \mu_2 < 5.7$.

9.43 $n_A = n_B = 12, \bar{x}_A = 36,300, \bar{x}_B = 38,100, s_A = 5,000, s_B = 6,100$, and

$$v = \frac{5000^2/12 + 6100^2/12}{\frac{(5000^2/12)^2}{11} + \frac{(6100^2/12)^2}{11}} = 21,$$

with $t_{0.025} = 2.080$ with 21 degrees of freedom. So,

$$(36,300 - 38,100) \pm (2.080)\sqrt{\frac{5000^2}{12} + \frac{6100^2}{12}} = -1,800 \pm 4,736,$$

which yields $-6,536 < \mu_A - \mu_B < 2,936$.

9.44 $n = 8, \bar{d} = -1112.5, s_d = 1454$, with $t_{0.005} = 3.499$ with 7 degrees of freedom. So,

$$-1112.5 \pm (3.499)\frac{1454}{\sqrt{8}} = -1112.5 \pm 1798.7,$$

which yields $-2911.2 < \mu_D < 686.2$.

9.45 $n = 9, \bar{d} = 2.778, s_d = 4.5765$, with $t_{0.025} = 2.306$ with 8 degrees of freedom. So,

$$2.778 \pm (2.306)\frac{4.5765}{\sqrt{9}} = 2.778 \pm 3.518,$$

which yields $-0.74 < \mu_D < 6.30$.

9.46 $n_I = 5, n_{II} = 7, \bar{x}_I = 98.4, \bar{x}_{II} = 110.7, s_I = 8.375$, and $s_{II} = 32.185$, with

$$v = \frac{(8.735^2/5 + 32.185^2/7)2}{\frac{(8.735^2/5)^2}{4} + \frac{(32.185^2/7)^2}{6}} = 7$$

So, $t_{0.05} = 1.895$ with 7 degrees of freedom.

$$(110.7 - 98.4) \pm 1.895\sqrt{8.735^2/5 + 32.185^2/7} = 12.3 \pm 24.2,$$

which yields $-11.9 < \mu_{II} - \mu_I < 36.5$.

9.47 $n = 10, \bar{d} = 14.89\%$, and $s_d = 30.4868$, with $t_{0.025} = 2.262$ with 9 degrees of freedom. So,

$$14.89 \pm (2.262)\frac{30.4868}{\sqrt{10}} = 14.89 \pm 21.81,$$

which yields $-6.92 < \mu_D < 36.70$.

9.48 $n_A = n_B = 20, \bar{x}_A = 32.91, \bar{x}_B = 30.47, s_A = 1.57, s_B = 1.74$, and $S_p = 1.657$.

(a) $t_{0.025} \approx 2.042$ with 38 degrees of freedom. So,

$$(32.91 - 30.47) \pm (2.042)(1.657)\sqrt{1/20 + 1/20} = 2.44 \pm 1.07,$$

which yields $1.37 < \mu_A - \mu_B < 3.51$.

(b) Since it is apparent that type A battery has longer life, it should be adopted.

9.49 $n_A = n_B = 15, \bar{x}_A = 3.82, \bar{x}_B = 4.94, s_A = 0.7794, s_B = 0.7538$, and $s_p = 0.7667$ with $t_{0.025} = 2.048$ with 28 degrees of freedom. So,

$$(4.94 - 3.82) \pm (2.048)(0.7667)\sqrt{1/15 + 1/15} = 1.12 \pm 0.57,$$

which yields $0.55 < \mu_B - \mu_A < 1.69$.

9.50 $n_1 = 8, n_2 = 13, \bar{x}_1 = 1.98, \bar{x}_2 = 1.30, s_1 = 0.51, s_2 = 0.35$, and $s_p = 0.416$. $t_{0.025} = 2.093$ with 19 degrees of freedom. So,

$$(1.98 - 1.30) \pm (2.093)(0.416)\sqrt{1/8 + 1/13} = 0.68 \pm 0.39,$$

which yields $0.29 < \mu_1 - \mu_2 < 1.07$.

9.51 (a) $n = 200, \hat{p} = 0.57, \hat{q} = 0.43$, and $z_{0.02} = 2.05$. So,

$$0.57 \pm (2.05)\sqrt{\frac{(0.57)(0.43)}{200}} = 0.57 \pm 0.072,$$

which yields $0.498 < p < 0.642$.

(b) Error $\leq (2.05)\sqrt{\frac{(0.57)(0.43)}{200}} = 0.072$.

9.52 $n = 500.\hat{p} = \frac{485}{500} = 0.97, \hat{q} = 0.03$, and $z_{0.05} = 1.645$. So,

$$0.97 \pm (1.645)\sqrt{\frac{(0.97)(0.03)}{500}} = 0.97 \pm 0.013,$$

which yields $0.957 < p < 0.983$.

9.53 $n = 1000, \hat{p} = \frac{228}{1000} = 0.228, \hat{q} = 0.772$, and $z_{0.005} = 2.575$. So,

$$0.228 \pm (2.575)\sqrt{\frac{(0.228)(0.772)}{1000}} = 0.228 \pm 0.034,$$

which yields $0.194 < p < 0.262$.

9.54 $n = 100, \hat{p} = \frac{8}{100} = 0.08, \hat{q} = 0.92$, and $z_{0.01} = 2.33$. So,

$$0.08 \pm (2.33)\sqrt{\frac{(0.08)(0.92)}{100}} = 0.08 \pm 0.063,$$

which yields $0.017 < p < 0.143$.

9.55 (a) $n = 40, \hat{p} = \frac{34}{40} = 0.85, \hat{q} = 0.15$, and $z_{0.025} = 1.96$. So,

$$0.85 \pm (1.96)\sqrt{\frac{(0.85)(0.15)}{40}} = 0.85 \pm 0.111,$$

which yields $0.739 < p < 0.961$.

(b) Since $p = 0.8$ falls in the confidence interval, we can not conclude that the new system is better.

9.56 $n = 100, \hat{p} = \frac{24}{100} = 0.24, \hat{q} = 0.76$, and $z_{0.005} = 2.575$.

(a) $0.24 \pm (2.575)\sqrt{\frac{(0.24)(0.76)}{100}} = 0.24 \pm 0.110$, which yields $0.130 < p < 0.350$.

(b) Error $\leq (2.575)\sqrt{\frac{(0.24)(0.76)}{100}} = 0.110$.

9.57 $n = 1600, \hat{p} = \frac{2}{3}, \hat{q} = \frac{1}{3}$, and $z_{0.025} = 1.96$.

(a) $\frac{2}{3} \pm (1.96)\sqrt{\frac{(2/3)(1/3)}{1600}} = \frac{2}{3} \pm 0.023$, which yields $0.644 < p < 0.690$.

(b) Error $\leq (1.96)\sqrt{\frac{(2/3)(1/3)}{1600}} = 0.023$.

9.58 $n = \frac{(1.96)^2(0.32)(0.68)}{(0.02)^2} = 2090$ when round up.

9.59 $n = \frac{(2.05)^2(0.57)(0.43)}{(0.02)^2} = 2576$ when round up.

9.60 $n = \frac{(2.575)^2(0.228)(0.772)}{(0.05)^2} = 467$ when round up.

9.61 $n = \frac{(2.33)^2(0.08)(0.92)}{(0.05)^2} = 160$ when round up.

9.62 $n = \frac{(1.96)^2}{(4)(0.01)^2} = 9604$ when round up.

9.63 $n = \frac{(2.575)^2}{(4)(0.01)^2} = 16577$ when round up.

9.64 $n = \frac{(1.96)^2}{(4)(0.04)^2} = 601$ when round up.

9.65 $n_M = n_F = 1000, \hat{p}_M = 0.250, \hat{q}_M = 0.750, \hat{p}_F = 0.275, \hat{q}_F = 0.725$, and $z_{0.025} = 1.96$. So

$$(0.275 - 0.250) \pm (1.96)\sqrt{\frac{(0.250)(0.750)}{1000} + \frac{(0.275)(0.725)}{1000}} = 0.025 \pm 0.039,$$

which yields $-0.0136 < p_F - p_M < 0.0636$.

9.66 $n_1 = 250, n_2 = 175, \hat{p}_1 = \frac{80}{250} = 0.32, \hat{p}_2 = \frac{40}{175} = 0.2286$, and $z_{0.05} = 1.645$. So,

$$(0.32 - 0.2286) \pm (1.645)\sqrt{\frac{(0.32)(0.68)}{250} + \frac{(0.2286)(0.7714)}{175}} = 0.0914 \pm 0.0713,$$

which yields $0.0201 < p_1 - p_2 < 0.1627$. From this study we conclude that there is a significantly higher proportion of women in electrical engineering than there is in chemical engineering.

9.67 $n_1 = n_2 = 500, \hat{p}_1 = \frac{120}{500} = 0.24, \hat{p}_2 = \frac{98}{500} = 0.196$, and $z_{0.05} = 1.645$. So,

$$(0.24 - 0.196) \pm (1.645)\sqrt{\frac{(0.24)(0.76)}{500} + \frac{(0.196)(0.804)}{500}} = 0.044 \pm 0.0429,$$

which yields $0.0011 < p_1 - p_2 < 0.0869$. Since 0 is not in this confidence interval, we conclude, at the level of 90% confidence, that inoculation has an effect on the incidence of the disease.

9.68 $n_{5^\circ\text{C}} = n_{15^\circ\text{C}} = 20, \hat{p}_{5^\circ\text{C}} = 0.50, \hat{p}_{15^\circ\text{C}} = 0.75$, and $z_{0.025} = 1.96$. So,

$$(0.5 - 0.75) \pm (1.96)\sqrt{\frac{(0.50)(0.50)}{20} + \frac{(0.75)(0.25)}{20}} = -0.25 \pm 0.2899,$$

which yields $-0.5399 < p_{5^\circ\text{C}} - p_{15^\circ\text{C}} < 0.0399$. Since this interval includes 0, the significance of the difference cannot be shown at the confidence level of 95%.

9.69 $n_{\text{now}} = 1000, \hat{p}_{\text{now}} = 0.2740, n_{91} = 760, \hat{p}_{91} = 0.3158$, and $z_{0.025} = 1.96$. So,

$$(0.2740 - 0.3158) \pm (1.96)\sqrt{\frac{(0.2740)(0.7260)}{1000} + \frac{(0.3158)(0.6842)}{760}} = -0.0418 \pm 0.0431,$$

which yields $-0.0849 < p_{\text{now}} - p_{91} < 0.0013$. Hence, at the confidence level of 95%, the significance cannot be shown.

9.70 $n_{90} = n_{94} = 20, \hat{p}_{90} = 0.337$, and $\hat{0}_{94} = 0.362$

(a) $n_{90}\hat{p}_{90} = (20)(0.337) \approx 7$ and $n_{94}\hat{p}_{94} = (20)(0.362) \approx 7$.

(b) Since $z_{0.025} = 1.96$, $(0.337 - 0.362) \pm (1.96)\sqrt{\frac{(0.337)(0.663)}{20} + \frac{(0.362)(0.638)}{20}} = -0.025 \pm 0.295$, which yields $-0.320 < p_{90} - p_{94} < 0.270$. Hence there is no evidence, at the confidence level of 95%, that there is a change in the proportions.

9.71 $s^2 = 0.815$ with $v = 4$ degrees of freedom. Also, $\chi^2_{0.025} = 11.143$ and $\chi^2_{0.975} = 0.484$. So,

$$\frac{(4)(0.815)}{11.143} < \sigma^2 < \frac{(4)(0.815)}{0.484}, \quad \text{which yields} \quad 0.293 < \sigma^2 < 6.736.$$

Since this interval contains 1, the claim that σ^2 seems valid.

9.72 $s^2 = 16$ with $v = 19$ degrees of freedom. It is known $\chi^2_{0.01} = 36.191$ and $\chi^2_{0.99} = 7.633$. Hence

$$\frac{(19)(16)}{36.191} < \sigma^2 < \frac{(19)(16)}{7.633}, \text{ or } 8.400 < \sigma^2 < 39.827.$$

9.73 $s^2 = 6.0025$ with $v = 19$ degrees of freedom. Also, $\chi^2_{0.025} = 32.852$ and $\chi^2_{0.975} = 8.907$. Hence,

$$\frac{(19)(6.0025)}{32.852} < \sigma^2 < \frac{(19)(6.0025)}{8.907}, \text{ or } 3.472 < \sigma^2 < 12.804,$$

9.74 $s^2 = 0.0006$ with $v = 8$ degrees of freedom. Also, $\chi^2_{0.005} = 21.955$ and $\chi^2_{0.995} = 1.344$. Hence,

$$\frac{(8)(0.0006)}{21.955} < \sigma^2 < \frac{(8)(0.0006)}{1.344}, \text{ or } 0.00022 < \sigma^2 < 0.00357.$$

9.75 $s^2 = 225$ with $v = 9$ degrees of freedom. Also, $\chi^2_{0.005} = 23.589$ and $\chi^2_{0.995} = 1.735$. Hence,

$$\frac{(9)(225)}{23.589} < \sigma^2 < \frac{(9)(225)}{1.735}, \text{ or } 85.845 < \sigma^2 < 1167.147,$$

which yields $9.27 < \sigma < 34.16$.

9.76 $s^2 = 2.25$ with $v = 11$ degrees of freedom. Also, $\chi^2_{0.05} = 19.675$ and $\chi^2_{0.95} = 4.575$. Hence,

$$\frac{(11)(2.25)}{19.675} < \sigma^2 < \frac{(11)(2.25)}{4.575}, \text{ or } 1.258 < \sigma^2 < 5.410.$$

9.77 $s_1^2 = 1.00, s_2^2 = 0.64$, $f_{0.01}(11, 9) = 5.19$, and $f_{0.01}(9, 11) = 4.63$. So,

$$\frac{1.00/0.64}{5.19} < \frac{\sigma_1^2}{\sigma_2^2} < (1.00/0.64)(4.63), \text{ or } 0.301 < \frac{\sigma_1^2}{\sigma_2^2} < 7.234,$$

which yields $0.549 < \frac{\sigma_1}{\sigma_2} < 2.690$.

9.78 $s_1^2 = 5000^2, s_2^2 = 6100^2$, and $f_{0.05}(11, 11) = 2.82$. (Note: this value can be found by using "=finv(0.05,11,11)" in Microsoft Excel.) So,

$$\left(\frac{5000}{6100}\right)^2 \frac{1}{2.82} < \frac{\sigma_1^2}{\sigma_2^2} < \left(\frac{5000}{6100}\right)^2 (2.82), \text{ or } 0.238 < \frac{\sigma_1^2}{\sigma_2^2} < 1.895.$$

Since the interval contains 1, it is reasonable to assume that $\sigma_1^2 = \sigma_2^2$.

9.79 $s_I^2 = 76.3, s_{II}^2 = 1035.905$, $f_{0.05}(4, 6) = 4.53$, and $f_{0.05}(6, 4) = 6.16$. So,

$$\left(\frac{76.3}{1035.905}\right)\left(\frac{1}{4.53}\right) < \frac{\sigma_I^2}{\sigma_{II}^2} < \left(\frac{76.3}{1035.905}\right)(6.16), \text{ or } 0.016 < \frac{\sigma_I^2}{\sigma_{II}^2} < 0.454.$$

Hence, we may assume that $\sigma_I^2 \neq \sigma_{II}^2$.

9.80 $s_A = 0.7794, s_B = 0.7538$, and $f_{0.025}(14, 14) = 2.98$ (Note: this value can be found by using "=finv(0.025,14,14)" in Microsoft Excel.) So,

$$\left(\frac{0.7794}{0.7538}\right)^2 \left(\frac{1}{2.98}\right) < \frac{\sigma_A^2}{\sigma_B^2} < \left(\frac{0.7794}{0.7538}\right)^2 (2.98), \text{ or } 0.623 < \frac{\sigma_A^2}{\sigma_B^2} < 3.186.$$

Hence, it is reasonable to assume the equality of the variances.

9.81 The likelihood function is

$$L(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i; p) = \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{n\bar{x}}(1-p)^{n(1-\bar{x})}.$$

Hence, $\ln L = n[\bar{x}\ln(p) + (1-\bar{x})\ln(1-p)]$. Taking derivative with respect to p and setting the derivative to zero, we obtain $\frac{\partial \ln(L)}{\partial p} = n\left(\frac{\bar{x}}{p} - \frac{1-\bar{x}}{1-p}\right) = 0$, which yields $\frac{\bar{x}}{p} - \frac{1-\bar{x}}{1-p} = 0$. Therefore, $\hat{p} = \bar{x}$.

9.82 (a) The likelihood function is

$$\begin{aligned} L(x_1, \dots, x_n) &= \prod_{i=1}^n f(x_i; \alpha, \beta) = (\alpha\beta)^n \prod_{i=1}^n x_i^{\beta-1} e^{-\alpha x_i^\beta} \\ &= (\alpha\beta)^n e^{-\alpha \sum\limits_{i=1}^n x_i^\beta} \left(\prod_{i=1}^n x_i\right)^{\beta-1}. \end{aligned}$$

(b) So, the log-likelihood can be expressed as

$$\ln L = n[\ln(\alpha) + \ln(\beta)] - \alpha \sum_{i=1}^n x_i^\beta + (\beta - 1)\sum_{i=1}^n \ln(x_i).$$

To solve for the maximum likelihood estimate, we need to solve the following two equations

$$\frac{\partial \ln L}{\partial \alpha} = 0, \text{ and } \frac{\partial \ln L}{\partial \beta} = 0.$$

9.83 (a) The likelihood function is

$$\begin{aligned} L(x_1, \dots, x_n) &= \prod_{i=1}^n f(x_i; \mu, \sigma) = \prod_{i=1}^n \left\{ \frac{1}{\sqrt{2\pi}\sigma x_i} e^{-\frac{[\ln(x_i)-\mu]^2}{2\sigma^2}} \right\} \\ &= \frac{1}{(2\pi)^{n/2}\sigma^n \prod\limits_{i=1}^n x_i} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n [\ln(x_i) - \mu]^2 \right\}. \end{aligned}$$

(b) It is easy to obtain

$$\ln L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \sum_{i=1}^{n}\ln(x_i) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}[\ln(x_i)-\mu]^2.$$

So, setting $0 = \frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}[\ln(x_i)-\mu]$, we obtain $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\ln(x_i)$, and setting $0 = \frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}[\ln(x_i)-\mu]^2$, we get $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}[\ln(x_i)-\hat{\mu}]^2$.

9.84 (a) The likelihood function is

$$L(x_1,\dots,x_n) = \frac{1}{\beta^{n\alpha}\Gamma(\alpha)^n}\prod_{i=1}^{n}\left(x_i^{\alpha-1}e^{-x_i/\beta}\right) = \frac{1}{\beta^{n\alpha}\Gamma(\alpha)^n}\left(\prod_{i=1}^{n}x_i\right)^{\alpha-1} e^{-\sum_{i=1}^{n}(x_i/\beta)}.$$

(b) Hence

$$\ln L = -n\alpha\ln(\beta) - n\ln(\Gamma(\alpha)) + (\alpha-1)\sum_{i=1}^{n}\ln(x_i) - \frac{1}{\beta}\sum_{i=1}^{n}x_i.$$

Taking derivatives of $\ln L$ with respect to α and β, respectively and setting both as zeros. Then solve them to obtain the maximum likelihood estimates.

9.85 $L(x) = p^x(1-p)^{1-x}$, and $\ln L = x\ln(p) + (1-x)\ln(1-p)$, with $\frac{\partial \ln L}{\partial p} = \frac{x}{p} - \frac{1-x}{1-p} = 0$, we obtain $\hat{p} = x = 1$.

9.86 From the density function $b^*(x;p) = \binom{x-1}{k-1}p^k(1-p)^{x-k}$, we obtain

$$\ln L = \ln\binom{x-1}{k-1} + k\ln p + (n-k)\ln(1-p).$$

Setting $\frac{\partial \ln L}{\partial p} = \frac{k}{p} - \frac{n-k}{1-p} = 0$, we obtain $\hat{p} = \frac{k}{n}$.

9.87 For the estimator S^2,

$$\begin{aligned} Var(S^2) &= \frac{1}{(n-1)^2}Var\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right] = \frac{1}{(n-1)^2}Var(\sigma^2\chi^2_{n-1}) \\ &= \frac{1}{(n-1)^2}\sigma^4[2(n-1)] = \frac{2\sigma^4}{n-1}. \end{aligned}$$

For the estimator $\widehat{\sigma^2}$, we have

$$Var(\widehat{\sigma^2}) = \frac{2\sigma^4(n-1)}{n^2}.$$

9.88 $n = 7, \bar{d} = 3.557, s_d = 2.776$, and $t_{0.025} = 2.447$ with 6 degrees of freedom. So,

$$3.557 \pm (2.447)\frac{2.776}{\sqrt{7}} = 3.557 \pm 2.567,$$

which yields $0.99 < \mu_D < 6.12$. Since 0 is not in the interval, the claim appears valid.

9.89 $n = 75, x = 28$, hence $\hat{p} = \frac{28}{75} = 0.3733$. Since $z_{0.025} = 1.96$, a 95% confidence interval for p can be calculate as

$$0.3733 \pm (1.96)\sqrt{\frac{(0.3733)(0.6267)}{75}} = 0.3733 \pm 0.1095,$$

which yields $0.2638 < p < 0.4828$. Since the interval contains 0.421, the claim made by the *Roanoke Times* seems reasonable.

9.90 $n = 12, \bar{d} = 40.58, s_d = 15.791$, and $t_{0.025} = 2.201$ with 11 degrees of freedom. So,

$$40.58 \pm (2.201)\frac{15.791}{\sqrt{12}} = 40.58 \pm 10.03,$$

which yields $30.55 < \mu_D < 50.61$.

9.91 $n = 6, \bar{d} = 1.5, s_d = 1.543$, and $t_{0.025} = 2.571$ with 5 degrees of freedom. So,

$$1.5 \pm (2.571)\frac{1.543}{\sqrt{6}} = 1.5 \pm 1.62,$$

which yields $-0.12 < \mu_D < 3.12$.

9.92 $n = 12, \bar{d} = 417.5, s_d = 1186.643$, and $t_{0.05} = 1.796$ with 11 degrees of freedom. So,

$$417.5 \pm (1.796)\frac{1186.643}{\sqrt{12}} = 417.5 \pm 615.23,$$

which yields $-197.73 < \mu_D < 1032.73$.

9.93 $n_p = n_u = 8, \bar{x}_p = 86,250.000, \bar{x}_u = 79,837.500, \sigma_p = \sigma_u = 4,000$, and $z_{0.025} = 1.96$. So,

$$(86250 - 79837.5) \pm (1.96)(4000)\sqrt{1/8 + 1/8} = 6412.5 \pm 3920,$$

which yields $2,492.5 < \mu_p - \mu_u < 10,332.5$. Hence, polishing does increase the average endurance limit.

9.94 $n_A = 100, n_B = 120, \hat{p}_A = \frac{24}{100} = 0.24, \hat{p}_B = \frac{36}{120} = 0.30$, and $z_{0.025} = 1.96$. So,

$$(0.30 - 0.24) \pm (1.96)\sqrt{\frac{(0.24)(0.76)}{100} + \frac{(0.30)(0.70)}{120}} = 0.06 \pm 0.117,$$

which yields $-0.057 < p_B - p_A < 0.177$.

9.95 $n_N = n_O = 23, s_N^2 = 105.9271, s_O^2 = 77.4138$, and $f_{0.025}(22, 22) = 2.358$. So,

$$\frac{105.9271}{77.4138}\frac{1}{2.358} < \frac{\sigma_N^2}{\sigma_O^2} < \frac{105.9271}{77.4138}(2.358), \text{ or } 0.58 < \frac{\sigma_N^2}{\sigma_O^2} < 3.23.$$

For the ratio of the standard deviations, the 95% confidence interval is approximately

$$0.76 < \frac{\sigma_N}{\sigma_O} < 1.80.$$

Since the intervals contain 1 we will assume that the variability did not change with the local supplier.

9.96 $n_A = n_B = 6, \bar{x}_A = 0.1407, \bar{x}_B = 0.1385, s_A = 0.002805, s_B = 0.002665$, and $s_p = 002736$. Using a 90% confidence interval for the difference in the population means, $t_{0.05} = 1.812$ with 10 degrees of freedom, we obtain

$$(0.1407 - 0.1385) \pm (1.812)(0.002736)\sqrt{1/6 + 1/6} = 0.0022 \pm 0.0029,$$

which yields $-0.0007 < \mu_A - \mu_B < 0.0051$. Since the 90% confidence interval contains 0, we conclude that wire A was not shown to be better than wire B, with 90% confidence.

9.97 To calculate the maximum likelihood estimator, we need to use

$$\ln L = \ln\left(\frac{e^{-n\mu}\mu^{\sum\limits_{i=1}^{n} x_i}}{\prod\limits_{i=1}^{n} x_i!}\right) = -n\mu + \ln(\mu)\sum_{i=1}^{n} x_i - \ln(\prod_{i=1}^{n} x_i!).$$

Taking derivative with respect to μ and setting it to zero, we obtain $\hat{\mu} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i = \bar{x}$. On the other hand, using the method of moments, we also get $\hat{\mu} = \bar{x}$.

9.98 $\hat{\mu} = \bar{x}$ and $\hat{\sigma}^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2$.

9.99 Equating $\bar{x} = e^{\mu+\sigma^2/2}$ and $s^2 = (e^{2\mu+\sigma^2})(e^{\sigma^2}-1)$, we get $\ln(\bar{x}) = \mu + \frac{\sigma^2}{2}$, or $\hat{\mu} = \ln(\bar{x}) - \frac{\hat{\sigma}^2}{2}$. On the other hand, $\ln s^2 = 2\mu + \sigma^2 + \ln(e^{\sigma^2} - 1)$. Plug in the form of $\hat{\mu}$, we obtain $\hat{\sigma}^2 = \ln\left(1 + \frac{s^2}{\bar{x}^2}\right)$.

9.100 Setting $\bar{x} = \alpha\beta$ and $s^2 = \alpha\beta^2$, we get $\hat{\alpha} = \frac{\bar{x}^2}{s^2}$, and $\hat{\beta} = \frac{s^2}{\bar{x}}$.

9.101 $n_1 = n_2 = 300, \bar{x}_1 = 102300, \bar{x}_2 = 98500, s_1 = 5700$, and $s_2 = 3800$.

(a) $z_{0.005} = 2.575$. Hence,

$$(102300 - 98500) \pm (2.575)\sqrt{\frac{5700^2}{300} + \frac{3800^2}{300}} = 3800 \pm 1018.46,$$

which yields $2781.54 < \mu_1 - \mu_2 < 4818.46$. There is a significant difference in salaries between the two regions.

(b) Since the sample sizes are large enough, it is not necessary to assume the normality due to the Central Limit Theorem.

(c) We assumed that the two variances are not equal. Here we are going to obtain a 95% confidence interval for the ratio of the two variances. It is known that $f_{0.025}(299, 299) = 1.255$. So,

$$\left(\frac{5700}{3800}\right)^2 \frac{1}{1.255} < \frac{\sigma_1^2}{\sigma_2^2} < \left(\frac{5700}{3800}\right)^2 (1.255), \text{ or } 1.793 < \frac{\sigma_1^2}{\sigma_2^2} < 2.824.$$

Since the confidence interval does not contain 1, the difference between the variances is significant.

9.102 The error in estimation, with 95% confidence, is $(1.96)(4000)\sqrt{\frac{2}{n}}$. Equating this quantity to 1000, we obtain

$$n = 2\left(\frac{(1.96)(4000)}{1000}\right)^2 = 123,$$

when round up. Hence, the sample sizes in Review Exercise 9.101 is sufficient to produce a 95% confidence interval on $\mu_1 - \mu_2$ having a width of \$1,000.

9.103 $n = 300, \bar{x} = 6.5$ and $s = 2.5$. Also, $1 - \alpha = 0.99$ and $1 - \gamma = 0.95$. Using Table A.7, $k = 2.522$. So, the limit of the one-sided tolerance interval is $6.5 + (2.522)(2.5) = 12.805$. Since this interval contains 10, the claim by the union leaders appears valid.

9.104 $n = 30, x = 8$, and $z_{0.025} = 1.96$. So,

$$\frac{4}{15} \pm (1.96)\sqrt{\frac{(4/15)(11/15)}{30}} = \frac{4}{15} \pm 0.158,$$

which yields $0.108 < p < 0.425$.

9.105 $n = \frac{(1.96)^2(4/15)(11/15)}{0.05^2} = 301$, when round up.

9.106 $n_1 = n_2 = 100, \hat{p}_1 = 0.1$, and $\hat{p}_2 = 0.06$.

(a) $z_{0.025} = 1.96$. So,

$$(0.1 - 0.06) \pm (1.96)\sqrt{\frac{(0.1)(0.9)}{100} + \frac{(0.06)(0.94)}{100}} = 0.04 \pm 0.075,$$

which yields $-0.035 < p_1 - p_2 < 0.115$.

(b) Since the confidence interval contains 0, it does not show sufficient evidence that $p_1 > p_2$.

9.107 $n = 20$ and $s^2 = 0.045$. It is known that $\chi^2_{0.025} = 32.825$ and $\chi^2_{0.975} = 8.907$ with 19 degrees of freedom. Hence the 95% confidence interval for σ^2 can be expressed as

$$\frac{(19)(0.045)}{32.825} < \sigma^2 < \frac{(19)(0.045)}{8.907}, \text{ or } 0.012 < \sigma^2 < 0.045.$$

Therefore, the 95% confidence interval for σ can be approximated as

$$0.110 < \sigma < 0.212.$$

Since 0.3 falls outside of the confidence interval, there is strong evidence that the process has been improved in variability.

9.108 $n_A = n_B = 15, \bar{y}_A = 87, s_A = 5.99, \bar{y}_B = 75, s_B = 4.85, s_p = 5.450$, and $t_{0.025} = 2.048$ with 28 degrees of freedom. So,

$$(87 - 75) \pm (2.048)(5.450)\sqrt{\frac{1}{15} + \frac{1}{15}} = 12 \pm 4.076,$$

which yields $7.924 < \mu_A - \mu_B < 16.076$. Apparently, the mean operating costs of type A engines are higher than those of type B engines.

9.109 Since the unbiased estimators of σ_1^2 and σ_2^2 are S_1^2 and S_2^2, respectively,

$$E(S^2) = \frac{1}{n_1 + n_2 - 2}[(n_1 - 1)E(S_1^2) + (n_2 - 1)E(S_2^2)] = \frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}.$$

If we assume $\sigma_1^2 = \sigma_2^2 = \sigma^2$, the right hand side of the above is σ^2, which means that S^2 is unbiased for σ^2.

9.110 $n = 15, \bar{x} = 3.2$, and $s = 0.6$.

(a) $t_{0.01} = 2.624$ with 14 degrees of freedom. So, a 99% left-sided confidence interval has an upper bound of $3.2 + (2.624)\frac{0.6}{\sqrt{15}} = 3.607$ seconds. We assumed normality in the calculation.

(b) $3.2 + (2.624)(0.6)\sqrt{1 + \frac{1}{15}} = 4.826$. Still, we need to assume normality in the distribution.

(c) $1 - \alpha = 0.99$ and $1 - \gamma = 0.95$. So, $k = 3.520$ with $n = 15$. So, the upper bound is $3.2 + (3.520)(0.6) = 5.312$. Hence, we are 99% confident to claim that 95% of the pilot will have reaction time less than 5.312 seconds.

9.111 $n = 400, x = 17$, so $\hat{p} = \frac{17}{400} = 0.0425$.

(a) $z_{0.025} = 1.96$. So,

$$0.0425 \pm (1.96)\sqrt{\frac{(0.0425)(0.9575)}{400}} = 0.0425 \pm 0.0198,$$

which yields $0.0227 < p < 0.0623$.

(b) $z_{0.05} = 1.645$. So, the upper bound of a left-sided 95% confidence interval is $0.0425 + (1.645)\sqrt{\frac{(0.0425)(0.9575)}{400}} = 0.0591$.

(c) Using both intervals, we do not have evidence to dispute suppliers' claim.

Chapter 10

One- and Two-Sample Tests of Hypotheses

10.1 (a) Conclude that fewer than 30% of the public are allergic to some cheese products when, in fact, 30% or more are allergic.

(b) Conclude that at least 30% of the public are allergic to some cheese products when, in fact, fewer than 30% are allergic.

10.2 (a) The training course is effective.

(b) The training course is effective.

10.3 (a) The firm is not guilty.

(b) The firm is guilty.

10.4 (a) $\alpha = P(X \leq 5 \mid p = 0.6) + P(X \geq 13 \mid p = 0.6) = 0.0338 + (1 - 0.9729) = 0.0609$.

(b) $\beta = P(6 \leq X \leq 12 \mid p = 0.5) = 0.9963 - 0.1509 = 0.8454$.
$\beta = P(6 \leq X \leq 12 \mid p = 0.7) = 0.8732 - 0.0037 = 0.8695$.

(c) This test procedure is not good for detecting differences of 0.1 in p.

10.5 (a) $\alpha = P(X < 110 \mid p = 0.6) + P(X > 130 \mid p = 0.6) = P(Z < -1.52) + P(Z > 1.52) = 2(0.0643) = 0.1286$.

(b) $\beta = P(110 < X < 130 \mid p = 0.5) = P(1.34 < Z < 4.31) = 0.0901$.
$\beta = P(110 < X < 130 \mid p = 0.7) = P(-4.71 < Z < -1.47) = 0.0708$.

(c) The probability of a Type I error is somewhat high for this procedure, although Type II errors are reduced dramatically.

10.6 (a) $\alpha = P(X \leq 3 \mid p = 0.6) = 0.0548$.

(b) $\beta = P(X > 3 \mid p = 0.3) = 1 - 0.6496 = 0.3504$.
$\beta = P(X > 3 \mid p = 0.4) = 1 - 0.3823 = 0.6177$.
$\beta = P(X > 3 \mid p = 0.5) = 1 - 0.1719 = 0.8281$.

10.7 (a) $\alpha = P(X \leq 24 \mid p = 0.6) = P(Z < -1.59) = 0.0559$.

(b) $\beta = P(X > 24 \mid p = 0.3) = P(Z > 2.93) = 1 - 0.9983 = 0.0017$.
$\beta = P(X > 24 \mid p = 0.4) = P(Z > 1.30) = 1 - 0.9032 = 0.0968$.
$\beta = P(X > 24 \mid p = 0.5) = P(Z > -0.14) = 1 - 0.4443 = 0.5557$.

10.8 (a) $n = 12$, $p = 0.7$, and $\alpha = P(X > 11) = 0.0712 + 0.0138 = 0.0850$.

(b) $n = 12$, $p = 0.9$, and $\beta = P(X \leq 10) = 0.3410$.

10.9 (a) $n = 100$, $p = 0.7$, $\mu = np = 70$, and $\sigma = \sqrt{npq} = \sqrt{(100)(0.7)(0.3)} = 4.583$. Hence $z = \frac{82.5-70}{4.583} = 0.3410$. Therefore,

$$\alpha = P(X > 82) = P(Z > 2.73) = 1 - 0.9968 = 0.0032.$$

(b) $n = 100$, $p = 0.9$, $\mu = np = 90$, and $\sigma = \sqrt{npq} = \sqrt{(100)(0.9)(0.1)} = 3$. Hence $z = \frac{82.5-90}{3} = -2.5$. So,

$$\beta = P(X \leq 82) = P(X < -2.5) = 0.0062.$$

10.10 (a) $n = 7$, $p = 0.4$, $\alpha = P(X \leq 2) = 0.4199$.

(b) $n = 7$, $p = 0.3$, $\beta = P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0.6471 = 0.3529$.

10.11 (a) $n = 70$, $p = 0.4$, $\mu = np = 28$, and $\sigma = \sqrt{npq} = 4.099$, with $z = \frac{23.5-28}{4.099} = -1.10$. Then $\alpha = P(X < 24) = P(Z < -1.10) = 0.1357$.

(b) $n = 70$, $p = 0.3$, $\mu = np = 21$, and $\sigma = \sqrt{npq} = 3.834$, with $z = \frac{23.5-21}{3.834} = 0.65$ Then $\beta = P(X \geq 24) = P(Z > 0.65) = 0.2578$.

10.12 (a) $n = 400$, $p = 0.6$, $\mu = np = 240$, and $\sigma = \sqrt{npq} = 9.798$, with

$$z_1 = \frac{259.5 - 240}{9.978} = 1.990, \quad \text{and} \quad z_2 = \frac{220.5 - 240}{9.978} = -1.990.$$

Hence,

$$\alpha = 2P(Z < -1.990) = (2)(0.0233) = 0.0466.$$

(b) When $p = 0.48$, then $\mu = 192$ and $\sigma = 9.992$, with

$$z_1 = \frac{220.5 - 192}{9.992} = 2.852, \quad \text{and} \quad z_2 = \frac{259.5 - 192}{9.992} = 6.755.$$

Therefore,

$$\beta = P(2.852 < Z < 6.755) = 1 - 0.9978 = 0.0022.$$

10.13 From Exercise 10.12(a) we have $\mu = 240$ and $\sigma = 9.798$. We then obtain

$$z_1 = \frac{214.5 - 240}{9.978} = -2.60, \quad \text{and} \quad z_2 = \frac{265.5 - 240}{9.978} = 2.60.$$

So

$$\alpha = 2P(Z < -2.60) = (2)(0.0047) = 0.0094.$$

Also, from Exercise 10.12(b) we have $\mu = 192$ and $\sigma = 9.992$, with

$$z_1 = \frac{214.5 - 192}{9.992} = 2.25, \quad \text{and} \quad z_2 = \frac{265.5 - 192}{9.992} = 7.36.$$

Therefore,

$$\beta = P(2.25 < Z < 7.36) = 1 - 0.9878 = 0.0122.$$

10.14 (a) $n = 50$, $\mu = 15$, $\sigma = 0.5$, and $\sigma_{\bar{X}} = \frac{0.5}{\sqrt{50}} = 0.071$, with $z = \frac{14.9-15}{0.071} = -1.41$. Hence, $\alpha = P(Z < -1.41) = 0.0793$.

(b) If $\mu = 14.8$, $z = \frac{14.9-14.8}{0.071} = 1.41$. So, $\beta = P(Z > 1.41) = 0.0793$.
If $\mu = 14.9$, then $z = 0$ and $\beta = P(Z > 0) = 0.5$.

10.15 (a) $\mu = 200$, $n = 9$, $\sigma = 15$ and $\sigma_{\bar{X}} = \frac{15}{3} = 5$. So,

$$z_1 = \frac{191 - 200}{5} = -1.8, \quad \text{and} \quad z_2 = \frac{209 - 200}{5} = 1.8,$$

with $\alpha = 2P(Z < -1.8) = (2)(0.0359) = 0.0718$.

(b) If $\mu = 215$, then $z - 1 = \frac{191-215}{5} = -4.8$ and $z_2 = \frac{209-215}{5} = -1.2$, with

$$\beta = P(-4.8 < Z < -1.2) = 0.1151 - 0 = 0.1151.$$

10.16 (a) When $n = 15$, then $\sigma_{\bar{X}} = \frac{15}{5} = 3$, with $\mu = 200$ and $n = 25$. Hence

$$z_1 = \frac{191 - 200}{3} = -3, \quad \text{and} \quad z_2 = \frac{209 - 200}{3} = 3,$$

with $\alpha = 2P(Z < -3) = (2)(0.0013) = 0.0026$.

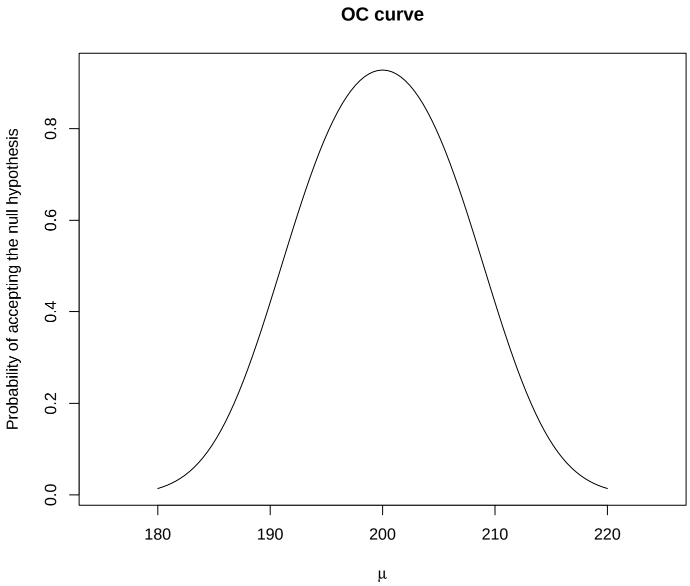
(b) When $\mu = 215$, then $z - 1 = \frac{191-215}{3} = -8$ and $z_2 = \frac{209-215}{3} = -2$, with

$$\beta = P(-8 < Z < -2) = 0.0228 - 0 = 0.0228.$$

10.17 (a) $n = 50$, $\mu = 5000$, $\sigma = 120$, and $\sigma_{\bar{X}} = \frac{120}{\sqrt{50}} = 16.971$, with $z = \frac{4970-5000}{16.971} = -1.77$ and $\alpha = P(Z < -1.77) = 0.0384$.

(b) If $\mu = 4970$, then $z = 0$ and hence $\beta = P(Z > 0) = 0.5$.
If $\mu = 4960$, then $z = \frac{4970-4960}{16.971} = 0.59$ and $\beta = P(Z > 0.59) = 0.2776$.

10.18 The OC curve is shown next.



10.19 The hypotheses are

$$H_0 : \mu = 800,$$
$$H_1 : \mu \neq 800.$$

Now, $z = \frac{788-800}{40/\sqrt{30}} = -1.64$, and P-value= $2P(Z < -1.64) = (2)(0.0505) = 0.1010$. Hence, the mean is not significantly different from 800 for $\alpha < 0.101$.

10.20 The hypotheses are

$$H_0 : \mu = 5.5,$$
$$H_1 : \mu < 5.5.$$

Now, $z = \frac{5.23-5.5}{0.24/\sqrt{64}} = -9.0$, and P-value= $P(Z < -9.0) \approx 0$. The White Cheddar Popcorn, on average, weighs less than 5.5oz.

10.21 The hypotheses are

$$H_0 : \mu = 40 \text{ months},$$
$$H_1 : \mu < 40 \text{ months}.$$

Now, $z = \frac{38-40}{5.8/\sqrt{64}} = -2.76$, and P-value= $P(Z < -2.76) = 0.0029$. Decision: reject H_0.

10.22 The hypotheses are

$$H_0 : \mu = 162.5 \text{ centimeters},$$
$$H_1 : \mu \neq 162.5 \text{ centimeters}.$$

Now, $z = \frac{165.2-162.5}{6.9/\sqrt{50}} = 2.77$, and P-value= $2P(Z > 2.77) = (2)(0.0028) = 0.0056$. Decision: reject H_0 and conclude that $\mu \neq 162.5$.

10.23 The hypotheses are

$$H_0 : \mu = 20,000 \text{ kilometers},$$
$$H_1 : \mu > 20,000 \text{ kilometers}.$$

Now, $z = \frac{23{,}500-20{,}000}{3900/\sqrt{100}} = 8.97$, and P-value$= P(Z > 8.97) \approx 0$. Decision: reject H_0 and conclude that $\mu \neq 20,000$ kilometers.

10.24 The hypotheses are

$$H_0 : \mu = 8,$$
$$H_1 : \mu > 8.$$

Now, $z = \frac{8.5-8}{2.25/\sqrt{225}} = 3.33$, and P-value$= P(Z > 3.33) = 0.0004$. Decision: Reject H_0 and conclude that men who use TM, on average, mediate more than 8 hours per week.

10.25 The hypotheses are

$$H_0 : \mu = 10,$$
$$H_1 : \mu \neq 10.$$

$\alpha = 0.01$ and $df = 9$.
Critical region: $t < -3.25$ or $t > 3.25$.
Computation: $t = \frac{10.06-10}{0.246/\sqrt{10}} = 0.77$.
Decision: Fail to reject H_0.

10.26 The hypotheses are

$$H_0 : \mu = 220 \text{ milligrams},$$
$$H_1 : \mu > 220 \text{ milligrams}.$$

$\alpha = 0.01$ and $df = 9$.
Critical region: $t > 1.729$.
Computation: $t = \frac{224-220}{24.5/\sqrt{20}} = 4.38$.
Decision: Reject H_0 and claim $\mu > 220$ milligrams.

10.27 The hypotheses are

$$H_0 : \mu_1 = \mu_2,$$
$$H_1 : \mu_1 > \mu_2.$$

Since $s_p = \sqrt{\frac{(29)(10.5)^2+(29)(10.2)^2}{58}} = 10.35$, then

$$P\left[T > \frac{34.0}{10.35\sqrt{1/30 + 1/30}}\right] = P(Z > 12.72) \approx 0.$$

Hence, the conclusion is that running increases the mean RMR in older women.

10.28 The hypotheses are

$$H_0: \mu_C = \mu_A,$$
$$H_1: \mu_C > \mu_A,$$

with $s_p = \sqrt{\frac{(24)(1.5)^2+(24)(1.25)^2}{48}} = 1.3807$. We obtain $t = \frac{20.0-12.0}{1.3807\sqrt{2/25}} = 20.48$. Since $P(T > 20.48) \approx 0$, we conclude that the mean percent absorbency for the cotton fiber is significantly higher than the mean percent absorbency for acetate.

10.29 The hypotheses are

$$H_0: \mu = 35 \text{ minutes},$$
$$H_1: \mu < 35 \text{ minutes}.$$

$\alpha = 0.05$ and $df = 19$.
Critical region: $t < -1.729$.
Computation: $t = \frac{33.1-35}{4.3/\sqrt{20}} = -1.98$.
Decision: Reject H_0 and conclude that it takes less than 35 minutes, on the average, to take the test.

10.30 The hypotheses are

$$H_0: \mu_1 = \mu_2,$$
$$H_1: \mu_1 \neq \mu_2.$$

Since the variances are known, we obtain $z = \frac{81-76}{\sqrt{5.2^2/25+3.5^2/36}} = 4.22$. So, P-value≈ 0 and we conclude that $\mu_1 > \mu_2$.

10.31 The hypotheses are

$$H_0: \mu_A - \mu_B = 12 \text{ kilograms},$$
$$H_1: \mu_A - \mu_B > 12 \text{ kilograms}.$$

$\alpha = 0.05$.
Critical region: $z > 1.645$.
Computation: $z = \frac{(86.7-77.8)-12}{\sqrt{(6.28)^2/50+(5.61)^2/50}} = -2.60$. So, fail to reject H_0 and conclude that the average tensile strength of thread A does not exceed the average tensile strength of thread B by 12 kilograms.

10.32 The hypotheses are

$$H_0: \mu_1 - \mu_2 = \$2,000,$$
$$H_1: \mu_1 - \mu_2 > \$2,000.$$

$\alpha = 0.01$.
Critical region: $z > 2.33$.
Computation: $z = \frac{(70750-65200)-2000}{\sqrt{(6000)^2/200+(5000)^2/200}} = 6.43$, with a P-value$= P(Z > 6.43) \approx 0$. Reject H_0 and conclude that the mean salary for associate professors in research institutions is \$2000 higher than for those in other institutions.

10.33 The hypotheses are

$$H_0: \mu_1 - \mu_2 = 0.5 \text{ micromoles per 30 minutes,}$$
$$H_1: \mu_1 - \mu_2 > 0.5 \text{ micromoles per 30 minutes.}$$

$\alpha = 0.01$.
Critical region: $t > 2.485$ with 25 degrees of freedom.
Computation: $s_p^2 = \frac{(14)(1.5)^2+(11)(1.2)^2}{25} = 1.8936$, and $t = \frac{(8.8-7.5)-0.5}{\sqrt{1.8936}\sqrt{1/15+1/12}} = 1.50$. Do not reject H_0.

10.34 The hypotheses are

$$H_0: \mu_1 - \mu_2 = 8,$$
$$H_1: \mu_1 - \mu_2 < 8.$$

Computation: $s_p^2 = \frac{(10)(4.7)^2+(16)(6.1)^2}{26} = 31.395$, and $t = \frac{(85-79)-8}{\sqrt{31.395}\sqrt{1/11+1/17}} = -0.92$. Using 28 degrees of freedom and Table A.4, we obtain that $0.15 < P$-value < 0.20. Decision: Do not reject H_0.

10.35 The hypotheses are

$$H_0: \mu_1 - \mu_2 = 0,$$
$$H_1: \mu_1 - \mu_2 < 0.$$

$\alpha = 0.05$
Critical region: $t < -1.895$ with 7 degrees of freedom.
Computation: $s_p = \sqrt{\frac{(3)(1.363)+(4)(3.883)}{7}} = 1.674$, and $t = \frac{2.075-2.860}{1.674\sqrt{1/4+1/5}} = -0.70$.
Decision: Do not reject H_0.

10.36 The hypotheses are

$$H_0: \mu_1 = \mu_2,$$
$$H_1: \mu_1 \neq \mu_2.$$

Computation: $s_p = \sqrt{\frac{5100^2+5900^2}{2}} = 5515$, and $t = \frac{37{,}900-39{,}800}{5515\sqrt{1/12+1/12}} = -0.84$.
Using 22 degrees of freedom and since $0.20 < P(T < -0.84) < 0.3$, we obtain $0.4 < P$-value < 0.6. Decision: Do not reject H_0.

10.37 The hypotheses are

$$H_0: \mu_1 - \mu_2 = 4 \text{ kilometers},$$
$$H_1: \mu_1 - \mu_2 \neq 4 \text{ kilometers}.$$

$\alpha = 0.10$ and the critical regions are $t < -1.725$ or $t > 1.725$ with 20 degrees of freedom.
Computation: $t = \frac{5-4}{(0.915)\sqrt{1/12+1/10}} = 2.55$.
Decision: Reject H_0.

10.38 The hypotheses are

$$H_0: \mu_1 - \mu_2 = 8,$$
$$H_1: \mu_1 - \mu_2 < 8.$$

$\alpha = 0.05$ and the critical region is $t < -1.714$ with 23 degrees of freedom.
Computation: $s_p = \sqrt{\frac{(9)(3.2)^2+(14)(2.8)^2}{23}} = 2.963$, and $t = \frac{5.5-8}{2.963\sqrt{1/10+1/15}} = -2.07$.
Decision: Reject H_0 and conclude that $\mu_1 - \mu_2 < 8$ months.

10.39 The hypotheses are

$$H_0: \mu_{II} - \mu_I = 10,$$
$$H_1: \mu_{II} - \mu_I > 10.$$

$\alpha = 0.1$.
Degrees of freedom is calculated as

$$v = \frac{(78.8/5 + 913.333/7)^2}{(78.8/5)^2/4 + (913/333/7)^2/6} = 7.38,$$

hence we use 7 degrees of freedom with the critical region $t > 2.998$.
Computation: $t = \frac{(110-97.4)-10}{\sqrt{78.800/5+913.333/7}} = 0.22$.
Decision: Fail to reject H_0.

10.40 The hypotheses are

$$H_0: \mu_S = \mu_N,$$
$$H_1: \mu_S \neq \mu_N.$$

Degrees of freedom is calculated as

$$v = \frac{(0.391478^2/8 + 0.214414^2/24)^2}{(0.391478^2/8)^2/7 + (0.214414^2/24)^2/23} = 8.$$

Computation: $t = \frac{0.97625-0.91583}{\sqrt{0.391478^2/8+0.214414^2/24}} = -0.42$. Since $0.3 < P(T < -0.42) < 0.4$, we obtain $0.6 < P\text{-value} < 0.8$.
Decision: Fail to reject H_0.

10.41 The hypotheses are

$$H_0 : \mu_1 = \mu_2,$$
$$H_1 : \mu_1 \neq \mu_2.$$

$\alpha = 0.05$.
Degrees of freedom is calculated as

$$v = \frac{(7874.329^2/16 + 2479/503^2/12)^2}{(7874.329^2/16)^2/15 + (2479.503^2/12)^2/11} = 19 \text{ degrees of freedom.}$$

Critical regions $t < -2.093$ or $t > 2.093$.
Computation: $t = \frac{9897.500-4120.833}{\sqrt{7874.329^2/16+2479.503^2/12}} = 2.76$.
Decision: Reject H_0 and conclude that $\mu_1 > \mu_2$.

10.42 The hypotheses are

$$H_0 : \mu_1 = \mu_2,$$
$$H_1 : \mu_1 \neq \mu_2.$$

$\alpha = 0.05$.
Critical regions $t < -2.776$ or $t > 2.776$, with 4 degrees of freedom.
Computation: $\bar{d} = -0.1$, $s_d = 0.1414$, so $t = \frac{-0.1}{0.1414/\sqrt{5}} = -1.58$.
Decision: Do not reject H_0 and conclude that the two methods are not significantly different.

10.43 The hypotheses are

$$H_0 : \mu_1 = \mu_2,$$
$$H_1 : \mu_1 > \mu_2.$$

Computation: $\bar{d} = 0.1417$, $s_d = 0.198$, $t = \frac{0.1417}{0.198/\sqrt{12}} = 2.48$ and $0.015 < P\text{-value} < 0.02$ with 11 degrees of freedom.
Decision: Reject H_0 when a significance level is above 0.02.

10.44 The hypotheses are

$$H_0 : \mu_1 - \mu_2 = 4.5 \text{ kilograms},$$
$$H_1 : \mu_1 - \mu_2 < 4.5 \text{ kilograms}.$$

Computation: $\bar{d} = 3.557$, $s_d = 2.776$, $t = \frac{3.557-4.5}{2.778/\sqrt{7}} = -0.896$, and $0.2 < P\text{-value} < 0.3$ with 6 degrees of freedom.
Decision: Do not reject H_0.

10.45 The hypotheses are

$$H_0: \mu_1 = \mu_2,$$
$$H_1: \mu_1 < \mu_2.$$

Computation: $\bar{d} = -54.13$, $s_d = 83.002$, $t = \frac{-54.13}{83.002/\sqrt{15}} = -2.53$, and $0.01 < P$-value < 0.015 with 14 degrees of freedom.
Decision: Reject H_0.

10.46 The hypotheses are

$$H_0: \mu_1 = \mu_2,$$
$$H_1: \mu_1 \neq \mu_2.$$

$\alpha = 0.05$.
Critical regions are $t < -2.365$ or $t > 2.365$ with 7 degrees of freedom.
Computation: $\bar{d} = 198.625$, $s_d = 210.165$, $t = \frac{198.625}{210.165/\sqrt{8}} = 2.67$.
Decision: Reject H_0; length of storage influences sorbic acid residual concentrations.

10.47 $n = \frac{(1.645+1.282)^2(0.24)^2}{0.3^2} = 5.48$. The sample size needed is 6.

10.48 $\beta = 0.1$, $\sigma = 5.8$, $\delta = 35.9 - 40 = -4.1$. Assume $\alpha = 0.05$ then $z_{0.05} = 1.645$, $z_{0.10} = 1.28$. Therefore,

$$n = \frac{(1.645 + 1.28)^2(5.8)^2}{(-4.1)^2} = 17.12 \approx 18 \text{ due to round up.}$$

10.49 $1 - \beta = 0.95$ so $\beta = 0.05$, $\delta = 3.1$ and $z_{0.01} = 2.33$. Therefore,

$$n = \frac{(1.645 + 2.33)^2(6.9)^2}{3.1^2} = 78.28 \approx 79 \text{ due to round up.}$$

10.50 $\beta = 0.05$, $\delta = 8$, $\alpha = 0.05$, $z_{0.05} = 1.645$, $\sigma_1 = 6.28$ and $\sigma_2 = 5.61$. Therefore,

$$n = \frac{(1.645 + 1.645)^2(6.28^2 + 5.61^2)}{8^2} = 11.99 \approx 12 \text{ due to round up.}$$

10.51 $n = \frac{1.645+0.842)^2(2.25)^2}{[(1.2)(2.25)]^2} = 4.29$. The sample size would be 5.

10.52 $\sigma = 1.25$, $\alpha = 0.05$, $\beta = 0.1$, $\delta = 0.5$, so $\Delta = \frac{0.5}{1.25} = 0.4$. Using Table A.8 we find $n = 68$.

10.53 (a) The hypotheses are

$$H_0: M_{\text{hot}} - M_{\text{cold}} = 0,$$
$$H_1: M_{\text{hot}} - M_{\text{cold}} \neq 0.$$

(b) Use paired T-test and find out $t = 0.99$ with $0.3 < P\text{-value} < 0.4$. Hence, fail to reject H_0.

10.54 Using paired T-test, we find out $t = 2.4$ with 8 degrees of freedom. So, $0.02 < P\text{-value} < 0.025$. Reject H_0; breathing frequency significantly higher in the presence of CO.

10.55 The hypotheses are

$$H_0 : p = 0.40,$$
$$H_1 : p > 0.40.$$

Denote by X for those who choose lasagna.

$$P\text{-value} = P(X \geq 9 \mid p = 0.40) = 0.4044.$$

The claim that $p = 0.40$ is not refuted.

10.56 The hypotheses are

$$H_0 : p = 0.40,$$
$$H_1 : p > 0.40.$$

$\alpha = 0.05$.
Test statistic: binomial variable X with $p = 0.4$ and $n = 15$.
Computation: $x = 8$ and $np_0 = (15)(0.4) = 6$. Therefore, from Table A.1,

$$P\text{-value} = P(X \geq 8 \mid p = 0.4) = 1 - P(X \leq 7 \mid p = 0.4) = 0.2131,$$

which is larger than 0.05.
Decision: Do not reject H_0.

10.57 The hypotheses are

$$H_0 : p = 0.5,$$
$$H_1 : p < 0.5.$$

$$P\text{-value} = P(X \leq 5 \mid p = 0.05) = 0.0207.$$

Decision: Reject H_0.

10.58 The hypotheses are

$$H_0 : p = 0.6,$$
$$H_1 : p < 0.6.$$

So

$$P\text{-value} \approx P\left(Z < \frac{110 - (200)(0.6)}{\sqrt{(200)(0.6)(0.4)}}\right) = P(Z < -1.44) = 0.0749.$$

Decision: Fail to reject H_0.

10.59 The hypotheses are

$$H_0 : p = 0.2,$$
$$H_1 : p < 0.2.$$

Then

$$P\text{-value} \approx P\left(Z < \frac{136 - (1000)(0.2)}{\sqrt{(1000)(0.2)(0.8)}}\right) = P(Z < -5.06) \approx 0.$$

Decision: Reject H_0; less than 1/5 of the homes in the city are heated by oil.

10.60 The hypotheses are

$$H_0 : p = 0.25,$$
$$H_1 : p > 0.25.$$

$\alpha = 0.05$.
Computation:

$$P\text{-value} \approx P\left(Z > \frac{28 - (90)(0.25)}{\sqrt{(90)(0.25)(0.75)}}\right) = P(Z > 1,34) = 0.091.$$

Decision: Fail to reject H_0; No sufficient evidence to conclude that $p > 0.25$.

10.61 The hypotheses are

$$H_0 : p = 0.8,$$
$$H_1 : p > 0.8.$$

$\alpha = 0.04$.
Critical region: $z > 1.75$.
Computation: $z = \frac{250-(300)(0.8)}{\sqrt{(300)(0.8)(0.2)}} = 1.44$.
Decision: Fail to reject H_0; it cannot conclude that the new missile system is more accurate.

10.62 The hypotheses are

$$H_0 : p = 0.25,$$
$$H_1 : p > 0.25.$$

$\alpha = 0.05$.
Critical region: $z > 1.645$.
Computation: $z = \frac{16-(48)(0.25)}{\sqrt{(48)(0.25)(0.75)}} = 1.333$.
Decision: Fail to reject H_0. On the other hand, we can calculate

$$P\text{-value} = P(Z > 1.33) = 0.0918.$$

10.63 The hypotheses are

$$H_0 : p_1 = p_2,$$
$$H_1 : p_1 \neq p_2.$$

Computation: $\hat{p} = \frac{63+59}{100+125} = 0.5422$, $z = \frac{(63/100)-(59/125)}{\sqrt{(0.5422)(0.4578)(1/100+1/125)}} = 2.36$, with P-value $= 2P(Z > 2.36) = 0.0182$.
Decision: Reject H_0 at level 0.0182. The proportion of urban residents who favor the nuclear plant is larger than the proportion of suburban residents who favor the nuclear plant.

10.64 The hypotheses are

$$H_0 : p_1 = p_2,$$
$$H_1 : p_1 > p_2.$$

Computation: $\hat{p} = \frac{240+288}{300+400} = 0.7543$, $z = \frac{(240/300)-(288/400)}{\sqrt{(0.7543)(0.2457)(1/300+1/400)}} = 2.44$, with P-value $= P(Z > 2.44) = 0.0073$.
Decision: Reject H_0. The proportion of couples married less than 2 years and planning to have children is significantly higher than that of couples married 5 years and planning to have children.

10.65 The hypotheses are

$$H_0 : p_U = p_R,$$
$$H_1 : p_U > p_R.$$

Computation: $\hat{p} = \frac{20+10}{200+150} = 0.085714$, $z = \frac{(20/200)-(10/150)}{\sqrt{(0.085714)(0.914286)(1/200+1/150)}} = 1.10$, with P-value $= P(Z > 1.10) = 0.1357$.
Decision: Fail to reject H_0. It cannot be shown that breast cancer is more prevalent in the urban community.

10.66 The hypotheses are

$$H_0 : p_1 = p_2,$$
$$H_1 : p_1 > p_2.$$

Computation: $\hat{p} = \frac{29+56}{120+280} = 0.2125$, $z = \frac{(29/120)-(56/280))}{\sqrt{(0.2125)(0.7875)(1/120+1/280)}} = 0.93$, with P-value $= P(Z > 0.93) = 0.1762$.
Decision: Fail to reject H_0. There is no significant evidence to conclude that the new medicine is more effective.

10.67 The hypotheses are

$$H_0 : \sigma^2 = 0.03,$$
$$H_1 : \sigma^2 \neq 0.03.$$

Computation: $\chi^2 = \frac{(9)(0.24585)^2}{0.03} = 18.13$. Since $0.025 < P(\chi^2 > 18.13) < 0.05$ with 9 degrees of freedom, $0.05 < P$-value $= 2P(\chi^2 > 18.13) < 0.10$.
Decision: Fail to reject H_0; the sample of 10 containers is not sufficient to show that σ^2 is not equal to 0.03.

10.68 The hypotheses are

$$H_0 : \sigma = 6,$$
$$H_1 : \sigma < 6.$$

Computation: $\chi^2 = \frac{(19)(4.51)^2}{36} = 10.74$. Using the table, $1-0.95 < P(\chi^2 < 10.74) < 0.1$ with 19 degrees of freedom, we obtain $0.05 < P$-value < 0.1.
Decision: Fail to reject H_0; there was not sufficient evidence to conclude that the standard deviation is less then 6 at level $\alpha = 0.05$ level of significance.

10.69 The hypotheses are

$$H_0 : \sigma^2 = 4.2 \text{ ppm},$$
$$H_1 : \sigma^2 \neq 4.2 \text{ ppm}.$$

Computation: $\chi^2 = \frac{(63)(4.25)^2}{4.2} = 63.75$. Since $0.3 < P(\chi^2 > 63.75) < 0.5$ with 63 degrees of freedom, P-value $= 2P(\chi^2 > 18.13) > 0.6$ (In Microsoft Excel, if you type "=2*chidist(63.75,63)", you will get the P-value as 0.8898.
Decision: Fail to reject H_0; the variance of aflotoxins is not significantly different from 4.2 ppm.

10.70 The hypotheses are

$$H_0 : \sigma = 1.40,$$
$$H_1 : \sigma > 1.40.$$

Computation: $\chi^2 = \frac{(11)(1.75)^2}{1.4} = 17.19$. Using the table, $0.1 < P(\chi^2 > 17.19) < 0.2$ with 11 degrees of freedom, we obtain $0.1 < P$-value < 0.2.
Decision: Fail to reject H_0; the standard deviation of the contributions from the sanitation department is not significantly greater than \$1.40 at the $\alpha = 0.01$ level of significance.

10.71 The hypotheses are

$$H_0 : \sigma^2 = 1.15,$$
$$H_1 : \sigma^2 > 1.15.$$

Computation: $\chi^2 = \frac{(24)(2.03)^2}{1.15} = 42.37$. Since $0.01 < P(\chi^2 > 42.37) < 0.02$ with 24 degrees of freedom, $0.01 < P$-value < 0.02.
Decision: Reject H_0; there is sufficient evidence to conclude, at level $\alpha = 0.05$, that the soft drink machine is out of control.

10.72 (a) The hypotheses are

$$H_0: \sigma = 10.0,$$
$$H_1: \sigma \neq 10.0.$$

Computation: $z = \frac{11.9-10.0}{10.0/\sqrt{200}} = 2.69$. So P-value $= P(Z < -2.69) + P(Z > 2.69) = 0.0072$. There is sufficient evidence to conclude that the standard deviation is different from 10.0.

(b) The hypotheses are

$$H_0: \sigma^2 = 6.25,$$
$$H_1: \sigma^2 < 6.25.$$

Computation: $z = \frac{2.1-2.5}{2.5/\sqrt{144}} = -1.92$. P-value $= P(Z < -1.92) = 0.0274$.
Decision: Reject H_0; the variance of the distance achieved by the diesel model is less than the variance of the distance achieved by the gasoline model.

10.73 The hypotheses are

$$H_0: \sigma_1^2 = \sigma_2^2,$$
$$H_1: \sigma_1^2 > \sigma_2^2.$$

Computation: $f = \frac{(6.1)^2}{(5.3)^2} = 1.33$. Since $f_{0.05}(10, 13) = 2.67 > 1.33$, we fail to reject H_0 at level $\alpha = 0.05$. So, the variability of the time to assemble the product is not significantly greater for men. On the other hand, if you use "=fdist(1.33,10,13)", you will obtain the P-value $= 0.3095$.

10.74 The hypotheses are

$$H_0: \sigma_1^2 = \sigma_2^2,$$
$$H_1: \sigma_1^2 \neq \sigma_2^2.$$

Computation: $f = \frac{(7874.329)^2}{(2479.503)^2} = 10.09$. Since $f_{0.01}(15, 11) = 4.25$, the P-value $> (2)(0.01) = 0.02$. Hence we reject H_0 at level $\alpha = 0.02$ and claim that the variances for the two locations are significantly different. The P-value $= 0.0004$.

10.75 The hypotheses are

$$H_0: \sigma_1^2 = \sigma_2^2,$$
$$H_1: \sigma_1^2 \neq \sigma_2^2.$$

Computation: $f = \frac{78.800}{913.333} = 0.086$. Since P-value $= 2P(f < 0.086) = (2)(0.0164) = 0.0328$ for 4 and 6 degrees of freedom, the variability of running time for company 1 is significantly less than, at level 0.0328, the variability of running time for company 2.

10.76 The hypotheses are

$$H_0 : \sigma_A = \sigma_B,$$
$$H_1 : \sigma_A \neq \sigma_B.$$

Computation: $f = \frac{(0.0125)}{0.0108} = 1.15$. Since P-value $= 2P(f > 1.15) = (2)(0.424) = 0.848$ for 8 and 8 degrees of freedom, the two instruments appear to have similar variability.

10.77 The hypotheses are

$$H_0 : \sigma_1 = \sigma_2,$$
$$H_1 : \sigma_1 \neq \sigma_2.$$

Computation: $f = \frac{(0.0553)^2}{(0.0125)^2} = 19.67$. Since P-value $= 2P(f > 19.67) = (2)(0.0004) = 0.0008$ for 7 and 7 degrees of freedom, production line 1 is not producing as consistently as production 2.

10.78 The hypotheses are

$$H_0 : \sigma_1 = \sigma_2,$$
$$H_1 : \sigma_1 \neq \sigma_2.$$

Computation: $s_1 = 291.0667$ and $s_2 = 119.3946$, $f = \frac{(291.0667)^2}{(119.3946)^2} = 5.54$. Since P-value $= 2P(f > 5.54) = (2)(0.0002) = 0.0004$ for 19 and 19 degrees of freedom, hydrocarbon emissions are more consistent in the 1990 model cars.

10.79 The hypotheses are

$$H_0 : \text{die is balanced},$$
$$H_1 : \text{die is unbalanced}.$$

$\alpha = 0.01$.
Critical region: $\chi^2 > 15.086$ with 5 degrees of freedom.
Computation: Since $e_i = 30$, for $i = 1, 2, \ldots, 6$, then

$$\chi^2 = \frac{(28-30)^2}{30} + \frac{(36-30)^2}{30} + \cdots + \frac{(23-30)^2}{30} = 4.47.$$

Decision: Fail to reject H_0; the die is balanced.

10.80 The hypotheses are

$$H_0 : \text{coin is balanced},$$
$$H_1 : \text{coin is not balanced}.$$

$\alpha = 0.05$.
Critical region: $\chi^2 > 3.841$ with 1 degrees of freedom.
Computation: Since $e_i = 30$, for $i = 1, 2, \ldots, 6$, then

$$\chi^2 = \frac{(63-50)^2}{50} + \frac{(37-50)^2}{50} = 6.76.$$

Decision: Reject H_0; the coin is not balanced.

10.81 The hypotheses are

$$H_0 : \text{nuts are mixed in the ratio 5:2:2:1},$$
$$H_1 : \text{nuts are not mixed in the ratio 5:2:2:1}.$$

$\alpha = 0.05$.
Critical region: $\chi^2 > 7.815$ with 3 degrees of freedom.
Computation:

Observed	269	112	74	45
Expected	250	100	100	50

$$\chi^2 = \frac{(269-250)^2}{250} + \frac{(112-100)^2}{100} + \frac{(74-100)^2}{100} + \frac{(45-50)^2}{50} = 10.14.$$

Decision: Reject H_0; the nuts are not mixed in the ratio 5:2:2:1.

10.82 The hypotheses are

$$H_0 : \text{Distribution of grades is uniform},$$
$$H_1 : \text{Distribution of grades is not uniform}.$$

$\alpha = 0.05$.
Critical region: $\chi^2 > 9.488$ with 4 degrees of freedom.
Computation: Since $e_i = 20$, for $i = 1, 2, \ldots, 5$, then

$$\chi^2 = \frac{(14-20)^2}{20} + \frac{(18-20)^2}{20} + \cdots \frac{(16-20)^2}{20} = 10.0.$$

Decision: Reject H_0; the distribution of grades is not uniform.

10.83 The hypotheses are

$$H_0: \text{Data follows the binomial distribution } b(y;3,1/4),$$

$$H_1: \text{Data does not follows the binomial distribution.}$$

$\alpha = 0.01$.
Computation: $b(0;3,1/4) = 27/64$, $b(1;3,1/4) = 27/64$, $b(2;3,1/4) = 9/64$, and $b(3;3,1/4) = 1/64$. Hence $e_1 = 27$, $e_2 = 27$, $e_3 = 9$ and $e_4 = 1$. Combining the last two classes together, we obtain

$$\chi^2 = \frac{(21-27)^2}{27} + \frac{(31-27)^2}{27} + \frac{(12-10)^2}{10} = 2.33.$$

Critical region: $\chi^2 > 9.210$ with 2 degrees of freedom.
Decision: Fail to reject H_0; the data is from a distribution not significantly different from $b(y;3,1/4)$.

10.84 The hypotheses are

$$H_0: \text{Data follows the hypergeometric distribution } h(x;8,3,5),$$

$$H_1: \text{Data does not follows the hypergeometric distribution.}$$

$\alpha = 0.05$.
Computation: $h(0;8,3,5) = 1/56$, $b(1;8,3,5) = 15/56$, $b(2;8,3,5) = 30/56$, and $b(3;8,3,5) = 10/56$. Hence $e_1 = 2$, $e_2 = 30$, $e_3 = 60$ and $e_4 = 20$. Combining the first two classes together, we obtain

$$\chi^2 = \frac{(32-32)^2}{32} + \frac{(55-60)^2}{60} + \frac{(25-20)^2}{20} = 1.67.$$

Critical region: $\chi^2 > 5.991$ with 2 degrees of freedom.
Decision: Fail to reject H_0; the data is from a distribution not significantly different from $h(y;8,3,5)$.

10.85 The hypotheses are

$$H_0: f(x) = g(x;1/2) \text{ for } x = 1,2,\dots,$$
$$H_1: f(x) \neq g(x;1/2).$$

$\alpha = 0.05$.
Computation: $g(x;1/2) = \frac{1}{2^x}$, for $x = 1,2,\dots,7$ and $P(X \geq 8) = \frac{1}{2^7}$. Hence $e_1 = 128$, $e_2 = 64$, $e_3 = 32$, $e_4 = 16$, $e_5 = 8$, $e_6 = 4$, $e_7 = 2$ and $e_8 = 2$. Combining the last three classes together, we obtain

$$\chi^2 = \frac{(136-128)^2}{128} + \frac{(60-64)^2}{64} + \frac{(34-32)^2}{32} + \frac{(12-16)^2}{16} + \frac{(9-8)^2}{8} + \frac{(5-8)^2}{8} = 3.125$$

Critical region: $\chi^2 > 11.070$ with 5 degrees of freedom.
Decision: Fail to reject H_0; $f(x) = g(x;1/2)$, for $x = 1,2,\dots$

10.88 The hypotheses are

$$H_0: \text{Distribution of grades is normal } n(x; 65, 21),$$

$$H_1: \text{Distribution of grades is not normal.}$$

$\alpha = 0.05$.
Computation:

z values	$P(Z < z)$	$P(z_{i-1} < Z < z_i)$	e_i		o_i	
$z_1 = \frac{19.5-65}{21} = -2.17$	0.0150	0.0150	0.9		3	
$z_2 = \frac{29.5-65}{21} = -1.69$	0.0454	0.0305	1.8	6.8	2	8
$z_3 = \frac{39.5-65}{21} = -1.21$	0.1131	0.0676	4.1		3	
$z_4 = \frac{49.5-65}{21} = -0.74$	0.2296	0.1165	7.0		4	
$z_5 = \frac{59.5-65}{21} = -0.26$	0.3974	0.1678	10.1		5	
$z_6 = \frac{69.5-65}{21} = 0.21$	0.5832	0.1858	11.1		11	
$z_7 = \frac{79.5-65}{21} = 0.69$	0.7549	0.1717	10.3		14	
$z_8 = \frac{89.5-65}{21} = 1.17$	0.8790	0.1241	7.4		14	
$z_9 = \infty$	1.0000	0.1210	7.3		4	

A goodness-of-fit test with 6 degrees of freedom is based on the following data:

o_i	8	4	5	11	14	14	4
e_i	6.8	7.0	10.1	11.1	10.3	7.4	7.3

Critical region: $\chi^2 > 12.592$.

$$\chi^2 = \frac{(8-6.8)^2}{6.8} + \frac{(4-7.0)^2}{7.0} + \cdots + \frac{(4-7.3)^2}{7.3} = 12.78.$$

Decision: Reject H_0; distribution of grades is not normal.

10.89 From the data we have

z values	$P(Z < z)$	$P(z_{i-1} < Z < z_i)$	e_i		o_i	
$z_1 = \frac{0.795-1.8}{0.4} = -2.51$	0.0060	0.0060	0.2		1	
$z_2 = \frac{0.995-1.8}{0.4} = -2.01$	0.0222	0.0162	0.6	6.1	1	5
$z_3 = \frac{1.195-1.8}{0.4} = -1.51$	0.0655	0.0433	1.7		1	
$z_4 = \frac{1.395-1.8}{0.4} = -1.01$	0.1562	0.0907	3.6		2	
$z_5 = \frac{1.595-1.8}{0.4} = -0.51$	0.3050	0.1488	6.0		4	
$z_6 = \frac{1.795-1.8}{0.4} = -0.01$	0.4960	0.1910	7.6		13	
$z_7 = \frac{1.995-1.8}{0.4} = 0.49$	0.6879	0.1919	7.7		8	
$z_8 = \frac{2.195-1.8}{0.4} = 0.99$	0.8389	0.1510	6.0		5	
$z_9 = \frac{2.395-1.8}{0.4} = 1.49$	0.9319	0.0930	3.7	6.4	3	5
$z_{10} = \infty$	1.0000	0.0681	2.7		2	

The hypotheses are

$$H_0: \text{Distribution of nicotine contents is normal } n(x; 1.8, 0.4),$$

$$H_1: \text{Distribution of nicotine contents is not normal.}$$

$\alpha = 0.01$.
Computation: A goodness-of-fit test with 5 degrees of freedom is based on the following data:

o_i	5	4	13	8	5	5
e_i	6.1	6.0	7.6	7.7	6.0	6.4

Critical region: $\chi^2 > 15.086$.

$$\chi^2 = \frac{(5-6.1)^2}{6.1} + \frac{(4-6.0)^2}{6.0} + \cdots + \frac{(5-6.4)^2}{6.4} = 5.19.$$

Decision: Fail to reject H_0; distribution of nicotine contents is not significantly different from $n(x; 1.8, 0.4)$.

10.90 The hypotheses are

$$H_0: \text{Presence or absence of hypertension is independent of smoking habits,}$$

$$H_1: \text{Presence or absence of hypertension is not independent of smoking habits.}$$

$\alpha = 0.05$.
Critical region: $\chi^2 > 5.991$ with 2 degrees of freedom.
Computation:

Observed and expected frequencies				
	Nonsmokers	Moderate Smokers	Heavy Smokers	Total
Hypertension	21 (33.4)	36 (30.0)	30 (23.6)	87
No Hypertension	48 (35.6)	26 (32.0)	19 (25.4)	93
Total	69	62	49	180

$$\chi^2 = \frac{(21-33.4)^2}{33.4} + \cdots + \frac{(19-25.4)^2}{25.4} = 14.60.$$

Decision: Reject H_0; presence or absence of hypertension and smoking habits are not independent.

10.91 The hypotheses are

$$H_0: \text{A person's gender and time spent watching television are independent,}$$

$$H_1: \text{A person's gender and time spent watching television are not independent.}$$

$\alpha = 0.01$.
Critical region: $\chi^2 > 6.635$ with 1 degrees of freedom.
Computation:

Observed and expected frequencies			
	Male	Female	Total
Over 25 hours	15 (20.5)	29 (23.5)	44
Under 25 hours	27 (21.5)	19 (24.5)	46
Total	42	48	90

$$\chi^2 = \frac{(15-20.5)^2}{20.5} + \frac{(29-23.5)^2}{23.5} + \frac{(27-21.5)^2}{21.5} + \frac{(19-24.5)^2}{24.5} = 5.47.$$

Decision: Fail to reject H_0; a person's gender and time spent watching television are independent.

10.92 The hypotheses are

H_0 : Size of family is independent of level of education of father,

H_1 : Size of family and the education level of father are not independent.

$\alpha = 0.05$.
Critical region: $\chi^2 > 9.488$ with 4 degrees of freedom.
Computation:

Observed and expected frequencies				
	Number of Children			
Education	0–1	2–3	Over 3	Total
Elementary	14 (18.7)	37 (39.8)	32 (24.5)	83
Secondary	19 (17.6)	42 (37.4)	17 (23.0)	78
College	12 (8.7)	17 (18.8)	10 (11.5)	39
Total	45	96	59	200

$$\chi^2 = \frac{(14-18.7)^2}{18.7} + \frac{(37-39.8)^2}{39.8} + \cdots + \frac{(10-11.5)^2}{11.5} = 7.54.$$

Decision: Fail to reject H_0; size of family is independent of level of education of father.

10.93 The hypotheses are

H_0 : Occurrence of types of crime is independent of city district,

H_1 : Occurrence of types of crime is dependent upon city district.

$\alpha = 0.01$.
Critical region: $\chi^2 > 21.666$ with 9 degrees of freedom.
Computation:

Observed and expected frequencies					
District	Assault	Burglary	Larceny	Homicide	Total
1	162 (186.4)	118 (125.8)	451 (423.5)	18 (13.3)	749
2	310 (380.0)	196 (256.6)	996 (863.4)	25 (27.1)	1527
3	258 (228.7)	193 (154.4)	458 (519.6)	10 (16.3)	919
4	280 (214.9)	175 (145.2)	390 (488.5)	19 (15.3)	864
Total	1010	682	2295	72	4059

$$\chi^2 = \frac{(162-186.4)^2}{186.4} + \frac{(118-125.8)^2}{125.8} + \cdots + \frac{(19-15.3)^2}{15.3} = 124.59.$$

Decision: Reject H_0; occurrence of types of crime is dependent upon city district.

10.94 The hypotheses are

H_0 : The three cough remedies are equally effective,

H_1 : The three cough remedies are not equally effective.

$\alpha = 0.05$.
Critical region: $\chi^2 > 9.488$ with 4 degrees of freedom.
Computation:

Observed and expected frequencies				
	NyQuil	Robitussin	Triaminic	Total
No Relief	11 (11)	13 (11)	9 (11)	33
Some Relief	32 (29)	28 (29)	27 (29)	87
Total Relief	7 (10)	9 (10)	14 (10)	30
Total	50	50	50	150

$$\chi^2 = \frac{(11-11)^2}{11} + \frac{(13-11)^2}{11} + \cdots + \frac{(14-10)^2}{10} = 3.81.$$

Decision: Fail to reject H_0; the three cough remedies are equally effective.

10.95 The hypotheses are

H_0 : The attitudes among the four counties are homogeneous,

H_1 : The attitudes among the four counties are not homogeneous.

Computation:

Observed and expected frequencies					
	County				
Attitude	Craig	Giles	Franklin	Montgomery	Total
Favor	65 (74.5)	66 (55.9)	40 (37.3)	34 (37.3)	205
Oppose	42 (53.5)	30 (40.1)	33 (26.7)	42 (26.7)	147
No Opinion	93 (72.0)	54 (54.0)	27 (36.0)	24 (36.0)	198
Total	200	150	100	100	550

$$\chi^2 = \frac{(65-74.5)^2}{74.5} + \frac{(66-55.9)^2}{55.9} + \cdots + \frac{(24-36.0)^2}{36.0} = 31.17.$$

Since P-value $= P(\chi^2 > 31.17) < 0.001$ with 6 degrees of freedom, we reject H_0 and conclude that the attitudes among the four counties are not homogeneous.

10.96 The hypotheses are

H_0 : The proportions of widows and widowers are equal with respect to the different time period,

H_1 : The proportions of widows and widowers are not equal with respect to the different time period.

$\alpha = 0.05$.
Critical region: $\chi^2 > 5.991$ with 2 degrees of freedom.
Computation:

Observed and expected frequencies			
Years Lived	Widow	Widower	Total
Less than 5	25 (32)	39 (32)	64
5 to 10	42 (41)	40 (41)	82
More than 10	33 (26)	21 (26)	54
Total	100	100	200

$$\chi^2 = \frac{(25-32)^2}{32} + \frac{(39-32)^2}{32} + \cdots + \frac{(21-26)^2}{26} = 5.78.$$

Decision: Fail to reject H_0; the proportions of widows and widowers are equal with respect to the different time period.

10.97 The hypotheses are

H_0 : Proportions of household within each standard of living category are equal,

H_1 : Proportions of household within each standard of living category are not equal.

$\alpha = 0.05$.
Critical region: $\chi^2 > 12.592$ with 6 degrees of freedom.
Computation:

Observed and expected frequencies				
Period	Somewhat Better	Same	Not as Good	Total
1980: Jan.	72 (66.6)	144 (145.2)	84 (88.2)	300
May.	63 (66.6)	135 (145.2)	102 (88.2)	300
Sept.	47 (44.4)	100 (96.8)	53 (58.8)	200
1981: Jan.	40 (44.4)	105 (96.8)	55 (58.8)	200
Total	222	484	294	1000

$$\chi^2 = \frac{(72-66.6)^2}{66.6} + \frac{(144-145.2)^2}{145.2} + \cdots + \frac{(55-58.8)^2}{58.8} = 5.92.$$

Decision: Fail to reject H_0; proportions of household within each standard of living category are equal.

10.98 The hypotheses are

H_0 : Proportions of voters within each attitude category are the same for each of the three states,

H_1 : Proportions of voters within each attitude category are not the same for each of the three states.

$\alpha = 0.05$.
Critical region: $\chi^2 > 9.488$ with 4 degrees of freedom.
Computation:

Observed and expected frequencies				
	Support	Do not Support	Undecided	Total
Indiana	82 (94)	97 (79)	21 (27)	200
Kentucky	107 (94)	66 (79)	27 (27)	200
Ohio	93 (94)	74 (79)	33 (27)	200
Total	282	237	81	600

$$\chi^2 = \frac{(82-94)^2}{94} + \frac{(97-79)^2}{79} + \cdots + \frac{(33-27)^2}{27} = 12.56.$$

Decision: Reject H_0; the proportions of voters within each attitude category are not the same for each of the three states.

10.99 The hypotheses are

H_0 : Proportions of voters favoring candidate A, candidate B, or undecided are the same for each city,

H_1 : Proportions of voters favoring candidate A, candidate B, or undecided are not the same for each city.

$\alpha = 0.05$.
Critical region: $\chi^2 > 5.991$ with 2 degrees of freedom.
Computation:

Observed and expected frequencies			
	Richmond	Norfolk	Total
Favor A	204 (214.5)	225 (214.5)	429
Favor B	211 (204.5)	198 (204.5)	409
Undecided	85 (81)	77 (81)	162
Total	500	500	1000

$$\chi^2 = \frac{(204 - 214.5)^2}{214.5} + \frac{(225 - 214.5)^2}{214.5} + \cdots + \frac{(77 - 81)^2}{81} = 1.84.$$

Decision: Fail to reject H_0; the proportions of voters favoring candidate A, candidate B, or undecided are not the same for each city.

10.100 The hypotheses are

$$H_0 : p_1 = p_2 = p_3,$$
$$H_1 : p_1, p_2, \text{ and } p_3 \text{ are not all equal.}$$

$\alpha = 0.05$.
Critical region: $\chi^2 > 5.991$ with 2 degrees of freedom.
Computation:

Observed and expected frequencies				
	Denver	Phoenix	Rochester	Total
Watch Soap Operas	52 (48)	31 (36)	37 (36)	120
Do not Watch	148 (152)	119 (114)	113 (114)	380
Total	200	150	150	500

$$\chi^2 = \frac{(52 - 48)^2}{48} + \frac{(31 - 36)^2}{36} + \cdots + \frac{(113 - 114)^2}{114} = 1.39.$$

Decision: Fail to reject H_0; no difference among the proportions.

10.101 The hypotheses are

$$H_0 : p_1 = p_2,$$
$$H_1 : p_1 > p_2.$$

$\alpha = 0.01$.
Critical region: $z > 2.33$.
Computation: $\hat{p}_1 = 0.31$, $\hat{p}_2 = 0.24$, $\hat{p} = 0.275$, and

$$z = \frac{0.31 - 0.24}{\sqrt{(0.275)(0.725)(1/100 + 1/100)}} = 1.11.$$

Decision: Fail to reject H_0; proportions are the same.

10.102 Using paired t-test, we observe that $t = 1.55$ with P-value > 0.05. Hence, the data was not sufficient to show that the oxygen consumptions was higher when there was little or not CO.

10.103 (a) $H_0: \mu = 21.8$, $H_1: \mu \neq 21.8$; critical region in both tails.

(b) $H_0: p = 0.2$, $H_1: p > 0.2$; critical region in right tail.

(c) $H_0: \ \mu = 6.2$, $H_1: \ \mu > 6.2$; critical region in right tail.

(d) $H_0: \ p = 0.7$, $H_1: \ p < 0.7$; critical region in left tail.

(e) $H_0: \ p = 0.58$, $H_1: \ p \neq 0.58$; critical region in both tails.

(f) $H_0: \ \mu = 340$, $H_1: \ \mu < 340$; critical region in left tail.

10.104 The hypotheses are

$$H_0 : p_1 = p_2,$$
$$H_1 : p_1 > p_2.$$

$\alpha = 0.05$.
Critical region: $z > 1.645$.
Computation: $\hat{p}_1 = 0.24$, $\hat{p}_2 = 0.175$, $\hat{p} = 0.203$, and

$$z = \frac{0.24 - 0.175}{\sqrt{(0.203)(0.797)(1/300 + 1/400)}} = 2.12.$$

Decision: Reject H_0; there is statistical evidence to conclude that more Italians prefer white champagne at weddings.

10.105 $n_1 = n_2 = 5$, $\bar{x}_1 = 165.0$, $s_1 = 6.442$, $\bar{x}_2 = 139.8$, $s_2 = 12.617$, and $s_p = 10.02$. Hence

$$t = \frac{165 - 139.8}{(10.02)\sqrt{1/5 + 1/5}} = 3.98.$$

This is a one-sided test. Therefore, $0.0025 < P$-value < 0.005 with 8 degrees of freedom. Reject H_0; the speed is increased by using the facilitation tools.

10.106 (a) $H_0: \ p = 0.2$, $H_1: \ p > 0.2$; critical region in right tail.

(b) $H_0: \ \mu = 3$, $H_1: \ \mu \neq 3$; critical region in both tails.

(c) $H_0: \ p = 0.15$, $H_1: \ p < 0.15$; critical region in left tail.

(d) $H_0: \ \mu = \$10$, $H_1: \ \mu > \$10$; critical region in right tail.

(e) $H_0: \ \mu = 9$, $H_1: \ \mu \neq 9$; critical region in both tails.

10.107 The hypotheses are

$$H_0 : p_1 = p_2 = p_3,$$
$$H_1 : p_1, p_2, \text{ and } p_3 \text{ are not all equal.}$$

$\alpha = 0.01$.
Critical region: $\chi^2 > 9.210$ with 2 degrees of freedom.
Computation:

Observed and expected frequencies				
	Distributor			
Nuts	1	2	3	Total
Peanuts	345 (339)	313 (339)	359 (339)	1017
Other	155 (161)	187 (161)	141 (161)	483
Total	500	500	500	1500

$$\chi^2 = \frac{(345-339)^2}{339} + \frac{(313-339)^2}{339} + \cdots + \frac{(141-161)^2}{161} = 10.19.$$

Decision: Reject H_0; the proportions of peanuts for the three distributors are not equal.

10.108 The hypotheses are

$$H_0 : p_1 - p_2 = 0.03,$$
$$H_1 : p_1 - p_2 > 0.03.$$

Computation: $\hat{p}_1 = 0.60$ and $\hat{p}_2 = 0.48$.

$$z = \frac{(0.60 - 0.48) - 0.03}{\sqrt{(0.60)(0.40)/200 + (0.48)(0.52)/500}} = 2.18.$$

P-value $= P(Z > 2.18) = 0.0146$.
Decision: Reject H_0 at level higher than 0.0146; the difference in votes favoring the proposal exceeds 3%.

10.109 The hypotheses are

$$H_0 : p_1 = p_2 = p_3 = p_4,$$
$$H_1 : p_1, p_2, p_3, \text{ and } p_4 \text{ are not all equal.}$$

$\alpha = 0.01$.
Critical region: $\chi^2 > 11.345$ with 3 degrees of freedom.
Computation:

Observed and expected frequencies					
Preference	Maryland	Virginia	Georgia	Alabama	Total
Yes	65 (74)	71 (74)	78 (74)	82 (74)	296
No	35 (26)	29 (26)	22 (26)	18 (26)	104
Total	100	100	100	100	400

$$\chi^2 = \frac{(65-74)^2}{74} + \frac{(71-74)^2}{74} + \cdots + \frac{(18-26)^2}{26} = 8.84.$$

Decision: Fail to reject H_0; the proportions of parents favoring Bibles in elementary schools are the same across states.

10.110 $\bar{d} = -2.905$, $s_d = 3.3557$, and $t = \frac{\bar{d}}{s_d/\sqrt{n}} = -2.12$. Since $0.025 < P(T > 2.12) < 0.05$ with 5 degrees of freedom, we have $0.05 < P$-value < 0.10. There is no significant change in WBC leukograms.

10.111 $n_1 = 15$, $\bar{x}_1 = 156.33$, $s_1 = 33.09$, $n_2 = 18$, $\bar{x}_2 = 170.00$ and $s_2 = 30.79$. First we do the f-test to test equality of the variances. Since $f = \frac{s_1^2}{s_2^2} = 1.16$ and $f_{0.05}(15, 18) = 2.27$, we conclude that the two variances are equal.
To test the difference of the means, we first calculate $s_p = 31.85$. Therefore, $t = \frac{156.33-170.00}{(31.85)\sqrt{1/15+1/18}} = -1.23$ with a P-value > 0.10.
Decision: H_0 cannot be rejected at 0.05 level of significance.

10.112 $n_1 = n_2 = 10$, $\bar{x}_1 = 7.95$, $s_1 = 1.10$, $\bar{x}_2 = 10.26$ and $s_2 = 0.57$. First we do the f-test to test equality of the variances. Since $f = \frac{s_1^2}{s_2^2} = 3.72$ and $f_{0.05}(9, 9) = 3.18$, we conclude that the two variances are not equal at level 0.10.
To test the difference of the means, we first find the degrees of freedom $v = 13$ when round up. Also, $t = \frac{7.95-10.26}{\sqrt{1.10^2/10+0.57^2/10}} = -5.90$ with a P-value < 0.0005.
Decision: Reject H_0; there is a significant difference in the steel rods.

10.113 $n_1 = n_2 = 10$, $\bar{x}_1 = 21.5$, $s_1 = 5.3177$, $\bar{x}_2 = 28.3$ and $s_2 = 5.8699$. Since $f = \frac{s_1^2}{s_2^2} = 0.8207$ and $f_{0.05}(9, 9) = 3.18$, we conclude that the two variances are equal.
$s_p = 5.6001$ and hence $t = \frac{21.5-28.3}{(5.6001)\sqrt{1/10+1/10}} = -2.71$ with $0.005 < P$-value < 0.0075.
Decision: Reject H_0; the high income neighborhood produces significantly more wastewater to be treated.

10.114 $n_1 = n_2 = 16$, $\bar{x}_1 = 48.1875$, $s_1 = 4.9962$, $\bar{x}_2 = 43.7500$ and $s_2 = 4.6833$. Since $f = \frac{s_1^2}{s_2^2} = 1.1381$ and $f_{0.05}(15, 15) = 2.40$, we conclude that the two variances are equal.
$s_p = 4.8423$ and hence $t = \frac{48.1875-43.7500}{(4.8423)\sqrt{1/16+1/16}} = 2.59$. This is a two-sided test. Since $0.005 < P(T > 2.59) < 0.0075$, we have $0.01 < P$-value < 0.015.
Decision: Reject H_0; there is a significant difference in the number of defects.

10.115 The hypotheses are:

$$H_0 : \mu = 24 \times 10^{-4} \text{ gm},$$
$$H_1 : \mu < 24 \times 10^{-4} \text{ gm}.$$

$t = \frac{22.8-24}{4.8/\sqrt{50}} = -1.77$ with $0.025 < P$-value < 0.05. Hence, at significance level of $\alpha = 0.05$, the mean concentration of PCB in malignant breast tissue is less than 24×10^{-4} gm.

Chapter 11

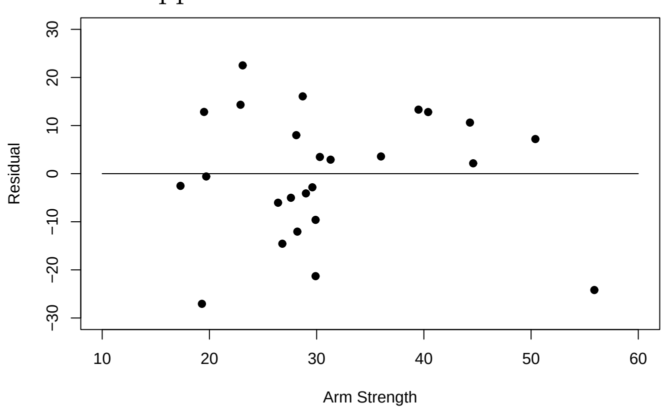
Simple Linear Regression and Correlation

11.1 (a) $\sum_i x_i = 778.7$, $\sum_i y_i = 2050.0$, $\sum_i x_i^2 = 26,591.63$, $\sum_i x_i y_i = 65,164.04$, $n = 25$. Therefore,

$$b = \frac{(25)(65,164.04) - (778.7)(2050.0)}{(25)(26,591.63) - (778.7)^2} = 0.5609,$$
$$a = \frac{2050 - (0.5609)(778.7)}{25} = 64.53.$$

(b) Using the equation $\hat{y} = 64.53 + 0.5609x$ with $x = 30$, we find $\hat{y} = 64.53 + (0.5609)(30) = 81.40$.

(c) Residuals appear to be random as desired.



11.2 (a) $\sum_i x_i = 707$, $\sum_i y_i = 658$, $\sum_i x_i^2 = 57,557$, $\sum_i x_i y_i = 53,258$, $n = 9$.

$$b = \frac{(9)(53,258) - (707)(658)}{(9)(57,557) - (707)^2} = 0.7771,$$
$$a = \frac{658 - (0.7771)(707)}{9} = 12.0623.$$

Hence $\hat{y} = 12.0623 + 0.7771x$.

(b) For $x = 85$, $\hat{y} = 12.0623 + (0.7771)(85) = 78$.

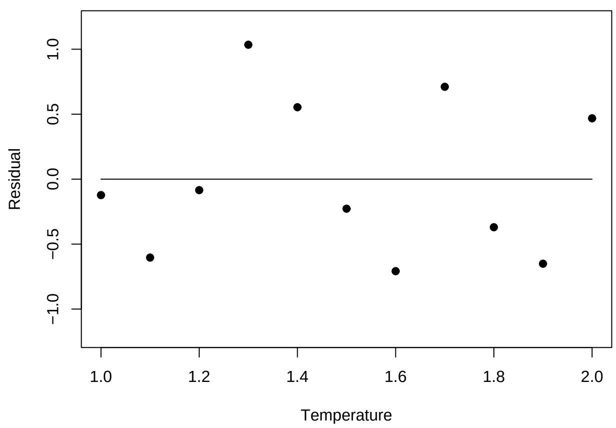
11.3 (a) $\sum_i x_i = 16.5$, $\sum_i y_i = 100.4$, $\sum_i x_i^2 = 25.85$, $\sum_i x_i y_i = 152.59$, $n = 11$. Therefore,

$$b = \frac{(11)(152.59) - (16.5)(100.4)}{(11)(25.85) - (16.5)^2} = 1.8091,$$
$$a = \frac{100.4 - (1.8091)(16.5)}{11} = 6.4136.$$

Hence $\hat{y} = 6.4136 + 1.8091x$

(b) For $x = 1.75$, $\hat{y} = 6.4136 + (1.8091)(1.75) = 9.580$.

(c) Residuals appear to be random as desired.



11.4 (a) $\sum_i x_i = 311.6$, $\sum_i y_i = 297.2$, $\sum_i x_i^2 = 8134.26$, $\sum_i x_i y_i = 7687.76$, $n = 12$.

$$b = \frac{(12)(7687.26) - (311.6)(297.2)^2}{=} - 0.6861,$$
$$a = \frac{297.2 - (-0.6861)(311.6)}{12} = 42.582.$$

Hence $\hat{y} = 42.582 - 0.6861x$.

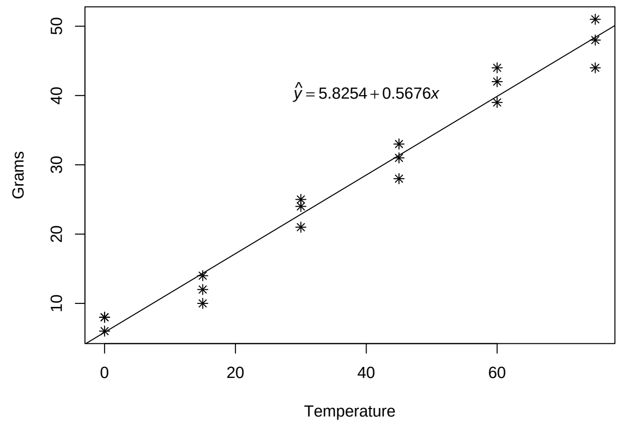
(b) At $x = 24.5$, $\hat{y} = 42.582 - (0.6861)(24.5) = 25.772$.

11.5 (a) $\sum_i x_i = 675$, $\sum_i y_i = 488$, $\sum_i x_i^2 = 37,125$, $\sum_i x_i y_i = 25,005$, $n = 18$. Therefore,

$$b = \frac{(18)(25,005) - (675)(488)}{(18)(37,125) - (675)^2} = 0.5676,$$
$$a = \frac{488 - (0.5676)(675)}{18} = 5.8254.$$

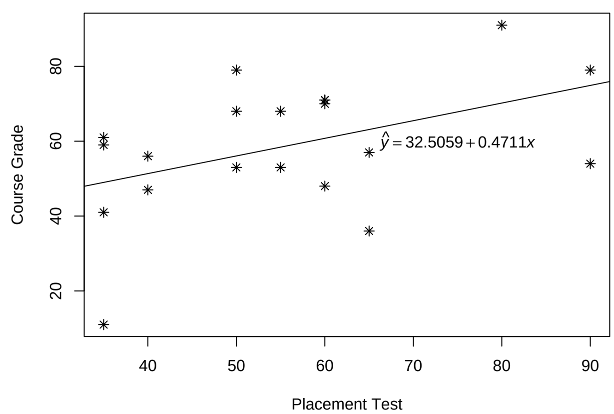
Hence $\hat{y} = 5.8254 + 0.5676x$

(b) The scatter plot and the regression line are shown below.



(c) For $x = 50$, $\hat{y} = 5.8254 + (0.5676)(50) = 34.205$ grams.

11.6 (a) The scatter plot and the regression line are shown below.



(b) $\sum_i x_i = 1110$, $\sum_i y_i = 1173$, $\sum_i x_i^2 = 67,100$, $\sum_i x_i y_i = 67,690$, $n = 20$. Therefore,

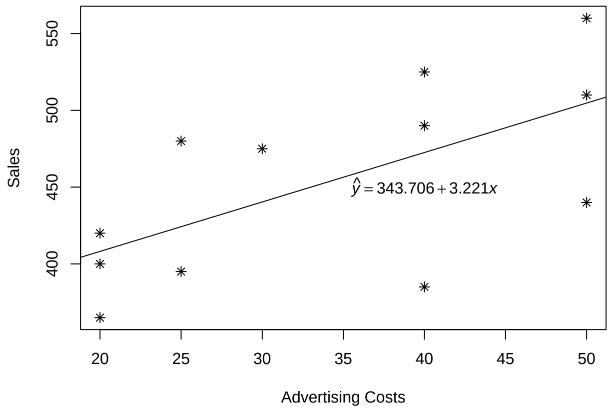
$$b = \frac{(20)(67,690) - (1110)(1173)}{(20)(67,100) - (1110)^2} = 0.4711,$$
$$a = \frac{1173 - (0.4711)(1110)}{20} = 32.5059.$$

Hence $\hat{y} = 32.5059 + 0.4711x$

(c) See part (a).

(d) For $\hat{y} = 60$, we solve $60 = 32.5059 + 0.4711x$ to obtain $x = 58.466$. Therefore, students scoring below 59 should be denied admission.

11.7 (a) The scatter plot and the regression line are shown here.



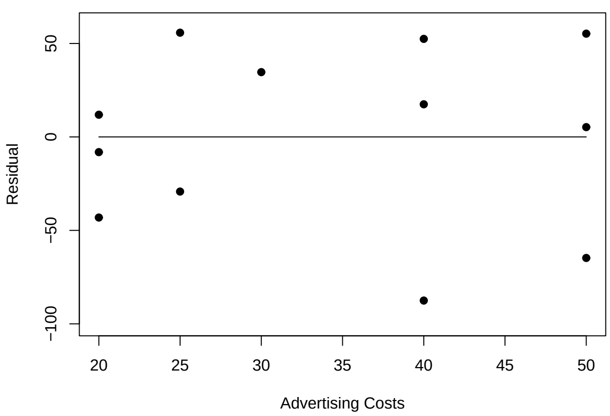
(b) $\sum_i x_i = 410$, $\sum_i y_i = 5445$, $\sum_i x_i^2 = 15,650$, $\sum_i x_i y_i = 191,325$, $n = 12$. Therefore,

$$b = \frac{(12)(191,325) - (410)(5445)}{(12)(15,650) - (410)^2} = 3.2208,$$
$$a = \frac{5445 - (3.2208)(410)}{12} = 343.7056.$$

Hence $\hat{y} = 343.7056 + 3.2208x$

(c) When $x = \$35$, $\hat{y} = 343.7056 + (3.2208)(35) = \$456.43$.

(d) Residuals appear to be random as desired.



11.8 (a) $\hat{y} = -1.70 + 1.81x$.

(b) $\hat{x} = (54 + 1.71)/1.81 = 30.78$.

11.9 (a) $\sum_i x_i = 45$, $\sum_i y_i = 1094$, $\sum_i x_i^2 = 244.26$, $\sum_i x_i y_i = 5348.2$, $n = 9$.

$$b = \frac{(9)(5348.2) - (45)(1094)}{(9)(244.26) - (45)^2} = -6.3240,$$
$$a = \frac{1094 - (-6.3240)(45)}{9} = 153.1755.$$

Hence $\hat{y} = 153.1755 - 6.3240x$.

(b) For $x = 4.8$, $\hat{y} = 153.1755 - (6.3240)(4.8) = 123$.

11.10 (a) $\hat{z} = cd^w$, $\ln \hat{z} = \ln c + (\ln d)w$; setting $\hat{y} = \ln z$, $a = \ln c$, $b = \ln d$, and $\hat{y} = a + bx$, we have

$x = w$	1	2	2	3	5	5
$y = \ln z$	8.7562	8.6473	8.6570	8.5932	8.5142	8.4960

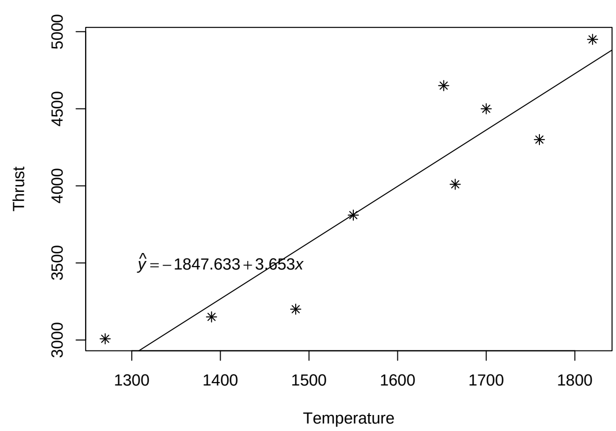
$\sum_i x_i = 18$, $\sum_i y_i = 51.6639$, $\sum_i x_i^2 = 68$, $\sum_i x_i y_i = 154.1954$, $n = 6$.

$$b = \ln d = \frac{(6)(154.1954) - (18)(51.6639)}{(6)(68) - (18)^2} = -0.0569,$$
$$a = \ln c = \frac{51.6639 - (-0.0569)(18)}{6} = 8.7813.$$

Now $c = e^{8.7813} = 6511.3364$, $d = e^{-0.0569} = 0.9447$, and $\hat{z} = 6511.3364 \times 0.9447^w$.

(b) For $w = 4$, $\hat{z} = 6511.3364 \times 0.9447^4 = \$5186.16$.

11.11 (a) The scatter plot and the regression line are shown here.

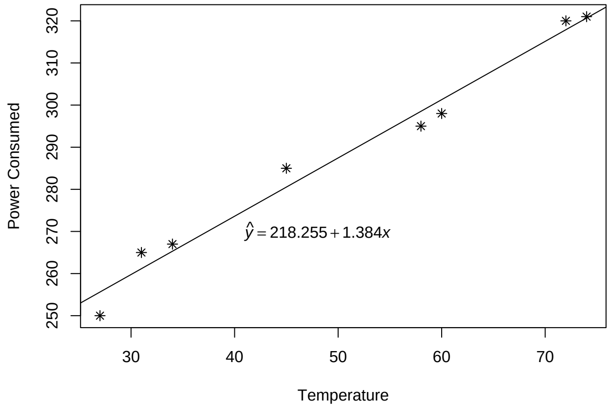


(b) $\sum_i x_i = 14,292$, $\sum_i y_i = 35,578$, $\sum_i x_i^2 = 22,954,054$, $\sum_i x_i y_i = 57,441,610$, $n = 9$. Therefore,

$$b = \frac{(9)(57,441,610) - (14,292)(35,578)}{(9)(22,954,054) - (14,292)^2} = 3.6529,$$
$$a = \frac{35,578 - (3.6529)(14,292)}{9} = -1847.69.$$

Hence $\hat{y} = -1847.69 + 3.6529x$.

11.12 (a) The scatter plot and the regression line are shown here.



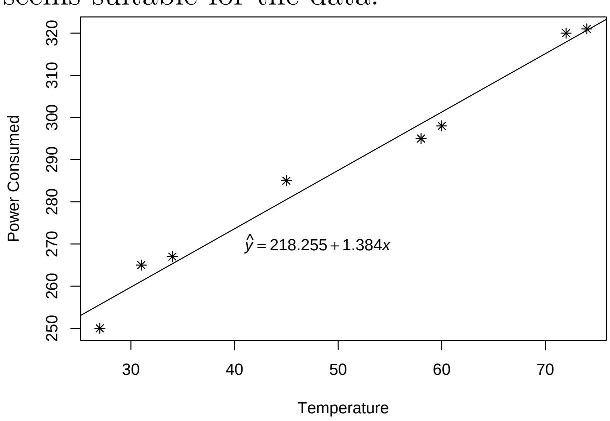
(b) $\sum_i x_i = 401$, $\sum_i y_i = 2301$, $\sum_i x_i^2 = 22,495$, $\sum_i x_i y_i = 118,652$, $n = 8$. Therefore,

$$b = \frac{(8)(118,652) - (401)(2301))}{(8)(22,495) - (401)^2} = 1.3839,$$
$$a = \frac{2301 - (1.3839)(401)}{8} = 218.26.$$

Hence $\hat{y} = 218.26 + 1.3839x$.

(c) For $x = 65°\text{F}$, $\hat{y} = 218.26 + (1.3839)(65) = 308.21$.

11.13 (a) The scatter plot and the regression line are shown here. A simple linear model seems suitable for the data.



(b) $\sum_i x_i = 999$, $\sum_i y_i = 670$, $\sum_i x_i^2 = 119,969$, $\sum_i x_i y_i = 74,058$, $n = 10$. Therefore,

$$b = \frac{(10)(74,058) - (999)(670)}{(10)(119,969) - (999)^2} = 0.3533,$$
$$a = \frac{670 - (0.3533)(999)}{10} = 31.71.$$

Hence $\hat{y} = 31.71 + 0.3533x$.

(c) See (a).

11.14 From the data summary, we obtain

$$b = \frac{(12)(318) - [(4)(12)][(12)(12)]}{(12)(232) - [(4)(12)]^2} = -6.45,$$
$$a = 12 - (-6.45)(4) = 37.8.$$

Hence, $\hat{y} = 37.8 - 6.45x$. It appears that attending professional meetings would not result in publishing more papers.

11.15 The least squares estimator A of α is a linear combination of normally distributed random variables and is thus normal as well.

$$E(A) = E(\bar{Y} - B\bar{x}) = E(\bar{Y}) - \bar{x}E(B) = \alpha + \beta\bar{x} - \beta\bar{x} = \alpha,$$
$$\sigma_A^2 = \sigma_{\bar{Y}-B\bar{x}}^2 = \sigma_{\bar{Y}}^2 + \bar{x}^2\sigma_B^2 - 2\bar{x}\sigma_{\bar{Y}B} = \frac{\sigma^2}{n} + \frac{\bar{x}^2\sigma^2}{\sum\limits_{i=1}^n (x_i - \bar{x})^2}, \text{ since } \sigma_{\bar{Y}B} = 0.$$

Hence

$$\sigma_A^2 = \frac{\sum\limits_{i=1}^n x_i^2}{n\sum\limits_{i=1}^n (x_i - \bar{x})^2}\sigma^2.$$

11.16 We have the following:

$$Cov(\bar{Y}, B) = E\left\{\left(\frac{1}{n}\sum_{i=1}^n Y_i - \frac{1}{n}\sum_{i=1}^n \mu_{Y_i}\right)\left[\frac{\sum\limits_{i=1}^n (x_i - \bar{x})Y_i}{\sum\limits_{i=1}^n (x_i - \bar{x})^2} - \frac{\sum\limits_{i=1}^n (x_i - \bar{x})\mu_{Y_i}}{\sum\limits_{i=1}^n (x_i - \bar{x})^2}\right]\right\}$$
$$= \frac{\sum\limits_{i=1}^n (x_i - \bar{x})E(Y_i - \mu_{Y_i})^2 + \sum\limits_{i\neq j}(x_i - \bar{x})E(Y_i - \mu_{Y_i})(Y_j - \mu_{Y_j})}{n\sum\limits_{i=1}^n (x_i - \bar{x})^2}$$
$$= \frac{\sum\limits_{i=1}^n (x_i - \bar{x})\sigma_{Y_i}^2 + \sum\limits_{i\neq j} Cov(Y_i, Y_j)}{n\sum\limits_{i=1}^n (x_i - \bar{x})^2}.$$

Now, $\sigma_{Y_i}^2 = \sigma^2$ for all i, and $Cov(Y_i, Y_j) = 0$ for $i \neq j$. Therefore,

$$Cov(\bar{Y}, B) = \frac{\sigma^2 \sum\limits_{i=1}^n (x_i - \bar{x})}{n\sum\limits_{i=1}^n (x_i - \bar{x})^2} = 0.$$

11.17 $S_{xx} = 26,591.63 - 778.7^2/25 = 2336.6824$, $S_{yy} = 172,891.46 - 2050^2/25 = 4791.46$, $S_{xy} = 65,164.04 - (778.7)(2050)/25 = 1310.64$, and $b = 0.5609$.

(a) $s^2 = \frac{4791.46-(0.5609)(1310.64)}{23} = 176.362$.

(b) The hypotheses are

$$H_0 : \beta = 0,$$
$$H_1 : \beta \neq 0.$$

$\alpha = 0.05$.
Critical region: $t < -2.069$ or $t > 2.069$.
Computation: $t = \frac{0.5609}{\sqrt{176.362/2336.6824}} = 2.04$.
Decision: Do not reject H_0.

11.18 $S_{xx} = 57,557 - 707^2/9 = 2018.2222$, $S_{yy} = 51,980 - 658^2/9 = 3872.8889$, $S_{xy} = 53,258 - (707)(658)/9 = 1568.4444$, $a = 12.0623$ and $b = 0.7771$.

(a) $s^2 = \frac{3872.8889-(0.7771)(1568.4444)}{7} = 379.150$.

(b) Since $s = 19.472$ and $t_{0.025} = 2.365$ for 7 degrees of freedom, then a 95% confidence interval is

$$12.0623 \pm (2.365)\sqrt{\frac{(379.150)(57,557)}{(9)(2018.222)}} = 12.0623 \pm 81.975,$$

which implies $-69.91 < \alpha < 94.04$.

(c) $0.7771 \pm (2.365)\sqrt{\frac{379.150}{2018.2222}}$ implies $-0.25 < \beta < 1.80$.

11.19 $S_{xx} = 25.85 - 16.5^2/11 = 1.1$, $S_{yy} = 923.58 - 100.4^2/11 = 7.2018$, $S_{xy} = 152.59 - (165)(100.4)/11 = 1.99$, $a = 6.4136$ and $b = 1.8091$.

(a) $s^2 = \frac{7.2018-(1.8091)(1.99)}{9} = 0.40$.

(b) Since $s = 0.632$ and $t_{0.025} = 2.262$ for 9 degrees of freedom, then a 95% confidence interval is

$$6.4136 \pm (2.262)(0.632)\sqrt{\frac{25.85}{(11)(1.1)}} = 6.4136 \pm 2.0895,$$

which implies $4.324 < \alpha < 8.503$.

(c) $1.8091 \pm (2.262)(0.632)/\sqrt{1.1}$ implies $0.446 < \beta < 3.172$.

11.20 $S_{xx} = 8134.26 - 311.6^2/12 = 43.0467$, $S_{yy} = 7407.80 - 297.2^2/12 = 47.1467$, $S_{xy} = 7687.76 - (311.6)(297.2)/12 = -29.5333$, $a = 42.5818$ and $b = -0.6861$.

(a) $s^2 = \frac{47.1467-(-0.6861)(-29.5333)}{10} = 2.688$.

(b) Since $s = 1.640$ and $t_{0.005} = 3.169$ for 10 degrees of freedom, then a 99% confidence interval is

$$42.5818 \pm (3.169)(1.640)\sqrt{\frac{8134.26}{(12)(43.0467)}} = 42.5818 \pm 20.6236,$$

which implies $21.958 < \alpha < 63.205$.

(c) $-0.6861 \pm (3.169)(1.640)/\sqrt{43.0467}$ implies $-1.478 < \beta < 0.106$.

11.21 $S_{xx} = 37,125 - 675^2/18 = 11,812.5$, $S_{yy} = 17,142 - 488^2/18 = 3911.7778$, $S_{xy} = 25,005 - (675)(488)/18 = 6705$, $a = 5.8254$ and $b = 0.5676$.

(a) $s^2 = \frac{3911.7778-(0.5676)(6705)}{16} = 6.626$.

(b) Since $s = 2.574$ and $t_{0.005} = 2.921$ for 16 degrees of freedom, then a 99% confidence interval is

$$5.8261 \pm (2.921)(2.574)\sqrt{\frac{37,125}{(18)(11,812.5)}} = 5.8261 \pm 3.1417,$$

which implies $2.686 < \alpha < 8.968$.

(c) $0.5676 \pm (2.921)(2.574)/\sqrt{11,812.5}$ implies $0.498 < \beta < 0.637$.

11.22 The hypotheses are

$$H_0 : \alpha = 10,$$
$$H_1 : \alpha > 10.$$

$\alpha = 0.05$.
Critical region: $t > 1.734$.
Computations: $S_{xx} = 67,100 - 1110^2/20 = 5495$, $S_{yy} = 74,725 - 1173^2/20 = 5928.55$, $S_{xy} = 67,690 - (1110)(1173)/20 = 2588.5$, $s^2 = \frac{5928.55-(0.4711)(2588.5)}{18} = 261.617$ and then $s = 16.175$. Now

$$t = \frac{32.51 - 10}{16.175\sqrt{67,100/(20)(5495)}} = 1.78.$$

Decision: Reject H_0 and claim $\alpha > 10$.

11.23 The hypotheses are

$$H_0 : \beta = 6,$$
$$H_1 : \beta < 6.$$

$\alpha = 0.025$.
Critical region: $t = -2.228$.

Computations: $S_{xx} = 15,650 - 410^2/12 = 1641.667$, $S_{yy} = 2,512.925 - 5445^2/12 = 42,256.25$, $S_{xy} = 191,325 - (410)(5445)/12 = 5,287.5$, $s^2 = \frac{42,256.25-(3,221)(5,287.5)}{10} = 2,522.521$ and then $s = 50.225$. Now

$$t = \frac{3.221 - 6}{50.225/\sqrt{1641.667}} = -2.24.$$

Decision: Reject H_0 and claim $\beta < 6$.

11.24 Using the value $s = 19.472$ from Exercise 11.18(a) and the fact that $\bar{y}_0 = 74.230$ when $x_0 = 80$, and $\bar{x} = 78.556$, we have

$$74.230 \pm (2.365)(19.472)\sqrt{\frac{1}{9} + \frac{1.444^2}{2018.222}} = 74.230 \pm 15.4216.$$

Simplifying it we get $58.808 < \mu_{Y \mid 80} < 89.652$.

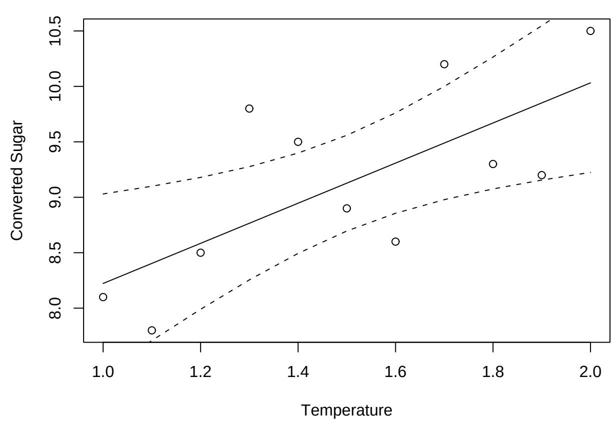
11.25 Using the value $s = 1.64$ from Exercise 11.20(a) and the fact that $y_0 = 25.7724$ when $x_0 = 24.5$, and $\bar{x} = 25.9667$, we have

(a) $25.7724 \pm (2.228)(1.640)\sqrt{\frac{1}{12} + \frac{(-1.4667)^2}{43.0467}} = 25.7724 \pm 1.3341$ implies $24.438 < \mu_{Y \mid 24.5} < 27.106$.

(b) $25.7724 \pm (2.228)(1.640)\sqrt{1 + \frac{1}{12} + \frac{(-1.4667)^2}{43.0467}} = 25.7724 \pm 3.8898$ implies $21.883 < y_0 < 29.662$.

11.26 95% confidence bands are obtained by plotting the limits

$$(6.4136 + 1.809x) \pm (2.262)(0.632)\sqrt{\frac{1}{11} + \frac{(x - 1.5)^2}{1.1}}.$$



11.27 Using the value $s = 0.632$ from Exercise 11.19(a) and the fact that $y_0 = 9.308$ when $x_0 = 1.6$, and $\bar{x} = 1.5$, we have

$$9.308 \pm (2.262)(0.632)\sqrt{1 + \frac{1}{11} + \frac{0.1^2}{1.1}} = 9.308 \pm 1.4994$$

implies $7.809 < y_0 < 10.807$.

11.28 sing the value $s = 2.574$ from Exercise 11.21(a) and the fact that $y_0 = 34.205$ when $x_0 = 50$, and $\bar{x} = 37.5$, we have

(a) $34.205 \pm (2.921)(2.574)\sqrt{\frac{1}{18} + \frac{12.5^2}{11{,}812.5}} = 34.205 \pm 1.9719$ implies $32.23 < \mu_{Y \mid 50} < 36.18$.

(b) $34.205 \pm (2.921)(2.574)\sqrt{1 + \frac{1}{18} + \frac{12.5^2}{11{,}812.5}} = 34.205 \pm 7.7729$ implies $26.43 < y_0 < 41.98$.

11.29 (a) 17.1812.

(b) The goal of 30 mpg is unlikely based on the confidence interval for mean mpg, $(27.95, 29.60)$.

(c) Based on the prediction interval, the Lexus ES300 should exceed 18 mpg.

11.30 It is easy to see that

$$\sum_{i=1}^{n}(y_i - \hat{y}_i) = \sum_{i=1}^{n}(y_i - a - bx_i) = \sum_{i=1}^{n}[y_i - (\bar{y} - b\bar{x}) - bx_i)$$
$$= \sum_{i=1}^{n}(y_i - \bar{y}_i) - b\sum_{i=1}^{n}(x_i - \bar{x}) = 0,$$

since $a = \bar{y} - b\bar{x}$.

11.31 When there are only two data points $x_1 \neq x_2$, using Exercise 11.30 we know that $(y_1 - \hat{y}_1) + (y_2 - \hat{y}_2) = 0$. On the other hand, by the method of least squares on page 395, we also know that $x_1(y_1 - \hat{y}_1) + x_2(y_2 - \hat{y}_2) = 0$. Both of these equations yield $(x_2 - x_1)(y_2 - \hat{y}_2) = 0$ and hence $y_2 - \hat{y}_2 = 0$. Therefore, $y_1 - \hat{y}_1 = 0$. So,

$$SSE = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 = 0.$$

Since $R^2 = 1 - \frac{SSE}{SST}$, we have $R^2 = 1$.

11.32 (a) Suppose that the fitted model is $\hat{y} = bx$. Then

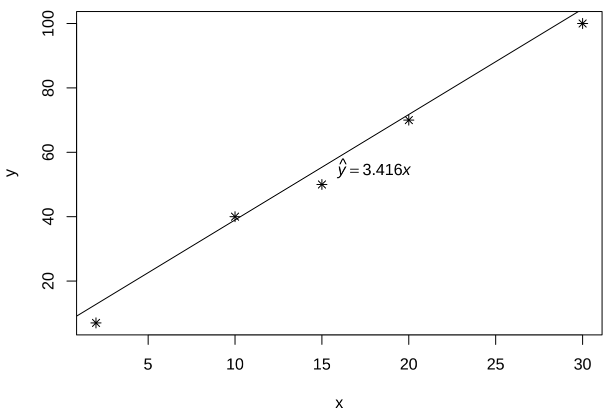
$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - bx_i)^2.$$

Taking derivative of the above with respect to b and setting the derivative to zero, we have $-2\sum_{i=1}^{n} x_i(y_i - bx_i) = 0$, which implies $b = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$.

(b) $\sigma_B^2 = \frac{Var\left(\sum_{i=1}^{n} x_i Y_i\right)}{(\sum_{i=1}^{n} x_i^2)^2} = \frac{\sum_{i=1}^{n} x_i^2 \sigma_{Y_i}^2}{(\sum_{i=1}^{n} x_i^2)^2} = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}$, since Y_i's are independent.

(c) $E(B) = \frac{E\left(\sum_{i=1}^{n} x_i Y_i\right)}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} x_i(\beta x_i)}{\sum_{i=1}^{n} x_i^2} = \beta.$

11.33 (a) The scatter plot of the data is shown next.



(b) $\sum_{i=1}^{n} x_i^2 = 1629$ and $\sum_{i=1}^{n} x_i y_i = 5564$. Hence $b = \frac{5564}{1629} = 3.4156$. So, $\hat{y} = 3.4156x$.

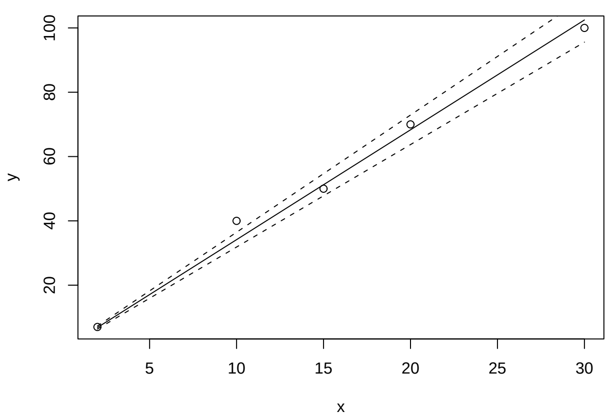
(c) See (a).

(d) Since there is only one regression coefficient, β, to be estimated, the degrees of freedom in estimating σ^2 is $n-1$. So,

$$\hat{\sigma}^2 = s^2 = \frac{SSE}{n-1} = \frac{\sum_{i=1}^{n}(y_i - bx_i)^2}{n-1}.$$

(e) $Var(\hat{y}_i) = Var(Bx_i) = x_i^2 Var(B) = \frac{x_i^2\sigma^2}{\sum_{i=1}^{n} x_i^2}$.

(f) The plot is shown next.



11.34 Using part (e) of Exercise 11.33, we can see that the variance of a prediction y_0 at x_0

is $\sigma^2_{y_0} = \sigma^2 \left(1 + \frac{x_0^2}{\sum\limits_{i=1}^{n} x_i^2}\right)$. Hence the 95% prediction limits are given as

$$(3.4145)(25) \pm (2.776)\sqrt{11.16132}\sqrt{1 + \frac{25^2}{1629}} = 85.3625 \pm 10.9092,$$

which implies $74.45 < y_0 < 96.27$.

11.35 (a) As shown in Exercise 11.32, the least squares estimator of β is $b = \frac{\sum\limits_{i=1}^{n} x_i y_i}{\sum\limits_{i=1}^{n} x_i^2}$.

(b) Since $\sum\limits_{i=1}^{n} x_i y_i = 197.59$, and $\sum\limits_{i=1}^{n} x_i^2 = 98.64$, then $b = \frac{197.59}{98.64} = 2.003$ and $\hat{y} = 2.003x$.

11.36 It can be calculated that $b = 1.929$ and $a = 0.349$ and hence $\hat{y} = 0.349 + 1.929x$ when intercept is in the model. To test the hypotheses

$$H_0 : \alpha = 0,$$
$$H_1 : \alpha \neq 0,$$

with 0.10 level of significance, we have the critical regions as $t < -2.132$ or $t > 2.132$.
Computations: $s^2 = 0.0957$ and $t = \frac{0.349}{\sqrt{(0.0957)(98.64)/(6)(25.14)}} = 1.40$.
Decision: Fail to reject H_0; the intercept appears to be zero.

11.37 Now since the true model has been changed,

$$E(B) = \frac{\sum\limits_{i=1}^{n}(x_{1i} - \bar{x}_1)E(Y_i)}{\sum\limits_{i=1}^{n}(x_{1i} - \bar{x}_i)^2} = \frac{\sum\limits_{i=1}^{n}(x_{1i} - \bar{x}_1)(\alpha + \beta x_{1i} + \gamma x_{2i})}{\sum\limits_{i=1}^{n}(x_{1i} - \bar{x}_1)^2}$$
$$= \frac{\beta \sum\limits_{i=1}^{n}(x_{1i} - \bar{x}_1)^2 + \gamma \sum\limits_{i=1}^{n}(x_{1i} - \bar{x}_1)x_{2i}}{\sum\limits_{i=1}^{n}(x_{1i} - \bar{x}_1)^2} = \beta + \gamma \frac{\sum\limits_{i=1}^{n}(x_{1i} - \bar{x}_1)x_{2i}}{\sum\limits_{i=1}^{n}(x_{1i} - \bar{x}_1)^2}.$$

11.38 The hypotheses are

$$H_0 : \beta = 0,$$
$$H_1 : \beta \neq 0.$$

Level of significance: 0.05.
Critical regions: $f > 5.12$.
Computations: $SSR = bS_{xy} = \frac{1.8091}{1.99} = 3.60$ and $SSE = S_{yy} - SSR = 7.20 - 3.60 = 3.60$.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Regression	3.60	1	3.60	9.00
Error	3.60	9	0.40	
Total	7.20	10		

Decision: Reject H_0.

11.39 (a) $S_{xx} = 1058$, $S_{yy} = 198.76$, $S_{xy} = -363.63$, $b = \frac{S_{xy}}{S_{xx}} = -0.34370$, and $a = \frac{210-(-0.34370)(172.5)}{25} = 10.81153$.

(b) The hypotheses are

$$H_0: \text{ The regression is linear in } x,$$

$$H_1: \text{ The regression is nonlinear in } x.$$

$\alpha = 0.05$.
Critical regions: $f > 3.10$ with 3 and 20 degrees of freedom.
Computations: $SST = S_{yy} = 198.76$, $SSR = bS_{xy} = 124.98$ and $SSE = S_{yy} - SSR = 73.98$. Since

$$T_{1.} = 51.1, T_{2.} = 51.5, T_{3.} = 49.3, T_{4.} = 37.0 \text{ and } T_{5.} = 22.1,$$

then

$$SSE(\text{pure}) = \sum_{i=1}^{5}\sum_{j=1}^{5} y_{ij}^2 - \sum_{i=1}^{5} \frac{T_{i.}^2}{5} = 1979.60 - 1910.272 = 69.33.$$

Hence the "Lack-of-fit SS" is $73.78 - 69.33 = 4.45$.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Regression	124.98	1	124.98	
Error	73.98	23	3.22	
Lack of fit	4.45	3	1.48	0.43
Pure error	69.33	20	3.47	
Total	198.76	24		

Decision: Do not reject H_0.

11.40 The hypotheses are

$$H_0: \text{ The regression is linear in } x,$$

$$H_1: \text{ The regression is nonlinear in } x.$$

$\alpha = 0.05$.
Critical regions: $f > 3.26$ with 4 and 12 degrees of freedom.

Computations: $SST = S_{yy} = 3911.78$, $SSR = bS_{xy} = 3805.89$ and $SSE = S_{yy} - SSR = 105.89$. $SSE(\text{pure}) = \sum\limits_{i=1}^{6}\sum\limits_{j=1}^{3} y_{ij}^2 - \sum\limits_{i=1}^{6} \frac{T_{i.}^2}{3} = 69.33$, and the "Lack-of-fit SS" is $105.89 - 69.33 = 36.56$.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Regression	3805.89	1	3805.89	
Error	105.89	16	6.62	
Lack of fit	36.56	4	9.14	1.58
Pure error	69.33	12	5.78	
Total	3911.78	17		

Decision: Do not reject H_0; the lack-of-fit test is insignificant.

11.41 The hypotheses are

$$H_0: \text{ The regression is linear in } x,$$

$$H_1: \text{ The regression is nonlinear in } x.$$

$\alpha = 0.05$.
Critical regions: $f > 3.00$ with 6 and 12 degrees of freedom.
Computations: $SST = S_{yy} = 5928.55$, $SSR = bS_{xy} = 1219.35$ and $SSE = S_{yy} - SSR = 4709.20$. $SSE(\text{pure}) = \sum\limits_{i=1}^{8}\sum\limits_{j=1}^{n_i} y_{ij}^2 - \sum\limits_{i=1}^{8} \frac{T_{i.}^2}{n_i} = 3020.67$, and the "Lack-of-fit SS" is $4709.20 - 3020.67 = 1688.53$.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Regression	1219.35	1	1219.35	
Error	4709.20	18	261.62	
Lack of fit	1688.53	6	281.42	1.12
Pure error	3020.67	12	251.72	
Total	5928.55	19		

Decision: Do not reject H_0; the lack-of-fit test is insignificant.

11.42 (a) $t = 2.679$ and $0.01 < P(T > 2.679) < 0.015$, hence $0.02 < P\text{-value} < 0.03$. There is a strong evidence that the slope is not 0. Hence emitter drive-in time influences gain in a positive linear fashion.

(b) $f = 56.41$ which results in a strong evidence that the lack-of-fit test is significant. Hence the linear model is not adequate.

(c) Emitter does does not influence gain in a linear fashion. A better model is a quadratic one using emitter drive-in time to explain the variability in gain.

11.43 $\hat{y} = -21.0280 + 0.4072x$; $f_{\text{LOF}} = 1.71$ with a P-value $= 0.2517$. Hence, lack-of-fit test is insignificant and the linear model is adequate.

11.44 (a) $\hat{y} = 0.011571 + 0.006462x$ with $t = 7.532$ and $P(T > 7.532) < 0.0005$ Hence P-value < 0.001; the slope is significantly different from 0 in the linear regression model.

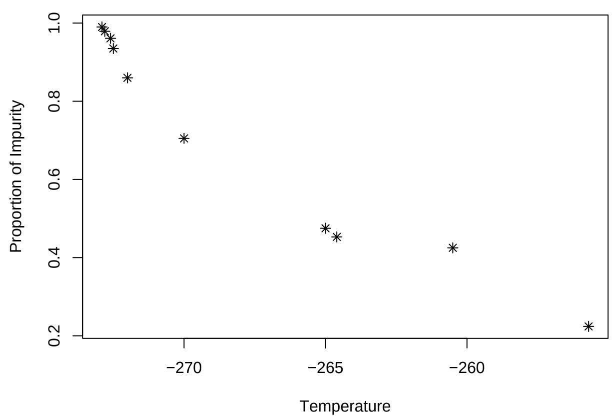
(b) $f_{\text{LOF}} = 14.02$ with P-value < 0.0001. The lack-of-fit test is significant and the linear model does not appear to be the best model.

11.45 (a) $\hat{y} = -11.3251 - 0.0449$ temperature.

(b) Yes.

(c) 0.9355.

(d) The proportion of impurities does depend on temperature.



However, based on the plot, it does not appear that the dependence is in linear fashion. If there were replicates, a lack-of-fit test could be performed.

11.46 (a) $\hat{y} = 125.9729 + 1.7337$ population; P-value for the regression is 0.0023.

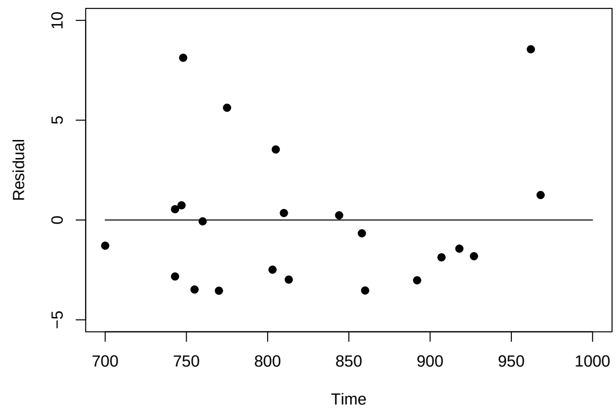
(b) $f_{6,2} = 0.49$ and P-value $= 0.7912$; the linear model appears to be adequate based on the lack-of-fit test.

(c) $f_{1,2} = 11.96$ and P-value $= 0.0744$. The results do not change. The pure error test is not as sensitive because the loss of error degrees of freedom.

11.47 (a) The figure is shown next.

(b) $\hat{y} = -175.9025 + 0.0902$ year; $R^2 = 0.3322$.

(c) There is definitely a relationship between year and nitrogen oxide. It does not appear to be linear.



11.48 The ANOVA model is:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Regression	135.2000	1	135.2000	
Error	10.4700	14	0.7479	
Lack of fit	6.5150	2	3.2575	9.88
Pure error	3.9550	12	0.3296	
Total	145.6700	15		

The P-value $= 0.0029$ with $f = 9.88$.
Decision: Reject H_0; the lack-of-fit test is significant.

11.49 $S_{xx} = 36,354 - 35,882.667 = 471.333$, $S_{yy} = 38,254 - 37,762.667 = 491.333$, and $S_{xy} = 36,926 - 36,810.667 = 115.333$. So, $r = \frac{115}{\sqrt{(471.333)(491.333)}} = 0.240$.

11.50 The hypotheses are

$$H_0 : \rho = 0,$$
$$H_1 : \rho \neq 0.$$

$\alpha = 0.05$.
Critical regions: $t < -2.776$ or $t > 2.776$.
Computations: $t = \frac{0.240\sqrt{4}}{\sqrt{1-0.240^2}} = 0.51$.
Decision: Do not reject H_0.

11.51 Since $b = \frac{S_{xy}}{S_{xx}}$, we can write $s^2 = \frac{S_{yy}-bS_{xy}}{n-2} = \frac{S_{yy}-b^2S_{xx}}{n-2}$. Also, $b = r\sqrt{\frac{S_{yy}}{S_{xx}}}$ so that $s^2 = \frac{S_{yy}-r^2S_{yy}}{n-2} = \frac{(1-r^2)S_{yy}}{n-2}$, and hence

$$t = \frac{b}{s\sqrt{S_{xx}}} = \frac{r\sqrt{S_{yy}/S_{xx}}}{\sqrt{S_{yy}S_{xx}(1-r^2)/(n-2)}} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

11.52 (a) $S_{xx} = 128.6602 - 32.68^2/9 = 9.9955$, $S_{yy} = 7980.83 - 266.7^2/9 = 77.62$, and $S_{xy} = 990.268 - (32.68)(266.7)/9 = 21.8507$. So, $r = \frac{21.8507}{\sqrt{(9.9955)(77.62)}} = 0.784$.

(b) The hypotheses are

$$H_0 : \rho = 0,$$
$$H_1 : \rho > 0.$$

$\alpha = 0.01$.
Critical regions: $t > 2.998$.
Computations: $t = \frac{0.784\sqrt{7}}{\sqrt{1-0.784^2}} = 3.34$.
Decision: Reject H_0; $\rho > 0$.

(c) $(0.784)^2(100\%) = 61.5\%$.

11.53 (a) From the data of Exercise 11.1 we can calculate

$$S_{xx} = 26,591.63 - (778.7)^2/25 = 2336.6824,$$
$$S_{yy} = 172,891.46 - (2050)^2/25 = 4791.46,$$
$$S_{xy} = 65,164.04 - (778.7)(2050)/25 = 1310.64.$$

Therefore, $r = \frac{1310.64}{\sqrt{(2236.6824)(4791.46)}} = 0.392$.

(b) The hypotheses are

$$H_0 : \rho = 0,$$
$$H_1 : \rho \neq 0.$$

$\alpha = 0.05$.
Critical regions: $t < -2.069$ or $t > 2.069$.
Computations: $t = \frac{0.392\sqrt{23}}{\sqrt{1-0.392^2}} = 2.04$.
Decision: Fail to reject H_0 at level 0.05. However, the P-value $= 0.0530$ which is marginal.

11.54 (a) From the data of Exercise 11.9 we find $S_{xx} = 244.26 - 45^2/9 = 19.26$, $S_{yy} = 133,786 - 1094^2/9 = 804.2222$, and $S_{xy} = 5348.2 - (45)(1094)/9 = -121.8$. So, $r = \frac{-121.8}{\sqrt{(19.26)(804.2222)}} = -0.979$.

(b) The hypotheses are

$$H_0 : \rho = -0.5,$$
$$H_1 : \rho < -0.5.$$

$\alpha = 0.025$.
Critical regions: $z < -1.96$.
Computations: $z = \frac{\sqrt{6}}{2} \ln\left[\frac{(0.021)(1.5)}{(1.979)(0.5)}\right] = -4.22$.
Decision: Reject H_0; $\rho < -0.5$.

(c) $(-0.979)^2(100\%) = 95.8\%$.

11.55 Using the value $s = 16.175$ from Exercise 11.6 and the fact that $\hat{y}_0 = 48.994$ when $x_0 = 35$, and $\bar{x} = 55.5$, we have

(a) $48.994 \pm (2.101)(16.175)\sqrt{1/20 + (-20.5)^2/5495}$ which implies to $36.908 < \mu_{Y \mid 35} < 61.080$.

(b) $48.994 \pm (2.101)(16.175)\sqrt{1 + 1/20 + (-20.5)^2/5495}$ which implies to $12.925 < y_0 < 85.063$.

11.56 The fitted model can be derived as $\hat{y} = 3667.3968 - 47.3289x$.
The hypotheses are

$$H_0 : \beta = 0,$$
$$H_1 : \beta \neq 0.$$

$t = -0.30$ with P-value $= 0.77$. Hence H_0 cannot be rejected.

11.57 (a) $S_{xx} = 729.18 - 118.6^2/20 = 25.882$, $S_{xy} = 1714.62 - (118.6)(281.1)/20 = 47.697$, so $b = \frac{S_{xy}}{S_{xx}} = 1.8429$, and $a = \bar{y} - b\bar{x} = 3.1266$. Hence $\hat{y} = 3.1266 + 1.8429x$.

(b) The hypotheses are

$$H_0: \text{ the regression is linear in } x,$$
$$H_1: \text{ the regression is not linear in } x.$$

$\alpha = 0.05$.
Critical region: $f > 3.07$ with 8 and 10 degrees of freedom.
Computations: $SST = 13.3695$, $SSR = 87.9008$, $SSE = 50.4687$, $SSE(\text{pure}) = 16.375$, and Lack-of-fit SS $= 34.0937$.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Regression	87.9008	1	87.9008	
Error	50.4687	18	2.8038	
Lack of fit	34.0937	8	4.2617	2.60
Pure error	16.375	10	1.6375	
Total	138.3695	19		

The P-value $= 0.0791$. The linear model is adequate at the level 0.05.

11.58 Using the value $s = 50.225$ and the fact that $\hat{y}_0 = \$448.644$ when $x_0 = \$45$, and $\bar{x} = \$34.167$, we have

(a) $488.644 \pm (1.812)(50.225)\sqrt{1/12 + \frac{10.833^2}{1641.667}}$, which implies $452.835 < \mu_{Y \mid 45} < 524.453$.

(b) $488.644 \pm (1.812)(50.225)\sqrt{1 + 1/12 + \frac{10.833^2}{1641.667}}$, which implies $390.845 < y_0 < 586.443$.

11.59 (a) $\hat{y} = 7.3598 + 135.4034x$.

(b) $SS(\text{Pure Error}) = 52,941.06$; $f_{\text{LOF}} = 0.46$ with P-value $= 0.64$. The lack-of-fit test is insignificant.

(c) No.

11.60 (a) $S_{xx} = 672.9167$, $S_{yy} = 728.25$, $S_{xy} = 603.75$ and $r = \frac{603.75}{\sqrt{(672.9167)(728.25)}} = 0.862$, which means that $(0.862)^2(100\%) = 74.3\%$ of the total variation of the values of Y in our sample is accounted for by a linear relationship with the values of X.

(b) To estimate and test hypotheses on ρ, X and Y are assumed to be random variables from a bivariate normal distribution.

(c) The hypotheses are

$$H_0 : \rho = 0.5,$$
$$H_1 : \rho > 0.5.$$

$\alpha = 0.01$.
Critical regions: $z > 2.33$.
Computations: $z = \frac{\sqrt{9}}{2} \ln\left[\frac{(1.862)(0.5)}{(0.138)(1.5)}\right] = 2.26$.
Decision: Reject H_0; $\rho > 0.5$.

11.61 $s^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}$. Using the centered model, $\hat{y}_i = \bar{y} + b(x_i - \bar{x}) + \epsilon_i$.

$$\begin{aligned}
(n-2)E(S^2) &= E\sum_{i=1}^{n}[\alpha + \beta(x_i - \bar{x}) + \epsilon_i - (\bar{y} + b(x_i - \bar{x}))]^2 \\
&= \sum_{i=1}^{n} E\left[(\alpha - \bar{y})^2 + (\beta - b)^2(x_i - \bar{x})^2 + \epsilon_i^2 - 2b(x_i - \bar{x})\epsilon_i - 2\bar{y}\epsilon_i\right], \\
&\quad \text{(other cross product terms go to 0)} \\
&= \frac{n\sigma^2}{n} + \frac{\sigma^2 S_{xx}}{S_{xx}} + n\sigma^2 - 2\frac{\sigma^2 S_{xx}}{S_{xx}} - 2\frac{n\sigma^2}{n} \\
&= (n-2)\sigma^2.
\end{aligned}$$

11.62 (a) The confidence interval is an interval on the mean sale price for a given buyer's bid. The prediction interval is an interval on a future observed sale price for a given buyer's bid.

(b) The standard errors of the prediction of sale price depend on the value of the buyer's bid.

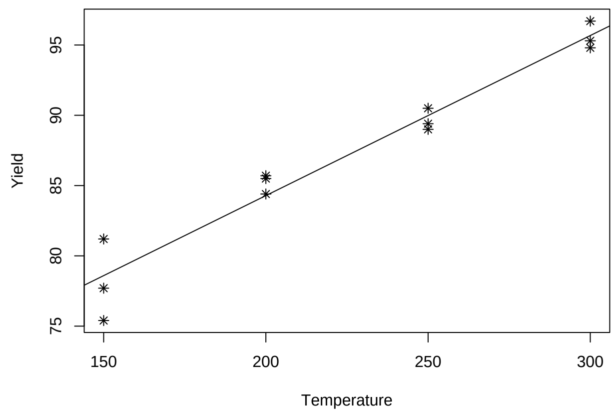
(c) Observations 4, 9, 10, and 17 have the lowest standard errors of prediction. These observations have buyer's bids very close to the mean.

11.63 (a) The residual plot appears to have a pattern and not random scatter. The R^2 is only 0.82.

(b) The log model has an R^2 of 0.84. There is still a pattern in the residuals.

(c) The model using gallons per 100 miles has the best R^2 with a 0.85. The residuals appear to be more random. This model is the best of the three models attempted. Perhaps a better model could be found.

11.64 (a) The plot of the data and an added least squares fitted line are given here.



(b) Yes.

(c) $\hat{y} = 61.5133 + 0.1139x$.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Regression	486.21	1	486.21	
Error	24.80	10	2.48	
Lack of fit	3.61	2	1.81	0.68
Pure error	21.19	8	2.65	
Total	511.01	11		

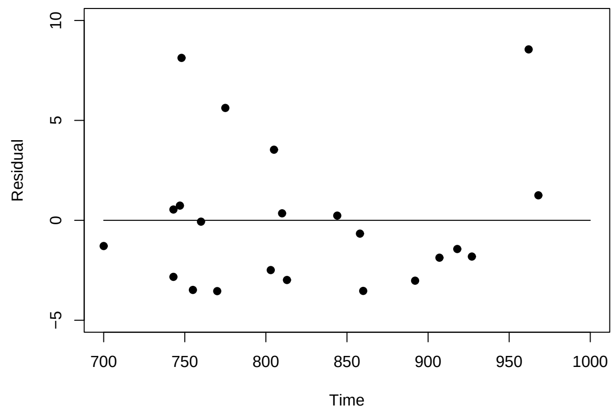
The P-value $= 0.533$.

(d) The results in (c) show that the linear model is adequate.

11.65 (a) $\hat{y} = 90.8904 - 0.0513x$.

(b) The t-value in testing $H_0: \beta = 0$ is -6.533 which results in a P-value < 0.0001. Hence, the time it takes to run two miles has a significant influence on maximum oxygen uptake.

(c) The residual graph shows that there may be some systematic behavior of the residuals and hence the residuals are not completely random



11.66 Let $Y_i^* = Y_i - \alpha$, for $i = 1, 2, \ldots, n$. The model $Y_i = \alpha + \beta x_i + \epsilon_i$ is equivalent to $Y_i^* = \beta x_i + \epsilon_i$. This is a "regression through the origin" model that is studied in Exercise 11.32.

(a) Using the result from Exercise 11.32(a), we have

$$b = \frac{\sum_{i=1}^{n} x_i(y_i - \alpha)}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\alpha}{\sum_{i=1}^{n} x_i^2}.$$

(b) Also from Exercise 11.32(b) we have $\sigma_B^2 = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}$.

11.67 $SSE = \sum_{i=1}^{n}(y_i - \beta x_i)^2$. Taking derivative with respect to β and setting this as 0, we get $\sum_{i=1}^{n} x_i(y_i - bx_i) = 0$, or $\sum_{i=1}^{n} x_i(y_i - \hat{y}_i) = 0$. This is the only equation we can get using the least squares method. Hence in general, $\sum_{i=1}^{n}(y_i - \hat{y}_i) = 0$ does not hold for a regression model with zero intercept.

11.68 No solution is provided.

Chapter 12

Multiple Linear Regression and Certain Nonlinear Regression Models

12.1 (a) $\widehat{y} = 27.5467 + 0.9217x_1 + 0.2842x_2$.

(b) When $x_1 = 60$ and $x_2 = 4$, the predicted value of the chemistry grade is $\hat{y} = 27.5467 + (0.9217)(60) + (0.2842)(4) = 84$.

12.2 $\hat{y} = -3.3727 + 0.0036x_1 + 0.9476x_2$.

12.3 $\hat{y} = 0.7800 + 2.7122x_1 + 2.0497x_2$.

12.4 (a) $\hat{y} = -22.99316 + 1.39567x_1 + 0.21761x_2$.

(b) $\hat{y} = -22.99316 + (1.39567)(35) + (0.21761)(250) = 80.25874$.

12.5 (a) $\hat{y} = 56.46333 + 0.15253x - 0.00008x^2$.

(b) $\hat{y} = 56.46333 + (0.15253)(225) - (0.00008)(225)^2 = 86.73333\%$.

12.6 (a) $\hat{d} = 13.35875 - 0.33944v - 0.01183v^2$.

(b) $\hat{d} = 13.35875 - (-0.33944)(70) - (0.01183)(70)^2 = 47.54206$.

12.7 $\hat{y} = 141.61178 - 0.28193x + 0.00031x^2$.

12.8 (a) $\hat{y} = 19.03333 + 1.0086x - 0.02038x^2$.

(b) $SSE = 24.47619$ with 12 degrees of freedom and $SS(\text{pure error}) = 24.36667$ with 10 degrees of freedom. So, $SSLOF = 24.47619 - 24.36667 = 0.10952$ with 2 degrees of freedom. Hence $f = \frac{0.10952/2}{24.36667/10} = 0.02$ with a P-value of 0.9778. Therefore, there is no lack of fit and the quadratic model fits the data well.

12.9 (a) $\hat{y} = -102.71324 + 0.60537x_1 + 8.92364x_2 + 1.43746x_3 + 0.01361x_4$.

(b) $\hat{y} = -102.71324 + (0.60537)(75) + (8.92364)(24) + (1.43746)(90) + (0.01361)(98) = 287.56183$.

12.10 (a) $\hat{y} = 1.07143 + 4.60317x - 1.84524x^2 + 0.19444x^3$.

(b) $\hat{y} = 1.07143 + (4.60317)(2) - (1.84524)(2)^2 + (0.19444)(2)^3 = 4.45238$.

12.11 $\hat{y} = 3.3205 + 0.42105x_1 - 0.29578x_2 + 0.01638x_3 + 0.12465x_4$.

12.12 $\hat{y} = 1,962.94816 - 15.85168x_1 + 0.05593x_2 + 1.58962x_3 - 4.21867x_4 - 394.31412x_5$.

12.13 $\hat{y} = -6.51221 + 1.99941x_1 - 3.67510x_2 + 2.52449x_3 + 5.15808x_4 + 14.40116x_5$.

12.14 $\hat{y} = -884.667 - 3 - 0.83813x_1 + 4.90661x_2 + 1.33113x_3 + 11.93129x_4$.

12.15 (a) $\hat{y} = 350.99427 - 1.27199x_1 - 0.15390x_2$.

(b) $\hat{y} = 350.99427 - (1.27199)(20) - (0.15390)(1200) = 140.86930$.

12.16 (a) $\hat{y} = -21.46964 - 3.32434x_1 + 0.24649x_2 + 20.34481x_3$.

(b) $\hat{y} = -21.46964 - (3.32434)(14) + (0.24649)(220) + (20.34481)(5) = 87.94123$.

12.17 $s^2 = 0.16508$.

12.18 $s^2 = 0.43161$.

12.19 $s^2 = 242.71561$.

12.20 Using *SAS* output, we obtain

$$\hat{\sigma}^2_{b_1} = 3.747 \times 10^{-7}, \quad \hat{\sigma}^2_{b_2} = 0.13024, \quad \hat{\sigma}_{b_1 b_2} = -4.165 \times 10^{-7}.$$

12.21 Using *SAS* output, we obtain

(a) $\hat{\sigma}^2_{b_2} = 28.09554$.

(b) $\hat{\sigma}_{b_1 b_4} = -0.00958$.

12.22 Using *SAS* output, we obtain
$0.4516 < \mu_{Y|x_1=900,x_2=1} < 1.2083$, and $-0.1640 < y_0 < 1.8239$.

12.23 Using *SAS* output, we obtain a 90% confidence interval for the mean response when $x = 19.5$ as $29.9284 < \mu_{Y|x=19.5} < 31.9729$.

12.24 Using *SAS* output, we obtain
$263.7879 < \mu_{Y|x_1=75,x_2=24,x_3=90,x_4=98} < 311.3357$, and $243.7175 < y_0 < 331.4062$.

12.25 The hypotheses are

$$H_0: \beta_2 = 0,$$
$$H_1: \beta_2 \neq 0.$$

The test statistic value is $t = 2.86$ with a P-value $= 0.0145$. Hence, we reject H_0 and conclude $\beta_2 \neq 0$.

12.26 The test statistic is $t = \frac{0.00362}{0.000612} = 5.91$ with P-value $= 0.0002$. Reject H_0 and claim that $\beta_1 \neq 0$.

12.27 The hypotheses are

$$H_0 : \beta_1 = 2,$$
$$H_1 : \beta_1 \neq 2.$$

The test statistics is $t = \frac{2.71224-2}{0.20209} = 3.524$ with P-value $= 0.0097$. Reject H_0 and conclude that $\beta_1 \neq 2$.

12.28 Using SAS output, we obtain

(a) $s^2 = 650.1408$.

(b) $\hat{y} = 171.6501$, $135.8735 < \mu_{Y|x_1=20,x_2=1000} < 207.4268$, and $82.9677 < y_0 < 260.3326$.

12.29 (a) P-value $= 0.3562$. Hence, fail to reject H_0.

(b) P-value $= 0.1841$. Again, fail to reject H_0.

(c) There is not sufficient evidence that the regressors x_1 and x_2 significantly influence the response with the described linear model.

12.30 (a) $s^2 = 17.22858$.

(b) $\hat{y} = 104.9617$ and $95.5660 < y_0 < 114.3574$.

12.31 $R^2 = \frac{SSR}{SST} = \frac{10953}{10956} = 99.97\%$. Hence, 99.97% of the variation in the response Y in our sample can be explained by the linear model.

12.32 The hypotheses are:

$$H_0 : \beta_1 = \beta_2 = 0,$$
$$H_1 : \text{At least one of the } \beta_i\text{'s is not zero, for } i = 1, 2.$$

Using the f-test, we obtain that $f = \frac{MSR}{MSE} = \frac{5476.60129}{0.43161} = 12688.7$ with P-value $<$ 0.0001. Hence, we reject H_0. The regression explained by the model is significant.

12.33 $f = 5.11$ with P-value $= 0.0303$. At level of 0.01, we fail to reject H_0 and we cannot claim that the regression is significant.

12.34 The hypotheses are:

$$H_0 : \beta_1 = \beta_2 = 0,$$
$$H_1 : \text{At least one of the } \beta_i\text{'s is not zero, for } i = 1, 2.$$

The partial f-test statistic is

$$f = \frac{(160.93598 - 145.88354)/2}{1.49331} = 5.04, \quad \text{with 2 and 7 degrees of freedom.}$$

The resulting P-value $= 0.0441$. Therefore, we reject H_0 and claim that at least one of β_1 and β_2 is not zero.

12.35 $f = \frac{(6.90079-1.13811)/1}{0.16508} = 34.91$ with 1 and 9 degrees of freedom. The P-value $= 0.0002$ which implies that H_0 is rejected.

12.36 (a) $\hat{y} = 0.900 + 0.575x_1 + 0.550x_2 + 1.150x_3$.

(b) For the model in (a), $SSR = 15.645$, $SSE = 1.375$ and $SST = 17.020$. The ANOVA table for all these single-degree-of-freedom components can be displayed as:

Source of Variation	Degrees of Freedom	Mean Square	Computed f	P-value
x_1	1	2.645	7.69	0.0501
x_2	1	2.420	7.04	0.0568
x_3	1	10.580	30.78	0.0052
Error	4	0.34375		
Total	7			

β_3 is found to be significant at the 0.01 level and β_1 and β_2 are not significant.

12.37 The hypotheses are:

$$H_0 : \beta_1 = \beta_2 = 0,$$

$$H_1 : \text{At least one of the } \beta_i\text{'s is not zero, for } i = 1, 2.$$

The partial f-test statistic is

$$f = \frac{(4957.24074 - 17.02338)/2}{242.71561} = 10.18, \quad \text{with 2 and 7 degrees of freedom.}$$

The resulting P-value $= 0.0085$. Therefore, we reject H_0 and claim that at least one of β_1 and β_2 is not zero.

12.38 Using computer software, we obtain the following.

$$\begin{aligned}
R(\beta_1 \mid \beta_0) &= 2.645, \\
R(\beta_1 \mid \beta_0, \beta_2, \beta_3) &= R(\beta_0, \beta_1, \beta_2, \beta_3) - R(\beta_0, \beta_2, \beta_3) = 15.645 - 13.000 = 2.645. \\
R(\beta_2 \mid \beta_0, \beta_1) &= R(\beta_0, \beta_1, \beta_2) - R(\beta_0, \beta_1) = 5.065 - 2.645 = 2.420, \\
R(\beta_2 \mid \beta_0, \beta_1, \beta_3) &= R(\beta_0, \beta_1, \beta_2, \beta_3) - R(\beta_0, \beta_1, \beta_3) = 15.645 - 13.225 = 2.420, \\
R(\beta_3 \mid \beta_0, \beta_1, \beta_2) &= R(\beta_0, \beta_1, \beta_2, \beta_3) - R(\beta_0, \beta_1, \beta_2) = 15.645 - 5.065 = 10.580.
\end{aligned}$$

12.39 The following is the summary.

	s^2	R^2	R^2_{adj}
The model using weight alone	8.12709	0.8155	0.8104
The model using weight and drive ratio	4.78022	0.8945	0.8885

The above favor the model using both explanatory variables. Furthermore, in the model with two independent variables, the t-test for β_2, the coefficient of drive ratio, shows P-value < 0.0001. Hence, the drive ratio variable is important.

12.40 The following is the summary:

	s^2	C.V.	R^2_{adj}	Average Length of the CIs
The model with x_3	4.29738	7.13885	0.8823	5.03528
The model without x_3	4.00063	6.88796	0.8904	4.11769

These numbers favor the model without using x_3. Hence, variable x_3 appears to be unimportant.

12.41 The following is the summary:

	s^2	C.V.	R^2_{adj}
The model with 3 terms	0.41966	4.62867	0.9807
The model without 3 terms	1.60019	9.03847	0.9266

Furthermore, to test $\beta_{11} = \beta_{12} = \beta_{22} = 0$ using the full model, $f = 15.07$ with P-value $= 0.0002$. Hence, the model with interaction and pure quadratic terms is better.

12.42 (a) Full model: $\hat{y} = 121.75 + 2.50x_1 + 14.75x_2 + 21.75x_3$, with $R^2_{\text{adj}} = 0.9714$.
Reduced model: $\hat{y} = 121.75 + 14.75x_2 + 21.75x_3$, with $R^2_{\text{adj}} = 0.9648$.
There appears to be little advantage using the full model.

(b) The average prediction interval widths are:
full model: 32.70; and reduced model: 32.18. Hence, the model without using x_1 is very competitive.

12.43 The following is the summary:

	s^2	C.V.	R^2_{adj}	Average Length of the CIs
x_1, x_2	650.14075	16.55705	0.7696	106.60577
x_1	967.90773	20.20209	0.6571	94.31092
x_2	679.99655	16.93295	0.7591	78.81977

In addition, in the full model when the individual coefficients are tested, we obtain P-value $= 0.3562$ for testing $\beta_1 = 0$ and P-value $= 0.1841$ for testing $\beta_2 = 0$.
In comparing the three models, it appears that the model with x_2 only is slightly better.

12.44 Here is the summary for all four models (including the full model)

	s^2	C.V.	R^2_{adj}
x_1, x_2, x_3	17.22858	3.78513	0.9899
x_1, x_2	297.97747	15.74156	0.8250
x_1, x_3	17.01876	3.76201	0.9900
x_2, x_3	17.07575	3.76830	0.9900

It appears that a two-variable model is very competitive with the full model as long as the model contains x_3.

12.45 (a) $\hat{y} = 5.95931 - 0.00003773$ odometer $+0.33735$ octane -12.62656 van -12.98455 suv.

(b) Since the coefficients of *van* and *suv* are both negative, sedan should have the best gas mileage.

(c) The parameter estimates (standard errors) for *van* and *suv* are -12.63 (1.07) and $-12, 98$ (1.11), respectively. So, the difference between the estimates are smaller than one standard error of each. So, no significant difference in a *van* and an *suv* in terms of gas mileage performance.

12.46 The parameter estimates are given here.

Variable	DF	Estimate	Standar Error	t	P-value
Intercept	1	−206.64625	163.70943	−1.26	0.2249
Income	1	0.00543	0.00274	1.98	0.0649
Family	1	−49.24044	51.95786	−0.95	0.3574
Female	1	236.72576	110.57158	2.14	0.0480

(a) $\hat{y} = -206.64625 + 0.00543\text{Income} - 49.24044\text{Family} + 236/72576\text{Female}$. The company would prefer female customers.

(b) Since the P-value $= 0.0649$ for the coefficient of the "Income," it is at least marginally important. Note that the $R^2 = 0.3075$ which is not very high. Perhaps other variables need to be considered.

12.47 (a) $\widehat{Hang\ Time} = 1.10765 + 0.01370\ \text{LLS} + 0.00429\ \text{Power}$.

(b) $\widehat{Hang\ Time} = 1.10765 + (0.01370)(180) + (0.00429)(260) = 4.6900$.

(c) $4.4502 < \mu_{Hang\ Time\ |\ \text{LLS}=180,\ \text{Power}=260} < 4.9299$.

12.48 (a) For forward selection, variable x_1 is entered first, and no other variables are entered at 0.05 level. Hence the final model is $\hat{y} = -6.33592 + 0.33738x_1$.

(b) For the backward elimination, variable x_3 is eliminated first, then variable x_4 and then variable x_2, all at 0.05 level of significance. Hence only x_1 remains in the model and the final model is the same one as in (a).

(c) For the stepwise regression, after x_1 is entered, no other variables are entered. Hence the final model is still the same one as in (a) and (b).

12.49 Using computer output, with $\alpha = 0.05$, x_4 was removed first, and then x_1. Neither x_2 nor x_3 were removed and the final model is $\hat{y} = 2.18332 + 0.95758x_2 + 3.32533x_3$.

12.50 (a) $\hat{y} = -29.59244 + 0.27872x_1 + 0.06967x_2 + 1.24195x_3 - 0.39554x_4 + 0.22365x_5$.

(b) The variables x_3 and x_5 were entered consecutively and the final model is $\hat{y} = -56.94371 + 1.63447x_3 + 0.24859x_5$.

(c) We have a summary table displayed next.

Model	s^2	PRESS	R^2	$\sum_i \lvert\delta_i\rvert$
x_2x_5	176.735	2949.13	0.7816	151.681
x_1x_5	174.398	3022.18	0.7845	166.223
$x_1x_3x_5$	174.600	3207.34	0.8058	174.819
x_3x_5	194.369	3563.40	0.7598	189.881
$x_3x_4x_5$	192.006	3637.70	0.7865	190.564
$x_2x_4x_5$	196.211	3694.97	0.7818	170.847
$x_2x_3x_5$	186.096	3702.90	0.7931	184.285
x_3x_4	249.165	3803.00	0.6921	192.172
$x_1x_2x_5$	184.446	3956.41	0.7949	189.107
x_5	269.355	3998.77	0.6339	189.373
x_3	257.352	4086.47	0.6502	199.520
$x_2x_3x_4x_5$	197.782	4131.88	0.8045	192.000
x_1	274.853	4558.34	0.6264	202.533
x_2x_3	264.670	4721.55	0.6730	210.853
x_1x_3	226.777	4736.02	0.7198	219.630
$x_1x_3x_4x_5$	188.333	4873.16	0.8138	207.542
x_2	328.434	4998.07	0.5536	217.814
x_4x_5	289.633	5136.91	0.6421	209.232
$x_1x_2x_3x_5$	195.344	5394.56	0.8069	216.934
$x_2x_3x_4$	269.800	5563.87	0.7000	234.565
x_1x_2	297.294	5784.20	0.6327	231.374
$x_1x_4x_5$	192.822	5824.58	0.7856	216.583
$x_1x_3x_4$	240.828	6564.79	0.7322	248.123
x_2x_4	352.781	6902.14	0.5641	248.621
$x_1x_2x_4x_5$	207.477	7675.70	0.7949	249.604
$x_1x_2x_3x_4x_5$	214.602	7691.30	0.8144	257.732
x_1x_4	287.794	7714.86	0.6444	249.221
$x_1x_2x_3$	249.038	7752.69	0.7231	264.324
x_4	613.411	8445.98	0.1663	259.968
$x_1x_2x_3x_4$	266.542	10089.94	0.7365	297.640
$x_1x_2x_4$	317.783	10591.58	0.6466	294.044

(d) It appears that the model with $x_2 =$ LLS and $x_5 =$ Power is the best in terms of PRESS, s^2, and $\sum_i |\delta_i|$.

12.51 (a) $\hat{y} = -587.21085 + 428.43313x$.

(b) $\hat{y} = 1180.00032 - 192.69121x + 35.20945x^2$.

(c) The summary of the two models are given as:

Model	s^2	R^2	PRESS
$\mu_Y = \beta_0 + \beta_1 x$	1,105,054	0.8378	18,811,057.08
$\mu_Y = \beta_0 + \beta_1 x + \beta_{11} x^2$	430,712	0.9421	8,706,973.57

It appears that the model with a quadratic term is preferable.

12.52 The parameter estimate for β_4 is 0.22365 with a standard error of 0.13052. Hence, $t = 1.71$ with P-value $= 0.6117$. Fail to reject H_0.

12.53 $\hat{\sigma}^2_{b_1} = 20,588.038$, $\hat{\sigma}^2_{b_{11}} = 62.650$, and $\hat{\sigma}_{b_1 b_{11}} = -1,103.423$.

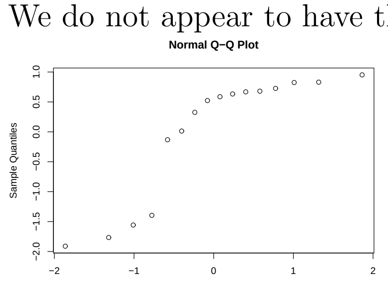
12.54 (a) The following is the summary of the models.

Model	s^2	R^2	PRESS	C_p
x_2x_3	8094.15	0.51235	282194.34	2.0337
x_2	8240.05	0.48702	282275.98	1.5422
x_1x_2	8392.51	0.49438	289650.65	3.1039
$x_1x_2x_3$	8363.55	0.51292	294620.94	4.0000
x_3	8584.27	0.46559	297242.74	2.8181
x_1	8727.47	0.45667	304663.57	3.3489
x_1x_3	8632.45	0.47992	306820.37	3.9645

(b) The model with $\ln(x_2)$ appears to have the smallest C_p with a small PRESS. Also, the model $\ln(x_2)$ and $\ln(x_3)$ has the smallest PRESS. Both models appear to better than the full model.

12.55 (a) There are many models here so the model summary is not displayed. By using MSE criterion, the best model, contains variables x_1 and x_3 with $s^2 = 313.491$. If PRESS criterion is used, the best model contains only the constant term with $s^2 = 317.51$. When the C_p method is used, the best model is model with the constant term.

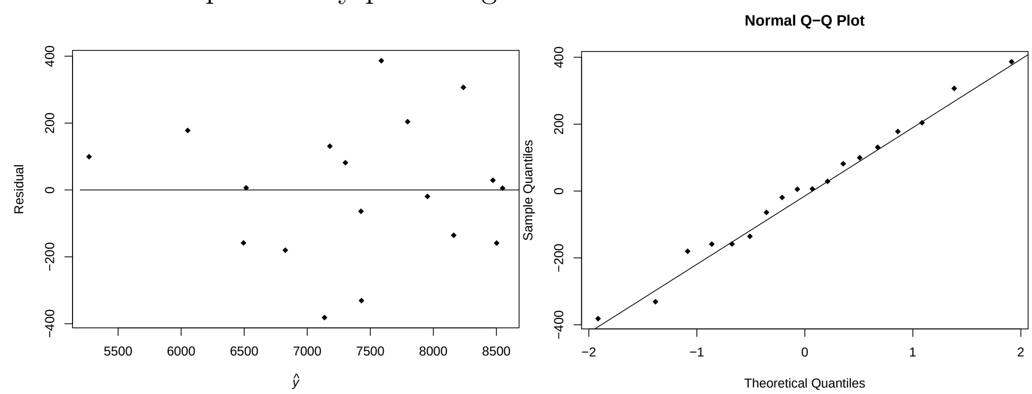
(b) The normal probability plot, for the model using intercept only, is shown next. We do not appear to have the normality.



12.56 (a) $\widehat{\text{Volt}} = -1.64129 + 0.000556 \text{ Speed } - 67.39589 \text{ Extension}$.

(b) P-values for the t-tests of the coefficients are all < 0.0001.

(c) The $R^2 = 0.9607$ and the model appears to have a good fit. The residual plot and a normal probability plot are given here.



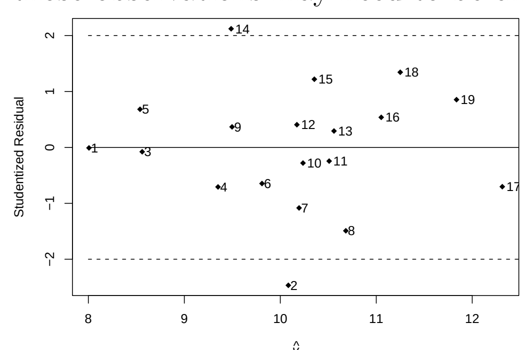
12.57 (a) $\hat{y} = 3.13682 + 0.64443x_1 - 0.01042x_2 + 0.50465x_3 - 0.11967x_4 - 2.46177x_5 + 1.50441x_6$.

(b) The final model using the stepwise regression is

$$\hat{y} = 4.65631 + 0.51133x_3 - 0.12418x_4.$$

(c) Using C_p criterion (smaller the better), the best model is still the model stated in (b) with $s^2 = 0.73173$ and $R^2 = 0.64758$. Using the s^2 criterion, the model with x_1, x_3 and x_4 has the smallest value of 0.72507 and $R^2 = 0.67262$. These two models are quite competitive. However, the model with two variables has one less variable, and thus may be more appealing.

(d) Using the model in part (b), displayed next is the Studentized residual plot. Note that observations 2 and 14 are beyond the 2 standard deviation lines. Both of those observations may need to be checked.



12.58 The partial F-test shows a value of 0.82, with 2 and 12 degrees of freedom. Consequently, the P-value $= 0.4622$, which indicates that variables x_1 and x_6 can be excluded from the model.

12.59 (a) $\hat{y} = 125.86555 + 7.75864x_1 + 0.09430x_2 - 0.00919x_1x_2$.

(b) The following is the summary of the models.

Model	s^2	R^2	PRESS	C_p
x_2	680.00	0.80726	7624.66	2.8460
x_1	967.91	0.72565	12310.33	4.8978
x_1x_2	650.14	0.86179	12696.66	3.4749
$x_1x_2x_3$	561.28	0.92045	15556.11	4.0000

It appears that the model with x_2 alone is the best.

12.60 (a) The fitted model is $\hat{y} = 85.75037 - 15.93334x_1 + 2.42280x_2 + 1.82754x_3 + 3.07379x_4$.

(b) The summary of the models are given next.

Model	s^2	PRESS	R^2	C_p
$x_1x_2x_4$	9148.76	447,884.34	0.9603	3.308
x_4	19170.97	453,304.54	0.8890	8.831
x_3x_4	21745.08	474,992.22	0.8899	10.719
$x_1x_2x_3x_4$	10341.20	482,210.53	0.9626	5.000
x_1x_4	10578.94	488,928.91	0.9464	3.161
x_2x_4	21630.42	512,749.78	0.8905	10.642
$x_2x_3x_4$	25160.18	532,065.42	0.8908	12.598
$x_1x_3x_4$	12341.87	614,553.42	0.9464	5.161
x_2	160756.81	1,658,507.38	0.0695	118.362
x_3	171264.68	1,888,447.43	0.0087	126.491
x_2x_3	183701.86	1,896,221.30	0.0696	120.349
x_1	95574.16	2,213,985.42	0.4468	67.937
x_1x_3	107287.63	2,261,725.49	0.4566	68.623
x_1x_2	109137.20	2,456,103.03	0.4473	69.875
$x_1x_2x_3$	125126.59	2,744,659.14	0.4568	70.599

When using PRESS as well as the s^2 criterion, a model with x_1, x_2 and x_4 appears to be the best, while when using the C_p criterion, the model with x_1 and x_4 is the best. When using the model with x_1, x_2 and x_4, we find out that the P-value for testing $\beta_2 = 0$ is 0.1980 which implies that perhaps x_2 can be excluded from the model.

(c) The model in part (b) has smaller C_p as well as competitive $PRESS$ in comparison to the full model.

12.61 Since $H = X(X`X)^{-1}X`$, and $\sum\limits_{i=1}^{n} h_i i = \text{tr}(H)$, we have

$$\sum_{i=1}^{n} h_{ii} = \text{tr}(X(X`X)^{-1}X`) = \text{tr}(X`X(X`X)^{-1}) = \text{tr}(I_p) = p,$$

where I_p is the $p \times p$ identity matrix. Here we use the property of $\text{tr}(AB) = \text{tr}(BA)$ in linear algebra.

12.62 (a) $\hat{y} = 9.9375 + 0.6125x_1 + 1.3125x_2 + 1.4625x_3$.

(b) The ANOVA table for all these single-degree-of-freedom components can be displayed as:

Source of Variation	Degrees of Freedom	Mean Square	Computed f	P-value
x_1	1	3.00125	0.95	0.3847
x_2	1	13.78125	4.37	0.1049
x_3	1	17.11125	5.42	0.0804
Error	4	3.15625		
Total	7			

Only x_3 is near significant.

12.63 (a) For the completed second-order model, we have

$$\text{PRESS} = 9,657,641.55, \qquad \sum_{i=1}^{n} |y_i - \hat{y}_{i,-i}| = 5,211.37.$$

(b) When the model does not include any term involving x_4,

$$\text{PRESS} = 6,954.49, \qquad \sum_{i=1}^{n} |y_i - \hat{y}_{i,-i}| = 277.292.$$

Apparently, the model without x_4 is much better.

(c) For the model with x_4:

$$\text{PRESS} = 312,963.71, \qquad \sum_{i=1}^{n} |y_i - \hat{y}_{i,-i}| = 762.57.$$

For the model without x_4:

$$\text{PRESS} = 3,879.89, \qquad \sum_{i=1}^{n} |y_i - \hat{y}_{i,-i}| = 220.12$$

Once again, the model without x_4 performs better in terms of $PRESS$ and $\sum_{i=1}^{n} |y_i - \hat{y}_{i,-i}|$.

12.64 (a) The stepwise regression results in the following fitted model:

$$\hat{y} = 102.20428 - 0.21962x_1 - 0.0723 - x_2 - 2.68252x_3 - 0.37340x_5 + 0.30491x_6.$$

(b) Using the C_p criterion, the best model is the same as the one in (a).

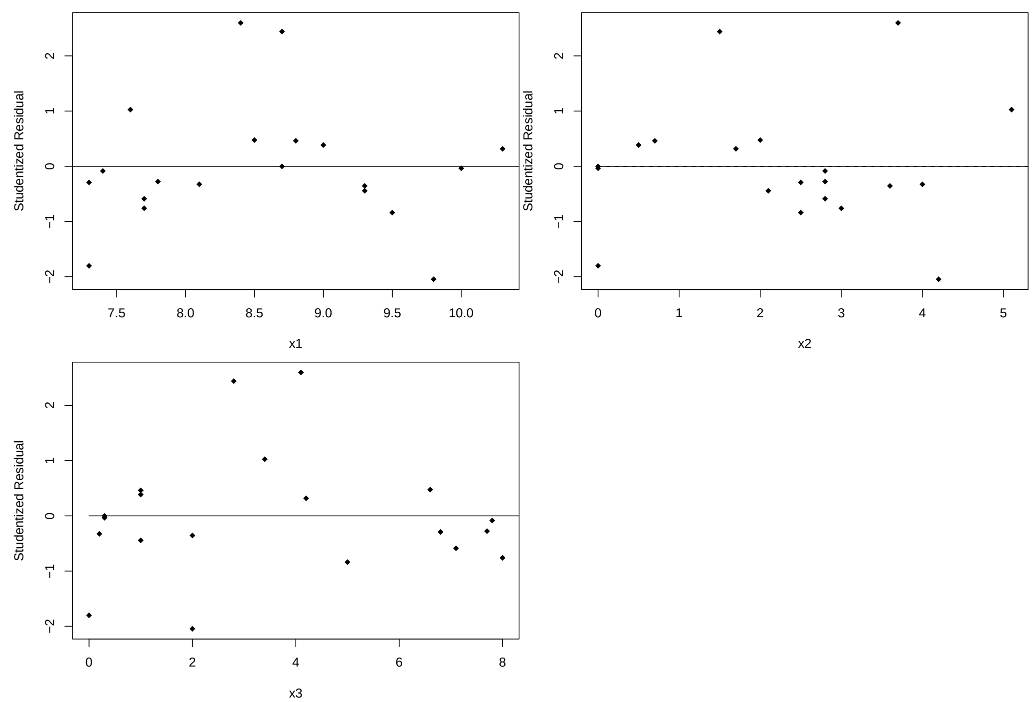
12.65 (a) Yes. The orthogonality is maintained with the use of interaction terms.

(b) No. There are no degrees of freedom available for computing the standard error.

12.66 The fitted model is $\hat{y} = 26.19333 + 0.04772x_1 + 0.76011x_2 - 0.00001334x_{11} - 0.00687x_{22} + 0.00011333x_{12}$. The t-tests for each coefficient show that x_{12} and x_{22} may be eliminated. So, we ran a test for $\beta_{12} = \beta_{22} = 0$ which yields P-value $= 0.2222$. Therefore, both x_{12} and x_{22} may be dropped out from the model.

12.67 (a) The fitted model is $\hat{y} = -0.26891 + 0.07755x_1 + 0.02532x_2 - 0.03575x_3$. The f-value of the test for $H_0: \beta_1 = \beta_2 = \beta_3 = 0$ is 35,28 with P-value < 0.0001. Hence, we reject H_0.

(b) The residual plots are shown below and they all display random residuals.



(c) The following is the summary of these three models.

Model	PRESS	C_p
x_1, x_2, x_3	0.091748	24.7365
$x_1, x_2, x_3, x_1^2, x_2^2, x_3^2$	0.08446	12.3997
$x_1, x_2, x_3, x_1^2, x_2^2, x_3^2, x_{12}, x_{13}, x_{23}$	0.088065	10

It is difficult to decide which of the three models is the best. Model I contains all the significant variables while models II and III contain insignificant variables. However, the C_p value and $PRESS$ for model are not so satisfactory. Therefore, some other models may be explored.

12.68 Denote by $Z_1 = 1$ when Group=1, and $Z_1 = 0$ otherwise;
Denote by $Z_2 = 1$ when Group=2, and $Z_2 = 0$ otherwise;
Denote by $Z_3 = 1$ when Group=3, and $Z_3 = 0$ otherwise;

(a) The parameter estimates are:

Variable	DF	Parameter Estimate	P-value
Intercept	1	46.34694	0.0525
BMI	1	−1.79090	0.0515
z_1	1	−23.84705	0.0018
z_2	1	−17.46248	0.0109

Yes, Group I has a mean change in blood pressure that was significantly lower than the control group. It is about 23.85 points lower.

(b) The parameter estimates are:

Variable	DF	Parameter Estimate	P-value
Intercept	1	28.88446	0.1732
BMI	1	−1.79090	0.0515
z_1	1	−6.38457	0.2660
z_3	1	17.46248	0.0109

Although Group I has a mean change in blood pressure that was 6.38 points lower than that of Group II, the difference is not very significant due to a high P-value.

12.69 (b) All possible regressions should be run. $R^2 = 0.9908$ and there is only one significant variable.

(c) The model including x_2, x_3 and x_5 is the best in terms of C_p, PRESS and has all variables significant.

12.70 Using the formula of R^2_{adj} on page 467, we have

$$R^2_{\text{adj}} = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)} = 1 - \frac{MSE}{MST}.$$

Since MST is fixed, maximizing R^2_{adj} is thus equivalent to minimizing MSE.

12.71 (a) The fitted model is $\hat{p} = \frac{1}{1+e^{2.7620-0.0308x}}$.

(b) The χ^2-values for testing $b_0 = 0$ and $b_1 = 0$ are 471.4872 and 243.4111, respectively. Their corresponding P-values are < 0.0001. Hence, both coefficients are significant.

(c) $\text{ED}_{50} = -\frac{-2.7620}{0.0308} = 89.675$.

12.72 (a) The fitted model is $\hat{p} = \frac{1}{1+e^{2.9949-0.0308x}}$.

(b) The increase in odds of failure that results by increasing the load by 20 lb/in.^{2} is $e^{(20)(0.0308)} = 1.8515$.

Chapter 13

One-Factor Experiments: General

13.1 Using the formula of SSE, we have

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n}(\epsilon_{ij} - \bar{\epsilon}_{i.})^2 = \sum_{i=1}^{k}\left[\sum_{j=1}^{n}\epsilon_{ij}^2 - n\bar{\epsilon}_{i.}^2\right].$$

Hence

$$E(SSE) = \sum_{i=1}^{k}\left[\sum_{j=1}^{n}E(\epsilon_{ij}^2) - nE(\bar{\epsilon}_{i.}^2)\right] = \sum_{i=1}^{k}\left[n\sigma^2 - n\frac{\sigma^2}{n}\right] = k(n-1)\sigma^2.$$

Thus $E\left[\frac{SSE}{k(n-1)}\right] = \frac{k(n-1)\sigma^2}{k(n-1)} = \sigma^2$.

13.2 Since $SSA = n\sum\limits_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..})^2 = n\sum\limits_{i=1}^{k}\bar{y}_{i.}^2 - kn\bar{y}_{..}^2$, $y_{ij} \sim n(y;\mu+\alpha_i,\sigma^2)$, and hence $\bar{y}_{i.} \sim n\left(y;\mu+\alpha_i,\frac{\sigma}{\sqrt{n}}\right)$ and $\bar{y}_{..} \sim n\left(\mu+\bar{\alpha},\frac{\sigma}{\sqrt{kn}}\right)$, then

$$E(\bar{y}_{i.}^2) = Var(\bar{y}_{i.}) + [E(\bar{y}_{i.})]^2 = \frac{\sigma^2}{n} + (\mu+\alpha_i)^2,$$

and

$$E[\bar{y}_{..}^2] = \frac{\sigma^2}{kn} + (\mu+\bar{\alpha})^2 = \frac{\sigma^2}{kn} + \mu^2,$$

due to the constraint on the α's. Therefore,

$$\begin{aligned} E(SSA) &= n\sum_{i=1}^{k}E(\bar{y}_{i.}^2) - knE(\bar{y}_{..}^2) = k\sigma^2 + n\sum_{i=1}^{k}(\mu+\alpha_i)^2 - (\sigma^2 + kn\mu^2) \\ &= (k-1)\sigma^2 + n\sum_{i=1}^{k}\alpha_i^2. \end{aligned}$$

13.3 The hypotheses are

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_6,$$
$$H_1 : \text{At least two of the means are not equal.}$$

$\alpha = 0.05$.
Critical region: $f > 2.77$ with $v_1 = 5$ and $v_2 = 18$ degrees of freedom.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Treatment	5.34	5	1.07	0.31
Error	62.64	18	3.48	
Total	67.98	23		

with P-value=0.9024.
Decision: The treatment means do not differ significantly.

13.4 The hypotheses are

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_5,$$
$$H_1 : \text{At least two of the means are not equal.}$$

$\alpha = 0.05$.
Critical region: $f > 2.87$ with $v_1 = 4$ and $v_2 = 20$ degrees of freedom.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Tablets	78.422	4	19.605	6.59
Error	59.532	20	2.977	
Total	137.954	24		

with P-value=0.0015.
Decision: Reject H_0. The mean number of hours of relief differ significantly.

13.5 The hypotheses are

$$H_0 : \mu_1 = \mu_2 = \mu_3,$$
$$H_1 : \text{At least two of the means are not equal.}$$

$\alpha = 0.01$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Shelf Height	399.3	2	199.63	14.52
Error	288.8	21	13.75	
Total	688.0	23		

with P-value=0.0001.
Decision: Reject H_0. The amount of money spent on dog food differs with the shelf height of the display.

13.6 The hypotheses are

$$H_0 : \mu_A = \mu_B = \mu_C,$$
$$H_1 : \text{At least two of the means are not equal.}$$

$\alpha = 0.01$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Drugs	158.867	2	79.433	5.46
Error	393.000	27	14.556	
Total	551.867	29		

with P-value=0.0102.
Decision: Since $\alpha = 0.01$, we fail to reject H_0. However, this decision is very marginal since the P-value is very close to the significance level.

13.7 The hypotheses are

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4,$$
$$H_1 : \text{At least two of the means are not equal.}$$

$\alpha = 0.05$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Treatments	119.787	3	39.929	2.25
Error	638.248	36	17.729	
Total	758.035	39		

with P-value=0.0989.
Decision: Fail to reject H_0 at level $\alpha = 0.05$.

13.8 The hypotheses are

$$H_0 : \mu_1 = \mu_2 = \mu_3,$$
$$H_1 : \text{At least two of the means are not equal.}$$

Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Solvents	3.3054	2	1.6527	24.51
Error	1.9553	29	0.0674	
Total	5.2608	31		

with P-value< 0.0001.
Decision: There is significant difference in the mean sorption rate for the three solvents. The mean sorption for the solvent Chloroalkanes is the highest. We know that it is significantly higher than the rate of Esters.

13.9 The hypotheses are

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4,$$
$$H_1 : \text{At least two of the means are not equal.}$$

$\alpha = 0.01$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Treatments	27.5506	3	9.1835	8.38
Error	18.6360	17	1.0962	
Total	46.1865	20		

with P-value$= 0.0012$.
Decision: Reject H_0. Average specific activities differ.

13.10 $s_{50} = 3.2098, s_{100} = 4.5253, s_{200} = 5.1788$, and $s_{400} = 3.6490$. Since the sample sizes are all the same,

$$s_p^2 = \frac{1}{4}\sum_{i=1}^{4} s_i^2 = 17.7291.$$

Therefore, the Bartlett's statistic is

$$b = \frac{\left(\prod_{i=1}^{4} s_i^2\right)^{1/4}}{s_p^2} = 0.9335.$$

Using Table A.10, the critical value of the Bartlett's test with $k = 4$ and $\alpha = 0.05$ is 0.7970. Since $b > 0.7970$, we fail to reject H_0 and hence the variances can be assumed equal.

13.11 Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
B vs. A, C, D	30.6735	1	30.6735	14.28
C vs. A, D	49.9230	1	49.9230	23.23
A vs. D	5.3290	1	5.3290	2.48
Error	34.3800	16	2.1488	

(a) P-value=0.0016. B is significantly different from the average of A, C, and D.

(b) P-value=0.0002. C is significantly different from the average of A and D.

(c) P-value=0.1349. A can not be shown to differ significantly from D.

13.12 (a) The hypotheses are

$$H_0 : \mu_{29} = \mu_{54} = \mu_{84},$$
$$H_1 : \text{At least two of the means are not equal.}$$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Protein Levels	32,974.87	2	16,487.43	5.15
Error	28,815.80	9	3,201.76	
Total	61,790.67	11		

with P-value= 0.0323.
Decision: Reject H_0. The mean nitrogen loss was significantly different for the three protein levels.

(b) For testing the contrast $L = 2\mu_{29} - \mu_{54} - \mu_{84}$ at level $\alpha = 0.05$, we have $SSw = 31,576.42$ and $f = 9.86$, with P-value=0.0119. Hence, the mean nitrogen loss for 29 grams of protein was different from the average of the two higher protein levels.

13.13 (a) The hypotheses are

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4,$$
$$H_1 : \text{At least two of the means are not equal.}$$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Treatments	1083.60	3	361.20	13.50
Error	1177.68	44	26.77	
Total	2261.28	47		

with P-value< 0.0001.
Decision: Reject H_0. The treatment means are different.

(b) For testing two contrasts $L_1 = \mu_1 - \mu_2$ and $L_2 = \mu_3 - \mu_4$ at level $\alpha = 0.01$, we have the following

Contrast	Sum of Squares	Computed f	P-value
1 vs. 2	785.47	29.35	< 0.0001
3 vs. 4	96.00	3.59	0.0648

Hence, Bath I and Bath II were significantly different for 5 launderings, and Bath I and Bath II were not different for 10 launderings.

13.14 The means of the treatments are:

$$\bar{y}_{1.} = 5.44, \bar{y}_{2.} = 7.90, \bar{y}_{3.} = 4.30, \bar{y}_{4.} = 2.98, \text{ and } \bar{y}_{5.} = 6.96.$$

Since $q(0.05, 5, 20) = 4.24$, the critical difference is $(4.24)\sqrt{\frac{2.9766}{5}} = 3.27$. Therefore, the Tukey's result may be summarized as follows:

$\bar{y}_{5.}$	$\bar{y}_{3.}$	$\bar{y}_{1.}$	$\bar{y}_{5.}$	$\bar{y}_{2.}$
2.98	4.30	5.44	6.96	7.90

13.15 Since $q(0.05, 4, 16) = 4.05$, the critical difference is $(4.05)\sqrt{\frac{2.14875}{5}} = 2.655$. Hence

$\bar{y}_{3.}$	$\bar{y}_{1.}$	$\bar{y}_{4.}$	$\bar{y}_{2.}$
56.52	59.66	61.12	61.96

13.16 (a) The hypotheses are

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4,$$
$$H_1 : \text{At least two of the means are not equal.}$$

$\alpha = 0.05$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Blends	119.649	3	39.883	7.10
Error	44.920	8	5.615	
Total	164.569	11		

with P-value$= 0.0121$.
Decision: Reject H_0. There is a significant difference in mean yield reduction for the 4 preselected blends.

(b) Since $\sqrt{s^2/3} = 1.368$ we get

p	2	3	4
r_p	3.261	3.399	3.475
R_p	4.46	4.65	4.75

Therefore,

$\bar{y}_{3.}$	$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\bar{y}_{4.}$
23.23	25.93	26.17	31.90

(c) Since $q(0.05, 4, 8) = 4.53$, the critical difference is 6.20. Hence

$\bar{y}_{3.}$	$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\bar{y}_{4.}$
23.23	25.93	26.17	31.90

13.17 (a) The hypotheses are

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_5,$$
$$H_1: \text{At least two of the means are not equal.}$$

Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Procedures	7828.30	4	1957.08	9.01
Error	3256.50	15	217.10	
Total	11084.80	19		

with P-value$= 0.0006$.
Decision: Reject H_0. There is a significant difference in the average species count for the different procedures.

(b) Since $q(0.05, 5, 15) = 4.373$ and $\sqrt{\frac{217.10}{4}} = 7.367$, the critical difference is 32.2. Hence

$\bar{y}_K$	$\bar{y}_S$	$\bar{y}_{Sub}$	$\bar{y}_M$	$\bar{y}_D$
12.50	24.25	26.50	55.50	64.25

13.18 The hypotheses are

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_5,$$
$$H_1: \text{At least two of the means are not equal.}$$

Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Angles	99.024	4	24.756	21.40
Error	23.136	20	1.157	
Total	122.160	24		

with P-value< 0.0001.
Decision: Reject H_0. There is a significant difference in mean pressure for the different angles.

13.19 When we obtain the ANOVA table, we derive $s^2 = 0.2174$. Hence

$$\sqrt{2s^2/n} = \sqrt{(2)(0.2174)/5} = 0.2949.$$

The sample means for each treatment levels are

$$\bar{y}_C = 6.88, \quad \bar{y}_{1.} = 8.82, \quad \bar{y}_{2.} = 8.16, \quad \bar{y}_{3.} = 6.82, \quad \bar{y}_{4.} = 6.14.$$

Hence

$$d_1 = \frac{8.82 - 6.88}{0.2949} = 6.579, \quad d_2 = \frac{8.16 - 6.88}{0.2949} = 4.340,$$
$$d_3 = \frac{6.82 - 6.88}{0.2949} = -0.203, \quad d_4 = \frac{6.14 - 6.88}{0.2949} = -2.509.$$

From Table A.14, we have $d_{0.025}(4, 20) = 2.65$. Therefore, concentrations 1 and 2 are significantly different from the control.

13.20 The ANOVA table can be obtained as follows:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Cables	1924.296	8	240.537	9.07
Error	2626.917	99	26.535	
Total	4551.213	107		

with P-value< 0.0001.
The results from Tukey's procedure can be obtained as follows:

$\bar{y}_{2.}$	$\bar{y}_{3.}$	$\bar{y}_{1.}$	$\bar{y}_{4.}$	$\bar{y}_{6.}$	$\bar{y}_{7.}$	$\bar{y}_{5.}$	$\bar{y}_{8.}$	$\bar{y}_{9.}$
-7.000	-6.083	-4.083	-2.667	0.833	0.917	1.917	3.333	6.250

The cables are significantly different:
9 is different from 4, 1, 2, 3
8 is different from 1, 3, 2
5, 7, 6 are different from 3, 2.

13.21 Aggregate 4 has a significantly lower absorption rate than the other aggregates.

13.22 (a) The hypotheses are

$$H_0: \mu_C = \mu_L = \mu_M = \mu_H,$$
$$H_1: \text{At least two of the means are not equal.}$$

Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Finance Leverages	80.7683	3	26.9228	5.34
Error	100.8700	20	5.0435	
Total	181.6383	23		

with P-value= 0.0073.
Decision: Reject H_0. The means are not all equal for the different financial leverages.

(b) $\sqrt{2s^2/n} = \sqrt{(2)(5.0435)/6} = 1.2966$. The sample means for each treatment levels are

$$\bar{y}_C = 4.3833, \quad \bar{y}_L = 5.1000, \quad \bar{y}_H = 8.3333, \quad \bar{y}_M = 8.4167.$$

Hence

$$d_L = \frac{5.1000 - 4.3833}{1.2966} = 0.5528, \quad d_M = \frac{8.4167 - 4.3833}{1.2966} = 3.1108,$$
$$d_L = \frac{8.333 - 4.3833}{1.2966} = 3.0464.$$

From Table A.14, we have $d_{0.025}(3, 20) = 2.54$. Therefore, the mean rate of return are significantly higher for median and high financial leverage than for control.

13.23 The ANOVA table can be obtained as follows:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Temperatures	1268.5333	4	317.1333	70.27
Error	112.8333	25	4.5133	
Total	1381.3667	29		

with P-value< 0.0001.
The results from Tukey's procedure can be obtained as follows:

$\bar{y}_0$	$\bar{y}_{25}$	$\bar{y}_{100}$	$\bar{y}_{75}$	$\bar{y}_{50}$
55.167	60.167	64.167	70.500	72.833

The batteries activated at temperature 50 and 75 have significantly longer activated life.

13.24 The Duncan's procedure shows the following results:

$$\begin{array}{ccc} \bar{y}_E & \bar{y}_A & \bar{y}_C \\ \underline{0.3300} & \underline{0.9422 \quad 1.0063} & \end{array}$$

Hence, the sorption rate using the Esters is significantly lower than the sorption rate using the Aromatics or the Chloroalkanes.

13.25 Based on the definition, we have the following.

$$SSB = k\sum_{j=1}^{b}(\bar{y}_{.j}-\bar{y}_{..})^2 = k\sum_{j=1}^{b}\left(\frac{T_{.j}}{k}-\frac{T_{..}}{bk}\right)^2 = \sum_{j=1}^{b}\frac{T_{.j}^2}{k} - 2\frac{T_{..}^2}{bk}+\frac{T_{..}^2}{bk} = \sum_{j=1}^{b}\frac{T_{.j}^2}{k}-\frac{T_{..}^2}{bk}.$$

13.26 From the model

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij},$$

and the constraints

$$\sum_{i=1}^{k}\alpha_i = 0 \quad \text{and} \quad \sum_{j=1}^{b}\beta_j = 0,$$

we obtain

$$\bar{y}_{.j} = \mu + \beta_j + \bar{\epsilon}_{.j} \quad \text{and} \quad \bar{y}_{..} = \mu + \bar{\epsilon}_{..}.$$

Hence

$$SSB = k\sum_{j=1}^{b}(\bar{y}_{.j}-\bar{y}_{..})^2 = k\sum_{j=1}^{b}(\beta_j+\bar{\epsilon}_{.j}-\bar{\epsilon}_{..})^2.$$

Since $E(\bar{\epsilon}_{.j}) = 0$ and $E(\bar{\epsilon}_{..}) = 0$, we obtain

$$E(SSB) = k\sum_{j=1}^{b}\beta_j^2 + k\sum_{j=1}^{b}E(\bar{\epsilon}_{.j}^2) - bkE(\bar{\epsilon}_{..}^2).$$

We know that $E(\bar{\epsilon}_{.j}^2) = \frac{\sigma^2}{k}$ and $E(\bar{\epsilon}_{..}^2) = \frac{\sigma^2}{bk}$. Then

$$E(SSB) = k\sum_{j=1}^{b}\beta_j^2 + b\sigma^2 - \sigma^2 = (b-1)\sigma^2 + k\sum_{j=1}^{b}\beta_j^2.$$

13.27 (a) The hypotheses are

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0, \text{ fertilizer effects are zero}$$

$$H_1: \text{At least one of the } \alpha_i\text{'s is not equal to zero.}$$

$\alpha = 0.05$.
Critical region: $f > 4.76$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Fertilizers	218.1933	3	72.7311	6.11
Blocks	197.6317	2	98.8158	
Error	71.4017	6	11.9003	
Total	487.2267	11		

P-value= 0.0296. Decision: Reject H_0. The means are not all equal.

(b) The results of testing the contrasts are shown as:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
$(f_1,\ f_3)$ vs $(f_2.\ f_4)$	206.6700	1	206.6700	17.37
f_1 vs f_3	11.4817	1	11.4817	0.96
Error	71.4017	6	11.9003	

The corresponding P-values for the above contrast tests are 0.0059 and 0.3639, respectively. Hence, for the first contrast, the test is significant and the for the second contrast, the test is insignificant.

13.28 The hypotheses are

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0, \text{ no differences in the varieties}$$

$$H_1 : \text{At least one of the } \alpha_i\text{'s is not equal to zero.}$$

$\alpha = 0.05$.
Critical region: $f > 5.14$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Treatments	24.500	2	12.250	1.74
Blocks	171.333	3	57.111	
Error	42.167	6	7.028	
Total	238.000	11		

P-value=0.2535. Decision: Do not reject H_0; could not show that the varieties of potatoes differ in field.

13.29 The hypotheses are

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0, \text{ brand effects are zero}$$

$$H_1 : \text{At least one of the } \alpha_i\text{'s is not equal to zero.}$$

$\alpha = 0.05$.
Critical region: $f > 3.84$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Treatments	27.797	2	13.899	5.99
Blocks	16.536	4	4.134	
Error	18.556	8	2.320	
Total	62.889	14		

P-value=0.0257. Decision: Reject H_0; mean percent of foreign additives is not the same for all three brand of jam. The means are:

$$\text{Jam } A\text{: } 2.36, \qquad \text{Jam } B\text{: } 3.48, \qquad \text{Jam } C\text{: } 5.64.$$

Based on the means, Jam A appears to have the smallest amount of foreign additives.

13.30 The hypotheses are

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0, \text{ courses are equal difficulty}$$

$$H_1 : \text{At least one of the } \alpha_i\text{'s is not equal to zero.}$$

$\alpha = 0.05$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Subjects	42.150	3	14.050	0.15
Students	1618.700	4	404.675	
Error	1112.100	12	92.675	
Total	2772.950	19		

P-value=0.9267. Decision: Fail to reject H_0; there is no significant evidence to conclude that courses are of different difficulty.

13.31 The hypotheses are

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_6 = 0, \text{ station effects are zero}$$

$$H_1 : \text{At least one of the } \alpha_i\text{'s is not equal to zero.}$$

$\alpha = 0.01$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Stations	230.127	5	46.025	26.14
Dates	3.259	5	0.652	
Error	44.018	25	1.761	
Total	277.405	35		

P-value< 0.0001. Decision: Reject H_0; the mean concentration is different at the different stations.

13.32 The hypotheses are

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0, \text{ station effects are zero}$$

$$H_1 : \text{At least one of the } \alpha_i\text{'s is not equal to zero.}$$

$\alpha = 0.05$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Stations	10.115	2	5.057	0.15
Months	537.030	11	48.821	
Error	744.416	22	33.837	
Total	1291.561	35		

P-value$= 0.8620$. Decision: Do not reject H_0; the treatment means do not differ significantly.

13.33 The hypotheses are

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0, \text{ diet effects are zero}$$

$$H_1 : \text{At least one of the } \alpha_i\text{'s is not equal to zero.}$$

$\alpha = 0.01$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Diets	4297.000	2	2148.500	11.86
Subjects	6033.333	5	1206.667	
Error	1811.667	10	181.167	
Total	12142.000	17		

P-value$= 0.0023$. Decision: Reject H_0; differences among the diets are significant.

13.34 The hypotheses are

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0, \text{ analyst effects are zero}$$

$$H_1 : \text{At least one of the } \alpha_i\text{'s is not equal to zero.}$$

$\alpha = 0.05$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Analysts	0.001400	2	0.000700	3.00
Individuals	0.021225	3	0.007075	
Error	0.001400	6	0.000233	
Total	0.024025	11		

P-value= 0.1250. Decision: Do not reject H_0; cannot show that the analysts differ significantly.

13.35 The hypotheses are

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0, \text{ treatment effects are zero}$$

$$H_1 : \text{At least one of the } \alpha_i\text{'s is not equal to zero.}$$

$\alpha = 0.01$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Treatments	79630.133	4	19907.533	0.58
Locations	634334.667	5	126866.933	
Error	689106.667	20	34455.333	
Total	1403071.467	29		

P-value= 0.6821. Decision: Do not reject H_0; the treatment means do not differ significantly.

13.36 The hypotheses are

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0, \text{ treatment effects are zero}$$

$$H_1 : \text{At least one of the } \alpha_i\text{'s is not equal to zero.}$$

$\alpha = 0.01$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Treatments	203.2792	2	101.6396	8.60
Subjects	188.2271	9	20.9141	
Error	212.8042	18	11.8225	
Total	604.3104	29		

P-value= 0.0024. Decision: Reject H_0; the mean weight losses are different for different treatments and the therapists had the greatest effect on the weight loss.

13.37 The total sums of squares can be written as

$$\begin{aligned}\sum_i\sum_j\sum_k(y_{ijk}-\bar{y}_{...})^2 =&\sum_i\sum_j\sum_k[(\bar{y}_{i..}-\bar{y}_{...})+(\bar{y}_{.j.}-\bar{y}_{...})+(\bar{y}_{..k}-\bar{y}_{...})\\&+(y_{ijk}-\bar{y}_{i..}-\bar{y}_{.j.}-\bar{y}_{..k}+2\bar{y}_{...})]^2\\=&r\sum_i(\bar{y}_{i..}-\bar{y}_{...})^2+r\sum_j(\bar{y}_{.j.}-\bar{y}_{...})^2+r\sum_k(\bar{y}_{..k}-\bar{y}_{...})^2\\&+\sum_i\sum_j\sum_k(y_{ijk}-\bar{y}_{i..}-\bar{y}_{.j.}-\bar{y}_{..k}+2\bar{y}_{...})^2\\&+6\text{ cross-product terms,}\end{aligned}$$

and all cross-product terms are equal to zeroes.

13.38 For the model $y_{ijk}=\mu+\alpha_i+\beta_j+\tau_k+\epsilon_{ijk}$, we have

$$\bar{y}_{..k}=\mu+\tau_k+\bar{\epsilon}_{..k},\quad\text{and}\quad\bar{y}_{...}=\mu+\bar{\epsilon}_{...}.$$

Hence $SSTr=r\sum_k(\bar{y}_{..k}-\bar{y}_{...})^2=r\sum_k(\tau_k+\bar{\epsilon}_{..k}-\bar{\epsilon}_{...})^2$, and

$$\begin{aligned}E(SSTr)&=r\sum_k\tau_k^2+r\sum_k E(\bar{\epsilon}_{..k}^2)-r^2\sum_k E(\bar{\epsilon}_{...}^2)\\&=r\sum_k\tau_k^2+r\sum_k\frac{\sigma^2}{r}-r^2\frac{\sigma^2}{r^2}=r\sum_k\tau_k^2+r\sigma^2-\sigma^2\\&=(r-1)\sigma^2+r\sum_k\tau_k^2.\end{aligned}$$

13.39 The hypotheses are

$$H_0:\tau_1=\tau_2=\tau_3=\tau_4=0,\ \text{professor effects are zero}$$
$$H_1:\text{At least one of the }\tau_i\text{'s is not equal to zero.}$$

$\alpha=0.05$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Time Periods	474.50	3	158.17	
Courses	252.50	3	84.17	
Professors	723.50	3	241.17	5.03
Error	287.50	6	47.92	
Total	1738.00	15		

P-value$=0.0446$. Decision: Reject H_0; grades are affected by different professors.

13.40 The hypotheses are

$$H_0 : \tau_A = \tau_B = \tau_C = \tau_D = \tau_E = 0, \text{ color additive effects are zero}$$

$$H_1 : \text{At least one of the } \tau_i\text{'s is not equal to zero.}$$

$\alpha = 0.05$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Workers	12.4344	4	3.1086	
Days	14.7944	4	3.6986	
Additives	3.9864	4	0.9966	1.29
Error	9.2712	12	0.7726	
Total	40.4864	24		

P-value$= 0.3280$. Decision: Do not reject H_0; color additives could not be shown to have an effect on setting time.

13.41 The hypotheses are

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0, \text{ dye effects are zero}$$

$$H_1 : \text{At least one of the } \alpha_i\text{'s is not equal to zero.}$$

$\alpha = 0.05$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Amounts	1238.8825	2	619.4413	122.37
Plants	53.7004	1	53.7004	
Error	101.2433	20	5.0622	
Total	1393.8263	23		

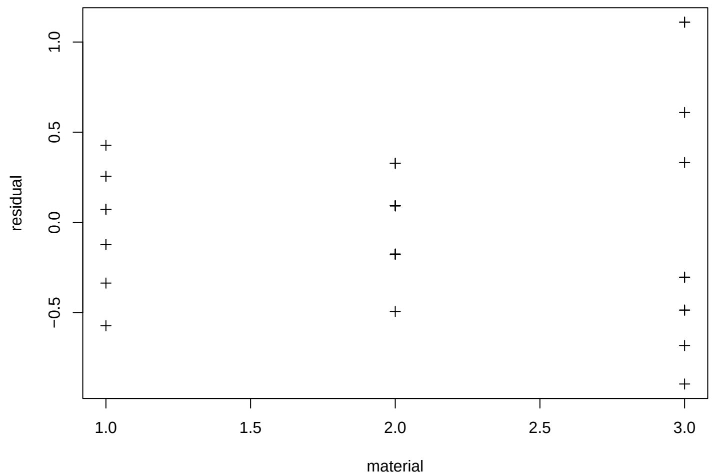
P-value< 0.0001. Decision: Reject H_0; color densities of fabric differ significantly for three levels of dyes.

13.42 (a) After a transformation $g(y) = \sqrt{y}$, we come up with an ANOVA table as follows.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Materials	7.5123	2	3.7561	16.20
Error	6.2616	27	0.2319	
Total	13.7739	29		

(b) The P-value< 0.0001. Hence, there is significant difference in flaws among three materials.

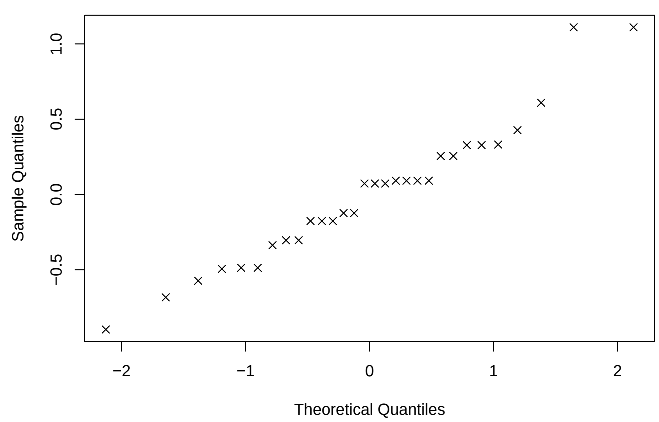
(c) A residual plot is given below and it does show some heterogeneity of the variance among three treatment levels.



(d) The purpose of the transformation is to stabilize the variances.

(e) One could be the distribution assumption itself. Once the data is transformed, it is not necessary that the data would follow a normal distribution.

(f) Here the normal probability plot on residuals is shown.



It appears to be close to a straight line. So, it is likely that the transformed data are normally distributed.

13.43 (a) The hypotheses are

$$H_0 : \sigma_\alpha^2 = 0,$$
$$H_1 : \sigma_\alpha^2 \neq 0$$

$\alpha = 0.05$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Operators	371.8719	3	123.9573	14.91
Error	99.7925	12	8.3160	
Total	471.6644	15		

P-value= 0.0002. Decision: Reject H_0; operators are different.

(b) $\hat{\sigma}^2 = 8.316$ and $\hat{\sigma}^2_\alpha = \frac{123.9573-8.3160}{4} = 28.910$.

13.44 The model is $y_{ij} = \mu + A_i + B_j + \epsilon_{ij}$. Hence

$$\bar{y}_{.j} = \mu + \bar{A}_. + B_j + \bar{\epsilon}_{.j}, \quad \text{and} \quad \bar{y}_{..} = \mu + \bar{A}_. + \bar{B}_. + \bar{\epsilon}_{..}.$$

Therefore,

$$SSB = k\sum_{j=1}^{b}(\bar{y}_{.j} - \bar{y}_{..})^2 = k\sum_{j=1}^{b}[(B_j - \bar{B}_.) + (\bar{\epsilon}_{.j} - \bar{\epsilon}_{..})]^2,$$

and

$$\begin{aligned} E(SSB) &= k\sum_{j=1}^{b} E(B_j^2) - kbE(\bar{B}_.^2) + k\sum_{j=1}^{b} E(\bar{\epsilon}_{.j}^2) - kbE(\bar{\epsilon}_{..}^2) \\ &= kb\sigma_\beta^2 - k\sigma_\beta^2 + b\sigma^2 - \sigma^2 = (b-1)\sigma^2 + k(b-1)\sigma_\beta^2. \end{aligned}$$

13.45 (a) The hypotheses are

$$H_0 : \sigma_\alpha^2 = 0,$$
$$H_1 : \sigma_\alpha^2 \neq 0$$

$\alpha = 0.05$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Treatments	23.238	3	7.746	3.33
Blocks	45.283	4	11.321	
Error	27.937	12	2.328	
Total	96.458	19		

P-value= 0.0565. Decision: Not able to show a significant difference in the random treatments at 0.05 level, although the P-value shows marginal significance.

(b) $\sigma_\alpha^2 = \frac{7.746-2.328}{5} = 1.084$, and $\sigma_\beta^2 = \frac{11.321-2.328}{4} = 2.248$.

13.46 From the model

$$y_{ijk} = \mu + A_i + B_j + T_k + \epsilon_{ij},$$

we have

$$\bar{y}_{..k} = \mu + \bar{A}_{.} + \bar{B}_{.} + T_k + \bar{\epsilon}_{..k}, \quad \text{and} \quad \bar{y}_{...} = \mu + \bar{A}_{.} + \bar{B}_{.} + \bar{T}_{.} + \bar{\epsilon}_{...}.$$

Hence,

$$SSTr = r\sum_k (\bar{y}_{..k} - \bar{y}_{...})^2 = r\sum_k [(T_j - \bar{T}_{.}) + (\bar{\epsilon}_{..k} - \bar{\epsilon}_{...})]^2,$$

and

$$\begin{aligned} E(SSTr) &= r\sum_k E(T_k^2) - r^2 E(\bar{T}_{.}^2) + r\sum_k E(\bar{\epsilon}_{..k}^2) - r^2 E(\bar{\epsilon}_{...}^2) \\ &= r^2\sigma_\tau^2 - r\sigma_\tau^2 + r\sigma^2 - \sigma^2 = (r-1)(\sigma^2 + r\sigma_\beta^2). \end{aligned}$$

13.47 (a) The matrix is

$$\mathbf{A} = \begin{bmatrix} bk & b & b & \cdots & b & k & k & \cdots & k \\ b & b & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \\ b & 0 & b & \cdots & 0 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ b & 0 & 0 & \cdots & b & 1 & 1 & \cdots & 1 \\ k & 1 & 1 & \cdots & k & 0 & 0 & \cdots & 0 \\ k & 1 & 1 & \cdots & 0 & k & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ k & 1 & 1 & \cdots & 0 & 0 & 0 & \cdots & k \end{bmatrix},$$

where $b =$ number of blocks and $k =$ number of treatments. The vectors are

$$\begin{aligned} \mathbf{b}' &= (\mu, \alpha_1, \alpha_2, \cdots, \alpha_k, \beta_1, \beta_2, \cdots, \beta_b)', \quad \text{and} \\ \mathbf{g}' &= (T_{..}, T_{1.}, T_{2.}, \cdots, T_{k.}, T_{.1}, T_{.2}, \cdots, T_{.b})'. \end{aligned}$$

(b) Solving the system $\mathbf{Ab} = \mathbf{g}$ with the constraints $\sum\limits_{i=1}^{k} \alpha_i = 0$ and $\sum\limits_{j=1}^{b} \beta_j = 0$, we have

$$\begin{aligned} \hat{\mu} &= \bar{y}_{..}, \\ \hat{\alpha}_i &= \bar{y}_{i.} - \bar{y}_{..}, \quad \text{for } i = 1, 2, \dots, k, \\ \hat{\beta}_j &= \bar{y}_{.j} - \bar{y}_{..}, \quad \text{for } j = 1, 2, \dots, b. \end{aligned}$$

Therefore,

$$\begin{aligned} R(\alpha_1, \alpha_2, \dots, \alpha_k, \beta_1, \beta_2, \dots, \beta_b) &= \mathbf{b}'\mathbf{g} - \frac{T_{..}^2}{bk} \\ &= \sum_{i=1}^{k} \frac{T_{i.}^2}{b} + \sum_{j=1}^{b} \frac{T_{.j}^2}{k} - 2\frac{T_{..}^2}{bk}. \end{aligned}$$

To find $R(\beta_1, \beta_2, \ldots, \beta_b \mid \alpha_1, \alpha_2, \ldots, \alpha_k)$ we first find $R(\alpha_1, \alpha_2, \ldots, \alpha_k)$. Setting $\beta_j = 0$ in the model, we obtain the estimates (after applying the constraint $\sum_{i=1}^{k} \alpha_i = 0$)

$$\hat{\mu} = \bar{y}_{..}, \quad \text{and} \quad \hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..}, \quad \text{for } i = 1, 2, \ldots, k.$$

The $\mathbf{g}$ vector is the same as in part (a) with the exception that $T_{.1}, T_{.2}, \ldots, T_{.b}$ do not appear. Thus one obtains

$$R(\alpha_1, \alpha_2, \ldots, \alpha_k) = \sum_{i=1}^{k} \frac{T_{i.}^2}{b} - \frac{T_{..}^2}{bk}$$

and thus

$$R(\beta_1, \beta_2, \ldots, \beta_b \mid \alpha_1, \alpha_2, \ldots, \alpha_k) = R(\alpha_1, \alpha_2, \ldots, \alpha_k, \beta_1, \beta_2, \ldots, \beta_b) - R(\alpha_1, \alpha_2, \ldots, \alpha_k) = \sum_{j=1}^{b} \frac{T_{.j}^2}{k} - \frac{T_{..}^2}{bk} = SSB.$$

13.48 Since

$$1 - \beta = P\left[F(3, 12) > \frac{3.49}{1 + (4)(1.5)}\right] = P[F(12, 3) < 2.006] < 0.95.$$

Hence we do not have large enough samples. We then find, by trial and error, that $n = 16$ is sufficient since

$$1 - \beta = P\left[F(3, 60) > \frac{2.76}{1 + (16)(1.5)}\right] = P[F(60, 3) < 9.07] > 0.95.$$

13.49 We know $\phi^2 = b \sum_{i=1}^{4} \frac{\alpha_i^2}{4\sigma^2} = \frac{b}{2}$, when $\sum_{i=1}^{4} \frac{\alpha_i^2}{\sigma^2} = 2.0$.
If $b = 10$, $\phi = 2.24$; $v_1 = 3$ and $v_2 = 27$ degrees of freedom.
If $b = 9$, $\phi = 2.12$; $v_1 = 3$ and $v_2 = 24$ degrees of freedom.
If $b = 8$, $\phi = 2.00$; $v_1 = 3$ and $v_2 = 21$ degrees of freedom.
From Table A.16 we see that $b = 9$ gives the desired result.

13.50 For the randomized complete block design we have

$$E(S_1^2) = E\left(\frac{SSA}{k-1}\right) = \sigma^2 + b \sum_{i=1}^{k} \frac{\alpha_i^2}{k-1}.$$

Therefore,

$$\lambda = \frac{v_1[E(S_1^2)]}{2\sigma^2} - \frac{v_1}{2} = \frac{(k-1)\left[\sigma^2 + b \sum_{i=1}^{k} \alpha_i^2/(k-1)\right]}{2\sigma^2} - \frac{k-1}{2} \quad = b \sum_{i=1}^{k} \frac{\alpha_i^2}{2\sigma^2},$$

and then

$$\phi^2 = \frac{E(S_1^2) - \sigma^2}{\sigma^2} \cdot \frac{v_1}{v_1 + 1} = \frac{[\sigma^2 + b\sum\limits_{i=1}^{k} \alpha_i^2/(k-1)] - \sigma^2}{\sigma^2} \cdot \frac{k-1}{k} = b\sum_{i=1}^{k} \frac{\alpha_i^2}{k\sigma^2}.$$

13.51 (a) The model is $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$, where $\alpha_i \sim n(0, \sigma_\alpha^2)$.

(b) Since $s^2 = 0.02056$ and $s_1^2 = 0.01791$, we have $\hat{\sigma}^2 = 0.02056$ and $\frac{s_1^2 - s^2}{10} = \frac{0.01791 - 0.02056}{10} = -0.00027$, which implies $\hat{\sigma}_\alpha^2 = 0$.

13.52 (a) The P-value of the test result is 0.7830. Hence, the variance component of pour is not significantly different from 0.

(b) We have the ANOVA table as follows:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Pours	0.08906	4	0.02227	0.43
Error	1.02788	20	0.05139	
Total	1.11694	24		

Since $\frac{s_1^2 - s^2}{5} = \frac{0.02227 - 0.05139}{5} < 0$, we have $\hat{\sigma}_\alpha^2 = 0$.

13.53 (a) $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$, where $\alpha_i \sim n(x; 0, \sigma_\alpha^2)$.

(b) Running an ANOVA analysis, we obtain the P-value as 0.0121. Hence, the loom variance component is significantly different from 0 at level 0.05.

(c) The suspicion is supported by the data.

13.54 The hypotheses are

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4,$$

$$H_1: \text{At least two of the } \mu_i\text{'s are not equal.}$$

$\alpha = 0.05$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Garlon levels	3.7289	3	1.2430	1.57
Error	9.5213	12	0.7934	
Total	13.2502	15		

P-value$= 0.2487$. Decision: Do not reject H_0; there is insufficient evidence to claim that the concentration levels of Garlon would impact the heights of shoots.

13.55 Bartlett's statistic is $b = 0.8254$. Conclusion: do not reject homogeneous variance assumption.

13.56 The hypotheses are

$$H_0 : \tau_A = \tau_B = \tau_C = \tau_D = \tau_E = 0, \text{ ration effects are zero}$$

$$H_1 : \text{At least one of the } \tau_i\text{'s is not zero.}$$

$\alpha = 0.01$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Lactation Periods	245.8224	4	61.4556	
Cows	353.1224	4	88.2806	
Rations	89.2624	4	22.3156	12.55
Error	21.3392	12	1.7783	
Total	709.5464	24		

P-value= 0.0003. Decision: Reject H_0; different rations have an effect on the daily milk production.

13.57 It can be shown that $\bar{y}_C = 76.16$, $\bar{y}_1 = 81.20$, $\bar{y}_2 = 81.54$ and $\bar{y}_3 = 80.98$. Since this is a one-sided test, we find $d_{0.01}(3, 16) = 3.05$ and

$$\sqrt{\frac{2s^2}{n}} = \sqrt{\frac{(2)(3.52575)}{5}} = 1.18756.$$

Hence

$$d_1 = \frac{81.20 - 76.16}{1.18756} = 4.244, \; d_2 = \frac{81.54 - 76.16}{1.18756} = 4.532, \; d_3 = \frac{80.98 - 76.16}{1.18756} = 4.059,$$

which are all larger than the critical value. Hence, significantly higher yields are obtained with the catalysts than with no catalyst.

13.58 (a) The hypotheses for the Bartlett's test are

$$H_0 : \sigma_A^2 = \sigma_B^2 = \sigma_C^2 = \sigma_D^2,$$
$$H_1 : \text{The variances are not all equal.}$$

$\alpha = 0.05$.
Critical region: We have $n_1 = n_2 = n_3 = n_4 = 5$, $N = 20$, and $k = 4$. Therefore, we reject H_0 when $b < b_4(0.05, 5) = 0.5850$.
Computation: $s_A = 1.40819$, $s_B = 2.16056$, $s_C = 1.16276$, $s_D = 0.76942$ and hence $s_p = 1.46586$. From these, we can obtain that $b = 0.7678$.
Decision: Do not reject H_0; there is no sufficient evidence to conclude that the variances are not equal.

(b) The hypotheses are

$$H_0 : \mu_A = \mu_B = \mu_C = \mu_D,$$

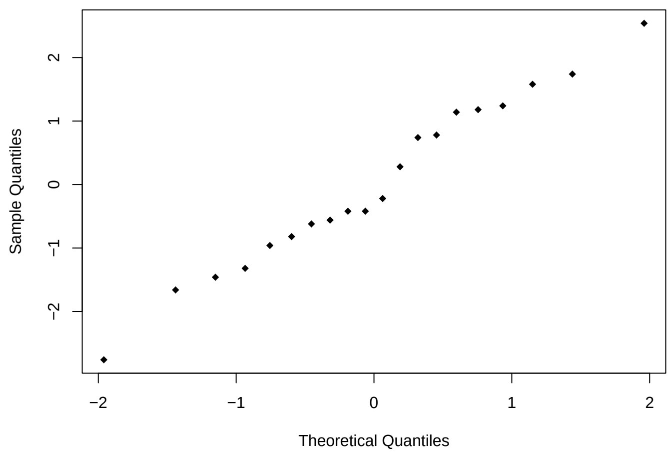
$$H_1 : \text{At least two of the } \mu_i\text{'s are not equal.}$$

$\alpha = 0.05$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Laboratories	85.9255	3	28.6418	13.33
Error	34.3800	16	2.1488	
Total	120.3055	19		

P-value= 0.0001. Decision: Reject H_0; the laboratory means are significantly different.

(c) The normal probability plot is given as follows:



13.59 The hypotheses for the Bartlett's test are

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2,$$
$$H_1 : \text{The variances are not all equal.}$$

$\alpha = 0.01$.
Critical region: We have $n_1 = n_2 = n_3 = 4$, $n_4 = 9$ $N = 21$, and $k = 4$. Therefore, we reject H_0 when

$$b < b_4(0.01, 4, 4, 4, 9)$$
$$= \frac{(4)(0.3475) + (4)(0.3475) + (4)(0.3475) + (9)(0.6892)}{21} = 0.4939.$$

Computation: $s_1^2 = 0.41709$, $s_2^2 = 0.93857$, $s_3^2 = 0.25673$, $s_4^2 = 1.72451$ and hence $s_p^2 = 1.0962$. Therefore,

$$b = \frac{[(0.41709)^3(0.93857)^3(0.25763)^3(1.72451)^8]^{1/17}}{1.0962} = 0.79.$$

Decision: Do not reject H_0; the variances are not significantly different.

13.60 The hypotheses for the Cochran's test are

$$H_0 : \sigma_A^2 = \sigma_B^2 = \sigma_C^2,$$
$$H_1 : \text{The variances are not all equal.}$$

$\alpha = 0.01$.
Critical region: $g > 0.6912$.
Computation: $s_A^2 = 29.5667$, $s_B^2 = 10.8889$, $s_C^2 = 3.2111$, and hence $\sum s_i^2 = 43.6667$.
Now, $g = \frac{29.5667}{43.6667} = 0.6771$.
Decision: Do not reject H_0; the variances are not significantly different.

13.61 The hypotheses for the Bartlett's test are

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2,$$
$$H_1 : \text{The variances are not all equal.}$$

$\alpha = 0.05$.
Critical region: reject H_0 when

$$b < b_4(0.05, 9, 8, 15) = \frac{(9)(0.7686) + (8)(0.7387) + (15)(0.8632)}{32} = 0.8055.$$

Computation: $b = \frac{[(0.02832)^8(0.16077)^7(0.04310)^{14}]^{1/29}}{0.067426} = 0.7822$.
Decision: Reject H_0; the variances are significantly different.

13.62 (a) The hypotheses are

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0,$$
$$H_1 : \text{At least one of the } \alpha_i\text{'s is not zero.}$$

$\alpha = 0.05$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Diets	822.1360	3	274.0453	38.65
Blocks	17.1038	5	3.4208	
Error	106.3597	15	7.0906	
Total	945.5995	23		

with P-value< 0.0001.
Decision: Reject H_0; diets do have a significant effect on mean percent dry matter.

(b) We know that $\bar{y}_C = 35.8483$, $\bar{y}_F = 36.4217$, $\bar{y}_T = 45.1833$, $\bar{y}_A = 49.6250$, and

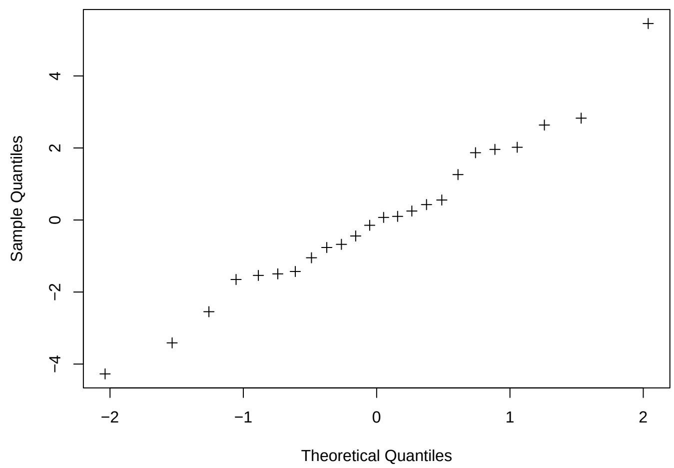
$$\sqrt{\frac{2s^2}{n}} = \sqrt{\frac{(2)(7.0906)}{6}} = 1.5374.$$

Hence,

$$\begin{aligned} d_{\text{Ammonia}} &= \frac{49.6250 - 35.8483}{1.5374} = 8.961, \\ d_{\text{Urea Feeding}} &= \frac{36.4217 - 35.8483}{1.5374} = 0.3730, \\ d_{\text{Urea Treated}} &= \frac{45.1833 - 35.8483}{1.5374} = 6.0719. \end{aligned}$$

Using the critical value $d_{0.05}(3, 15) = 2.61$, we obtain that only "Urea Feeding" is not significantly different from the control, at level 0.05.

(c) The normal probability plot for the residuals are given below.



13.63 The hypotheses are

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0,$$

$$H_1 : \text{At least one of the } \alpha_i\text{'s is not zero.}$$

$\alpha = 0.05$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Diet	0.32356	2	0.16178	9.33
Error	0.20808	12	0.01734	
Total	0.53164	14		

with P-value= 0.0036.
Decision: Reject H_0; zinc is significantly different among the diets.

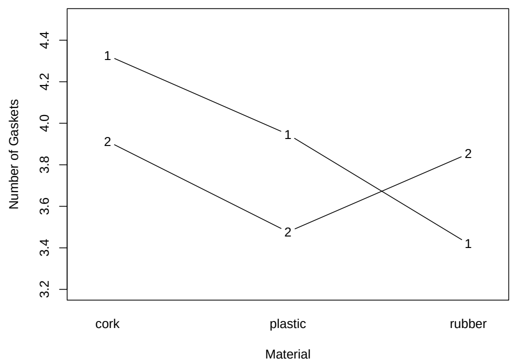
13.64 (a) The gasoline manufacturers would want to apply their results to more than one model of car.

(b) Yes, there is a significant difference in the miles per gallon for the three brands of gasoline.

(c) I would choose brand C for the best miles per gallon.

13.65 (a) The process would include more than one stamping machine and the results might differ with different machines.

(b) The mean plot is shown below.



(c) Material 1 appears to be the best.

(d) Yes, there is interaction. Materials 1 and 3 have better results with machine 1 but material 2 has better results with machine 2.

13.66 (a) The hypotheses are

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0,$$
$$H_1 : \text{At least one of the } \alpha_i\text{'s is not zero.}$$

Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Paint Types	227875.11	2	113937.57	5.08
Error	336361.83	15	22424.12	
Total	564236.94	17		

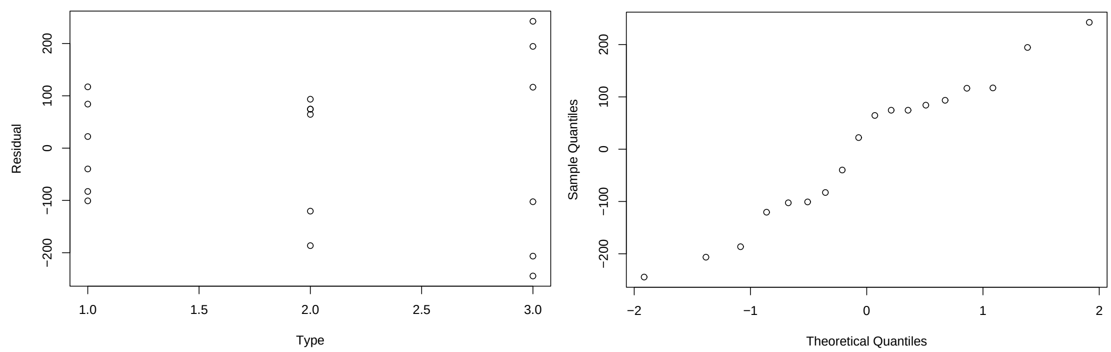
with P-value= 0.0207.
Decision: Reject H_0 at level 0.05; the average wearing quality differs significantly for three paints.

(b) Using Tukey's test, it turns out the following.

$\bar{y}_{1.}$	$\bar{y}_{3.}$	$\bar{y}_{2.}$
197.83	419.50	450.50

Types 2 and 3 are not significantly different, while Type 1 is significantly different from Type 2.

(c) We plot the residual plot and the normal probability plot for the residuals as follows.

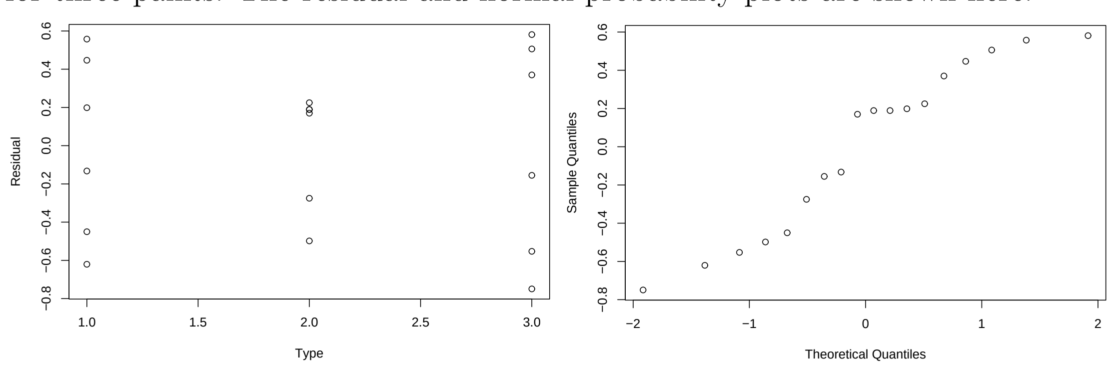


It appears that the heterogeneity in variances may be violated, as is the normality assumption.

(d) We do a log transformation of the data, i.e., $y' = \log(y)$. The ANOVA result has changed as follows.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Paint Types	2.6308	2	1.3154	6.07
Error	3.2516	15	0.2168	
Total	5.8824	17		

with P-value= 0.0117.
Decision: Reject H_0 at level 0.05; the average wearing quality differ significantly for three paints. The residual and normal probability plots are shown here:



While the homogeneity of the variances seem to be a little better, the normality assumption may still be invalid.

13.67 (a) The hypotheses are

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0,$$

$$H_1 : \text{At least one of the } \alpha_i\text{'s is not zero.}$$

Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Locations	0.01594	3	0.00531	13.80
Error	0.00616	16	0.00039	
Total	0.02210	19		

with P-value= 0.0001.
Decision: Reject H_0; the mean ozone levels differ significantly across the locations.

(b) Using Tukey's test, the results are as follows.

$\bar{y}_{4.}$	$\bar{y}_{1.}$	$\bar{y}_{3.}$	$\bar{y}_{2.}$
0.078	0.092	0.096	0.152

Location 2 appears to have much higher ozone measurements than other locations.

Chapter 14

Factorial Experiments (Two or More Factors)

14.1 The hypotheses of the three parts are,

(a) for the main effects temperature,

$$H_0^{'} : \alpha_1 = \alpha_2 = \alpha_3 = 0,$$
$$H_1^{'} : \text{At least one of the } \alpha_i\text{'s is not zero;}$$

(b) for the main effects ovens,

$$H_0^{''} : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0,$$
$$H_1^{''} : \text{At least one of the } \beta_i\text{'s is not zero;}$$

(c) and for the interactions,

$$H_0^{'''} : (\alpha\beta)_{11} = (\alpha\beta)_{12} = \cdots = (\alpha\beta)_{34} = 0,$$
$$H_1^{'''} : \text{At least one of the } (\alpha\beta)_{ij}\text{'s is not zero.}$$

$\alpha = 0.05$.
Critical regions: (a) $f_1 > 3.00$; (b) $f_2 > 3.89$; and (c) $f_3 > 3.49$.
Computations: From the computer printout we have

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Temperatures	5194.08	2	2597.0400	8.13
Ovens	4963.12	3	1654.3733	5.18
Interaction	3126.26	6	521.0433	1.63
Error	3833.50	12	319.4583	
Total	17,116.96	23		

Decision: (a) Reject H_0'; (b) Reject H_0''; (c) Do not reject H_0'''.

14.2 The hypotheses of the three parts are,

(a) for the main effects brands,

$$H_0' : \alpha_1 = \alpha_2 = \alpha_3 = 0,$$
$$H_1' : \text{At least one of the } \alpha_i\text{'s is not zero;}$$

(b) for the main effects times,

$$H_0'' : \beta_1 = \beta_2 = \beta_3 = 0,$$
$$H_1'' : \text{At least one of the } \beta_i\text{'s is not zero;}$$

(c) and for the interactions,

$$H_0''' : (\alpha\beta)_{11} = (\alpha\beta)_{12} = \cdots = (\alpha\beta)_{33} = 0,$$
$$H_1''' : \text{At least one of the } (\alpha\beta)_{ij}\text{'s is not zero.}$$

$\alpha = 0.05$.
Critical regions: (a) $f_1 > 3.35$; (b) $f_2 > 3.35$; and (c) $f_3 > 2.73$.
Computations: From the computer printout we have

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Brands	32.7517	2	16.3758	1.74
Times	227.2117	2	113.6058	12.04
Interaction	17.3217	4	4.3304	0.46
Error	254.7025	27	9.4334	
Total	531.9875	35		

Decision: (a) Do not reject H_0'; (b) Reject H_0''; (c) Do not reject H_0'''.

14.3 The hypotheses of the three parts are,

(a) for the main effects environments,

$$H_0' : \alpha_1 = \alpha_2 = 0, \text{ (no differences in the environment)}$$
$$H_1' : \text{At least one of the } \alpha_i\text{'s is not zero;}$$

(b) for the main effects strains,

$$H_0'' : \beta_1 = \beta_2 = \beta_3 = 0, \text{ (no differences in the strains)}$$
$$H_1'' : \text{At least one of the } \beta_i\text{'s is not zero;}$$

(c) and for the interactions,

$H_0^{'''}$: $(\alpha\beta)_{11} = (\alpha\beta)_{12} = \cdots = (\alpha\beta)_{23} = 0$, (environments and strains do not interact)

$H_1^{'''}$: At least one of the $(\alpha\beta)_{ij}$'s is not zero.

$\alpha = 0.01$.
Critical regions: (a) $f_1 > 7.29$; (b) $f_2 > 5.16$; and (c) $f_3 > 5.16$.
Computations: From the computer printout we have

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Environments	14,875.521	1	14,875.521	14.81
Strains	18,154.167	2	9,077.083	9.04
Interaction	1,235.167	2	617.583	0.61
Error	42,192.625	42	1004.586	
Total	76,457.479	47		

Decision: (a) Reject $H_0^{'}$; (b) Reject $H_0^{''}$; (c) Do not reject $H_0^{'''}$. Interaction is not significant, while both main effects, environment and strain, are all significant.

14.4 (a) The hypotheses of the three parts are,

$$H_0^{'} : \alpha_1 = \alpha_2 = \alpha_3 = 0$$

$$H_1^{'} : \text{At least one of the } \alpha_i\text{'s is not zero;}$$

$$H_0^{''} : \beta_1 = \beta_2 = \beta_3 = 0,$$

$$H_1^{''} : \text{At least one of the } \beta_i\text{'s is not zero;}$$

$$H_0^{'''} : (\alpha\beta)_{11} = (\alpha\beta)_{12} = \cdots = (\alpha\beta)_{33} = 0,$$

$$H_1^{'''} : \text{At least one of the } (\alpha\beta)_{ij}\text{'s is not zero.}$$

$\alpha = 0.01$.
Critical regions: for $H_0^{'}$, $f_1 > 3.21$; for $H_0^{''}$, $f_2 > 3.21$; and for $H_0^{'''}$, $f_3 > 2.59$.
Computations: From the computer printout we have

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Coating	1,535,021.37	2	767,510.69	6.87
Humidity	1,020,639.15	2	510,319.57	4.57
Interaction	1,089,989.63	4	272,497.41	2.44
Error	5,028,396.67	45	111,742.15	
Total	76,457.479	47		

Decision: Reject $H_0^{'}$; Reject $H_0^{''}$; Do not reject $H_0^{'''}$. Coating and humidity do not interact, while both main effects are all significant.

(b) The three means for the humidity are $\bar{y}_L = 733.78$, $\bar{y}_M = 406.39$ and $\bar{y}_H = 638.39$. Using Duncan's test, the means can be grouped as

$$\begin{array}{ccc} \bar{y}_M & \bar{y}_L & \bar{y}_H \\ \underline{406.39 \quad 638.39} & & \underline{733.78} \end{array}$$

Therefore, corrosion damage is different for medium humidity than for low or high humidity.

14.5 The hypotheses of the three parts are,

(a) for the main effects subjects,

$$H_0^{'} : \alpha_1 = \alpha_2 = \alpha_3 = 0,$$

$$H_1^{'} : \text{At least one of the } \alpha_i\text{'s is not zero;}$$

(b) for the main effects muscles,

$$H_0^{''} : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0,$$

$$H_1^{''} : \text{At least one of the } \beta_i\text{'s is not zero;}$$

(c) and for the interactions,

$$H_0^{'''} : (\alpha\beta)_{11} = (\alpha\beta)_{12} = \cdots = (\alpha\beta)_{35} = 0,$$

$$H_1^{'''} : \text{At least one of the } (\alpha\beta)_{ij}\text{'s is not zero.}$$

$\alpha = 0.01$.
Critical regions: (a) $f_1 > 5.39$; (b) $f_2 > 4.02$; and (c) $f_3 > 3.17$.
Computations: From the computer printout we have

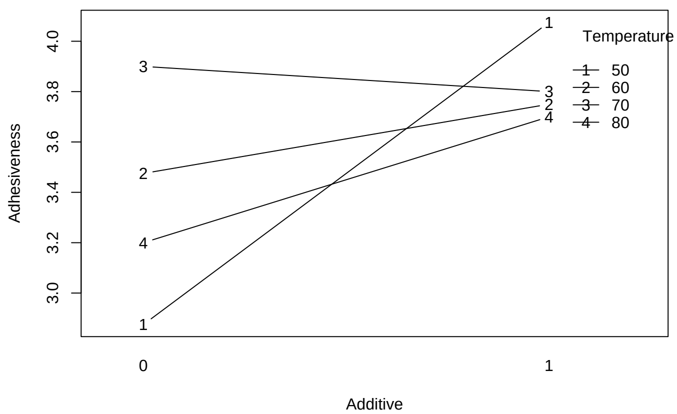
Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Subjects	4,814.74	2	2,407.37	34.40
Muscles	7,543.87	4	1,885.97	26.95
Interaction	11,362.20	8	1,420.28	20.30
Error	2,099.17	30	69.97	
Total	25,819.98	44		

Decision: (a) Reject $H_0^{'}$; (b) Reject $H_0^{''}$; (c) Reject $H_0^{'''}$.

14.6 The ANOVA table is shown as

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Additive	1.7578	1	1.7578	22.29	< 0.0001
Temperature	0.8059	3	0.2686	3.41	0.0338
Interaction	1.7934	3	0.5978	7.58	0.0010
Error	1.8925	24	0.0789		
Total	6.2497	32			

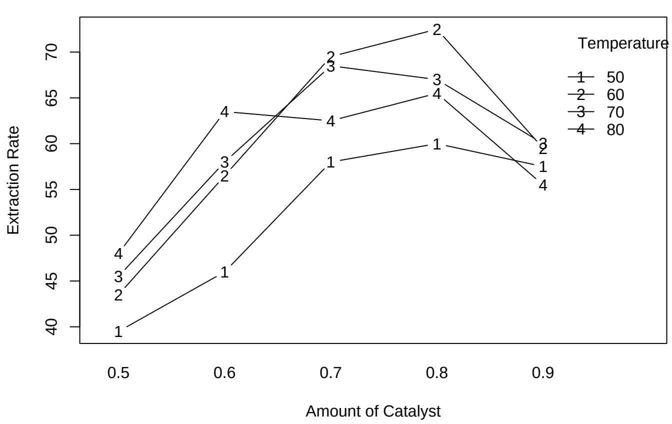
Decision: All main effects and interaction are significant.
An interaction plot is given here.



14.7 The ANOVA table is

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Temperature	430.475	3	143.492	10.85	0.0003
Catalyst	2,466.650	4	616.663	46.63	< 0.0001
Interaction	326.150	12	27.179	2.06	0.0745
Error	264.500	20	13.225		
Total	3,487.775	39			

Decision: All main effects are significant and the interaction is significant at level 0.0745. Hence, if 0.05 significance level is used, interaction is not significant.
An interaction plot is given here.



Duncan's tests, at level 0.05, for both main effects result in the following.

(a) For Temperature:

$\bar{y}_{50}$	$\bar{y}_{80}$	$\bar{y}_{70}$	$\bar{y}_{60}$
52.200	59.000	59.800	60.300

(b) For Amount of Catalyst:

$$\underline{\bar{y}_{0.5}\atop 44.125}\quad \underline{\bar{y}_{0.6}\quad \bar{y}_{0.9}\atop 56.000\ 58.125}\quad \underline{\bar{y}_{0.7}\quad \bar{y}_{0.8}\atop 64.625\ 66.250}$$

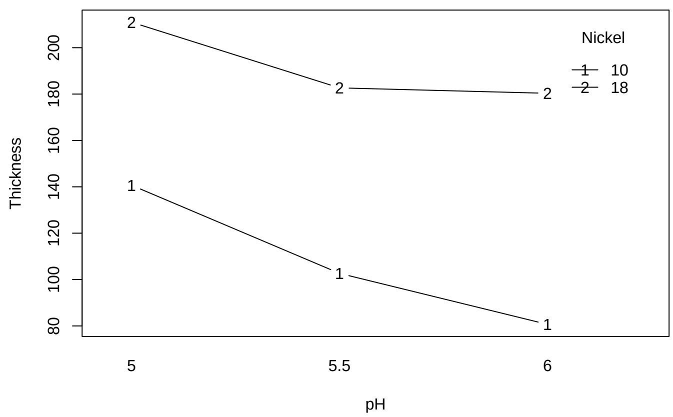
14.8 (a) The ANOVA table is

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Nickel	$31,250.00$	1	$31,250.00$	44.52	< 0.0001
pH	$6,606.33$	2	$3,303.17$	4.71	0.0310
Nickel*pH	670.33	2	335.17	0.48	0.6316
Error	$8,423.33$	12	701.94		
Total	$3,487.775$	39			

Decision: Nickel contents and levels of pH do not interact to each other, while both main effects of nickel contents and levels of pH are all significant, at level higher than 0.0310.

(b) In comparing the means of the six treatment combinations, a nickel content level of 18 and a pH level of 5 resulted in the largest values of thickness.

(c) The interaction plot is given here and it shows no apparent interactions.

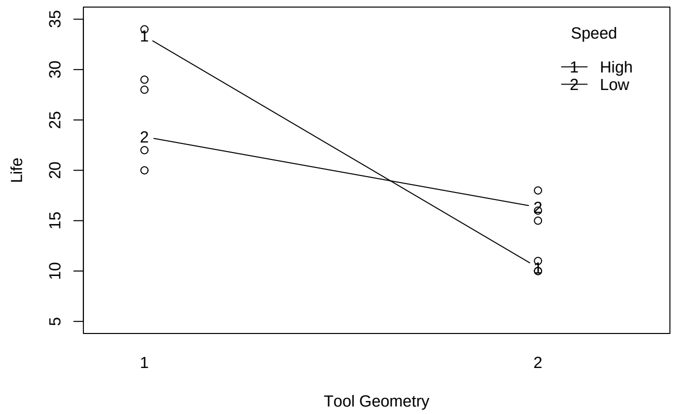


14.9 (a) The ANOVA table is

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Tool	675.00	1	675.00	74.31	< 0.0001
Speed	12.00	1	12.00	1.32	0.2836
Tool*Speed	192.00	1	192.00	21.14	0.0018
Error	72.67	8	9.08		
Total	951.67	11			

Decision: The interaction effects are significant. Although the main effects of speed showed insignificance, we might not make such a conclusion since its effects might be masked by significant interaction.

(b) In the graph shown, we claim that the cutting speed that results in the longest life of the machine tool depends on the tool geometry, although the variability of the life is greater with tool geometry at level 1.



(c) Since interaction effects are significant, we do the analysis of variance for separate tool geometry.

(i) For tool geometry 1, an f-test for the cutting speed resulted in a P-value = 0.0405 with the mean life (standard deviation) of the machine tool at 33.33 (4.04) for high speed and 23.33 (4.16) for low speed. Hence, a high cutting speed has longer life for tool geometry 1.

(ii) For tool geometry 2, an f-test for the cutting speed resulted in a P-value = 0.0031 with the mean life (standard deviation) of the machine tool at 10.33 (0.58) for high speed and 16.33 (1.53) for low speed. Hence, a low cutting speed has longer life for tool geometry 2.

For the above detailed analysis, we note that the standard deviations for the mean life are much higher at tool geometry 1.

(d) See part (b).

14.10 (a) $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$, $i = 1, 2, \dots, a$; $j = 1, 2, \dots, b$; $k = 1, 2, \dots, n$.

(b) The ANOVA table is

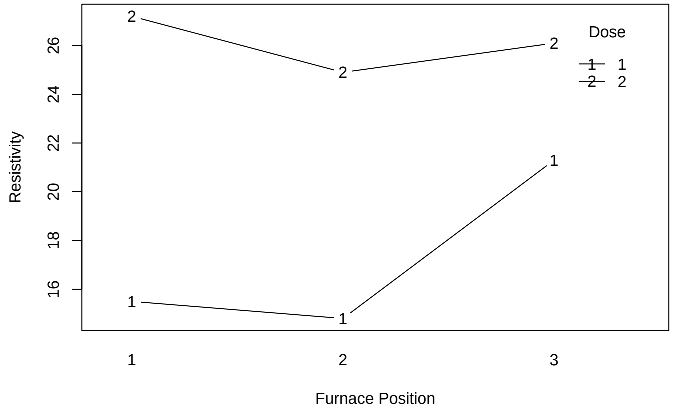
Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Dose	117.9267	1	117.9267	18.08	0.0511
Position	15.0633	2	7.5317	1.15	0.4641
Error	13.0433	2	6.5217		
Total	146.0333	5			

(c) $(n-1) - (a-1) - (b-1) = 5 - 1 - 2 = 2$.

(d) At level 0.05, Tukey's result for the furnace position is shown here:

$\bar{y}_2$	$\bar{y}_1$	$\bar{y}_3$
19.850	21.350	23.700

Although Tukey's multiple comparisons resulted in insignificant differences among the furnace position levels, based on a P-value of 0.0511 for the Dose and on the plot shown we can see that Dose=2 results in higher resistivity.



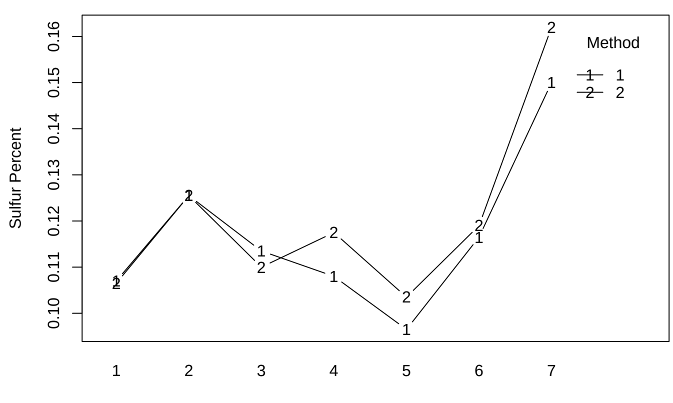
14.11 (a) The ANOVA table is

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Method	0.000104	1	0.000104	6.57	0.0226
Lab	0.008058	6	0.001343	84.70	< 0.0001
Method*Lab	0.000198	6	0.000033	2.08	0.1215
Error	0.000222	14	0.000016		
Total	0.00858243	27			

(b) Since the P-value = 0.1215 for the interaction, the interaction is not significant. Hence, the results on the main effects can be considered meaningful to the scientist.

(c) Both main effects, method of analysis and laboratory, are all significant.

(d) The interaction plot is show here.



(e) When the tests are done separately, i.e., we only use the data for Lab 1, or Lab 2 alone, the P-values for testing the differences of the methods at Lab 1 and 7 are 0.8600 and 0.1557, respectively. In this case, usually the degrees of freedom of errors are small. If we compare the mean differences of the method within the overall ANOVA model, we obtain the P-values for testing the differences of the methods at Lab 1 and 7 as 0.9010 and 0.0093, respectively. Hence, methods are no difference in Lab 1 and are significantly different in Lab 7. Similar results may be found in the interaction plot in (d).

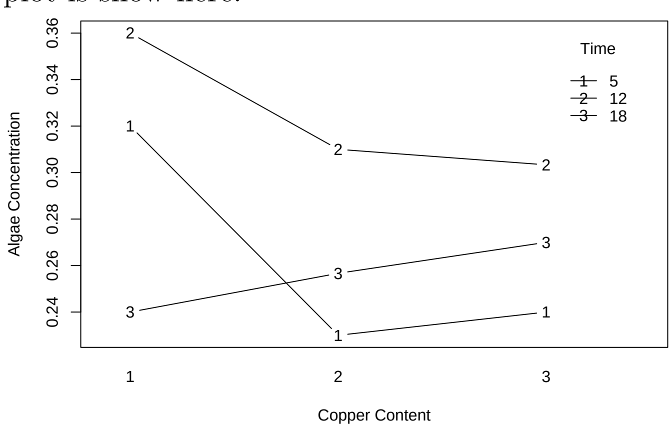
14.12 (a) The ANOVA table is

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Time	0.025622	2	0.012811	38.87	< 0.0001
Copper	0.008956	2	0.004478	13.58	0.0003
Time*Copper	0.012756	4	0.003189	9.67	0.0002
Error	0.005933	18	0.000330		
Total	0.053267	26			

(b) The P-value < 0.0001. There is a significant time effect.

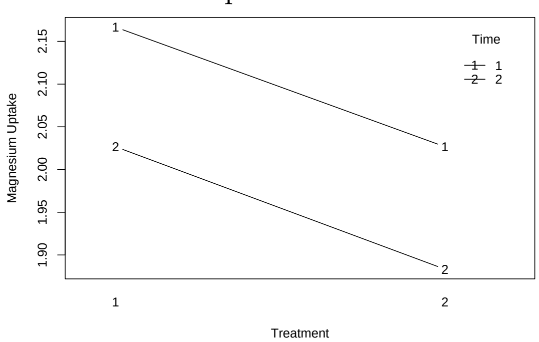
(c) The P-value $= 0.0003$. There is a significant copper effect.

(d) The interaction effect is significant since the P-value $= 0.0002$. The interaction plot is show here.



The algae concentrations for the various copper contents are all clearly influenced by the time effect shown in the graph.

14.13 (a) The interaction plot is show here. There seems no interaction effect.



(b) The ANOVA table is

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Treatment	0.060208	1	0.060208	157.07	< 0.0001
Time	0.060208	1	0.060208	157.07	< 0.0001
Treatment*Time	0.000008	1	0.000008	0.02	0.8864
Error	0.003067	8	0.000383		
Total	0.123492	11			

(c) The magnesium uptake are lower using treatment 2 than using treatment 1, no matter what the times are. Also, time 2 has lower magnesium uptake than time 1. All the main effects are significant.

(d) Using the regression model and making "Treatment" as categorical, we have the following fitted model:

$$\hat{y} = 2.4433 - 0.13667 \text{ Treatment} - 0.13667 \text{ Time} - 0.00333\text{Treatment} \times \text{Time}.$$

(e) The P-value of the interaction for the above regression model is 0.8864 and hence it is insignificant.

14.14 (a) A natural linear model with interaction would be

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2.$$

The fitted model would be

$$\hat{y} = 0.41772 - 0.06631 x_1 - 0.00866 x_2 + 0.00416 x_1 x_2,$$

with the P-values of the t-tests on each of the coefficients as 0.0092, 0.0379 and 0.0318 for x_1, x_2, and $x_1 x_2$, respectively. They are all significant at a level larger than 0.0379. Furthermore, $R^2_{\text{adj}} = 0.1788$.

(b) The new fitted model is

$$\hat{y} = 0.3368 - 0.15965 x_1 + 0.02684 x_2 + 0.00416 x_1 x_2 + 0.02333 x_1^2 - 0.00155 x_2^2,$$

with P-values of the t-tests on each of the coefficients as 0.0004, < 0.0001, 0.0003, 0.0156, and < 0.0001 for x_1, x_2. $x_1 x_2$, x_1^2, and x_2^2, respectively. Furthermore, $R^2_{\text{adj}} = 0.7700$ which is much higher than that of the model in (a). Model in (b) would be more appropriate.

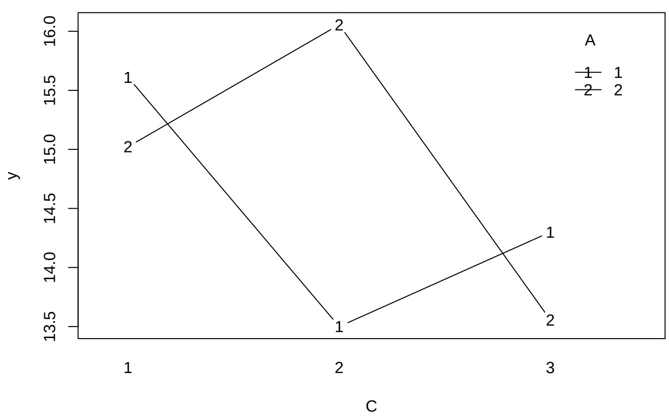
14.15 The ANOVA table is given here.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Main effect					
A	2.24074	1	2.24074	0.54	0.4652
B	56.31815	2	28.15907	6.85	0.0030
C	17.65148	2	8.82574	3.83	0.1316
Two-factor Interaction					
AB	31.47148	2	15.73574	3.83	0.0311
AC	31.20259	2	15.60130	3.79	0.0320
BC	2156074	4	5.39019	1.31	0.2845
Three-factor Interaction					
ABC	26.79852	4	6.69963	1.63	0.1881
Error	148.04000	36	4.11221		
Total	335.28370	53			

(a) Based on the P-values, only AB and AC interactions are significant.

(b) The main effect B is significant. However, due to significant interactions mentioned in (a), the insignificance of A and C cannot be counted.

(c) Look at the interaction plot of the mean responses versus C for different cases of A.



Apparently, the mean responses at different levels of C varies in different patterns for the different levels of A. Hence, although the overall test on factor C is insignificant, it is misleading since the significance of the effect C is masked by the significant interaction between A and C.

14.16 (a) When only A, B, C, and BC factors are in the model, the P-value for BC interaction is 0.0806. Hence at level of 0.05, the interaction is insignificant.

(b) When the sum of squares of the BC term is pooled with the sum of squares of the error, we increase the degrees of freedom of the error term. The P-values of the main effects of A, B, and C are 0.0275, 0.0224, and 0.0131, respectively. All these are significant.

14.17 Letting A, B, and C designate coating, humidity, and stress, respectively, the ANOVA table is given here.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Main effect					
A	216,384.1	1	216,384.1	0.05	0.8299
B	19,876,891.0	2	9,938,445.5	2.13	0.1257
C	427,993,946.4	2	213,996,973.2	45.96	< 0.0001
Two-factor Interaction					
AB	31,736,625	2	15,868,312.9	3.41	0.0385
AC	699,830.1	2	349,915.0	0.08	0.9277
BC	58,623,693.2	4	13,655,923.3	3.15	0.0192
Three-factor Interaction					
ABC	36,034,808.9	4	9,008,702.2	1.93	0.1138
Error	335,213,133.6	72	4,655,738.0		
Total	910,395,313.1	89			

(a) The Coating and Humidity interaction, and the Humidity and Stress interaction have the P-values of 0.0385 and 0.0192, respectively. Hence, they are all significant. On the other hand, the Stress main effect is strongly significant as well. However, both other main effects, Coating and Humidity, cannot be claimed as insignificant, since they are all part of the two significant interactions.

(b) A Stress level of 20 consistently produces low fatigue. It appears to work best with medium humidity and an uncoated surface.

14.18 The ANOVA table is given here.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
A	1.90591	3	0.63530	8.28	0.0003
B	0.02210	1	0.02212	0.29	0.5951
C	38.93402	1	38.93402	507.57	< 0.0001
AB	0.88632	3	0.29544	3.85	0.0185
AC	0.53594	3	0.17865	2.33	0.0931
BC	0.45435	1	0.45435	5.92	0.0207
ABC	0.42421	3	0.14140	1.84	0.1592
Error	2.45460	32	0.07671		
Total	45.61745	47			

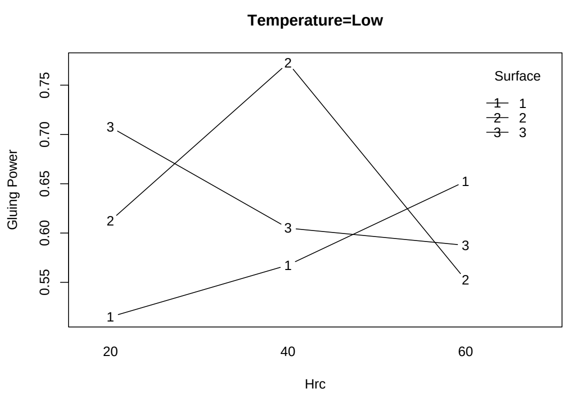
(a) Two-way interactions of AB and BC are all significant and main effect of A and C are all significant. The insignificance of the main effect B may not be valid due to the significant BC interaction.

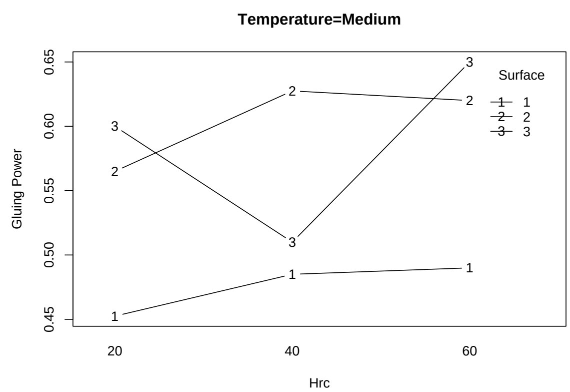
(b) Based on the P-values, Duncan's tests and the interaction means, the most important factor is C and using $C = 2$ is the most important way to increase percent weight. Also, using factor A at level 1 is the best.

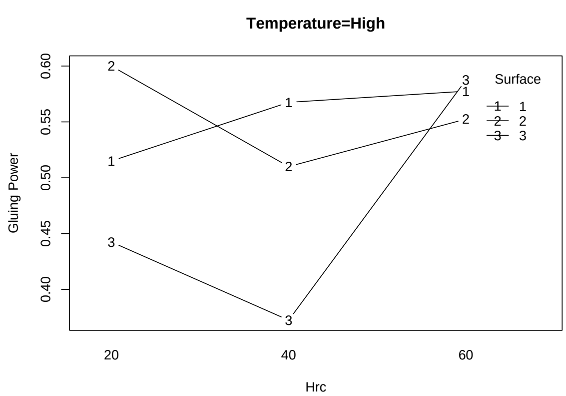
14.19 The ANOVA table shows:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
A	0.16617	2	0.08308	14.22	< 0.0001
B	0.07825	2	0.03913	6.70	0.0020
C	0.01947	2	0.00973	1.67	0.1954
AB	0.12845	4	0.03211	5.50	0.0006
AC	0.06280	4	0.01570	2.69	0.0369
BC	0.12644	4	0.03161	5.41	0.0007
ABC	0.14224	8	0.01765	3.02	0.0051
Error	0.47323	81	0.00584		
Total	1.19603	107			

There is a significant three-way interaction by Temperature, Surface, and Hrc. A plot of each Temperature is given to illustrate the interaction





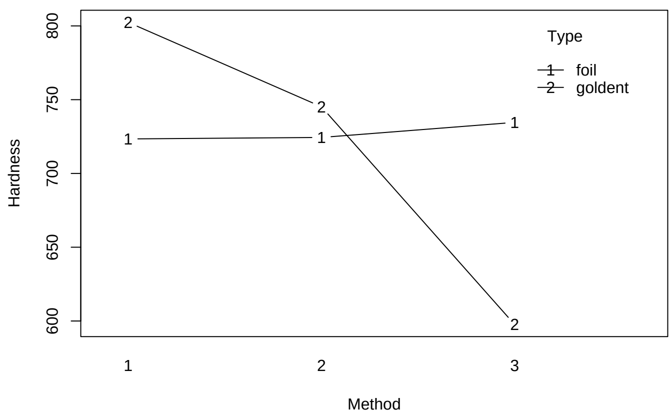


14.20 (a) $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\beta\gamma)_{jk} + \epsilon_{ijk}$; $\sum_j \beta_j = 0$, $\sum_k \gamma_k = 0$, $\sum_j (\beta\gamma)_{jk} = 0$, $\sum_k (\beta\gamma)_{jk} = 0$, and $\epsilon_{ijk} \sim n(x; 0, \sigma^2)$.

(b) The P-value of the Method and Type of Gold interaction is 0.10587. Hence, the interaction is at least marginally significant.

(c) The best method depends on the type of gold used.
The tests of the method effect for different type of gold yields the P-values as 0.9801 and 0.0099 for "Gold Foil" and "Goldent", respectively. Hence, the methods are significantly different for the "Goldent" type.

Here is an interaction plot.



It appears that when Type is "Goldent" and Method is 1, it yields the best hardness.

14.21 (a) Yes, the P-values for $Brand * Type$ and $Brand * Temp$ are both < 0.0001.

(b) The main effect of Brand has a P-value < 0.0001. So, three brands averaged across the other two factore are significantly different.

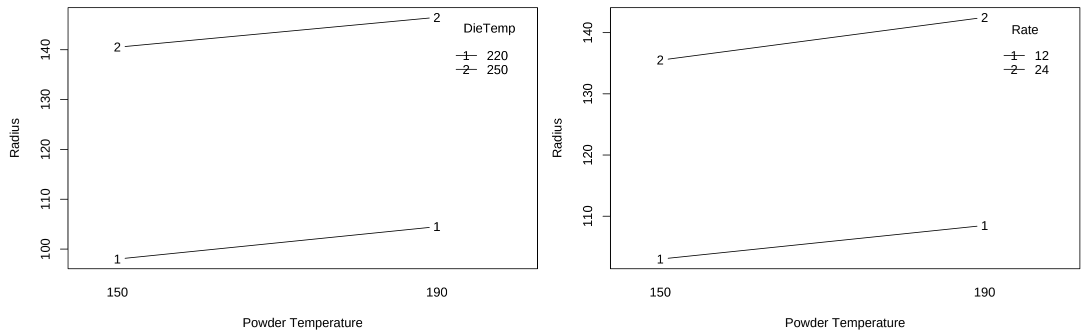
(c) Using brand Y, powdered detergent and hot water yields the highest percent removal of dirt.

14.22 (a) Define A, B, and C as "Powder Temperature," "Die Temperature," and "Extrusion Rate," respectively. The ANOVA table shows:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
A	78.125	1	78.125	125.00	0.0079
B	3570.125	1	3570.125	5712.20	0.0002
C	2211.125	1	2211.125	3537.80	0.0003
AB	0.125	1	0.125	0.20	0.6985
AC	1.125	1	1.125	1.80	0.3118
Error	1.25	2	0.625		
Total	5861.875	7			

The ANOVA results only show that the main effects are all significant and no two-way interactions are significant.

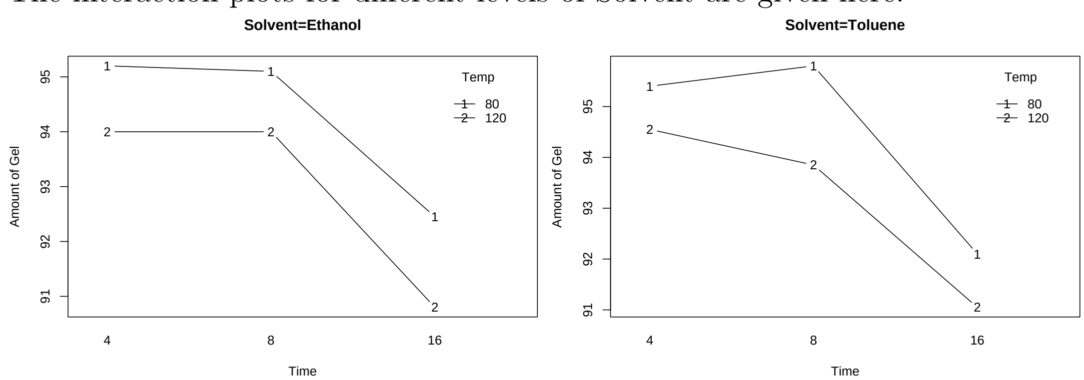
(b) The interaction plots are shown here.



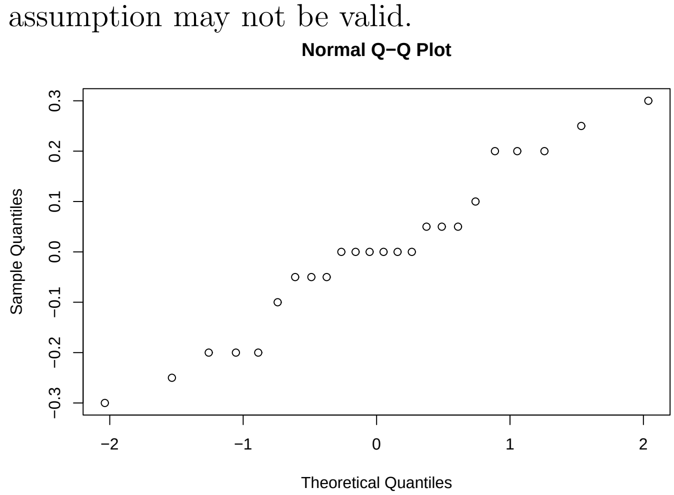
(c) The interaction plots in part (b) are consistent with the findings in part (a) that no two-way interactions present.

14.23 (a) The P-values of two-way interactions Time×Temperature, Time×Solvent, Temperature × Solvent, and the P-value of the three-way interaction Time×Temperature×Solvent are 0.1103, 0.1723, 0.8558, and 0.0140, respectively.

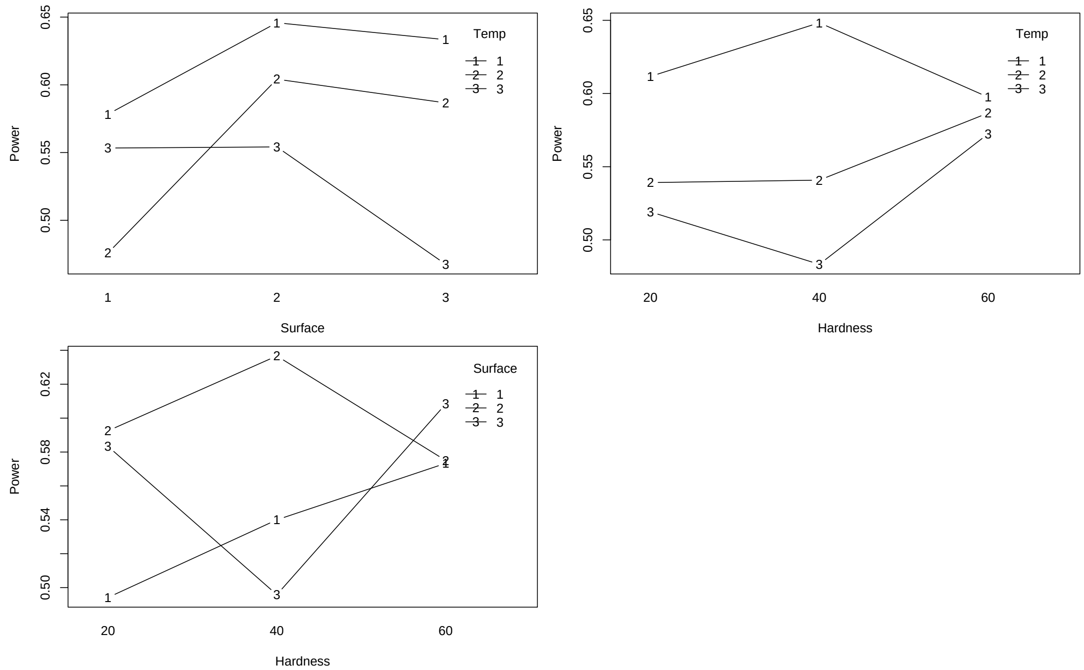
(b) The interaction plots for different levels of Solvent are given here.



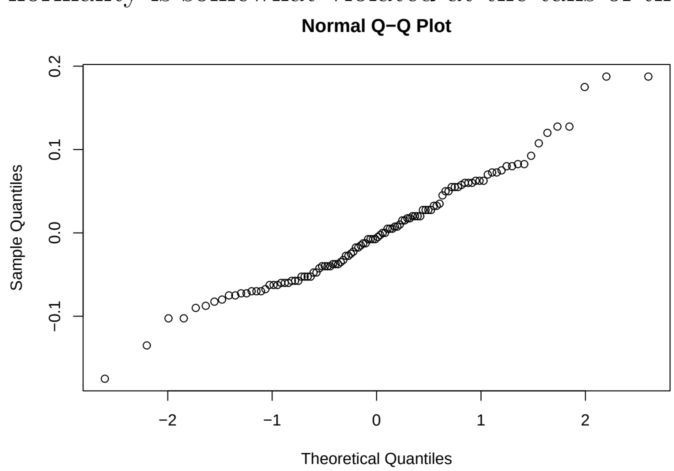
(c) A normal probability plot of the residuals is given and it appears that normality assumption may not be valid.



14.24 (a) The two-way interaction plots are given here and they all show significant interactions.



(b) The normal probability plot of the residuals is shown here and it appears that normality is somewhat violated at the tails of the distribution.



14.25 The ANOVA table is given.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Filters	4.63389	2	2.31694	8.39	0.0183
Operators	10.31778	3	3.43926	12.45	0.0055
Interaction	1.65722	6	0.27620	1.49	0.2229
Error	4.44000	24	0.18500		
Total	21.04889	35			

Note the f values for the main effects are using the interaction term as the denominator.

(a) The hypotheses are

$$H_0: \sigma^2_{\alpha\beta} = 0,$$
$$H_1: \sigma^2_{\alpha\beta} \neq 0.$$

Decision: Since P-value $= 0.2229$, the null hypothesis cannot be rejected. There is no significant interaction variance component.

(b) The hypotheses are

$$H_0^{'}: \sigma^2_{\alpha} = 0. \qquad H_0^{''}: \sigma^2_{\beta} = 0.$$
$$H_1^{'}: \sigma^2_{\alpha} \neq 0. \qquad H_1^{''}: \sigma^2_{\beta} \neq 0.$$

Decisions: Based on the P-values of 0.0183, and 0.0055 for $H_0^{'}$ and $H_1^{'}$, respectively, we reject both $H_0^{'}$ and $H_0^{''}$. Both σ^2_α and σ^2_β are significantly different from zero.

(c) $\hat{\sigma}^2 = s^2 = 0.185$; $\hat{\sigma}^2_\alpha = \frac{2.31691-0.185}{12} = 0.17766$, and $\hat{\sigma}^2_\beta = \frac{3.43926-0.185}{9} = 0.35158$.

14.26 $\hat{\sigma}^2_\alpha = \frac{(42.6289/2-14.8011/4}{12} = 1.4678$ Brand.
$\hat{\sigma}^2_\beta = \frac{(299.3422/2-14.8011/4}{12} = 12.1642$ Time.
$s^2 = 0.9237$.

14.27 The ANOVA table with expected mean squares is given here.

Source of Variation	Degrees of Freedom	Mean Square	(a) Computed f	(b) Computed f
A	3	140	$f_1 = s_1^2/s_5^2 = 5.83$	$f_1 = s_1^2/s_5^2 = 5.83$
B	1	480	$f_2 = s_2^2/s_{p_1}^2 = 78.82$	$f_2 = s_2^2/s_6^2 = 26.67$
C	2	325	$f_3 = s_3^2/s_5^2 = 13.54$	$f_3 = s_3^2/s_5^2 = 13.54$
AB	3	15	$f_4 = s_4^2/s_{p_2}^2 = 2.86$	$f_4 = s_4^2/s_7^2 = 7.50$
AC	6	24	$f_5 = s_5^2/s_{p_2}^2 = 4.57$	$f_5 = s_5^2/s_7^2 = 12.00$
BC	2	18	$f_6 = s_6^2/s_{p_1}^2 = 4.09$	$f_6 = s_6^2/s_7^2 = 9.00$
ABC	6	2	$f_7 = s_7^2/s^2 = 0.40$	$f_7 = s_7^2/s^2 = 0.40$
Error	24	5		
Total	47			

In column (a) we have found the following main effects and interaction effects significant using the pooled estimates: σ^2_β, σ^2_γ, and $\sigma^2_{\alpha\gamma}$.
$s^2_{p_1} = (12+120)/30 = 4.4$ with 30 degrees of freedom.
$s^2_{p_2} = (12+120+36)/32 = 5.25$ with 32 degrees of freedom.
$s^2_{p_3} = (12+120+36+45)/35 = 6.09$ with 35 degrees of freedom.

In column (b) we have found the following main effect and interaction effect significant when sums of squares of insignificant effects were not pooled: σ^2_γ and $\sigma^2_{\alpha\gamma}$.

14.28 $\sum_{i=1}^{4} \gamma_k^2 = 0.24$ and $\phi = \sqrt{\frac{(16)(0.24)}{(3)(0.197)}} = 2.55$. With $\alpha = 0.05$, $v_1 = 2$ and $v_2 = 39$ we find from A.16 that the power is approximately 0.975. Therefore, 2 observations for each treatment combination are sufficient.

14.29 The power can be calculated as

$$1 - \beta = P\left[F(2,6) > f_{0.05}(2,6)\frac{\sigma^2 + 3\sigma_{\alpha\beta}^2}{\sigma^2 + 3\sigma_{\alpha\beta}^2 + 12\sigma_\beta^2}\right]$$
$$= P\left[F(2,6) > \frac{(5.14)(0.2762)}{2.3169}\right] = P[F(2,6) > 0.6127] = 0.57.$$

14.30 (a) A mixed model. Inspectors (α_i in the model) are random effects. Inspection level (β_j in the model) is a fixed effect.

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk};$$
$$\alpha_i \sim n(x; 0, \sigma_\alpha^2), \quad (\alpha\beta)_{ij} \sim n(x; 0, \sigma_{\alpha\beta}^2), \quad \epsilon_{ijk} \sim n(x; 0, \sigma^2), \quad \sum_j \beta_j = 0.$$

(b) The hypotheses are

$$H_0: \sigma_\alpha^2 = 0. \qquad H_0: \sigma_{\alpha\beta}^2 = 0. \qquad H_0: \beta_1 = \beta_2 = \beta_3 = 0,$$
$$H_1: \sigma_\alpha^2 \neq 0. \qquad H_1: \sigma_{\alpha\beta}^2 \neq 0 \qquad H_1: \text{At least one } \beta_i\text{'s is not 0.}$$

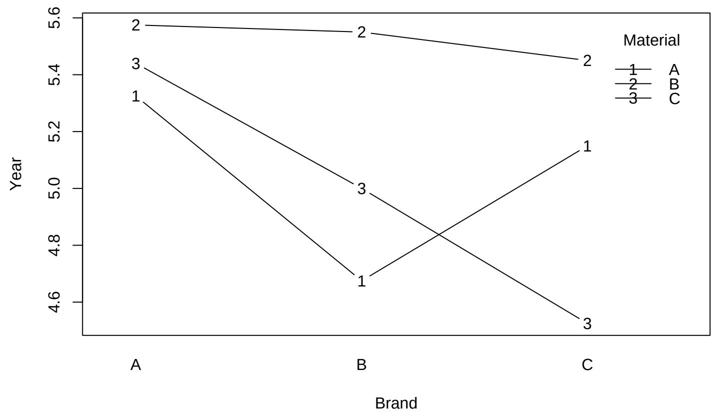
$f_{2,36} = 0.02$ with P-value $= 0.9806$. There does not appear to be an effect due to the inspector.
$f_{4,36} = 0.04$ with P-value $= 0.9973$. There does not appear to be an effect due to the inspector by inspector level.
$f_{2,4} = 54.77$ with P-value $= 0.0012$. Mean inspection levels were significantly different in determining failures per 1000 pieces.

14.31 (a) A mixed model.

(b) The ANOVA table is

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Material	1.03488	2	0.51744	47.42	< 0.0001
Brand	0.60654	2	0.30327	1.73	0.2875
Material*Brand	9, 70109	4	0.17527	16.06	0.0004
Error	0.09820	9	0.01091		
Total	2.44071	17			

(c) No, the main effect of Brand is not significant. An interaction plot is given.



Although brand A has highest means in general, it is not always significant, especially for Material 2.

14.32 (a) Operators (α_i) and time of day (β_j) are random effects.

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk};$$
$$\alpha_i \sim n(x;0,\sigma_\alpha^2), \quad \beta_j \sim n(x;0,\sigma_\beta^2), \quad (\alpha\beta)_{ij} \sim n(x;0,\sigma_{\alpha\beta}^2), \quad \epsilon_{ijk} \sim n(x;0,\sigma^2).$$

(b) $\sigma_\alpha^2 = \sigma_\beta^2 = 0$ (both estimates of the variance components were negative).

(c) The yield does not appear to depend on operator or time.

14.33 (a) A mixed model. Power setting (α_i in the model) is a fixed effect. Cereal type (β_j in the model) is a random effect.

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk};$$
$$\sum_i \alpha_i = 0, \quad \beta_j \sim n(x;0,\sigma_\beta^2), \quad (\alpha\beta)_{ij} \sim n(x;0,\sigma_{\alpha\beta}^2), \quad \epsilon_{ijk} \sim n(x;0,\sigma^2).$$

(b) No. $f_{2,4} = 1.37$ and P-value $= 0.3524$.

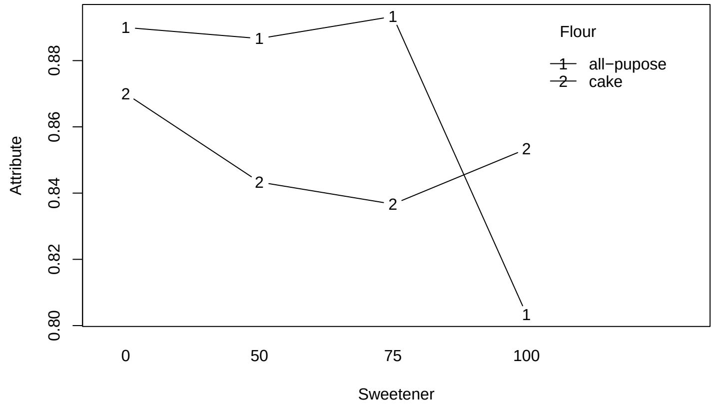
(c) No. The estimate of σ_β^2 is negative.

14.34 (a) The ANOVA table is given.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Sweetener	0.00871	3	0.00290	2.90	0.0670
Flour	0.00184	1	0.00184	1.84	0.1941
Interaction	0.01015	3	0.00338	3.38	0.0442
Error	0.01600	16	0.00100		
Total	0.03670	23			

Sweetener factor is close to be significant, while the P-value of the Flour shows insignificance. However, the interaction effects appear to be significant.

(b) Since the interaction is significant with a P-value $= 0.0442$, testing the effect of sweetener on the specific gravity of the cake samples by flour type we get P-value $= 0.0077$ for "All Purpose" flour and P-value $= 0.6059$ for "Cake" flour. We also have the interaction plot which shows that sweetener at 100% concentration yielded a specific gravity significantly lower than the other concentrations for all-purpose flour. For cake flour, however, there were no big differences in the effect of sweetener concentration.



14.35 (a) The ANOVA table is given.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Sauce	1,031.3603	1	1,031.3603	7.32	0.0123
Fish	16,505.8640	2	8,252.9320	58.58	< 0.0001
Sauce*Fish	724.6107	2	362.3053	2.57	0.0973
Error	3,381.1480	24	140.8812		
Total	21,642.9830	29			

Interaction effect is not significant.

(b) Both P-values of Sauce and Fish Type are all small enough to call significance.

14.36 (a) The ANOVA table is given here.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Plastic Type	142.6533	2	71.3267	16.92	0.0003
Humidity	143.7413	3	47.9138	11.36	0.0008
Interaction	133.9400	6	22.3233	5.29	0.0070
Error	50.5950	12	4.2163		
Total	470.9296	23			

The interaction is significant.

(b) The SS for AB with only Plastic Type A and B is 24.8900 with 3 degrees of freedom. Hence $f = \frac{24.8900/3}{4.2163} = 1.97$ with P-value $= 0.1727$. Hence, there is no significant interaction when only A and B are considered.

(c) The SS for the single-degree-of-freedom contrast is 143.0868. Hence $f = 33.94$ with P-value < 0.0001. Therefore, the contrast is significant.

(d) The SS for *Humidity* when only C is considered in Plastic Type is 2.10042. So, $f = 0.50$ with P-value $= 0.4938$. Hence, Humidity effect is insignificant when Type C is used.

14.37 (a) The ANOVA table is given here.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Environment	0.8624	1	0.8624	15.25	0.0021
Stress	40.8140	2	20.4020	360.94	< 0.0001
Interaction	0.0326	2	0.0163	0.29	0.7547
Error	0.6785	12	0.0565		
Total	42.3875	17			

The interaction is insignificant.

(b) The mean fatigue life for the two main effects are all significant.

14.38 The ANOVA table is given here.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Sweetener	1.26893	3	0.42298	2.65	0.0843
Flour	1.77127	1	1.77127	11.09	0.0042
Interaction	0.14647	3	0.04882	0.31	0.8209
Error	2.55547	16	0.15972		
Total	5.74213	23			

The interaction effect is insignificant. The main effect of Sweetener is somewhat insignificant, since the P-value $= 0.0843$. The main effect of Flour is strongly significant.

14.39 The ANOVA table is given here.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
	1133.5926	2	566.7963	8.30	0.0016
A	26896.2593	2	13448.1296	196.91	< 0.0001
B	40.1482	2	20.0741	0.29	0.7477
C	216.5185	4	54.1296	0.79	0.5403
AB	1.6296	4	0.4074	0.01	0.9999
AC	2.2963	4	0.5741	0.01	0.9999
Error	2.5926	8	0.3241	0.00	1.0000
	1844.0000	27	68.2963		
Total	30137.0370	53			

All the two-way and three-way interactions are insignificant. In the main effects, only A and B are significant.

14.40 (a) Treating Solvent as a class variable and Temperature and Time as continuous variable, only three terms in the ANOVA model show significance. They are (1) Intercept; (2) Coefficient for Temperature and (3) Coefficient for Time.

(b) Due to the factor that none of the interactions are significant, we can claim that the models for ethanol and toluene are equivalent apart from the intercept.

(c) The three-way interaction in Exercise 14.23 was significant. However, the general patterns of the gel generated are pretty similar for the two Solvent levels.

14.41 The ANOVA table is displayed.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Surface	2.22111	2	1.11056	0.02	0.9825
Pressure	39.10778	2	19.55389	0.31	0.7402
Interaction	112.62222	4	28.15556	0.45	0.7718
Error	565.72000	9	62.85778		
Total	719.67111	17			

All effects are insignificant.

14.42 (a) This is a two-factor fixed-effects model with interaction.

$$y_{ijk} = \mu + \alpha_i + \beta_j +)\alpha\beta)_{ij} + \epsilon_{ijk},$$
$$\sum_i \alpha_i = 0,\ \sum_j \beta_j = 0,\ \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0,\ \epsilon_{ijk} \sim n(x; 0, \sigma)$$

(b) The ANOVA table is displayed.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Time	0.16668	3	0.05556	48.67	< 0.0001
Temperature	0.27151	2	0.13575	118.91	< 0.0001
Interaction	0.03209	6	0.00535	4.68	0.0111
Error	0.01370	12	0.00114		
Total	0.48398	23			

The interaction is insignificant, while two main effects are significant.

(c) It appears that using a temperature of $-20°C$ with drying time of 2 hours would speed up the process and still yield a flavorful coffee. It might be useful to try some additional runs at this combination.

14.43 (a) Since it is more reasonable to assume the data come from Poisson distribution, it would be dangerous to use standard analysis of variance because the normality assumption would be violated. It would be better to transform the data to get at least stable variance.

(b) The ANOVA table is displayed.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Teller	40.45833	3	13.48611	6.35	0.0080
Time	97.33333	2	48.66667	22.90	< 0.0001
Interaction	8.66667	6	1.44444	0.68	0.6694
Error	25.50000	12	2.12500		
Total	171.95833	23			

The interaction effect is insignificant. Two main effects, Teller and Time, are all significant.

(c) The ANOVA table using a squared-root transformation on the response is given.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f	P-value
Teller	0.05254	3	0.17514	5.86	0.0106
Time	1.32190	2	0.66095	22.11	< 0.0001
Interaction	0.11502	6	0.01917	0.64	0.6965
Error	0.35876	12	0.02990		
Total	2.32110	23			

Same conclusions as in (b) can be reached. To check on whether the assumption of the standard analysis of variance is violated, residual analysis may used to do diagnostics.

Chapter 15

2^k Factorial Experiments and Fractions

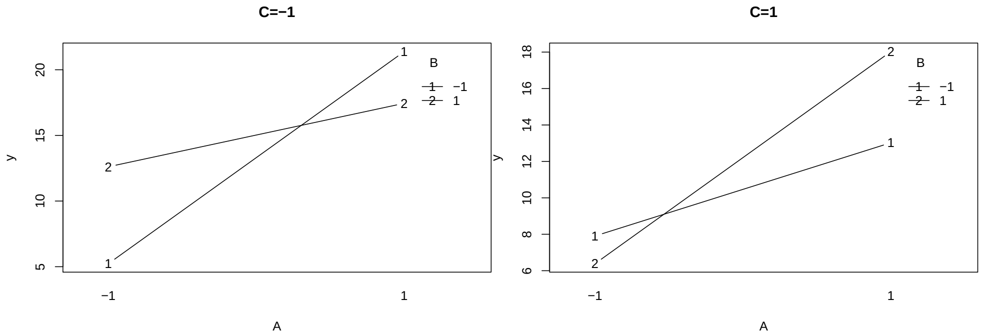
15.1 Either using Table 15.5 (e.g., $SSA = \frac{(-41+51-57-63+67+54-76+73)^2}{24} = 2.6667$) or running an analysis of variance, we can get the Sums of Squares for all the factorial effects.

$$SSA = 2.6667, \quad SSB = 170.6667, \quad SSC = 104.1667, \quad SS(AB) = 1.500.$$
$$SS(AC) = 42.6667, \quad SS(BC) = 0.0000, \quad SS(ABC) = 1.5000.$$

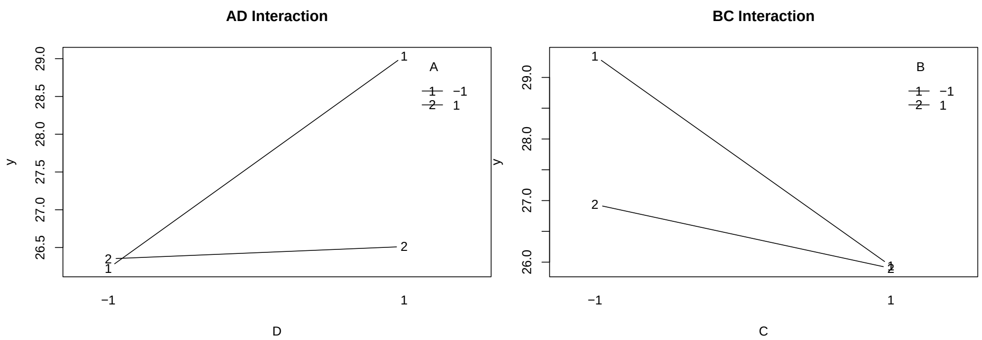
15.2 A simplified ANOVA table is given.

Source of Variation	Degrees of Freedom	Computed f	P-value
A	1	1294.65	< 0.0001
B	1	43.56	0.0002
AB	1	20.88	0.0018
C	1	116.49	< 0.0001
AC	1	16.21	0.0038
BC	1	0.00	0.9668
ABC	1	289.23	< 0.0001
Error	8		
Total	15		

All the main and interaction effects are significant, other than BC effect. However, due to the significance of the 3-way interaction, the insignificance of BC effect cannot be counted. Interaction plots are given.



15.3 The AD and BC interaction plots are printed here. The AD plot varies with levels of C since the ACD interaction is significant, or with levels of B since ABD interaction is significant.



15.4 The ANOVA table is displayed.

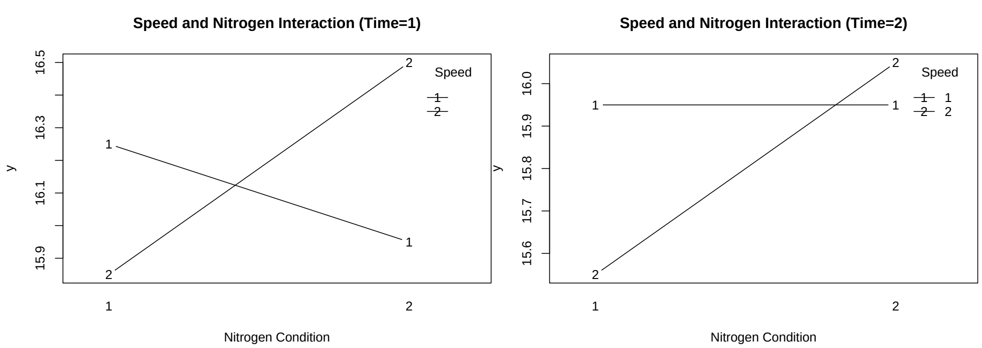
Source of Variation	Degrees of Freedom	Computed f	P-value
A	1	57.85	< 0.0001
B	1	7.52	0.0145
AB	1	6.94	0.0180
C	1	127.86	< 0.0001
AC	1	7.08	0.0171
BC	1	10.96	0.0044
ABC	1	1.26	0.2787
D	1	44.72	< 0.0001
AD	1	4.85	0.0427
BD	1	4.85	0.0427
ABD	1	1.14	0.3017
CD	1	6.52	0.0213
ACD	1	1.72	0.2085
BCD	1	1.20	0.2900
ABCD	1	0.87	0.3651
Error	16		
Total	31		

All main effects and two-way interactions are significant, while all higher order interactions are insignificant.

15.5 The ANOVA table is displayed.

Source of Variation	Degrees of Freedom	Computed f	P-value
A	1	9.98	0.0251
B	1	0.20	0.6707
C	1	6.54	0.0508
D	1	0.02	0.8863
AB	1	1.83	0.2338
AC	1	0.20	0.6707
AD	1	0.57	0.4859
BC	1	19.03	0.0073
BD	1	1.83	0.2338
CD	1	0.02	0.8863
Error	5		
Total	15		

One two-factor interaction BC, which is the interaction of Blade Speed and Condition of Nitrogen, is significant. As of the main effects, Mixing time (A) and Nitrogen Condition (C) are significant. Since BC is significant, the insignificant main effect B, the Blade Speed, cannot be declared insignificant. Interaction plots for BC at different levels of A are given here.



15.6 (a) The three effects are given as

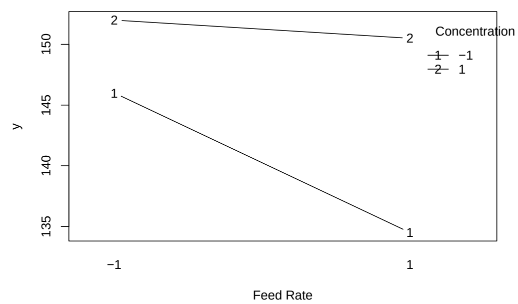
$$w_A = \frac{301 + 304 - 269 - 292}{4} = 11, \quad w_B = \frac{301 + 269 - 304 - 292}{4} = -6.5,$$
$$w_{AB} = \frac{301 - 304 - 269 + 292}{4} = 5.$$

There are no clear interpretation at this time.

(b) The ANOVA table is displayed.

Source of Variation	Degrees of Freedom	Computed f	P-value
Concentration	1	35.85	0.0039
Feed Rate	1	12.52	0.0241
Interaction	1	7.41	0.0529
Error	4		
Total	7		

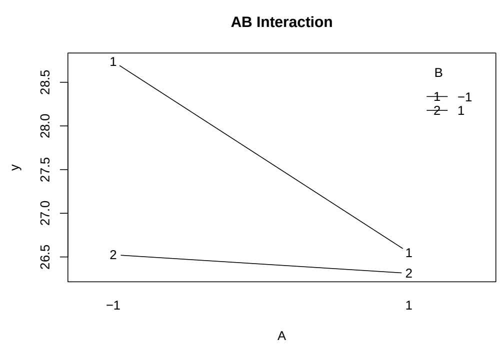
The interaction between the Feed Rate and Concentration is closed to be significant at 0.0529 level. An interaction plot is given here.



The mean viscosity does not change much at high level of concentration, while it changes a lot at low concentration.

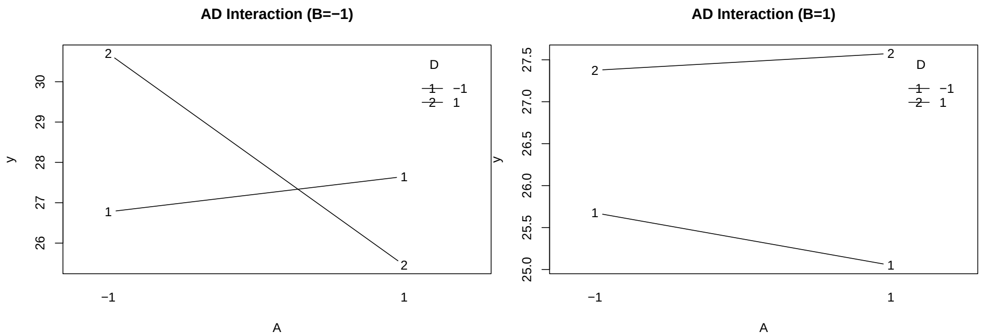
(c) Both main effects are significant. Averaged across Feed Rate a high concentration of reagent yields significantly higher viscosity, and averaged across concentration a low level of Feed Rate yields a higher level of viscosity.

15.7 Both AD and BC interaction plots are shown in Exercise 15.3. Here is the interaction plot of AB.



For AD, at the high level of A, Factor D essentially has no effect, but at the low level of A, D has a strong positive effect. For BC, at the low level of B, Factor C has a strong negative effect, but at the high level of B, the negative effect of C is not as pronounced. For AB, at the high level of B, A clearly has no effect. At the low level of B, A has a strong negative effect.

15.8 The two interaction plots are displayed.



It can be argued that when $B = 1$ that there is essentially no interaction between A and D. Clearly when $B = -1$, the presence of a high level of D produces a strong negative effect of Factor A on the response.

15.9 (a) The parameter estimates for x_1, x_2 and x_1x_2 are given as follows.

Variable	Degrees of Freedom	Estimate	f	P-value
x_1	1	5.50	5.99	0.0039
x_2	1	−3.25	−3.54	0.0241
x_1x_2	1	2.50	2.72	0.0529

(b) The coefficients of b_1, b_2, and b_{12} are $w_A/2$, $w_B/2$, and $w_{AB}/2$, respectively.

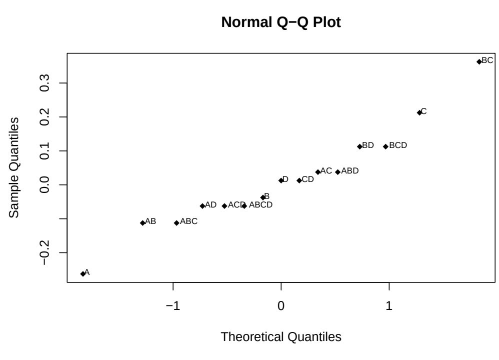
(c) The P-values are matched exactly.

15.10 The effects are given here.

A	B	C	D	AB	AC	AD	BC
−0.2625	−0.0375	0.2125	0.0125	−0.1125	0.0375	−0.0625	0.3625

BD	CD	ABC	ABD	ACD	BCD	$ABCD$
0.1125	0.0125	−0.1125	0.0375	−0.0625	0.1125	−0.0625

The normal probability plot of the effects is displayed.



(a) It appears that all three- and four-factor interactions are not significant.

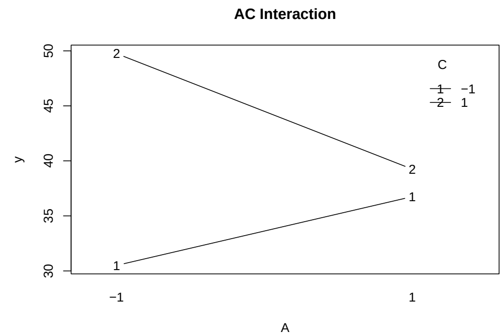
(b) From the plot, it appears that A and BC are significant and C is somewhat significant.

15.11 (a) The effects are given here and it appears that B, C, and AC are all important.

A	B	C	AB	AC	BC	ABC
-0.875	5.875	9.625	-3.375	-9.625	0.125	-1.125

(b) The ANOVA table is given.

Source of Variation	Degrees of Freedom	Computed f	P-value
A	1	0.11	0.7528
B	1	4.79	0.0600
AB	1	12.86	0.0071
C	1	1.58	0.2440
AC	1	12.86	0.0071
BC	1	0.00	0.9640
ABC	1	0.18	0.6861
Error	8		
Total	15		



(c) Yes, they do agree.

(d) For the low level of Cutting Angle, C, Cutting Speed, A, has a positive effect on the life of a machine tool. When the Cutting Angle is large, Cutting Speed has a negative effect.

15.12 A is not orthogonal to BC, B is not orthogonal to AC, and C is not orthogonal to AB. If we assume that interactions are negligible, we may use this experiment to estimate the main effects. Using the data, the effects can be obtained as A: 1.5; B: -6.5; C: 2.5. Hence Factor B, Tool Geometry, seems more significant than the other two factors.

15.13 Here is the block arrangement.

Block 1	Block 2	Block 1	Block 2	Block 1	Block 2
(1)	a	(1)	a	(1)	a
c	b	c	b	c	b
ab	ac	ab	ac	ab	ac
abc	bc	abc	bc	abc	bc
Replicate 1		Replicate 2		Replicate 3	
AB Confounded		AB Confounded		AB Confounded	

Analysis of Variance

Source of Variation	Degrees of Freedom
Blocks	5
A	1
B	1
C	1
AC	1
BC	1
ABC	1
Error	12
Total	23

15.14 (a) ABC is confounded with blocks in the first replication and $ABCD$ is confounded with blocks in second replication.

(b) Computing the sums of squares by the contrast method yields the following ANOVA table.

Source of Variation	Degrees of Freedom	Mean Square	Computed f	P-value
Blocks	3	2.32		
A	1	2.00	3.34	0.0907
B	1	0.50	0.83	0.3775
C	1	4.50	7.51	0.0168
D	1	8.00	13.36	0.0029
AB	1	0.50	0.83	0.3775
AC	1	0.32	0.53	0.4778
BC	1	0.50	0.83	0.3775
AD	1	0.72	1.20	0.2928
BD	1	0.32	0.53	0.4778
CD	1	0.18	0.30	0.5928
ABC	1	1.16	1.93	0.1882
ABD	1	0.32	0.53	0.4778
ACD	1	0.02	0.03	0.8578
BCD	1	0.18	0.30	0.5928
ABCD	1	0.53	0.88	0.3659
Error	13	0.60		
Total	31			

Only the main effects C and D are significant.

15.15 $L_1 = \gamma_1 + \gamma_2 + \gamma_3$ and $L_2 = \gamma_1 + \gamma_2 + \gamma_4$. For treatment combination (1) we find $L_1 \pmod 2 = 0$. For treatment combination a we find $L_1 \pmod 2 = 1$ and $L_2 \pmod 2 = 1$. After evaluating L_1 and L_2 for all 16 treatment combinations we obtain the following blocking scheme:

Block 1	Block 2	Block 3	Block 4
(1)	c	d	a
ab	abc	ac	b
acd	ad	bc	cd
bcd	bd	abd	$abcd$
$L_1 = 0$	$L_1 = 1$	$L_1 = 0$	$L_1 = 1$
$L_2 = 0$	$L_2 = 0$	$L_2 = 1$	$L_2 = 1$

Since $(ABC)(ABD) = A^2B^2CD = CD$ (mod 2), then CD is the other effect confounded.

15.16 (a) $L_1 = \gamma_1 + \gamma_2 + \gamma_4 + \gamma_5$, $L_2 = \gamma_1 + \gamma_5$. We find that the following treatment combinations are in the principal block ($L_1 = 0$, $L_2 = 0$): $(1), c, ae, bd, ace, abde, abcde$. The other blocks are constructed by multiplying the treatment combinations in the principal block modulo 2 by a, b, and ab, respectively, to give the following blocking arrangement:

Block 1	Block 2	Block 3	Block 4
(1)	a	b	ab
c	ac	bc	abc
ae	e	abe	bce
bd	abd	d	ad
ace	ce	$abce$	bce
bcd	$abcd$	cd	acd
$abde$	bde	ade	de
$abcde$	$bcde$	$acde$	cde

(b) $(ABDE)(AE) = BD$ (mod 2). Therefore BD is also confounded with days.

(c) Yates' technique gives the following sums of squares for the main effects:

$$SSA = 21.9453, \quad SSB = 40.2753, \quad SSC = 2.4753,$$
$$SSD = 7.7028, \quad SSE = 1.0878.$$

15.17 $L_1 = \gamma_1 + \gamma_2 + \gamma_3$, $L_2 = \gamma_1 + \gamma_2$.

Block		Block		Block	
1	2	1	2	1	2
abc	ab	abc	ab	(1)	a
a	ac	a	ac	c	b
b	bc	b	bc	ab	ac
c	(1)	c	(1)	abc	bc
Rep 1		Rep 2		Rep 3	
ABC Confounded		ABC Confounded		AB Confounded	

For treatment combination (1) we find L_1 (mod 2) = 0 and L_2 (mod 2) = 0. For treatment combination a we find L_1 (mod 2) = 1 and L_2 (mod 2) = 1. Replicate 1 and Replicate 2 have $L_1 = 0$ in one block and $L_1 = 1$ in the other. Replicate 3 has $L_2 = 0$ in one block and $L_2 = 1$ in the other.

Analysis of Variance

Source of Variation	Degrees of Freedom
Blocks	5
A	1
B	1
C	1
AB	$1'$
AC	1
BC	1
ABC	$1'$
Error	11
Total	23

Relative information on $ABC = \frac{1}{3}$ and relative information on $AB = \frac{2}{3}$.

15.18 (a) The ANOVA table is shown here.

Source of Variation	Degrees of Freedom	Mean Square	Computed f	P-value
Operators	1	0.1225		
A	1	4.4100	0.04	0.2413
B	1	3.6100	1.55	0.2861
C	1	9.9225	1.27	0.0912
D	1	2.2500	3.49	0.3945
Error	10	2.8423	0.79	
Total	15			

None of the main effects is significant at 0.05 level.

(b) ABC is confounded with operators since all treatments with positive signs in the ABC contrast are in one block and those with negative signs are in the other block.

15.19 (a) One possible design would be:

Machine								
1	(1)	ab	ce	$abce$	acd	bde	ade	bcd
2	a	b	ace	bce	cd	$abde$	de	$abcd$
3	c	abc	e	abe	ad	$bcde$	$acde$	bd
4	d	abd	cde	$abcde$	ac	be	ae	bc

(b) ABD, CDE, and $ABCE$.

15.20 (a) $\hat{y} = 43.9 + 1.625x_1 - 8.625x_2 + 0.375x_3 + 9.125x_1x_2 + 0.625x_1x_3 + 0.875x_2x_3$.

(b) The Lack-of-fit test results in a P-value of 0.0493. There are possible quadratic terms missing in the model.

15.21 (a) The P-values of the regression coefficients are:

Parameter	Intercept	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$
P-value	< 0.0001	0.5054	0.0772	0.0570	0.0125	0.0205	0.7984	0.6161

and $s^2 = 0.57487$ with 4 degrees of freedom. So x_2, x_3, x_1x_2 and x_1x_3 are important in the model.

(b) $t = \frac{\bar{y}_f - \bar{y}_C}{\sqrt{s^2(1/n_f + 1/n_C)}} = \frac{52.075 - 49.275}{\sqrt{(0.57487)(1/8+1/4)}} = 6.0306$. Hence the P-value $= 0.0038$ for testing quadratic curvature. It is significant.

(c) Need one additional design point different from the original ones.

15.22 (a) No.

(b) It could be as follows.

Machine								
1	(1)	*ad*	*bc*	*abce*	*acd*	*abd*	*cde*	*bde*
2	*a*	*e*	*abc*	*bce*	*cd*	*bd*	*acde*	*abde*
3	*b*	*abe*	*c*	*ace*	*abcd*	*ad*	*bcde*	*de*
4	*d*	*ade*	*bcd*	*abcde*	*ac*	*ab*	*ce*	*be*

ADE, BCD and $ABCE$ are confounded with blocks.

(c) Partial confounding.

15.23 To estimate the quadratic terms, it might be good to add points in the middle of the edges. Hence $(-1, 0)$, $(0, -1)$, $(1, 0)$, and $(0, 1)$ might be added.

15.24 The alias for each effect is obtained by multiplying each effect by the defining contrast and reducing the exponents modulo 2.

$$\begin{array}{lllll} A \equiv CDE, & AB \equiv BCDE, & BD \equiv ABCE, & B \equiv ABCDE, & AC \equiv DE, \\ BE \equiv ABCD, & C \equiv ADE, & AD \equiv CE, & ABC \equiv BDE, & D \equiv ACE, \\ AE \equiv CD, & ABD \equiv BCE, & E \equiv ACD, & BC \equiv ABDE, & ABE \equiv BCD, \end{array}$$

15.25 (a) With BCD as the defining contrast, we have $L = \gamma_2 + \gamma_3 + \gamma_4$. The $\frac{1}{2}$ fraction corresponding to $L = 0$ (mod 2 is the principal block: $\{(1), a, bc, abc, bd, abd, cd, acd\}$.

(b) To obtain 2 blocks for the $\frac{1}{2}$ fraction the interaction ABC is confounded using $L = \gamma_1 + \gamma_2 + \gamma_3$:

Block 1	Block 2
(1)	a
bc	abc
abd	bd
acd	cd

(c) Using BCD as the defining contrast we have the following aliases:

$$\begin{array}{llll} A{\equiv}ABCD, & AB{\equiv}ACD, & B{\equiv}CD, & AC{\equiv}ABD, \\ C{\equiv}BD, & AD{\equiv}ABC, & D{\equiv}BC. & \end{array}$$

Since AD and ABC are confounded with blocks there are only 2 degrees of freedom for error from the unconfounded interactions.

Analysis of Variance

Source of Variation	Degrees of Freedom
Blocks	1
A	1
B	1
C	1
D	1
Error	2
Total	7

15.26 With $ABCD$ and $BDEF$ as defining contrasts, we have

$$L_1 = \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4, \qquad L_2 = \gamma_2 + \gamma_4 + \gamma_5 + \gamma_6.$$

The following treatment combinations give $L_1 = 0,\ L_2 = 0$ (mod 2) and thereby suffice as the $\frac{1}{4}$ fraction:

$$\{(1), ac, bd, abcd, abe, bce, ade, abf, bcf, adf, cdf, ef, acef, bdef, abcdef\}.$$

The third defining contrast is given by

$$(ABCD)(BDEF) = AB^2CD^2EF = ACEF \text{ (mod 2)}.$$

The effects that are aliased with the six main effects are:

$$\begin{array}{llll} A{\equiv}BCD & {\equiv}ABDEF{\equiv}CEF, & B{\equiv}ACD & {\equiv}DEF{\equiv}ABCEF, \\ C{\equiv}ABD & {\equiv}BCDEF{\equiv}AEF, & D{\equiv}ABC & {\equiv}BEF{\equiv}ACDEF, \\ E{\equiv}ABCDE{\equiv}BDF & {\equiv}ACF, & F{\equiv}ABCDF{\equiv}BDE{\equiv}ACE. & \end{array}$$

15.27 (a) With $ABCE$ and $ABDF$, and hence $(ABCE)(ABDF) = CDEF$ as the defining contrasts, we have

$$L_1 = \gamma_1 + \gamma_2 + \gamma_3 + \gamma_5, \qquad L_2 = \gamma_1 + \gamma_2 + \gamma_4 + \gamma_6.$$

The principal block, for which $L_1 = 0$, and $L_2 = 0$, is as follows:

$$\{(1), ab, acd, bcd, ce, abce, ade, bde, acf, bcf, df, abdf, aef, bef, cdef, abcdef\}.$$

(b) The aliases for each effect are obtained by multiplying each effect by the three defining contrasts and reducing the exponents modulo 2.

$$\begin{array}{llllllll}
A & \equiv BCE & \equiv BDF & \equiv ACDEF, & B & \equiv ACE & \equiv ADF & \equiv BCDEF, \\
C & \equiv ABE & \equiv ABCDF & \equiv DEF, & D & \equiv ABCDE & \equiv ABF & \equiv CEF, \\
E & \equiv ABC & \equiv ABDEF & \equiv CDF, & F & \equiv ABCEF & \equiv ABD & \equiv CDE, \\
AB & \equiv CE & \equiv DF & \equiv ABCDEF, & AC & \equiv BE & \equiv BCDF & \equiv ADEF, \\
AD & \equiv BCDE & \equiv BF & \equiv ACEF, & AE & \equiv BC & \equiv BDEF & \equiv ACDF, \\
AF & \equiv BCEF & \equiv BD & \equiv ACDE, & CD & \equiv ABDE & \equiv ABCF & \equiv EF, \\
DE & \equiv ABCD & \equiv ABEF & \equiv CF, & BCD & \equiv ADE & \equiv ACF & \equiv BEF, \\
DCF & \equiv AEF & \equiv ACD & \equiv BDE, & & & & .
\end{array}$$

Since E and F do not interact and all three-factor and higher interactions are negligible, we obtain the following ANOVA table:

Source of Variation	Degrees of Freedom
A	1
B	1
C	1
D	1
E	1
F	1
AB	1
AC	1
AD	1
BC	1
BD	1
CD	1
Error	3
Total	15

15.28 The ANOVA table is shown here and the error term is computed by pooling all the interaction effects. Factor E is the only significant effect, at level 0.05, although the decision on factor G is marginal.

Source of Variation	Degrees of Freedom	Mean Square	Computed f	P-value
A	1	1.44	0.48	0.5060
B	1	4.00	1.35	0.2793
C	1	9.00	3.03	0.1199
D	1	5.76	1.94	0.2012
E	1	16.00	5.39	0.0488
F	1	3.24	1.09	0.3268
G	1	12.96	4.36	0.0701
Error	8	2.97		
Total	15			

15.29 All two-factor interactions are aliased with each other. So, assuming that two-factor as well as higher order interactions are negligible, a test on the main effects is given in the ANOVA table.

Source of Variation	Degrees of Freedom	Mean Square	Computed f	P-value
A	1	6.125	5.81	0.0949
B	1	0.605	0.57	0.5036
C	1	4.805	4.56	0.1223
D	1	0.245	0.23	0.6626
Error	3	1.053		
Total	7			

Apparently no main effects is significant at level 0.05. Comparatively factors A and C are more significant than the other two. Note that the degrees of freedom on the error term is only 3, the test is not very powerful.

15.30 Two-factor interactions are aliased with each other. There are total 7 two-factor interactions that can be estimated. Among those 7, we picked the three, which are AC, AF, and BD, that have largest SS values and pool the other 2-way interactions to the error term. An ANOVA can be obtained.

Source of Variation	Degrees of Freedom	Mean Square	Computed f	P-value
A	1	81.54	0.37	0.5643
B	1	166.54	0.76	0.4169
C	1	5.64	0.03	0.8778
D	1	4.41	0.02	0.8918
E	1	40.20	0.18	0.6834
F	1	1678.54	7.66	0.0325
AC	1	978.75	4.47	0.0790
AF	1	625.00	2.85	0.1423
BD	1	429.53	1.96	0.2111
Error	6	219.18		
Total	15			

Main effect F, the location of detection, appears to be the only significant effect. The AC interaction, which is aliased with BE, has a P-value closed to 0.05.

15.31 To get all main effects and two-way interactions in the model, this is a saturated design, with no degrees of freedom left for error. Hence, we first get all SS of these effects and pick the 2-way interactions with large SS values, which are AD, AE, BD and BE. An ANOVA table is obtained.

Source of Variation	Degrees of Freedom	Mean Square	Computed f	P-value
A	1	$388,129.00$	$3,585.49$	< 0.0001
B	1	$277,202.25$	$2,560.76$	< 0.0001
C	1	$4,692.25$	43.35	0.0006
D	1	$9,702.25$	89.63	< 0.0001
E	1	$1,806.25$	16.69	0.0065
AD	1	$1,406.25$	12.99	0.0113
AE	1	462.25	4.27	0.0843
BD	1	$1,156.25$	10.68	0.0171
BE	1	961.00	8.88	0.0247
Error	6	108.25		
Total	15			

All main effects, plus AD, BD and BE two-way interactions, are significant at 0.05 level.

15.32 Consider a 2^4 design with letters A, B, C, and D, with design points

$$\{(1), a, b, c, d, ab, ac, ad, bc, bd, cd, abc, abd, acd, bcd, abcd\}$$

. Using $E = ABCD$, we have the following design:

$$\{e, a, b, c, d, abe, ace, ade, bce, bde, cde, abc, abd, acd, bcd, abcde\}.$$

15.33 Begin with a 2^3 with design points

$$\{(1), a, b, c, ab, ac, bc, abc\}.$$

Now, use the generator $D = AB$, $E = AC$, and $F = BC$. We have the following result:

$$\{def, af, be, cd, abd, ace, bcf, abcdef\}.$$

15.34 We can use the $D = AB$, $E = -AC$ and $F = BC$ as generators and obtain the result:

$$\{df, aef, b, cde, abde, ac, bcef, abcdf\}.$$

15.35 Here are all the aliases

$$\begin{aligned}
&A{\equiv}BD{\equiv}CE{\equiv}CDF{\equiv}BEF{\equiv}\equiv ABCF \equiv ADEF \equiv ABCDE;\\
&B{\equiv}AD{\equiv}CF{\equiv}CDE{\equiv}AEF{\equiv}\equiv ABCE \equiv BDEF \equiv ABCDF;\\
&C{\equiv}AE{\equiv}BF{\equiv}BDE{\equiv}ADF{\equiv}\equiv CDEF \equiv ABCD \equiv ABCEF;\\
&D{\equiv}AB{\equiv}EF{\equiv}BCE{\equiv}ACF{\equiv}\equiv BCDF \equiv ACDE \equiv ABDEF;\\
&E{\equiv}AC{\equiv}DF{\equiv}ABF{\equiv}BCD{\equiv}\equiv ABDE \equiv BCEF \equiv ACDEF;\\
&F{\equiv}BC{\equiv}DE{\equiv}ACD{\equiv}ABE{\equiv}\equiv ACEF \equiv ABDF \equiv BCDEF.
\end{aligned}$$

15.36 (a) The defining relation is $ABC = -I$.

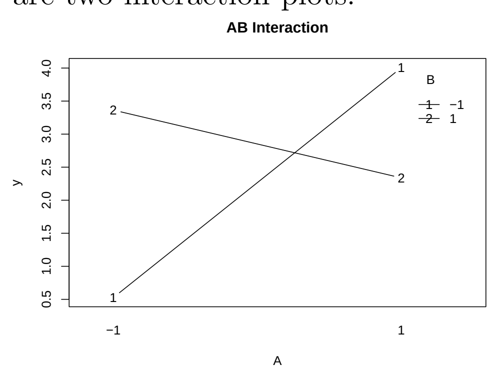
(b) $A = -BC$, $B = -AC$, and $C = -AB$.

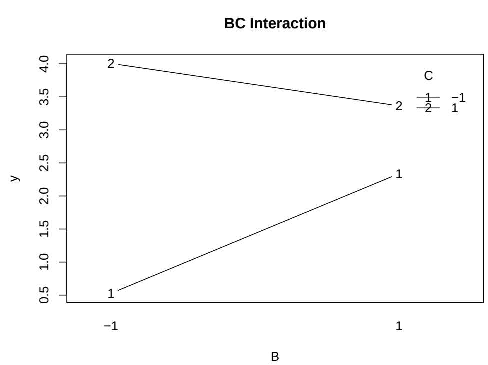
(c) The mean squares for A, B, and C are 1.50, 0.34, and 5.07, respectively. So, factor C, the amount of grain refiner, appears to be most important.

(d) Low level of C.

(e) All at the "low" level.

(f) A hazard here is that the interactions may play significant roles. The following are two interaction plots.





15.37 When the variables are centered and scaled, the fitted model is

$$\hat{y} = 12.7519 + 4.7194x_1 + 0.8656x_2 - 1.4156x_3.$$

The lack-of-fit test results in an f-value of 81.58 with P-value < 0.0001. Hence, higher-order terms are needed in the model.

15.38 The ANOVA table for the regression model looks like the following.

Coefficients	Degrees of Freedom
Intercept	1
β_1	1
β_2	1
β_3	1
β_4	1
β_5	1
Two-factor interactions	10
Lack of fit	16
Pure error	32
Total	63

15.39 The defining contrasts are

AFG, $CEFG$, $ACDF$, BEG, $BDFG$, CDG, $BCDE$, $ABCDEFG$, DEF, $ADEG$.

15.40 Begin with the basic line for $N = 24$; permute as described in Section 15.12 until 18 columns are formed.

15.41 The fitted model is

$$\begin{aligned}\hat{y} =& 190,056.67 + 181,343.33x_1 + 40,395.00x_2 + 16,133.67x_3 + 45,593.67x_4 \\ &- 29,412.33x_5 + 8,405.00x_6.\end{aligned}$$

The t-tests are given as

Variable	t	P-value
Intercept	4.48	0.0065
x_1	4.27	0.0079
x_2	0.95	0.3852
x_3	0.38	0.7196
x_4	1.07	0.3321
x_5	−0.69	0.5194
x_6	0.20	0.8509

Only x_1 and x_2 are significant.

15.42 An ANOVA table is obtained.

Source of Variation	Degrees of Freedom	Mean Square	Computed f	P-value
Polymer 1	1	172.98	1048.36	< 0.0001
Polymer 2	1	180.50	1093.94	< 0.0001
Polymer 1*Polymer 2	1	1.62	9.82	0.0351
Error	4	0.17		
Total	7			

All main effects and interactions are significant.

15.43 An ANOVA table is obtained.

Source of Variation	Degrees of Freedom	Mean Square	Computed f	P-value
Mode	1	$2,054.36$	74.25	< 0.0001
Type	1	$4,805.96$	173.71	< 0.0001
Mode*Type	1	482.90	17.45	0.0013
Error	12	27.67		
Total	15			

All main effects and interactions are significant.

15.44 Two factors at two levels each can be used with three replications of the experiment, giving 12 observations. The requirement that there must be tests on main effects and the interactions suggests that partial confounding be used The following design is indicated:

	Rep 1		Rep 2		Rep 3	
Block	1	2	1	2	1	2
	(1)	a	a	(1)	(1)	ab
	ab	b	ab	b	a	a

15.45 Using the contrast method and compute sums of squares, we have

Source of Variation	d.f.	MS	f
A	1	0.0248	1.24
B	1	0.0322	1.61
C	1	0.0234	1.17
D	1	0.0676	3.37
E	1	0.0028	0.14
F	1	0.0006	0.03
Error	8	0.0201	

15.46 With the defining contrasts $ABCD$, $CDEFG$, and BDF, we have

$$L_1 = \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4,$$
$$L_2 = \gamma_3 + \gamma_4 + \gamma_5 + \gamma_6 + \gamma_7.$$
$$L_3 = \gamma_2 + \gamma_4 + \gamma_6.$$

The principal block and the remaining 7 blocks are given by

Block 1	Block 2	Block 3	Block 4
(1), *eg*	*a, aeg*	*b, beg*	*c, ceg*
abcd, bdg	*bcd, abdg*	*acd, dg*	*abd, bcdg*
adf, bcf	*df, abcf*	*abdf, cf*	*acdf, bf*
cdef, abcdeg	*acdef, bcdeg*	*bcdef, acdeg*	*def, abdeg*
bde, adefg	*abde, defg*	*de, abdefg*	*bcde, acdefg*
bcefg, cdfg	*abcefg, acdfg*	*cefg, bcdfg*	*befg, dfg*
acg, abef	*cg, bef*	*abcg, aef*	*ag, abcef*
ace, abfg	*ce, bfg*	*abce, afg*	*ae, abcfg*

Block 5	Block 6	Block 7	Block 8
d, deg	*e, g*	*f, efg*	*ab, abeg*
abc, bg	*abcde, bdeg*	*abcdf, bdfg*	*cd, adg*
af, bcdf	*adef, bcef*	*ad, bc*	*bdf, acf*
cef, abceg	*cdf, abcdg*	*cde, abcdefg*	*abcdef, cdeg*
be, aefg	*bd, adfg*	*bdef, adeg*	*ade, bdefg*
bcdefg, cfg	*bcfg, cdefg*	*bceg, cdg*	*acefg, abcdfg*
acdg, abdef	*aceg, abf*	*acfg, abe*	*bcg, ef*
acde, abdfg	*ac, abefg*	*acef, abg*	*bce, fg*

The two-way interactions $AB \equiv CD$, $AC \equiv BD$, $AD \equiv BC$, $BD \equiv F$, $BF \equiv D$ and $DF \equiv B$.

15.47 A design (where $L_1 = L_2 = L_3 = L_4 = 0$ (mod 2) are used) is:

$$\{(1), abcg, abdh, abef, acdf, aceh, adeg, afgh,$$
$$bcde, bcfh, bdfg, cdgh, cefg, defh, degh, abcdefgh\}$$

15.48 In the four defining contrasts, $BCDE$, $ACDF$, $ABCG$, and $ABDH$, the length of interactions are all 4. Hence, it must be a resolution IV design.

15.49 Assuming three factors the design is a 2^3 design with 4 center runs.

15.50 (a) Consider a 2^{3-1}_{III} design with $ABC \equiv I$ as defining contrast. Then the design points are

x_1	x_2	x_3
-1	-1	-1
1	-1	1
-1	1	1
1	1	-1
0	0	0
0	0	0

For the noncentral design points, $\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = 0$ and $\bar{x}_1^2 = \bar{x}_2^2 = \bar{x}_3^2 = 1$. Hence

$$E(\bar{y}_f - \bar{y}_0) = \beta_0 + \beta_1\bar{x}_1 + \beta_2\bar{x}_2 + \beta_3\bar{x}_3 + \beta_{11}\bar{x}_1^2 + \beta_{22}\bar{x}_2^2 + \beta_{33}\bar{x}_3^2 - \beta_0 = \beta_{11} + \beta_{22} + \beta_{33}.$$

(b) It is learned that the test for curvature that involves $\bar{y}_f - \bar{y}_0$ actually is testing the hypothesis $\beta_{11} + \beta_{22} + \beta_{33} = 0$.

Chapter 16

Nonparametric Statistics

16.1 The hypotheses

$$H_0 : \tilde{\mu} = 20 \text{ minutes}$$
$$H_1 : \tilde{\mu} > 20 \text{ minutes.}$$

$\alpha = 0.05$.
Test statistic: binomial variable X with $p = 1/2$.
Computations: Subtracting 20 from each observation and discarding the zeroes. We obtain the signs

$$- \quad + \quad + \quad - \quad + \quad + \quad - \quad + \quad + \quad +$$

for which $n = 10$ and $x = 7$. Therefore, the P-value is

$$P = P(X \geq 7 \mid p = 1/2) = \sum_{x=7}^{10} b(x; 10, 1/2)$$
$$= 1 - \sum_{x=0}^{6} b(x; 10, 1/2) = 1 - 0.8281 = 0.1719 > 0.05.$$

Decision: Do not reject H_0.

16.2 The hypotheses

$$H_0 : \tilde{\mu} = 12$$
$$H_1 : \tilde{\mu} \neq 12.$$

$\alpha = 0.02$.
Test statistic: binomial variable X with $p = 1/2$.
Computations: Replacing each value above and below 12 by the symbol "+" and "−", respectively, and discarding the two values which equal to 12. We obtain the sequence

$$- \quad + \quad - \quad - \quad + \quad + \quad + \quad + \quad - \quad + \quad + \quad - \quad + \quad + \quad - \quad +$$

for which $n = 16$, $x = 10$ and $n/2 = 8$. Therefore, the P-value is

$$P = 2P(X \geq 10 \mid p = 1/2) = 2\sum_{x=10}^{16} b(x; 16, 1/2)$$
$$= 2(1 - \sum_{x=0}^{9} b(x; 16, 1/2)) = 2(1 - 0.7728) = 0.4544 > 0.02.$$

Decision: Do not reject H_0.

16.3 The hypotheses

$$H_0 : \tilde{\mu} = 2.5$$
$$H_1 : \tilde{\mu} \neq 2.5.$$

$\alpha = 0.05$.
Test statistic: binomial variable X with $p = 1/2$.
Computations: Replacing each value above and below 2.5 by the symbol "+" and "−", respectively. We obtain the sequence

$$- \quad - \quad - \quad - \quad - \quad - \quad - \quad + \quad + \quad - \quad + \quad - \quad - \quad - \quad - \quad -$$

for which $n = 16$, $x = 3$. Therefore, $\mu = np = (16)(0.5) = 8$ and $\sigma = \sqrt{(16)(0.5)(0.5)} = 2$. Hence $z = (3.5 - 8)/2 = -2.25$, and then

$$P = 2P(X \leq 3 \mid p = 1/2) \approx 2P(Z < -2.25) = (2)(0.0122) = 0.0244 < 0.05.$$

Decision: Reject H_0.

16.4 The hypotheses

$$H_0 : \tilde{\mu}_1 = \tilde{\mu}_2$$
$$H_1 : \tilde{\mu}_1 < \tilde{\mu}_2.$$

$\alpha = 0.05$.
Test statistic: binomial variable X with $p = 1/2$.
Computations: After replacing each positive difference by a "+" symbol and negative difference by a "−" symbol, respectively, and discarding the two zero differences, we have $n = 10$ and $x = 2$. Therefore, the P-value is

$$P = P(X \leq 2 \mid p = 1/2) = \sum_{x=0}^{2} b(x; 10, 1/2) = 0.0547 > 0.05.$$

Decision: Do not reject H_0.

16.5 The hypotheses

$$H_0: \tilde{\mu}_1 - \tilde{\mu}_2 = 4.5$$
$$H_1: \tilde{\mu}_1 - \tilde{\mu}_2 < 4.5.$$

$\alpha = 0.05$.
Test statistic: binomial variable X with $p = 1/2$.
Computations: We have $n = 10$ and $x = 4$ plus signs. Therefore, the P-value is

$$P = P(X \leq 4 \mid p = 1/2) = \sum_{x=0}^{4} b(x; 10, 1/2) = 0.3770 > 0.05.$$

Decision: Do not reject H_0.

16.6 The hypotheses

$$H_0: \tilde{\mu}_A = \tilde{\mu}_B$$
$$H_1: \tilde{\mu}_A \neq \tilde{\mu}_B.$$

$\alpha = 0.05$.
Test statistic: binomial variable X with $p = 1/2$.
Computations: We have $n = 14$ and $x = 12$. Therefore, $\mu = np = (14)(1/2) = 7$ and $\sigma = \sqrt{(14)(1/2)(1/2)} = 1.8708$. Hence, $z = (11.5 - 7)/1.8708 = 2.41$, and then

$$P = 2P(X \geq 12 \mid p = 1/2) = 2P(Z > 2.41) = (2)(0.0080) = 0.0160 < 0.05.$$

Decision: Reject H_0.

16.7 The hypotheses

$$H_0: \tilde{\mu}_2 - \tilde{\mu}_1 = 8$$
$$H_1: \tilde{\mu}_2 - \tilde{\mu}_1 < 8.$$

$\alpha = 0.05$.
Test statistic: binomial variable X with $p = 1/2$.
Computations: We have $n = 13$ and $x = 4$. Therefore, $\mu = np = (13)(1/2) = 6.5$ and $\sigma = \sqrt{(13)(1/2)(1/2)} = 1.803$. Hence, $z = (4.5 - 6.5)/1.803 = -1.11$, and then

$$P = P(X \geq 4 \mid p = 1/2) = P(Z < -1.11) = 0.1335 > 0.05.$$

Decision: Do not reject H_0.

16.8 The hypotheses

$$H_0: \tilde{\mu} = 20$$
$$H_1: \tilde{\mu} > 20.$$

$\alpha = 0.05$.
Critical region: $w_{\leq} 11$ for $n = 10$.
Computations:

d_i	-3	12	5	-5	8	5	-8	15	6	4
Rank	1	9	4	4	7.5	4	7.5	10	6	2

Therefore, $w_=12.5$.
Decision: Do not reject H_0.

16.9 The hypotheses

$$H_0: \tilde{\mu} = 12$$
$$H_1: \tilde{\mu} \neq 12.$$

$\alpha = 0.02$.
Critical region: $w_\leq 20$ for $n = 15$.
Computations:

d_i	-3	1	-2	-1	6	4	1	2	-1	3	-3	1	2	-1	2
Rank	12	3.5	8.5	3.5	15	14	3.5	8.5	3.5	12	12	3.5	8.5	3.5	8.5

Now, $w_=43$ and $w_+ = 77$, so that $w = 43$.
Decision: Do not reject H_0.

16.10 The hypotheses

$$H_0: \tilde{\mu}_1 - \tilde{\mu}_2 = 0$$
$$H_1: \tilde{\mu}_1 - \tilde{\mu}_2 < 0.$$

$\alpha = 0.02$.
Critiral region: $w+ \leq 1$ for $n = 5$.
Computations:

Pair	1	2	3	4	5
d_i	-5	-2	1	-4	2
Rank	5	2.5	1	4	2.5

Therefore, $w_+ = 3.5$.
Decision: Do not reject H_0.

16.11 The hypotheses

$$H_0: \tilde{\mu}_1 - \tilde{\mu}_2 = 4.5$$
$$H_1: \tilde{\mu}_1 - \tilde{\mu}_2 < 4.5.$$

$\alpha = 0.05$.
Critiral region: $w+ \leq 11$.
Computations:

Woman	1	2	3	4	5	6	7	8	9	10
d_i	−1.5	5.4	3.6	6.9	5.5	2.7	2.3	3.4	5.9	0.7
$d_i - d_0$	−6.0	0.9	−0.9	2.4	1.0	−1.8	−2.2	−1.1	1.4	−3.8
Rank	10	1.5	1.5	8	3	6	7	4	5	9

Therefore, $w_+ = 17.5$.
Decision: Do not reject H_0.

16.12 The hypotheses

$$H_0 : \tilde{\mu}_A - \tilde{\mu}_B = 0$$
$$H_1 : \tilde{\mu}_A - \tilde{\mu}_B > 0.$$

$\alpha = 0.01$.
Critiral region: $z > 2.575$.
Computations:

Day	1	2	3	4	5	6	7	8	9	10
d_i	2	6	3	5	8	−3	8	1	6	−3
Rank	4	15.5	7.5	13	19.5	7.5	19.5	1.5	15.5	7.5
Day	11	12	13	14	15	16	17	18	19	20
d_i	4	6	6	2	−4	3	7	1	−2	4
Rank	11	15.5	15.5	4	11	7.5	18	1.5	4	11

Now $w = 180$, $n = 20$, $\mu_{W_+} = (20)(21)/4 = 105$, and $\sigma_{W_+} = \sqrt{(20)(21)(41)/24} = 26.786$. Therefore, $z = (180 - 105)/26.786 = 2.80$
Decision: Reject H_0; on average, Pharmacy A fills more prescriptions than Pharmacy B.

16.13 The hypotheses

$$H_0 : \tilde{\mu}_1 - \tilde{\mu}_2 = 8$$
$$H_1 : \tilde{\mu}_1 - \tilde{\mu}_2 < 8.$$

$\alpha = 0.05$.
Critiral region: $z < -1.645$.
Computations:

d_i	6	9	3	5	8	9	4	10
$d_i - d_0$	−2	1	−5	−3	0	1	−4	2
Rank	4.5	1.5	10.5	7.5	–	1.5	9	4.5
d_i	8	2	6	3	1	6	8	11
$d_i - d_0$	0	−6	−2	−5	−7	−2	0	3
Rank	–	12	4.5	10.5	13	4.5	–	7.5

Discarding zero differences, we have $w_+ = 15$, $n = 13$, $\mu_{W_+} = (13)(14)/4 = 45, 5$, and $\sigma_{W_+} = \sqrt{(13)(14)(27)/24} = 15.309$. Therefore, $z = (15 - 45.5)/14.309 = -2.13$
Decision: Reject H_0; the average increase is less than 8 points.

16.14 The hypotheses

$$H_0 : \tilde{\mu}_A - \tilde{\mu}_B = 0$$
$$H_1 : \tilde{\mu}_A - \tilde{\mu}_B \neq 0.$$

$\alpha = 0.05$.
Critiral region: $w \leq 21$ for $n = 14$.
Computations:

d_i	0.09	0.08	0.12	0.06	0.13	−0.06	0.12
Rank	7	5.5	10	2.5	12	2.5	10
d_i	0.11	0.12	−0.04	0.08	0.15	0.07	0.14
Rank	8	10	1	5.5	14	4	13

Hence, $w_+ = 101.5$, $w_- = 3.5$, so $w = 3.5$.
Decision: Reject H_0; the different instruments lead to different results.

16.15 The hypotheses

$$H_0 : \tilde{\mu}_B = \tilde{\mu}_A$$
$$H_1 : \tilde{\mu}_B < \tilde{\mu}_A.$$

$\alpha = 0.05$.
Critiral region: $n_1 = 3$, $n_2 = 6$ so $u_1 \leq 2$.
Computations:

Original data	1	7	8	9	10	11	12	13	14
Rank	1	2*	3*	4	5*	6	7	8	9

Now $w_1 = 10$ and hence $u_1 = 10 - (3)(4)/2 = 4$
Decision: Do not reject H_0; the claim that the tar content of brand B cigarettes is lower than that of brand A is not statistically supported.

16.16 The hypotheses

$$H_0 : \tilde{\mu}_1 = \tilde{\mu}_2$$
$$H_1 : \tilde{\mu}_1 < \tilde{\mu}_2.$$

$\alpha = 0.05$.
Critiral region: $u_1 \leq 2$.
Computations:

Original data	0.5	0.9	1.4	1.9	2.1	2.8	3.1	4.6	5.3
Rank	1^*	2	3	4^*	5	6^*	7^*	8	9

Now $w_1 = 18$ and hence $u_1 = 18 - (4)(5)/2 = 8$
Decision: Do not reject H_0.

16.17 The hypotheses

$$H_0 : \tilde{\mu}_A = \tilde{\mu}_B$$
$$H_1 : \tilde{\mu}_A > \tilde{\mu}_B.$$

$\alpha = 0.01$.
Critiral region: $u_2 \le 14$.
Computations:

Original data	3.8	4.0	4.2	4.3	4.5	4.5	4.6	4.8	4.9
Rank	1^*	2^*	3^*	4^*	5.5^*	5.5^*	7	8^*	9^*
Original Data	5.0	5.1	5.2	5.3	5.5	5.6	5.8	6.2	6.3
Rank	10	11	12	13	14	15	16	17	18

Now $w_2 = 50$ and hence $u_2 = 50 - (9)(10)/2 = 5$
Decision: Reject H_0; calculator A operates longer.

16.18 The hypotheses

$$H_0 : \tilde{\mu}_1 = \tilde{\mu}_2$$
$$H_1 : \tilde{\mu}_1 \neq \tilde{\mu}_2.$$

$\alpha = 0.01$.
Critiral region: $u \le 27$.
Computations:

Original data	8.7	9.3	9.5	9.6	9.8	9.8	9.8	9.9	9.9	10.0
Rank	1^*	2	3^*	4	6^*	6^*	6^*	8.5^*	8.5	10
Original Data	10.1	10.4	10.5	10.7	10.8	10.9	11.0	11.2	11.5	11.8
Rank	11^* 12	13^*	14	15^*	16	17^*	18^*	19	20	

Here "*" is for process 2. Now $w_1 = 111.5$ for process 1 and $w_2 = 98.5$ for process 2. Therefore, $u_1 = 111.5 - (10)(11)/2 = 56.5$ and $u_2 = 98.5 - (10)(11)/2 = 43.5$, so that $u = 43.5$.
Decision: Do not reject H_0.

16.19 The hypotheses

$$H_0 : \tilde{\mu}_1 = \tilde{\mu}_2$$
$$H_1 : \tilde{\mu}_1 \neq \tilde{\mu}_2.$$

$\alpha = 0.05$.
Critiral region: $u \leq 5$.
Computations:

Original data	64	67	69	75	78	79	80	82	87	88	91	93
Rank	1	2	3*	4	5*	6	7*	8	9*	10	11*	12

Now $w_1 = 35$ and $w_2 = 43$. Therefore, $u_1 = 35 - (5)(6)/2 = 20$ and $u_2 = 43 - (7)(8)/2 = 15$, so that $u = 15$.
Decision: Do not reject H_0.

16.20 The hypotheses

$$H_0 : \tilde{\mu}_1 = \tilde{\mu}_2$$
$$H_1 : \tilde{\mu}_1 \neq \tilde{\mu}_2.$$

$\alpha = 0.05$.
Critiral region: $Z < -1.96$ or $z > 1.96$.
Computations:

Observation	12.7	13.2	13.6	13.6	14.1	14.1	14.5	14.8	15.0	15.0	15.4
Rank	1*	2	3.5*	3.5	5.5*	5.5	7	8	9.5*	9.5	11.5*
Observation	15.4	15.6	15.9	15.9	16.3	16.3	16.3	16.3	16.5	16.8	17.2
Rank	11.5	13*	14.5*	14.5	17.5*	17.5*	17.5	17.5	20	21*	22
Observation	17.4	17.7	17.7	18.1	18.1	18.3	18.6	18.6	18.6	19.1	20.0
Rank	23	24.5	24.5*	26.5*	26.5	28	30	30*	30	32	33

Now $w_1 = 181.5$ and $u_1 = 181.5 - (12)(13)/2 = 103.5$. Then with $\mu_{U_1} = (21)(12)/2 = 126$ and $\sigma_{U_1} = \sqrt{(21)(12)(34)/12} = 26.721$, we find $z = (103.5 - 126)/26.721 = -0.84$.
Decision: Do not reject H_0.

16.21 The hypotheses

$$H_0 : \text{Operating times for all three calculators are equal.}$$
$$H_1 : \text{Operating times are not all equal.}$$

$\alpha = 0.01$.
Critiral region: $h > \chi^2_{0.01} = 9.210$ with $v = 2$ degrees of freedom.
Computations:

Ranks for Calculators		
A	B	C
4	8.5	15
12	7	18
1	13	10
2	11	16
6	8.5	14
$r_1 = 25$	5	17
	3	$r_3 = 90$
	$r_2 = 56$	

Now $h = \frac{12}{(18)(19)}\left[\frac{25^2}{5} + \frac{56^2}{7} + \frac{90^2}{6}\right] - (3)(19) = 10.47$.
Decision: Reject H_0; the operating times for all three calculators are not equal.

16.22 Kruskal-Wallis test (Chi-square approximation)

$$h = \frac{12}{(32)(33)}\left[\frac{210.5^2}{9} + \frac{189^2}{8} + \frac{128.5^2}{15}\right] - (3)(33) = 20.21.$$

$\chi^2_{0.05} = 5.991$ with 2 degrees of freedom. So, we reject H_0 and claim that the mean sorptions are not the same for all three solvents.

16.23 The hypotheses

$$H_0: \text{Sample is random.}$$
$$H_1: \text{Sample is not random.}$$

$\alpha = 0.1$.
Test statistics: V, the total number of runs.
Computations: for the given sequence we obtain $n_1 = 5$, $n_2 = 10$, and $v = 7$. Therefore, from Table A.18, the P-value is

$$P = 2P(V \leq 7 \text{ when } H_0 \text{ is true}) = (2)(0.455) = 0.910 > 0.1$$

Decision: Do not reject H_0; the sample is random.

16.24 The hypotheses

$$H_0: \text{Fluctuations are random.}$$
$$H_1: \text{Fluctuations are not random.}$$

$\alpha = 0.05$.
Test statistics: V, the total number of runs.
Computations: for the given sequence we find $\tilde{x} = 0.021$. Replacing each measurement

by the symbol "+" if it falls above 0.021 and by the symbol "−" if it falls below 0.021 and omitting the two measurements that equal 0.021, we obtain the sequence

$$- \quad - \quad - \quad - \quad - \quad + \quad + \quad + \quad + \quad +$$

for which $n_1 = 5$, $n_2 = 5$, and $v = 2$. Therefore, the P-value is

$$P = 2P(V \leq 2 \text{ when } H_0 \text{ is true}) = (2)(0.008) = 0.016 < 0.05$$

Decision: Reject H_0; the fluctuations are not random.

16.25 The hypotheses

$$H_0 : \mu_A = \mu_B$$
$$H_1 : \mu_A > \mu_B.$$

$\alpha = 0.01$.
Test statistics: V, the total number of runs.
Computations: from Exercise 16.17 we can write the sequence

$$B \quad B \quad B \quad B \quad B \quad B \quad A \quad B \quad B \quad A \quad A \quad B \quad A \quad A \quad A \quad A \quad A \quad A$$

for which $n_1 = 9$, $n_2 = 9$, and $v = 6$. Therefore, the P-value is

$$P = P(V \leq 6 \text{ when } H_0 \text{ is true}) = 0.044 > 0.01$$

Decision: Do not reject H_0.

16.26 The hypotheses

$$H_0 : \text{Defectives occur at random.}$$
$$H_1 : \text{Defectives do not occur at random.}$$

$\alpha = 0.05$.
Critical region: $z < -1.96$ or $z > 1.96$.
Computations: $n_1 = 11$, $n_2 = 17$, and $v = 13$. Therefore,

$$\mu_V = \frac{(2)(11)(17)}{28} + 1 = 14.357,$$
$$\sigma_V^2 = \frac{(2)(11)(17)[(2)(11)(17) - 11 - 17]}{(28^2)(27)} = 6.113,$$

and hence $\sigma_V = 2.472$. Finally,

$$z = (13 - 14.357)/2.472 = -0.55.$$

Decision: Do not reject H_0.

16.27 The hypotheses

$$H_0: \text{Sample is random.}$$
$$H_1: \text{Sample is not random.}$$

$\alpha = 0.05$.
Critical region: $z < -1.96$ or $z > 1.96$.
Computations: we find $\bar{x} = 2.15$. Assigning "+" and "−" signs for observations above and below the median, respectively, we obtain $n_1 = 15$, $n_2 = 15$, and $v = 19$. Hence,

$$\mu_V = \frac{(2)(15)(15)}{30} + 1 = 16,$$
$$\sigma_V^2 = \frac{(2)(15)(15)[(2)(15)(15) - 15 - 15]}{(30^2)(29)} = 7.241,$$

which yields $\sigma_V = 2.691$. Therefore,

$$z = (19 - 16)/2.691 = 1.11.$$

Decision: Do not reject H_0.

16.28 $1 - \gamma = 0.95$, $1 - \alpha = 0.85$. From Table A.20, $n = 30$.

16.29 $n = 24$, $1 - \alpha = 0.90$. From Table A.20, $1 - \gamma = 0.70$.

16.30 $1 - \gamma = 0.99$, $1 - \alpha = 0.80$. From Table A.21, $n = 21$.

16.31 $n = 135$, $1 - \alpha = 0.95$. From Table A.21, $1 - \gamma = 0.995$.

16.32 (a) Using the computations, we have

Student	Test	Exam	d_i
L.S.A.	4	4	0
W.P.B.	10	2	8
R.W.K.	7	8	−1
J.R.L.	2	3	−1
J.K.L.	5	6.5	−1.5
D.L.P.	9	6.5	2.5
B.L.P.	3	10	−7
D.W.M.	1	1	0
M.N.M.	8	9	−1
R.H.S.	6	5	−1

$$r_S = 1 - \frac{(6)(125.5)}{(10)(100-1)} = 0.24.$$

(b) The hypotheses

$$H_0: \rho = 0$$
$$H_1: \rho > 0$$

$\alpha = 0.025$.
Critical region: $r_S > 0.648$.
Decision: Do not reject H_0.

16.33 (a) Using the following

Ranks x	Ranks y	d	Ranks x	Ranks y	d
1	6	-5	14	12	2
2	1	1	15	2	13
3	16	-13	16	6	10
4	9.5	-5.5	17	13.5	3.5
5	18.5	-13.5	18	13.5	4.5
6	23	-17	19	16	3
7	8	-1	20	23	-3
8	3	5	21	23	-2
9	9.5	-0.5	22	23	-1
10	16	-6	23	18.5	4.5
11	4	7	24	23	1
12	20	-8	25	6	19
13	11	2			

we obtain $r_S = 1 - \frac{(6)(1586.5)}{(25)(625-1)} = 0.39$.

(b) The hypotheses

$$H_0: \rho = 0$$
$$H_1: \rho \neq 0$$

$\alpha = 0.05$.
Critical region: $r_S < -0.400$ or $r_s > 0.400$.
Decision: Do not reject H_0.

16.34 The numbers come up as follows

Ranks x	Ranks y	d	Ranks x	Ranks y	d	Ranks x	Ranks y	d
3	7	-4	4	6	-2	7	3	4
6	4.5	1.5	8	2	6	5	4.5	0.5
2	8	-6	1	9	-8	9	1	8

$$\sum d^2 = 238.5, \quad r_S = 1 - \frac{(6)(238.5)}{(9)(80)} = -0.99.$$

16.35 (a) We have the following table:

Weight	Chest Size	d_i	Weight	Chest Size	d_i	Weight	Chest Size	d_i
3	6	-3	1	1	0	8	8	0
9	9	0	4	2	2	7	3	4
2	4	-2	6	7	-1	5	5	0

$$r_S = 1 - \frac{(6)(34)}{(9)(80)} = 0.72.$$

(b) The hypotheses

$$H_0 : \rho = 0$$
$$H_1 : \rho > 0$$

$\alpha = 0.025$.
Critical region: $r_S > 0.683$.
Decision: Reject H_0 and claim $\rho > 0$.

16.36 The hypotheses

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

$\alpha = 0.05$.
Critical region: $r_S < -0.683$ or $r_S > 0.683$.
Computations:

Manufacture	A	B	C	D	E	F	G	H	I
Panel rating	6	9	2	8	5	1	7	4	3
Price rank	5	1	9	8	6	7	2	4	3
d_i	1	8	-7	0	-1	-6	5	0	0

Therefore, $r_S = 1 - \frac{(6)(176)}{(9)(80)} = -0.47$.
Decision: Do not reject H_0.

16.37 (a) $\sum d^2 = 24$, $r_S = 1 - \frac{(6)(24)}{(8)(63)} = 0.71$.

(b) The hypotheses

$$H_0 : \rho = 0$$
$$H_1 : \rho > 0$$

$\alpha = 0.05$.
Critical region: $r_S > 0.643$.
Computations: $r_S = 0.71$.
Decision: Reject H_0, $\rho > 0$.

16.38 (a) $\sum d^2 = 1828$, $r_S = 1 - \frac{(6)(1828)}{(30)(899)} = 0.59$.

(b) The hypotheses

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

$\alpha = 0.05$.
Critical region: $r_S < -0.364$ or $r_S > 0.364$.
Computations: $r_S = 0.59$.
Decision: Reject H_0, $\rho \neq 0$.

16.39 (a) The hypotheses

$$H_0 : \mu_A = \mu_B$$
$$H_1 : \mu_A \neq \mu_B$$

Test statistic: binomial variable X with $p = 1/2$.
Computations: $n = 9$, omitting the identical pair, so $x = 3$ and P-value is $P = P(X \leq 3) = 0.2539$.
Decision: Do not reject H_0.

(b) $w_+ = 15.5$, $n = 9$.
Decision: Do not reject H_0.

16.40 The hypotheses:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4.$$
H_1 : At least two of the means are not equal.

$\alpha = 0.05$.
Critical region: $h > \chi^2_{0.05} = 7.815$ with 3 degrees of freedom.
Computaions:

Ranks for the Laboratories			
A	B	C	D
7	18	2	12
15.5	20	3	10.5
13.5	19	4	13.5
8	9	1	15.5
6	10.5	5	17
$r_1 = 50$	$r_2 = 76.5$	$r_3 = 15$	$r_4 = 68.5$

Now
$$h = \frac{12}{(20)(21)}\left[\frac{50^2 + 76.5^2 + 15^2 + 68.5^2}{5}\right] - (3)(21) = 12.83.$$
Decision: Reject H_0.

16.41 The hypotheses:

$$H_0 : \mu_{29} = \mu_{54} = \mu_{84}.$$
$$H_1 : \text{At least two of the means are not equal.}$$

Kruskal-Wallis test (Chi-squared approximation)

$$h = \frac{12}{(12)(13)}\left[\frac{6^2}{3} + \frac{38^2}{5} + \frac{34^2}{4}\right] - (3)(13) = 6.37,$$

with 2 degrees of freedom. $\chi^2_{0.05} = 5.991$.
Decision: reject H_0. Mean nitrogen loss is different for different levels of dietary protein.

Chapter 17

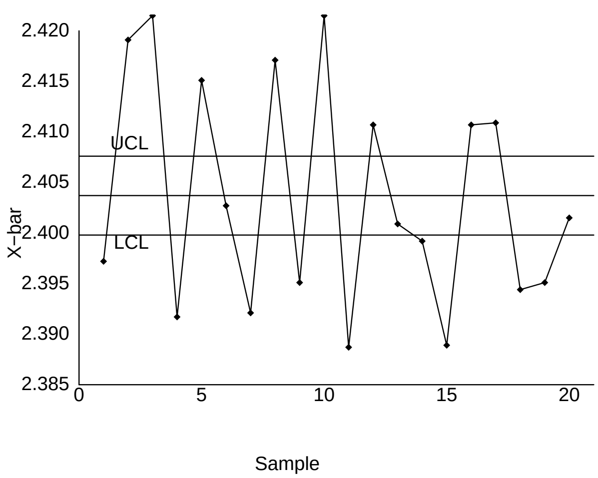
Statistical Quality Control

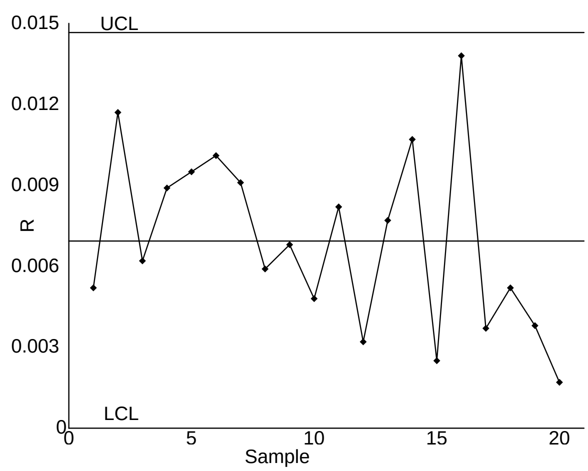
17.1 Let $Y = X_1 + X_2 + \cdots + X_n$. The moment generating function of a Poisson random variable is given by $M_X(t) = e^{\mu(e^t - 1)}$. By Theorem 7.10,

$$M_Y(t) = e^{\mu_1(e^t-1)} \cdot e^{\mu_2(e^t-1)} \cdots e^{\mu_n(e^t-1)} = e^{(\mu_1+\mu_2+\cdots+\mu_n)(e^t-1)},$$

which we recognize as the moment generating function of a Poisson random variable with mean and variance given by $\sum_{i=1}^{n} \mu_i$.

17.2 The charts are shown as follows.





Although none of the points in R-chart is outside of the limits, there are many values fall outside control limits in the $\bar{X}$-chart.

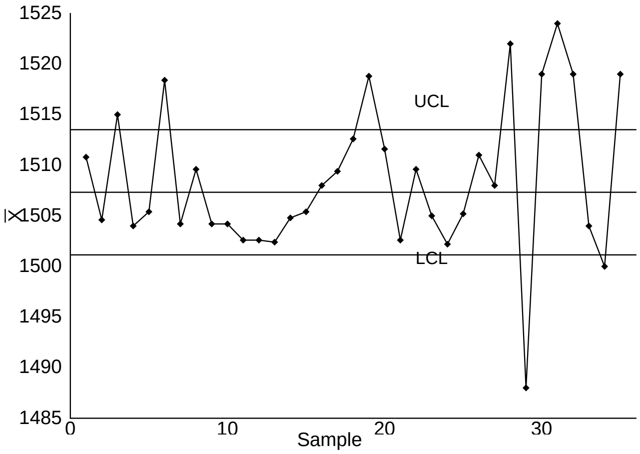
17.3 There are 10 values, out of 20, fall outside the specification ranges. So, 50% of the units produced by this process will not confirm the specifications.

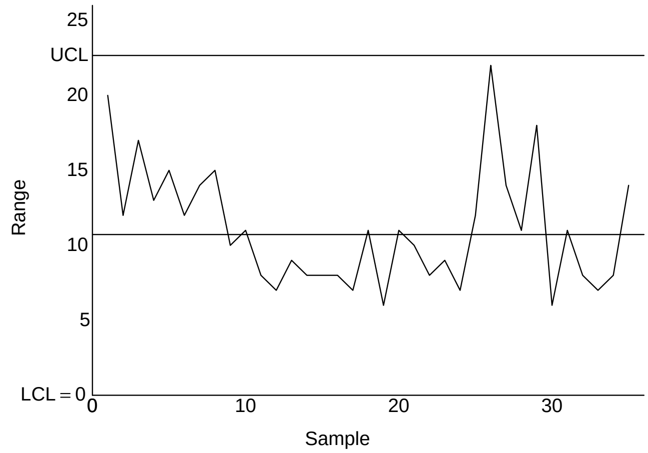
17.4 $\bar{\bar{X}} = 2.4037$ and $\hat{\sigma} = \frac{\bar{R}}{d_2} = \frac{0.006935}{2.326} = 0.00298$.

17.5 Combining all 35 data values, we have

$$\bar{\bar{x}} = 1508.491, \quad \bar{R} = 11.057,$$

so for $\bar{X}$-chart, LCL $= 1508.491 - (0.577)(11.057) = 1502.111$, and UCL $= 1514.871$; and for R-chart, LCL $= (11.057)(0) = 0$, and UCL $= (11.057)(2.114) = 23.374$. Both charts are given below.





The process appears to be out of control.

17.6

$$\begin{aligned}\beta &= P(Z < 3 - 1.5\sqrt{5}) - P(Z < -3 - 1.5\sqrt{5}) \\ &= P(Z < -0.35) - P(Z < -6.35) \approx 0.3632.\end{aligned}$$

So,

$$E(S) = 1/(1 - 0.3632) = 1.57, \quad \text{and} \quad \sigma_S = \sqrt{\beta}(1-\beta)^2 = 0.896.$$

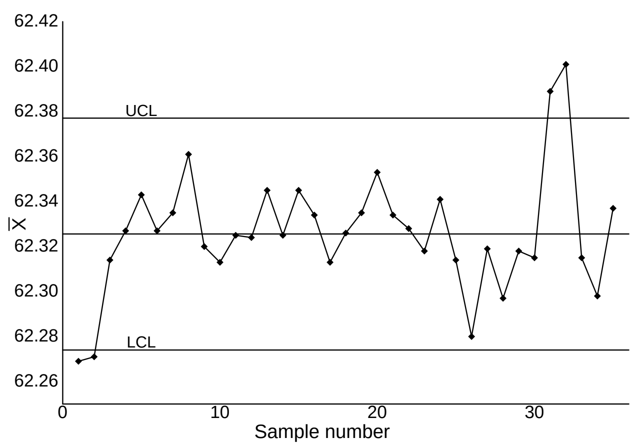
17.7 From Example 17.2, it is known than

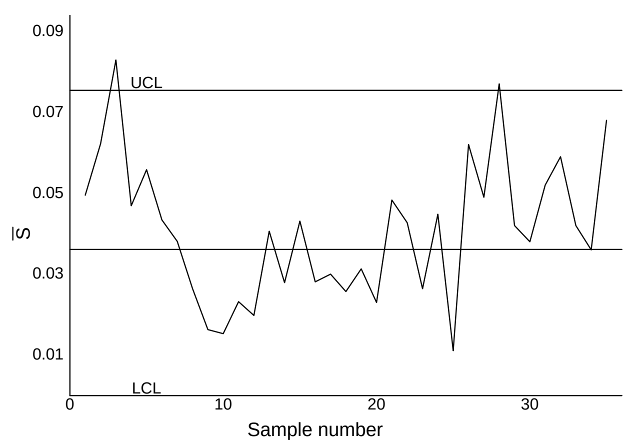
$$\text{LCL} = 62.2740, \quad \text{and} \quad \text{UCL} = 62.3771,$$

for the $\bar{X}$-chart and

$$\text{LCL} = 0, \quad \text{and} \quad \text{UCL} = 0.0754,$$

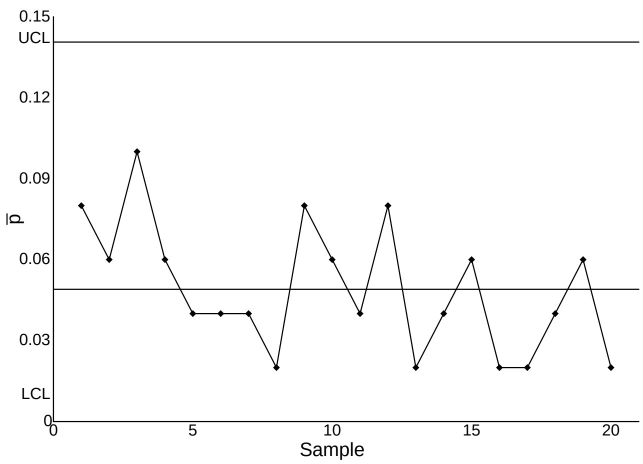
for the S-chart. The charts are given below.



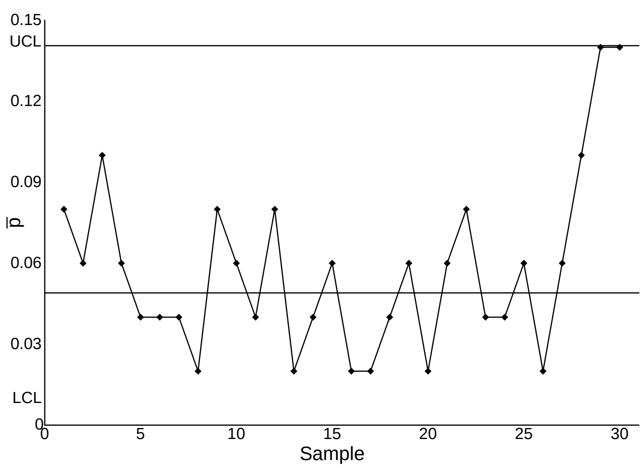


The process appears to be out of control.

17.8 Based on the data, we obtain $\hat{p} = 0.049$, LCL $= 0.049 - 3\sqrt{\frac{(0.049)(0.951)}{50}} = -0.043$, and LCL $= 0.049 + 3\sqrt{\frac{(0.049)(0.951)}{50}} = 0.1406$. Based on the chart shown below, it appears that the process is in control.

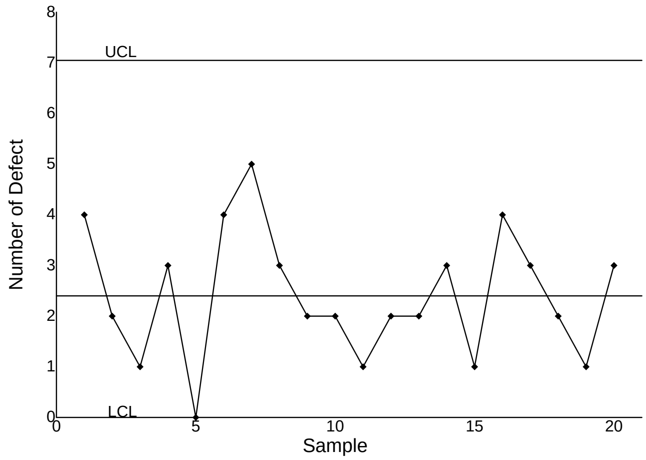


17.9 The chart is given below.



Although there are a few points closed to the upper limit, the process appears to be in control as well.

17.10 We use the Poisson distribution. The estimate of the parameter λ is $\hat{\lambda} = 2.4$. So, the control limits are LCL $= 2.4 - 3\sqrt{2.4} = -2.25$ and UCL $= 2.4 + 3\sqrt{2.4} = 7.048$. The control chart is shown below.



The process appears in control.

Chapter 18

Bayesian Statistics

18.1 For $p = 0.1, b(2; 2, 0.1) = \binom{2}{2}(0.1)^2 = 0.01$.
For $p = 0.2, b(2; 2, 0.2) = \binom{2}{2}(0.2)^2 = 0.04$. Denote by

$$\begin{aligned} A:&\quad \text{number of defectives in our sample is 2;} \\ B_1:&\quad \text{proportion of defective is } p = 0.1; \\ B_2:&\quad \text{proportion of defective is } p = 0.2. \end{aligned}$$

Then

$$P(B_1|A) = \frac{(0.6)(0.01)}{(0.6)(0.01) + (0.4)(0.04)} = 0.27,$$

and then by subtraction $P(B_2|A) = 1 - 0.27 = 0.73$. Therefore, the posterior distribution of p after observing A is

p	0.1	0.2
$\pi(p\|x = 2)$	0.27	0.73

for which we get $p^* = (0.1)(0.27) + (0.2)(0.73) = 0.173$.

18.2 (a) For $p = 0.05, b(2; 9, 0.05) = \binom{9}{2}(0.05)^2(0.95)^7 = 0.0629$.
For $p = 0.10, b(2; 9, 0.10) = \binom{9}{2}(0.10)^2(0.90)^7 = 0.1722$.
For $p = 0.15, b(2; 9, 0.15) = \binom{9}{2}(0.15)^2(0.85)^7 = 0.2597$.
Denote the following events:

$$\begin{aligned} A:&\quad \text{2 drinks overflow;} \\ B_1:&\quad \text{proportion of drinks overflowing is } p = 0.05; \\ B_2:&\quad \text{proportion of drinks overflowing is } p = 0.10; \\ B_3:&\quad \text{proportion of drinks overflowing is } p = 0.15. \end{aligned}$$

Then

$$P(B_1|A) = \frac{(0.3)(0.0629)}{(0.3)(0.0629) + (0.5)(0.1722) + (0.2)(0.2597)} = 0.12,$$

$$P(B_2|A) = \frac{(0.5)(0.1722)}{(0.3)(0.0629) + (0.5)(0.1722) + (0.2)(0.2597)} = 0.55,$$

and $P(B_3|A) = 1 - 0.12 - 0.55 = 0.33$. Hence the posterior distribution is

p	0.05	0.10	0.15
$\pi(p\|x=2)$	0.12	0.55	0.33

(b) $p^* = (0.05)(0.12) + (0.10)(0.55) + (0.15)(0.33) = 0.111$.

18.3 (a) Let X = the number of drinks that overflow. Then

$$f(x|p) = b(x;4,p) = \binom{4}{x} p^x (1-p)^{4-x}, \quad \text{for } x = 0,1,2,3,4.$$

Since

$$f(1,p) = f(1|p)\pi(p) = 10\binom{4}{1} p(1-p)^3 = 40p(1-p)^3, \quad \text{for } 0.05 < p < 0.15,$$

then

$$g(1) = 40\int_{0.05}^{0.15} p(1-p)^3\,dp = -2(1-p)^4\,(4p+1)\Big|_{0.05}^{0.15} = 0.2844,$$

and

$$\pi(p|x=1) = 40p(1-p)^3/0.2844.$$

(b) The Bayes estimator

$$\begin{aligned} p^* &= \frac{40}{0.2844}\int_{0.05}^{0.15} p^2(1-p)^3\,dp \\ &= \frac{40}{(0.2844)(60)} p^3\,(20 - 45p + 36p^2 - 10p^3)\Big|_{0.05}^{0.15} = 0.106. \end{aligned}$$

18.4 Denote by

A : 12 condominiums sold are units;
B_1 : proportion of two-bedroom condominiums sold 0.60;
B_2 : proportion of two-bedroom condominiums sold 0.70.

For $p = 0.6$, $b(12;15,0.6) = 0.0634$ and for $p = 0.7$, $b(12;15,0.7) = 0.1701$. The prior distribution is given by

p	0.6	0.7
$\pi(p)$	1/3	2/3

So, $P(B_1|A) = \frac{(1/3)(0.0634)}{(1/3)(0.0634)+(2/3)(0.1701)} = 0.157$ and $P(B_2|A) = 1 - 0.157 = 0.843$. Therefore, the posterior distribution is

p	0.6	0.7
$\pi(p\|x=12)$	0.157	0.843

(b) The Bayes estimator is $p^* = (0.6)(0.157) + (0.7)(0.843) = 0.614$.

18.5 $n = 10, \bar{x} = 9, \sigma = 0.8, \mu_0 = 8, \sigma_0 = 0.2$, and $z_{0.025} = 1.96$. So,

$$\mu_1 = \frac{(10)(9)(0.04) + (8)(0.64)}{(10)(0.04) + 0.64} = 8.3846, \quad \sigma_1 = \sqrt{\frac{(0.04)(0.64)}{(10)(0.04) + 0.64}} = 0.1569.$$

To calculate Bayes interval, we use $8.3846 \pm (1.96)(0.1569) = 8.3846 \pm 0.3075$ which yields $(8.0771, 8.6921)$. Hence, the probability that the population mean is between 8.0771 and 8.6921 is 95%.

18.6 $n = 30, \bar{x} = 24.90, s = 2.10, \mu_0 = 30$ and $\sigma_0 = 1.75$.

(a) $\mu^* = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2} = \frac{2419.988}{96.285} = 25.1336$.

(b) $\sigma^* = \sqrt{\frac{\sigma_0^2\sigma^2}{n\sigma_0^2+\sigma^2}} = \sqrt{\frac{13.5056}{96.285}} = 0.3745$, and $z_{0.025} = 1.96$. Hence, the 95% Bayes interval is calculated by $25.13 \pm (1.96)(0.3745)$ which yields $\$23.40 < \mu < \$25.86$.

(c) $P(24 < \mu < 26) = P\left(\frac{24-25.13}{0.3745} < Z < \frac{26-25.13}{0.3745}\right) = P(-3.02 < Z < 2.32) = 0.9898 - 0.0013 = 0.9885$.

18.7 (a) $P(71.8 < \mu < 73.4) = P\left(\frac{71.8-72}{\sqrt{5.76}} < Z < \frac{73.4-72}{\sqrt{5.76}}\right) = P(-0.08 < Z < 0.58) = 0.2509$.

(b) $n = 100, \bar{x} = 70, s^2 = 64, \mu_0 = 72$ and $\sigma_0^2 = 5.76$. Hence,

$$\mu_1 = \frac{(100)(70)(5.76) + (72)(64)}{(100)(5.76) + 64} = 70.2,$$

$$\sigma_1 = \sqrt{\frac{(5.76)(64)}{(100)(5.76) + 64}} = 0.759.$$

Hence, the 95% Bayes interval can be calculated as $70.2 \pm (1.96)(0.759)$ which yields $68.71 < \mu < 71.69$.

(c) $P(71.8 < \mu < 73.4) = P\left(\frac{71.8-70.2}{0.759} < Z < \frac{73.4-70.2}{0.759}\right) = P(2.11 < Z < 4.22) = 0.0174$.

18.8 Multiplying the likelihood function

$$f(x_1, x_2, \ldots, x_n|\mu) = \frac{1}{(2\pi)^{25/2}100^{25}} \exp\left[-\frac{1}{2}\sum_{i=1}^{25}\left(\frac{x_i - \mu}{100}\right)^2\right]$$

by the prior $\pi(\mu) = \frac{1}{60}$ for $770 < \mu < 830$, we obtain

$$f(x_1, x_2, \ldots, x_n, \mu) = \frac{1}{(60)(2\pi)^{25/2}100^{25}} \exp\left[-\frac{1}{2}\sum_{i=1}^{25}\left(\frac{x_i - \mu}{100}\right)^2\right] = Ke^{-\frac{1}{2}\left(\frac{\mu-780}{20}\right)^2},$$

where K is a function of the sample values. Since the marginal distribution

$$g(x_1, x_2, \ldots, x_n) = \sqrt{2\pi}(20)K \left[\frac{1}{\sqrt{2\pi}20} \int_{770}^{830} e^{-\frac{1}{2}\left(\frac{\mu-780}{100}\right)^2} d\mu \right] = \sqrt{2\pi}(13.706)K.$$

Hence, the posterior distribution

$$\pi(\mu|x_1, x_2, \ldots, x_n) = \frac{f(x_1, x_2, \ldots, x_n)}{g(x_1, x_2, \ldots, x_n)} = \frac{1}{\sqrt{2\pi}(13.706)} e^{-\frac{1}{2}\left(\frac{\mu-780}{20}\right)^2},$$

for $770 < \mu < 830$.

18.9 Multiplying the likelihood function and the prior distribution together, we get the joint density function of θ as

$$f(t_1, t_2, \ldots, t_n, \theta) = 2\theta^n \exp\left[-\theta \left(\sum_{i=1}^{n} t_i + 2 \right) \right], \quad \text{for } \theta > 0.$$

Then the marginal distribution of $(T_1, T_2, \ldots, T_n)$ is

$$\begin{aligned} g(t_1, t_2, \ldots, t_n) &= 2 \int_0^\infty \theta^n \exp\left[-\theta \left(\sum_{i=1}^{n} t_i + 2 \right) \right] d\theta \\ &= \frac{2\Gamma(n+1)}{\left(\sum_{i=1}^{n} t_i + 2 \right)^{n+1}} \int_0^\infty \frac{\theta^n \exp\left[-\theta \left(\sum_{i=1}^{n} t_i + 2 \right) \right]}{\Gamma(n+1) \left(\sum_{i=1}^{n} t_i + 2 \right)^{-(n+1)}} d\theta \\ &= \frac{2\Gamma(n+1)}{\left(\sum_{i=1}^{n} t_i + 2 \right)^{n+1}}, \end{aligned}$$

since the integrand in the last term constitutes a gamma density function with parameters $\alpha = n + 1$ and $\beta = 1/\left(\sum_{i=1}^{n} t_i + 2 \right)$. Hence, the posterior distribution of θ is

$$\pi(\theta|t_1, \ldots, t_n) = \frac{f(t_1, \ldots, t_n, \theta)}{g(t_1, \ldots, t_n)} = \frac{\left(\sum_{i=1}^{n} t_i + 2 \right)^{n+1}}{\Gamma(n+1)} \theta^n \exp\left[-\theta \left(\sum_{i=1}^{n} t_i + 2 \right) \right],$$

for $\theta > 0$, which is a gamma distribution with parameters $\alpha = n + 1$ and $\beta = 1/\left(\sum_{i=1}^{n} t_i + 2 \right)$.

18.10 Assume that $p(x_i|\lambda) = e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}$, $x_i = 0, 1, \ldots,$ for $i = 1, 2, \ldots, n$ and $\pi(\lambda) = \frac{1}{2^4}\lambda^2 e^{-\lambda/2}$, for $\lambda > 0$. The posterior distribution of λ is calculated as

$$\pi(\lambda|x_1, \ldots, x_n) = \frac{e^{-(n+1/2)\lambda}\lambda^{\sum\limits_{i=1}^{n} x_i+2}}{\int_0^\infty e^{-(n+1/2)\lambda}\lambda^{\sum\limits_{i=1}^{n} x_i+2}\ d\lambda} = \frac{(n+1/2)^{n\bar{x}+3}}{\Gamma(n\bar{x}+3)}\lambda^{\sum\limits_{i=1}^{n} x_i+2} e^{-(n+1/2)\lambda},$$

which is a gamma distribution with parameters $\alpha = n\bar{x} + 3$ and $\beta = (n + 1/2)^{-1}$, with mean $\frac{n\bar{x}+3}{n+1/2}$. Hence, plug the data in we obtain the Bayes estimator of λ, under squared-error loss, is $\lambda^* = \frac{57+3}{10+1/2} = 5.7143$.

18.11 The likelihood function of p is $\binom{x-1}{4}p^5(1-p)^{x-5}$ and the prior distribution is $\pi(p) = 1$. Hence the posterior distribution of p is

$$\pi(p|x) = \frac{p^5(1-p)^{x-5}}{\int_0^1 p^5(1-p)^{x-5}\ dp} = \frac{\Gamma(x+2)}{\Gamma(6)\Gamma(x-4)}p^5(1-p)^{x-5},$$

which is a Beta distribution with parameters $\alpha = 6$ and $\beta = x - 4$. Hence the Bayes estimator, under the squared-error loss, is $p^* = \frac{6}{x+2}$.