CHAPTER 1

Section 1.1

1.

- **a.** Houston Chronicle, Des Moines Register, Chicago Tribune, Washington Post
- **b.** Capital One, Campbell Soup, Merrill Lynch, Pulitzer
- **c.** Bill Jasper, Kay Reinke, Helen Ford, David Menedez
- **d.** 1.78, 2.44, 3.5, 3.04

2.

- **a.** 29.1 yd., 28.3 yd., 24.7 yd., 31.0 yd.
- **b.** 432, 196, 184, 321
- **c.** 2.1, 4.0, 3.2, 6.3
- **d.** 0.07 g, 1.58 g, 7.1 g, 27.2 g

3.

- **a.** In a sample of 100 VCRs, what are the chances that more than 20 need service while under warrantee? What are the chances than none need service while still under warrantee?
- **b.** What proportion of all VCRs of this brand and model will need service within the warrantee period?

4.

a. Concrete: All living U.S. Citizens, all mutual funds marketed in the U.S., all books published in 1980.

Hypothetical: All grade point averages for University of California undergraduates during the next academic year. Page lengths for all books published during the next calendar year. Batting averages for all major league players during the next baseball season.

b. Concrete: Probability: In a sample of 5 mutual funds, what is the chance that all 5 have rates of return which exceeded 10% last year?

Statistics: If previous year rates-of-return for 5 mutual funds were 9.6, 14.5, 8.3, 9.9 and 10.2, can we conclude that the average rate for all funds was below 10%?

Conceptual: Probability: In a sample of 10 books to be published next year, how likely is it that the average number of pages for the 10 is between 200 and 250?

Statistics: If the sample average number of pages for 10 books is 227, can we be highly confident that the average for all books is between 200 and 245?

5.

a. No, the relevant conceptual population is all scores of all students who participate in the SI in conjunction with this particular statistics course.

b. The advantage to randomly choosing students to participate in the two groups is that we are more likely to get a sample representative of the population at large. If it were left to students to choose, there may be a division of abilities in the two groups which could unnecessarily affect the outcome of the experiment.

c. If all students were put in the treatment group there would be no results with which to compare the treatments.

6. One could take a simple random sample of students from all students in the California State University system and ask each student in the sample to report the distance form their hometown to campus. Alternatively, the sample could be generated by taking a stratified random sample by taking a simple random sample from each of the 23 campuses and again asking each student in the sample to report the distance from their hometown to campus. Certain problems might arise with self reporting of distances, such as recording error or poor recall. This study is enumerative because there exists a finite, identifiable population of objects from which to sample.

7. One could generate a simple random sample of all single family homes in the city or a stratified random sample by taking a simple random sample from each of the 10 district neighborhoods. From each of the homes in the sample the necessary variables would be collected. This would be an enumerative study because there exists a finite, identifiable population of objects from which to sample.

8.

a. Number observations equal 2 x 2 x 2 = 8

b. This could be called an analytic study because the data would be collected on an existing process. There is no sampling frame.

9.

a. There could be several explanations for the variability of the measurements. Among them could be measuring error, (due to mechanical or technical changes across measurements), recording error, differences in weather conditions at time of measurements, etc.

b. This could be called an analytic study because there is no sampling frame.

Section 1.2

10.

a. Minitab generates the following stem-and-leaf display of this data:

5	9	
6	33588	
7	00234677889	
8	127	
9	077	stem: ones
10	7	leaf: tenths
11	368	

What constitutes large or small variation usually depends on the application at hand, but an often-used rule of thumb is: the variation tends to be large whenever the spread of the data (the difference between the largest and smallest observations) is large compared to a representative value. Here, 'large' means that the percentage is closer to 100% than it is to 0%. For this data, the spread is 11 - 5 = 6, which constitutes 6/8 = .75, or, 75%, of the typical data value of 8. Most researchers would call this a large amount of variation.

b. The data display is not perfectly symmetric around some middle/representative value. There tends to be some positive skewness in this data.

c. In Chapter 1, outliers are data points that appear to be *very* different from the pack. Looking at the stem-and-leaf display in part (a), there appear to be no outliers in this data. (Chapter 2 gives a more precise definition of what constitutes an outlier).

d. From the stem-and-leaf display in part (a), there are 4 values greater than 10. Therefore, the proportion of data values that exceed 10 is 4/27 = .148, or, about 15%.

11.

6l	034
6h	667899
7l	00122244
7h	
8l	001111122344
8h	5557899
9l	03
9h	58

Stem=Tens
Leaf=Ones

This display brings out the gap in the data:
There are no scores in the high 70's.

12. One method of denoting the pairs of stems having equal values is to denote the first stem by L, for 'low', and the second stem by H, for 'high'. Using this notation, the stem-and-leaf display would appear as follows:

3L	1
3H	56678
4L	000112222234
4H	5667888
5L	144
5H	58
6L	2
6H	6678
7L	
7H	5

stem: tenths
leaf: hundredths

The stem-and-leaf display on the previous page shows that .45 is a good representative value for the data. In addition, the display is not symmetric and appears to be positively skewed. The spread of the data is .75 - .31 = .44, which is.44/.45 = .978, or about 98% of the typical value of .45. This constitutes a reasonably large amount of variation in the data. The data value .75 is a possible outlier

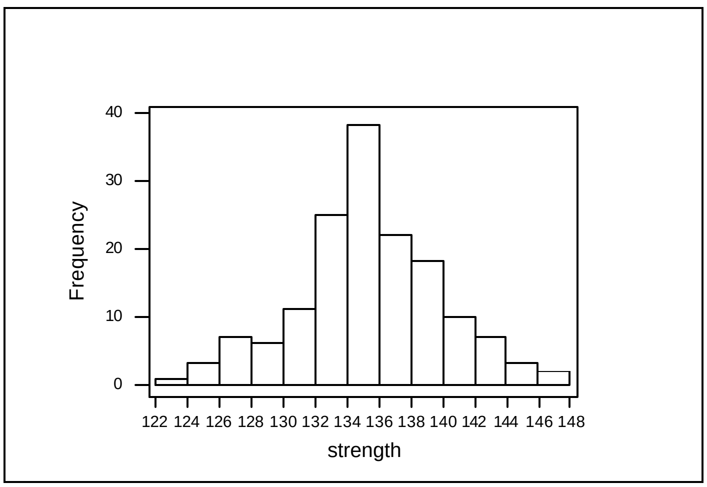
13.

a.

12	2	Leaf = ones
12	445	Stem = tens
12	6667777	
12	889999	
13	00011111111	
13	2222222222333333333333333	
13	44444444444444444455555555555555555555	
13	6666666666667777777777	
13	888888888888999999	
14	0000001111	
14	2333333	
14	444	
14	77	

The observations are highly concentrated at 134 – 135, where the display suggests the typical value falls.

b.



The histogram is symmetric and unimodal, with the point of symmetry at approximately 135.

14.

a.

Stem	Leaf
2	23
3	2344567789
4	01356889
5	00001114455666789
6	0000122223344456667789999
7	00012233455555668
8	02233448
9	012233335666788
10	2344455688
11	2335999
12	37
13	8
14	36
15	0035
16	
17	
18	9

stem units: 1.0
leaf units: .10

b. A representative value could be the median, 7.0.

c. The data appear to be highly concentrated, except for a few values on the positive side.

d. No, the data is skewed to the right, or positively skewed.

e. The value 18.9 appears to be an outlier, being more than two stem units from the previous value.

15.

Crunchy		Creamy
	2	2
644	3	69
77220	4	145
6320	5	3666
222	6	258
55	7	
0	8	

Both sets of scores are reasonably spread out. There appear to be no outliers. The three highest scores are for the crunchy peanut butter, the three lowest for the creamy peanut butter.

16.

a.

beams		cylinders
9	5	8
88533	6	16
98877643200	7	012488
721	8	13359
770	9	278
7	10	
863	11	2
	12	6
	13	
	14	1

The data appears to be slightly skewed to the right, or positively skewed. The value of 14.1 appears to be an outlier. Three out of the twenty, 3/20 or .15 of the observations exceed 10 Mpa.

b. The majority of observations are between 5 and 9 Mpa for both beams and cylinders, with the modal class in the 7 Mpa range. The observations for cylinders are more variable, or spread out, and the maximum value of the cylinder observations is higher.

c. Dot Plot

```
          . .  .  :..  : .: . . .   :         .        .         .
          -+---------+---------+---------+---------+---------+-----
cylinder
         6.0       7.5       9.0      10.5      12.0      13.5
```

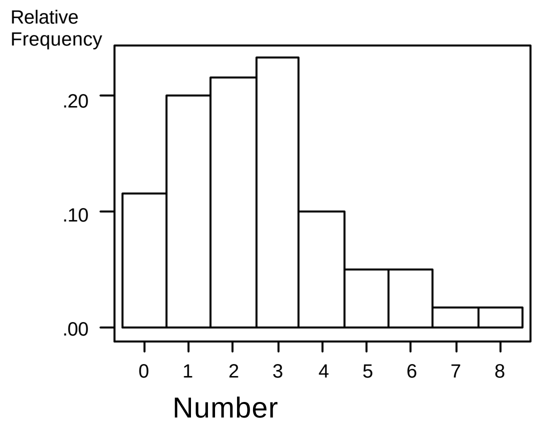
17.

a.

Number Nonconforming	Frequency	RelativeFrequency(Freq/60)
0	7	0.117
1	12	0.200
2	13	0.217
3	14	0.233
4	6	0.100
5	3	0.050
6	3	0.050
7	1	0.017
8	1	0.017
doesn't add exactly to 1 because relative frequencies have been rounded		1.001

b. The number of batches with at most 5 nonconforming items is 7+12+13+14+6+3 = 55, which is a proportion of 55/60 = .917. The proportion of batches with (strictly) fewer than 5 nonconforming items is 52/60 = .867. Notice that these proportions could also have been computed by using the relative frequencies: e.g., proportion of batches with 5 or fewer nonconforming items = 1- (.05+.017+.017) = .916; proportion of batches with fewer than 5 nonconforming items = 1 - (.05+.05+.017+.017) = .866.

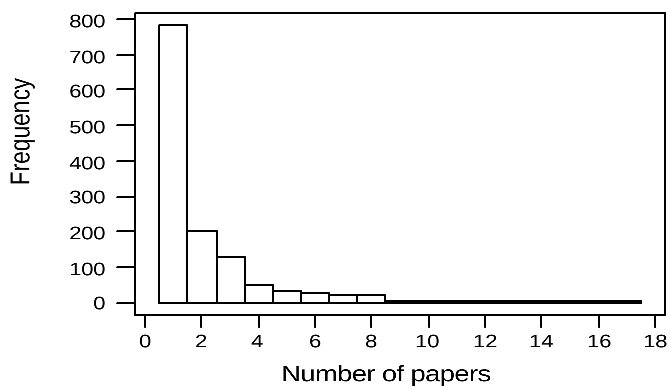
c. The following is a Minitab histogram of this data. The center of the histogram is somewhere around 2 or 3 and it shows that there is some positive skewness in the data. Using the rule of thumb in Exercise 1, the histogram also shows that there is a lot of spread/variation in this data.



18.

a.

The following histogram was constructed using Minitab:



The most interesting feature of the histogram is the heavy positive skewness of the data.

Note: One way to have Minitab automatically construct a histogram from grouped data such as this is to use Minitab's ability to enter multiple copies of the same number by typing, for example, 784(1) to enter 784 copies of the number 1. The frequency data in this exercise was entered using the following Minitab commands:

```
MTB > set c1
DATA> 784(1) 204(2) 127(3) 50(4) 33(5) 28(6) 19(7) 19(8)
DATA> 6(9) 7(10) 6(11) 7(12) 4(13) 4(14) 5(15) 3(16) 3(17)
DATA> end
```

b. From the frequency distribution (or from the histogram), the number of authors who published at least 5 papers is 33+28+19+…+5+3+3 = 144, so the proportion who published 5 or more papers is 144/1309 = .11, or 11%. Similarly, by adding frequencies and dividing by n = 1309, the proportion who published 10 or more papers is 39/1309 = .0298, or about 3%. The proportion who published more than 10 papers (i.e., 11 or more) is 32/1309 = .0245, or about 2.5%.

c. No. Strictly speaking, the class described by ' ≥15 ' has no upper boundary, so it is impossible to draw a rectangle above it having finite area (i.e., frequency).

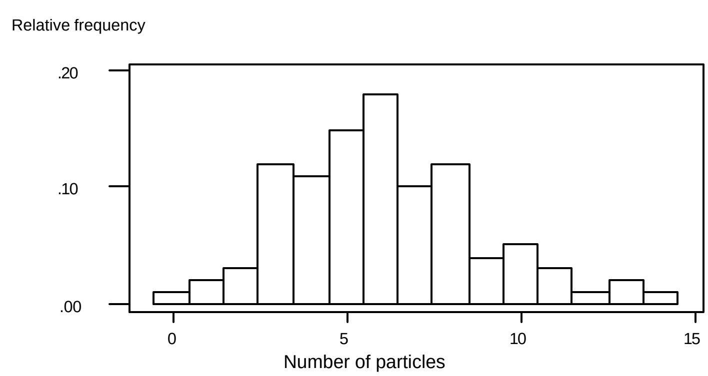
d. The category 15-17 does have a finite width of 2, so the cumulated frequency of 11 can be plotted as a rectangle of height 6.5 over this interval. The basic rule is to make the area of the bar equal to the class frequency, so area = 11 = (width)(height) = 2(height) yields a height of 6.5.

19.

a. From this frequency distribution, the proportion of wafers that contained at least one particle is (100-1)/100 = .99, or 99%. Note that it is much easier to subtract 1 (which is the number of wafers that contain 0 particles) from 100 than it would be to add all the frequencies for 1, 2, 3,… particles. In a similar fashion, the proportion containing at least 5 particles is (100 - 1-2-3-12-11)/100 = 71/100 = .71, or, 71%.

b. The proportion containing between 5 and 10 particles is (15+18+10+12+4+5)/100 = 64/100 = .64, or 64%. The proportion that contain strictly between 5 and 10 (meaning strictly *more* than 5 and strictly *less* than 10) is (18+10+12+4)/100 = 44/100 = .44, or 44%.

c. The following histogram was constructed using Minitab. The data was entered using the same technique mentioned in the answer to exercise 8(a). The histogram is *almost* symmetric and unimodal; however, it has a few relative maxima (i.e., modes) and has a very slight positive skew.



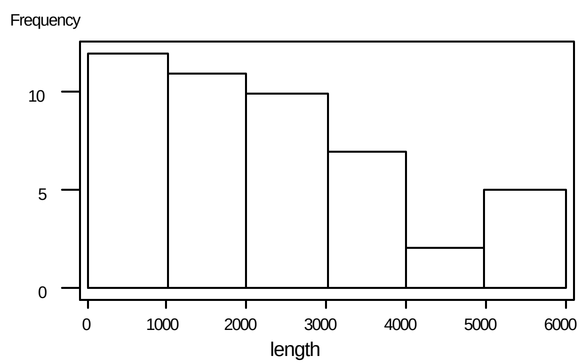
20.

a. The following stem-and-leaf display was constructed:

0	123334555599	
1	00122234688	stem: thousands
2	1112344477	leaf: hundreds
3	0113338	
4	37	
5	23778	

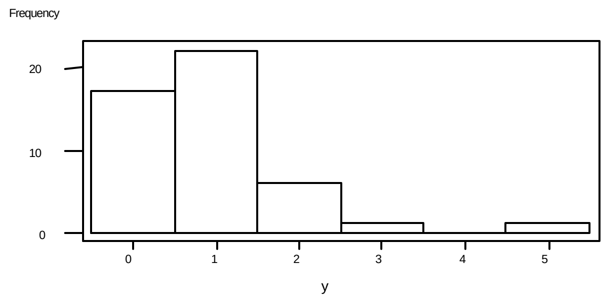
A typical data value is somewhere in the low 2000's. The display is almost unimodal (the stem at 5 would be considered a mode, the stem at 0 another) and has a positive skew.

b. A histogram of this data, using classes of width 1000 centered at 0, 1000, 2000, 6000 is shown below. The proportion of subdivis ions with total length less than 2000 is (12+11)/47 = .489, or 48.9%. Between 200 and 4000, the proportion is (7 + 2)/47 = .191, or 19.1%. The histogram shows the same general shape as depicted by the stem-and-leaf in part (a).

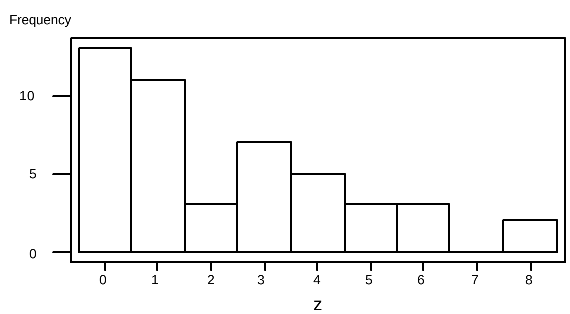


21.

a. A histogram of the y data appears below. From this histogram, the number of subdivisions having no cul-de-sacs (i.e., $y = 0$) is 17/47 = .362, or 36.2%. The proportion having at least one cul-de-sac ($y \geq 1$) is (47-17)/47 = 30/47 = .638, or 63.8%. Note that subtracting the number of cul-de-sacs with $y = 0$ from the total, 47, is an easy way to find the number of subdivisions with $y \geq 1$.



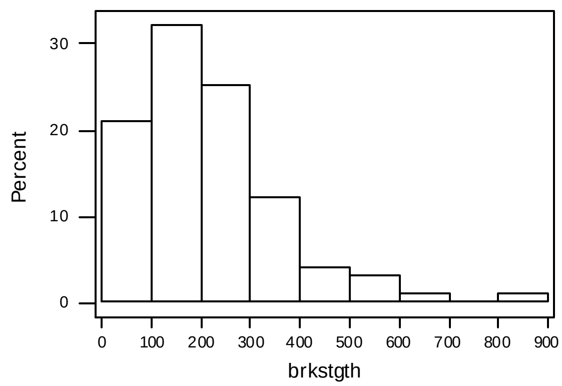
b. A histogram of the z data appears below. From this histogram, the number of subdivisions with at most 5 intersections (i.e., $z \leq 5$) is 42/47 = .894, or 89.4%. The proportion having fewer than 5 intersections ($z < 5$) is 39/47 = .830, or 83.0%.



22. A very large percentage of the data values are greater than 0, which indicates that most, but not all, runners do slow down at the end of the race. The histogram is also positively skewed, which means that some runners slow down a *lot* compared to the others. A typical value for this data would be in the neighborhood of 200 seconds. The proportion of the runners who ran the last 5 km faster than they did the first 5 km is very small, about 1% or so.

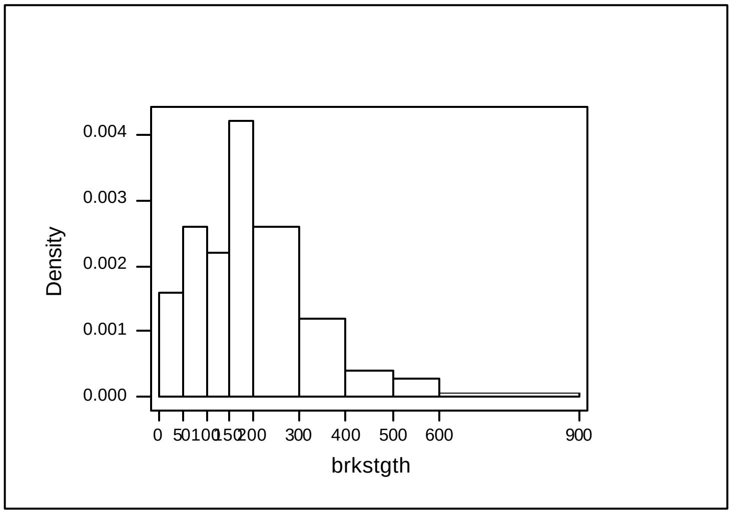
23.

a.



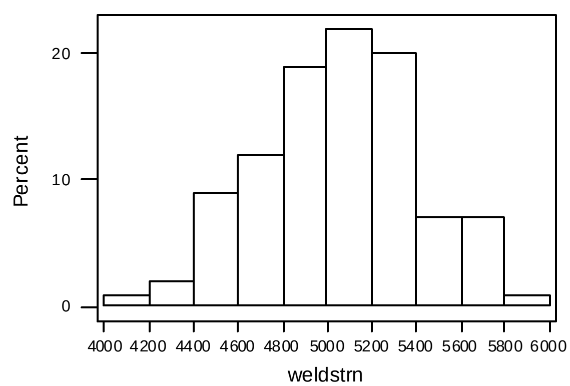
The histogram is skewed right, with a majority of observations between 0 and 300 cycles. The class holding the most observations is between 100 and 200 cycles.

b.

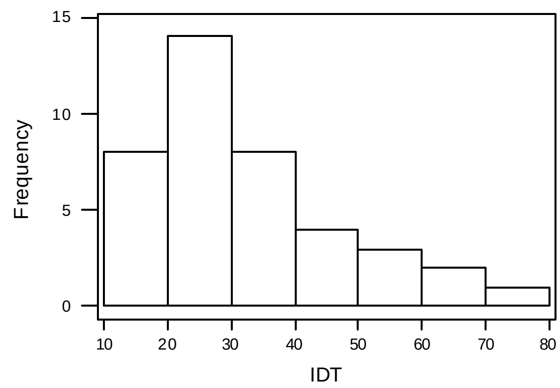


c [proportion ≥ 100] = 1 – [proportion < 100] = 1 - .21 = .79

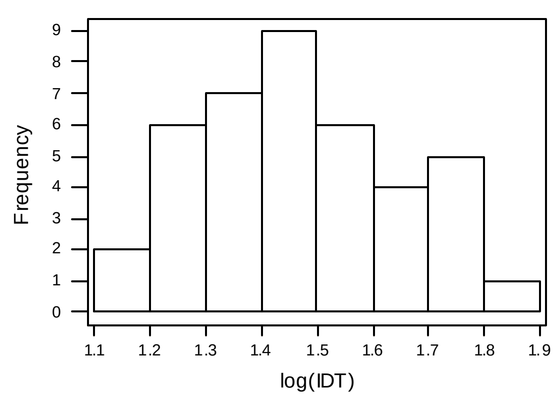
24.



25. Histogram of original data:



Histogram of transformed data:

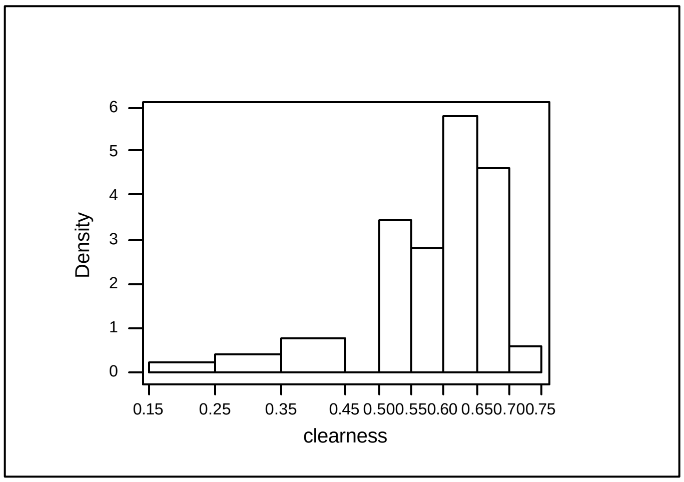


The transformation creates a much more symmetric, mound-shaped histogram.

26.

a.

Class Intervals	**Frequency**	**Rel. Freq.**
.15 -< .25	8	0.02192
.25 -< .35	14	0.03836
.35 -< .45	28	0.07671
.45 -< .50	24	0.06575
.50 -< .55	39	0.10685
.55 -< .60	51	0.13973
.60 -< .65	106	0.29041
.65 -< .70	84	0.23014
.70 -< .75	11	0.03014
	n=365	1.00001



b. The proportion of days with a clearness index smaller than .35 is $\frac{(8+4)}{365} = .06$, or 6%.

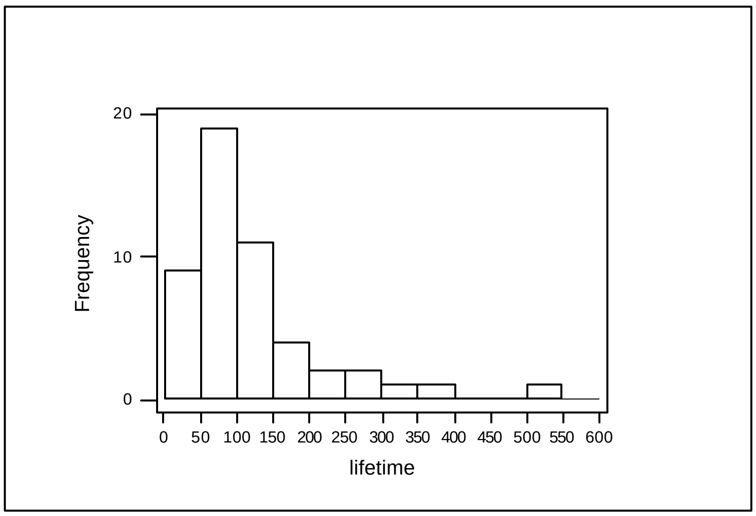
c. The proportion of days with a clearness index of at least .65 is $\frac{(84+11)}{365} = .26$, or 26%.

27.

a. The endpoints of the class intervals overlap. For example, the value 50 falls in both of the intervals ‘0 – 50’ and ’50 – 100’.

b.

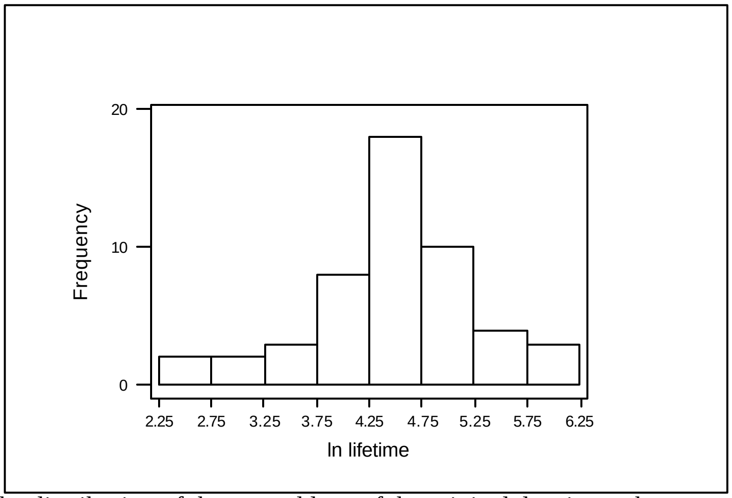
Class Interval	Frequency	Relative Frequency
0 - < 50	9	0.18
50 - < 100	19	0.38
100 - < 150	11	0.22
150 - < 200	4	0.08
200 - < 250	2	0.04
250 - < 300	2	0.04
300 - < 350	1	0.02
350 - < 400	1	0.02
>= 400	1	0.02
	50	1.00



The distribution is skewed to the right, or positively skewed. There is a gap in the histogram, and what appears to be an outlier in the ‘500 – 550’ interval.

c.

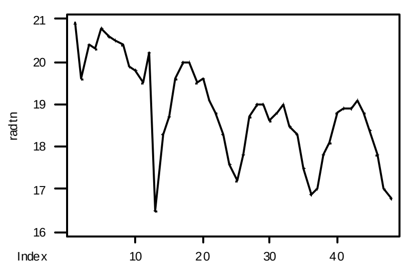
Class Interval	Frequency	Relative Frequency
2.25 - < 2.75	2	0.04
2.75 - < 3.25	2	0.04
3.25 - < 3.75	3	0.06
3.75 - < 4.25	8	0.16
4.25 - < 4.75	18	0.36
4.75 - < 5.25	10	0.20
5.25 - < 5.75	4	0.08
5.75 - < 6.25	3	0.06



The distribution of the natural logs of the original data is much more symmetric than the original.

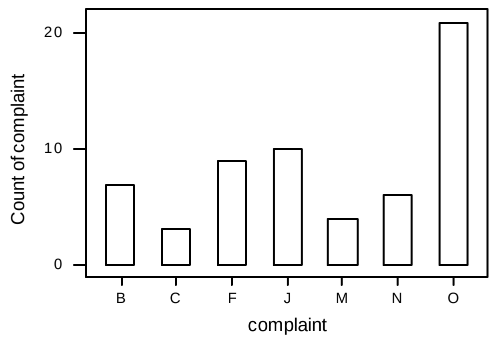
d. The proportion of lifetime observations in this sample that are less than 100 is .18 + .38 = .56, and the proportion that is at least 200 is .04 + .04 + .02 + .02 + .02 = .14.

28. There are seasonal trends with lows and highs 12 months apart.

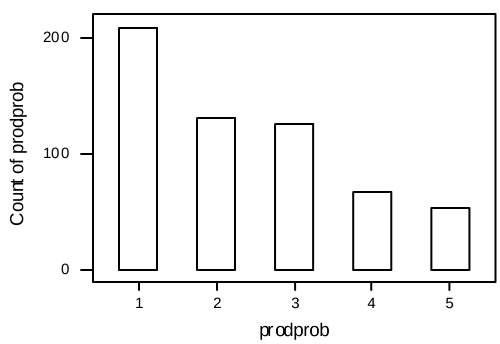


29.

Complaint	Frequency	Relative Frequency
B	7	0.1167
C	3	0.0500
F	9	0.1500
J	10	0.1667
M	4	0.0667
N	6	0.1000
O	21	0.3500
	60	1.0000



30.



1. incorrect component
2. missing component
3. failed component
4. insufficient solder
5. excess solder

31.

Class	Frequency	Relative Frequency	Cumulative Relative Frequency
0.0 - under 4.0	2	2	0.050
4.0 - under 8.0	14	16	0.400
8.0 - under 12.0	11	27	0.675
12.0 - under 16.0	8	35	0.875
16.0 - under 20.0	4	39	0.975
20.0 - under 24.0	0	39	0.975
24.0 - under 28.0	1	40	1.000

32.

a. The frequency distribution is:

Class	Relative Frequency	Class	Relative Frequency
0-< 150	.193	900-<1050	.019
150-< 300	.183	1050-<1200	.029
300-< 450	.251	1200-<1350	.005
450-< 600	.148	1350-<1500	.004
600-< 750	.097	1500-<1650	.001
750-< 900	.066	1650-<1800	.002
		1800-<1950	.002

The relative frequency distribution is almost unimodal and exhibits a large positive skew. The typical middle value is somewhere between 400 and 450, although the skewness makes it difficult to pinpoint more exactly than this.

b. The proportion of the fire loads less than 600 is .193+.183+.251+.148 = .775. The proportion of loads that are at least 1200 is .005+.004+.001+.002+.002 = .014.

c. The proportion of loads between 600 and 1200 is 1 - .775 - .014 = .211.

Section 1.3

33.

a. $\bar{x} = 192.57$, $\tilde{x} = 189$. The mean is larger than the median, but they are still fairly close together.

b. Changing the one value, $\bar{x} = 189.71$, $\tilde{x} = 189$. The mean is lowered, the median stays the same.

c. $\bar{x}_{tr} = 191.0$. $\frac{1}{14} = .07$ or 7% trimmed from each tail.

d. For n = 13, Σx = (119.7692) x 13 = 1,557
For n = 14, Σx = 1,557 + 159 = 1,716

$$\bar{x} = \frac{1716}{14} = 122.5714 \text{ or } 122.6$$

34.

a. The sum of the n = 11 data points is 514.90, so $\bar{x}$ = 514.90/11 = 46.81.

b. The sample size (n = 11) is odd, so there will be a middle value. Sorting from smallest to largest: 4.4 16.4 22.2 30.0 33.1 36.6 40.4 66.7 73.7 81.5 109.9. The sixth value, 36.6 is the middle, or median, value. The mean differs from the median because the largest sample observations are much further from the median than are the smallest values.

c. Deleting the smallest (x = 4.4) and largest (x = 109.9) values, the sum of the remaining 9 observations is 400.6. The trimmed mean $\bar{x}_{tr}$ is 400.6/9 = 44.51. The trimming percentage is 100(1/11) ≈ 9.1%. $\bar{x}_{tr}$ lies between the mean and median.

35.

a. The sample mean is $\bar{x}$ = (100.4/8) = 12.55.

The sample size (n = 8) is even. Therefore, the sample median is the average of the (n/2) and (n/2) + 1 values. By sorting the 8 values in order, from smallest to largest: 8.0 8.9 11.0 12.0 13.0 14.5 15.0 18.0, the forth and fifth values are 12 and 13. The sample median is (12.0 + 13.0)/2 = 12.5.

The 12.5% trimmed mean requires that we first trim (.125)(n) or 1 value from the ends of the ordered data set. Then we average the remaining 6 values. The 12.5% trimmed mean $\bar{x}_{tr(12.5)}$ is 74.4/6 = 12.4.

All three measures of center are similar, indicating little skewness to the data set.

b. The smallest value (8.0) could be increased to any number below 12.0 (a change of less than 4.0) without affecting the value of the sample median.

c. The values obtained in part (a) can be used directly. For example, the sample mean of 12.55 psi could be re-expressed as

$$(12.55 \text{ psi}) \text{ x} \left(\frac{1ksi}{2.2\,psi}\right) = 5.70\,ksi\,.$$

36.

a. A stem-and leaf display of this data appears below:

32	55	stem: ones
33	49	leaf: tenths
34		
35	6699	
36	34469	
37	03345	
38	9	
39	2347	
40	23	
41		
42	4	

The display is reasonably symmetric, so the mean and median will be close.

b. The sample mean is $\bar{x}$ = 9638/26 = 370.7. The sample median is
$\tilde{x}$ = (369+370)/2 = 369.50.

c. The largest value (currently 424) could be increased by any amount. Doing so will not change the fact that the middle two observations are 369 and 170, and hence, the median will not change. However, the value x = 424 can not be changed to a number less than 370 (a change of 424-370 = 54) since that *will* lower the values(s) of the two middle observations.

d. Expressed in minutes, the mean is (370.7 sec)/(60 sec) = 6.18 min; the median is 6.16 min.

37. $\bar{x} = 12.01$, $\tilde{x} = 11.35$, $\bar{x}_{tr(10)} = 11.46$. The median or the trimmed mean would be good choices because of the outlier 21.9.

38.

a. The reported values are (in increasing order) 110, 115, 120, 120, 125, 130, 130, 135, and 140. Thus the median of the reported values is 125.

b. 127.6 is reported as 130, so the median is now 130, a very substantial change. When there is rounding or grouping, the median can be highly sensitive to small change.

39.

a. $\Sigma x_l = 16.475$ so $\bar{x} = \dfrac{16.475}{16} = 1.0297$

$$\tilde{x} = \frac{(1.007 + 1.011)}{2} = 1.009$$

b. 1.394 can be decreased until it reaches 1.011(the largest of the 2 middle values) – i.e. by 1.394 – 1.011 = .383, If it is decreased by more than .383, the median will change.

40. $\tilde{x} = 60.8$

$\bar{x}_{tr(25)} = 59.3083$

$\bar{x}_{tr(10)} = 58.3475$

$\bar{x} = 58.54$

All four measures of center have about the same value.

41.

a. $7/10 = .70$

b. $\bar{x} = .70$ = proportion of successes

c. $\dfrac{s}{25} = .80$ so s = (0.80)(25) = 20

total of 20 successes

20 – 7 = 13 of the new cars would have to be successes

42.

a. $\bar{y} = \dfrac{\Sigma y_i}{n} = \dfrac{\Sigma(x_i + c)}{n} = \dfrac{\Sigma x_i}{n} + \dfrac{nc}{n} = \bar{x} + c$

$\tilde{y} =$ the median of $(x_1 + c, x_2 + c, ..., x_n + c) =$ median of $(x_1, x_2, ..., x_n) + c = \tilde{x} + c$

b. $\bar{y} = \dfrac{\Sigma y_i}{n} = \dfrac{\Sigma(x_i \cdot c)}{n} = \dfrac{c\Sigma x_i}{n} = c\bar{x}$

$\tilde{y} = (cx_1, cx_2, ..., cx_n) = c \cdot median(x_1, x_2, ..., x_n) = c\tilde{x}$

43. median = $\dfrac{(57 + 79)}{2} = 68.0$, 20% trimmed mean = 66.2, 30% trimmed mean = 67.5.

Section 1.4

44.

a. range = 49.3 – 23.5 = 25.8

b.

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	x_i^2
29.5	-1.53	2.3409	870.25
49.3	18.27	333.7929	2430.49
30.6	-0.43	0.1849	936.36
28.2	-2.83	8.0089	795.24
28.0	-3.03	9.1809	784.00
26.3	-4.73	22.3729	691.69
33.9	2.87	8.2369	1149.21
29.4	-1.63	2.6569	864.36
23.5	-7.53	56.7009	552.25
31.6	0.57	0.3249	998.56
$\Sigma x = 310.3$	$\Sigma(x_i - \bar{x}) = 0$	$\Sigma(x_i - \bar{x})^2 = 443.801$	$\Sigma(x_i^2) = 10{,}072.41$

$$\bar{x} = 31.03$$

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{443.801}{9} = 49.3112$$

c. $s = \sqrt{s^2} = 7.0222$

d. $s^2 = \dfrac{\Sigma x^2 - (\Sigma x)^2/n}{n-1} = \dfrac{10{,}072.41 - (310.3)^2/10}{9} = 49.3112$

45.

a. $\bar{x} = \frac{1}{n}\sum_i x_i$ = 577.9/5 = 115.58. Deviations from the mean:
116.4 - 115.58 = .82, 115.9 - 115.58 = .32, 114.6 -115.58 = -.98,
115.2 - 115.58 = -.38, and 115.8-115.58 = .22.

b. $s^2 = [(.82)^2 + (.32)^2 + (-.98)^2 + (-.38)^2 + (.22)^2]/(5-1) = 1.928/4 = .482$,
so s = .694.

c. $\sum_i x_i^2 = 66{,}795.61$, so $s^2 = \frac{1}{n-1}\left[\sum_i x_i^2 - \frac{1}{n}\left(\sum_i x_i\right)^2\right] =$
$[66{,}795.61 - (577.9)^2/5]/4 = 1.928/4 = .482$.

d. Subtracting 100 from all values gives $\bar{x} = 15.58$, all deviations are the same as in part b, and the transformed variance is identical to that of part b.

46.

a. $\bar{x} = \frac{1}{n}\sum_i x_i$ = 14438/5 = 2887.6. The sorted data is: 2781 2856 2888 2900 3013, so the sample median is $\tilde{x}$ = 2888.

b. Subtracting a constant from each observation shifts the data, but does not change its sample variance (Exercise 16). For example, by subtracting 2700 from each observation we get the values 81, 200, 313, 156, and 188, which are smaller (fewer digits) and easier to work with. The sum of squares of this transformed data is 204210 and its sum is 938, so the computational formula for the variance gives s^2 = [204210-(938)^2/5]/(5-1) = 7060.3.

47. The sample mean, $\bar{x} = \frac{1}{n}\sum x_i = \frac{1}{10}(1{,}162) = \bar{x} = 116.2$.

The sample standard deviation, $s = \sqrt{\dfrac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\dfrac{140{,}992 - \frac{(1{,}162)^2}{10}}{9}} = 25.75$

On average, we would expect a fracture strength of 116.2. In general, the size of a typical deviation from the sample mean (116.2) is about 25.75. Some observations may deviate from 116.2 by more than this and some by less.

48. Using the computational formula, $s^2 = \frac{1}{n-1}\left[\sum_i x_i^2 - \frac{1}{n}\left(\sum_i x_i\right)^2\right] =$

[3,587,566-(9638)^2/26]/(26-1) = 593.3415, so s = 24.36. In general, the size of a typical deviation from the sample mean (370.7) is about 24.4. Some observations may deviate from 370.7 by a little more than this, some by less.

49.

a. $\Sigma x = 2.75 + ... + 3.01 = 56.80$, $\Sigma x^2 = (2.75)^2 + ... + (3.01)^2 = 197.8040$

b. $s^2 = \dfrac{197.8040 - (56.80)^2/17}{16} = \dfrac{8.0252}{16} = .5016, \quad s = .708$

50. First, we need $\bar{x} = \frac{1}{n}\sum x_i = \frac{1}{27}(20{,}179) = 747.37$. Then we need the sample standard deviation $s = \sqrt{\dfrac{24{,}657{,}511 - \frac{(20{,}179)^2}{27}}{26}} = 606.89$. The maximum award should be $\bar{x} + 2s = 747.37 + 2(606.89) = 1961.16$, or in dollar units, $1,961,160. This is quite a bit less than the $3.5 million that was awarded originally.

51.

a. $\Sigma x = 2563$ and $\Sigma x^2 = 368{,}501$, so

$$s^2 = \frac{[368{,}501 - (2563)^2/19]}{18} = 1264.766 \text{ and } s = 35.564$$

b. If y = time in minutes, then y = cx where $c = \frac{1}{60}$, so

$$s_y^2 = c^2 s_x^2 = \frac{1264.766}{3600} = .351 \text{ and } s_y = cs_x = \frac{35.564}{60} = .593$$

52. Let *d* denote the fifth deviation. Then $.3 + .9 + 1.0 + 1.3 + d = 0$ or $3.5 + d = 0$, so $d = -3.5$. One sample for which these are the deviations is $x_1 = 3.8,\ x_2 = 4.4,\ x_3 = 4.5,\ x_4 = 4.8,\ x_5 = 0.$ (obtained by adding 3.5 to each deviation; adding any other number will produce a different sample with the desired property)

53.

a. lower half: 2.34 2.43 2.62 2.74 2.74 2.75 2.78 3.01 3.46
upper half: 3.46 3.56 3.65 3.85 3.88 3.93 4.21 4.33 4.52
Thus the lower fourth is 2.74 and the upper fourth is 3.88.

b. $f_s = 3.88 - 2.74 = 1.14$

c. f_s wouldn't change, since increasing the two largest values does not affect the upper fourth.

d. By at most .40 (that is, to anything not exceeding 2.74), since then it will not change the lower fourth.

e. Since n is now even, the lower half consists of the smallest 9 observations and the upper half consists of the largest 9. With the lower fourth = 2.74 and the upper fourth = 3.93, $f_s = 1.19$.

54.

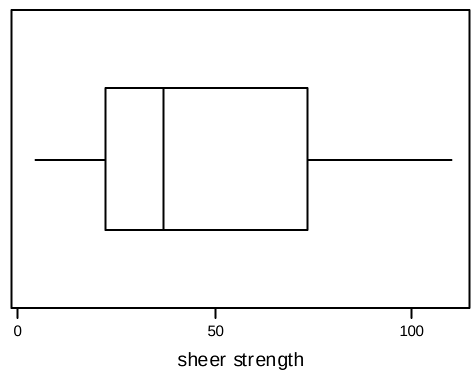
a. The lower half of the data set: 4.4 16.4 22.2 30.0 33.1 36.6, whose median, and therefore, the lower quartile, is $\frac{(22.2+30.0)}{2}+26.1.$

The top half of the data set: 36.6 40.4 66.7 73.7 81.5 109.9, whose median, and therefore, the upper quartile, is $\frac{(66.7+73.7)}{2}=70.2$.

So, the IQR = (70.2 – 26.1) = 44.1

b.

A boxplot (created in Minitab) of this data appears below:



There is a slight positive skew to the data. The variation seems quite large. There are no outliers.

c. An observation would need to be further than 1.5(44.1) = 66.15 units below the lower quartile $[(26.1-66.15)=-40.05\ units]$ or above the upper quartile $[(70.2+66.15)=136.35\ units]$ to be classified as a mild outlier. Notice that, in this case, an outlier on the lower side would not be possible since the sheer strength variable cannot have a negative value.

An extreme outlier would fall (3)44.1) = 132.3 or more units below the lower, or above the upper quartile. Since the minimum and maximum observations in the data are 4.4 and 109.9 respectively, we conclude that there are no outliers, of either type, in this data set.

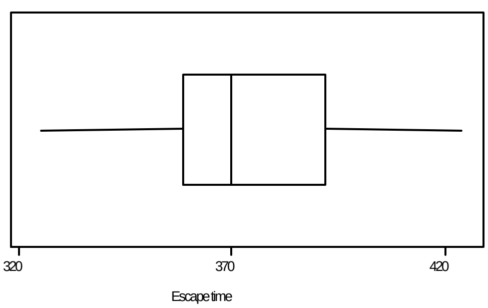
d. Not until the value x = 109.9 is lowered below 73.7 would there be any change in the value of the upper quartile. That is, the value x = 109.9 could not be decreased by more than (109.9 – 73.7) = 36.2 units.

55.

a. Lower half of the data set: 325 325 334 339 356 356 359 359 363 364 364 366 369, whose median, and therefore the lower quartile, is 359 (the 7^{th} observation in the sorted list).
The top half of the data is 370 373 373 374 375 389 392 393 394 397 402 403 424, whose median, and therefore the upper quartile is 392. So, the IQR = 392 - 359 = 33.

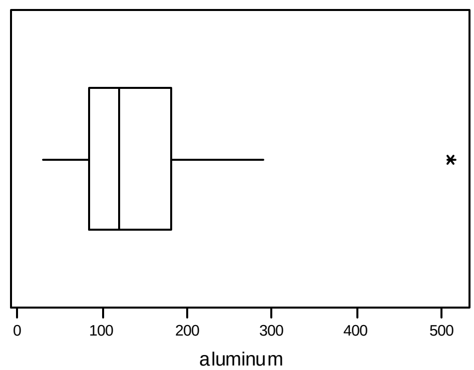
b. 1.5(IQR) = 1.5(33) = 49.5 and 3(IQR) = 3(33) = 99. Observations that are further than 49.5 below the lower quartile (i.e., 359-49.5 = 309.5 or less) or more than 49.5 units above the upper quartile (greater than 392+49.5 = 441.5) are classified as 'mild' outliers. 'Extreme' outliers would fall 99 or more units below the lower, or above the upper, quartile. Since the minimum and maximum observations in the data are 325 and 424, we conclude that there are no mild outliers in this data (and therefore, no 'extreme' outliers either).

c. A boxplot (created by Minitab) of this data appears below. There is a slight positive skew to the data, but it is not far from being symmetric. The variation, however, seems large (the spread 424-325 = 99 is a large percentage of the median/typical value)



d. Not until the value x = 424 is lowered below the upper quartile value of 392 would there be any change in the value of the upper quartile. That is, the value x = 424 could not be decreased by more than 424-392 = 32 units.

56. A boxplot (created in Minitab) of this data appears below.

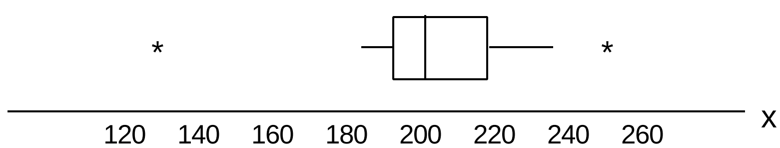


There is a slight positive skew to this data. There is one extreme outler (x=511). Even when removing the outlier, the variation is still moderately large.

57.

a. 1.5(IQR) = 1.5(216.8-196.0) = 31.2 and 3(IQR) = 3(216.8-196.0) = 62.4.
Mild outliers: observations below 196-31.2 = 164.6 or above 216.8+31.2 = 248.
Extreme outliers: observations below 196-62.4 = 133.6 or above 216.8+62.4 = 279.2. Of the observations given, 125.8 is an extreme outlier and 250.2 is a mild outlier.

b. A boxplot of this data appears below. There is a bit of positive skew to the data but, except for the two outliers identified in part (a), the variation in the data is relatively small.



58. The most noticeable feature of the comparative boxplots is that machine 2's sample values have considerably more variation than does machine 1's sample values. However, a typical value, as measured by the median, seems to be about the same for the two machines. The only outlier that exists is from machine 1.

59.

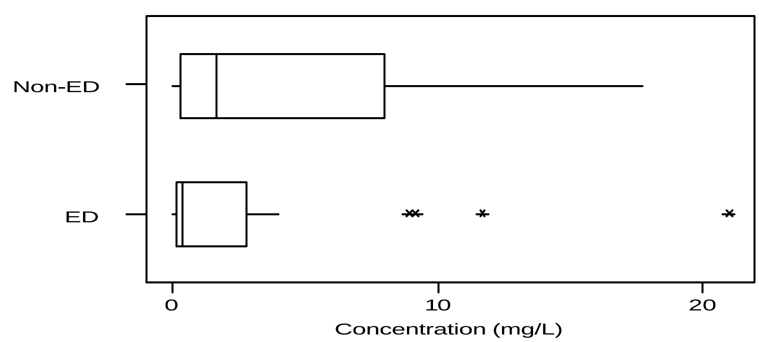
a. ED: median = .4 (the 14th value in the *sorted* list of data). The lower quartile (median of the lower half of the data, including the median, since n is odd) is (.1+.1)/2 = .1. The upper quartile is (2.7+2.8)/2 = 2.75. Therefore, IQR = 2.75 - .1 = 2.65.

Non-ED: median = (1.5+1.7)/2 = 1.6. The lower quartile (median of the lower 25 observations) is .3; the upper quartile (median of the upper half of the data) is 7.9. Therefore, IQR = 7.9 - .3 = 7.6.

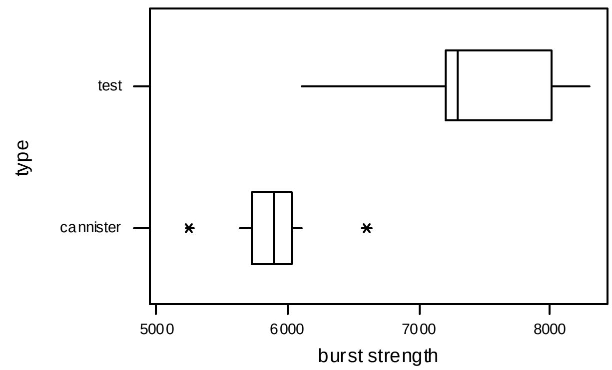
b. ED: mild outliers are less than .1 - 1.5(2.65) = -3.875 or greater than 2.75 + 1.5(2.65) = 6.725. Extreme outliers are less than .1 - 3(2.65) = -7.85 or greater than 2.75 + 3(2.65) = 10.7. So, the two largest observations (11.7, 21.0) are extreme outliers and the next two largest values (8.9, 9.2) are mild outliers. There are no outliers at the lower end of the data.

Non-ED: mild outliers are less than .3 - 1.5(7.6) = -11.1 or greater than 7.9 + 1.5(7.6) = 19.3. Note that there are no mild outliers in the data, hence there can not be any extreme outliers either.

c. A comparative boxplot appears below. The outliers in the ED data are clearly visible. There is noticeable positive skewness in both samples; the Non-Ed data has more variability then the Ed data; the typical values of the ED data tend to be smaller than those for the Non-ED data.



60. A comparative boxplot (created in Minitab) of this data appears below.



The burst strengths for the test nozzle closure welds are quite different from the burst strengths of the production canister nozzle welds.

The test welds have much higher burst strengths and the burst strengths are much more variable.

The production welds have more consistent burst strength and are consistently lower than the test welds. The production welds data does contain 2 outliers.

61. Outliers occur in the 6 a.m. data. The distributions at the other times are fairly symmetric. Variability and the 'typical' values in the data increase a little at the 12 noon and 2 p.m. times.

Supplementary Exercises

62. To somewhat simplify the algebra, begin by subtracting 76,000 from the original data. This transformation will affect each date value and the mean. It will not affect the standard deviation.

$x_1 = 683, \quad x_2 = 1{,}048, \;\; \bar{y} = 831$

$n\bar{x} = (4)(831) = 3{,}324$ so, $x_1 + x_2 + x_3 + x_4 = 3{,}324$

and $x_2 + x_3 = 3{,}324 - x_1 - x_4 = 1{,}593$ and $x_3 = (1{,}593 - x_2)$

Next, $s^2 = (180)^2 = \left[\dfrac{\sum x_i^2 - \dfrac{(3324)^2}{4}}{3} \right]$

So, $\sum x_i^2 = 2{,}859{,}444$, $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 2{,}859{,}444$ and

$x_2^2 + x_3^2 = 2{,}859{,}444 - x_1^2 + x_4^2 = 1{,}294{,}651$

By substituting $x_3 = (1593 - x_2)$ we obtain the equation

$x_2^2 + (1{,}593 - x_2)^2 - 1{,}294{,}651 = 0$.

$x_x^2 - 1{,}593 x_2 + 621{,}499 = 0$

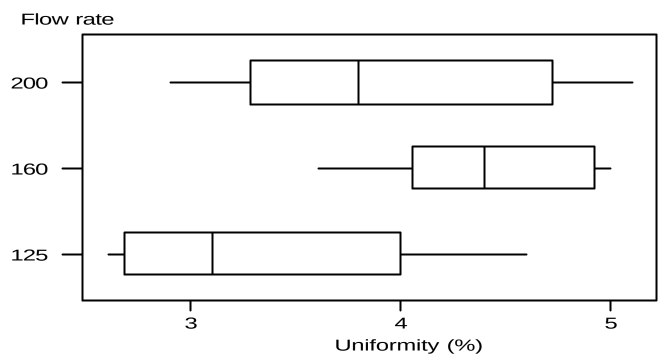
Evaluating for x_2 we obtain $x_2 = 682.8635$ and $x_3 = 1{,}593 - 682.8635 = 910.1365$.

Thus, $x_2 = 76{,}683 \quad x_3 = 76{,}910$.

63.

Flow rate	Median	Lower quartile	Upper quartile	IQR	1.5(IQR)	3(IQR)
125	3.1	2.7	3.8	1.1	1.65	.3
160	4.4	4.2	4.9	.7	1.05	.1
200	3.8	3.4	4.6	1.2	1.80	3.6

There are no outliers in the three data sets. However, as the comparative boxplot below shows, the three data sets differ with respect to their central values (the medians are different) and the data for flow rate 160 is somewhat less variable than the other data sets. Flow rates 125 and 200 also exhibit a small degree of positive skewness.



64.

6	34	stem=ones
7	17	leaf=tenths
8	4589	
9	1	
10	12667789	
11	122499	
12	2	
13	1	

$\bar{x} = 9.9556, \tilde{x} = 10.6$

$s = 1.7594$

$n = 27$

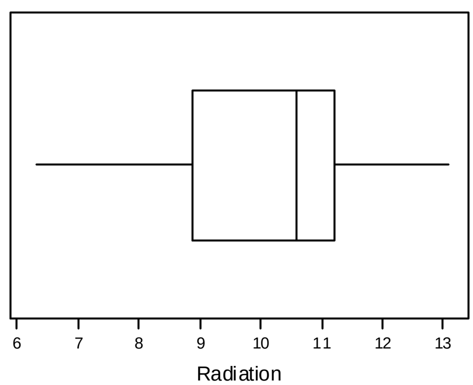
$f_s = 2.3$

lower fourth = 8.85, upper fourth = 11.15

$8.85 - (1.5)(2.3) = 5.4$

$11.15 + (1.5)(2.3) = 14.6$

no outliers



There are no outliers. The distribution is skewed to the left.

65.

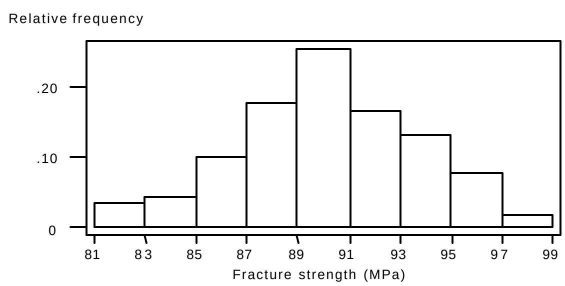
a. HC data: $\sum_i x_i^2 = 2618.42$ and $\sum_i x_i = 96.8$,
so $s^2 = [2618.42 - (96.8)^2/4]/3 = 91.953$
and the sample standard deviation is $s = 9.59$.

CO data: $\sum_i x_i^2 = 145645$ and $\sum_i x_i = 735$, so $s^2 = [145645 - (735)^2/4]/3 = 3529.583$ and the sample standard deviation is $s = 59.41$.

b. The mean of the HC data is 96.8/4 = 24.2; the mean of the CO data is 735/4 = 183.75. Therefore, the coefficient of variation of the HC data is 9.59/24.2 = .3963, or 39.63%. The coefficient of variation of the CO data is 59.41/183.75 = .3233, or 32.33%. Thus, even though the CO data has a larger standard deviation than does the HC data, it actually exhibits *less* variability (in percentage terms) around its average than does the HC data.

66.

a. The histogram appears below. A representative value for this data would be x = 90. The histogram is reasonably symmetric, unimodal, and somewhat bell-shaped. The variation in the data is not small since the spread of the data (99-81 = 18) constitutes about 20% of the typical value of 90.



b. The proportion of the observations that are at least 85 is 1 - (6+7)/169 = .9231. The proportion less than 95 is 1 - (22+13+3)/169 = .7751.

c. x = 90 is the midpoint of the class 89-<91, which contains 43 observations (a relative frequency of 43/169 = .2544. Therefore about half of this frequency, .1272, should be added to the relative frequencies for the classes to the left of x = 90. That is, the approximate proportion of observations that are less than 90 is .0355 + .0414 + .1006 + .1775 + .1272 = .4822.

67.

$$\sum x_i = 163.2$$

$$100\left(\frac{1}{15}\right)\% trimmed mean = \frac{163.2 - 8.5 - 15.6}{13} = 10.70$$

$$100\left(\frac{2}{15}\right)\% trimmed mean = \frac{163.2 - 8.5 - 8.8 - 15.6 - 13.7}{11} = 10.60$$

$$\therefore \frac{1}{2}(100)\left(\frac{1}{15}\right) + \frac{1}{2}(100)\left(\frac{2}{15}\right) = 100\left(\frac{1}{10}\right) = 10\% trimmed mean$$

$$= \frac{1}{2}(10.70) + \frac{1}{2}(10.60) = 10.65$$

68.

a.

$$\frac{d}{dc\left\{\sum (x_i - c)^2\right\}} = \frac{\sum d}{dc(x_i - c)^2} = -2\sum (x_i - c) = 0 \Rightarrow \sum (x_i - c) = 0$$

$$\Rightarrow \sum x_i - \sum c = 0 \Rightarrow \sum x_i - nc = 0 \Rightarrow nc = \sum x_i \Rightarrow c = \frac{\sum x_i}{n} = \bar{x}.$$

b. $\sum (x_i - \bar{x})^2 is smaller than \sum (x_i - m)^2.$

69.

a.

$$\bar{y} = \frac{\sum y_i}{n} = \frac{\sum (ax_i + b)}{n} = \frac{a\sum x_i + b}{n} = a\bar{x} + b.$$

$$s_y^2 = \frac{\sum (y_i - \bar{y})^2}{n-1} = \frac{\sum (ax_i + b - (a\bar{x} + b))^2}{n-1} = \frac{\sum (ax_i - a\bar{x})^2}{n-1}$$

$$= \frac{a^2\sum (x_i - \bar{x})^2}{n-1} = a^2 s_x^2.$$

b.

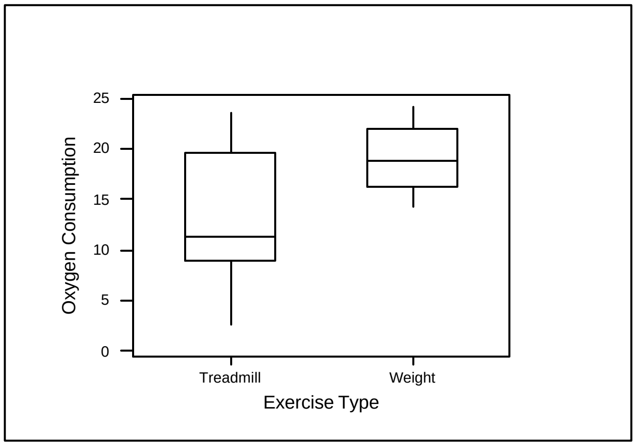
$$x = {}^{\circ}C, y = {}^{\circ}F$$

$$\bar{y} = \frac{9}{5}(87.3) + 32 = 189.14$$

$$s_y = \sqrt{s_y^2} = \sqrt{\left(\frac{9}{5}\right)^2 (1.04)^2} = \sqrt{3.5044} = 1.872$$

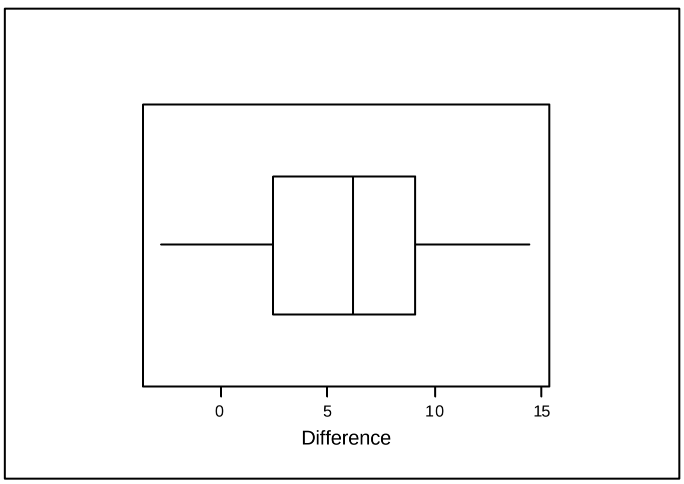
70.

a.



There is a significant difference in the variability of the two samples. The weight training produced much higher oxygen consumption, on average, than the treadmill exercise, with the median consumptions being approximately 20 and 11 liters, respectively.

b. Subtracting the y from the x for each subject, the differences are 3.3, 9.1, 10.4, 9.1, 6.2, 2.5, 2.2, 8.4, 8.7, 14.4, 2.5, -2.8, -0.4, 5.0, and 11.5.

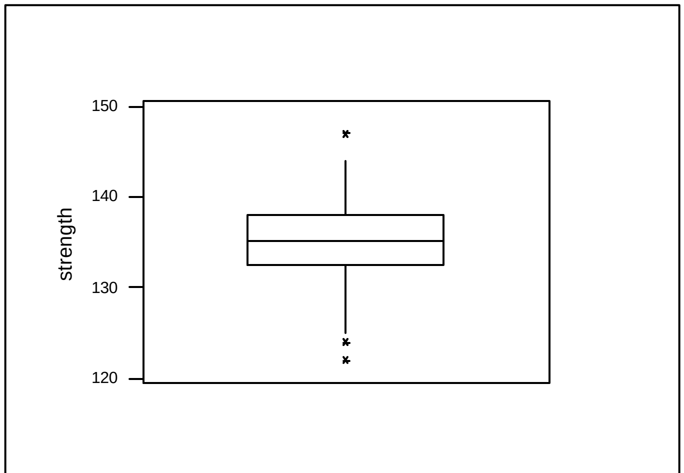


The majority of the differences are positive, which suggests that the weight training produced higher oxygen consumption for most subjects. The median difference is about 6 liters.

71.

a. The mean, median, and trimmed mean are virtually identical, which suggests symmetry. If there are outliers, they are balanced. The range of values is only 25.5, but half of the values are between 132.95 and 138.25.

b.

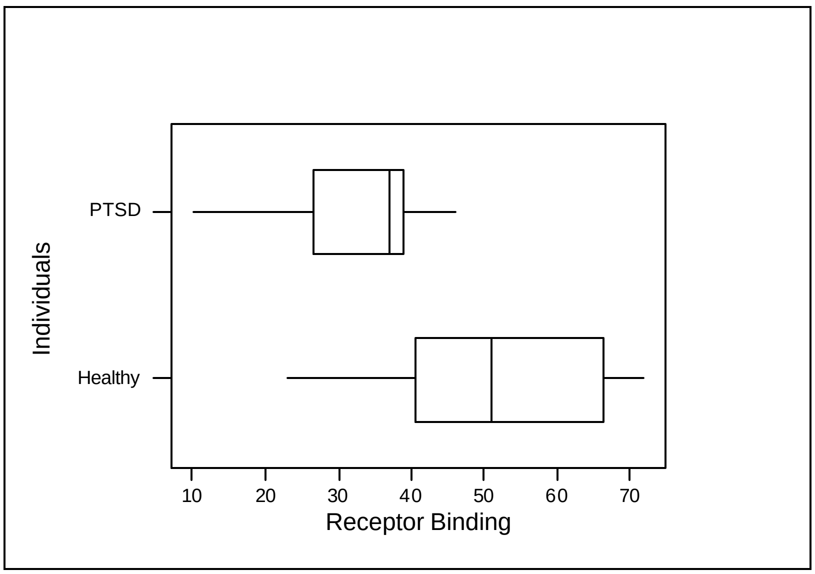


The boxplot also displays the symmetry, and adds a visual of the outliers, two on the lower end, and one on the upper.

72. A table of summary statistics, a stem and leaf display, and a comparative boxplot are below. The healthy individuals have higher receptor binding measure on average than the individuals with PTSD. There is also more variation in the healthy individuals' values. The distribution of values for the healthy is reasonably symmetric, while the distribution for the PTSD individuals is negatively skewed. The box plot indicates that there are no outliers, and confirms the above comments regarding symmetry and skewness.

	PTSD	Healthy
Mean	32.92	52.23
Median	37	51
Std Dev	9.93	14.86
Min	10	23
Max	46	72

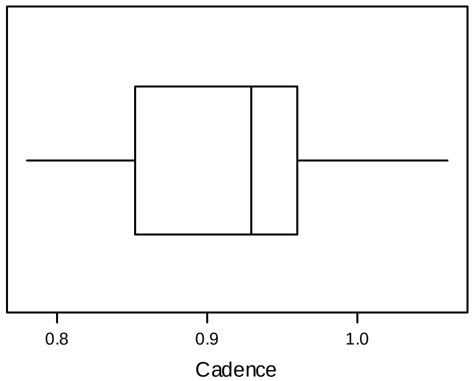
	1	0	stem = tens
3	2	058	leaf = ones
9	3	1578899	
7310	4	26	
81	5		
9763	6		
2	7		



73.

```
0.7 | 8              stem=tenths
0.8 | 11556          leaf=hundredths
0.9 | 2233335566
1.0 | 0566
```

$$\bar{x} = .9255, s = .0809, \tilde{x} = .93$$
$$lowerfourth = .855, upperfourth = .96$$



The data appears to be a bit skewed toward smaller values (negatively skewed). There are no outliers. The mean and the median are close in value.

74.

a. Mode = .93. It occurs four times in the data set.

b. The Modal Category is the one in which the most observations occur.

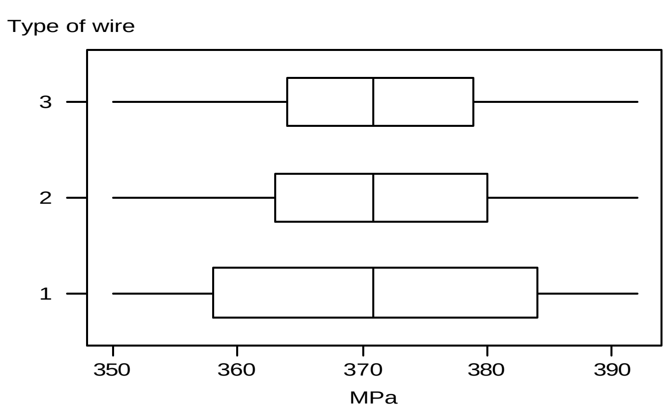
75.

a. The median is the same (371) in each plot and all three data sets are very symmetric. In addition, all three have the same minimum value (350) and same maximum value (392). Moreover, all three data sets have the same lower (364) and upper quartiles (378). So, all three boxplots will be *identical*.

b. A comparative dotplot is shown below. These graphs show that there are differences in the variability of the three data sets. They also show differences in the way the values are distributed in the three data sets.

```
  .            .
  :    .      :::        .     :.
-----+---------+---------+---------+---------+---------+- Type 1

  .   .   .  . .  .. . . . .   .  .   .   .
-----+---------+---------+---------+---------+---------+- Type 2
                         .
  .       . . :. .   .   : .:         .
-----+---------+---------+---------+---------+---------+- Type 3
   352.0    360.0    368.0    376.0    384.0    392.0
```

c. The boxplot in (a) is not capable of detecting the differences among the data sets. The primary reason is that boxplots give up some detail in describing data because they use only 5 summary numbers for comparing data sets. Note: The definition of lower and upper quartile used in this text is slightly different than the one used by some other authors (and software packages). Technically speaking, the median of the lower half of the data is not really the first quartile, although it is generally *very close*. Instead, the medians of the lower and upper halves of the data are often called the **lower** and **upper hinges.** Our boxplots use the lower and upper hinges to define the spread of the middle 50% of the data, but other authors sometimes use the *actual* quartiles for this purpose. The difference is usually very slight, usually unnoticeable, but not always. For example in the data sets of this exercise, a comparative boxplot based on the actual quartiles (as computed by Minitab) is shown below. The graph shows substantially the same type of information as those described in (a) except the graphs based on quartiles are able to detect the slight differences in variation between the three data sets.



76. The measures that are sensitive to outliers are: the mean and the midrange. The mean is sensitive because all values are used in computing it. The midrange is sensitive because it uses only the most extreme values in its computation.

The median, the trimmed mean, and the midhinge are not sensitive to outliers.

The median is the most resistant to outliers because it uses only the middle value (or values) in its computation.

The trimmed mean is somewhat resistant to outliers. The larger the trimming percentage, the more resistant the trimmed mean becomes.

The midhinge, which uses the quartiles, is reasonably resistant to outliers because both quartiles are resistant to outliers.

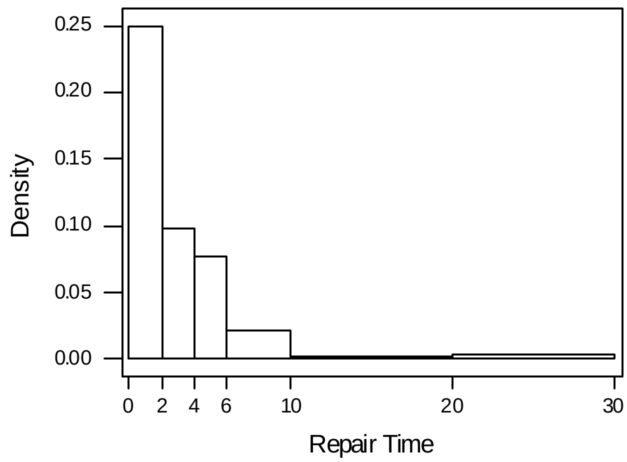
77.

a.

0	2355566777888	
1	0000135555	
2	00257	
3	0033	
4	0057	
5	044	
6		stem: ones
7	05	leaf: tenths
8	8	
9	0	
10	3	
HI	22.0 24.5	

b.

Interval	Frequency	Rel. Freq.	Density
0 -< 2	23	.500	.250
2 -< 4	9	.196	.098
4 -< 6	7	.152	.076
6 -< 10	4	.087	.022
10 -< 20	1	.022	.002
20 -< 30	2	.043	.004



78.

a. Since the constant $\bar{x}$ is subtracted from each x value to obtain each y value, and addition or subtraction of a constant doesn't affect variability, $s_y^2 = s_x^2$ and $s_y = s_x$

b. Let $c = 1/s$, where s is the sample standard deviation of the x's and also (by a) of the y's. Then $s_z = cs_y = (1/s)s = 1$, and $s_z^2 = 1$. That is, the "standardized" quantities $z_1, \ldots, z_n$ have a sample variance and standard deviation of 1.

79.

a. $\sum_{i=1}^{n+1} x_i = \sum_{i=1}^{n} x_i + x_{n+1} = n\bar{x}_n + x_{n+1}, so \bar{x}_{n+1} = \frac{[n\bar{x}_n + x_{n+1}]}{(n+1)}$

b.

$$ns_{n+1}^2 = \sum_{i=1}^{n+1} (x_i - \bar{x}_{n+1})^2 = \sum_{i=1}^{n+1} x_i^2 - (n+1)\bar{x}_{n+1}^2$$

$$= \sum_{i=1}^{n} x_i^2 - n\bar{x}_n^2 + x_{n+1}^2 + n\bar{x}_n^2 - (n+1)\bar{x}_{n+1}^2$$

$$= (n-1)s_n^2 + \left\{x_{n+1}^2 + n\bar{x}_n^2 - (n+1)\bar{x}_{n+1}^2\right\}$$

When the expression for $\bar{x}_{n+1}$ from **a** is substituted, the expression in braces simplifies to

the following, as desired: $\frac{n(x_{n+1} - \bar{x}_n)^2}{(n+1)}$

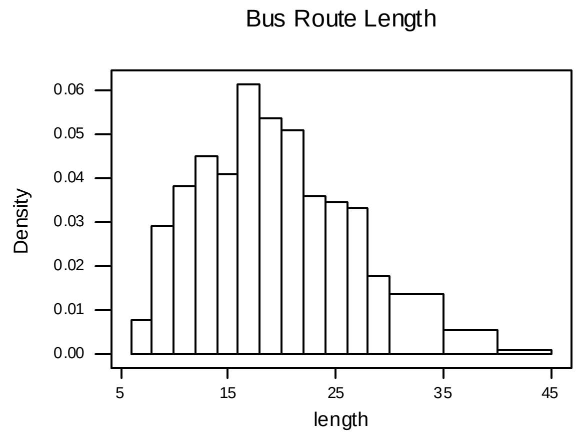
c. $\bar{x}_{n+1} = \frac{15(12.58) + 11.8}{16} = \frac{200.5}{16} = 12.53$

$$s_{n+1}^2 = \frac{n-1}{n}\left(s_n^2\right) + \frac{(x_{n+1} - \bar{x}_n)^2}{(n+1)} = \frac{14}{15}\left(.512^2\right) + \frac{(11.8 - 12.58)^2}{(16)}$$

$= .245 + .038 = .238$. So the standard deviation $s_{n+1} = \sqrt{.238} = .532$

80.

a.



b. Proportion less than $20 = \left(\frac{216}{391}\right) = .552$

Proportion at least $30 = \left(\frac{40}{391}\right) = .102$

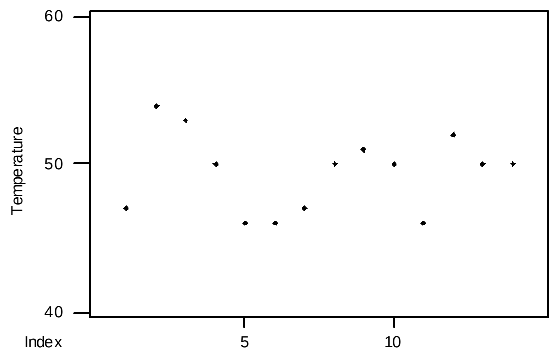
c. First compute (.90)(391 + 1) = 352.8. Thus, the 90th percentile should be about the 352nd ordered value. The 351st ordered value lies in the interval 28 - < 30. The 352nd ordered value lies in the interval 30 - < 35. There are 27 values in the interval 30 - < 35. We do not know how these values are distributed, however, the smallest value (i.e., the 352nd value in the data set) cannot be smaller than 30. So, the 90th percentile is roughly 30.

d. First compute (.50)(391 + 1) = 196. Thus the median (50th percentile) should be the 196 ordered value. The 174th ordered value lies in the interval 16 -< 18. The next 42 observation lie in the interval 18 - < 20. So, ordered observation 175 to 216 lie in the intervals 18 - < 20. The 196th observation is about in the middle of these. Thus, we would say, the median is roughly 19.

81. Assuming that the histogram is unimodal, then there is evidence of positive skewness in the data since the median lies to the left of the mean (for a symmetric distribution, the mean and median would coincide). For more evidence of skewness, compare the distances of the 5th and 95th percentiles from the median: median - 5th percentile = 500 - 400 = 100 while 95th percentile -median = 720 - 500 = 220. Thus, the largest 5% of the values (above the 95th percentile) are further from the median than are the lowest 5%. The same skewness is evident when comparing the 10th and 90th percentiles to the median: median - 10th percentile = 500 - 430 = 70 while 90th percentile -median = 640 - 500 = 140. Finally, note that the largest value (925) is much further from the median (925-500 = 425) than is the smallest value (500 - 220 = 280), again an indication of positive skewness.

82.

a. There is some evidence of a cyclical pattern.



b.
$$\bar{x}_2 = .1x_2 + .9\bar{x}_1 = (.1)(54) + (.9)(47) = 47.7$$
$$\bar{x}_3 = .1x_3 + .9\bar{x}_2 = (.1)(53) + (.9)(47.7) = 48.23 \approx 48.2, etc.$$

t	$\bar{x}_t$ for.$a = .1$	$\bar{x}_t$ for.$a = .5$
1	47.0	47.0
2	47.7	50.5
3	48.2	51.8
4	48.4	50.9
5	48.2	48.4
6	48.0	47.2
7	47.9	47.1
8	48.1	48.6
9	48.4	49.8
10	48.5	49.9
11	48.3	47.9
12	48.6	50.0
13	48.8	50.0
14	48.9	50.0

α= .1 gives a smoother series.

c.
$$\bar{x}_t = ax_t + (1-a)\bar{x}_{t-1}$$
$$= ax_t + (1-a)[ax_{t-1} + (1-a)\bar{x}_{t-2}]$$
$$= ax_t + a(1-a)x_{t-1} + (1-a)^2[ax_{t-2} + (1-a)\bar{x}_{t-3}]$$
$$= ... = ax_t + a(1-a)x_{t-1} + a(1-a)^2 x_{t-2} + ... + a(1-a)^{t-2} x_2 + (1-a)^{t-1}\bar{x}_1$$

Thus, (x bar)$_t$ depends on x_t and all previous values. As k increases, the coefficient on x_{t-k} decreases (further back in time implies less weight).

d. Not very sensitive, since $(1-\alpha)^{t-1}$ will be very small.

83.

a. When there is perfect symmetry, the smallest observation y_1 and the largest observation y_n will be equidistant from the median, so $y_n - \overline{x} = \overline{x} - y_1$.

Similarly, the second smallest and second largest will be equidistant from the median, so $y_{n-1} - \overline{x} = \overline{x} - y_2$

and so on. Thus, the first and second numbers in each pair will be equal, so that each point in the plot will fall exactly on the 45 degree line. When the data is positively skewed, y_n will be much further from the median than is y_1, so $y_n - \tilde{x}$ will considerably exceed $\tilde{x} - y_1$ and the point $(y_n - \tilde{x}, \tilde{x} - y_1)$ will fall considerably below the 45 degree line. A similar comment aplies to other points in the plot.

b. The first point in the plot is (2745.6 – 221.6, 221.6 0- 4.1) = (2524.0, 217.5). The others are: (1476.2, 213.9), (1434.4, 204.1), (756.4, 190.2), (481.8, 188.9), (267.5, 181.0), (208.4, 129.2), (112.5, 106.3), (81.2, 103.3), (53.1, 102.6), (53.1, 92.0), (33.4, 23.0), and (20.9, 20.9). The first number in each of the first seven pairs greatly exceed the second number, so each point falls well below the 45 degree line. A substantial positive skew (stretched upper tail) is indicated.

CHAPTER 2

Section 2.1

1.

a. S = { 1324, 1342, 1423, 1432, 2314, 2341, 2413, 2431, 3124, 3142, 4123, 4132, 3214, 3241, 4213, 4231 }

b. Event A contains the outcomes where 1 is first in the list:
A = { 1324, 1342, 1423, 1432 }

c. Event B contains the outcomes where 2 is first or second:
B = { 2314, 2341, 2413, 2431, 3214, 3241, 4213, 4231 }

d. The compound event $A \cup B$ contains the outcomes in A or B or both:
$A \cup B$ = {1324, 1342, 1423, 1432, 2314, 2341, 2413, 2431, 3214, 3241, 4213, 4231 }

2.

a. Event A = { RRR, LLL, SSS }

b. Event B = { RLS, RSL, LRS, LSR, SRL, SLR }

c. Event C = { RRL, RRS, RLR, RSR, LRR, SRR }

d. Event D = { RRL, RRS, RLR, RSR, LRR, SRR, LLR, LLS, LRL, LSL, RLL, SLL, SSR, SSL, SRS, SLS, RSS, LSS }

e. Event D′ contains outcomes where all cars go the same direction, or they all go different directions:
D′ = { RRR, LLL, SSS, RLS, RSL, LRS, LSR, SRL, SLR }

Because Event D totally encloses Event C, the compound event $C \cup D = D$:
$C \cup D$ = { RRL, RRS, RLR, RSR, LRR, SRR, LLR, LLS, LRL, LSL, RLL, SLL, SSR, SSL, SRS, SLS, RSS, LSS }

Using similar reasoning, we see that the compound event $C \cap D = C$:
$C \cap D$ = { RRL, RRS, RLR, RSR, LRR, SRR }

3.

a. Event A = { SSF, SFS, FSS }

b. Event B = { SSS, SSF, SFS, FSS }

c. For Event C, the system must have component 1 working (S in the first position), then at least one of the other two components must work (at least one S in the 2nd and 3rd positions: Event C = { SSS, SSF, SFS }

d. Event C′ = { SFF, FSS, FSF, FFS, FFF }
Event A∪C = { SSS, SSF, SFS, FSS }
Event A∩C = { SSF, SFS }
Event B∪C = { SSS, SSF, SFS, FSS }
Event B∩C = { SSS SSF, SFS }

4.

a.

	Home Mortgage Number			
Outcome	1	2	3	4
1	F	F	F	F
2	F	F	F	V
3	F	F	V	F
4	F	F	V	V
5	F	V	F	F
6	F	V	F	V
7	F	V	V	F
8	F	V	V	V
9	V	F	F	F
10	V	F	F	V
11	V	F	V	F
12	V	F	V	V
13	V	V	F	F
14	V	V	F	V
15	V	V	V	F
16	V	V	V	V

b. Outcome numbers 2, 3, 5 ,9

c. Outcome numbers 1, 16

d. Outcome numbers 1, 2, 3, 5, 9

e. In words, the UNION described is the event that either all of the mortgages are variable, or that at most all of them are variable: outcomes 1,2,3,5,9,16. The INTERSECTION described is the event that all of the mortgages are fixed: outcome 1.

f. The UNION described is the event that either exactly three are fixed, or that all four are the same: outcomes 1, 2, 3, 5, 9, 16. The INTERSECTION in words is the event that exactly three are fixed AND that all four are the same. This cannot happen. (There are no outcomes in common) : $\mathbf{b} \cap \mathbf{c} = \varnothing$.

5.

a.

Outcome Number	Outcome
1	111
2	112
3	113
4	121
5	122
6	123
7	131
8	132
9	133
10	211
11	212
12	213
13	221
14	222
15	223
16	231
17	232
18	233
19	311
20	312
21	313
22	321
23	322
24	323
25	331
26	332
27	333

b. Outcome Numbers 1, 14, 27

c. Outcome Numbers 6, 8, 12, 16, 20, 22

d. Outcome Numbers 1, 3, 7, 9, 19, 21, 25, 27

6.

a.

Outcome Number	Outcome
1	123
2	124
3	125
4	213
5	214
6	215
7	13
8	14
9	15
10	23
11	24
12	25
13	3
14	4
15	5

b. Outcomes 13, 14, 15

c. Outcomes 3, 6, 9, 12, 15

d. Outcomes 10, 11, 12, 13, 14, 15

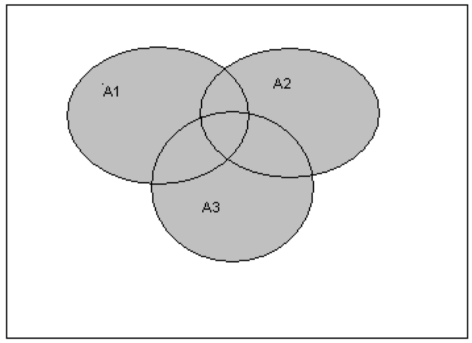
7.

a. S = {BBBAAAA, BBABAAA, BBAABAA, BBAAABA, BBAAAAB, BABBAAA, BABABAA, BABAABA, BABAAAB, BAABBAA, BAABABA, BAABAAB, BAAABBA, BAAABAB, BAAAABB, ABBBAAA, ABBABAA, ABBAABA, ABBAAAB, ABABBAA, ABABABA, ABABAAB, ABAABBA, ABAABAB, ABAAABB, AABBBAA, AABBABA, AABBAAB, AABABBA, AABABAB, AABAABB, AAABBBA, AAABBAB, AAABABB, AAAABBB}

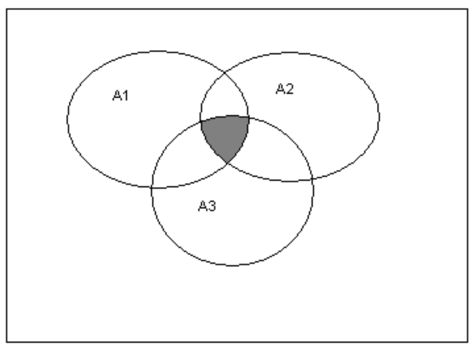
b. {AAAABBB, AAABABB, AAABBAB, AABAABB, AABABAB}

8.

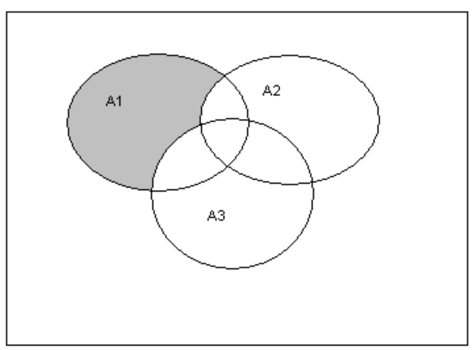
a. $A_1 \cup A_2 \cup A_3$



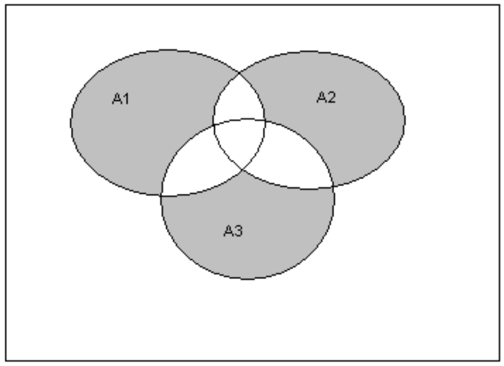
b. $A_1 \cap A_2 \cap A_3$



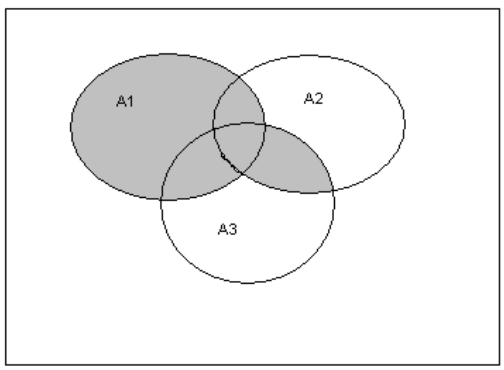
c. $A_1 \cap A_2' \cap A_3'$



d. $(A_1 \cap A_2' \cap A_3') \cup (A_1' \cap A_2 \cap A_3') \cup (A_1' \cap A_2' \cap A_3)$

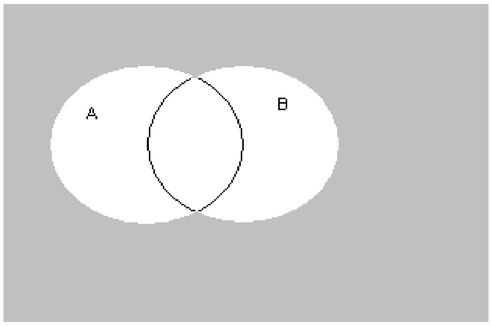


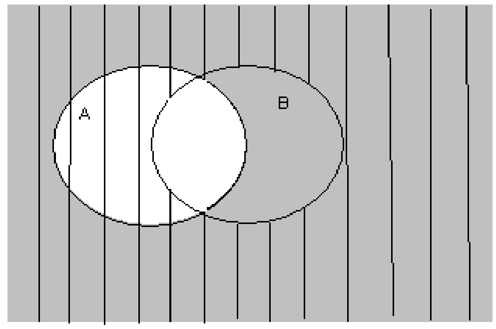
e. $A_1 \cup (A_2 \cap A_3)$



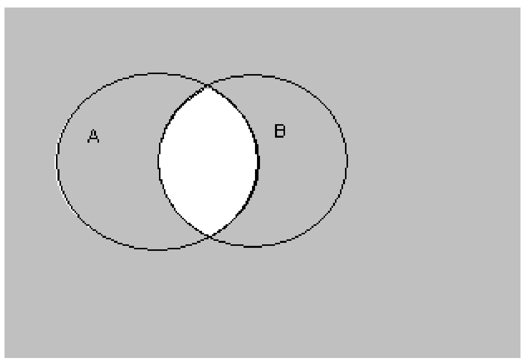
9.

a. In the diagram on the left, the shaded area is $(A \cup B)'$. On the right, the shaded area is A', the striped area is B', and the intersection $A' \cap B'$ occurs where there is BOTH shading and stripes. These two diagrams display the same area.





b. In the diagram below, the shaded area represents $(A \cap B)'$. Using the diagram on the right above, the union of A' and B' is represented by the areas that have either shading or stripes or both. Both of the diagrams display the same area.



10.

a. A = {Chev, Pont, Buick}, B = {Ford, Merc}, C = {Plym, Chrys} are three mutually exclusive events.

b. No, let E = {Chev, Pont}, F = {Pont, Buick}, G = {Buick, Ford}. These events are not mutually exclusive (e.g. E and F have an outcome in common), yet there is no outcome common to all three events.

Section 2.2

11.

a. .07

b. .15 + .10 + .05 = .30

c. Let event A = selected customer owns stocks. Then the probability that a selected customer does not own a stock can be represented by
$P(A') = 1 - P(A) = 1 – (.18 + .25) = 1 - .43 = .57$. This could also have been done easily by adding the probabilities of the funds that are not stocks.

12.

a. $P(A \cup B) = .50 + .40 - .25 = .65$

b. $P(A \cup B)' = 1 - .65 = .35$

c. $A \cap B'$; $P(A \cap B') = P(A) – P(A \cap B) = .50 - .25 = .25$

13.

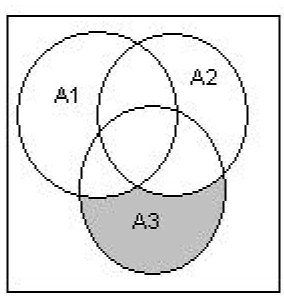
a. awarded either #1 or #2 (or both):
$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = .22 + .25 - .11 = .36$

b. awarded neither #1 or #2:
$P(A_1' \cap A_2') = P[(A_1 \cup A_2)'] = 1 - P(A_1 \cup A_2) = 1 - .36 = .64$

c. awarded at least one of #1, #2, #3:
$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) –$
$P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$
$= .22 +.25 + .28 - .11 -.05 - .07 + .01 = .53$

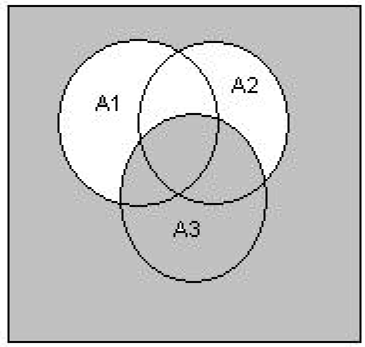
d. awarded none of the three projects:
$P(A_1' \cap A_2' \cap A_3') = 1 – P(\text{awarded at least one}) = 1 - .53 = .47.$

e. awarded #3 but neither #1 nor #2:
$P(A_1' \cap A_2' \cap A_3) = P(A_3) - P(A_1 \cap A_3) – P(A_2 \cap A_3)$
$+ P(A_1 \cap A_2 \cap A_3)$
$= .28 - .05 - .07+ .01 = .17$

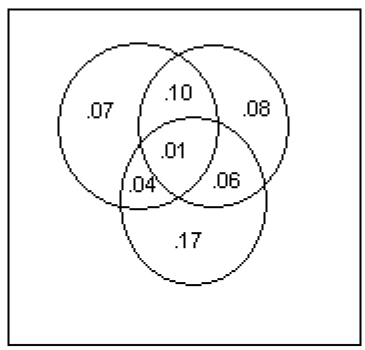


f. either (neither #1 nor #2) or #3:

$$P[(A_1' \cap A_2') \cup A_3] = P(\text{shaded region}) = P(\text{awarded none}) + P(A_3)$$
$$= .47 + .28 = .75$$



Alternatively, answers to **a – f** can be obtained from probabilities on the accompanying Venn diagram



14.

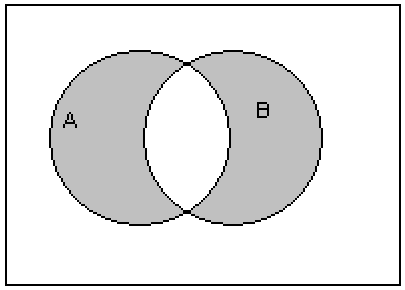
a. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$,

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

$= .8 + .7 - .9 = .6$

b. P(shaded region) = $P(A \cup B) - P(A \cap B) = .9 - .6 = .3$

Shaded region = event of interest = $(A \cap B') \cup (A' \cap B)$



15.

a. Let event E be the event that at most one purchases an electric dryer. Then E′ is the event that at least two purchase electric dryers.

$P(E') = 1 - P(E) = 1 - .428 = .572$

b. Let event A be the event that all five purchase gas. Let event B be the event that all five purchase electric. All other possible outcomes are those in which at least one of each type is purchased. Thus, the desired probability =

$1 - P(A) - P(B) = 1 - .116 - .005 = .879$

16.

a. There are six simple events, corresponding to the outcomes CDP, CPD, DCP, DPC, PCD, and PDC. The probability assigned to each is $\frac{1}{6}$.

b. P(C ranked first) = P({CPD, CDP}) = $\frac{1}{6} + \frac{1}{6} = \frac{2}{6} = .333$

c. P(C ranked first and D last) = P({CPD}) = $\frac{1}{6}$

17.

a. The probabilities do not add to 1 because there are other software packages besides SPSS and SAS for which requests could be made.

b. $P(A') = 1 - P(A) = 1 - .30 = .70$

c. $P(A \cup B) = P(A) + P(B) = .30 + .50 = .80$
(since A and B are mutually exclusive events)

d. $P(A' \cap B') = P[(A \cup B)']$ (De Morgan's law)
$= 1 - P(A \cup B)$
$= 1 - .80 = .20$

18. This situation requires the complement concept. The only way for the desired event NOT to happen is if a 75 W bulb is selected first. Let event A be that a 75 W bulb is selected first, and $P(A) = \frac{6}{15}$. Then the desired event is event A′.
So $P(A') = 1 - P(A) = 1 - \frac{6}{15} = \frac{9}{15} = .60$

19. Let event A be that the selected joint was found defective by inspector A. $P(A) = \frac{724}{10{,}000}$. Let event B be analogous for inspector B. $P(B) = \frac{751}{10{,}000}$. Compound event $A \cup B$ is the event that the selected joint was found defective by at least one of the two inspectors. $P(A \cup B) = \frac{1159}{10{,}000}$.

a. The desired event is $(A \cup B)'$, so we use the complement rule:
$P(A \cup B)' = 1 - P(A \cup B) = 1 - \frac{1159}{10{,}000} = \frac{8841}{10{,}000} = .8841$

b. The desired event is $B \cap A'$. $P(B \cap A') = P(B) - P(A \cap B)$.
$P(A \cap B) = P(A) + P(B) - P(A \cup B)$,
$= .0724 + .0751 - .1159 = .0316$
So $P(B \cap A') = P(B) - P(A \cap B)$
$= .0751 - .0316 = .0435$

20. Let S1, S2 and S3 represent the swing and night shifts, respectively. Let C1 and C2 represent the unsafe conditions and unrelated to conditions, respectively.

a. The simple events are {S1,C1}, {S1,C2}, {S2,C1}, {S2,C2},{S3,C1}, {S3,C2}.

b. P({C1})= P({S1,C1},{S2,C1},{S3,C1})= .10 + .08 + .05 = .23

c. P({S1}′) = 1 - P({S1,C1}, {S1,C2}) = 1 – (.10 + .35) = .55

21.

a. P({M,H}) = .10

b. P(low auto) = P[{(L,N}, (L,L), (L,M), (L,H)}] = .04 + .06 + .05 + .03 = .18 Following a similar pattern, P(low homeowner's) = .06 + .10 + .03 = .19

c. P(same deductible for both) = P[{ LL, MM, HH }] = .06 + .20 + .15 = .41

d. P(deductibles are different) = 1 – P(same deductibles) = 1 - .41 = .59

e. P(at least one low deductible) = P[{ LN, LL, LM, LH, ML, HL }]
= .04 + .06 + .05 + .03 + .10 + .03 = .31

f. P(neither low) = 1 – P(at least one low) = 1 - .31 = .69

22.

a. $P(A_1 \cap A_2) = P(A_1) + P(A_2) - P(A_1 \cup A_2) = .4 + .5 - .6 = .3$

b. $P(A_1 \cap A_2') = P(A_1) - P(A_1 \cap A_2) = .4 - .3 = .1$

c. P(exactly one) = $P(A_1 \cup A_2) - P(A_1 \cap A_2) = .6 - .3 = .3$

23. Assume that the computers are numbered 1 – 6 as described. Also assume that computers 1 and 2 are the laptops. Possible outcomes are (1,2) (1,3) (1,4) (1,5) (1,6) (2,3) (2,4) (2,5) (2,6) (3,4) (3,5) (3,6) (4,5) (4,6) and (5,6).

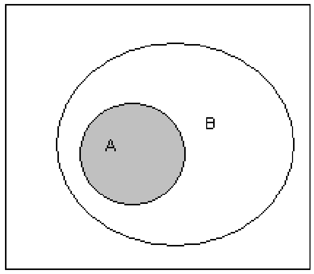
a. P(both are laptops) = P[{ (1,2)}] = $\frac{1}{15}$ =.067

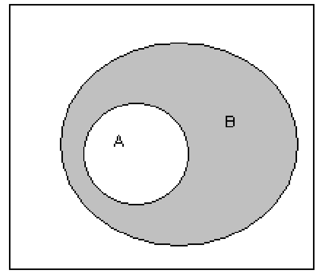
b. P(both are desktops) = P[{(3,4) (3,5) (3,6) (4,5) (4,6) (5,6)}] = $\frac{6}{15}$ = .40

c. P(at least one desktop) = 1 – P(no desktops)
= 1 – P(both are laptops)
= 1 – .067 = .933

d. P(at least one of each type) = 1 – P(both are the same)
= 1 – P(both laptops) – P(both desktops)
= 1 - .067 - .40 = .533

24. Since A is contained in B, then B can be written as the union of A and $(B \cap A')$, two mutually exclusive events. (See diagram).



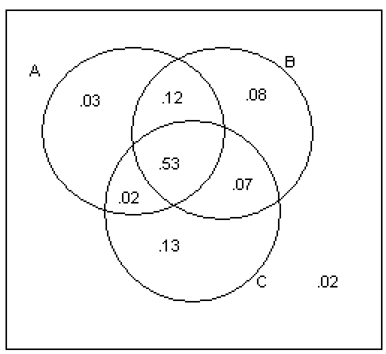


From Axiom 3, $P[A \cup (B \cap A')] = P(A) + P(B \cap A')$. Substituting P(B), $P(B) = P(A) + P(B \cap A')$ or $P(B) - P(A) = P(B \cap A')$. From Axiom 1, $P(B \cap A') \geq 0$, so $P(B) \geq P(A)$ or $P(A) \leq P(B)$. For general events A and B, $P(A \cap B) \leq P(A)$, and $P(A \cup B) \geq P(A)$.

25. $P(A \cap B) = P(A) + P(B) - P(A \cup B) = .65$

$P(A \cap C) = .55$, $P(B \cap C) = .60$

$$P(A \cap B \cap C) = P(A \cup B \cup C) - P(A) - P(B) - P(C) + P(A \cap B) + P(A \cap C) + P(B \cap C)$$
$$= .98 - .7 - .8 - .75 + .65 + .55 + .60$$
$$= .53$$



a. $P(A \cup B \cup C) = .98$, as given.

b. P(none selected) = $1 - P(A \cup B \cup C) = 1 - .98 = .02$

c. P(only automatic transmission selected) = .03 from the Venn Diagram

d. P(exactly one of the three) = .03 + .08 + .13 = .24

26.

a. $P(A_1') = 1 – P(A_1) = 1 - .12 = .88$

b. $P(A_1 \cap A_2) = P(A_1) + P(A_2) - P(A_1 \cup A_2) = .12 + .07 - .13 = .06$

c. $P(A_1 \cap A_2 \cap A_3') = P(A_1 \cap A_2) - P(A_1 \cap A_2 \cap A_3) = .06 - .01 = .05$

d. P(at most two errors) = 1 – P(all three types)
$= 1 - P(A_1 \cap A_2 \cap A_3)$
$= 1 - .01 = .99$

27. Outcomes: $(A,B)\ (A,C_1)\ (A,C_2)\ (A,F)\ (B,A)\ (B,C_1)\ (B,C_2)\ (B,F)$
$(C_1,A)\ (C_1,B)\ (C_1,C_2)\ (C_1,F)\ (C_2,A)\ (C_2,B)\ (C_2,C_1)\ (C_2,F)$
$(F,A)\ (F,B)\ (F,C_1)\ (F,C_2)$

a. P[(A,B) or (B,A)] = $\frac{2}{20} = \frac{1}{10} = .1$

b. P(at least one C) = $\frac{14}{20} = \frac{7}{10} = .7$

c. P(at least 15 years) = 1 – P(at most 14 years)
= 1 – P[(3,6) or (6,3) or (3,7) or (7,3) or (3,10) or (10,3) or (6,7) or (7,6)]
= $1 - \frac{8}{20} = 1 - .4 = .6$

28. There are 27 equally likely outcomes.

a. P(all the same) = P[(1,1,1) or (2,2,2) or (3,3,3)] = $\frac{3}{27} = \frac{1}{9}$

b. P(at most 2 are assigned to the same station) = 1 – P(all 3 are the same)
= $1 - \frac{3}{27} = \frac{24}{27} = \frac{8}{9}$

c. P(all different) = [{(1,2,3) (1,3,2) (2,1,3) (2,3,1) (3,1,2) (3,2,1)}]
= $\frac{6}{27} = \frac{2}{9}$

Section 2.3

29.

a. (5)(4) = 20 (5 choices for president, 4 remain for vice president)

b. (5)(4)(3) = 60

c. $\binom{5}{2} = \frac{5!}{2!3!} = 10$ (No ordering is implied in the choice)

30.

a. Because order is important, we'll use $P_{8,3} = 8(7)(6) = 336$.

b. Order doesn't matter here, so we use $C_{30,6} = 593{,}775$.

c. From each group we choose 2: $\binom{8}{2} \bullet \binom{10}{2} \bullet \binom{12}{2} = 83{,}160$

d. The numerator comes from part c and the denominator from part b: $\frac{83{,}160}{593{,}775} = .14$

e. We use the same denominator as in part d. We can have all zinfandel, all merlot, or all cabernet, so P(all same) = P(all z) + P(all m) + P(all c) =

$$\frac{\binom{8}{6} + \binom{10}{6} + \binom{12}{6}}{\binom{30}{6}} = \frac{1162}{593{,}775} = .002$$

31.

a. $(n_1)(n_2) = (9)(27) = 243$

b. $(n_1)(n_2)(n_3) = (9)(27)(15) = 3645$, so such a policy could be carried out for 3645 successive nights, or approximately 10 years, without repeating exactly the same program.

32.

a. $5\times4\times3\times4 = 240$

b. $1\times1\times3\times4 = 12$

c. $4\times3\times3\times3 = 108$

d. # with at least on Sony = total # - # with no Sony = 240 – 108 = 132

e. P(at least one Sony) = $\frac{132}{240} = .55$

P(exactly one Sony) = P(only Sony is receiver)
+ P(only Sony is CD player)
+ P(only Sony is deck)

$$= \frac{1\times3\times3\times3}{240} + \frac{4\times1\times3\times3}{240} + \frac{4\times3\times3\times1}{240} = \frac{27+36+36}{240}$$

$$= \frac{99}{240} = .413$$

33.

a. $\binom{25}{5} = \frac{25!}{5!20!} = 53{,}130$

b. $\binom{8}{4}\bullet\binom{17}{1} = 1190$

c. P(exactly 4 have cracks) = $\dfrac{\binom{8}{4}\binom{17}{1}}{\binom{25}{5}} = \dfrac{1190}{53{,}130} = .022$

d. P(at least 4) = P(exactly 4) + P(exactly 5)

$$= \frac{\binom{8}{4}\binom{17}{1}}{\binom{25}{5}} + \frac{\binom{8}{5}\binom{17}{0}}{\binom{25}{5}} = .022 + .001 = .023$$

34.

a. $\binom{20}{6} = 38{,}760$. P(all from day shift) = $\dfrac{\binom{20}{6}\binom{25}{0}}{\binom{45}{6}} = \dfrac{38{,}760}{8{,}145{,}060} = .0048$

b. P(all from same shift) = $\dfrac{\binom{20}{6}\binom{25}{0}}{\binom{45}{6}} + \dfrac{\binom{15}{6}\binom{30}{0}}{\binom{45}{6}} + \dfrac{\binom{10}{6}\binom{35}{0}}{\binom{45}{6}}$

= .0048 + .0006 + .0000 = .0054

c. P(at least two shifts represented) = 1 – P(all from same shift)
= 1 - .0054 = .9946

d. Let A_1 = day shift unrepresented, A_2 = swing shift unrepresented, and A_3 = graveyard shift unrepresented. Then we wish $P(A_1 \cup A_2 \cup A_3)$.
$P(A_1)$ = P(day unrepresented) = P(all from swing and graveyard)

$P(A_1) = \dfrac{\binom{25}{6}}{\binom{45}{6}}$, $P(A_2) = \dfrac{\binom{30}{6}}{\binom{45}{6}}$, $P(A_3) = \dfrac{\binom{35}{6}}{\binom{45}{6}}$,

$P(A_1 \cap A_2)$ = P(all from graveyard) = $\dfrac{\binom{10}{6}}{\binom{45}{6}}$

$P(A_1 \cap A_3) = \dfrac{\binom{15}{6}}{\binom{45}{6}}$, $P(A_2 \cap A_3) = \dfrac{\binom{20}{6}}{\binom{45}{6}}$, $P(A_1 \cap A_2 \cap A_3) = 0$,

So $P(A_1 \cup A_2 \cup A_3) = \dfrac{\binom{25}{6}}{\binom{45}{6}} + \dfrac{\binom{30}{6}}{\binom{45}{6}} + \dfrac{\binom{35}{6}}{\binom{45}{6}} - \dfrac{\binom{10}{6}}{\binom{45}{6}} - \dfrac{\binom{15}{6}}{\binom{45}{6}} - \dfrac{\binom{20}{6}}{\binom{45}{6}}$

= .2939 - .0054 = .2885

35. There are 10 possible outcomes -- $\binom{5}{2}$ ways to select the positions for B's votes: BBAAA, BABAA, BAABA, BAAAB, ABBAA, ABABA, ABAAB, AABBA, AABAB, and AAABB. Only the last two have A ahead of B throughout the vote count. Since the outcomes are equally likely, the desired probability is $\frac{2}{10} = .20$.

36.

a. $n_1 = 3$, $n_2 = 4$, $n_3 = 5$, so $n_1 \times n_2 \times n_3 = 60$ runs

b. $n_1 = 1$, (just one temperature), $n_2 = 2$, $n_3 = 5$ implies that there are 10 such runs.

37. There are $\binom{60}{5}$ ways to select the 5 runs. Each catalyst is used in 12 different runs, so the number of ways of selecting one run from each of these 5 groups is 12^5. Thus the desired probability is $\dfrac{12^5}{\binom{60}{5}} = .0456$.

38.

a. P(selecting 2 - 75 watt bulbs) = $\dfrac{\binom{6}{2}\binom{9}{1}}{\binom{15}{3}} = \dfrac{15 \cdot 9}{455} = .2967$

b. P(all three are the same) = $\dfrac{\binom{4}{3}+\binom{5}{3}+\binom{6}{3}}{\binom{15}{3}} = \dfrac{4+10+20}{455} = .0747$

c. $\binom{4}{1}\binom{5}{1}\binom{6}{1} = \dfrac{120}{455} = .2637$

d. To examine exactly one, a 75 watt bulb must be chosen first. (6 ways to accomplish this). To examine exactly two, we must choose another wattage first, then a 75 watt. (9×6 ways). Following the pattern, for exactly three, $9 \times 8 \times 6$ ways; for four, $9 \times 8 \times 7 \times 6$; for five, $9 \times 8 \times 7 \times 6 \times 6$.

P(examine at least 6 bulbs) = 1 – P(examine 5 or less)
= 1 – P(examine exactly 1 or 2 or 3 or 4 or 5)
= 1 – [P(one) + P(two) + … + P(five)]

$$= 1 - \left[\frac{6}{15} + \frac{9\times6}{15\times14} + \frac{9\times8\times6}{15\times14\times13} + \frac{9\times8\times7\times6}{15\times14\times13\times12} + \frac{9\times8\times7\times6\times6}{15\times14\times13\times12\times11}\right]$$

= 1 – [.4 + .2571 + .1582 + .0923 + .0503]
= 1 - .9579 = .0421

39.

a. We want to choose all of the 5 cordless, and 5 of the 10 others, to be among the first 10 serviced, so the desired probability is $\dfrac{\binom{5}{5}\binom{10}{5}}{\binom{15}{10}} = \dfrac{252}{3003} = .0839$

b. Isolating one group, say the cordless phones, we want the other two groups represented in the last 5 serviced. So we choose 5 of the 10 others, except that we don't want to include the outcomes where the last five are all the same.

So we have $\dfrac{\binom{10}{5} - 2}{\binom{15}{5}}$. But we have three groups of phones, so the desired probability is

$$\frac{3 \cdot \left[\binom{10}{5} - 2\right]}{\binom{15}{5}} = \frac{3(250)}{3003} = .2498.$$

c. We want to choose 2 of the 5 cordless, 2 of the 5 cellular, and 2 of the corded phones:

$$\frac{\binom{5}{2}\binom{5}{2}\binom{5}{2}}{\binom{15}{6}} = \frac{1000}{5005} = .1998$$

40.

a. If the A's are distinguishable from one another, and similarly for the B's, C's and D's, then there are 12! Possible chain molecules. Six of these are:

$A_1A_2A_3B_2C_3C_1D_3C_2D_1D_2B_3B_1$, $A_1A_3A_2B_2C_3C_1D_3C_2D_1D_2B_3B_1$
$A_2A_1A_3B_2C_3C_1D_3C_2D_1D_2B_3B_1$, $A_2A_3A_1B_2C_3C_1D_3C_2D_1D_2B_3B_1$
$A_3A_1A_2B_2C_3C_1D_3C_2D_1D_2B_3B_1$, $A_3A_2A_1B_2C_3C_1D_3C_2D_1D_2B_3B_1$

These 6 (=3!) differ only with respect to ordering of the 3 A's. In general, groups of 6 chain molecules can be created such that within each group only the ordering of the A's is different. When the A subscripts are suppressed, each group of 6 "collapses" into a single molecule (B's, C's and D's are still distinguishable). At this point there are $\frac{12!}{3!}$ molecules. Now suppressing subscripts on the B's, C's and D's in turn gives ultimately $\frac{12!}{(3!)^4} = 369{,}600$ chain molecules.

b. Think of the group of 3 A's as a single entity, and similarly for the B's, C's, and D's. Then there are 4! Ways to order these entities, and thus 4! Molecules in which the A's are contiguous, the B's, C's, and D's are also. Thus, P(all together) = $\frac{4!}{369.600} = .00006494$.

41.

a. P(at least one F among 1^{st} 3) = 1 – P(no F's among 1^{st} 3)

$$= 1 - \frac{4\times3\times2}{8\times7\times6} = 1 - \frac{24}{336} = 1 - .0714 = .9286$$

An alternative method to calculate P(no F's among 1^{st} 3) would be to choose none of the females and 3 of the 4 males, as follows:

$\dfrac{\binom{4}{0}\binom{4}{3}}{\binom{8}{3}} = \dfrac{4}{56} = .0714$, obviously producing the same result.

b. P(all F's among 1^{st} 5) = $\dfrac{\binom{4}{4}\binom{4}{1}}{\binom{8}{5}} = \dfrac{4}{56} = .0714$

c. P(orderings are different) = 1 – P(orderings are the same for both semesters)
= 1 – (# orderings such that the orders are the same each semester)/(total # of possible orderings for 2 semesters)

$$= 1 - \frac{8\times7\times6\times5\times4\times3\times2\times1}{(8\times7\times6\times5\times4\times3\times2\times1)\times(8\times7\times6\times5\times4\times3\times2\times1)} = .99997520$$

42. Seats:

1	2	3	4	5	6

P(J&P in 1&2) $= \dfrac{2\times1\times4\times3\times2\times1}{6\times5\times4\times3\times2\times1} = \dfrac{1}{15} = .0667$

P(J&P next to each other) = P(J&P in 1&2) + … + P(J&P in 5&6)

$= 5\times\dfrac{1}{15} = \dfrac{1}{3} = .333$

P(at least one H next to his W) = 1 – P(no H next to his W)
We count the # of ways of no H next to his W as follows:
if orderings without a H-W pair in seats #1 and 3 and no H next to his W = $6^* \times 4 \times 1^* \times 2^{\#} \times 1 \times 1 = 48$
*= pair, # =can't put the mate of seat #2 here or else a H-W pair would be in #5 and 6.

# of orderings without a H-W pair in seats #1 and 3, and no H next to his W = $6 \times 4 \times 2^{\#} \times 2 \times 2 \times 1 = 192$
#= can't be mate of person in seat #1 or #2.
So, # of seating arrangements with no H next to W = 48 + 192 = 240

And P(no H next to his W) = $= \dfrac{240}{6\times5\times4\times3\times2\times1} = \dfrac{1}{3}$, so

P(at least one H next to his W) = 1 - $\dfrac{1}{3} = \dfrac{2}{3}$

43. # of 10 high straights = 4×4×4×4×4 (4 – 10's, 4 – 9's , etc)

P(10 high straight) = $\dfrac{4^5}{\binom{52}{5}} = \dfrac{1024}{2{,}598{,}960} = .000394$

P(straight) = $10\times\dfrac{4^5}{\binom{52}{5}} = .003940$ (Multiply by 10 because there are 10 different card values that could be high: Ace, King, etc.) There are only 40 straight flushes (10 in each suit), so

P(straight flush) = $\dfrac{40}{\binom{52}{5}} = .00001539$

44. $\binom{n}{k} = \dfrac{n!}{k!(n-k)!} = \dfrac{n!}{(n-k)!k!} = \binom{n}{n-k}$

The number of subsets of size k = the number of subsets of size n-k, because to each subset of size k there corresponds exactly one subset of size n-k (the n-k objects not in the subset of size k).

Section 2.4

45.

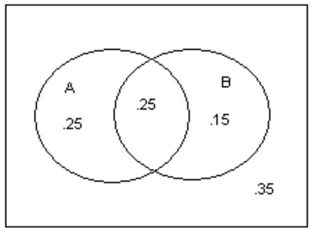
a. P(A) = .106 + .141 + .200 = .447, P(C) =.215 + .200 + .065 + .020 = .500 P(A ∩ C) = .200

b. P(A|C) = $\dfrac{P(A \cap C)}{P(C)} = \dfrac{.200}{.500} = .400$. If we know that the individual came from ethnic group 3, the probability that he has type A blood is .40. P(C|A) = $\dfrac{P(A \cap C)}{P(A)} = \dfrac{.200}{.447} = .447$. If a person has type A blood, the probability that he is from ethnic group 3 is .447

c. Define event D = {ethnic group 1 selected}. We are asked for P(D|B′) = $\dfrac{P(D \cap B')}{P(B')} = \dfrac{.200}{.500} = .400$. P(D∩B′)=.082 + .106 + .004 = .192, P(B′) = 1 – P(B) = 1 – [.008 + .018 + .065] = .909

46. Let event A be that the individual is more than 6 feet tall. Let event B be that the individual is a professional basketball player. Then P (A | B) = the probability of the individual being more than 6 feet tall, knowing that the individual is a professional basketball player, and P (B | A) = the probability of the individual being a professional basketball player, knowing that the individual is more than 6 feet tall. P (A | B) will be larger. Most professional BB players are tall, so the probability of an individual in that reduced sample space being more than 6 feet tall is very large. The number of individuals that are pro BB players is small in relation to the # of males more than 6 feet tall.

47.



a. P(B | A) = $\dfrac{P(A \cap B)}{P(A)} = \dfrac{.25}{.50} = .50$

b. P(B′ | A) = $\dfrac{P(A \cap B')}{P(A)} = \dfrac{.25}{.50} = .50$

c. P(A | B) = $\dfrac{P(A \cap B)}{P(B)} = \dfrac{.25}{.40} = .6125$

d. P(A′ | B) = $\dfrac{P(A' \cap B)}{P(B)} = \dfrac{.15}{.40} = .3875$

e. P(A | A∪B) = $\dfrac{P[A \cap (A \cup B)]}{P(A \cup B)} = \dfrac{.50}{.65} = .7692$

48.

a. P(A_2 | A_1) = $\dfrac{P(A_1 \cap A_2)}{P(A_1)} = \dfrac{.06}{.12} = .50$

b. P($A_1 \cap A_2 \cap A_3$ | A_1) = $\dfrac{.01}{.12} = .0833$

c. We want P[(exactly one) | (at least one)].

P(at least one) = P($A_1 \cup A_2 \cup A_3$)
= .12 + .07 + .05 - .06 - .03 - .02 + .01 = .14

Also notice that the intersection of the two events is just the 1st event, since "exactly one" is totally contained in "at least one."

So P[(exactly one) | (at least one)]= $\dfrac{.04 + .01}{.14} = .3571$

d. The pieces of this equation can be found in your answers to exercise 26 (section 2.2):

$$P(A_3' \mid A_1 \cap A_2) = \frac{P(A_1 \cap A_2 \cap A_3')}{P(A_1 \cap A_2)} = \frac{.05}{.06} = .833$$

49. The first desired probability is P(both bulbs are 75 watt $|$ at least one is 75 watt).
P(at least one is 75 watt) = 1 – P(none are 75 watt)

$$= 1 - \frac{\binom{9}{2}}{\binom{15}{2}} = 1 - \frac{36}{105} = \frac{69}{105}.$$

Notice that P[(both are 75 watt)$\cap$(at least one is 75 watt)]

$$= \text{P(both are 75 watt)} = \frac{\binom{6}{2}}{\binom{15}{2}} = \frac{15}{105}.$$

So P(both bulbs are 75 watt $|$ at least one is 75 watt) = $\dfrac{\frac{15}{105}}{\frac{69}{105}} = \dfrac{15}{69} = .2174$

Second, we want P(same rating $|$ at least one NOT 75 watt).
P(at least one NOT 75 watt) = 1 – P(both are 75 watt)

$$= 1 - \frac{15}{105} = \frac{90}{105}.$$

Now, P[(same rating)$\cap$(at least one not 75 watt)] = P(both 40 watt or both 60 watt).

$$\text{P(both 40 watt or both 60 watt)} = \frac{\binom{4}{2} + \binom{5}{2}}{\binom{15}{2}} = \frac{16}{105}$$

Now, the desired conditional probability is $\dfrac{\frac{16}{105}}{\frac{90}{105}} = \dfrac{16}{90} = .1778$

50.

a. P(M $\cap$ LS $\cap$ PR) = .05, directly from the table of probabilities

b. P(M $\cap$ Pr) = P(M,Pr,LS) + P(M,Pr,SS) = .05+.07=.12

c. P(SS) = sum of 9 probabilities in SS table = 56, P(LS) = 1 = .56 = .44

d. P(M) = .08+.07+.12+.10+.05+.07 = .49
P(Pr) = .02+.07+.07+.02+.05+.02 = .25

e. $P(M|SS \cap Pl) = \dfrac{P(M \cap SS \cap Pl)}{P(SS \cap Pl)} = \dfrac{.08}{.04 + .08 + .03} = .533$

f. $P(SS|M \cap Pl) = \dfrac{P(SS \cap M \cap Pl)}{P(M \cap Pl)} = \dfrac{.08}{.08 + .10} = .444$

P(LS|M Pl) = 1 - P(SS|M Pl) = 1 - .444 = .556

51.

a. P(R from 1st $\cap$ R from 2nd) = P(R from 2nd | R from 1st) $\bullet$ P(R from 1st)

$= \dfrac{8}{11} \bullet \dfrac{6}{10} = .436$

b. P(same numbers) = P(both selected balls are the same color)

= P(both red) + P(both green) = $.436 + \dfrac{4}{11} \bullet \dfrac{4}{10} = .581$

52. Let A_1 be the event that #1 fails and A_2 be the event that #2 fails. We assume that $P(A_1) = P(A_2) = q$ and that $P(A_1 \mid A_2) = P(A_2 \mid A_1) = r$. Then one approach is as follows:

$P(A_1 \cap A_2) = P(A_2 \mid A_1) \bullet P(A_1) = rq = .01$

$P(A_1 \cup A_2) = P(A_1 \cap A_2) + P(A_1' \cap A_2) + P(A_1 \cap A_2') = rq + 2(1-r)q = .07$

These two equations give $2q - .01 = .07$, from which $q = .04$ and $r = .25$. Alternatively, with $t = P(A_1' \cap A_2) = P(A_1 \cap A_2')$, $t + .01 + t = .07$, implying $t = .03$ and thus $q = .04$ without reference to conditional probability.

53. $P(B \mid A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{P(B)}{P(A)}$ (since B is contained in A, $A \cap B = B$)

$= \dfrac{.05}{.60} = .0833$

54. $P(A_1) = .22$, $P(A_2) = .25$, $P(A_3) = .28$, $P(A_1 \cap A_2) = .11$, $P(A_1 \cap A_3) = .05$, $P(A_2 \cap A_3) = .07$, $P(A_1 \cap A_2 \cap A_3) = .01$

a. $P(A_2 \mid A_1) = \dfrac{P(A_1 \cap A_2)}{P(A_1)} = \dfrac{.11}{.22} = .50$

b. $P(A_2 \cap A_3 \mid A_1) = \dfrac{P(A_1 \cap A_2 \cap A_3)}{P(A_1)} = \dfrac{.01}{.22} = .0455$

c. $$P(A_2 \cup A_3 \mid A_1) = \frac{P[A_1 \cap (A_2 \cup A_3)]}{P(A_1)} = \frac{P[(A_1 \cap A_2) \cup (A_1 \cap A_3)]}{P(A_1)}$$
$$= \frac{P(A_1 \cap A_2) + P(A_1 \cap A_3) - P(A_1 \cap A_2 \cap A_3)}{P(A_1)} = \frac{.15}{.22} = .682$$

d. $$P(A_1 \cap A_2 \cap A_3 \mid A_1 \cup A_2 \cup A_3) = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cup A_2 \cup A_3)} = \frac{.01}{.53} = .0189$$

This is the probability of being awarded all three projects given that at least one project was awarded.

55.

a. P(A B) = P(B|A)•P(A) = $\dfrac{2\times 1}{4\times 3} \times \dfrac{2\times 1}{6\times 5} = .0111$

b. P(two other H's next to their wives | J and M together in the middle)

$$\frac{P[(H - W.or.W - H) and (J - M.or.M - J) and (H - W.or.W - H)]}{P(J - M.or.M - J.in.the.middle)}$$

numerator = $\dfrac{4\times 1\times 2\times 1\times 2\times 1}{6\times 5\times 4\times 3\times 2\times 1} = \dfrac{16}{6!}$

denominator = $\dfrac{4\times 3\times 2\times 1\times 2\times 1}{6\times 5\times 4\times 3\times 2\times 1} = \dfrac{48}{6!}$

so the desired probability = $\dfrac{16}{48} = \dfrac{1}{3}$.

c. P(all H's next to W's | J & M together)
= P(all H's next to W's – including J&M)/P(J&M together)

$$= \frac{\dfrac{6\times1\times4\times1\times2\times1}{6!}}{\dfrac{5\times2\times1\times4\times3\times2\times1}{6!}} = \frac{48}{240} = .2$$

56. If P(B|A) > P(B), then P(B'|A) < P(B').
Proof by contradiction.
Assume P(B'|A) ≥ P(B').
Then 1 – P(B|A) ≥ 1 – P(B).
- P(B|A) ≥ – P(B).
P(B|A) ≤ P(B).
This contradicts the initial condition, therefore P(B'|A) < P(B').

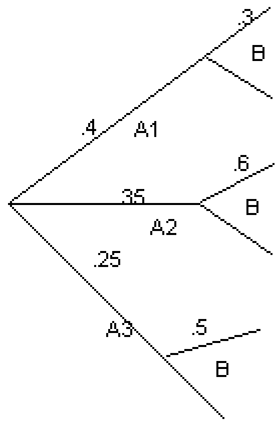
57. $$P(A\mid B)+P(A'\mid B)=\frac{P(A\cap B)}{P(B)}+\frac{P(A'\cap B)}{P(B)}=\frac{P(A\cap B)+P(A'\cap B)}{P(B)}=\frac{P(B)}{P(B)}=1$$

58. $$P(A\cup B\mid C)=\frac{P[(A\cup B)\cap C)}{P(C)}=\frac{P[(A\cap C)\cup(B\cap C)]}{P(C)}$$

$$=\frac{P(A\cap C)+P(B\cap C)-P(A\cap B\cap C)}{P(C)}$$

= P(A|C) + P(B|C) – P(A ∩ B | C)

59.



$.4 \times .3 = .12 = P(A_1 \cap B) = P(A_1) \bullet P(B \mid A)$

$.35 \times .6 = .21 = P(A_2 \cap B)$

$.25 \times .5 = .125 = P(A_3 \cap B)$

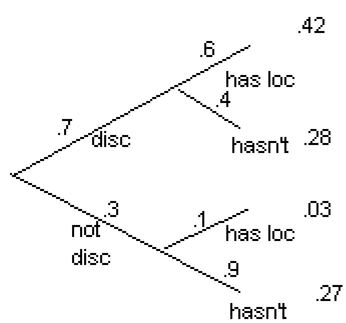
a. $P(A_2 \cap B) = .21$

b. $P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) = .455$

c. $P(A_1|B) = \dfrac{P(A_1 \cap B)}{P(B)} = \dfrac{.12}{.455} = .264$

$P(A_2|B) = \dfrac{.21}{.455} = .462$, $P(A_3|B) = 1 - .264 - .462 = .274$

60.



a. P(not disc | has loc) = $\dfrac{P(not.disc \cap has.loc)}{P(has.loc)} = \dfrac{.03}{.03 + .42} = .067$

b. P(disc | no loc) = $\dfrac{P(disc \cap no.loc)}{P(no.loc)} = \dfrac{.28}{.55} = .509$

61. P(0 def in sample | 0 def in batch) = 1

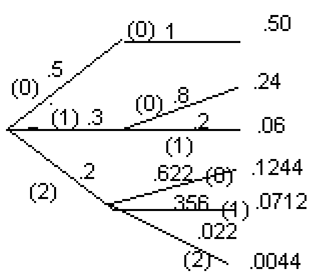
P(0 def in sample | 1 def in batch) = $\dfrac{\binom{9}{2}}{\binom{10}{2}} = .800$

P(1 def in sample | 1 def in batch) = $\dfrac{\binom{9}{1}}{\binom{10}{2}} = .200$

P(0 def in sample | 2 def in batch) = $\dfrac{\binom{8}{2}}{\binom{10}{2}} = .622$

P(1 def in sample | 2 def in batch) = $\dfrac{\binom{2}{1}\binom{8}{1}}{\binom{10}{2}} = .356$

P(2 def in sample | 2 def in batch) = $\dfrac{1}{\binom{10}{2}} = .022$



a. P(0 def in batch | 0 def in sample) = $\dfrac{.5}{.5+.24+.1244} = .578$

P(1 def in batch | 0 def in sample) = $\dfrac{.24}{.5+.24+.1244} = .278$

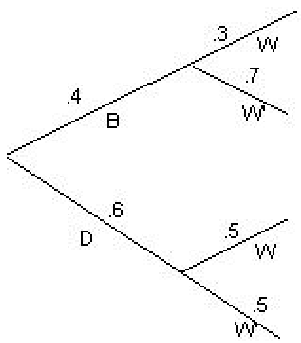
P(2 def in batch | 0 def in sample) = $\dfrac{.1244}{.5+.24+.1244} = .144$

b. P(0 def in batch | 1 def in sample) = 0

P(1 def in batch | 1 def in sample) = $\dfrac{.06}{.06+.0712}=.457$

P(2 def in batch | 1 def in sample) = $\dfrac{.0712}{.06+.0712}=.543$

62. Using a tree diagram, B = basic, D = deluxe, W = warranty purchase, W' = no warranty



$.4\times.3=.12=P(B\cap W)$

$.4\times.7=.28=P(B\cap W')$

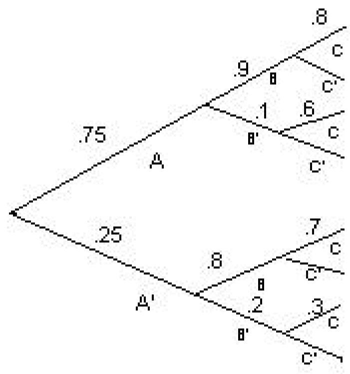
$.6\times.5=.30=P(D\cap W)$

$.6\times.5=.30=P(D\cap W')$

We want P(B|W) = $\dfrac{P(B\cap W)}{P(W)}=\dfrac{.12}{.30+.12}=\dfrac{.12}{.42}=.2857$

63.

a.



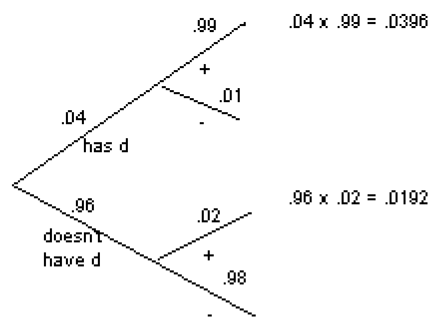
b. P(A ∩ B ∩ C) = .75 × .9 × .8 = .5400

c. P(B ∩ C) = P(A ∩ B ∩ C) + P(A′ ∩ B ∩ C)
=.5400+.25×.8×.7 = .6800

d. P(C) = P(A ∩ B ∩ C)+P(A′ ∩ B ∩ C) + P(A ∩ B′ ∩ C) + P(A′ ∩ B′ ∩ C)
= .54+.045+.14+.015 = .74

e. P(A|B ∩ C) = $\frac{P(A \cap B \cap C)}{P(B \cap C)} = \frac{.54}{.68} = .7941$

64.

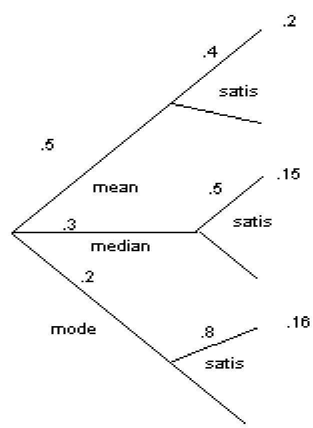


a. P(+) = .0588

b. P(has d | +) = $\dfrac{.0396}{.0588} = .6735$

c. P(doesn't have d | -) = $\dfrac{.9408}{.9412} = .9996$

65.



P(satis) = .51

P(mean | satis) = $\dfrac{.2}{.51} = .3922$

P(median | satis) = .2941

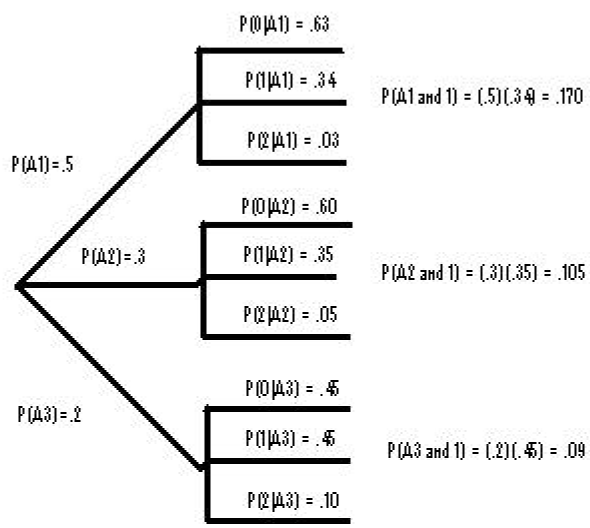
P(mode | satis) = .3137

So Mean (and not Mode!) is the most likely author, while Median is least.

66. Define events A1, A2, and A3 as flying with airline 1, 2, and 3, respectively. Events 0, 1, and 2 are 0, 1, and 2 flights are late, respectively. Event DC = the event that the flight to DC is late, and event LA = the event that the flight to LA is late. Creating a tree diagram as described in the hint, the probabilities of the second generation branches are calculated as follows: For the A1 branch, $P(0|A1) = P[DC' \cap LA'] = P[DC'] \cdot P[LA'] = (.7)(.9) = .63$; $P(1|A1) = P[(DC' \cap LA) \cup (DC \cap LA')] = (.7)(.1) + (.3)(.9) = .07 + .27 = .34$; $P(2|A1) = P[DC \cap LA] = P[DC] \cdot P[LA] = (.3)(.1) = .03$
Follow a similar pattern for A2 and A3.

From the law of total probability, we know that
$P(1) = P(A1 \cap 1) + P(A2 \cap 1) + P(A2 \cap 1)$
= (from tree diagram below) $.170 + .105 + .09 = .365$.

We wish to find $P(A1|1)$, $P(A2|1)$, and $P(A2|1)$.

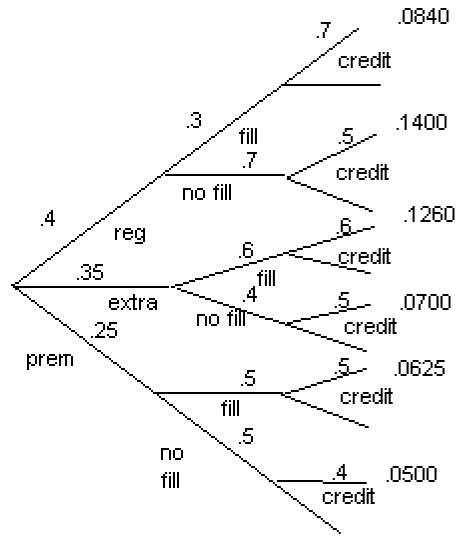


P(A1/1) = $\dfrac{P(A1 \cap 1)}{P(1)} = \dfrac{.170}{.365} = .466$;

P(A2/1) = $\dfrac{P(A2 \cap 1)}{P(1)} = \dfrac{.105}{.365} = .288$;

P(A3/1) = $\dfrac{P(A3 \cap 1)}{P(1)} = \dfrac{.090}{.365} = .247$;

67.



a. P(U ∩ F ∩ Cr) = .1260

b. P(Pr ∩ NF ∩ Cr) = .05

c. P(Pr ∩ Cr) = .0625 + .05 = .1125

d. P(F ∩ Cr) = .0840 + .1260 + .0625 = .2725

e. P(Cr) = .5325

f. P(PR | Cr) = $\dfrac{P(\text{Pr} \cap Cr)}{P(Cr)} = \dfrac{.1125}{.5325} = .2113$

Section 2.5

68. Using the definition, two events A and B are independent if $P(A|B) = P(A)$;
$P(A|B) = .6125$; $P(A) = .50$; $.6125 \neq .50$, so A and B are dependent.
Using the multiplication rule, the events are independent if
$P(A \cap B)=P(A)\bullet P(B)$;
$P(A \cap B) = .25$; $P(A) \bullet P(B) = (.5)(.4) = .2$. $.25 \neq .2$, so A and B are dependent.

69.

a. Since the events are independent, then A' and B' are independent, too. (see paragraph below equation 2.7. $P(B'|A') = .$ $P(B') = 1 - .7 = .3$

b. $P(A \cup B)=P(A) + P(B) - P(A)\cdot P(B) = .4 + .7 + (.4)(.7) = .82$

c. $P(AB' | A \cup B) = \dfrac{P(AB' \cap (A \cup B))}{P(A \cup B)} = \dfrac{P(AB')}{P(A \cup B)} = \dfrac{.12}{.82} = .146$

70. $P(A_1 \cap A_2) = .11$, $P(A_1) \bullet P(A_2) = .055$. A_1 and A_2 are not independent.
$P(A_1 \cap A_3) = .05$, $P(A_1) \bullet P(A_3) = .0616$. A_1 and A_3 are not independent.
$P(A_2 \cap A_3) = .07$, $P(A_1) \bullet P(A_3) = .07$. A_2 and A_3 are independent.

71. $P(A' \cap B) = P(B) - P(A \cap B) = P(B) - P(A) \bullet P(B) = [1 - P(A)] \bullet P(B) = P(A')\bullet P(B)$.
Alternatively, $P(A' | B) = \dfrac{P(A' \cap B)}{P(B)} = \dfrac{P(B) - P(A \cap B)}{P(B)}$

$$= \frac{P(B) - P(A) \cdot P(B)}{P(B)} = 1 - P(A) = P(A').$$

72. Using subscripts to differentiate between the selected individuals,
$P(O_1 \cap O_2) = P(O_1)\bullet P(O_2) = (.44)(.44) = .1936$
P(two individuals match) = $P(A_1 \cap A_2)+P(B_1 \cap B_2) + P(AB_1 \cap AB_2) + P(O_1 \cap O_2)$
$= .42^2 + .10^2 + .04^2 + .44^2 = .3816$

73. Let event E be the event that an error was signaled incorrectly. We want P(at least one signaled incorrectly) = $P(E_1 \cup E_2 \cup \ldots \cup E_{10}) = 1 - P(E_1' \cap E_2' \cap \ldots \cap E_{10}')$. $P(E') = 1 - .05 = .95$. For 10 independent points, $P(E_1' \cap E_2' \cap \ldots \cap E_{10}') = P(E_1')P(E_2')\ldots P(E_{10}')$ so $= P(E_1 \cup E_2 \cup \ldots \cup E_{10}) = 1 - [.95]^{10} = .401$. Similarly, for 25 points, the desired probability is $=1 - [P(E')]^{25} = 1 - (.95)^{25} = .723$

74. P(no error on any particular question) = .9, so P(no error on any of the 10 questions) =$(.9)^{10}$ = .3487. Then P(at least one error) = 1 – $(.9)^{10}$ = .6513. For **p** replacing .1, the two probabilities are $(1-\mathbf{p})^n$ and $1 - (1-\mathbf{p})^n$.

75. Let q denote the probability that a rivet is defective.

a. P(seam need rework) = .20 = 1 – P(seam doesn't need rework)
= 1 – P(no rivets are defective)
= 1 – P(1^{st} isn't def $\cap \ldots \cap$ 25^{th} isn't def)
= $1 - (1 - q)^{25}$, so $.80 = (1 - q)^{25}$, $1 - q = (.80)^{1/25}$, and thus q = 1 - .99111 = .00889.

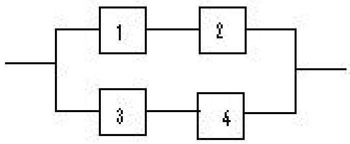
b. The desired condition is $.10 = 1 - (1 - q)^{25}$, i.e. $(1 - q)^{25} = .90$, from which q = 1 - .99579 = .00421.

76. P(at least one opens) = 1 – P(none open) = $1 - (.05)^5$ = .99999969
P(at least one fails to open) = 1 = P(all open) = $1 - (.95)^5$ = .2262

77. Let A_1 = older pump fails, A_2 = newer pump fails, and $x = P(A_1 \cap A_2)$. Then $P(A_1) = .10 + x$, $P(A_2) = .05 + x$, and $x = P(A_1 \cap A_2) = P(A_1) \bullet P(A_2) = (.10 + x)(.05 + x)$. The resulting quadratic equation, $x^2 - .85x + .005 = 0$, has roots x = .0059 and x = .8441. Hopefully the smaller root is the actual probability of system failure.

78. P(system works) = P(1 – 2 works $\cup$ 3 – 4 works)
= P(1 – 2 works) + P(3 – 4 works) - P(1 – 2 works $\cap$ 3 – 4 works)
= P(1 works $\cup$ 2 works) + P(3 works $\cap$ 4 works) – P(1 – 2) $\bullet$ P(3 – 4)
= (.9+.9-.81) + (.9)(.9) – (.9+.9-.81)(.9)(.9)
= .99 + .81 - .8019 = .9981

79.



Using the hints, let $P(A_i) = p$, and $x = p^2$, then P(system lifetime exceeds t_0) $= p^2 + p^2 - p^4 = 2p^2 - p^4 = 2x - x^2$. Now, set this equal to .99, or $2x - x^2 = .99 \Rightarrow x^2 - 2x + .99 = 0$. Use the quadratic formula to solve for x: $= \dfrac{2 \pm \sqrt{4-(4)(.99)}}{2} = \dfrac{2 \pm .2}{2} = 1 \pm .1 = .99$ or 1.01

Since the value we want is a probability, and has to be = 1, we use the value of .99.

80. Event A: { (3,1)(3,2)(3,3)(3,4)(3,5)(3,6) }, $P(A) = \frac{1}{6}$;

Event B: { (1,4)(2,4)(3,4)(4,4)(5,4)(6,4) }, $P(B) = \frac{1}{6}$;

Event C: { (1,6)(2,5)(3,4)(4,3)(5,2)(6,1) }, $P(C) = \frac{1}{6}$;

Event $A \cap B$: { (3,4) }; $P(A \cap B) = \frac{1}{36}$;

Event $A \cap C$: { (3,4) }; $P(A \cap C) = \frac{1}{36}$;

Event $B \cap C$: { (3,4) }; $P(A \cap C) = \frac{1}{36}$;

Event $A \cap B \cap C$: { (3,4) }; $P(A \cap B \cap C) = \frac{1}{36}$;

$P(A) \cdot P(B) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = P(A \cap B)$

$P(A) \cdot P(C) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = P(A \cap C)$

$P(B) \cdot P(C) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = P(B \cap C)$

The events are pairwise independent.

$P(A) \cdot P(B) \cdot P(C) = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{216} \neq \frac{1}{36} = P(A \cap B \cap C)$

The events are not mutually independent

81. P(both detect the defect) = 1 – P(at least one doesn't) = 1 - .2 = .8

a. P(1^{st} detects $\cap$ 2^{nd} doesn't) = P(1^{st} detects) – P(1^{st} does $\cap$ 2^{nd} does)
= .9 - .8 = .1
Similarly, P(1^{st} doesn't $\cap$ 2^{nd} does) = .1, so P(exactly one does)= .1+.1= .2

b. P(neither detects a defect) = 1 – [P(both do) + P(exactly 1 does)]
= 1 – [.8+.2] = 0
so P(all 3 escape) = (0)(0)(0) = 0.

82. P(pass) = .70

a. (.70)(.70)(.70) = .343

b. 1 – P(all pass) = 1 - .343 = .657

c. P(exactly one passes) = (.70)(.30)(.30) + (.30)(.70)(.30) + (.30)(.30)(.70) = .189

d. P(# pass $\le$ 1) = P(0 pass) + P(exactly one passes) = $(.3)^3$ + .189 = .216

e. P(3 pass | 1 or more pass) =

$$= \frac{P(3.pass \cap \ge 1.pass)}{P(\ge 1.pass)} = \frac{P(3.pass)}{P(\ge 1.pass)} = \frac{.343}{.973} = .353$$

83.

a. Let D_1 = detection on 1^{st} fixation, D_2 = detection on 2^{nd} fixation.
P(detection in at most 2 fixations) = $P(D_1) + P(D_1' \cap D_2)$
$= P(D_1) + P(D2 \mid D1')P(D_1)$
= p + p(1 – p) = p(2 – p).

b. Define $D_1, D_2, \dots, D_n$ as in **a**. Then P(at most n fixations)
$= P(D_1) + P(D_1' \cap D_2) + P(D_1' \cap D_2' \cap D_3) + \dots + P(D_1' \cap D_2' \cap \dots \cap D_{n-1}' \cap D_n)$
$= p + p(1 – p) + p(1 – p)^2 + \dots + p(1 – p)^{n-1}$

$$= p\,[\,1 + (1 – p) + (1 – p)^2 + \dots + (1 – p)^{n-1}] = p \bullet \frac{1-(1-p)^n}{1-(1-p)} = 1-(1-p)^n$$

Alternatively, P(at most n fixations) = 1 – P(at least n+1 are req'd)
= 1 – P(no detection in 1^{st} n fixations)
$= 1 – P(D_1' \cap D_2' \cap \dots \cap D_n')$
$= 1 – (1 – p)^n$

c. P(no detection in 3 fixations) = $(1 – p)^3$

d. P(passes inspection) = P({not flawed} $\cup$ {flawed and passes})
= P(not flawed) + P(flawed and passes)
= .9 + P(passes | flawed)$\bullet$ P(flawed) = .9+(1 – p)^3(.1)

e. P(flawed | passed) = $\dfrac{P(flawed \cap passed)}{P(passed)} = \dfrac{.1(1-p)^3}{.9+.1(1-p)^3}$

For p = .5, P(flawed | passed) = $\dfrac{.1(.5)^3}{.9+.1(.5)^3} = .0137$

84.

a. P(A) = $\dfrac{2000}{10{,}000} = .02$, P(B) = P(A $\cap$ B) + P(A′ $\cap$ B)

= P(B|A) P(A) + P(B|A′) P(A′) = $\dfrac{1999}{9999} \bullet (.2) + \dfrac{2000}{9999} \bullet (.8) = .2$

P(A $\cap$ B) = .039984; since P(A $\cap$ B) $\neq$ P(A)P(B), the events are not independent.

b. P(A $\cap$ B) = .04. Very little difference. Yes.

c. P(A) = P(B) = .2, P(A)P(B) = .04, but P(A $\cap$ B) = P(B|A)P(A) = $\frac{1}{9} \cdot \frac{2}{10} = .0222$, so the two numbers are quite different.
In **a**, the sample size is small relative to the “population” size, while here it is not.

85. P(system works) = P(1 – 2 works $\cap$ 3 – 4 – 5 – 6 works $\cap$ 7 works)
= P(1 – 2 works) $\bullet$ P(3 – 4 – 5 – 6 works) $\bullet$P(7 works)
= (.99) (.9639) (.9) = .8588
With the subsystem in figure 2.14 connected in parallel to this subsystem,
P(system works) = .8588+.927 – (.8588)(.927) = .9897

86.

a. For route #1, P(late) = P(stopped at 2 or 3 or 4 crossings)
$= 1 - P(\text{stopped at 0 or 1}) = 1 - [.9^4 + 4(.9)^3(.1)]$
$= .0523$

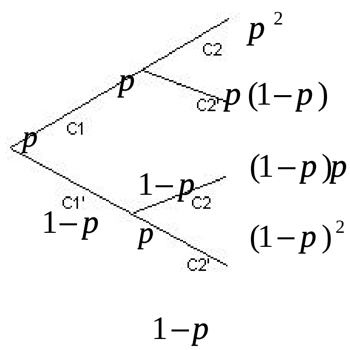
For route #2, P(late) = P(stopped at 1 or 2 crossings)
$= 1 - P(\text{stopped at none}) = 1 - .81 = .19$

thus route #1 should be taken.

b. P(4 crossing route | late) = $\dfrac{P(4 crossing \cap late)}{P(late)}$

$$= \frac{.(.5)(.0523)}{(.5)(.0523) + (.5)(.19)} = .216$$

87.



P(at most 1 is lost) = 1 – P(both lost)
$= 1 - \pi^2$

P(exactly 1 lost) = $2\pi(1 - \pi)$

P(exactly 1 | at most 1) = $\dfrac{P(exactly 1)}{P(at.most 1)} = \dfrac{2p(1-p)}{1-p^2}$

Supplementary Exercises

88.

a. $\binom{20}{3} = 1140$

b. $\binom{19}{3} = 969$

c. # having at least 1 of the 10 best = 1140 - # of crews having none of 10 best = 1140 - $\binom{10}{3} = 1140$ - 120 = 1020

d. P(best will not work) = $\dfrac{969}{1140} = .85$

89.

a. P(line 1) = $\dfrac{500}{1500} = .333$;

P(Crack) = $\dfrac{.50(500) + .44(400) + .40(600)}{1500} = \dfrac{666}{1500} = .444$

b. P(Blemish | line 1) = .15

c. P(Surface Defect) = $\dfrac{.10(500) + .08(400) + .15(600)}{1500} = \dfrac{172}{1500}$

P(line 1 and Surface Defect) = $\dfrac{.10(500)}{1500} = \dfrac{50}{1500}$

So P(line 1 | Surface Defect) = $= \dfrac{50/1500}{172/1500} = .291$

90.

a. The only way he will have one type of forms left is if they are all course substitution forms. He must choose all 6 of the withdrawal forms to pass to a subordinate. The desired probability is $\dfrac{\binom{6}{6}}{\binom{10}{6}} = .00476$

b. He can start with the wd forms: W-C-W-C or with the cs forms: C-W-C-W:

# of ways: $6 \times 4 \times 5 \times 3 + 4 \times 6 \times 3 \times 5 = 2(360) = 720$;

The total # ways to arrange the four forms: $10 \times 9 \times 8 \times 7 = 5040$.

The desired probability is 720/5040 = .1429

91. $P(A \cup B) = P(A) + P(B) - P(A)P(B)$

$.626 = P(A) + P(B) - .144$

So P(A) + P(B) = .770 and P(A)P(B) = .144.
Let x = P(A) and y = P(B), then using the first equation, y = .77 – x, and substituting this into the second equation, we get x (.77 – x) = .144 or
$x^2 - .77x + .144 = 0$. Use the quadratic formula to solve:

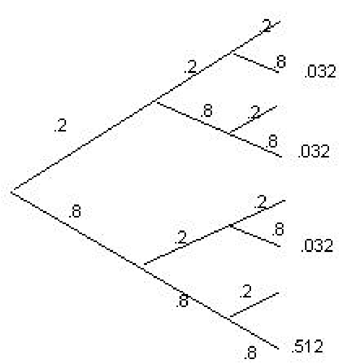
$$\frac{.77 \pm \sqrt{.77^2 - (4)(.144)}}{2} = \frac{.77 \pm .13}{2} = .32 \text{ or } .45$$

So P(A) = .45 and P(B) = .32

92.

a. (.8)(.8)(.8) = .512

b.



.512+.032+.023+.023 = .608

c. P(1 sent | 1 received) = $\dfrac{P(1sent \cap 1received)}{P(1received} = \dfrac{.4256}{.5432} = .7835$

93.

a. There are 5×4×3×2×1 = 120 possible orderings, so P(BCDEF) = $\frac{1}{120} = .0083$

b. # orderings in which F is 3rd = 4×3×1*×2×1 = 24, (* because F must be here), so P(F 3rd) = $\frac{24}{120} = .2$

c. P(F last) = $\dfrac{4\times3\times2\times1\times1}{120} = .2$

94. P(F hasn't heard after 10 times) = P(not on #1 ∩ not on #2 ∩…∩ not on #10)

$$= \left(\frac{4}{5}\right)^{10} = .1074$$

95. When three experiments are performed, there are 3 different ways in which detection can occur on exactly 2 of the experiments: (i) #1 and #2 and not #3 (ii) #1 and not #2 and #3; (iii) not#1 and #2 and #3. If the impurity is present, the probability of exactly 2 detections in three (independent) experiments is (.8)(.8)(.2) + (.8)(.2)(.8) + (.2)(.8)(.8) = .384. If the impurity is absent, the analogous probability is 3(.1)(.1)(.9) = .027. Thus

P(present | detected in exactly 2 out of 3) =

$$\frac{P(\det ected.in.exactly.2 \cap present)}{P(\det ected.in.exactly.2)}$$

$$= \frac{(.384)(.4)}{(.384)(.4) + (.027)(.6)} = .905$$

96. P(exactly 1 selects category #1 | all 3 are different)

$$= \frac{P(exactly.1.selects\#1 \cap all.are.different)}{P(all.are.different)}$$

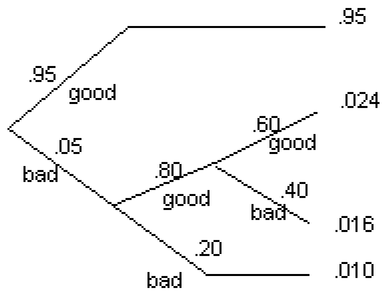
Denominator = $\dfrac{6\times5\times4}{6\times6\times6} = \dfrac{5}{9} = .5556$

Numerator = 3 P(contestant #1 selects category #1 and the other two select two different categories)

$$= 3\times\frac{1\times5\times4}{6\times6\times6} = \frac{5\times4\times3}{6\times6\times6}$$

The desired probability is then $\dfrac{5\times4\times3}{6\times5\times4} = \dfrac{1}{2} = .5$

97.



a. P(pass inspection) = P(pass initially $\cup$ passes after recrimping) = P(pass initially) + P(fails initially $\cap$ goes to recrimping $\cap$ is corrected after recrimping)
= .95 + (.05)(.80)(.60) (following path "bad-good-good" on tree diagram)
= .974

b. P(needed no recrimping | passed inspection) = $\dfrac{P(passed.initially)}{P(passed.inspection)}$

$$= \frac{.95}{.974} = .9754$$

98.

a. P(both +) = P(carrier $\cap$ both +) + P(not a carrier $\cap$ both +)
=P(both + | carrier) x P(carrier)
+ P(both + | not a carrier) x P(not a carrier)
= $(.90)^2(.01) + (.05)^2(.99) = .01058$
P(both –) = $(.10)^2(.01) + (.95)^2(.99) = .89358$
P(tests agree) = .01058 + .89358 = .90416

b. P(carrier | both + ve) = $\dfrac{P(carrier \cap both.positive)}{P(both.positive)} = \dfrac{(.90)^2(.01)}{.01058} = .7656$

99. Let A = 1st functions, B = 2nd functions, so P(B) = .9, P(A $\cup$ B) = .96, P(A $\cap$ B)=.75. Thus, P(A $\cup$ B) = P(A) + P(B) - P(A $\cap$ B) = P(A) + .9 - .75 = .96, implying P(A) = .81.

This gives P(B | A) = $\dfrac{P(B \cap A)}{P(A)} = \dfrac{.75}{.81} = .926$

100. $P(E_1 \cap \text{late}) = P(\text{late} \mid E_1)P(E_1) = (.02)(.40) = .008$

101.

a. The law of total probability gives

$$P(\text{late}) = \sum_{i=1}^{3} P(late \mid E_i)\cdot P(E_i)$$

$$= (.02)(.40) + (.01)(.50) + (.05)(.10) = .018$$

b. $P(E_1' \mid \text{on time}) = 1 - P(E_1 \mid \text{on time})$

$$= 1 - \frac{P(E_1 \cap on.time)}{P(on.time)} = 1 - \frac{(.98)(.4)}{.982} = .601$$

102. Let B denote the event that a component needs rework. Then

$$P(B) = \sum_{i=1}^{3} P(B \mid A_i)\cdot P(A_i) = (.05)(.50) + (.08)(.30) + (.10)(.20) = .069$$

Thus $P(A_1 \mid B) = \dfrac{(.05)(.50)}{.069} = .362$

$$P(A_2 \mid B) = \frac{(.08)(.30)}{.069} = .348$$

$$P(A_3 \mid B) = \frac{(.10)(.20)}{.069} = .290$$

103.

a. $P(\text{all different}) = \dfrac{(365)(364)...(356)}{(365)^{10}} = .883$

P(at least two the same) = 1 - .883 = .117

b. P(at least two the same) = .476 for k=22, and = .507 for k=23

c. P(at least two have the same SS number) = 1 – P(all different)

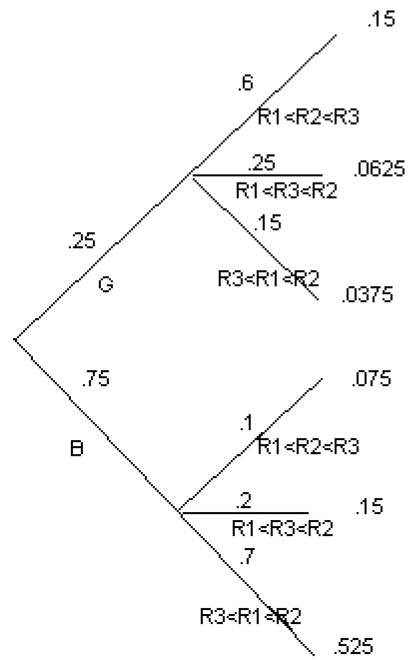
$$= 1 - \frac{(1000)(999)...(991)}{(1000)^{10}}$$

$$= 1 - .956 = .044$$

Thus P(at least one "coincidence") = P(BD coincidence $\cup$ SS coincidence)

$$= .117 + .044 - (.117)(.044) = .156$$

104.



a. P(G | $R_1 < R_2 < R_3$) = $\dfrac{.15}{.15+.075} = .67$, P(B | $R_1 < R_2 < R_3$) = .33, classify as granite.

b. P(G | $R_1 < R_3 < R_2$) = $\dfrac{.0625}{.2125} = .2941$ < .05, so classify as basalt.

P(G | $R_3 < R_1 < R_2$) = $\dfrac{.0375}{.5625} = .0667$, so classify as basalt.

c. P(erroneous classif) = P(B classif as G) + P(G classif as B)
= P(classif as G | B)P(B) + P(classif as B | G)P(G)
= P($R_1 < R_2 < R_3$ | B)(.75) + P($R_1 < R_3 < R_2$ or $R_3 < R_1 < R_2$ | G)(.25)
= (.10)(.75) + (.25 + .15)(.25) = .175

d. For what values of p will $P(G \mid R_1<R_2<R_3) > .5$, $P(G \mid R_1 < R_3 < R_2) > .5$, $P(G \mid R_3 < R_1 < R_2) > .5$?

$P(G \mid R_1 < R_2 < R_3) = \dfrac{.6p}{.6p + .1(1-p)} = \dfrac{.6p}{.1+.5p} > .5$ iff $p > \dfrac{1}{7}$

$P(G \mid R_1 < R_3 < R_2) = \dfrac{.25p}{.25p + .2(1-p)} > .5$ iff $p > \dfrac{4}{9}$

$P(G \mid R_3 < R_1 < R_2) = \dfrac{.15p}{.15p + .7(1-p)} > .5$ iff $p > \dfrac{14}{17}$ (most restrictive)

If $p > \dfrac{14}{17}$ always classify as granite.

105. P(detection by the end of the nth glimpse) = 1 – P(not detected in 1st n)

$= 1 - P(G_1' \cap G_2' \cap \ldots \cap G_n') = 1 - P(G_1')P(G_2') \ldots P(G_n')$

$= 1 - (1-p_1)(1-p_2) \ldots (1-p_n) = 1 - \prod_{i=1}^{n}(1-p_i)$

106.

a. P(walks on 4th pitch) = P(1st 4 pitches are balls) = $(.5)^4 = .0625$

b. P(walks on 6th) = P(2 of the 1st 5 are strikes, #6 is a ball)
= P(2 of the 1st 5 are strikes)P(#6 is a ball)
$= [10(.5)^5](.5) = .15625$

c. P(Batter walks) = P(walks on 4th) + P(walks on 5th) + P(walks on 6th)
= .0625 + .15625 + .15625 = .375

d. P(first batter scores while no one is out) = P(first 4 batters walk)
$=(.375)^4 = .0198$

107.

a. P(all in correct room) = $\dfrac{1}{4\times3\times2\times1} = \dfrac{1}{24} = .0417$

b. The 9 outcomes which yield incorrect assignments are: 2143, 2341, 2413, 3142, 3412, 3421, 4123, 4321, and 4312, so P(all incorrect) = $\dfrac{9}{24} = .375$

108.

a. P(all full) = $P(A \cap B \cap C)$ = (.6)(.5)(.4) = .12
P(at least one isn't full) = 1 – P(all full) = 1 - .12 = .88

b. P(only NY is full) = $P(A \cap B' \cap C') = P(A)P(B')P(C')$ = .18
Similarly, P(only Atlanta is full) = .12 and P(only LA is full) = .08
So P(exactly one full) = .18 + .12 + .08 = .38

109. Note: s = 0 means that the very first candidate interviewed is hired. Each entry below is the candidate hired for the given policy and outcome.

Outcome	s=0	s=1	s=2	s=3	**Outcome**	s=0	s=1	s=2	s=3
1234	1	4	4	4	3124	3	1	4	4
1243	1	3	3	3	3142	3	1	4	2
1324	1	4	4	4	3214	3	2	1	4
1342	1	2	2	2	3241	3	2	1	1
1423	1	3	3	3	3412	3	1	1	2
1432	1	2	2	2	3421	3	2	2	1
2134	2	1	4	4	4123	4	1	3	3
2143	2	1	3	3	4132	4	1	2	2
2314	2	1	1	4	4213	4	2	1	3
2341	2	1	1	1	4231	4	2	1	1
2413	2	1	1	3	4312	4	3	1	2
2431	2	1	1	1	4321	4	3	2	1

s	0	1	2	3
P(hire#1)	$\frac{6}{24}$	$\frac{11}{24}$	$\frac{10}{24}$	$\frac{6}{24}$

So s = 1 is best.

110. P(at least one occurs) = 1 – P(none occur)

$$= 1 - (1-p_1)(1-p_2)(1-p_3)(1-p_4)$$
$$= p_1p_2(1-p_3)(1-p_4) + \ldots + (1-p_1)(1-p_2)p_3p_4$$
$$+ (1-p_1)p_2p_3p_4 + \ldots + p_1p_2p_3(1-p_4) + p_1p_2p_3p_4$$

111. $P(A_1)$ = P(draw slip 1 or 4) = ½; $P(A_2)$ = P(draw slip 2 or 4) = ½;
$P(A_3)$ = P(draw slip 3 or 4) = ½; $P(A_1 \cap A_2)$ = P(draw slip 4) = ¼;
$P(A_2 \cap A_3)$ = P(draw slip 4) = ¼; $P(A_1 \cap A_3)$ = P(draw slip 4) = ¼
Hence $P(A_1 \cap A_2) = P(A_1)P(A_2)$ = ¼, $P(A_2 \cap A_3) = P(A_2)P(A_3)$ = ¼,
$P(A_1 \cap A_3) = P(A_1)P(A_3)$ = ¼, thus there exists pairwise independence
$P(A_1 \cap A_2 \cap A_3)$ = P(draw slip 4) = ¼ $\neq$ 1/8 = $P(A_1)p(A_2)P(A_3)$, so the events are not mutually independent.

CHAPTER 3

Section 3.1

1.

S:	FFF	SFF	FSF	FFS	FSS	SFS	SSF	SSS
X:	0	1	1	1	2	2	2	3

2. X = 1 if a randomly selected book is non-fiction and X = 0 otherwise
X = 1 if a randomly selected executive is a female and X = 0 otherwise
X = 1 if a randomly selected driver has automobile insurance and X = 0 otherwise

3. M = the difference between the large and the smaller outcome with possible values 0, 1, 2, 3, 4, or 5; W = 1 if the sum of the two resulting numbers is even and W = 0 otherwise, a Bernoulli random variable.

4. In my perusal of a zip code directory, I found no 00000, nor did I find any zip codes with four zeros, a fact which was not obvious. Thus possible X values are 2, 3, 4, 5 (and not 0 or 1). X = 5 for the outcome 15213, X = 4 for the outcome 44074, and X = 3 for 94322.

5. No. In the experiment in which a coin is tossed repeatedly until a H results, let Y = 1 if the experiment terminates with at most 5 tosses and Y = 0 otherwise. The sample space is infinite, yet Y has only two possible values.

6. Possible X values are1, 2, 3, 4, … (all positive integers)

Outcome:	RL	AL	RAARL	RRRRL	AARRL
X:	2	2	5	5	5

7.

a. Possible values are 0, 1, 2, …, 12; discrete

b. With N = # on the list, values are 0, 1, 2, … , N; discrete

c. Possible values are 1, 2, 3, 4, … ; discrete

d. $\{ x: 0 < x < \infty \}$ if we assume that a rattlesnake can be arbitrarily short or long; not discrete

e. With c = amount earned per book sold, possible values are 0, c, 2c, 3c, … , 10,000c; discrete

f. $\{ y: 0 < y < 14\}$ since 0 is the smallest possible pH and 14 is the largest possible pH; not discrete

g. With m and M denoting the minimum and maximum possible tension, respectively, possible values are $\{ x: m < x < M \}$; not discrete

h. Possible values are 3, 6, 9, 12, 15, … -- i.e. 3(1), 3(2), 3(3), 3(4), …giving a first element, etc,; discrete

8. Y = 3 : SSS; Y = 4: FSSS; Y = 5: FFSSS, SFSSS;
Y = 6: SSFSSS, SFFSSS, FSFSSS, FFFSSS;
Y = 7: SSFFS, SFSFSSS, SFFFSSS, FSSFSSS, FSFFSSS, FFSFSSS, FFFFSSS

9.

a. Returns to 0 can occur only after an even number of tosses; possible S values are 2, 4, 6, 8, …(i.e. 2(1), 2(2), 2(3), 2(4),…) an infinite sequence, so x is discrete.

b. Now a return to 0 is possible after any number of tosses greater than 1, so possible values are 2, 3, 4, 5, … (1+1,1+2, 1+3, 1+4, …, an infinite sequence) and X is discrete

10.

a. T = total number of pumps in use at both stations. Possible values: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

b. X: -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6

c. U: 0, 1, 2, 3, 4, 5, 6

d. Z: 0, 1, 2

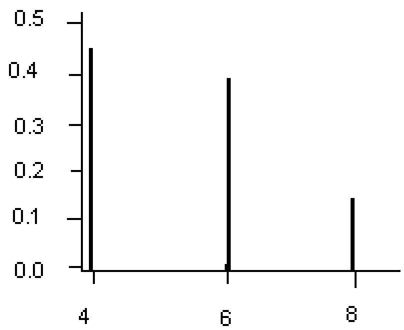
Section 3.2

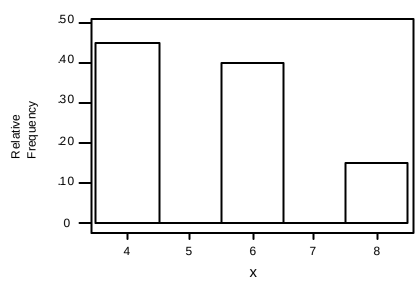
11.

a.

x	4	6	8
P(x)	.45	.40	.15

b.





c. P(x = 6) = .40 + .15 = .55 P(x > 6) = .15

12.

a. In order for the flight to accommodate all the ticketed passengers who show up, no more than 50 can show up. We need y = 50.
P(y = 50) = .05 + .10 + .12 + .14 + .25 + .17 = .83

b. Using the information in a. above, P(y > 50) = 1 - P(y = 50) = 1 - .83 = .17

c. For you to get on the flight, at most 49 of the ticketed passengers must show up. P(y = 49) = .05 + .10 + .12 + .14 + .25 = .66. For the 3rd person on the standby list, at most 47 of the ticketed passengers must show up. P(y = 44) = .05 + .10 + .12 = .27

13.

a. $P(X \le 3) = p(0) + p(1) + p(2) + p(3) = .10+.15+.20+.25 = .70$

b. $P(X < 3) = P(X \le 2) = p(0) + p(1) + p(2) = .45$

c. $P(3 \le X) = p(3) + p(4) + p(5) + p(6) = .55$

d. $P(2 \le X \le 5) = p(2) + p(3) + p(4) + p(5) = .71$

e. The number of lines not in use is $6 – X$, so $6 – X = 2$ is equivalent to $X = 4$, $6 – X = 3$ to $X = 3$, and $6 – X = 4$ to $X = 2$. Thus we desire $P(2 \le X \le 4) = p(2) + p(3) + p(4) = .65$

f. $6 – X \ge 4$ if $6 – 4 \ge X$, i.e. $2 \ge X$, or $X \le 2$, and $P(X \le 2) = .10+.15+.20 = .45$

14.

a. $\sum_{y=1}^{5} p(y)$ = K[1 + 2 + 3 + 4 + 5] = 15K = 1 $\Rightarrow K = \frac{1}{15}$

b. $P(Y \le 3) = p(1) + p(2) + p(3) = \frac{6}{15} = .4$

c. $P(2 \le Y \le 4) = p(2) + p(3) + p(4) = \frac{9}{15} = .6$

d. $\sum_{y=1}^{5}\left(\frac{y^2}{50}\right) = \frac{1}{50}[1+4+9+16+25] = \frac{55}{50} \ne 1$; No

15.

a. (1,2) (1,3) (1,4) (1,5) (2,3) (2,4) (2,5) (3,4) (3,5) (4,5)

b. $P(X = 0) = p(0) = P[\{ (3,4)\ (3,5)\ (4,5)\}] = \frac{3}{10} = .3$
$P(X = 2) = p(2) = P[\{ (1,2) \}] = \frac{1}{10} = .1$
$P(X = 1) = p(1) = 1 – [p(0) + p(2)] = .60$, and $p(x) = 0$ if $x \ne 0, 1, 2$

c. $F(0) = P(X \le 0) = P(X = 0) = .30$
$F(1) = P(X \le 1) = P(X = 0 \text{ or } 1) = .90$
$F(2) = P(X \le 2) = 1$

The c.d.f. is

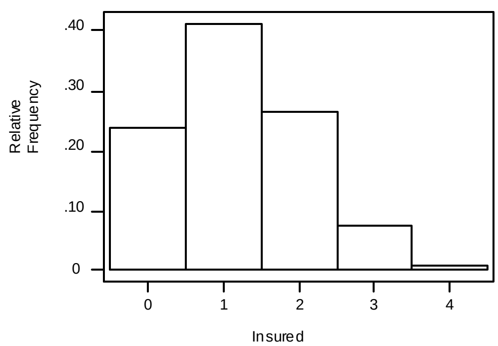
$$F(x) = \begin{cases} 0 & x < 0 \\ .30 & 0 \le x < 1 \\ .90 & 1 \le x < 2 \\ 1 & 2 \le x \end{cases}$$

16.

a.

x	Outcomes	p(x)	
0	FFFF	$(.7)^4$	=.2401
1	FFFS,FFSF,FSFF,SFFF	$4[(.7)^3(.3)]$	=.4116
2	FFSS,FSFS,SFFS,FSSF,SFSF,SSFF	$6[(.7)^2(.3)^2]$	=.2646
3	FSSS, SFSS,SSFS,SSSF	$4[(.7)(.3)^3]$	=.0756
4	SSSS	$(.3)^4$	=.0081

b.



c. p(x) is largest for X = 1

d. $P(X \geq 2) = p(2) + p(3) + p(4) = .2646+.0756+.0081 = .3483$
This could also be done using the complement.

17.

a. $P(2) = P(Y = 2) = P(1^{st}$ 2 batteries are acceptable)
$= P(AA) = (.9)(.9) = .81$

b. $p(3) = P(Y = 3) = P(UAA \text{ or } AUA) = (.1)(.9)^2 + (.1)(.9)^2 = 2[(.1)(.9)^2] = .162$

c. The fifth battery must be an A, and one of the first four must also be an A. Thus, p(5) = P(AUUUA or UAUUA or UUAUA or UUUAA) = $4[(.1)^3(.9)^2] = .00324$

d. $P(Y = y) = p(y) =$ P(the y^{th} is an A and so is exactly one of the first y – 1)
$=(y - 1)(.1)^{y-2}(.9)^2$, y = 2,3,4,5,…

18.

a. p(1) = P(M = 1) = P[(1,1)] = $\frac{1}{36}$

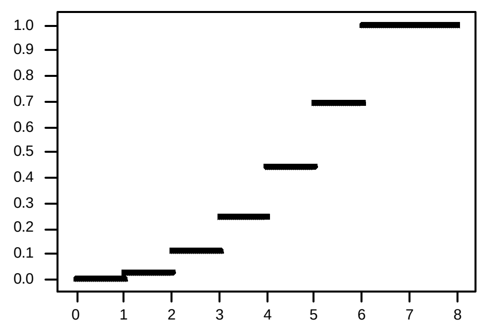
p(2) = P(M = 2) = P[(1,2) or (2,1) or (2,2)] = $\frac{3}{36}$

p(3) = P(M = 3) = P[(1,3) or (2,3) or (3,1) or (3,2) or (3,3)] = $\frac{5}{36}$

Similarly, p(4) = $\frac{7}{36}$, p(5) = $\frac{9}{36}$, and p(6) = $\frac{11}{36}$

b. F(m) = 0 for m < 1, $\frac{1}{36}$ for 1 ≤ m < 2,

$$F(m) = \begin{cases} 0 & m < 1 \\ \frac{1}{36} & 1 \le m < 2 \\ \frac{4}{36} & 2 \le m < 3 \\ \frac{9}{36} & 3 \le m < 4 \\ \frac{16}{36} & 4 \le m < 5 \\ \frac{25}{36} & 5 \le m < 6 \\ 1 & m \ge 6 \end{cases}$$



19. Let A denote the type O+ individual (type O positive blood) and B, C, D, the other 3 individuals. Then p(1) – P(Y = 1) = P(A first) = $\frac{1}{4}$ = .25

p(2) = P(Y = 2) = P(B, C, or D first and A next) = $\frac{3}{4} \cdot \frac{1}{3} = \frac{1}{4} = .25$

p(4) = P(Y = 3) = P(A last) = $\frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{4} = .25$

So p(3) = 1 – (.25+.25+.25) = .25

20. P(0) = P(Y = 0) = P(both arrive on Wed.) = (.3)(.3) = .09
P(1) = P(Y = 1) = P[(W,Th)or(Th,W)or(Th,Th)]
= (.3)(.4) + (.4)(.3) + (.4)(.4) = .40
P(2) = P(Y = 2) = P[(W,F)or(Th,F)or(F,W) or (F,Th) or (F,F)] = .32
P(3) = 1 – [.09 + .40 + .32] = .19

21. The jumps in F(x) occur at x = 0, 1, 2, 3, 4, 5, and 6, so we first calculate F( ) at each of these values:

$F(0) = P(X \le 0) = P(X = 0) = .10$
$F(1) = P(X \le 1) = p(0) + p(1) = .25$
$F(2) = P(X \le 2) = p(0) + p(1) + p(2) = .45$
$F(3) = .70$, $F(4) = .90$, $F(5) = .96$, and $F(6) = 1$.

The c.d.f. is

$$F(x) = \begin{cases} .00 & x < 0 \\ .10 & 0 \le x < 1 \\ .25 & 1 \le x < 2 \\ .45 & 2 \le x < 3 \\ .70 & 3 \le x < 4 \\ .90 & 4 \le x < 5 \\ .96 & 5 \le x < 6 \\ 1.00 & 6 \le x \end{cases}$$

Then $P(X \le 3) = F(3) = .70$, $P(X < 3) = P(X \le 2) = F(2) = .45$,
$P(3 \le X) = 1 - P(X \le 2) = 1 - F(2) = 1 - .45 = .55$,
and $P(2 \le X \le 5) = F(5) - F(1) = .96 - .25 = .71$

22.

a. $P(X = 2) = .39 - .19 = .20$

b. $P(X > 3) = 1 - .67 = .33$

c. $P(2 \le X \le 5) = .92 - .19 = .78$

d. $P(2 < X < 5) = .92 - .39 = .53$

23.

a. Possible X values are those values at which F(x) jumps, and the probability of any particular value is the size of the jump at that value. Thus we have:

x	1	3	4	6	12
p(x)	.30	.10	.05	.15	.40

b. $P(3 \le X \le 6) = F(6) - F(3\text{-}) = .60 - .30 = .30$
$P(4 \le X) = 1 - P(X < 4) = 1 - F(4\text{-}) = 1 - .40 = .60$

24. $P(0) = P(Y = 0) = P(\text{B first}) = p$
$P(1) = P(Y = 1) = P(\text{G first, then B}) = P(GB) = (1 - p)p$
$P(2) = P(Y = 2) = P(GGB) = (1 - p)^2 p$
Continuing, $p(y) = P(Y=y) = P(\text{y G's and then a B}) = (1 - p)^y p$ for $y = 0,1,2,3,\ldots$

25.

a. Possible X values are 1, 2, 3, …

P(1) = P(X = 1) = P(return home after just one visit) = $\frac{1}{3}$

P(2) = P(X = 2) = P(second visit and then return home) = $\frac{2}{3} \cdot \frac{1}{3}$

P(3) = P(X = 3) = P(three visits and then return home) = $\left(\frac{2}{3}\right)^2 \cdot \frac{1}{3}$

In general p(x) = $\left(\frac{2}{3}\right)^{x-1}\left(\frac{1}{3}\right)$ for x = 1, 2, 3, …

b. The number of straight line segments is Y = 1 + X (since the last segment traversed returns Alvie to O), so as in a, p(y) = $\left(\frac{2}{3}\right)^{y-2}\left(\frac{1}{3}\right)$ for y = 2, 3, …

c. Possible Z values are 0, 1, 2, 3 , …

p(0) = P(male first and then home) = $\frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$,

p(1) = P(exactly one visit to a female) = P(female 1st, then home) + P(F, M, home) + P(M, F, home) + P(M, F, M, home)

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

$= \left(\frac{1}{2}\right)\left(1 + \frac{2}{3}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{2}\right)\left(\frac{2}{3}\right)\left(\frac{2}{3} + 1\right)\left(\frac{1}{3}\right) = \left(\frac{1}{2}\right)\left(\frac{5}{3}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{2}\right)\left(\frac{2}{3}\right)\left(\frac{5}{3}\right)\left(\frac{1}{3}\right)$

where the first term corresponds to initially visiting a female and the second term corresponds to initially visiting a male. Similarly,

p(2) = $\left(\frac{1}{2}\right)\left(\frac{2}{3}\right)^2\left(\frac{5}{3}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{2}\right)\left(\frac{2}{3}\right)^2\left(\frac{5}{3}\right)\left(\frac{1}{3}\right)$. In general,

p(z) = $\left(\frac{1}{2}\right)\left(\frac{2}{3}\right)^{2z-2}\left(\frac{5}{3}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{2}\right)\left(\frac{2}{3}\right)^{2z-2}\left(\frac{5}{3}\right)\left(\frac{1}{3}\right) = \left(\frac{24}{54}\right)\left(\frac{2}{3}\right)^{2z-2}$ for z = 1, 2, 3, …

26.

a. The sample space consists of all possible permutations of the four numbers 1, 2, 3, 4:

outcome	y value	outcome	y value	outcome	y value
1234	4	2314	1	3412	0
1243	2	2341	0	3421	0
1324	2	2413	0	4132	1
1342	1	2431	1	4123	0
1423	1	3124	1	4213	1
1432	2	3142	0	4231	2
2134	2	3214	2	4312	0
2143	0	3241	1	4321	0

b. Thus p(0) = P(Y = 0) = $\frac{9}{24}$, p(1) = P(Y = 1) = $\frac{8}{24}$, p(2) = P(Y = 2) = $\frac{6}{24}$,
p(3) = P(Y = 3) = 0, p(3) = P(Y = 3) = $\frac{1}{24}$.

27. If $x_1 < x_2$, $F(x_2) = P(X \le x_2) = P(\{X \le x_1\} \cup \{x_1 < X \le x_2\})$
$= P(X \le x_1) + P(x_1 < X \le x_2) \ge P(X \le x_1) = F(x_1)$.
$F(x_1) = F(x_2)$ when $P(x_1 < X \le x_2) = 0$.

Section 3.3

28.

a. $E(X) = \sum_{x=0}^{4} x \cdot p(x)$
$= (0)(.08) + (1)(.15) + (2)(.45) + (3)(.27) + (4)(.05) = 2.06$

b. $V(X) = \sum_{x=0}^{4} (x - 2.06)^2 \cdot p(x) = (0 - 2.06)^2(.08) + \ldots + (4 - 2.06)^2(.05)$
$= .339488 + .168540 + .001620 + .238572 + .188180 = .9364$

c. $\sigma_x = \sqrt{.9364} = .9677$

d. $V(X) = \left[\sum_{x=0}^{4} x^2 \cdot p(x)\right] - (2.06)^2 = 5.1800 - 4.2436 = .9364$

29.

a. $E(Y) = \sum_{x=0}^{4} y \cdot p(y) = (0)(.60) + (1)(.25) + (2)(.10) + (3)(.05) = .60$

b. $E(100Y^2) = \sum_{x=0}^{4} 100y^2 \cdot p(y) = (0)(.60) + (100)(.25)$
$+ (400)(.10) + (900)(.05) = 110$

30. $E(Y) = .60$;
$E(Y^2) = 1.1$
$V(Y) = E(Y^2) - [E(Y)]^2 = 1.1 - (.60)^2 = .74$
$\sigma_y = \sqrt{.74} = .8602$
$E(Y) \pm \sigma_y = .60 \pm .8602 = (-.2602, 1.4602)$ or $(0, 1)$.
$P(Y = 0) + P(Y = 1) = .85$

31.

a. $E(X) = (13.5)(.2) + (15.9)(.5) + (19.1)(.3) = 16.38$,
$E(X^2) = (13.5)^2(.2) + (15.9)^2(.5) + (19.1)^2(.3) = 272.298$,
$V(X) = 272.298 – (16.38)^2 = 3.9936$

b. $E(25X – 8.5) = 25\,E(X) – 8.5 = (25)(16.38) – 8.5 = 401$

c. $V(25X – 8.5) = V(25X) = (25)^2V(X) = (625)(3.9936) = 2496$

d. $E[h(X)] = E[X - .01X^2] = E(X) - .01E(X^2) = 16.38 – 2.72 = 13.66$

32.

a. $E(X^2) = \sum_{x=0}^{1} x^2 \cdot p(x) = (0^2)((1 – p) + (1^2)(p) = (1)(p) = p$

b. $V(X) = E(X^2) – [E(X)]^2 = p – p^2 = p(1 – p)$

c. $E(x^{79}) = (0^{79})(1 – p) + (1^{79})(p) = p$

33. $E(X) = \sum_{x=1}^{\infty} x \cdot p(x) = \sum_{x=1}^{\infty} x \cdot \frac{c}{x^3} = c\sum_{x=1}^{\infty} \frac{1}{x^2}$, but it is a well-known result from the theory of infinite series that $\sum_{x=1}^{\infty} \frac{1}{x^2} < \infty$, so E(X) is finite.

34. Let h(X) denote the net revenue (sales revenue – order cost) as a function of X. Then $h_3(X)$ and $h_4(X)$ are the net revenue for 3 and 4 copies purchased, respectively. For x = 1 or 2 , $h_3(X) = 2x – 3$, but at x = 3,4,5,6 the revenue plateaus. Following similar reasoning, $h_4(X) = 2x – 4$ for x=1,2,3, but plateaus at 4 for x = 4,5,6.

x	1	2	3	4	5	6
$h_3(x)$	-1	1	3	3	3	3
$h_4(x)$	-2	0	2	4	4	4
p(x)	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{4}{15}$	$\frac{3}{15}$	$\frac{2}{15}$

$E[h_3(X)] = \sum_{x=1}^{6} h_3(x) \cdot p(x) = (-1)(\frac{1}{15}) + \ldots + (3)(\frac{2}{15}) = 2.4667$

Similarly, $E[h_4(X)] = \sum_{x=1}^{6} h_4(x) \cdot p(x) = (-2)(\frac{1}{15}) + \ldots + (4)(\frac{2}{15}) = 2.6667$

Ordering 4 copies gives slightly higher revenue, on the average.

35.

P(x)	.8	.1	.08	.02
x	0	1,000	5,000	10,000
H(x)	0	500	4,500	9,500

E[h(X)] = 600. Premium should be $100 plus expected value of damage minus deductible or $700.

36. E(X) = $\sum_{x=1}^{n} x \cdot \left(\frac{1}{n}\right) = \left(\frac{1}{n}\right)\sum_{x=1}^{n} x = \frac{1}{n}\left[\frac{n(n+1)}{2}\right] = \frac{n+1}{2}$

E(X^2) = $\sum_{x=1}^{n} x^2 \cdot \left(\frac{1}{n}\right) = \left(\frac{1}{n}\right)\sum_{x=1}^{n} x^2 = \frac{1}{n}\left[\frac{n(n+1)(2n+1)}{6}\right] = \frac{(n+1)(2n+1)}{6}$

So V(X) = $\frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2-1}{12}$

37. E[h(X)] = E$\left(\frac{1}{X}\right) = \sum_{x=1}^{6}\left(\frac{1}{x}\right) \cdot p(x) = \frac{1}{6}\sum_{x=1}^{6}\frac{1}{x} = .408$, whereas $\frac{1}{3.5} = .286$, so you expect to win more if you gamble.

38. E(X) = $\sum_{x=1}^{4} x \cdot p(x)$ = 2.3, E(X^2) = 6.1, so V(X) = 6.1 – $(2.3)^2$ = .81

Each lot weighs 5 lbs, so weight left = 100 – 5x.
Thus the expected weight left is 100 – 5E(X) = 88.5,
and the variance of the weight left is
V(100 – 5X) = V(-5X) = 25V(x) = 20.25.

39.

a. The line graph of the p.m.f. of –X is just the line graph of the p.m.f. of X reflected about zero, but both have the same degree of spread about their respective means, suggesting V(-X) = V(X).

b. With a = -1, b = 0, V(aX + b) = V(-X) = a^2V(X).

40. V(aX + b) = $\sum_x [aX + b - E(aX+b)]^2 \cdot p(x) = \sum_x [aX + b - (am+b)]^2 p(x)$

$$= \sum_x [aX - (am)]^2 p(x) = a^2 \sum_x [X - m]^2 p(x) = a^2 V(X).$$

41.

a. E[X(X-1)] = $E(X^2)$ – E(X), $\Rightarrow E(X^2)$ = E[X(X-1)] + E(X) = 32.5

b. V(X) = 32.5 – $(5)^2$ = 7.5

c. V(X) = E[X(X-1)] + E(X) – $[E(X)]^2$

42. With a = 1 and b = c, E(X – c) = E(aX + b) = aE(X) + b = E(X) – c. When c = μ, E(X - μ) = E(X) - μ = μ - μ = 0, so the expected deviation from the mean is zero.

43.

a.

k	2	3	4	5	10
$\frac{1}{k^2}$	.25	.11	.06	.04	.01

b. $m = \sum_{x=0}^{6} x \cdot p(x) = 2.64$, $s^2 = \left[\sum_{x=0}^{6} x^2 \cdot p(x)\right] - m^2 = 2.37,\ s = 1.54$

Thus μ - 2σ = -.44, and μ + 2σ = 5.72,
so P(|x-μ| ≥ 2σ) = P(X is lat least 2 s.d.'s from μ)
= P(x is either ≤-.44 or ≥ 5.72) = P(X = 6) = .04.
Chebyshev's bound of .025 is much too conservative. For K = 3,4,5, and 10, P(|x-μ| ≥ kσ) = 0, here again pointing to the very conservative nature of the bound $\frac{1}{k^2}$.

c. μ = 0 and $s = \frac{1}{3}$, so P(|x-μ| ≥ 3σ) = P(| X | ≥ 1)
= P(X = -1 or +1) = $\frac{1}{18} + \frac{1}{18} = \frac{1}{9}$, identical to the upper bound.

d. Let p(-1) = $\frac{1}{50}$, $p(+1) = \frac{1}{50}$, $p(0) = \frac{24}{25}$.

Section 3.4

44.

a. $b(3;8,.6) = \binom{8}{3}(.6)^3(.4)^5$ = (56)(.00221184) = .124

b. $b(5;8,.6) = \binom{8}{5}(.6)^5(.4)^3$ = (56)(.00497664) = .279

c. $P(3 \le X \le 5) = b(3;8,.6) + b(4;8,.6) + b(5;8,.6) = .635$

d. $P(1 \le X) = 1 - P(X = 0) = 1 - \binom{12}{0}(.1)^0(.9)^{12} = 1 - (.9)^{12} = .718$

45.

a. $B(4;10,.3) = .850$

b. $b(4;10,.3) = B(4;10,.3) - B(3;10,.3) = .200$

c. $b(6;10,.7) = B(6;10,.7) - B(5;10,.7) = .200$

d. $P(2 \le X \le 4) = B(4;10,.3) - B(1;10,.3) = .701$

e. $P(2 < X) = 1 - P(X \le 1) = 1 - B(1;10,.3) = .851$

f. $P(X \le 1) = B(1;10,.7) = .0000$

g. $P(2 < X < 6) = P(3 \le X \le 5) = B(5;10,.3) - B(2;10,.3) = .570$

46. $X \sim Bin(25, .05)$

a. $P(X \le 2) = B(2;25,.05) = .873$

b. $P(X \ge 5) = 1 - P(X \le 4) = 1 - B(4;25,.05) = .1 - .993 = .007$

c. $P(1 \le X \le 4) = P(X \le 4) - P(X \le 0) = .993 - .277 = .716$

d. $P(X = 0) = P(X \le 0) = .277$

e. $E(X) = np = (25)(.05) = 1.25$
$V(X) = np(1 - p) = (25)(.05)(.95) = 1.1875$
$\sigma_x = 1.0897$

47. X ~ Bin(6, .10)

a. P(X = 1) = $\binom{n}{x}(p)^x(1-p)^{n-x}$ = $\binom{6}{1}(.1)^1(.9)^5 = .3543$

b. P(X ≥ 2) = 1 – [P(X = 0) + P(X = 1)].

From **a** , we know P(X = 1) = .3543, and P(X = 0) = $\binom{6}{0}(.1)^0(.9)^6 = .5314$.

Hence P(X ≥ 2) = 1 – [.3543 + .5314] = .1143

c. Either 4 or 5 goblets must be selected

i) Select 4 goblets with zero defects: P(X = 0) = $\binom{4}{0}(.1)^0(.9)^4 = .6561$.

ii) Select 4 goblets, one of which has a defect, and the 5th is good:

$$\left[\binom{4}{1}(.1)^1(.9)^3\right]\times .9 = .26244$$

So the desired probability is .6561 + .26244 = .91854

48. Let S = comes to a complete stop, so p = .25 , n = 20

a. P(X ≤ 6) = B(6;20,.25) = .786

b. P(X = 6) = b(6;20,.20) = B(6;20,.25) - B(5;20,.25) = .786 - .617 = .169

c. P(X ≥ 6) = 1 – P(X ≤ 5) = 1 - B(5;20,.25) = 1 - .617 = .383

d. E(X) = (20)(.25) = 5. We expect 5 of the next 20 to stop.

49. Let S = has at least one citation. Then p = .4, n = 15

a. If at least 10 have no citations (Failure), then at most 5 have had at least one (Success):

P(X ≤ 5) = B(5;15,.40) = .403

b. P(X ≤ 7) = B(7;15,.40) = .787

c. P(5 ≤ X ≤ 10) = P(X ≤ 10) – P(X ≤ 4) = .991 - .217 = .774

50. $X \sim Bin(10, .60)$

a. $P(X \geq 6) = 1 - P(X \leq 5) = 1 - B(5;20,.60) = 1 - .367 = .633$

b. $E(X) = np = (10)(.6) = 6$; $V(X) = np(1-p) = (10)(.6)(.4) = 2.4$;
$\sigma_x = 1.55$
$E(X) \pm \sigma_x = (4.45, 7.55)$.
We desire $P(5 \leq X \leq 7) = P(X \leq 7) - P(X \leq 4) = .833 - .166 = .667$

c. $P(3 \leq X \leq 7) = P(X \leq 7) - P(X \leq 2) = .833 - .012 = .821$

51. Let S represent a telephone that is submitted for service while under warranty and must be replaced. Then $p = P(S) = P(\text{replaced} \mid \text{submitted}) \cdot P(\text{submitted}) = (.40)(.20) = .08$. Thus X, the number among the company's 10 phones that must be replaced, has a binomial distribution with $n = 10$, $p = .08$, so $p(2) = P(X=2) = \binom{10}{2}(.08)^2(.92)^8 = .1478$

52. $X \sim Bin(25, .02)$

a. $P(X=1) = 25(.02)(.98)^{24} = .308$

b. $P(X=1) = 1 - P(X=0) = 1 - (.98)^{25} = 1 - .603 = .397$

c. $P(X=2) = 1 - P(X=1) = 1 - [.308 + .397]$

d. $\bar{x} = 25(.02) = .5$; $s = \sqrt{npq} = \sqrt{25(.02)(.98)} = \sqrt{.49} = .7$
$\bar{x} + 2s = .5 + 1.4 = 1.9$ So $P(0 = X = 1.9 = P(X=1) = .705$

e. $\dfrac{.5(4.5) + 24.5(3)}{25} = 3.03$ hours

53. X = the number of flashlights that work.
Let event B = {battery has acceptable voltage}.
Then P(flashlight works) = P(both batteries work) = P(B)P(B) = (.9)(.9) = .81 We must assume that the batteries' voltage levels are independent.
$X \sim Bin(10, .81)$. $P(X=9) = P(X=9) + P(X=10)$

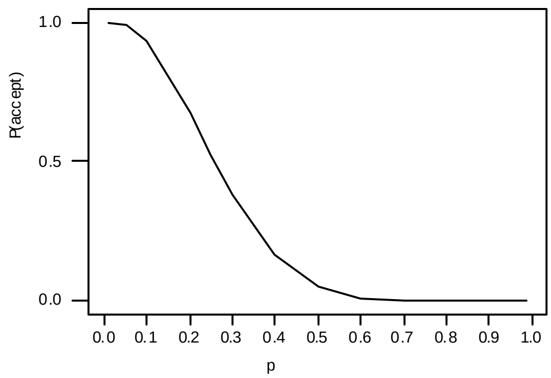
$$\binom{10}{9}(.81)^9(.19) + \binom{10}{10}(.81)^{10} = .285 + .122 = .407$$

54. Let p denote the actual proportion of defectives in the batch, and X denote the number of defectives in the sample.

a. P(the batch is accepted) = $P(X \le 2) = B(2;10,p)$

p	.01	.05	.10	.20	.25
P(accept)	1.00	.988	.930	.678	.526

b.



c. P(the batch is accepted) = $P(X \le 1) = B(1;10,p)$

p	.01	.05	.10	.20	.25
P(accept)	.996	.914	.736	.376	.244

d. P(the batch is accepted) = $P(X \le 2) = B(2;15,p)$

p	.01	.05	.10	.20	.25
P(accept)	1.00	.964	.816	.398	.236

e. We want a plan for which P(accept) is high for $p \le .1$ and low for $p > .1$
The plan in **d** seems most satisfactory in these respects.

55.

a. P(rejecting claim when p = .8) = B(15;25,.8) = .017

b. P(not rejecting claim when p = .7) = P(X ≥ 16 when p = .7)
= 1 - B(15;25,.7) = 1 - .189 = .811; for p = .6, this probability is
= 1 - B(15;25,.6) = 1 - .575 = .425.

c. The probability of rejecting the claim when p = .8 becomes B(14;25,.8) = .006, smaller than in **a** above. However, the probabilities of **b** above increase to .902 and .586, respectively.

56. h(x) = 1 · X + 2.25(25 – X) = 62.5 – 1.5X, so E(h(X)) = 62.5 – 1.5E(x)
= 62.5 – 1.5np – 62.5 – (1.5)(25)(.6) = $40.00

57. If topic A is chosen, when n = 2, P(at least half received)
= $P(X \geq 1) = 1 - P(X = 0) = 1 - (.1)^2 = .99$
If B is chosen, when n = 4, P(at least half received)
= $P(X \geq 2) = 1 - P(X \leq 1) = 1 - (0.1)^4 - 4(.1)^3(.9) = .9963$
Thus topic B should be chosen.
If p = .5, the probabilities are .75 for A and .6875 for B, so now A should be chosen.

58.

a. np(1 – p) = 0 if either p = 0 (whence every trial is a failure, so there is no variability in X) or if p = 1 (whence every trial is a success and again there is no variability in X)

b. $\frac{d}{dp}\left[np(1-p)\right]$ = n[(1 – p) + p(-1)] = n[1 – 2p = 0 ⇒ p = .5, which is easily seen to correspond to a maximum value of V(X).

59.

a. b(x; n, 1 – p) = $\binom{n}{x}(1-p)^x(p)^{n-x} = \binom{n}{n-x}(p)^{n-x}(1-p)^x$ = b(n-x; n, p)
Alternatively, P(x S's when P(S) = 1 – p) = P(n-x F's when P(F) = p), since the two events are identical), but the labels S and F are arbitrary so can be interchanged (if P(S) and P(F) are also interchanged), yielding P(n-x S's when P(S) = 1 – p) as desired.

b. B(x;n,1 – p) = P(at most x S's when P(S) = 1 – p)
= P(at least n-x F's when P(F) = p)
= P(at least n-x S's when P(S) = p)
= 1 – P(at most n-x-1 S's when P(S) = p)
= 1 – B(n-x-1;n,p)

c. Whenever p > .5, (1 – p) < .5 so probabilities involving X can be calculated using the results **a** and **b** in combination with tables giving probabilities only for p ≤ .5

60. Proof of E(X) = np:

$$E(X) = \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=1}^{n} x \cdot \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} = np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x}$$

$$= np \sum_{y=0}^{n} \frac{(n-1)!}{(y)!(n-1-y)!} p^y (1-p)^{n-1-y} \text{ (y replaces x-1)}$$

$$= np \left\{ \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y} \right\}$$

The expression in braces is the sum over all possible values y = 0, 1, 2, … , n-1 of a binomial p.m.f. based on n-1 trials, so equals 1, leaving only np, as desired.

61.

a. Although there are three payment methods, we are only concerned with S = uses a debit card and F = does not use a debit card. Thus we can use the binomial distribution. So n = 100 and p = .5. E(X) = np = 100(.5) = 50, and V(X) = 25.

b. With S = doesn't pay with cash, n = 100 and p = .7, E(X) = np = 100(.7) = 70, and V(X) = 21.

62.

a. Let X = the number with reservations who show, a binomial r.v. with n = 6 and p = .8. The desired probability is
P(X = 5 or 6) = b(5;6,.8) + b(6;6,.8) = .3932 + .2621 = .6553

b. Let h(X) = the number of available spaces. Then

When x is:	0	1	2	3	4	5	6
H(x) is:	4	3	2	1	0	0	0

$E[h(X)] = \sum_{x=0}^{6} h(x) \cdot b(x;6,.8)$ = 4(.000) + 3(.002) = 2(.015 + 3(.082) = .277

c. Possible X values are 0, 1, 2, 3, and 4. X = 0 if there are 3 reservations and none show or …or 6 reservations and none show, so
P(X = 0) = b(0;3,.8)(.1) + b(0;4,.8)(.2) + b(0;5,.8)(.3) + b(0;6,.8)(.4)
= .0080(.1) + .0016(.2) + .0003(.3) + .0001(.4) = .0013
P(X = 1) = b(1;3,.8)(.1) + … + b(1;6,.8)(.4) = .0172
P(X = 2) = .0906, P(X = 3) = .2273,
P(X = 4) = 1 – [.0013 + … + .2273] = .6636

63. When p = .5, μ = 10 and σ = 2.236, so 2σ = 4.472 and 3σ = 6.708.
The inequality $|X - 10| \geq 4.472$ is satisfied if either $X \leq 5$ or $X \geq 15$, or $P(|X - \mu| \geq 2\sigma) = P(X \leq 5 \text{ or } X \geq 15) = .021 + .021 = .042$.

In the case p = .75, μ = 15 and σ = 1.937, so 2σ = 3.874 and 3σ = 5.811. $P(|X - 15| \geq 3.874) = P(X \leq 11 \text{ or } X \geq 19) = .041 + .024 = .065$, whereas $P(|X - 15| \geq 5.811) = P(X \leq 9) = .004$. All these probabilities are considerably less than the upper bounds .25(for k = 2) and .11 (for k = 3) given by Chebyshev.

Section 3.5

64.

a. X ~ Hypergeometric N=15, n=5, M=6

b. P(X=2) = $\dfrac{\binom{6}{2}\binom{9}{3}}{\binom{15}{5}} = \dfrac{840}{3003} = .280$

P(X=2) = P(X=0) + P(X=1) + P(X=2)

$$= \frac{\binom{9}{5}}{\binom{15}{5}} + \frac{\binom{6}{1}\binom{9}{4}}{\binom{15}{5}} + \frac{840}{3003} = \frac{126 + 756 + 840}{3003} = \frac{1722}{3003} = .573$$

P(X=2) = 1 – P(X=1) = 1 – [P(X=0) + P(X=1)] = $1 - \dfrac{126 + 756}{3003} = .706$

c. E(X) = $5\left(\dfrac{6}{15}\right) = 2$; V(X) = $\left(\dfrac{15-5}{14}\right) \cdot 5 \cdot \left(\dfrac{6}{15}\right) \cdot \left(1 - \dfrac{6}{15}\right) = .857$;

$s = \sqrt{V(X)} = .926$

65. X~h(x; 6, 12, 7)

a. P(X=5) = $\dfrac{\binom{7}{5}\binom{5}{1}}{\binom{12}{6}} = \dfrac{105}{924} = .114$

b. P(X=4) = 1 – P(X=5) = 1 – [P(X=5) + P(X=6)] =

$$1-\left[\frac{\binom{7}{5}\binom{5}{1}}{\binom{12}{6}}+\frac{\binom{7}{6}}{\binom{12}{6}}\right]=1-\frac{105+7}{924}=1-.121=.879$$

c. E(X) = $\left(\dfrac{6\cdot 7}{12}\right)=3.5$; $s = \sqrt{\left(\frac{6}{11}\right)(6)\left(\frac{7}{12}\right)\left(\frac{5}{12}\right)} = \sqrt{.795} = .892$

P(X > 3.5 + .892) = P(X > 4.392) = P(X=5) = .121 (see part b)

d. We can approximate the hypergeometric distribution with the binomial if the population size and the number of successes are large: h(x;15,40,400) approaches b(x;15,.10). So P(X=5) ˜ B(5; 15, .10) from the binomial tables = .998

66.

a. P(X = 10) = h(10;15,30,50) = $\dfrac{\binom{30}{10}\binom{20}{5}}{\binom{50}{15}} = .2070$

b. P(X ≥ 10) = h(10;15,30,50) + h(11;15,30,50) + … + h(15;15,30,50)
= .2070+.1176+.0438+.0101+.0013+.0001 = .3799

c. P(at least 10 from the same class) = P(at least 10 from second class [answer from **b**]) + P(at least 10 from first class). But "at least 10 from 1^{st} class" is the same as "at most 5 from the second" or P(X ≤ 5).

P(X ≤ 5) = h(0;15,30,50) + h(1;15,30,50) + … + h(5;15,30,50)
= 11697+.002045+.000227+.000150+.000001+.000000
= .01412

So the desired probability = P(x ≥ 10) + P(X ≤ 5)
= .3799 + .01412 = .39402

d. E(X) = $n \cdot \frac{M}{N} = 15 \cdot \frac{30}{50} = 9$

V(X) = $\left(\frac{35}{49}\right)(9)\left(1 - \frac{30}{50}\right) = 2.5714$

$\sigma_x = 1.6036$

e. Let Y = 15 – X. Then E(Y) = 15 – E(X) = 15 – 9 = 6

V(Y) = V(15 – X) – V(X) = 2.5714, so $\sigma_Y = 1.6036$

67.

a. Possible values of X are 5, 6, 7, 8, 9, 10. (In order to have less than 5 of the granite, there would have to be more than 10 of the basaltic).

P(X = 5) = h(5; 15,10,20) = $\frac{\binom{10}{5}\binom{10}{10}}{\binom{20}{15}} = .0163$.

Following the same pattern for the other values, we arrive at the pmf, in table form below.

x	5	6	7	8	9	10
p(x)	.0163	.1354	.3483	.3483	.1354	.0163

b. P(all 10 of one kind or the other) = P(X = 5) + P(X = 10) = .0163 + .0163 = .0326

c. E(X) = $n \cdot \frac{M}{N} = 15 \cdot \frac{10}{20} = 7.5$; V(X) = $\left(\frac{5}{19}\right)(7.5)\left(1 - \frac{10}{20}\right) = .9868$;

$\sigma_x = .9934$

$\mu \pm \sigma = 7.5 \pm .9934 = (6.5066, 8.4934)$, so we want

P(X = 7) + P(X = 8) = .3483 + .3483 = .6966

68.

a. h(x; 6,4,11)

b. $6 \cdot \left(\frac{4}{11}\right) = 2.18$

69.

a. h(x; 10,10,20) (the successes here are the top 10 pairs, and a sample of 10 pairs is drawn from among the 20)

b. Let X = the number among the top 5 who play E-W. Then P(all of top 5 play the same direction) = P(X = 5) + P(X = 0) = h(5;10,5,20) + h(5;10,5,20)

$$= \frac{\binom{15}{5}}{\binom{20}{10}} + \frac{\binom{15}{10}}{\binom{20}{10}} = .033$$

c. N = 2n; M = n; n = n
h(x;n,n,2n)
E(X) = $n \cdot \frac{n}{2n} = \frac{1}{2}n$;
V(X) =

$$\left(\frac{2n-n}{2n-1}\right)\cdot n \cdot \frac{n}{2n} \cdot \left(1-\frac{n}{2n}\right) = \left(\frac{n}{2n-1}\right)\cdot \frac{n}{2} \cdot \left(1-\frac{n}{2n}\right) = \left(\frac{n}{2n-1}\right)\cdot \frac{n}{2} \cdot \left(\frac{1}{2}\right)$$

70.

a. h(x;10,15,50)

b. When N is large relative to n, h(x; n,M,N) $\approx b\left(x;n,\frac{M}{N}\right)$,

so h(x;10,150,500) $\approx b(x;10,.3)$

c. Using the hypergeometric model, E(X) = $10 \cdot \left(\frac{150}{500}\right) = 3$ and

V(X) = $\frac{490}{499}(10)(.3)(.7) = .982(2.1) = 2.06$
Using the binomial model, E(X) = (10)(.3) = 3, and
V(X) = 10(.3)(.7) = 2.1

71.

a. With S = a female child and F = a male child, let X = the number of F's before the 2^{nd} S. Then P(X = x) = nb(x;2, .5)

b. P(exactly 4 children) = P(exactly 2 males)
= nb(2;2,.5) = (3)(.0625) = .188

c. P(at most 4 children) = P(X ≤ 2)

$$= \sum_{x=0}^{2} nb(x;2,.5) = .25+2(.25)(.5) + 3(.0625) = .688$$

d. E(X) = $\frac{(2)(.5)}{.5} = 2$, so the expected number of children = E(X + 2)
= E(X) + 2 = 4

72. The only possible values of X are 3, 4, and 5.
p(3) = P(X = 3) = P(first 3 are B's or first 3 are G's) = $2(.5)^3$ = .250
p(4) = P(two among the 1^{st} three are B's and the 4th is a B) + P(two among the 1^{st} three are G's and the 4th is a G) = $2 \cdot \binom{3}{2}(.5)^4 = .375$
p(5) = 1 – p(3) – p(4) = .375

73. This is identical to an experiment in which a single family has children until exactly 6 females have been born(since p = .5 for each of the three families), so p(x) = nb(x;6,.5) and E(X) = 6 (= 2+2+2, the sum of the expected number of males born to each one.)

74. The interpretation of "roll" here is a pair of tosses of a single player's die(two tosses by A or two by B). With S = doubles on a particular roll, p = $\frac{1}{6}$. Furthermore, A and B are really identical (each die is fair), so we can equivalently imagine A rolling until 10 doubles appear. The P(x rolls) = P(9 doubles among the first x – 1 rolls and a double on the x^{th} roll =

$$\binom{x-1}{9}\left(\frac{5}{6}\right)^{x-10}\left(\frac{1}{6}\right)^{9} \cdot \left(\frac{1}{6}\right) = \binom{x-1}{9}\left(\frac{5}{6}\right)^{x-10}\left(\frac{1}{6}\right)^{10}$$

E(X) = $\frac{r(1-p)}{p} = \frac{10(\frac{5}{6})}{\frac{1}{6}} = 10(5) = 50$ V(X) = $\frac{r(1-p)}{p^2} = \frac{10(\frac{5}{6})}{(\frac{1}{6})^2} = 10(5)(6) = 300$

Section 3.6

75.

a. $P(X \le 8) = F(8;5) = .932$

b. $P(X = 8) = F(8;5) - F(7;5) = .065$

c. $P(X \ge 9) = 1 - P(X \le 8) = .068$

d. $P(5 \le X \le 8) = F(8;5) - F(4;5) = .492$

e. $P(5 < X < 8) = F(7;5) - F(5;5) = .867 - .616 = .251$

76.

a. $P(X \le 5) = F(5;8) = .191$

b. $P(6 \le X \le 9) = F(9;8) - F(5;8) = .526$

c. $P(X \ge 10) = 1 - P(X \le 9) = .283$

d. $E(X) = \lambda = 10$, $\sigma_X = \sqrt{l} = 2.83$, so $P(X > 12.83) = P(X \ge 13) = 1 - P(X \le 12) = 1 - .936 = .064$

77.

a. $P(X \le 10) = F(10;20) = .011$

b. $P(X > 20) = 1 - F(20;20) = 1 - .559 = .441$

c. $P(10 \le X \le 20) = F(20;20) - F(9;20) = .559 - .005 = .554$
$P(10 < X < 20) = F(19;20) - F(10;20) = .470 - .011 = .459$

d. $E(X) = \lambda = 20$, $\sigma_X = \sqrt{l} = 4.472$

$$\begin{aligned} P(\mu - 2\sigma < X < \mu + 2\sigma) &= P(20 - 8.944 < X < 20 + 8.944) \\ &= P(11.056 < X < 28.944) \\ &= P(X \le 28) - P(X \le 11) \\ &= F(28;20) - F(12;20)] \\ &= .966 - .021 = .945 \end{aligned}$$

78.

a. $P(X = 1) = F(1;2) - F(0;2) = .982 - .819 = .163$

b. $P(X \ge 2) = 1 - P(X \le 1) = 1 - F(1;2) = 1 - .982 = .018$

c. $P(1^{st} \text{ doesn't} \cap 2^{nd} \text{ doesn't}) = P(1^{st} \text{ doesn't}) \cdot P(2^{nd} \text{ doesn't})$
$= (.819)(.819) = .671$

79. $p = \frac{1}{200}$; n = 1000; λ = np = 5

a. $P(5 \le X \le 8) = F(8;5) - F(4;5) = .492$

b. $P(X \ge 8) = 1 - P(X \le 7) = 1 - .867 = .133$

80.

a. The experiment is binomial with n = 10,000 and p = .001,
so $\mu = np = 10$ and $\sigma = \sqrt{npq} = 3.161$.

b. X has approximately a Poisson distribution with λ = 10,
so $P(X > 10) \approx 1 - F(10;10) = 1 - .583 = .417$

c. $P(X = 0) \approx 0$

81.

a. λ = 8 when t = 1, so $P(X = 6) = F(6;8) - F(5;8) = .313 - .191 = .122$,
$P(X \ge 6) = 1 - F(5;8) = .809$, and $P(X \ge 10) = 1 - F(9;8) = .283$

b. t = 90 min = 1.5 hours, so λ = 12; thus the expected number of arrivals is 12 and the SD $= \sqrt{12} = 3.464$

c. t = 2.5 hours implies that λ = 20; in this case, $P(X \ge 20) = 1 - F(19;20) = .530$ and $P(X \le 10) = F(10;20) = .011$.

82.

a. $P(X = 4) = F(4;5) - F(3;5) = .440 - .265 = .175$

b. $P(X \ge 4) = 1 - P(X \le 3) = 1 - .265 = .735$

c. Arrivals occur at the rate of 5 per hour, so for a 45 minute period the rate is λ = (5)(.75) = 3.75, which is also the expected number of arrivals in a 45 minute period.

83.

a. For a two hour period the parameter of the distribution is λt = (4)(2) = 8,
so $P(X = 10) = F(10;8) - F(9;8) = .099$.

b. For a 30 minute period, λt = (4)(.5) = 2, so $P(X = 0) = F(0,2) = .135$

c. $E(X) = \lambda t = 2$

84. Let X = the number of diodes on a board that fail.

a. E(X) = np = (200)(.01) = 2, V(X) = npq = (200)(.01)(.99) = 1.98, σ_X = 1.407

b. X has approximately a Poisson distribution with λ = np = 2,
so $P(X \ge 4) = 1 - P(X \le 3) = 1 - F(3;2) = 1 - .857 = .143$

c. P(board works properly) = P(all diodes work) = P(X = 0) = F(0;2) = .135
Let Y = the number among the five boards that work, a binomial r.v. with n = 5 and p = .135. Then $P(Y \ge 4) = P(Y = 4) + P(Y = 5) =$

$$\binom{5}{4}(.135)^4(.865) + \binom{5}{5}(.135)^5(.865)^0 = .00144 + .00004 = .00148$$

85. α = 1/(mean time between occurrences) = $\frac{1}{.5} = 2$

a. $\alpha t = (2)(2) = 4$

b. $P(X > 5)\ 1 - P(X \le 5) = 1 - .785 = .215$

c. Solve for t, given $\alpha = 2$:

$.1 = e^{-\alpha t}$

$\ln(.1) = -\alpha t$

$t = \frac{2.3026}{2} \approx 1.15$ years

86. $$E(X) = \sum_{x=0}^{\infty} x\frac{e^{-l}\,l^x}{x!} = \sum_{x=1}^{\infty} x\frac{e^{-l}\,l^x}{x!} = l\sum_{x=1}^{\infty} x\frac{e^{-l}\,l^x}{x!} = l\sum_{y=0}^{\infty} x\frac{e^{-l}\,l^y}{y!} = l$$

87.

a. For a one-quarter acre plot, the parameter is (80)(.25) = 20,
so $P(X \le 16) = F(16;20) = .221$

b. The expected number of trees is $\lambda \cdot$(area) = 80(85,000) = 6,800,000.

c. The area of the circle is πr^2 = .031416 sq. miles or 20.106 acres. Thus X has a Poisson distribution with parameter 20.106

88.

a. P(X = 10 and no violations) = P(no violations | X = 10) · P(X = 10)
= $(.5)^{10}$ · [F(10;10) – F(9;10)]
= (.000977)(.125) = .000122

b. P(y arrive and exactly 10 have no violations)
= P(exactly 10 have no violations | y arrive) · P(y arrive)

= P(10 successes in y trials when p = .5) · $e^{-10}\dfrac{(10)^y}{y!}$

$$= \binom{y}{10}(.5)^{10}(.5)^{y-10}e^{-10}\frac{(10)^y}{y!} = \frac{e^{-10}(5)^y}{10!(y-10)!}$$

c. P(exactly 10 without a violation) = $\displaystyle\sum_{y=10}^{\infty}\frac{e^{-10}(5)^y}{10!(y-10)!}$

$$= \frac{e^{-10}\cdot 5^{10}}{10!}\sum_{y=10}^{\infty}\frac{(5)^{y-10}}{(y-10)!} = \frac{e^{-10}\cdot 5^{10}}{10!}\sum_{u=0}^{\infty}\frac{(5)^u}{(u)!} = \frac{e^{-10}\cdot 5^{10}}{10!}\cdot e^5$$

$$= \frac{e^{-5}\cdot 5^{10}}{10!} = p(10;5).$$

In fact, generalizing this argument shows that the number of “no-violation” arrivals within the hour has a Poisson distribution with parameter 5; the 5 results from λp = 10(.5).

89.

a. No events in (0, t+Δt) if and only if no events in (o, t) and no events in (t, t+Δt). Thus, $P_0(t+\Delta t) = P_0(t)$ ·P(no events in (t, t+Δt))
= $P_0(t)[1 - \lambda\cdot\Delta t - o(\Delta t)]$

b. $$\frac{P_0(t+\Delta t)-P_0(t)}{\Delta t} = -l P_0(t)\frac{\Delta' t}{\Delta' t} - P_0(t)\cdot\frac{o(\Delta t)}{\Delta t}$$

c. $\dfrac{d}{dt}\left[e^{-lt}\right] = -\lambda e^{-\lambda t} = -\lambda P_0(t)$, as desired.

d. $$\frac{d}{dt}\left[\frac{e^{-lt}(lt)^k}{k!}\right] = \frac{-le^{-lt}(lt)^k}{k!} + \frac{kle^{-lt}(lt)^{k-1}}{k!}$$

$$= -l\frac{e^{-lt}(lt)^k}{k!} + l\frac{e^{-lt}(lt)^{k-1}}{(k-1)!} = -\lambda P_k(t) + \lambda P_{k-1}(t) \text{ as desired.}$$

Supplementary Exercises

90. Outcomes are(1,2,3)(1,2,4) (1,2,5) … (5,6,7); there are 35 such outcomes. Each having probability $\frac{1}{35}$. The W values for these outcomes are 6 (=1+2+3), 7, 8, …, 18. Since there is just one outcome with W value 6, p(6) = P(W = 6) = $\frac{1}{35}$. Similarly, there are three outcomes with W value 9 [(1,2,6) (1,3,5) and 2,3,4)], so p(9) = $\frac{3}{35}$. Continuing in this manner yields the following distribution:

W	6	7	8	9	10	11	12	13	14	15	16	17	18
P(W)	$\frac{1}{35}$	$\frac{1}{35}$	$\frac{2}{35}$	$\frac{3}{35}$	$\frac{4}{35}$	$\frac{4}{35}$	$\frac{5}{35}$	$\frac{4}{35}$	$\frac{4}{35}$	$\frac{3}{35}$	$\frac{2}{35}$	$\frac{1}{35}$	$\frac{1}{35}$

Since the distribution is symmetric about 12, μ = 12, and $s^2 = \sum_{w=6}^{18}(w-12)^2 p(w)$

= $\frac{1}{35}$ $[(6)^2(1) + (5)^2(1) + \ldots + (5)^2(1) + (6)^2(1) = 8$

91.

a. p(1) = P(exactly one suit) = P(all spades) + P(all hearts) + P(all diamonds)

+ P(all clubs) = 4P(all spades) = $4 \cdot \dfrac{\binom{13}{5}}{\binom{52}{5}} = .00198$

p(2) = P(all hearts and spades with at least one of each) + …+ P(all diamonds and clubs with at least one of each)
= 6 P(all hearts and spades with at least one of each)
= 6 [P(1 h and 4 s) + P(2 h and 3 s) + P(3 h and 2 s) + P(4 h and 1 s)]

$$= 6 \cdot \left[2 \cdot \frac{\binom{13}{4}\binom{13}{1}}{\binom{52}{5}} + 2 \cdot \frac{\binom{13}{3}\binom{13}{2}}{\binom{52}{5}} \right] = 6\left[\frac{18{,}590 + 44{,}616}{2{,}598{,}960}\right] = .14592$$

p(4) = 4P(2 spades, 1 h, 1 d, 1 c) = $\dfrac{4 \cdot \binom{13}{2}(13)(13)(13)}{\binom{52}{5}} = .26375$

p(3) = 1 – [p(1) + p(2) + p(4)] = .58835

b. $\mu = \sum_{x=1}^{4} x \cdot p(x) = 3.114,\ s^2 = \left[\sum_{x=1}^{4} x^2 \cdot p(x)\right] - (3.114)^2 = .405,\ s = .636$

92. p(y) = P(Y = y) = P(y trials to achieve r S's) = P(y-r F's before rth S)
= nb(y – r;r,p) = $\binom{y-1}{r-1} p^r (1-p)^{y-r}$, y = r, r+1, r+2, …

93.

a. b(x;15,.75)

b. P(X > 10) = 1 - B(9;15, .75) = 1 - .148

c. B(10;15, .75) - B(5;15, .75) = .314 - .001 = .313

d. μ = (15)(.75) = 11.75, σ^2 = (15)(.75)(.25) = 2.81

e. Requests can all be met if and only if X ≤ 10, and 15 – X ≤ 8, i.e. if 7 ≤ X ≤ 10, so P(all requests met) = B(10; 15,.75) - B(6; 15,.75) = .310

94. P(6-v light works) = P(at least one 6-v battery works) = 1 – P(neither works)
= 1 –$(1 – p)^2$. P(D light works) = P(at least 2 d batteries work) = 1 – P(at most 1 D battery works) = 1 – [$(1 – p)^4$ + 4$(1 – p)^3$]. The 6-v should be taken if 1 –$(1 – p)^2$ ≥ 1 – [$(1 – p)^4$ + 4$(1 – p)^3$].
Simplifying, $1 \le (1-p)^2 + 4p(1-p) \Rightarrow 0 \le 2p - 3p^3 \Rightarrow p \le \frac{2}{3}$.

95. Let X ~ Bin(5, .9). Then P(X ≥ 3) = 1 – P(X ≤ 2) = 1 – B(2;5,.9) = .991

96.

a. P(X ≥ 5) = 1 - B(4;25,.05) = .007

b. P(X ≥ 5) = 1 - B(4;25,.10) = .098

c. P(X ≥ 5) = 1 - B(4;25,.20) = .579

d. All would decrease, which is bad if the % defective is large and good if the % is small.

97.

a. N = 500, p = .005, so np = 2.5 and b(x; 500, .005) $\approx$ p(x; 2.5), a Poisson p.m.f.

b. P(X = 5) = p(5; 2.5) - p(4; 2.5) = .9580 - .8912 = .0668

c. P(X ≥ 5) = 1 – p(4;2.5) = 1 - .8912 = .1088

98. $X \sim B(x; 25, p)$.

a. $B(18; 25, .5) – B(6; 25, .5) = .986$

b. $B(18; 25, .8) – B(6; 25, .8) = .220$

c. With p = .5, P(rejecting the claim) = $P(X \le 7) + P(X \ge 18) = .022 + [1 - .978] = .022 + .022 = .044$

d. The claim will not be rejected when $8 \le X \le 17$.
With p=.6, $P(8 \le X \le 17) = B(17;25,.6) – B(7;25,.6) = .846 - .001 = .845$.
With p=.8, $P(8 \le X \le 17) = B(17;25,.8) – B(7;25,.8) = .109 - .000 = .109$.

e. We want P(rejecting the claim) = .01. Using the decision rule "reject if X = 6 or X ≥ 19" gives the probability .014, which is too large. We should use "reject if X = 5 or X ≥ 20" which yields P(rejecting the claim) = .002 + .002 = .004.

99. Let Y denote the number of tests carried out. For n = 3, possible Y values are 1 and 4. P(Y = 1) = P(no one has the disease) = $(.9)^3 = .729$ and P(Y = 4) = .271, so E(Y) = (1)(.729) + (4)(.271) = 1.813, as contrasted with the 3 tests necessary without group testing.

100. Regard any particular symbol being received as constituting a trial. Then p = P(S) = P(symbol is sent correctly or is sent incorrectly and subsequently corrected) = $1 – p_1 + p_1p_2$. The block of n symbols gives a binomial experiment with n trials and $p = 1 – p_1 + p_1p_2$.

101. $p(2) = P(X = 2) = P(S \text{ on \#1 and S on \#2}) = p^2$
$p(3) = P(S \text{ on \#3 and S on \#2 and F on \#1}) = (1 – p)p^2$
$p(4) = P(S \text{ on \#4 and S on \#3 and F on \#2}) = (1 – p)p^2$
p(5) = P(S on #5 and S on #4 and F on #3 and no 2 consecutive S's on trials prior to #3) = $[\,1 – p(2)\,](1 – p)p^2$
p(6) = P(S on #6 and S on #5 and F on #4 and no 2 consecutive S's on trials prior to #4) = $[\,1 – p(2) – p(3)](1 – p)p^2$
In general, for x = 5, 6, 7, …: $p(x) = [\,1 – p(2) - \dots – p(x - 3)](1 – p)p^2$
For p = .9,

x	2	3	4	5	6	7	8
p(x)	.81	.081	.081	.0154	.0088	.0023	.0010

So $P(X \le 8) = p(2) + \dots + p(8) = .9995$

102.

a. With X ~ Bin(25, .1),$P(2 \le X \le 6) = B(6;25,.1 – B(1;25,.1) = .991 - .271 = 720$

b. $E(X) = np = 25(.1) = 2.5$, $\sigma_X = \sqrt{npq} = \sqrt{25(.1)(.9)} = \sqrt{2.25} = 1.50$

c. $P(X \ge 7 \text{ when } p = .1) = 1 – B(6;25,.1) = 1 - .991 = .009$

d. $P(X \le 6 \text{ when } p = .2) = B(6;25,.2) = = .780$, which is quite large

103.

a. Let event C = seed carries single spikelets, and event P = seed produces ears with single spikelets. Then P(P ∩ C) = P(P | C) · P(C) = .29 (.40) = .116. Let X = the number of seeds out of the 10 selected that meet the condition P ∩ C. Then X ~ Bin(10, .116).

P(X = 5) = $\binom{10}{5}(.116)^5(.884)^5 = .002857$

b. For 1 seed, the event of interest is P = seed produces ears with single spikelets.
P(P) = P(P ∩ C) + P(P ∩ C′) = .116 (from **a**) + P(P | C′) · P(C′)
= .116 + (.26)(.40) = .272.
Let Y = the number out of the 10 seeds that meet condition P.
Then Y ~ Bin(10, .272), and P(Y = 5) = .0767.
P(Y ≤ 5) = b(0;10,.272) + … + b(5;10,.272) = .041813 + … + .076719 = .97024

104. With S = favored acquittal, the population size is N = 12, the number of population S's is M = 4, the sample size is n = 4, and the p.m.f. of the number of interviewed jurors who favor acquittal is the hypergeometric p.m.f. h(x;4,4,12). E(X) = $4 \cdot \left(\frac{4}{12}\right) = 1.33$

105.

a. P(X = 0) = F(0;2) 0.135

b. Let S = an operator who receives no requests. Then p = .135 and we wish P(4 S's in 5 trials) = b(4;5,..135) = $\binom{5}{4}(.135)^4(.884)^1 = .00144$

c. P(all receive x) = P(first receives x) · … · P(fifth receives x) = $\left[\frac{e^{-2}2^x}{x!}\right]^5$, and P(all receive the same number) is the sum from x = 0 to ∞.

106. P(at least one) = 1 – P(none) = 1 - $e^{-lpR^2} \cdot \frac{(lpR^2)^0}{0!}$ = 1 - e^{-lpR^2} = .99 ⇒ e^{-lpR^2} = .01

$$\Rightarrow R^2 = \frac{-1n(.01)}{lp} = .7329 \Rightarrow R = .8561$$

107. The number sold is min (X, 5), so E[min(x, 5)] = $\sum^{\infty} \min(x,5)\,p(x;4)$

= (0)p(0;4) + (1) p(1;4) + (2) p(2;4) + (3) p(3;4) + (4) p(4;4) + $5\sum_{x=5}^{\infty} p(x;4)$

= 1.735 + 5[1 – F(4;4)] = 3.59

108.

a. P(X = x) = P(A wins in x games) + P(B wins in x games)
= P(9 S's in 1st x-1 $\cap$ S on the xth) + P(9 F's in 1st x-1 $\cap$ F on the xth)

$$= \binom{x-1}{9} p^9 (1-p)^{x-10}\, p + \binom{x-1}{9} (1-p)^9\, p^{x-10} (1-p)$$

$$= \binom{x-1}{9} \left[p^{10} (1-p)^{x-10} + (1-p)^{10} p^{x-10} \right]$$

b. Possible values of X are now 10, 11, 12, …(all positive integers $\geq$ 10). Now

$$P(X = x) = \binom{x-1}{9} \left[p^{10} (1-p)^{x-10} + q^{10} (1-q)^{x-10} \right] \text{ for } x = 10, \ldots, 19,$$

So $P(X \geq 20) = 1 - P(X < 20)$ and $P(X < 20) = \sum_{x=10}^{19} P(X = x)$

109.

a. No; probability of success is not the same for all tests

b. There are four ways exactly three could have positive results. Let D represent those with the disease and D′ represent those without the disease.

Combination D	Combination D′	Probability
0	3	$\left[\binom{5}{0}(.2)^0(.8)^5\right]\cdot\left[\binom{5}{3}(.9)^3(.1)^2\right]$ =(.32768)(.0729) = .02389
1	2	$\left[\binom{5}{1}(.2)^1(.8)^4\right]\cdot\left[\binom{5}{2}(.9)2(.1)^3\right]$ =(.4096)(.0081) = .00332
2	1	$\left[\binom{5}{2}(.2)^2(.8)^3\right]\cdot\left[\binom{5}{1}(.9)^1(.1)^4\right]$ =(.2048)(.00045) = .00009216
3	0	$\left[\binom{5}{3}(.2)^3(.8)^2\right]\cdot\left[\binom{5}{0}(.9)^0(.1)^5\right]$ =(.0512)(.00001) = .000000512

Adding up the probabilities associated with the four combinations yields 0.0273.

110. k(r,x) = $\dfrac{(x+r-1)(x+r-2)...(x+r-x)}{x!}$

With r = 2.5 and p = .3, p(4) = $\dfrac{(5.5)(4.5)(3.5)(2.5)}{4!}(.3)^{2.5}(.7)^{4} = .1068$

Using k(r,0) = 1, P(X ≥ 1) = 1 – p(0) = 1 – $(.3)^{2.5}$ = .9507

111.

a. p(x; λ,μ) = $\frac{1}{2}p(x;\,l\,)+\frac{1}{2}p(x;m)$ where both p(x;λ) and p(x; μ) are Poisson p.m.f.'s and thus ≥ 0, so p(x; λ,μ) ≥ 0. Further,

$$\sum_{x=0}^{\infty} p(x;\,l\,,m) = \frac{1}{2}\sum_{x=0}^{\infty} p(x;\,l\,) + \frac{1}{2}\sum_{x=0}^{\infty} p(x;m) = \frac{1}{2}+\frac{1}{2} = 1$$

b. $.6\,p(x;\,l\,)+.4\,p(x;m)$

c. E(X) = $\sum_{x=0}^{\infty} x[\frac{1}{2}p(x;\,l\,)+\frac{1}{2}p(x;m)] = \frac{1}{2}\sum_{x=0}^{\infty} xp(x;\,l\,)+\frac{1}{2}\sum_{x=0}^{\infty} xp(x;m)$

= $\frac{1}{2}\,l\,+\frac{1}{2}m = \frac{l+m}{2}$

d. E(X^2) = $\frac{1}{2}\sum_{x=0}^{\infty} x^2 p(x;\,l\,)+\frac{1}{2}\sum_{x=0}^{\infty} x^2 p(x;m) = \frac{1}{2}(\,l^2+l\,)+\frac{1}{2}(m^2+m)$ (since for a Poisson r.v., E(X^2) = V(X) + [E(X)]^2 = λ + λ^2),

so V(X) = $\frac{1}{2}\left[l^2+l+m^2+m\right]-\left[\frac{l+m}{2}\right]^2 = \left(\frac{l-m}{2}\right)^2+\frac{l+m}{2}$

112.

a. $\dfrac{b(x+1;n,p)}{b(x;n,p)} = \dfrac{(n-x)}{(x+1)}\cdot\dfrac{p}{(1-p)} > 1$ if np – (1 – p) > x, from which the stated conclusion follows.

b. $\dfrac{p(x+1;\,l\,)}{p(x;\,l\,)} = \dfrac{l}{(x+1)} > 1$ if x < λ - 1 , from which the stated conclusion follows. If λ is an integer, then λ - 1 is a mode, but p(λ,λ) = p(1 - λ, λ) so λ is also a mode[p(x; λ)] achieves its maximum for both x = λ - 1 and x = λ.

113. P(X = j) = $\sum_{i=1}^{10} P$ (arm on track $i \cap$ X = j) = $\sum_{i=1}^{10} P$ (X = j | arm on i) · p_i

= $\sum_{i=1}^{10} P$ (next seek at I+j+1 or I-j-1) · p_i = $\sum_{i=1}^{10}(p_{i+j+1} + P_{i-j-1})\,p_i$

where p_k = 0 if k < 0 or k > 10.

114. E(X) = $\sum_{x=0}^{n} x \cdot \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \sum_{x=1}^{n} \dfrac{\dfrac{M!}{(x-1)!(M-x)!} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}$

$$n \cdot \frac{M}{N} \sum_{x=1}^{n} \binom{M-1}{x-1} \frac{\binom{N-M}{n-x}}{\binom{N-1}{n-1}} = n \cdot \frac{M}{N} \sum_{y=0}^{n-1} \binom{M-1}{y} \frac{\binom{N-1-(M-1)}{n-1-y}}{\binom{N-1}{n-1}}$$

$$n \cdot \frac{M}{N} \sum_{y=0}^{n-1} h(y; n-1, M-1, N-1) = n \cdot \frac{M}{N}$$

115. Let A = {x: |x - μ| ≥ kσ}. Then σ^2 = $\sum_A (x-m)^2 p(x) \geq (ks)^2 \sum_A p(x)$. But $\sum_A p(x)$ = P(X is in A) = P(|X - μ| ≥ kσ), so $\sigma^2 \geq k^2\sigma^2 \cdot$ P(|X - μ| ≥ kσ), as desired.

116.

a. For [0,4], λ = $\int_0^4 e^{2+.6t}dt$ = 123.44, whereas for [2,6], λ = $\int_2^6 e^{2+.6t}dt$ = 409.82

b. λ = $\int_0^{0.9907} e^{2+.6t}dt$ = 9.9996 ≈ 10, so the desired probability is F(15, 10) = .951.

CHAPTER 4

Section 4.1

1.

a. P(x ≤ 1) = $\int_{-\infty}^{1} f(x)dx = \int_0^1 \frac{1}{2}xdx = \frac{1}{4}x^2\Big]_0^1 = .25$

b. P(.5 ≤ X ≤ 1.5) = $\int_{.5}^{1.5} \frac{1}{2}xdx = \frac{1}{4}x^2\Big]_{.5}^{1.5} = .5$

c. P(x > 1.5) = $\int_{1.5}^{\infty} f(x)dx = \int_{1.5}^{2} \frac{1}{2}xdx = \frac{1}{4}x^2\Big]_{1.5}^{2} = \frac{7}{16} \approx .438$

2. F(x) = $\frac{1}{10}$ for $-5 \le x \le 5$, and = 0 otherwise

a. P(X < 0) = $\int_{-5}^{0} \frac{1}{10}dx = .5$

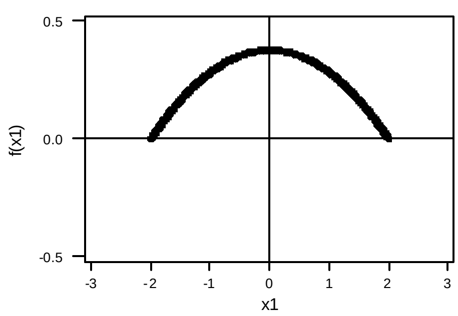
b. P(-2.5 < X < 2.5) = $\int_{-2.5}^{2.5} \frac{1}{10}dx = .5$

c. P(-2 ≤ X ≤ 3) = $\int_{-2}^{3} \frac{1}{10}dx = .5$

d. P(k < X < k + 4) = $\int_{k}^{k+4} \frac{1}{10}dx = \frac{x}{10}\Big]_k^{k+4} = \frac{1}{10}[(k+4)-k] = .4$

3.

a. Graph of f(x) = .09375(4 – x^2)



b. P(X > 0) = $\int_0^2 .09375(4 - x^2)dx = .09375(4x - \frac{x^3}{3})\Big]_0^2 = .5$

c. P(-1 < X < 1) = $\int_{-1}^1 .09375(4 - x^2)dx = .6875$

d. P(x < -.5 OR x > .5) = 1 – P(-.5 ≤ X ≤ .5) = 1 - $\int_{-.5}^{.5} .09375(4 - x^2)dx$
= 1 - .3672 = .6328

4.

a. $\int_{-\infty}^{\infty} f(x;q)dx = \int_0^{\infty} \frac{x}{q^2} e^{-x^2/2q^2} dx = -e^{-x^2/2q^2}\Big]_0^{\infty} = 0 - (-1) = 1$

b. P(X ≤ 200) = $\int_{-\infty}^{200} f(x;q)dx = \int_0^{200} \frac{x}{q^2} e^{-x^2/2q^2} dx$
$= -e^{-x^2/2q^2}\Big]_0^{200} \approx -.1353 + 1 = .8647$
P(X < 200) = P(X ≤ 200) ≈ .8647, since x is continuous.
P(X ≥ 200) = 1 - P(X ≤ 200) ≈ .1353

c. P(100 ≤ X ≤ 200) = $\int_{100}^{200} f(x;q)dx = -e^{-x^2/20,000}\Big]_{100}^{200} \approx .4712$

d. For x > 0, P(X ≤ x) =
$\int_{-\infty}^{x} f(y;q)dy = \int_0^{x} \frac{y}{e^2} e^{-y^2/2q^2} dx = -e^{-y^2/2q^2}\Big]_0^{x} = 1 - e^{-x^2/2q^2}$

5.

a. 1 = $\int_{-\infty}^{\infty} f(x)dx = \int_0^2 kx^2 dx = k\left(\frac{x^3}{3}\right)\Big]_0^2 = k\left(\frac{8}{3}\right) \Rightarrow k = \frac{3}{8}$

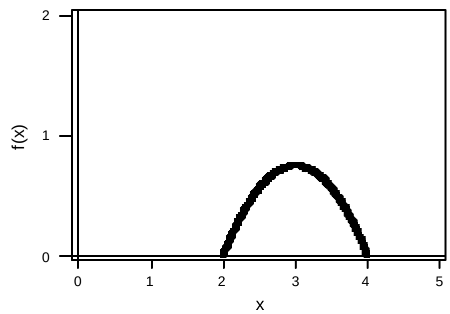
b. P(0 ≤ X ≤ 1) = $\int_0^1 \frac{3}{8}x^2 dx = \frac{1}{8}x^3\Big]_0^1 = \frac{1}{8} = .125$

c. P(1 ≤ X ≤ 1.5) = $\int_1^{1.5} \frac{3}{8}x^2 dx = \frac{1}{8}x^3\Big]_1^{1.5} = \frac{1}{8}\left(\frac{3}{2}\right)^3 - \frac{1}{8}(1)^3 = \frac{19}{64} \approx .2969$

d. P(X ≥ 1.5) = 1 - $\int_0^{1.5} \frac{3}{8}x^2 dx = \frac{1}{8}x^3\Big]_0^{1.5} = 1 - \left[\frac{1}{8}\left(\frac{3}{2}\right)^3 - 0\right] = 1 - \frac{27}{64} = \frac{37}{64} \approx .5781$

6.

a.



b. 1 = $\int_2^4 k[1-(x-3)^2]dx = \int_{-1}^1 k[1-u^2]du = \frac{4}{3} \Rightarrow k = \frac{3}{4}$

c. P(X > 3) = $\int_3^4 \frac{3}{4}[1-(x-3)^2]dx = .5$ by symmetry of the p.d.f

d. $P\left(\frac{11}{4} \le X \le \frac{13}{4}\right) = \int_{11/4}^{13/4} \frac{3}{4}[1-(x-3)^2]dx = \frac{3}{4}\int_{-1/4}^{1/4}[1-(u)^2]du = \frac{47}{128} \approx .367$

e. P(|X-3| > .5) = 1 – P(|X-3| ≤ .5) = 1 – P(2.5 ≤ X ≤ 3.5)

= 1 - $\int_{-.5}^{.5} \frac{3}{4}[1-(u)^2]du = \frac{5}{16} \approx .313$

7.

a. f(x) = $\frac{1}{10}$ for 25 ≤ x ≤ 35 and = 0 otherwise

b. P(X > 33) = $\int_{33}^{35} \frac{1}{10} dx = .2$

c. E(X) = $\int_{25}^{35} x \cdot \frac{1}{10} dx = \left.\frac{x^2}{20}\right]_{25}^{35} = 30$

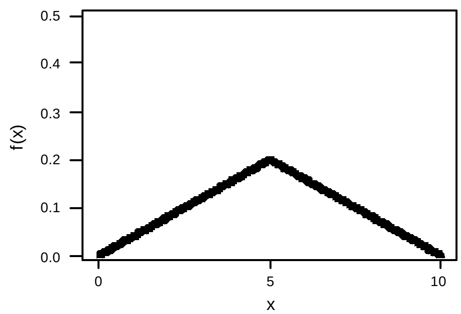
30 ± 2 is from 28 to 32 minutes:

P(28 < X < 32) = $\int_{28}^{32} \frac{1}{10} dx = \left.\frac{1}{10}x\right]_{28}^{32} = .4$

d. P(a ≤ x ≤ a+2) = $\int_a^{a+2} \frac{1}{10} dx = .2$, since the interval has length 2.

8.

a.



b. $\int_{-\infty}^{\infty} f(y)dy = \int_0^5 \frac{1}{25} y dy + \int_5^{10} (\frac{2}{5} - \frac{1}{25} y)dy = \left.\frac{y^2}{50}\right]_0^5 + \left.\left(\frac{2}{5}y - \frac{1}{50}y^2\right)\right]_5^{10}$

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

c. P(Y ≤ 3) = $\int_0^3 \frac{1}{25} y dy = \left.\frac{y^2}{50}\right]_0^5 = \frac{9}{50} \approx .18$

d. P(Y ≤ 8) = $= \int_0^5 \frac{1}{25} y dy + \int_5^8 (\frac{2}{5} - \frac{1}{25} y)dy = \frac{23}{25} \approx .92$

e. P(3 ≤ Y ≤ 8) = P(Y ≤ 8) - P(Y < 3) = $\frac{46}{50} - \frac{9}{50} = \frac{37}{50} = .74$

f. P(Y < 2 or Y > 6) = $= \int_0^3 \frac{1}{25} y dy + \int_6^{10} (\frac{2}{5} - \frac{1}{25} y)dy = \frac{2}{5} = .4$

9.

a. P(X ≤ 6) = $= \int_{.5}^6 .15 e^{-.15(x-5)} dx = .15 \int_0^{5.5} e^{-.15u} du$ (after u = x - .5)

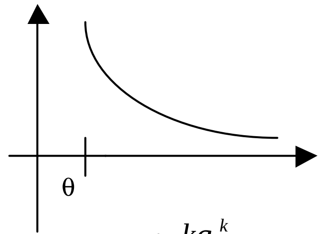
$= \left. e^{-.15u} \right]_0^{5.5} = 1 - e^{-.825} \approx .562$

b. 1 - .562 = .438; .438

c. P(5 ≤ Y ≤ 6) = P(Y ≤ 6) - P(Y ≤ 5) ≈ .562 - .491 = .071

10.

a.



b. $= \int_{-\infty}^{\infty} f(x;k,q)dx = \int_{q}^{\infty} \frac{kq^k}{x^{k+1}}dx = q^k \cdot \left(-\frac{1}{x^k}\right)\Bigg]_q^{\infty} = \frac{q^k}{q^k} = 1$

c. P(X ≤ b) = $\int_q^b \frac{kq^k}{x^{k+1}}dx = q^k \cdot \left(-\frac{1}{x^k}\right)\Bigg]_q^b = 1 - \left(\frac{q}{b}\right)^k$

d. P(a ≤ X ≤ b) = $\int_a^b \frac{kq^k}{x^{k+1}}dx = q^k \cdot \left(-\frac{1}{x^k}\right)\Bigg]_a^b = \left(\frac{q}{a}\right)^k - \left(\frac{q}{b}\right)^k$

Section 4.2

11.

a. P(X ≤ 1) = F(1) = $\frac{1}{4} = .25$

b. P(.5 ≤ X ≤ 1) = F(1) – F(.5) = $\frac{3}{16} = .1875$

c. P(X > .5) = 1 – P(X ≤ .5) = 1 – F(.5) = $\frac{15}{16} = .9375$

d. .5 = $F(\tilde{m}) = \frac{\tilde{m}^2}{4} \Rightarrow \tilde{m}^2 = 2 \Rightarrow \tilde{m} = \sqrt{2} \approx 1.414$

e. f(x) = F′(x) = $\frac{x}{2}$ for 0 ≤ x < 2, and = 0 otherwise

f. E(X) = $\int_{-\infty}^{\infty} x \cdot f(x)dx = \int_0^2 x \cdot \frac{1}{2}xdx = \frac{1}{2}\int_0^2 x^2 dx = \frac{x^3}{6}\Bigg]_0^2 = \frac{8}{6} \approx 1.333$

g. E(X^2) = $\int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^2 x^2 \frac{1}{2}xdx = \frac{1}{2}\int_0^2 x^3 dx = \frac{x^4}{8}\Bigg]_0^2 = 2,$

So Var(X) = E(X^2) – [E(X)]^2 = $2 - \left(\frac{8}{6}\right)^2 = \frac{8}{36} \approx .222$, $\sigma_x \approx .471$

h. From **g**, E(X^2) = 2

12.

a. P(X < 0) = F(0) = .5

b. P(-1 ≤ X ≤ 1) = F(1) – F(-1) = $\frac{11}{16} = .6875$

c. P(X > .5) = 1 – P(X ≤ .5) = 1 – F(.5) = 1 - .6836 = .3164

d. F(x) = F′(x) = $\frac{d}{dx}\left(\frac{1}{2}+\frac{3}{32}\left(4x-\frac{x^3}{3}\right)\right) = 0+\frac{3}{32}\left(4-\frac{3x^2}{3}\right) = .09375\left(4-x^2\right)$

e. $F(\tilde{m}) = .5$ by definition. F(0) = .5 from **a** above, which is as desired.

13.

a. $1=\int_1^\infty \frac{k}{x^4}dx \Rightarrow 1=\frac{-k}{3}x^{-3}\Big|_1^\infty \Rightarrow 1=0-(-\frac{k}{3})(1)\Rightarrow 1=\frac{k}{3}\Rightarrow k=3$

b. cdf: F(x)= $\int_{-\infty}^{x} f(y)dy = \int_1^x 3y^{-4}dy = -\frac{3}{3}y^{-3}\Big|_1^x = -x^{-3}+1 = 1-\frac{1}{x^3}$. So

$$F(x)=\begin{cases} 0, & x\le 1 \\ 1-x^{-3}, & x>1 \end{cases}$$

c. P(x > 2) = 1 – F(2) = 1 – $\left(1-\frac{1}{8}\right)=\frac{1}{8}$ or .125;

$$P(2<x<3)=F(3)-F(2)=\left(1-\tfrac{1}{27}\right)-\left(1-\tfrac{1}{8}\right)=.963-.875=.088$$

d. $E(x)=\int_1^\infty x\left(\frac{3}{x^4}\right)dx=\int_1^\infty\left(\frac{3}{x^3}\right)dx=-\frac{3}{2}x^{-2}\Big|_1^x=0+\frac{3}{2}=\frac{3}{2}$

$$E(x^2)=\int_1^\infty x^2\left(\frac{3}{x^4}\right)dx=\int_1^\infty\left(\frac{3}{x^2}\right)dx=-3x^{-1}\Big|_1^x=0+3=3$$

$V(x)=E(x^2)-[E(x)]^2=3-\left(\frac{3}{2}\right)^2=3-\frac{9}{4}=\frac{3}{4}$ or .75

$$s=\sqrt{V(x)}=\sqrt{3/4}=.866$$

e. $P(1.5-.866<x<1.5+.866)=P(x<2.366)=F(2.366)$

$$=1-(2.366^{-3})=.9245$$

14.

a. If X is uniformly distributed on the interval from A to B, then

$$E(X)=\int_A^B x\cdot\frac{1}{B-A}\,dx=\frac{A+B}{2},\ E(X^2)=\frac{A^2+AB+B^2}{3}$$

$V(X) = E(X^2) – [E(X)]^2 = \dfrac{(B-A)^2}{2}$.

With A = 7.5 and B = 20, E(X) = 13.75, V(X) = 13.02

b. $F(X) = \begin{cases} 0 & x<7.5 \\ \dfrac{x-7.5}{12.5} & 7.5\le x<20 \\ 1 & x\ge 20 \end{cases}$

c. $P(X \le 10) = F(10) = .200$; $P(10 \le X \le 15) = F(15) – F(10) = .4$

d. $\sigma = 3.61$, so $\mu \pm \sigma = (10.14, 17.36)$

Thus, $P(\mu - \sigma \le X \le \mu + \sigma) = F(17.36) – F(10.14) = .5776$

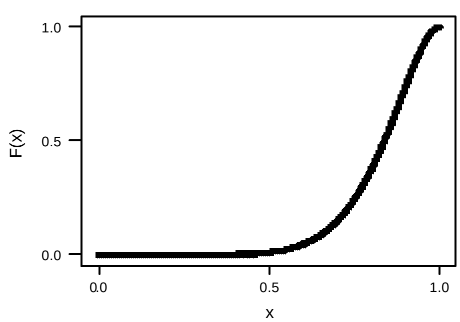
Similarly, $P(\mu - \sigma \le X \le \mu + \sigma) = P(6.53 \le X \le 20.97) = 1$

15.

a. F(X) = 0 for $x \le 0$, = 1 for $x \ge 1$, and for $0 < X < 1$,

$$F(X)=\int_{-\infty}^{x} f(y)dy=\int_0^x 90y^8(1-y)dy=90\int_0^x (y^8-y^9)dy$$

$$90\left(\tfrac{1}{9}y^9-\tfrac{1}{10}y^{10}\right)\Big]_0^x=10x^9-9x^{10}$$



b. $F(.5) = 10(.5)^9 – 9(.5)^{10} \approx .0107$

c. $P(.25 \le X \le .5) = F(.5) – F(.25) \approx .0107 – [10(.25)^9 – 9(.25)^{10}]$
$\approx .0107 – .0000 \approx .0107$

d. The 75th percentile is the value of x for which F(x) = .75
$\Rightarrow .75 = 10(x)^9 – 9(x)^{10}$ $\Rightarrow x \approx .9036$

e. E(X) = $\int_{-\infty}^{\infty} x \cdot f(x)dx = \int_0^1 x \cdot 90x^8(1-x)dx = 90\int_0^1 x^9(1-x)dx$

$= 9x^{10} - \frac{90}{11}x^{11}\Big]_0^1 = \frac{9}{11} \approx .8182$

E(X^2) = $\int_{-\infty}^{\infty} x^2 \cdot f(x)dx = \int_0^1 x^2 \cdot 90x^8(1-x)dx = 90\int_0^1 x^{10}(1-x)dx$

$= \frac{90}{11}x^{11} - \frac{90}{12}x^{12}\Big]_0^1 \approx .6818$

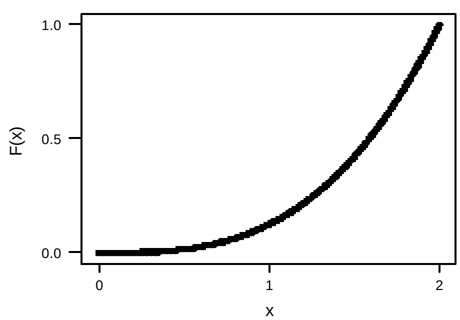
V(X) ≈ .6818 – $(.8182)^2$ = .0124, σ_x = .11134.

f. $\mu \pm \sigma$ = (.7068, .9295). Thus, P(μ - $\sigma \le X \le \mu + \sigma$) = F(.9295) – F(.7068)
= .8465 - .1602 = .6863

16.

a. F(x) = 0 for x < 0 and F(x) = 1 for x > 2. For $0 \le x \le 2$,

F(x) = $\int_0^x \frac{3}{8}y^2 dy = \frac{1}{8}y^3\Big]_0^x = \frac{1}{8}x^3$



b. P(x ≤ .5) = F(.5) = $\frac{1}{8}\left(\frac{1}{2}\right)^3 = \frac{1}{64}$

c. P(.25 ≤ X ≤ .5) = F(.5) – F(.25) $= \frac{1}{64} - \frac{1}{8}\left(\frac{1}{4}\right)^3 = \frac{7}{512} \approx .0137$

d. .75 = F(x) = $\frac{1}{8}x^3 \Rightarrow x^3 = 6 \Rightarrow x \approx 1.8171$

e. E(X) = $\int_{-\infty}^{\infty} x \cdot f(x)dx = \int_0^2 x \cdot \left(\frac{3}{8}x^2\right)dx = \frac{3}{8}\int_0^1 x^3 dx = \frac{3}{8}\left(\frac{1}{4}x^4\right)\Big]_0^2 = \frac{3}{2} = 1.5$

E(X^2) = $\int_0^2 x \cdot \left(\frac{3}{8}x^2\right)dx = \frac{3}{8}\int_0^1 x^4 dx = \frac{3}{8}\left(\frac{1}{5}x5\right)\Big]_0^2 = \frac{12}{5} = 2.4$

V(X) = $\frac{12}{5} - \left(\frac{3}{2}\right)^2 = \frac{3}{20} = .15$ σ_x = .3873

f. $\mu \pm \sigma$ = (1.1127, 1.8873). Thus, P(μ - $\sigma \le X \le \mu + \sigma$) = F(1.8873) – F(1.1127) = .8403 - .1722 = .6681

17.

a. For $2 \le X \le 4$, $F(X) = \int_{-\infty}^{x} f(y)dy = \int_{2}^{x} \frac{3}{4}[1-(y-3)^2]dy$ (let u = y-3)

$= \int_{-1}^{x-3} \frac{3}{4}[1-u^2]du = \frac{3}{4}\left[u - \frac{u^3}{3}\right]_{-1}^{x-3} = \frac{3}{4}\left[x - \frac{7}{3} - \frac{(x-3)^3}{3}\right]$. Thus

$$F(x) = \begin{cases} 0 & x < 2 \\ \frac{1}{4}[3x - 7 - (x-3)^3] & 2 \le x \le 4 \\ 1 & x > 4 \end{cases}$$

b. By symmetry of f(x), $\tilde{m} = 3$

c. E(X) = $\int_{2}^{4} x \cdot \frac{3}{4}[1-(x-3)^2]dx = \frac{3}{4}\int_{-1}^{1}(y+3)(1-y^2)dx$

$$= \frac{3}{4}\left[3y + \frac{y^2}{2} - y^3 - \frac{y^4}{4}\right]_{-1}^{1} = \frac{3}{4}\cdot 4 = 3$$

V(X) = $\int_{-\infty}^{\infty}(x-m)^2 f(x)dx = \frac{3}{4}\int_{2}^{4}(x-3)^2 \cdot [1-(x-3)^2]dx$

$$= \frac{3}{4}\int_{-1}^{1} y^2(1-y^2)dy = \frac{3}{4}\cdot\frac{4}{15} = \frac{1}{5} = .2$$

18.

a. F(X) = $\frac{x-A}{B-A}$ = p $\Rightarrow$ x = (100p)th percentile = A + (B - A)p

b. $E(X) = \int_{A}^{B} x \cdot \frac{1}{B-A}dx = \frac{1}{B-A}\cdot\left.\frac{x^2}{2}\right]_{A}^{B} = \frac{1}{2}\cdot\frac{1}{B-A}\cdot\left(B^2 - A^2\right) = \frac{A+B}{2}$

$$E(X^2) = \frac{1}{3}\cdot\frac{1}{B-A}\cdot\left(B^3 - A^3\right) = \frac{A^2 + AB + B^2}{3}$$

$$V(X) = \left(\frac{A^2 + AB + B^2}{3}\right) - \left(\frac{(A+B)}{2}\right)^2 = \frac{(B-A)^2}{12},\ s_x = \frac{(B-A)}{\sqrt{12}}$$

c. $E(X^n) = \int_{A}^{B} x^n \cdot \frac{1}{B-A}dx = \frac{B^{n+1} - A^{n+1}}{(n+1)(B-A)}$

19.

a. $P(X \le 1) = F(1) = .25[1 + \ln(4)] \approx .597$

b. $P(1 \le X \le 3) = F(3) - F(1) \approx .966 - .597 \approx .369$

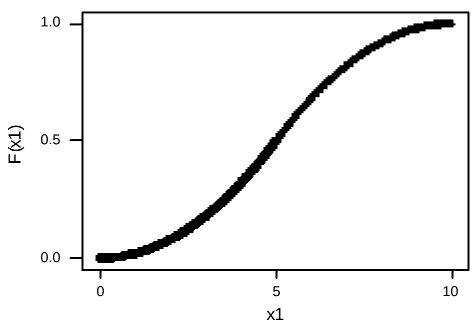
c. $f(x) = F'(x) = .25 \ln(4) - .25 \ln(x)$ for $o < x < 4$

20.

a. For $0 \le y \le 5$, $F(y) = \int_0^y \frac{1}{25} u\,du = \frac{y^2}{50}$

For $5 \le y \le 10$, $F(y) = \int_0^y f(u)du = \int_0^5 f(u)du + \int_5^y f(u)du$

$$= \frac{1}{2} + \int_0^y \left(\frac{2}{5} - \frac{u}{25}\right)du = \frac{2}{5}y - \frac{y^2}{50} - 1$$



b. For $0 < p \le .5$, $p = F(y_p) = \frac{y_p^2}{50} \Rightarrow y_p = (50p)^{1/2}$

For $.5 < p \le 1$, $p = \frac{2}{5}y_p - \frac{y_p^2}{50} - 1 \Rightarrow y_p = 10 - 5\sqrt{2(1-p)}$

c. $E(Y) = 5$ by straightforward integration (or by symmetry of $f(y)$), and similarly $V(Y)= \frac{50}{12} = 4.1667$. For the waiting time X for a single bus,

$E(X) = 2.5$ and $V(X) = \frac{25}{12}$

21. $E(\text{area}) = E(\pi R^2) = \int_{-\infty}^{\infty} pr^2 f(r)dr = \int_9^{11} pr^2 \left(\frac{3}{4}\right)\left(1 - (10-r)^2\right)dr$

$$= \left(\frac{3}{4}\right)p \int_9^{11} r^2\left(1 - (100 - 20r + r^2)\right)dr = \frac{3}{4}p \int_9^{11} -99r^2 + 20r^3 - r^4 dr = 100 \cdot 2p$$

22.

a. For $1 \le x \le 2$, F(x) = $\int_1^x 2\left(1-\frac{1}{y^2}\right)dy = 2\left(y+\frac{1}{y}\right)\Big]_1^x = 2\left(x+\frac{1}{x}\right)-4,$ so

$$F(x) = \begin{cases} 0 & x<1 \\ 2\left(x+\frac{1}{x}\right)-4 & 1\le x\le 2 \\ 1 & x>2 \end{cases}$$

b. $2\left(x_p+\frac{1}{x_p}\right)-4 = p \Rightarrow 2x_p^{\,2}-(4-p)x_p+2=0 \Rightarrow x_p = \frac{1}{4}\left[4+p+\sqrt{p^2+8p}\right]$ To find $\tilde{m}$, set p = .5 $\Rightarrow$ $\tilde{m}$ = 1.64

c. E(X) = $\int_1^2 x\cdot 2\left(1-\frac{1}{x^2}\right)dx = 2\int_1^2\left(x-\frac{1}{x}\right)dx = 2\left(\frac{x^2}{2}-\ln(x)\right)\Big]_1^2 = 1.614$

$E(X^2)$ = $2\int_1^2\left(x^2-1\right)dx = 2\left(\frac{x^3}{3}-x\right)\Big]_1^2 = \frac{8}{3} \Rightarrow$ Var(X) = .0626

d. Amount left = max(1.5 – X, 0), so

E(amount left) = $\int_1^2 \max(1.5-x,0)\,f(x)dx = 2\int_1^{1.5}(1.5-x)\left(1-\frac{1}{x^2}\right)dx = .061$

23. With X = temperature in °C, temperature in °F = $\frac{9}{5}X+32,$ so

$E\left[\frac{9}{5}X+32\right] = \frac{9}{5}(120)+32 = 248,\quad Var\left[\frac{9}{5}X+32\right] = \left(\frac{9}{5}\right)^2\cdot(2)^2 = 12.96,$ so $\sigma = 3.6$

24.

a. E(X) = $\int_q^\infty x \cdot \frac{kq^k}{x^{k+1}}dx = kq^k \int_q^\infty \frac{1}{x^k}dx = \frac{kq^k x^{-k+1}}{-k+1}\Big]_q^\infty = \frac{kq}{k-1}$

b. E(X) = ∞

c. E(X^2) = $kq^k \int_q^\infty \frac{1}{x^{k-1}}dx = \frac{kq^2}{k-2}$, so

Var(X) = $\left(\frac{kq^2}{k-2}\right) - \left(\frac{kq}{k-1}\right)^2 = \frac{kq^2}{(k-2)(k-1)^2}$

d. Var(x) = ∞, since E(X^2) = ∞.

e. E(X^n) = $kq^k \int_q^\infty x^{n-(k+1)}dx$, which will be finite if n – (k+1) < -1, i.e. if n<k.

25.

a. P(Y ≤ 1.8 $\tilde{m}$ + 32) = P(1.8X + 32 ≤ 1.8 $\tilde{m}$ + 32) = P(X ≤ $\tilde{m}$) = .5

b. 90th for Y = 1.8η(.9) + 32 where η(.9) is the 90th percentile for X, since
P(Y ≤ 1.8η(.9) + 32) = P(1.8X + 32 ≤ 1.8η(.9) + 32)
= (X ≤ η(.9)) = .9 as desired.

c. The (100p)th percentile for Y is 1.8η(p) + 32, verified by substituting p for .9 in the argument of **b**. When Y = aX + b, (i.e. a linear transformation of X), and the (100p)th percentile of the X distribution is η(p), then the corresponding (100p)th percentile of the Y distribution is a·η(p) + b. (same linear transformation applied to X's percentile)

Section 4.3

26.

a. $P(0 \le Z \le 2.17) = \Phi(2.17) - \Phi(0) = .4850$

b. $\Phi(1) - \Phi(0) = .3413$

c. $\Phi(0) - \Phi(-2.50) = .4938$

d. $\Phi(2.50) - \Phi(-2.50) = .9876$

e. $\Phi(1.37) = .9147$

f. $P(-1.75 < Z) + [1 – P(Z < -1.75)] = 1 - \Phi(-1.75) = .9599$

g. $\Phi(2) - \Phi(-1.50) = .9104$

h. $\Phi(2.50) - \Phi(1.37) = .0791$

i. $1 - \Phi(1.50) = .0668$

j. $P(|Z| \le 2.50) = P(-2.50 \le Z \le 2.50) = \Phi(2.50) - \Phi(-2.50) = .9876$

27.

a. .9838 is found in the 2.1 row and the .04 column of the standard normal table so c = 2.14.

b. $P(0 \le Z \le c) = .291 \Rightarrow \Phi(c) = .7910 \Rightarrow c = .81$

c. $P(c \le Z) = .121 \Rightarrow 1 - P(c \le Z) = P(Z < c) = \Phi(c) = 1 - .121 = .8790 \Rightarrow c = 1.17$

d. $P(-c \le Z \le c) = \Phi(c) - \Phi(-c) = \Phi(c) – (1 - \Phi(c)) = 2\Phi(c) – 1$
$\Rightarrow \Phi(c) = .9920 \Rightarrow c = .97$

e. $P(c \le |Z|) = .016 \Rightarrow 1 - .016 = .9840 = 1 – P(c \le |Z|) = P(|Z| < c)$
$= P(-c < Z < c) = \Phi(c) - \Phi(-c) = 2\Phi(c) – 1$
$\Rightarrow \Phi(c) = .9920 \Rightarrow c = 2.41$

28.

a. $\Phi(c) = .9100 \Rightarrow c \approx 1.34$ (.9099 is the entry in the 1.3 row, .04 column)

b. 9^{th} percentile = -91^{st} percentile = -1.34

c. $\Phi(c) = .7500 \Rightarrow c \approx .675$ since .7486 and .7517 are in the .67 and .68 entries, respectively.

d. $25^{th} = -75^{th} = -.675$

e. $\Phi(c) = .06 \Rightarrow c \approx .-1.555$ (both .0594 and .0606 appear as the –1.56 and –1.55 entries, respectively).

29.

a. Area under Z curve above $z_{.0055}$ is .0055, which implies that $\Phi(z_{.0055}) = 1 - .0055 = .9945$, so $z_{.0055} = 2.54$

b. $\Phi(z_{.09}) = .9100 \Rightarrow z = 1.34$ (since .9099 appears as the 1.34 entry).

c. $\Phi(z_{.633})$ = area below $z_{.633} = .3370 \Rightarrow z_{.633} \approx -.42$

30.

a. $P(X \le 100) = P\left(z \le \dfrac{100-80}{10}\right) = P(Z \le 2) = \Phi(2.00) = .9772$

b. $P(X \le 80) = P\left(z \le \dfrac{80-80}{10}\right) = P(Z \le 0) = \Phi(0.00) = .5$

c. $P(65 \le X \le 100) = P\left(\dfrac{65-80}{10} \le z \le \dfrac{100-80}{10}\right) = P(-1.50 \le Z \le 2)$
$= \Phi(2.00) - \Phi(-1.50) = .9772 - .0668 = .9104$

d. $P(70 \le X) = P(-1.00 \le Z) = 1 - \Phi(-1.00) = .8413$

e. $P(85 \le X \le 95) = P(.50 \le Z \le 1.50) = \Phi(1.50) - \Phi(.50) = .2417$

f. $P(|X - 80| \le 10) = P(-10 \le X - 80 \le 10) = P(70 \le X \le 90)$
$P(-1.00 \le Z \le 1.00) = .6826$

31.

a. $P(X \le 18) = P\left(z \le \frac{18-15}{1.25}\right) = P(Z \le 2.4) = \Phi(2.4) = .9452$

b. $P(10 \le X \le 12) = P(-4.00 \le Z \le -2.40) \approx P(Z \le -2.40) = \Phi(-2.40) = .0082$

c. $P(|X - 10| \le 2(1.25)) = P(-2.50 \le X\text{-}15 \le 2.50) = P(12.5 \le X \le 17.5)$
$P(-2.00 \le Z \le 2.00) = .9544$

32.

a. $P(X > .25) = P(Z > -.83) = 1 - .2033 = .7967$

b. $P(X \le .10) = \Phi(-3.33) = .0004$

c. We want the value of the distribution, c, that is the 95^{th} percentile (5% of the values are higher). The 95^{th} percentile of the standard normal distribution = 1.645. So c = .30 + (1.645)(.06) = .3987. The largest 5% of all concentration values are above .3987 mg/cm^3.

33.

a. $P(X \ge 10) = P(Z \ge .43) = 1 - \Phi(.43) = 1 - .6664 = .3336$.
$P(X > 10) = P(X \ge 10) = .3336$, since for any continuous distribution, $P(x = a) = 0$.

b. $P(X > 20) = P(Z > 4) \approx 0$

c. $P(5 \le X \le 10) = P(-1.36 \le Z \le .43) = \Phi(.43) - \Phi(-1.36) = .6664 - .0869 = .5795$

d. $P(8.8 - c \le X \le 8.8 + c) = .98$, so $8.8 - c$ and $8.8 + c$ are at the 1^{st} and the 99^{th} percentile of the given distribution, respectively. The 1^{st} percentile of the standard normal distribution has the value –2.33, so
$8.8 - c = \mu + (-2.33)\sigma = 8.8 - 2.33(2.8) \Rightarrow c = 2.33(2.8) = 6.524$.

e. From a, $P(x > 10) = .3336$. Define event A as {diameter > 10}, then P(at least one A_i) =
$1 - P(\text{no } A_i) = 1 - P(A')^4 = 1 - (1 - .3336)^4 = 1 - .1972 = .8028$

34. Let X denote the diameter of a randomly selected cork made by the first machine, and let Y be defined analogously for the second machine.
$P(2.9 \le X \le 3.1) = P(-1.00 \le Z \le 1.00) = .6826$
$P(2.9 \le Y \le 3.1) = P(-7.00 \le Z \le 3.00) = .9987$
So the second machine wins handily.

35.

a. $\mu + \sigma\cdot(91^{st}$ percentile from std normal) = 30 + 5(1.34) = 36.7

b. 30 + 5(-1.555) = 22.225

c. μ = 3.000 μm; σ = 0.140. We desire the 90^{th} percentile: 30 + 1.28(0.14) = 3.179

36. $\mu = 43; \sigma = 4.5$

a. P(X < 40) = $P\left(z \le \frac{40-43}{4.5}\right)$ = P(Z < -0.667) = .2514

P(X > 60) = $P\left(z > \frac{60-43}{4.5}\right)$ = P(Z > 3.778) ≈ 0

b. 43 + (-0.67)(4.5) = 39.985

37. P(damage) = P(X < 100) = $P\left(z < \frac{100-200}{300}\right)$ = P(Z < -3.33) = .0004

P(at least one among five is damaged) = 1 – P(none damaged)
= 1 – $(.9996)^5$ = 1 - .998 = .002

38. From Table A.3, P(-1.96 ≤ Z ≤ 1.96) = .95. Then P(μ - .1 ≤ X ≤ μ + .1) =
$P\left(\frac{-.1}{s} < z < \frac{.1}{s}\right)$ implies that $\frac{.1}{s}$ = 1.96, and thus that $s = \frac{.1}{1.96} = .0510$

39. Since 1.28 is the 90^{th} z percentile ($z_{.1}$ = 1.28) and –1.645 is the 5^{th} z percentile ($z_{.05}$ = 1.645), the given information implies that $\mu + \sigma(1.28) = 10.256$ and $\mu + \sigma(-1.645) = 9.671$, from which $\sigma(-2.925) = -.585$, $\sigma = .2000$, and $\mu = 10$.

40.

a. P(μ - 1.5σ ≤ X ≤ μ + 1.5σ) = P(-1.5 ≤ Z ≤ 1.5) = Φ(1.50) - Φ(-1.50) = .8664

b. P(X < μ - 2.5σ or X > μ + 2.5σ) = 1 - P(μ - 2.5σ ≤ X ≤ μ + 2.5σ)
= 1 - P(-2.5 ≤ Z ≤ 2.5) = 1 - .9876 = .0124

c. P(μ - 2σ ≤ X ≤ μ - σ or μ + σ ≤ X ≤ μ + 2σ) = P(within 2 sd's) – P(within 1 sd) = P(μ - 2σ ≤ X ≤ μ + 2σ) - P(μ - σ ≤ X ≤ μ + σ)
= .9544 - .6826 = .2718

41. With μ = .500 inches, the acceptable range for the diameter is between .496 and .504 inches, so unacceptable bearings will have diameters smaller than .496 or larger than .504. The new distribution has μ = .499 and σ =.002. P(x < .496 or x >.504) =

$$P\left(z < \frac{.496 - .499}{.002}\right) + P\left(z > \frac{.504 - .499}{.002}\right) = P(z < -1.5) + P(z > 2.5)$$

$\Phi(-1.5) + (1 - \Phi(2.5)) = .0068 + .0062 = .073$, or 7.3% of the bearings will be unacceptable.

42.

a. P(67 ≤ X ≤ 75) = P(-1.00 ≤ Z ≤ 1.67) = .7938

b. P(70 – c ≤ X ≤ 70 + c) = $P\left(\frac{-c}{3} \le Z \le \frac{c}{3}\right) = 2\Phi(\frac{c}{3}) - 1 = .95 \Rightarrow \Phi(\frac{c}{3}) = .9750$

$$\frac{c}{3} = 1.96 \Rightarrow c = 5.88$$

c. 10·P(a single one is acceptable) = 9.05

d. p = P(X < 73.84) = P(Z < 1.28) = .9, so P(Y ≤ 8) = B(8;10,.9) = .264

43. The stated condition implies that 99% of the area under the normal curve with μ = 10 and σ = 2 is to the left of c – 1, so c – 1 is the 99th percentile of the distribution. Thus c – 1 = μ + σ(2.33) = 20.155, and c = 21.155.

44.

a. By symmetry, P(-1.72 ≤ Z ≤ -.55) = P(.55 ≤ Z ≤ 1.72) = Φ(1.72) - Φ(.55)

b. P(-1.72 ≤ Z ≤ .55) = Φ(.55) - Φ(-1.72) = Φ(.55) – [1 - Φ(1.72)]
No, symmetry of the Z curve about 0.

45. X ~N(3432, 482)

a. $$P(x > 4000) = P\left(Z > \frac{4000 - 3432}{482}\right) = P(z > 1.18)$$
$$= 1 - \Phi(1.18) = 1 - .8810 = .1190$$
$$P(3000 < x < 4000) = P\left(\frac{3000 - 3432}{482} < Z < \frac{4000 - 3432}{482}\right)$$
$$= \Phi(1.18) - \Phi(-.90) = .8810 - .1841 = .6969$$

b. $$P(x < 2000 or x > 5000) = P\left(Z < \frac{2000 - 3432}{482}\right) + P\left(Z > \frac{5000 - 3432}{482}\right)$$
$$= \Phi(-2.97) + [1 - \Phi(3.25)] = .0015 + .0006 = .0021$$

c. We will use the conversion 1 lb = 454 g, then 7 lbs = 3178 grams, and we wish to find

$$P(x > 3178) = P\left(Z > \frac{3178 - 3432}{482}\right) = 1 - \Phi(-.53) = .7019$$

d. We need the top .0005 and the bottom .0005 of the distribution. Using the Z table, both .9995 and .0005 have multiple z values, so we will use a middle value, ±3.295. Then 3432±(482)3.295 = 1844 and 5020, or the most extreme .1% of all birth weights are less than 1844 g and more than 5020 g.

e. Converting to lbs yields mean 7.5595 and s.d. 1.0608. Then

$P(x > 7) = P\left(Z > \frac{7 - 7.5595}{1.0608}\right) = 1 - \Phi(-.53) = .7019$ This yields the same answer as in part c.

46. We use a Normal approximation to the Binomial distribution: X ~ b(x;1000,.03) ˜ N(30,5.394)

a. $P(x \geq 40) = 1 - P(x \leq 39) = 1 - P\left(Z \leq \frac{39.5 - 30}{5.394}\right)$

$= 1 - \Phi(1.76) = 1 - .9608 = .0392$

b. 5% of 1000 = 50: $P(x \leq 50) = P\left(Z \leq \frac{50.5 - 30}{5.394}\right) = \Phi(3.80) \approx 1.00$

47. $P(|X - \mu| \geq \sigma) = P(X \leq \mu - \sigma \text{ or } X \geq \mu + \sigma)$

$= 1 - P(\mu - \sigma \leq X \leq \mu + \sigma) = 1 - P(-1 \leq Z \leq 1) = .3174$

Similarly, $P(|X - \mu| \geq 2\sigma) = 1 - P(-2 \leq Z \leq 2) = .0456$

And $P(|X - \mu| \geq 3\sigma) = 1 - P(-3 \leq Z \leq 3) = .0026$

48.

a. $P(20 - .5 \leq X \leq 30 + .5) = P(19.5 \leq X \leq 30.5) = P(-1.1 \leq Z \leq 1.1) = .7286$

b. P(at most 30) = $P(X \leq 30 + .5) = P(Z \leq 1.1) = .8643$.

P(less than 30) = $P(X < 30 - .5) = P(Z < .9) = .8159$

49.

P:	.5	.6	.8
μ:	12.5	15	20
σ:	2.50	2.45	2.00

a.

	$P(15 \le X \le 20)$	$P(14.5 \le \text{normal} \le 20.5)$
.5	.212	$P(.80 \le Z \le 3.20) = .2112$
.6	.577	$P(-.20 \le Z \le 2.24) = .5668$
.8	.573	$P(-2.75 \le Z \le .25) = .5957$

b.

$P(X \le 15)$	$P(\text{normal} \le 15.5)$
.885	$P(Z \le 1.20) = .8849$
.575	$P(Z \le .20) = .5793$
.017	$P(Z \le -2.25) = .0122$

c.

$P(20 \le X)$	$P(19.5 \le \text{normal})$
.002	.0026
.029	.0329
.617	.5987

50. P = .10; n = 200; np = 20, npq = 18

a. $P(X \le 30) = \Phi\left(\frac{30+.5-20}{\sqrt{18}}\right) = \Phi(2.47) = .9932$

b. $P(X < 30) = P(X \le 29) = \Phi\left(\frac{29+.5-20}{\sqrt{18}}\right) = \Phi(2.24) = .9875$

c. $P(15 \le X \le 25) = P(X \le 25) - P(X \le 14) = \Phi\left(\frac{25+.5-20}{\sqrt{18}}\right) - \Phi\left(\frac{14+.5-20}{\sqrt{18}}\right)$

$\Phi(1.30) - \Phi(-1.30) = .9032 - .0968 = .8064$

51. N = 500, p = .4, μ = 200, σ = 10.9545

a. $P(180 \le X \le 230) = P(179.5 \le \text{normal} \le 230.5) = P(-1.87 \le Z \le 2.78) = .9666$

b. $P(X < 175) = P(X \le 174) = P(\text{normal} \le 174.5) = P(Z \le -2.33) = .0099$

52. P(X ≤ μ + σ[(100p)th percentile for std normal])

$P\left(\frac{X-\mu}{\sigma} \le [\ldots]\right)$ = P(Z ≤ […]) = p as desired

53.

a. $F_y(y) = P(Y \le y) = P(aX + b \le y) = P\left(X \le \frac{(y-b)}{a}\right)$ (for a > 0).

Now differentiate with respect to y to obtain

$f_y(y) = F_y{}'(y) = \frac{1}{\sqrt{2\pi}\, a\sigma} e^{-\frac{1}{2a^2\sigma^2}[y-(a\mu+b)]^2}$ so Y is normal with mean aμ + b and variance $a^2\sigma^2$.

b. Normal, mean $\frac{9}{5}(115)+32 = 239$, variance = 12.96

54.

a. $P(Z \ge 1) \approx .5 \cdot \exp\left(\frac{83+351+562}{703+165}\right) = .1587$

b. $P(Z > 3) \approx .5 \cdot \exp\left(\frac{-2362}{399.3333}\right) = .0013$

c. $P(Z > 4) \approx .5 \cdot \exp\left(\frac{-3294}{340.75}\right) = .0000317$, so

$P(-4 < Z < 4) \approx 1 - 2(.0000317) = .999937$

d. $P(Z > 5) \approx .5 \cdot \exp\left(\frac{-4392}{305.6}\right) = .00000029$

Section 4.4

55.

a. $\Gamma(6) = 5! = 120$

b. $\Gamma\left(\frac{5}{2}\right) = \frac{3}{2}\Gamma\left(\frac{1}{2}\right) = \frac{3}{2}\cdot\frac{1}{2}\cdot\Gamma\left(\frac{1}{2}\right) = \left(\frac{3}{4}\right)\sqrt{p} \approx 1.329$

c. F(4;5) = .371 from row 4, column 5 of Table A.4

d. F(5;4) = .735

e. F(0;4) = $P(X \le 0; \alpha = 4) = 0$

56.

a. $P(X \le 5) = F(5;7) = .238$

b. $P(X < 5) = P(X \le 5) = .238$

c. $P(X > 8) = 1 - P(X < 8) = 1 - F(8;7) = .313$

d. $P(3 \le X \le 8) = F(8;7) - F(3;7) = .653$

e. $P(3 < X < 8) = .653$

f. $P(X < 4 \text{ or } X > 6) = 1 - P(4 \le X \le 6) = 1 - [F(6;7) - F(4;7)] = .713$

57.

a. $\mu = 20,\ \sigma^2 = 80 \Rightarrow \alpha\beta = 20, \alpha\beta^2 = 80 \Rightarrow \beta = \frac{80}{20}, \alpha = 5$

b. $P(X \le 24) = F\left(\frac{24}{4};5\right) = F(6;5) = .715$

c. $P(20 \le X \le 40) = F(10;5) - F(5;5) = .411$

58. $\mu = 24,\ \sigma^2 = 144 \Rightarrow \alpha\beta = 24, \alpha\beta^2 = 144 \Rightarrow \beta = 6, \alpha = 4$

a. $P(12 \le X \le 24) = F(4;4) - F(2;4) = .424$

b. $P(X \le 24) = F(4;4) = .567$, so while the mean is 24, the median is less than 24. ($P(X \le \tilde{m}) = .5$); This is a result of the positive skew of the gamma distribution.

c. We want a value of X for which F(X;4)=.99. In table A.4, we see F(10;4)=.990. So with $\beta = 6$, the 99th percentile = 6(10)=60.

d. We want a value of X for which F(X;4)=.995. In the table, F(11;4)=.995, so t = 6(11)=66. At 66 weeks, only .5% of all transistors would still be operating.

59.

a. $E(X) = \dfrac{1}{l} = 1$

b. $s = \dfrac{1}{l} = 1$

c. $P(X \le 4) = 1 - e^{-(1)(4)} = 1 - e^{-4} = .982$

d. $P(2 \le X \le 5) = 1 - e^{-(1)(5)} - \left[1 - e^{-(1)(2)}\right] = e^{-2} - e^{-5} = .129$

60.

a. $P(X \le 100) = 1 - e^{-(100)(.01386)} = 1 - e^{-1.386} = .7499$

$P(X \le 200) = 1 - e^{-(200)(.01386)} = 1 - e^{-2.772} = .9375$

$P(100 \le X \le 200) = P(X \le 200) - P(X \le 100) = .9375 - .7499 = .1876$

b. $\mu = \dfrac{1}{.01386} = 72.15$, $\sigma = 72.15$

$P(X > \mu + 2\sigma) = P(X > 72.15 + 2(72.15)) = P(X > 216.45) =$

$1 - \left[1 - e^{-(216.45)(.01386)}\right] = e^{-2.9999} = .0498$

c. $.5 = P(X \le \tilde{m}) \Rightarrow 1 - e^{-(\tilde{m})(.01386)} = .5 \Rightarrow e^{-(\tilde{m})(.01386)} = .5$

$-\tilde{m}(.01386) = \ln(.5) = .693 \Rightarrow \tilde{m} = 50$

61. Mean = $\dfrac{1}{l} = 25{,}000$ implies $\lambda = .00004$

a. $P(X > 20{,}000) = 1 - P(X \le 20{,}000) = 1 - F(20{,}000; .00004) = e^{-(.00004)(20{,}000)} = .449$

$P(X \le 30{,}000) = F(30{,}000; .00004) = e^{-1.2} = .699$

$P(20{,}000 \le X \le 30{,}000) = .699 - .551 = .148$

b. $s = \dfrac{1}{l} = 25{,}000$, so $P(X > \mu + 2\sigma) = P(x > 75{,}000) =$

$1 - F(75{,}000;.00004) = .05.$

Similarly, $P(X > \mu + 3\sigma) = P(x > 100{,}000) = .018$

62.

a. E(X) = αβ = $n\frac{1}{l} = \frac{n}{l}$; for λ = .5, n = 10, E(X) = 20

b. P(X ≤ 30) = $F\left(\frac{30}{2};10\right)$ = F(15;10) = .930

c. P(X ≤ t) = P(at least n events in time t) = P(Y ≥ n) when Y ~ Poisson with parameter λt .
Thus P(X ≤ t) = 1 – P(Y < n) = 1 – P(Y ≤ n – 1) $= 1 - \sum_{k=0}^{n-1} \frac{e^{-lt}(lt)^k}{k!}$.

63.

a. {X ≥ t} = $A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5$

b. P(X ≥ t) =P(A_1) · P(A_2) · P(A_3) · P(A_4) · P(A_5) = $\left(e^{-lt}\right)^5 = e^{-.05t}$, so F_x(t) = P(X ≤ t) = 1 - $e^{-.05t}$, f_x(t) = $.05e^{-.05t}$ for t ≥ 0. Thus X also ha an exponential distribution , but with parameter λ = .05.

c. By the same reasoning, P(X ≤ t) = 1 - e^{-nlt}, so X has an exponential distribution with parameter nλ.

64. With x_p = (100p)th percentile, p = F(x_p) = 1 - $e^{-lx_p} \Rightarrow e^{-lx_p} = 1 - p$,
$\Rightarrow -lx_p = \ln(1-p) \Rightarrow x_p = \frac{-[\ln(1-p)]}{l}$. For p = .5, $x_{.5}$ = $\tilde{m} = \frac{.693}{l}$.

65.

a. {X^2 ≤ y} = $\left\{-\sqrt{y} \le X \le \sqrt{y}\right\}$

b. P(X^2 ≤ y) = $\int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2p}} e^{-z^2/2} dz$. Now differentiate with respect to y to obtain the chi-squared p.d.f. with ν = 1.

Section 4.5

66.

a. E(X) = $3\Gamma\left(1+\frac{1}{2}\right) = 3\cdot\frac{1}{2}\cdot\Gamma\left(\frac{1}{2}\right) = 2.66$,

Var(X) = $9\left[\Gamma(1+1) - \Gamma^2\left(1+\frac{1}{2}\right)\right] = 1.926$

b. P(X ≤ 6) = $1 - e^{-(6/b)^a} = 1 - e^{-(6/3)^2} = 1 - e^{-4} = .982$

c. P(1.5 ≤ X ≤ 6) = $1 - e^{-(6/3)^2} - \left[1 - e^{-(1.5/3)^2}\right] = e^{-.25} - e^{-4} = .760$

67.

a. P(X ≤ 250) = F(250;2.5, 200) = $1 - e^{-(250/200)^{2.5}} = 1 - e^{-1.75} \approx .8257$
P(X < 250) = P(X ≤ 250) ≈ .8257
P(X > 300) = 1 – F(300; 2.5, 200) = $e^{-(1.5)^{2.5}} = .0636$

b. P(100 ≤ X ≤ 250) = F(250;2.5, 200) - F(100;2.5, 200) ≈ .8257 - .162 = .6637

c. The median $\tilde{m}$ is requested. The equation F($\tilde{m}$) = .5 reduces to

.5 = $e^{-(\tilde{m}/200)^{2.5}}$, i.e., ln(.5) ≈ $-\left(\frac{\tilde{m}}{200}\right)^{2.5}$, so $\tilde{m}$ = $(.6931)^{.4}(200) = 172.727$.

68.

a. For x > 3.5, F(x) = P(X ≤ x) = P(X – 3.5 ≤ x – 3.5) = 1 - $e^{-\left[\frac{(x-3.5)}{1.5}\right]^2}$

b. E(X – 3.5) = $1.5\Gamma\left(\frac{3}{2}\right)$ = 1.329 so E(X) = 4.829

Var(X) = Var(X – 3.5) = $(1.5)^2\left[\Gamma(2) - \Gamma^2\left(\frac{3}{2}\right)\right] = .483$

c. P(X > 5) = 1 – P(X ≤ 5) = $1 - \left[1 - e^{-1}\right] = e^{-1} = .368$

d. P(5 ≤ X ≤ 8) = $1 - e^{-9} - \left[1 - e^{-1}\right] = e^{-1} - e^{-9} = .3679 - .0001 = .3678$

69. $m = \int_0^\infty x \cdot \frac{a}{b^a} x^{a-1} e^{-(x/b)^a}\, dx$ = (after y = $\left(\frac{x}{b}\right)^a$, dy = $\frac{ax^{a-1}}{b^a}$ dx)

$b \int_0^\infty y^{1/a} e^{-y} dy = b \cdot \Gamma\left(1 + \frac{1}{a}\right)$ by definition of the gamma function.

70.

a. $.5 = F(\tilde{m}) = 1 - e^{-(m/3)^2} \Rightarrow$

$e^{-m^2/9} = .5 \Rightarrow \tilde{m}^2 = -9\ln(.5) = 6.2383 \Rightarrow \tilde{m} = 2.50$

b. $1 - e^{-[(\tilde{m}-3.5)/1.5]^2} = .5 \Rightarrow (\tilde{m} - 3.5)^2$ = -2.25 ln(.5) = 1.5596 $\Rightarrow \tilde{m}$ = 4.75

c. P = F(x_p) = 1 - $e^{-(x_p/b)^a}$ $\Rightarrow$ $(x_p/\beta)^\alpha$ = -ln(1 – p) $\Rightarrow$ $x_p = \beta[-\ln(1-p)]^{1/\alpha}$

d. The desired value of t is the 90th percentile (since 90% will not be refused and 10% will be). From **c,** the 90th percentile of the distribution of X – 3.5 is 1.5[-ln(.1)]$^{1/2}$ = 2.27661, so t = 3.5 + 2.2761 = 5.7761

71. X ~ Weibull: α=20,β=100

a. $F(x,20,b) = 1 - e^{-\left(\frac{x}{b}\right)^a} = 1 - e^{-\left(\frac{105}{100}\right)^{20}} = 1 - .070 = .930$

b. $F(105) - F(100) = .930 - \left(1 - e^{-1}\right) = .930 - .632 = .298$

c. $.50 = 1 - e^{-\left(\frac{x}{100}\right)^{20}} \Rightarrow e^{-\left(\frac{x}{100}\right)^{20}} = .50 \Rightarrow -\left(\frac{x}{100}\right)^{20} = \ln(.50)$

$\left(\frac{-x}{100}\right) = \sqrt[20]{\ln(.50)} \Rightarrow -x = 100\left(\sqrt[20]{\ln(.50)}\right) \Rightarrow x = 98.18$

72.

a. $E(X) = e^{\left(m + \frac{s^2}{2}\right)} = e^{4.82} = 123.97$

$V(X) = \left(e^{\left(2(4.5)+.8^2\right)}\right) \cdot \left(e^{-.8} - 1\right) = (15{,}367.34)(.8964) = 13{,}776.53$

$s = 117.373$

b. $P(x \le 100) = P\left(z \le \frac{\ln(100) - 4.5}{.8}\right) = \Phi(0.13) = .5517$

c. $P(x \ge 200) = P\left(z \ge \frac{\ln(200) - 4.5}{.8}\right) = 1 - \Phi(1.00) = 1 - .8413 = .1587 = P(x > 200)$

73.

a. E(X) = $e^{3.5+(1.2)^2/2}$ = 68.0335; V(X) = $e^{2(3.5)+(1.2)^2} \cdot \left(e^{(1.2)^2} - 1\right) = 14907.168$;
σ_x = 122.0949

b. P(50 ≤ X ≤ 250) = $P\left(z \le \frac{\ln(250) - 3.5}{1.2}\right) - P\left(z \le \frac{\ln(50) - 3.5}{1.2}\right)$
P(Z ≤ 1.68) – P(Z ≤ .34) = .9535 - .6331 = .3204.

c. P(X ≤ 68.0335) = $P\left(z \le \frac{\ln(68.0335) - 3.5}{1.2}\right)$ = P(Z ≤ .60) = .7257. The lognormal distribution is not a symmetric distribution.

74.

a. .5 = F($\tilde{m}$) = $\Phi\left(\frac{\ln(\tilde{m}) - m}{s}\right)$, (where $\tilde{m}$ refers to the lognormal distribution and μ and σ to the normal distribution). Since the median of the standard normal distribution is 0, $\frac{\ln(\tilde{m}) - m}{s} = 0$, so ln($\tilde{m}$) = μ ⇒ $\tilde{m} = e^{m}$. For the power distribution,
$\tilde{m} = e^{3.5} = 33.12$

b. 1 - α = Φ(z_α) = P(Z ≤ z_α) = $\left(\frac{\ln(X) - m}{s} \le z_a\right) = P(\ln(X) \le m + sz_a)$
= $P(X \le e^{m+sz_a})$, so the 100(1 - α)th percentile is e^{m+sz_a}. For the power distribution, the 95th percentile is $e^{3.5+(1.645)(1.2)} = e^{5.474} = 238.41$

75.

a. E(X) = $e^{5+(.01)/2} = e^{5.005} = 149.157$; Var(X) = $e^{10+(.01)} \cdot \left(e^{.01} - 1\right) = 223.594$

b. P(X > 125) = 1 – P(X ≤ 125) =
$$= 1 - P\left(z \le \frac{\ln(125) - 5}{.1}\right) = 1 - \Phi(-1.72) = .9573$$

c. P(110 ≤ X ≤ 125) $= \Phi(-1.72) - \Phi\left(\frac{\ln(110) - 5}{.1}\right) = .0427 - .0013 = .0414$

d. $\tilde{m} = e^5 = 148.41$ (continued)

e. P(any particular one has X > 125) = .9573 ⇒ expected # = 10(.9573) = 9.573

f. We wish the 5th percentile, which is $e^{5+(-1.645)(.1)} = 125.90$

76.

a. E(X) = $e^{1.9+9^2/2} = 10.024$; Var(X) = $e^{3.8+(.81)} \cdot \left(e^{.81}-1\right) = 125.395$, σ_x = 11.20

b. P(X ≤ 10) = P(ln(X) ≤ 2.3026) = P(Z ≤ .45) = .6736
P(5 ≤ X ≤ 10) = P(1.6094 ≤ ln(X) ≤2.3026)
= P(-.32 ≤ Z ≤ .45) = .6736 - .3745 = .2991

77. The point of symmetry must be $\frac{1}{2}$, so we require that $f\left(\frac{1}{2}-m\right) = f\left(\frac{1}{2}+m\right)$, i.e., $\left(\frac{1}{2}-m\right)^{a-1}\left(\frac{1}{2}+m\right)^{b-1} = \left(\frac{1}{2}+m\right)^{a-1}\left(\frac{1}{2}-m\right)^{b-1}$, which in turn implies that $\alpha = \beta$.

78.

a. E(X) = $\frac{5}{(5+2)} = \frac{5}{7} = .714$, V(X) = $\frac{10}{(49)(8)} = .0255$

b. f(x) = $\frac{\Gamma(7)}{\Gamma(5)\Gamma(2)} \cdot x^4 \cdot (1-x) = 30\left(x^4 - x^5\right)$ for 0 ≤ X ≤ 1,

so P(X ≤ .2) = $\int_0^{.2} 30\left(x^4 - x^5\right)dx = .0016$

c. P(.2 ≤ X ≤ .4) = $\int_{.2}^{.4} 30\left(x^4 - x^5\right)dx = .03936$

d. E(1 – X) = 1 – E(X) = 1 - $\frac{5}{7} = \frac{2}{7} = .286$

79.

a. E(X) = $\int_0^1 x \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}dx = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 x^a(1-x)^{b-1}dx$

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} = \frac{a\Gamma(a)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+b)}{(a+b)\Gamma(a+b)} = \frac{a}{a+b}$$

b. E[(1 – X)^m] = $\int_0^1 (1-x)^m \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}dx$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 x^{a-1}(1-x)^{m+b-1}dx = \frac{\Gamma(a+b)\cdot\Gamma(m+b)}{\Gamma(a+b+m)\Gamma(b)}$$

For m = 1, E(1 – X) = $\frac{b}{a+b}$.

80.

a. E(Y) = 10 $\Rightarrow E\left(\dfrac{Y}{20}\right) = \dfrac{1}{2} = \dfrac{a}{a+b}$; Var(Y) = $\dfrac{100}{7} \Rightarrow Var\left(\dfrac{Y}{20}\right) = \dfrac{100}{2800} = \dfrac{1}{28}$ $\dfrac{ab}{(a+b)^2(a+b+1)} \Rightarrow a = 3,\ b = 3$, after some algebra.

b. P(8 ≤ X ≤ 12) = $F\left(\dfrac{12}{20};3,3\right) - F\left(\dfrac{8}{20};3,3\right)$ = F(.6;3,3) – F(.4; 3,3).
The standard density function here is $30y^2(1-y)^2$,
so P(8 ≤ X ≤ 12) = $\int_{.4}^{.6} 30y^2(1-y)^2\,dy = .365$.

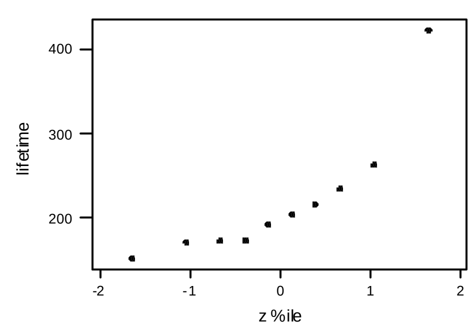
c. We expect it to snap at 10, so P(Y < 8 or Y > 12) = 1 - P(8 ≤ X ≤ 12)
= 1 - .365 = .665.

Section 4.6

81. The given probability plot is quite linear, and thus it is quite plausible that the tension distribution is normal.

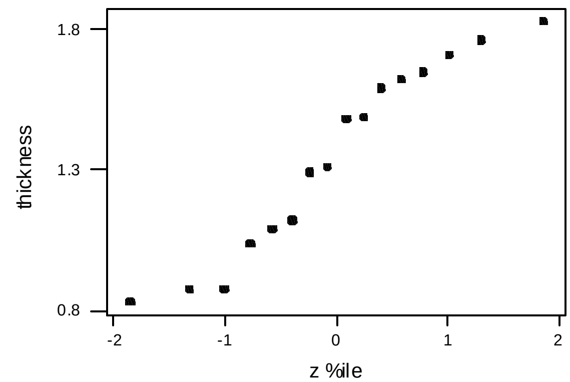
82. The z percentiles and observations are as follows:

percentile	observation
-1.645	152.7
-1.040	172.0
-0.670	172.5
-0.390	173.3
-0.130	193.0
0.130	204.7
0.390	216.5
0.670	234.9
1.040	262.6
1.645	422.6

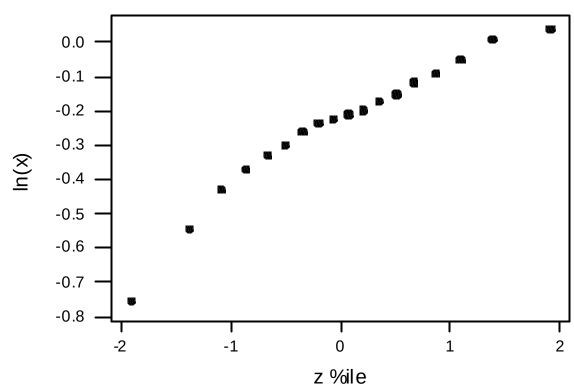


The accompanying plot is quite straight except for the point corresponding to the largest observation. This observation is clearly much larger than what would be expected in a normal random sample. Because of this outlier, it would be inadvisable to analyze the data using any inferential method that depended on assuming a normal population distribution.

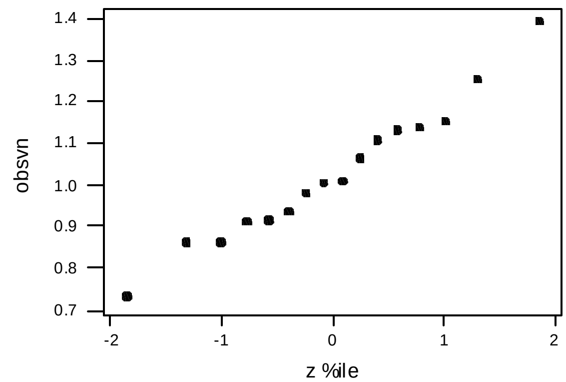
83. The z percentile values are as follows: -1.86, -1.32, -1.01, -0.78, -0.58, -0.40, -0.24,-0.08, 0.08, 0.24, 0.40, 0.58, 0.78, 1.01, 1.30, and 1.86. The accompanying probability plot is reasonably straight, and thus it would be reasonable to use estimating methods that assume a normal population distribution.



84. The Weibull plot uses ln(observations) and the z percentiles of the p_i values given. The accompanying probability plot appears sufficiently straight to lead us to agree with the argument that the distribution of fracture toughness in concrete specimens could well be modeled by a Weibull distribution.

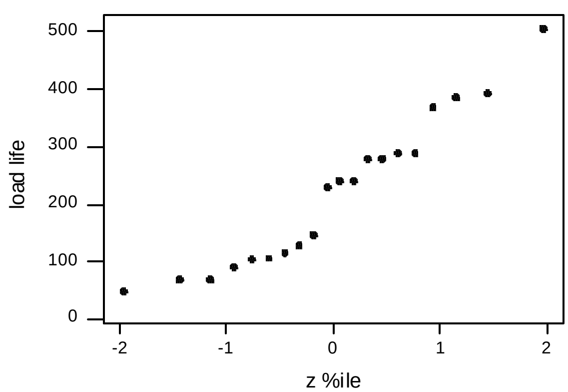


85. The (z percentile, observation) pairs are (-1.66, .736), (-1.32, .863), (-1.01, .865), (-.78, .913), (-.58, .915), (-.40, .937), (-.24, .983), (-.08, 1.007), (.08, 1.011), (.24, 1.064), (.40, 1.109), (.58, 1.132), (.78, 1.140), (1.01, 1.153), (1.32, 1.253), (1.86, 1.394). The accompanying probability plot is very straight, suggesting that an assumption of population normality is extremely plausible.

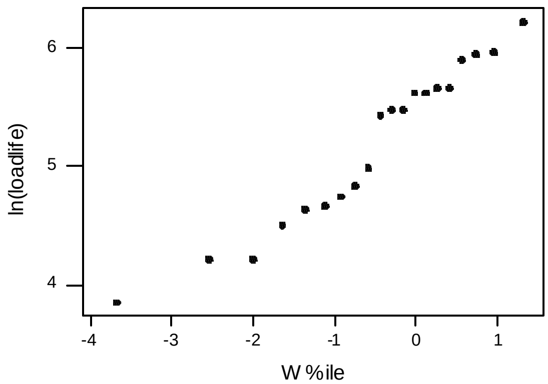


86.

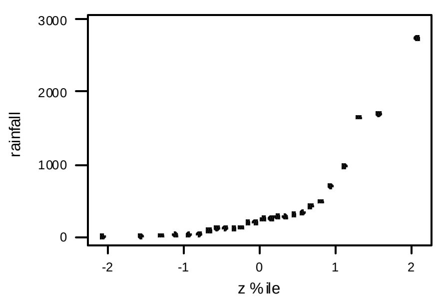
a. The 10 largest z percentiles are 1.96, 1.44, 1.15, .93, .76, .60, .45, .32, .19 and .06; the remaining 10 are the negatives of these values. The accompanying normal probability plot is reasonably straight. An assumption of population distribution normality is plausible.

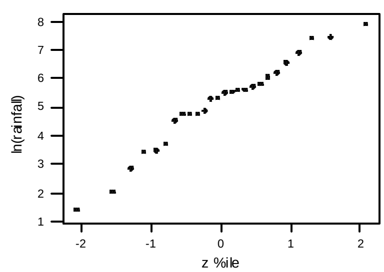


b. For a Weibull probability plot, the natural logs of the observations are plotted against extreme value percentiles; these percentiles are -3.68, -2.55, -2.01, -1.65, -1.37, -1.13, -.93, -.76, -.59, -.44, -.30, -.16, -.02, .12, .26, .40, .56, .73, .95, and 1.31. The accompanying probability plot is roughly as straight as the one for checking normality (a plot of ln(x) versus the z percentiles, appropriate for checking the plausibility of a lognormal distribution, is also reasonably straight - any of 3 different families of population distributions seems plausible.)



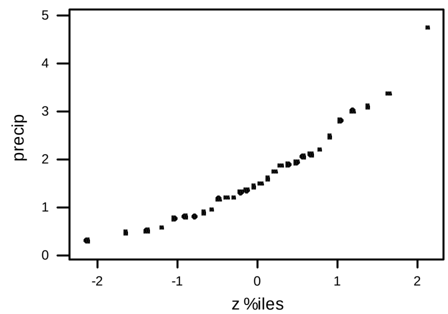
87. To check for plausibility of a lognormal population distribution for the rainfall data of Exercise 81 in Chapter 1, take the natural logs and construct a normal probability plot. This plot and a normal probability plot for the original data appear below. Clearly the log transformation gives quite a straight plot, so lognormality is plausible. The curvature in the plot for the original data implies a positively skewed population distribution - like the lognormal distribution.



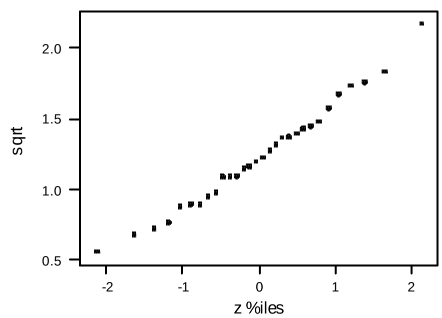


88.

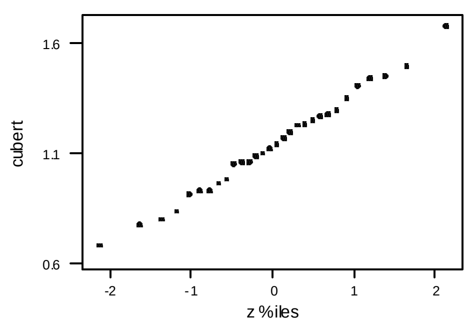
a. The plot of the original (untransformed) data appears somewhat curved.



b. The square root transformation results in a very straight plot. It is reasonable that this distribution is normally distributed.

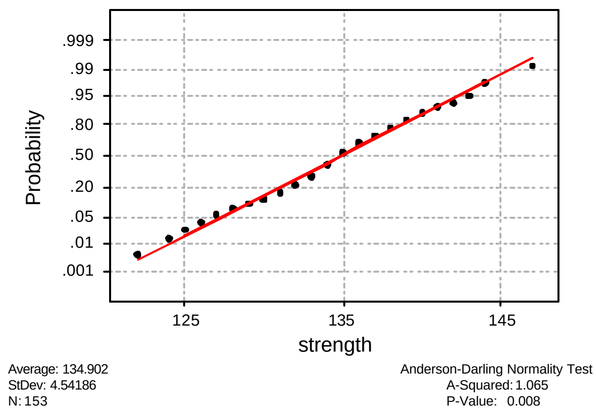


c. The cube root transformation also results in a very straight plot. It is very reasonable that the distribution is normally distributed.



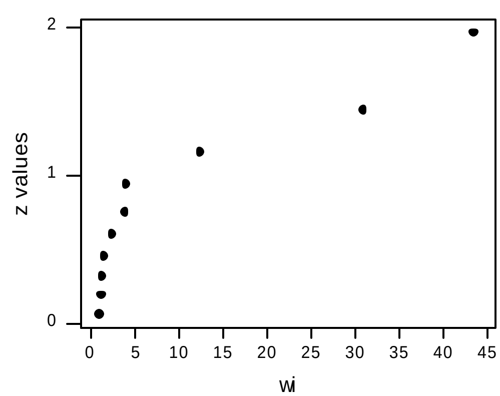
89. The pattern in the plot (below, generated by Minitab) is quite linear. It is very plausible that strength is normally distributed.

Normal Probability Plot



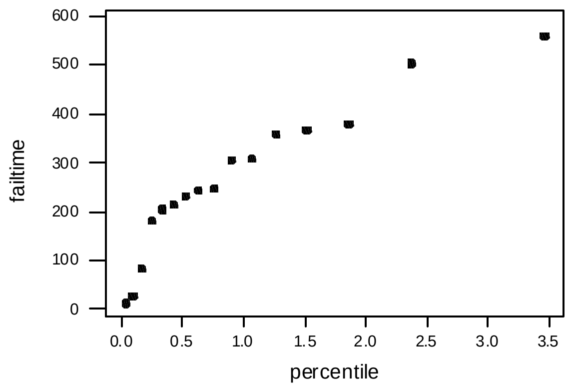
90. We use the data (table below) to create the desired plot.

ordered absolute values (w's)	probabilities	z values
0.89	0.525	0.063
1.15	0.575	0.19
1.27	0.625	0.32
1.44	0.675	0.454
2.34	0.725	0.6
3.78	0.775	0.755
3.96	0.825	0.935
12.38	0.875	1.15
30.84	0.925	1.44
43.4	0.975	1.96



This half-normal plot reveals some extreme values, without which the distribution may appear to be normal.

91. The $(100p)^{th}$ percentile $\eta(p)$ for the exponential distribution with $\lambda = 1$ satisfies $F(\eta(p)) = 1 - \exp[-\eta(p)] = p$, i.e., $\eta(p) = -\ln(1 - p)$. With n = 16, we need $\eta(p)$ for p = $\frac{5}{16}, \frac{1.5}{16}, \ldots, \frac{15.5}{16}$. These are .032, .398, .170, .247, .330, .421, .521, .633, .758, .901, 1.068, 1.269, 1.520, 1.856, 2.367, 3.466. this plot exhibits substantial curvature, casting doubt on the assumption of an exponential population distribution. Because λ is a scale parameter (as is σ for the normal family), $\lambda = 1$ can be used to assess the plausibility of the entire exponential family.



Supplementary Exercises

92.

a. $P(10 \le X \le 20) = \dfrac{10}{25} = .4$

b. $P(X \ge 10) = P(10 \le X \le 25) = \dfrac{15}{25} = .6$

c. For $0 \le X \le 25$, $F(x) = \int_0^x \dfrac{1}{25}\,dy = \dfrac{x}{25}$. F(x)=0 for x < 0 and = 1 for x > 25.

d. $E(X) = \dfrac{(A+B)}{2} = \dfrac{(0+25)}{2} = 12.5$; $Var(X) = \dfrac{(B-A)^2}{12} = \dfrac{625}{12} = 52.083$

93.

a. For $0 \le Y \le 25$, F(y) = $\frac{1}{24}\int_0^y \left(u - \frac{u^2}{12}\right) = \frac{1}{24}\left(\frac{u^2}{2} - \frac{u^3}{36}\right)\Bigg]_0^y$. Thus

$$F(y) = \begin{cases} 0 & y < 0 \\ \frac{1}{48}\left(y^2 - \frac{y^3}{18}\right) & 0 \le y \le 12 \\ 1 & y > 12 \end{cases}$$

b. $P(Y \le 4) = F(4) = .259$, $P(Y > 6) = 1 - F(6) = .5$
$P(4 \le X \le 6) = F(6) - F(4) = .5 - .259 = .241$

c. E(Y) = $\frac{1}{24}\int_0^{12} y^2\left(1 - \frac{y}{12}\right)dy = \frac{1}{24}\left[\frac{y^3}{3} - \frac{y^4}{48}\right]_0^{12} = 6$

$E(Y^2)$ = $\frac{1}{24}\int_0^{12} y^3\left(1 - \frac{y}{12}\right)dy = 43.2$, so V(Y) = 43.2 – 36 = 7.2

d. $P(Y < 4 \text{ or } Y > 8) = 1 - P(4 \le X \le 8) = .518$

e. the shorter segment has length min(Y, 12 – Y) so

E[min(Y, 12 – Y)] = $\int_0^{12} \min(y, 12-y)\cdot f(y)dy = \int_0^{6} \min(y, 12-y)\cdot f(y)dy$

$+ \int_6^{12} \min(y, 12-y)\cdot f(y)dy = \int_0^6 y \cdot f(y)dy + \int_6^{12} (12-y)\cdot f(y)dy = \frac{90}{24} = .3.75$

94.

a. Clearly $f(x) \ge 0$. The c.d.f. is , for x > 0,

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_0^x \frac{32}{(y+4)^3}dy = -\frac{1}{2}\cdot\frac{32}{(y+4)^2}\Bigg]_0^x = 1 - \frac{16}{(x+4)^2}$$

(F(x) = 0 for $x \le 0$.)

Since F(∞) = $\int_{-\infty}^{\infty} f(y)dy = 1$, f(x) is a legitimate pdf.

b. See above

c. $P(2 \le X \le 5) = F(5) - F(2) = 1 - \frac{16}{81} - \left(1 - \frac{16}{36}\right) = .247$

(continued)

d. $E(x) = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_{-\infty}^{\infty} x \cdot \frac{32}{(x+4)^3} dx = \int_0^{\infty} (x+4-4) \cdot \frac{32}{(x+4)^3} dx$

$= \int_0^{\infty} \frac{32}{(x+4)^2} dx - 4\int_0^{\infty} \frac{32}{(x+4)^3} dx = 8 - 4 = 4$

e. E(salvage value) = $= \int_0^{\infty} \frac{100}{x+4} \cdot \frac{32}{(y+4)^3} dx = 3200\int_0^{\infty} \frac{1}{(y+4)^4} dx = \frac{3200}{(3)(64)} = 16.67$

95.

a. By differentiation,

$$f(x) = \begin{cases} x^2 & 0 \le x < 1 \\ \frac{7}{4} - \frac{3}{4}x & 1 \le y \le \frac{7}{3} \\ 0 & otherwise \end{cases}$$

b. P(.5 ≤ X ≤ 2) = F(2) – F(.5) = $1 - \frac{1}{2}\left(\frac{7}{3} - 2\right)\left(\frac{7}{4} - \frac{3}{4} \cdot 2\right) - \frac{(.5)^3}{3} = \frac{11}{12} = .917$

c. E(X) = $\int_0^1 x \cdot x^2 dx + \int_1^{7/3} x \cdot \left(\frac{7}{4} - \frac{3}{4}x\right) dx = \frac{131}{108} = 1.213$

96. μ = 40 V; σ = 1.5 V

a. P(39 < X < 42) = $\Phi\left(\frac{42-40}{1.5}\right) - \Phi\left(\frac{39-40}{1.5}\right)$

$= \Phi(1.33) - \Phi(-.67) = .9082 - .2514 = .6568$

b. We desire the 85th percentile: 40 + (1.04)(1.5) = 41.56

c. P(X > 42) = 1 – P(X ≤ 42) = 1 $- \Phi\left(\frac{42-40}{1.5}\right)$ = 1 - Φ(1.33) = .0918

Let D represent the number of diodes out of 4 with voltage exceeding 42.

P(D ≥ 1) = 1 – P(D = 0) = $1 - \binom{4}{0}(.0918)^0(.9082)^4$ =1 - .6803 = .3197

97. $\mu = 137.2$ oz.; $\sigma = 1.6$ oz

a. $P(X > 135) = 1 - \Phi\left(\frac{135 - 137.2}{1.6}\right) = 1 - \Phi(-1.38) = 1 - .0838 = .9162$

b. With Y = the number among ten that contain more than 135 oz,
Y ~ Bin(10, .9162, so $P(Y \geq 8) = b(8; 10, .9162) + b(9; 10, .9162)$
$+ b(10; 10, .9162) = .9549$.

c. $\mu = 137.2$; $\frac{135 - 137.2}{s} = -1.65 \Rightarrow s = 1.33$

98.

a. Let S = defective. Then p = P(S) = .05; n = 250 $\Rightarrow \mu = np = 12.5$, $\sigma = 3.446$. The random variable X = the number of defectives in the batch of 250. X ~ Binomial. Since np = 12.5 ≥ 10, and nq = 237.5 ≥ 10, we can use the normal approximation.
$P(X_{bin} \geq 25) \approx 1 - \Phi\left(\frac{24.5 - 12.5}{3.446}\right) = 1 - \Phi(3.48) = 1 - .9997 = .0003$

b. $P(X_{bin} = 10) \approx P(X_{norm} \leq 10.5) - P(X_{norm} \leq 9.5)$
$= \Phi(-.58) - \Phi(-.87) = .2810 - .1922 = .0888$

99.

a. $P(X > 100) = 1 - \Phi\left(\frac{100 - 96}{14}\right) = 1 - \Phi(.29) = 1 - .6141 = .3859$

b. $P(50 < X < 80) = \Phi\left(\frac{80 - 96}{14}\right) - \Phi\left(\frac{50 - 96}{14}\right)$
$= \Phi(-1.5) - \Phi(-3.29) = .1271 - .0005 = .1266$.

c. a = 5^{th} percentile = 96 + (-1.645)(14) = 72.97.
b = 95^{th} percentile = 96 + (1.645)(14) = 119.03. The interval (72.97, 119.03) contains the central 90% of all grain sizes.

100.

a. F(X) = 0 for x < 1 and = 1 for x > 3. For $1 \le x \le 3$, $F(x) = \int_{-\infty}^{x} f(y)dy$

$$= \int_{-\infty}^{1} 0dy + \int_{1}^{x} \frac{3}{2} \cdot \frac{1}{y^2} dy = 1.5\left(1 - \frac{1}{x}\right)$$

b. $P(X \le 2.5) = F(2.5) = 1.5(1 - .4) = .9$; $P(1.5 \le x \le 2.5) =$ $F(2.5) - F(1.5) = .4$

c. E(X) = $= \int_{1}^{3} x \cdot \frac{3}{2} \cdot \frac{1}{x^2} dx = \frac{3}{2} \int_{1}^{3} \frac{1}{x} dx = 1.5 \ln(x)\Big]_1^3 = 1.648$

d. $E(X^2) = = \int_{1}^{3} x^2 \cdot \frac{3}{2} \cdot \frac{1}{x^2} dx = \frac{3}{2} \int_{1}^{3} dx = 3$, so $V(X) = E(X^2) - [E(X)]^2 = .284$,

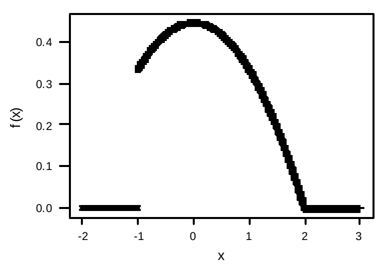
$\sigma = .553$

e. $h(x) = \begin{cases} 0 & 1 \le x \le 1.5 \\ x - 1.5 & 1.5 \le x \le 2.5 \\ 1 & 2.5 \le x \le 3 \end{cases}$

so E[h(X)] = $= \int_{1.5}^{2.5} (x - 1.5) \cdot \frac{3}{2} \cdot \frac{1}{x^2} dx + \int_{2.5}^{3} 1 \cdot \frac{3}{2} \cdot \frac{1}{x^2} dx = .267$

101.

a.



b. F(x) = 0 for x < -1 or == 1 for x > 2. For $-1 \le x \le 2$,

$$F(x) = \int_{-1}^{x} \frac{1}{9}(4 - y^2)dy = \frac{1}{9}\left(4x - \frac{x^3}{3}\right) + \frac{11}{27}$$

c. The median is 0 iff F(0) = .5. Since F(0) = $\frac{11}{27}$, this is not the case. Because $\frac{11}{27} < .5$, the median must be greater than 0.

d. Y is a binomial r.v. with n = 10 and p = P(X > 1) = 1 – F(1) = $\frac{5}{27}$

102.

a. E(X) = $\dfrac{1}{l}$ = 1.075, $s = \dfrac{1}{l}$ = 1.075

b. P(3.0 < X) = 1 – P(X ≤ 3.0) = 1 – F(3.0) = $3^{-.93(3.0)}$ = .0614
P(1.0 ≤ X ≤ 3.0) = F(3.0) – F(1.0) = .333

c. The 90th percentile is requested; denoting it by c, we have
.9 = F(c) = 1 – $e^{-(.93)c}$, whence c = $\dfrac{\ln(.1)}{(-.93)} = 2.476$

103.

a. P(X ≤ 150) = $\exp\left[-\exp\left(\dfrac{-(150-150)}{90}\right)\right] = \exp[-\exp(0)] = \exp(-1) = .368$, where exp(u) = e^u. P(X ≤ 300) = $\exp[-\exp(-1.6667)] = .828$,
and P(150 ≤ X ≤ 300) = .828 - .368 = .460.

b. The desired value c is the 90th percentile, so c satisfies
.9 = $\exp\left[-\exp\left(\dfrac{-(c-150)}{90}\right)\right]$. Taking the natural log of each side twice in succession yields ln[ln(.9)] = $\dfrac{-(c-150)}{90}$, so c = 90(2.250367) + 150 = 352.53.

c. f(x) = F′(X) = $\dfrac{1}{b}\cdot\exp\left[-\exp\left(\dfrac{-(x-a)}{b}\right)\right]\cdot\exp\left(\dfrac{-(x-a)}{b}\right)$

d. We wish the value of x for which f(x) is a maximum; this is the same as the value of x for which ln[f(x)] is a maximum. The equation of $\dfrac{d[\ln(f(x))]}{dx} = 0$ gives
$\exp\left(\dfrac{-(x-a)}{b}\right) = 1$, so $\dfrac{-(x-a)}{b} = 0$, which implies that x = α. Thus the mode is α.

e. E(X) = .5772β + α = 201.95, whereas the mode is 150 and the median is
–(90)ln[-ln(.5)] + 150 = 182.99. The distribution is positively skewed.

104.

a. E(cX) = cE(X) = $\dfrac{c}{l}$

b. E[c(1 - .5e^{ax})] = $\int_0^\infty c\left(1-.5e^{ax}\right)\cdot l e^{-lx}dx = \dfrac{c[.5l-a]}{l-a}$

105.

a. From a graph of f(x; μ, σ) or by differentiation, x* = μ.

b. No; the density function has constant height for A ≤ X ≤ B.

c. F(x; λ) is largest for x = 0 (the derivative at 0 does not exist since f is not continuous there) so x* = 0.

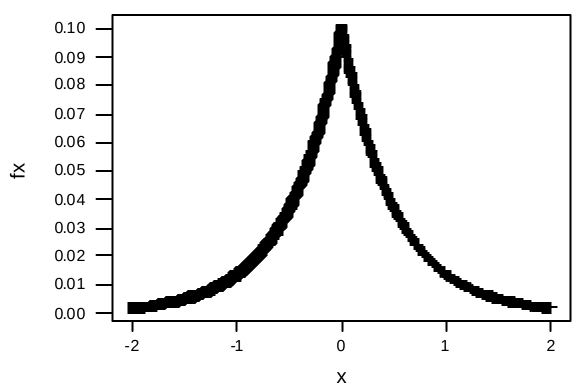
d. $\ln[f(x;a,b)] = -\ln(b^a) - \ln(\Gamma(a)) + (a-1)\ln(x) - \dfrac{x}{b;}$

$$\frac{d}{dx}\ln[f(x;a,b)] = \frac{a-1}{x} - \frac{1}{b} \Rightarrow x = x^* = (a-1)b$$

e. From **d** $x^* = \left(\dfrac{n}{2} - 1\right)(2) = n - 2.$

106.

a. $\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} .1e^{.2x}dx + \int_{0}^{\infty} .1e^{-.2x}dx = .5 + .5 = 1$



b. For x < 0, F(x) = $\int_{-\infty}^{x} .1e^{.2y}dy = \dfrac{1}{2}e^{.2x}$.

For x ≥ 0, F(x) = $\dfrac{1}{2} + \int_{0}^{x} .1e^{-.2y}dy = 1 - \dfrac{1}{2}e^{-.2x}$.

c. P(X < 0) = F(0) = $\dfrac{1}{2} = .5$, P(X < 2) = F(2) = 1 - .5e$^{-.4}$ = .665,

P(-1 ≤ X ≤ 2) – F(2) – F(-1) = .256, 1 - (-2 ≤ X ≤ 2) = .670

107.

a. Clearly f(x; λ_1, λ_2, p) ≥ 0 for all x, and $\int_{-\infty}^{\infty} f(x;\lambda_1,\lambda_2,p)dx$

$= \int_0^{\infty}\left[p\lambda_1 e^{-\lambda_1 x} + (1-p)\lambda_2 e^{-\lambda_2 x}\right]dx = p\int_0^{\infty}\lambda_1 e^{-\lambda_1 x}dx + (1-p)\int_0^{\infty}\lambda_2 e^{-\lambda_2 x}dx$

= p + (1 – p) = 1

b. For x > 0, F(x; λ_1, λ_2, p) = $\int_0^{x} f(y;\lambda_1,\lambda_2,p)dy = p(1-e^{-\lambda_1 x}) + (1-p)(1-e^{-\lambda_2 x})$.

c. E(X) = $\int_0^{\infty} x\cdot\left[p\lambda_1 e^{-\lambda_1 x}) + (1-p)\lambda_2 e^{-\lambda_2 x})\right]dx$

$$= p\int_0^{\infty} x\lambda_1 e^{-\lambda_1 x}dx + (1-p)\int_0^{\infty} x\lambda_2 e^{-\lambda_2 x}dx = \frac{p}{\lambda_1} + \frac{(1-p)}{\lambda_2}$$

d. E(X^2) = $\frac{2p}{\lambda_1^2} + \frac{2(1-p)}{\lambda_2^2}$, so Var(X) = $\frac{2p}{\lambda_1^2} + \frac{2(1-p)}{\lambda_2^2} - \left[\frac{p}{\lambda_1} + \frac{(1-p)}{\lambda_2}\right]^2$

e. For an exponential r.v., CV = $\frac{1/\lambda}{1/\lambda} = 1$. For X hyperexponential,

$$\text{CV} = \left[\frac{\frac{2p}{\lambda_1^2} + \frac{2(1-p)}{\lambda_2^2}}{\left[\frac{p}{\lambda_1} + \frac{(1-p)}{\lambda_2}\right]^2} - 1\right]^{1/2} = \left[\frac{2\left(p\lambda_2^2 + (1-p)\lambda_1^2\right)}{\left(p\lambda_2 + (1-p)\lambda_1\right)^2} - 1\right]^{1/2}$$

= $[2r - 1]^{1/2}$ where r = $\frac{\left(p\lambda_2^2 + (1-p)\lambda_1^2\right)}{\left(p\lambda_2 + (1-p)\lambda_1\right)^2}$. But straightforward algebra shows that r > 1 provided $\lambda_1 \neq \lambda_2$, so that CV > 1.

f. $\mu = \frac{n}{\lambda}$, $\sigma^2 = \frac{n}{\lambda^2}$, so $\sigma = \frac{\sqrt{n}}{\lambda}$ and CV = $\frac{1}{\sqrt{n}} < 1$ if n > 1.

108.

a. $1 = \int_5^\infty \frac{k}{x^a}dx = k \cdot \frac{5^{1-a}}{a-1} \Rightarrow k = (a-1)5^{1-a}$ where we must have α > 1.

b. For x ≥ 5, F(x) = $\int_5^x \frac{k}{y^a}dy = 5^{1-a}\left[\frac{1}{5^{1-a}} - \frac{1}{x^{a-1}}\right] = 1 - \left(\frac{5}{x}\right)^{a-1}$.

c. E(X) = $\int_5^\infty x \cdot \frac{k}{x^a}dx = \int_5^\infty x \cdot \frac{k}{x^{a-1}}dx = \frac{k}{5^{a-2} \cdot (a-2)}$, provided α > 2.

d. $P\left(\ln\left(\frac{X}{5}\right) \le y\right) = P\left(\frac{X}{5} \le e^y\right) = P(X \le 5e^y) = F(5e^y) = 1 - \left(\frac{5}{5e^y}\right)^{a-1}$
$1 - e^{-(a-1)y}$, the cdf of an exponential r.v. with parameter α - 1.

109.

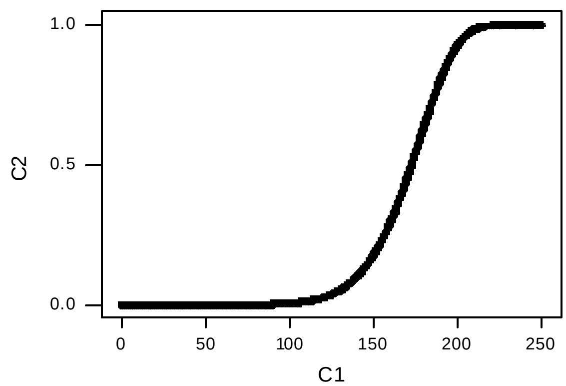
a. A lognormal distribution, since $\ln\left(\frac{I_o}{I_i}\right)$ is a normal r.v.

b. $P(I_o > 2I_i) = P\left(\frac{I_o}{I_i} > 2\right) = P\left(\ln\left(\frac{I_o}{I_i}\right) > \ln 2\right) = 1 - P\left(\ln\left(\frac{I_o}{I_i}\right) \le \ln 2\right)$
$1 - \Phi\left(\frac{\ln 2 - 1}{.05}\right) = 1 - \Phi(-6.14) = 1$

c. $E\left(\frac{I_o}{I_i}\right) = e^{1+.0025/2} = 2.72,\ Var\left(\frac{I_o}{I_i}\right) = e^{2+.0025} \cdot \left(e^{.0025} - 1\right) = .0185$

110.

a.



b. P(X > 175) = 1 – F(175; 9, 180) = $e^{-\left(\frac{175}{180}\right)^9} = .4602$
P(150 ≤ X ≤ 175) = F(175; 9, 180) - F(150; 9, 180)
= .5398 - .1762 = .3636

c. P(at least one) = 1 – P(none) = 1 – $(1 - .3636)^2$ = .5950

d. We want the 10th percentile: .10 = F(x; 9, 180) = $1 - e^{-\left(\frac{x}{180}\right)^9}$. A small bit of algebra leads us to x = 140.178. Thus 10% of all tensile strengths will be less than 140.178 MPa.

111. F(y) = P(Y ≤ y) = P(σZ + μ ≤ y) = $P\left(Z \le \frac{(y-\mu)}{\sigma}\right) = \int_{-\infty}^{\frac{(y-\mu)}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\,dz$. Now differentiate with respect to y to obtain a normal pdf with parameters μ and σ.

112.

a. $F_Y(y)$ = P(Y ≤ y) = P(60X ≤ y) = $P\left(X \le \frac{y}{60}\right) = F\left(\frac{y}{60\beta};\alpha\right)$. Thus $f_Y(y)$

$= f\left(\frac{y}{60\beta};\alpha\right)\cdot\frac{1}{60\beta} = \frac{y^{\alpha-1}e^{\frac{-y}{60\beta}}}{(60\beta)^{\alpha}\,\Gamma(\alpha)}$, which shows that Y has a gamma distribution with parameters α and 60β.

b. With c replacing 60 in **a**, the same argument shows that cX has a gamma distribution with parameters α and cβ.

113.

a. Y = -ln(X) $\Rightarrow$ x = e^{-y} = k(y), so k′(y) = $-e^{-y}$. Thus since f(x) = 1,
g(y) = 1 · | $-e^{-y}$ | = e^{-y} for 0 < y < ∞, so y has an exponential distribution with parameter λ = 1.

b. y = σZ + μ $\Rightarrow$ y = h(z) = σZ + μ $\Rightarrow$ z = k(y) = $\frac{(y-m)}{s}$ and k′(y) = $\frac{1}{s}$, from which the result follows easily.

c. y = h(x) = cx $\Rightarrow$ x = k(y) = $\frac{y}{c}$ and k′(y) = $\frac{1}{c}$, from which the result follows easily.

114.

a. If we let $a = 2$ and $b = \sqrt{2}s$, then we can manipulate f(v) as follows:

$$f(n) = \frac{n}{s^2} e^{-n^2/2s^2} = \frac{2}{2s^2} n e^{-n^2/2s^2} = \frac{2}{\left(\sqrt{2}s\right)^2} n^{2-1} e^{-\left(n/\sqrt{2}s\right)^2} = \frac{a}{b^a} n^{a-1} e^{-\left(n/b\right)^a},$$

which is in the Weibull family of distributions.

b. $F(n) = \int_0^{25} \frac{n}{400} e^{\frac{-n}{800}} dn$; cdf: $F\left(n;2,\sqrt{2}s\right) = 1 - e^{-\left(\frac{n}{\sqrt{2}s}\right)} = 1 - e^{\frac{-v^2}{800}}$, so

$$F\left(25;2,\sqrt{2}\right) = 1 - e^{\frac{-625}{800}} = 1 - .458 = .542$$

115.

a. Assuming independence, P(all 3 births occur on March 11) = $\left(\frac{1}{365}\right)^3 = .00000002$

b. $\left(\frac{1}{365}\right)^3 (365) = .0000073$

c. Let X = deviation from due date. X~N(0, 19.88). Then the baby due on March 15 was 4 days early. P(x = -4) ˜ P(-4.5 < x < -3.5)

$$= \Phi\left(\frac{-3.5}{19.88}\right) - \Phi\left(\frac{-4.5}{19.88}\right) = \Phi(-.18) - \Phi(-.237) = .4286 - .4090 = .0196 .$$

Similarly, the baby due on April 1 was 21 days early, and P(x = -21)

$$˜ \ \Phi\left(\frac{-20.5}{19.88}\right) - \Phi\left(\frac{-21.5}{19.88}\right) = \Phi(-1.03) - \Phi(-1.08) = .1515 - .1401 = .0114 .$$

The baby due on April 4 was 24 days early, and P(x = -24) ˜ .0097

Again, assuming independence, P(all 3 births occurred on March 11) =
$(.0196)(.0114)(.0097) = .00002145$

d. To calculate the probability of the three births happening on any day, we could make similar calculations as in part c for each possible day, and then add the probabilities.

116.

a. F(x) = le^{-lx} and F(x) = $1 - e^{-lx}$, so r(x) = $\dfrac{l e^{-lx}}{e^{-lx}} = l$, a constant (independent of X); this is consistent with the memoryless property of the exponential distribution.

b. r(x) = $\left(\dfrac{a}{b^{a}}\right)x^{a-1}$; for α > 1 this is increasing, while for α < 1 it is a decreasing function.

c. ln(1 – F(x)) = $-\int a\left(1-\dfrac{x}{b}\right)dx = -a\left[x - \dfrac{x^2}{2b}\right] \Rightarrow F(x) = 1 - e^{-a\left(x - \frac{x^2}{2b}\right)}$,

f(x) = $a\left(1-\dfrac{x}{b}\right)e^{-a\left(x-\frac{x^2}{2b}\right)}$ $\qquad 0 \le x \le \beta$

117.

a. $F_X(x) = P\left(-\dfrac{1}{l}\ln(1-U) \le x\right) = P(\ln(1-U) \ge -lx) = P(1 - U \ge e^{-lx})$

$= P(U \le 1 - e^{-lx}) = 1 - e^{-lx}$ since $F_U(u) = u$ (U is uniform on [0, 1]). Thus X has an exponential distribution with parameter λ.

b. By taking successive random numbers u_1, u_2, u_3, …and computing $x_i = -\dfrac{1}{10}\ln(1-u_i)$, … we obtain a sequence of values generated from an exponential distribution with parameter λ = 10.

118.

a. E(g(X)) ≈ E[g(μ) + g′(μ)(X - μ)] = E(g(μ)) + g′(μ)·E(X - μ), but E(X) - μ = 0 and E(g(μ)) = g(μ) (since g(μ) is constant), giving E(g(X)) ≈ g(μ).
V(g(X)) ≈ V[g(μ) + g′(μ)(X - μ)] = V[g′(μ)(X - μ)] = $(g′(μ))^2$·V(X - μ) = $(g′(μ))^2$·V(X).

b. $g(I) = \dfrac{v}{I}, g'(I) = \dfrac{-v}{I^2}$, so $E(g(I)) = m_R \approx \dfrac{v}{m_I} = \dfrac{v}{20}$

$$V(g(I)) \approx \left(\frac{-v}{m_I^2}\right)^2 \cdot V(I), s_{g(I)} \approx \frac{v}{20^2} \cdot s_I = \frac{v}{800}$$

119. g(μ) + g′(μ)(X - μ) ≤ g(X) implies that E[g(μ) + g′(μ)(X - μ)] = E(g(μ)) = g(μ) ≤ E(g(X)), i.e. that g(E(X)) ≤ E(g(X)).

120. For $y > 0$, $F(y) = P(Y \le y) = P\left(\frac{2X^2}{b^2} \le y\right) = P\left(X^2 \le \frac{b^2 y}{2}\right) = P\left(X \le \frac{b\sqrt{y}}{\sqrt{2}}\right)$. Now take the cdf of X (Weibull), replace x by $\frac{b\sqrt{y}}{\sqrt{2}}$, and then differentiate with respect to y to obtain the desired result $f_Y(y)$.

CHAPTER 5

Section 5.1

1.

a. $P(X = 1, Y = 1) = p(1,1) = .20$

b. $P(X \le 1 \text{ and } Y \le 1) = p(0,0) + p(0,1) + p(1,0) + p(1,1) = .42$

c. At least one hose is in use at both islands. $P(X \ne 0 \text{ and } Y \ne 0) = p(1,1) + p(1,2) + p(2,1) + p(2,2) = .70$

d. By summing row probabilities, $p_x(x) = .16, .34, .50$ for $x = 0, 1, 2$, and by summing column probabilities, $p_y(y) = .24, .38, .38$ for $y = 0, 1, 2$. $P(X \le 1) = p_x(0) + p_x(1) = .50$

e. $P(0,0) = .10$, but $p_x(0) \cdot p_y(0) = (.16)(.24) = .0384 \ne .10$, so X and Y are not independent.

2.

a.

	p(x,y)	y: 0	1	2	3	4	
	0	.30	.05	.025	.025	.10	.5
x	1	.18	.03	.015	.015	.06	.3
	2	.12	.02	.01	.01	.04	.2
		.6	.1	.05	.05	.2	

b. $P(X \le 1 \text{ and } Y \le 1) = p(0,0) + p(0,1) + p(1,0) + p(1,1) = .56$
$= (.8)(.7) = P(X \le 1) \cdot P(Y \le 1)$

c. $P(X + Y = 0) = P(X = 0 \text{ and } Y = 0) = p(0,0) = .30$

d. $P(X + Y \le 1) = p(0,0) + p(0,1) + p(1,0) = .53$

3.

a. $p(1,1) = .15$, the entry in the 1st row and 1st column of the joint probability table.

b. $P(X_1 = X_2) = p(0,0) + p(1,1) + p(2,2) + p(3,3) = .08+.15+.10+.07 = .40$

c. $A = \{ (x_1, x_2): x_1 \ge 2 + x_2 \} \cup \{ (x_1, x_2): x_2 \ge 2 + x_1 \}$
$P(A) = p(2,0) + p(3,0) + p(4,0) + p(3,1) + p(4,1) + p(4,2) + p(0,2) + p(0,3) + p(1,3) =.22$

d. $P(\text{ exactly } 4) = p(1,3) + p(2,2) + p(3,1) + p(4,0) = .17$
$P(\text{at least } 4) = P(\text{exactly } 4) + p(4,1) + p(4,2) + p(4,3) + p(3,2) + p(3,3) + p(2,3)=.46$

4.

a. $P_1(0) = P(X_1 = 0) = p(0,0) + p(0,1) + p(0,2) + p(0,3) = .19$
$P_1(1) = P(X_1 = 1) = p(1,0) + p(1,1) + p(1,2) + p(1,3) = .30$, etc.

x_1	0	1	2	3	4
$p_1(x_1)$	.19	.30	.25	.14	.12

b. $P_2(0) = P(X_2 = 0) = p(0,0) + p(1,0) + p(2,0) + p(3,0) + p(4,0) = .19$, etc

x_2	0	1	2	3
$p_2(x_2)$	.19	.30	.28	.23

c. $p(4,0) = 0$, yet $p_1(4) = .12 > 0$ and $p_2(0) = .19 > 0$, so $p(x_1 , x_2) \neq p_1(x_1) \cdot p_2(x_2)$ for every (x_1 , x_2), and the two variables are not independent.

5.

a. P(X = 3, Y = 3) = P(3 customers, each with 1 package)
= P(each has 1 package | 3 customers) · P(3 customers)
= $(.6)^3 \cdot (.25) = .054$

b. P(X = 4, Y = 11) = P(total of 11 packages | 4 customers) · P(4 customers)
Given that there are 4 customers, there are 4 different ways to have a total of 11 packages: 3, 3, 3, 2 or 3, 3, 2, 3 or 3, 2, 3 ,3 or 2, 3, 3, 3. Each way has probability $(.1)^3(.3)$, so $p(4, 11) = 4(.1)^3(.3)(.15) = .00018$

6.

a. $p(4,2) = P(Y = 2 \mid X = 4) \cdot P(X = 4) = \left[\binom{4}{2}(.6)^2(.4)^2\right] \cdot (.15) = .0518$

b. $P(X = Y) = p(0,0) + p(1,1) + p(2,2) + p(3,3) + p(4,4) = .1+(.2)(.6) + (.3)(.6)^2 + (.25)(.6)^3 + (.15)(.6)^4 = .4014$

c. p(x,y) = 0 unless y = 0, 1, …, x; x = 0, 1, 2, 3, 4. For any such pair,

p(x,y) = P(Y = y | X = x) · P(X = x) = $\binom{x}{y}(.6)^y(.4)^{x-y}\cdot p_x(x)$

$p_y(4) = p(y = 4) = p(x = 4, y = 4) = p(4,4) = (.6)^4\cdot(.15) = .0194$

$p_y(3) = p(3,3) + p(4,3) = (.6)^3(.25)+\binom{4}{3}(.6)^3(.4)(.15) = .1058$

$p_y(2) = p(2,2) + p(3,2) + p(4,2) = (.6)^2(.3)+\binom{3}{2}(.6)^2(.4)(.25)$

$$+\binom{4}{2}(.6)^2(.4)^2(.15) = .2678$$

$p_y(1) = p(1,1) + p(2,1) + p(3,1) + p(4,1) = (.6)(.2)+\binom{2}{1}(.6)(.4)(.3)$

$$\binom{3}{1}(.6)(.4)^2(.25)+\binom{4}{1}(.6)(.4)^3(.15) = .3590$$

$p_y(0) = 1 - [.3590+.2678+.1058+.0194] = .2480$

7.

a. p(1,1) = .030

b. P(X ≤ 1 and Y ≤ 1 = p(0,0) + p(0,1) + p(1,0) + p(1,1) = .120

c. P(X = 1) = p(1,0) + p(1,1) + p(1,2) = .100; P(Y = 1) = p(0,1) + … + p(5,1) = .300

d. P(overflow) = P(X + 3Y > 5) = 1 – P(X + 3Y ≤ 5) = 1 – P[(X,Y)=(0,0) or …or (5,0) or (0,1) or (1,1) or (2,1)] = 1 - .620 = .380

e. The marginal probabilities for X (row sums from the joint probability table) are $p_x(0)$ = .05, $p_x(1)$ = .10 , $p_x(2)$ = .25, $p_x(3)$ = .30, $p_x(4)$ = .20, $p_x(5)$ = .10; those for Y (column sums) are $p_y(0)$ = .5, $p_y(1)$ = .3, $p_y(2)$ = .2. It is now easily verified that for every (x,y), $p(x,y) = p_x(x)\cdot p_y(y)$, so X and Y are independent.

8.

a. numerator = $\binom{8}{3}\binom{10}{2}\binom{12}{1} = (56)(45)(12) = 30{,}240$

denominator = $\binom{30}{6} = 593{,}775$; p(3,2) = $\dfrac{30{,}240}{593{,}775} = .0509$

b. p(x,y) = $\begin{cases} \dfrac{\binom{8}{x}\binom{10}{y}\binom{12}{6-(x+y)}}{\binom{30}{6}} & x, y\_are\_non-negative \\ & \text{int } egers\_such\_that \\ & 0 \le x+y \le 6 \\ 0 & otherwise \end{cases}$

9.

a. $1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = \int_{20}^{30}\int_{20}^{30} K(x^2 + y^2)dxdy$

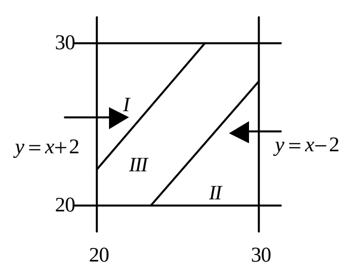
$= K\int_{20}^{30}\int_{20}^{30} x^2 dydx + K\int_{20}^{30}\int_{20}^{30} y^2 dxdy = 10K\int_{20}^{30} x^2 dx + 10K\int_{20}^{30} y^2 dy$

$= 20K \cdot \left(\dfrac{19{,}000}{3}\right) \Rightarrow K = \dfrac{3}{380{,}000}$

b. P(X < 26 and Y < 26) = $\int_{20}^{26}\int_{20}^{26} K(x^2 + y^2)dxdy = 12K\int_{20}^{26} x^2 dx$

$4Kx^3\Big|_{20}^{26} = 38{,}304K = .3024$

c.



P(| X – Y | ≤ 2) = $\iint_{\substack{region \\ III}} f(x,y)dxdy$

$1 - \iint_{I} f(x,y)dxdy - \iint_{II} f(x,y)dxdy$

$1 - \int_{20}^{28}\int_{x+2}^{30} f(x,y)dydx - \int_{22}^{30}\int_{20}^{x-2} f(x,y)dydx$

= (after much algebra) .3593

d. $f_x(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{20}^{30} K(x^2 + y^2)dy = 10Kx^2 + K\frac{y^3}{3}\Big|_{20}^{30}$

$= 10Kx^2 + .05, \qquad 20 \le x \le 30$

e. $f_y(y)$ is obtained by substituting y for x in (d); clearly $f(x,y) \ne f_x(x) \cdot f_y(y)$, so X and Y are not independent.

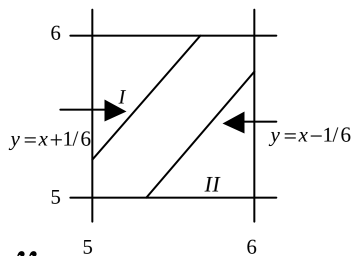
10.

a. $f(x,y) = \begin{cases} 1 & 5 \le x \le 6, 5 \le y \le 6 \\ 0 & otherwise \end{cases}$

since $f_x(x) = 1$, $f_y(y) = 1$ for $5 \le x \le 6$, $5 \le y \le 6$

b. $P(5.25 \le X \le 5.75, 5.25 \le Y \le 5.75) = P(5.25 \le X \le 5.75) \cdot P(5.25 \le Y \le 5.75)$ = (by independence) (.5)(.5) = .25

c.



$P((X,Y) \in A) = \iint_A 1dxdy$

= area of A = 1 – (area of I + area of II)

$= 1 - \frac{25}{36} = \frac{11}{36} = .306$

11.

a. $p(x,y) = \frac{e^{-l} l^x}{x!} \cdot \frac{e^{-m} m^y}{y!}$ for x = 0, 1, 2, …; y = 0, 1, 2, …

b. $p(0,0) + p(0,1) + p(1,0) = e^{-l-m}\left[1 + l + m\right]$

c. $P(X+Y=m) = \sum_{k=0}^{m} P(X = k, Y = m-k) = \sum_{k=0}^{m} e^{-l-m} \frac{l^k}{k!} \frac{m^{m=k}}{(m-k)!}$

$\frac{e^{-(l+m)}}{m!} \sum_{k=0}^{m} \binom{m}{k} l^k m^{m-k} = \frac{e^{-(l+m)}(l+m)^m}{m!}$, so the total # of errors X+Y also has a Poisson distribution with parameter $l + m$.

12.

a. P(X> 3) = $\int_3^\infty \int_0^\infty xe^{-x(1+y)}dydx = \int_3^\infty e^{-x}dx = .050$

b. The marginal pdf of X is $\int_0^\infty xe^{-x(1+y)}dy = e^{-x}$ for $0 \le x$; that of Y is $\int_3^\infty xe^{-x(1+y)}dx = \dfrac{1}{(1+y)^2}$ for $0 \le y$. It is now clear that f(x,y) is not the product of the marginal pdf's, so the two r.v's are not independent.

c. P(at least one exceeds 3) = 1 – P(X ≤ 3 and Y ≤ 3)

$$=1-\int_0^3\int_0^3 xe^{-x(1+y)}dydx = 1-\int_0^3\int_0^3 xe^{-x}e^{-xy}dy$$

$$=1-\int_0^3 e^{-x}(1-e^{-3x})dx = e^{-3}+.25-.25e^{-12} = .300$$

13.

a. f(x,y) = $f_x(x) \cdot f_y(y)$ = $\begin{cases} e^{-x-y} & x \ge 0, y \ge 0 \\ 0 & otherwise \end{cases}$

b. P(X ≤ 1 and Y ≤ 1) = P(X ≤ 1) · P(Y ≤ 1) = $(1 - e^{-1})(1 - e^{-1})$ = .400

c. P(X + Y ≤ 2) = $\int_0^2\int_0^{2-x} e^{-x-y}dydx = \int_0^2 e^{-x}\left[1-e^{-(2-x)}\right]dx$

$$= \int_0^2 (e^{-x}-e^{-2})dx = 1-e^{-2}-2e^{-2} = .594$$

d. P(X + Y ≤ 1) = $\int_0^1 e^{-x}\left[1-e^{-(1-x)}\right]dx = 1-2e^{-1} = .264$,

so P(1 ≤ X + Y ≤ 2) = P(X + Y ≤ 2) – P(X + Y ≤ 1) = .594 - .264 = .330

14.

a. P(X_1 < t, X_2 < t, … , X_{10} < t) = P(X_1 < t) … P(X_{10} < t) = $(1-e^{-lt})^{10}$

b. If "success" = {fail before t}, then p = P(success) = $1-e^{-lt}$,

and P(k successes among 10 trials) = $\binom{10}{k}1-e^{-lt^k}(e^{-lt})^{10-k}$

c. P(exactly 5 fail) = P(5 of l's fail and other 5 don't) + P(4 of l's fail, m fails, and other 5 don't) = $\binom{9}{5}(1-e^{-lt})^5(e^{-lt})^4(e^{-mt}) + \binom{9}{4}(1-e^{-lt})^4(1-e^{-mt})(e^{-lt})^5$

15.

a. F(y) = P(Y ≤ y) = P [(X_1 ≤y) ∪ ((X_2 ≤ y) ∩ (X_3 ≤ y))]
= P (X_1 ≤ y) + P[(X_2 ≤ y) ∩ (X_3 ≤ y)] - P[(X_1 ≤ y) ∩ (X_2 ≤ y) ∩ (X_3 ≤ y)]
= $(1-e^{-ly}) + (1-e^{-ly})^2 - (1-e^{-ly})^3$ for y ≥ 0

b. f(y) = F′(y) = $le^{-ly} + 2(1-e^{-ly})\left(le^{-ly}\right) - 3(1-e^{-ly})^2\left(le^{-ly}\right)$
= $4le^{-2ly} - 3le^{-3ly}$ for y ≥ 0

E(Y) = $\int_0^\infty y \cdot \left(4le^{-2ly} - 3le^{-3ly}\right)dy = 2\left(\frac{1}{2l}\right) - \frac{1}{3l} = \frac{2}{3l}$

16.

a. f(x_1, x_3) = $\int_{-\infty}^{\infty} f(x_1, x_2, x_3)dx_2 = \int_0^{1-x_1-x_3} kx_1x_2(1-x_3)dx_2$
$72x_1(1-x_3)(1-x_1-x_3)^2$ 0 ≤ x_1, 0 ≤ x_3, x_1 + x_3 ≤ 1

b. P(X_1 + X_3 ≤ .5) = $\int_0^{.5}\int_0^{.5-x_1} 72x_1(1-x_3)(1-x_1-x_3)^2\,dx_2dx_1$
= (after much algebra) .53125

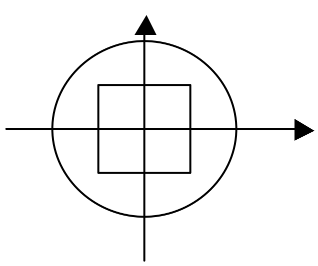
c. $f_{x_1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_3)dx_3 = \int 72x_1(1-x_3)(1-x_1-x_3)^2\,dx_3$
$18x_1 - 48x_1^2 + 36x_1^3 - 6x_1^5$ 0 ≤ x_1 ≤ 1

17.

a. $P\left((X,Y)\right.$ within a circle of radius $\left.\frac{R}{2}\right) = P(A) = \iint_A f(x,y)dxdy$

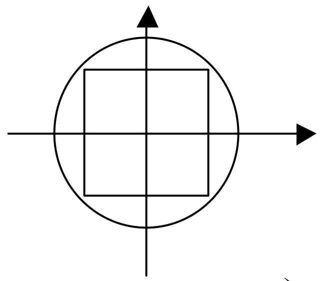
$$= \frac{1}{pR^2}\iint_A dxdy = \frac{area.of.A}{pR^2} = \frac{1}{4} = .25$$

b.



$$P\left(-\frac{R}{2} \le X \le \frac{R}{2}, -\frac{R}{2} \le Y \le \frac{R}{2}\right) = \frac{R^2}{pR^2} = \frac{1}{p}$$

c.



$$P\left(-\frac{R}{\sqrt{2}} \le X \le \frac{R}{\sqrt{2}}, -\frac{R}{\sqrt{2}} \le Y \le \frac{R}{\sqrt{2}}\right) = \frac{2R^2}{pR^2} = \frac{2}{p}$$

d. $f_x(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{pR^2} dy = \frac{2\sqrt{R^2-x^2}}{pR^2}$ for $-R \le x \le R$ and similarly for $f_Y(y)$. X and Y are not independent since e.g. $f_x(.9R) = f_Y(.9R) > 0$, yet $f(.9R, .9R) = 0$ since $(.9R, .9R)$ is outside the circle of radius R.

18.

a. $P_{y|X}(y|1)$ results from dividing each entry in x = 1 row of the joint probability table by $p_x(1) = .34$:

$$P_{y|x}(0|1) = \frac{.08}{.34} = .2353$$

$$P_{y|x}(1|1) = \frac{.20}{.34} = .5882$$

$$P_{y|x}(2|1) = \frac{.06}{.34} = .1765$$

b. $P_{y|X}(x|2)$ is requested; to obtain this divide each entry in the y = 2 row by $p_x(2) = .50$:

y	0	1	2
$P_{y\|X}(y\|2)$	.12	.28	.60

c. $P(Y \le 1 \mid x = 2) = P_{y|X}(0|2) + P_{y|X}(1|2) = .12 + .28 = .40$

d. $P_{X|Y}(x|2)$ results from dividing each entry in the y = 2 column by $p_y(2) = .38$:

x	0	1	2
$P_{x\|y}(x\|2)$	.0526	.1579	.7895

19.

a. $f_{Y|X}(y\mid x)=\dfrac{f(x,y)}{f_X(x)}=\dfrac{k(x^2+y^2)}{10kx^2+.05}$ $\quad 20 \le y \le 30$

$f_{X|Y}(x\mid y)=\dfrac{k(x^2+y^2)}{10ky^2+.05}$ $\quad 20 \le x \le 30 \quad \left(k=\dfrac{3}{380{,}000}\right)$

b. P(Y ≥ 25 | X = 22) = $\int_{25}^{30} f_{Y|X}(y\mid 22)dy$

$= \int_{25}^{30}\dfrac{k((22)^2+y^2)}{10k(22)^2+.05}dy = .783$

P(Y ≥ 25) = $\int_{25}^{30} f_Y(y)dy = \int_{25}^{30}(10ky^2+.05)dy = .75$

c. E(Y | X=22) = $\int_{-\infty}^{\infty} y\cdot f_{Y|X}(y\mid 22)dy = \int_{20}^{30} y\cdot\dfrac{k((22)^2+y^2)}{10k(22)^2+.05}dy$

= 25.372912

E(Y^2 | X=22) = $\int_{20}^{30} y^2\cdot\dfrac{k((22)^2+y^2)}{10k(22)^2+.05}dy = 652.028640$

V(Y| X = 22) = E(Y^2 | X=22) – [E(Y | X=22)]^2 = 8.243976

20.

a. $f_{x_3|x_1,x_2}(x_3\mid x_1,x_2)=\dfrac{f(x_1,x_2,x_3)}{f_{x_1,x_2}(x_1,x_2)}$ where $f_{x_1,x_2}(x_1,x_2)=$ the marginal joint pdf of (X_1, X_2) = $\int_{-\infty}^{\infty} f(x_1,x_2,x_3)dx_3$

b. $f_{x_2,x_3|x_1}(x_2,x_3\mid x_1)=\dfrac{f(x_1,x_2,x_3)}{f_{x_1}(x_1)}$ where

$f_{x_1}(x_1)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2,x_3)dx_2dx_3$

21. For every x and y, $f_{Y|X}(y|x) = f_y(y)$, since then $f(x,y) = f_{Y|X}(y|x)\cdot f_X(x) = f_Y(y)\cdot f_X(x)$, as required.

Section 5.2

22.

a. E(X + Y) = $\sum_x \sum_y (x+y)\,p(x,y) = (0+0)(.02)$

$+ (0+5)(.06) + ... + (10+15)(.01) = 14.10$

b. E[max (X,Y)] = $\sum_x \sum_y \max(x+y)\cdot p(x,y)$

$= (0)(.02) + (5)(.06) + ... + (15)(.01) = 9.60$

23. E(X_1 – X_2) = $\sum_{x_1=0}^{4} \sum_{x_2=0}^{3} (x_1 - x_2)\cdot p(x_1, x_2) =$

(0 – 0)(.08) + (0 – 1)(.07) + … + (4 – 3)(.06) = .15
(which also equals E(X_1) – E(X_2) = 1.70 – 1.55)

24. Let h(X,Y) = # of individuals who handle the message.

	h(x,y)	y: 1	2	3	4	5	6
	1	-	2	3	4	3	2
	2	2	-	2	3	4	3
x	3	3	2	-	2	3	4
	4	4	3	2	-	2	3
	5	3	4	3	2	-	2
	6	2	3	4	3	2	-

Since p(x,y) = $\frac{1}{30}$ for each possible (x,y), E[h(X,Y)] = $\sum_x \sum_y h(x,y)\cdot \frac{1}{30} = \frac{84}{30} = 2.80$

25. E(XY) = E(X) · E(Y) = L · L = L^2

26. Revenue = 3X + 10Y, so E (revenue) = E (3X + 10Y)

$$= \sum_{x=0}^{5} \sum_{y=0}^{2} (3x+10y)\cdot p(x,y) = 0\cdot p(0,0) + ... + 35\cdot p(5,2) = 15.4$$

27. E[h(X,Y)] = $\int_0^1\int_0^1 |x-y| \cdot 6x^2 y\,dx\,dy = 2\int_0^1\int_0^x (x-y)\cdot 6x^2 y\,dy\,dx$

$$12\int_0^1\int_0^x \left(x^3 y - x^2 y^2\right)dy\,dx = 12\int_0^1 \frac{x^5}{6}dx = \frac{1}{3}$$

28. E(XY) = $\sum_x\sum_y xy\cdot p(x,y) = \sum_x\sum_y xy\cdot p_x(x)\cdot p_y(y) = \sum_x x p_x(x)\cdot \sum_y y p_y(y)$

= E(X) · E(Y). (replace Σ with $\int$ in the continuous case)

29. Cov(X,Y) = $-\frac{2}{75}$ and $m_x = m_y = \frac{2}{5}$. E(X^2) = $\int_0^1 x^2 \cdot f_x(x)dx$

$= 12\int_0^1 x^3(1-x^2dx) = \frac{12}{60} = \frac{1}{5}$, so Var (X) = $\frac{1}{5} - \frac{4}{25} = \frac{1}{25}$

Similarly, Var(Y) = $\frac{1}{25}$, so $r_{X,Y} = \frac{-2/75}{\sqrt{\frac{1}{25}}\cdot\sqrt{\frac{1}{25}}} = -\frac{50}{75} = -.667$

30.

a. E(X) = 5.55, E(Y) = 8.55, E(XY) = (0)(.02) + (0)(.06) + … + (150)(.01) = 44.25, so Cov(X,Y) = 44.25 – (5.55)(8.55) = -3.20

b. $s_X^2 = 12.45, s_Y^2 = 19.15$, so $r_{X,Y} = \frac{-3.20}{\sqrt{(12.45)(19.15)}} = -.207$

31.

a. E(X) = $\int_{20}^{30} x f_x(x)dx = \int_{20}^{30} x\left[10Kx^2 + .05\right]dx = 25.329 = E(Y)$

E(XY) = $\int_{20}^{30}\int_{20}^{30} xy\cdot K(x^2+y^2)dx\,dy = 641.447$

$\Rightarrow Cov(X,Y) = 641.447 - (25.329)^2 = -.111$

b. E(X^2) = $\int_{20}^{30} x^2\left[10Kx^2 + .05\right]dx = 649.8246 = E(Y^2)$,

so Var (X) = Var(Y) = 649.8246 – (25.329)^2 = 8.2664

$\Rightarrow r = \frac{-.111}{\sqrt{(8.2664)(8.2664)}} = -.0134$

32. There is a difficulty here. Existence of r requires that both X and Y have finite means and variances. Yet since the marginal pdf of Y is $\dfrac{1}{(1-y)^2}$ for $y \ge 0$,

$$E(y) = \int_0^\infty \frac{y}{(1+y)^2}dy = \int_0^\infty \frac{(1+y-1)}{(1+y)^2}dy = \int_0^\infty \frac{1}{(1+y)}dy - \int_0^\infty \frac{1}{(1+y)^2}dy$$, and the first integral is not finite. Thus r itself is undefined.

33. Since E(XY) = E(X) · E(Y), Cov(X,Y) = E(XY) – E(X) · E(Y) = E(X) · E(Y) - E(X) · E(Y) = 0, and since Corr(X,Y) = $\dfrac{Cov(X,Y)}{s_x s_y}$, then Corr(X,Y) = 0

34.

a. In the discrete case, Var[h(X,Y)] = E{[h(X,Y) – E(h(X,Y))]^2} =

$$\sum_x \sum_y [h(x,y) - E(h(X,Y))]^2 p(x,y) = \sum_x \sum_y [h(x,y)^2 p(x,y)] - [E(h(X,Y))]^2$$

with $\iint$ replacing $\sum\sum$ in the continuous case.

b. E[h(X,Y)] = E[max(X,Y)] = 9.60, and E[h^2(X,Y)] = E[(max(X,Y))^2] = (0)^2(.02) +(5)^2(.06) + …+ (15)^2(.01) = 105.5, so Var[max(X,Y)] = 105.5 – (9.60)^2 = 13.34

35.

a. Cov(aX + b, cY + d) = E[(aX + b)(cY + d)] – E(aX + b) · E(cY + d)
= E[acXY + adX + bcY + bd] – (aE(X) + b)(cE(Y) + d)
= acE(XY) – acE(X)E(Y) = acCov(X,Y)

b. Corr(aX + b, cY + d) =

$$\frac{Cov(aX+b, cY+d)}{\sqrt{Var(aX+b)}\sqrt{Var(cY+d)}} = \frac{acCov(X,Y)}{|a|\cdot|c|\sqrt{Var(X)\cdot Var(Y)}}$$

= Corr(X,Y) when a and c have the same signs.

c. When a and c differ in sign, Corr(aX + b, cY + d) = -Corr(X,Y).

36. Cov(X,Y) = Cov(X, aX+b) = E[X·(aX+b)] – E(X) ·E(aX+b) = a Var(X),

so Corr(X,Y) = $\dfrac{aVar(X)}{\sqrt{Var(X)\cdot Var(Y)}} = \dfrac{aVar(X)}{\sqrt{Var(X)\cdot a^2 Var(X)}}$ = 1 if a > 0, and –1 if a < 0

Section 5.3

37.

	$P(x_1)$	.20	.50	.30
$P(x_2)$	$x_2 \mid x_1$	25	40	65
.20	25	.04	.10	.06
.50	40	.10	.25	.15
.30	65	.06	.15	.09

a.

$\bar{x}$	25	32.5	40	45	52.5	65
$p(\bar{x})$	.04	.20	.25	.12	.30	.09

$$E(\bar{x}) = (25)(.04) + 32.5(.20) + \ldots + 65(.09) = 44.5 = m$$

b.

s^2	0	112.5	312.5	800
$P(s^2)$	.38	.20	.30	.12

$E(s^2) = 212.25 = \sigma^2$

38.

a.

T_0	0	1	2	3	4
$P(T_0)$	.04	.20	.37	.30	.09

b. $m_{T_0} = E(T_0) = 2.2 = 2 \cdot m$

c. $s^2_{T_0} = E(T_0^2) - E(T_0)^2 = 5.82 - (2.2)^2 = .98 = 2 \cdot s^2$

39.

x	0	1	2	3	4	5	6	7	8	9	10
x/n	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
p(x/n)	.000	.000	.000	.001	.005	.027	.088	.201	.302	.269	.107

X is a binomial random variable with p = .8.

40.

a. Possible values of M are: 0, 5, 10. M = 0 when all 3 envelopes contain 0 money, hence $p(M = 0) = (.5)^3 = .125$. M = 10 when there is a single envelope with \$10, hence p(M = 10) = 1 – p(no envelopes with \$10) = $1 - (.8)^3 = .488$.
p(M = 5) = 1 – [.125 + .488] = .387.

M	0	5	10
p(M)	.125	.387	.488

An alternative solution would be to list all 27 possible combinations using a tree diagram and computing probabilities directly from the tree.

b. The statistic of interest is M, the maximum of x_1, x_2, or x_3, so that M = 0, 5, or 10. The population distribution is a s follows:

x	0	5	10
p(x)	1/2	3/10	1/5

Write a computer program to generate the digits 0 – 9 from a uniform distribution. Assign a value of 0 to the digits 0 – 4, a value of 5 to digits 5 – 7, and a value of 10 to digits 8 and 9. Generate samples of increasing sizes, keeping the number of replications constant and compute M from each sample. As n, the sample size, increases, p(M = 0) goes to zero, p(M = 10) goes to one. Furthermore, p(M = 5) goes to zero, but at a slower rate than p(M = 0).

41.

Outcome	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4
Probability	.16	.12	.08	.04	.12	.09	.06	.03
$\overline{x}$	1	1.5	2	2.5	1.5	2	2.5	3
r	0	1	2	3	1	0	1	2

Outcome	3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4
Probability	.08	.06	.04	.02	.04	.03	.02	.01
$\overline{x}$	2	2.5	3	3.5	2.5	3	3.5	4
r	2	1	0	1	3	2	1	2

a.

$\overline{x}$	1	1.5	2	2.5	3	3.5	4
$p(\overline{x})$	.16	.24	.25	.20	.10	.04	.01

b. $P\left(\overline{x} \le 2.5\right) = .8$

c.

r	0	1	2	3
p(r)	.30	.40	.22	.08

d. $P(\overline{X} \le 1.5)$ = P(1,1,1,1) + P(2,1,1,1) + … + P(1,1,1,2) + P(1,1,2,2) + … + P(2,2,1,1) + P(3,1,1,1) + … + P(1,1,1,3)
$= (.4)^4 + 4(.4)^3(.3) + 6(.4)^2(.3)^2 + 4(.4)^2(.2)^2 = .2400$

42.

a.

$\overline{x}$	27.75	28.0	29.7	29.95	31.65	31.9	33.6
$p(\overline{x})$	$\frac{4}{30}$	$\frac{2}{30}$	$\frac{6}{30}$	$\frac{4}{30}$	$\frac{8}{30}$	$\frac{4}{30}$	$\frac{2}{30}$

b.

$\overline{x}$	27.75	31.65	31.9
$p(\overline{x})$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

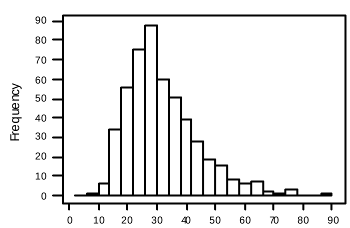
c. all three values are the same: 30.4333

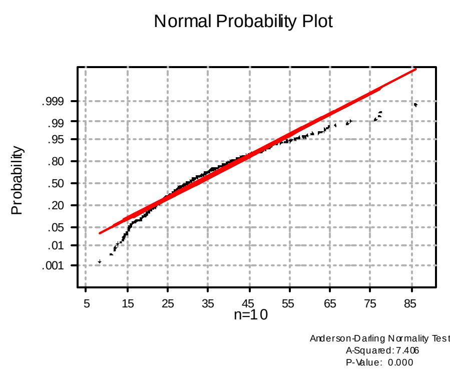
43. The statistic of interest is the fourth spread, or the difference between the medians of the upper and lower halves of the data. The population distribution is uniform with A = 8 and B = 10. Use a computer to generate samples of sizes n = 5, 10, 20, and 30 from a uniform distribution with A = 8 and B = 10. Keep the number of replications the same (say 500, for example). For each sample, compute the upper and lower fourth, then compute the difference. Plot the sampling distributions on separate histograms for n = 5, 10, 20, and 30.

44. Use a computer to generate samples of sizes n = 5, 10, 20, and 30 from a Weibull distribution with parameters as given, keeping the number of replications the same, as in problem 43 above. For each sample, calculate the mean. Below is a histogram, and a normal probability plot for the sampling distribution of $\bar{x}$ for n = 5, both generated by Minitab. This sampling distribution appears to be normal, so since larger sample sizes will produce distributions that are closer to normal, the others will also appear normal.

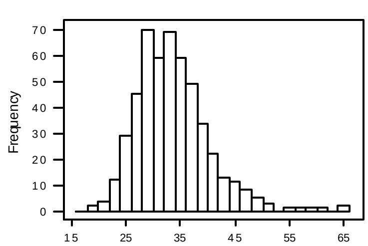
45. Using Minitab to generate the necessary sampling distribution, we can see that as n increases, the distribution slowly moves toward normality. However, even the sampling distribution for n = 50 is not yet approximately normal.

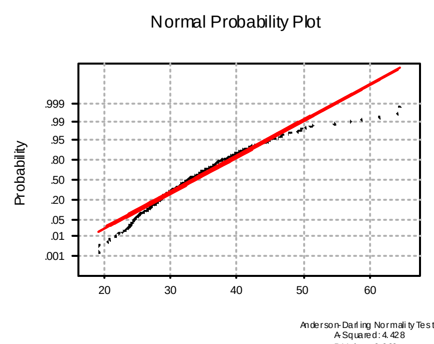
n = 10





n = 50





Section 5.4

46. μ = 12 cm σ = .04 cm

a. n = 16 $E(\overline{X}) = m = 12cm \quad s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{.04}{4} = .01cm$

b. n = 64 $E(\overline{X}) = m = 12cm \quad s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{.04}{8} = .005cm$

c. $\overline{X}$ is more likely to be within .01 cm of the mean (12 cm) with the second, larger, sample. This is due to the decreased variability of $\overline{X}$ with a larger sample size.

47. μ = 12 cm σ = .04 cm

a. n = 16 P(11.99 ≤ $\overline{X}$ ≤ 12.01) = $P\left(\frac{11.99-12}{.01} \le Z \le \frac{12.01-12}{.01}\right)$

= P(-1 ≤ Z ≤ 1)
= Φ(1) - Φ(-1)
=.8413 - .1587
=.6826

b. n = 25 P($\overline{X}$ > 12.01) = $P\left(Z > \frac{12.01-12}{.04/5}\right)$ = P(Z > 1.25)

= 1 - Φ(1.25)
= 1 - .8944
=.1056

48.

a. $m_{\overline{X}} = m = 50$, $s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{1}{\sqrt{100}} = .10$

P(49.75 ≤ $\overline{X}$ ≤ 50.25) = $P\left(\frac{49.75-50}{.10} \le Z \le \frac{50.25-50}{.10}\right)$

= P(-2.5 ≤ Z ≤ 2.5) = .9876

b. P(49.75 ≤ $\overline{X}$ ≤ 50.25) ≈ $P\left(\frac{49.75-49.8}{.10} \le Z \le \frac{50.25-49.8}{.10}\right)$

= P(-.5 ≤ Z ≤ 4.5) = .6915

49.

a. 11 P.M. – 6:50 P.M. = 250 minutes. With $T_0 = X_1 + \dots + X_{40}$ = total grading time,

$m_{T_0} = nm = (40)(6) = 240$ and $s_{T_0} = s\sqrt{n} = 37.95,$ so $P(T_0 \le 250) \approx$

$$P\left(Z \le \frac{250-240}{37.95}\right) = P(Z \le .26) = .6026$$

b. $$P(T_0 > 260) = P\left(Z > \frac{260-240}{37.95}\right) = P(Z > .53) = .2981$$

50. μ = 10,000 psi σ = 500 psi

a. n = 40

$$P(9{,}900 \le \overline{X} \le 10{,}200) \approx P\left(\frac{9{,}900-10{,}000}{500/\sqrt{40}} \le Z \le \frac{10{,}200-10{,}000}{500/\sqrt{40}}\right)$$

$= P(-1.26 \le Z \le 2.53)$
$= \Phi(2.53) - \Phi(-1.26)$
$= .9943 - .1038$
$= .8905$

b. According to the Rule of Thumb given in Section 5.4, n should be greater than 30 in order to apply the C.L.T., thus using the same procedure for n = 15 as was used for n = 40 would not be appropriate.

51. X ~ N(10,4). For day 1, n = 5

$$P(\overline{X} \le 11) = P\left(Z \le \frac{11-10}{2/\sqrt{5}}\right) = P(Z \le 1.12) = .8686$$

For day 2, n = 6

$$P(\overline{X} \le 11) = P\left(Z \le \frac{11-10}{2/\sqrt{6}}\right) = P(Z \le 1.22) = .8888$$

For both days,

$P(\overline{X} \le 11) = (.8686)(.8888) = .7720$

52. X ~ N(10), n =4

$m_{T_0} = nm = (4)(10) = 40$ and $s_{T_0} = s\sqrt{n} = (2)(1) = 2,$

We desire the 95th percentile: 40 + (1.645)(2) = 43.29

53. $\mu = 50, \quad \sigma = 1.2$

a. n = 9

$$P(\overline{X} \geq 51) = P\left(Z \geq \frac{51-50}{1.2/\sqrt{9}} \right) = P(Z \geq 2.5) = 1 - .9938 = .0062$$

b. n = 40

$$P(\overline{X} \geq 51) = P\left(Z \geq \frac{51-50}{1.2/\sqrt{40}} \right) = P(Z \geq 5.27) \approx 0$$

54.

a. $m_{\overline{X}} = m = 2.65$, $s_{\overline{x}} = \frac{s_x}{\sqrt{n}} = \frac{.85}{5} = .17$

$$P(\overline{X} \leq 3.00) = P\left(Z \leq \frac{3.00-2.65}{.17} \right) = P(Z \leq 2.06) = .9803$$

$$P(2.65 \leq \overline{X} \leq 3.00) == P(\overline{X} \leq 3.00) - P(\overline{X} \leq 2.65) = .4803$$

b. $P(\overline{X} \leq 3.00) = P\left(Z \leq \frac{3.00-2.65}{.85/\sqrt{n}} \right) = .99$ implies that $\frac{.35}{85/\sqrt{n}} = 2.33,$ from which n = 32.02. Thus n = 33 will suffice.

55. $m = np = 20 \qquad s = \sqrt{npq} = 3.464$

a. $P(25 \leq X) \approx P\left(\frac{24.5-20}{3.464} \leq Z \right) = P(1.30 \leq Z) = .0968$

b. $P(15 \leq X \leq 25) \approx P\left(\frac{14.5-20}{3.464} \leq Z \leq \frac{25.5-20}{3.464} \right)$
$= P(-1.59 \leq Z \leq 1.59) = .8882$

56.

a. With Y = # of tickets, Y has approximately a normal distribution with $m = l = 50$, $s = \sqrt{l} = 7.071$, so $P(35 \leq Y \leq 70) \approx P\left(\frac{34.5-50}{7.071} \leq Z \leq \frac{70.5-50}{7.071} \right)$ = P(-2.19 ≤ Z ≤ 2.90) = .9838

b. Here $m = 250$, $s^2 = 250, s = 15.811$, so P(225 ≤ Y ≤ 275) ≈
$P\left(\frac{224.5-250}{15.811} \leq Z \leq \frac{275.5-250}{15.811} \right)$ = P(-1.61 ≤ Z ≤ 1.61) = .8926

57. E(X) = 100, Var(X) = 200, $s_x = 14.14$, so P(X ≤ 125) ≈ $P\left(Z \le \frac{125-100}{14.14}\right)$

= P(Z ≤ 1.77) = .9616

Section 5.5

58.

a. $E(27X_1 + 125X_2 + 512X_3) = 27\,E(X_1) + 125\,E(X_2) + 512\,E(X_3)$

$= 27(200) + 125(250) + 512(100) = 87{,}850$

$V(27X_1 + 125X_2 + 512X_3) = 27^2\,V(X_1) + 125^2\,V(X_2) + 512^2\,V(X_3)$

$= 27^2\,(10)^2 + 125^2\,(12)^2 + 512^2\,(8)^2 = 19{,}100{,}116$

b. The expected value is still correct, but the variance is not because the covariances now also contribute to the variance.

59.

a. $E(X_1 + X_2 + X_3) = 180$, $V(X_1 + X_2 + X_3) = 45$, $s_{x_1+x_2+x_3} = 6.708$

$P(X_1 + X_2 + X_3 \le 200) = P\left(Z \le \frac{200-180}{6.708}\right) = P(Z \le 2.98) = .9986$

$P(150 \le X_1 + X_2 + X_3 \le 200) = P(-4.47 \le Z \le 2.98) \approx .9986$

b. $m_{\overline{X}} = m = 60$, $s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{\sqrt{15}}{\sqrt{3}} = 2.236$

$P(\overline{X} \ge 55) = P\left(Z \ge \frac{55-60}{2.236}\right) = P(Z \ge -2.236) = .9875$

$P(58 \le \overline{X} \le 62) = P(-.89 \le Z \le .89) = .6266$

c. $E(X_1 - .5X_2 - .5X_3) = 0$;

$V(X_1 - .5X_2 - .5X_3) = s_1^2 + .25s_2^2 + .25s_3^2 = 22.5$, sd = 4.7434

$P(-10 \le X_1 - .5X_2 - .5X_3 \le 5) = P\left(\frac{-10-0}{4.7434} \le Z \le \frac{5-0}{4.7434}\right)$

$= P(-2.11 \le Z \le 1.05)$ = .8531 - .0174 = .8357

d. E($X_1 + X_2 + X_3$) = 150, V($X_1 + X_2 + X_3$) = 36, $s_{x_1+x_2+x_3} = 6$

P($X_1 + X_2 + X_3 \le 200$) = $P\left(Z \le \dfrac{160-150}{6}\right) = P(Z \le 1.67) = .9525$

We want P($X_1 + X_2 \ge 2X_3$), or written another way, P($X_1 + X_2 - 2X_3 \ge 0$)
E($X_1 + X_2 - 2X_3$) = 40 + 50 – 2(60) = -30,
V($X_1 + X_2 - 2X_3$) = $s_1^2 + s_2^2 + 4s_3^2 = 78,$ 36, sd = 8.832, so

P($X_1 + X_2 - 2X_3 \ge 0$) = $P\left(Z \ge \dfrac{0-(-30)}{8.832}\right) = P(Z \ge 3.40) = .0003$

60. Y is normally distributed with $m_Y = \dfrac{1}{2}(m_1 + m_2) - \dfrac{1}{3}(m_3 + m_4 + m_5) = -1$, and

$$s_Y^2 = \frac{1}{4}s_1^2 + \frac{1}{4}s_2^2 + \frac{1}{9}s_3^2 + \frac{1}{9}s_4^2 + \frac{1}{9}s_5^2 = 3.167, s_Y = 1.7795.$$

Thus, $P(0 \le Y) = P\left(\dfrac{0-(-1)}{1.7795} \le Z\right) = P(.56 \le Z) = .2877$ and

$$P(-1 \le Y \le 1) = P\left(0 \le Z \le \frac{2}{1.7795}\right) = P(0 \le Z \le 1.12) = .3686$$

61.

a. The marginal pmf's of X and Y are given in the solution to Exercise 7, from which E(X) = 2.8, E(Y) = .7, V(X) = 1.66, V(Y) = .61. Thus E(X+Y) = E(X) + E(Y) = 3.5, V(X+Y) = V(X) + V(Y) = 2.27, and the standard deviation of X + Y is 1.51

b. E(3X+10Y) = 3E(X) + 10E(Y) = 15.4, V(3X+10Y) = 9V(X) + 100V(Y) = 75.94, and the standard deviation of revenue is 8.71

62. E($X_1 + X_2 + X_3$) = E(X_1) + E(X_2) + E(X_3) = 15 + 30 + 20 = 65 min.,
V($X_1 + X_2 + X_3$) = $1^2 + 2^2 + 1.5^2$ = 7.25, $s_{x_1+x_2+x_3} = \sqrt{7.25} = 2.6926$

Thus, P($X_1 + X_2 + X_3 \le 60$) = $P\left(Z \le \dfrac{60-65}{2.6926}\right) = P(Z \le -1.86) = .0314$

63.

a. E(X_1) = 1.70, E(X_2) = 1.55, E(X_1X_2) = $\sum_{x_1}\sum_{x_2} x_1 x_2 p(x_1, x_2) = 3.33$, so Cov($X_1$,$X_2$) = E($X_1X_2$) - E($X_1$) E($X_2$) = 3.33 – 2.635 = .695

b. V($X_1 + X_2$) = V(X_1) + V(X_2) + 2 Cov(X_1,X_2)
= 1.59 + 1.0875 + 2(.695) = 4.0675

64. Let $X_1, \ldots, X_5$ denote morning times and $X_6, \ldots, X_{10}$ denote evening times.

a. $E(X_1 + \ldots + X_{10}) = E(X_1) + \ldots + E(X_{10}) = 5\,E(X_1) + 5\,E(X_6)$
$= 5(4) + 5(5) = 45$

b. $Var(X_1 + \ldots + X_{10}) = Var(X_1) + \ldots + Var(X_{10}) = 5\,Var(X_1) + 5Var(X_6)$

$$= 5\left[\frac{64}{12} + \frac{100}{12}\right] = \frac{820}{12} = 68.33$$

c. $E(X_1 - X_6) = E(X_1) - E(X_6) = 4 - 5 = -1$

$Var(X_1 - X_6) = Var(X_1) + Var(X_6) = \dfrac{64}{12} + \dfrac{100}{12} = \dfrac{164}{12} = 13.67$

d. $E[(X_1 + \ldots + X_5) - (X_6 + \ldots + X_{10})] = 5(4) - 5(5) = -5$
$Var[(X_1 + \ldots + X_5) - (X_6 + \ldots + X_{10})]$
$= Var(X_1 + \ldots + X_5) + Var(X_6 + \ldots + X_{10})] = 68.33$

65. $\mu = 5.00$, $\sigma = .2$

a. $E(\overline{X} - \overline{Y}) = 0; \qquad V(\overline{X} - \overline{Y}) = \dfrac{s^2}{25} + \dfrac{s^2}{25} = .0032,\ s_{\overline{X}-\overline{Y}} = .0566$

$$\Rightarrow P\left(-.1 \le \overline{X} - \overline{Y} \le .1\right) \approx P\left(-1.77 \le Z \le 1.77\right) = .9232 \text{ (by the CLT)}$$

b. $V(\overline{X} - \overline{Y}) = \dfrac{s^2}{36} + \dfrac{s^2}{36} = .0022222,\ s_{\overline{X}-\overline{Y}} = .0471$

$$\Rightarrow P\left(-.1 \le \overline{X} - \overline{Y} \le .1\right) \approx P\left(-2.12 \le Z \le 2.12\right) = .9660$$

66.

a. With $M = 5X_1 + 10X_2$, $E(M) = 5(2) + 10(4) = 50$,
$Var(M) = 5^2\,(.5)^2 + 10^2\,(1)^2 = 106.25$, $\sigma_M = 10.308$.

b. $P(\,75 < M\,) = P\left(\dfrac{75 - 50}{10.308} < Z\right) = P(2.43 < Z) = .0075$

c. $M = A_1X_1 + A_2X_2$ with the A_I's and X_I's all independent, so
$E(M) = E(A_1X_1) + E(A_2X_2) = E(A_1)E(X_1) + E(A_2)E(X_2) = 50$

d. $Var(M) = E(M^2) - [E(M)]^2$. Recall that for any r.v. Y,
$E(Y^2) = Var(Y) + [E(Y)]^2$. Thus, $E(M^2) = E\left(A_1^2X_1^2 + 2A_1X_1A_2X_2 + A_2^2X_2^2\right)$
$= E\left(A_1^2\right)E\left(X_1^2\right) + 2E\left(A_1\right)E\left(X_1\right)E\left(A_2\right)E\left(X_2\right) + E\left(A_2^2\right)E\left(X_2^2\right)$
(by independence)
$= (.25 + 25)(.25 + 4) + 2(5)(2)(10)(4) + (.25 + 100)(1 + 16) = 2611.5625$, so $Var(M) =$
$2611.5625 - (50)^2 = 111.5625$

e. E(M) = 50 still, but now

$$Var(M) = a_1^2 Var(X_1) + 2a_1 a_2 Cov(X_1, X_2) + a_2^2 Var(X_2)$$

= 6.25 + 2(5)(10)(-.25) + 100 = 81.25

67. Letting X_1, X_2, and X_3 denote the lengths of the three pieces, the total length is $X_1 + X_2 - X_3$. This has a normal distribution with mean value 20 + 15 – 1 = 34, variance .25+.16+.01 = .42, and standard deviation .6481. Standardizing gives $P(34.5 \le X_1 + X_2 - X_3 \le 35) = P(.77 \le Z \le 1.54) = .1588$

68. Let X_1 and X_2 denote the (constant) speeds of the two planes.

a. After two hours, the planes have traveled $2X_1$ km. and $2X_2$ km., respectively, so the second will not have caught the first if $2X_1 + 10 > 2X_2$, i.e. if $X_2 - X_1 < 5$. $X_2 - X_1$ has a mean 500 – 520 = -20, variance 100 + 100 = 200, and standard deviation 14.14. Thus,

$$P(X_2 - X_1 < 5) = P\left(Z < \frac{5-(-20)}{14.14}\right) = P(Z < 1.77) = .9616.$$

b. After two hours, #1 will be $10 + 2X_1$ km from where #2 started, whereas #2 will be $2X_2$ from where it started. Thus the separation distance will be al most 10 if $|2X_2 - 10 - 2X_1| \le 10$, i.e. $-10 \le 2X_2 - 10 - 2X_1 \le 10$,
i.e. $0 \le X_2 - X_1 \le 10$. The corresponding probability is
$P(0 \le X_2 - X_1 \le 10) = P(1.41 \le Z \le 2.12) = .9830 - .9207 = .0623$.

69.

a. $E(X_1 + X_2 + X_3) = 800 + 1000 + 600 = 2400$.

b. Assuming independence of X_1, X_2, X_3, $Var(X_1 + X_2 + X_3)$
$= (16)^2 + (25)^2 + (18)^2 = 12.05$

c. $E(X_1 + X_2 + X_3) = 2400$ as before, but now $Var(X_1 + X_2 + X_3)$
$= Var(X_1) + Var(X_2) + Var(X_3) + 2Cov(X_1,X_2) + 2Cov(X_1, X_3) + 2Cov(X_2, X_3) = 1745$,
with sd = 41.77

70.

a. $E(Y_i) = .5,$ so $E(W) = \sum_{i=1}^{n} i \cdot E(Y_i) = .5\sum_{i=1}^{n} i = \frac{n(n+1)}{4}$

b. $Var(Y_i) = .25,$ so $Var(W) = \sum_{i=1}^{n} i^2 \cdot Var(Y_i) = .25\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{24}$

71.

a. $M = a_1X_1 + a_2X_2 + W\int_0^{12} x\,dx = a_1X_1 + a_2X_2 + 72W$, so

E(M) = (5)(2) + (10)(4) + (72)(1.5) = 158m

$\sigma_M^2 = (5)^2(.5)^2 + (10)^2(1)^2 + (72)^2(.25)^2 = 430.25$, $\sigma_M = 20.74$

b. $P(M \le 200) = P\left(Z \le \dfrac{200-158}{20.74}\right) = P(Z \le 2.03) = .9788$

72. The total elapsed time between leaving and returning is $T_o = X_1 + X_2 + X_3 + X_4$, with $E(T_o) = 40$, $\sigma_{T_o}^2 = 40$, $\sigma_{T_o} = 5.477$. T_o is normally distributed, and the desired value t is the 99th percentile of the lapsed time distribution added to 10 A.M.: 10:00 + [40+(5.477)(2.33)] = 10:52.76

73.

a. Both approximately normal by the C.L.T.

b. The difference of two r.v.'s is just a special linear combination, and a linear combination of normal r.v's has a normal distribution, so $\overline{X} - \overline{Y}$ has approximately a normal distribution with $\mu_{\overline{X}-\overline{Y}} = 5$ and $\sigma_{\overline{X}-\overline{Y}}^2 = \dfrac{8^2}{40} + \dfrac{6^2}{35} = 2.629$, $\sigma_{\overline{X}-\overline{Y}} = 1.621$

c. $P\left(-1 \le \overline{X} - \overline{Y} \le 1\right) \approx P\left(\dfrac{-1-5}{1.6213} \le Z \le \dfrac{1-5}{1.6213}\right)$

$= P(-3.70 \le Z \le -2.47) \approx .0068$

d. $P\left(\overline{X} - \overline{Y} \ge 10\right) \approx P\left(Z \ge \dfrac{10-5}{1.6213}\right) = P(Z \ge 3.08) = .0010.$ This probability is quite small, so such an occurrence is unlikely if $\mu_1 - \mu_2 = 5$, and we would thus doubt this claim.

74. X is approximately normal with $\mu_1 = (50)(.7) = 35$ and $\sigma_1^2 = (50)(.7)(.3) = 10.5$, as is Y with $\mu_2 = 30$ and $\sigma_2^2 = 12$. Thus $\mu_{X-Y} = 5$ and $\sigma_{X-Y}^2 = 22.5$, so

$$p(-5 \le X - Y \le 5) \approx P\left(\frac{-10}{4.74} \le Z \le \frac{0}{4.74}\right) = P(-2.11 \le Z \le 0) = .4826$$

Supplementary Exercises

75.

a. $p_X(x)$ is obtained by adding joint probabilities across the row labeled x, resulting in $p_X(x)$ = .2, .5, .3 for x = 12, 15, 20 respectively. Similarly, from column sums $p_y(y)$ = .1, .35, .55 for y = 12, 15, 20 respectively.

b. $P(X \le 15 \text{ and } Y \le 15) = p(12,12) + p(12,15) + p(15,12) + p(15,15) = .25$

c. $p_x(12) \cdot p_y(12) = (.2)(.1) \ne .05 = p(12,12)$, so X and Y are not independent. (Almost any other (x,y) pair yields the same conclusion).

d. $E(X+Y) = \sum\sum (x+y)p(x,y) = 33.35$ (or = E(X) + E(Y) = 33.35)

e. $E(|X-Y|) = \sum\sum |x+y| p(x,y) = 3.85$

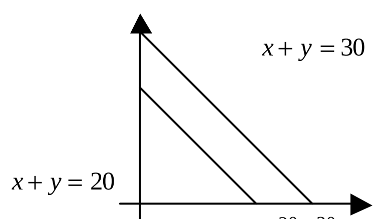
76. The roll-up procedure is not valid for the 75th percentile unless $s_1 = 0$ or $s_2 = 0$ or both s_1 and $s_2 = 0$, as described below.

Sum of percentiles: $m_1 + (Z)s_1 + m_2 + (Z)s_2 = m_1 + m_2 + (Z)(s_1 + s_2)$

Percentile of sums: $m_1 + m_2 + (Z)\sqrt{s_1^2 + s_2^2}$

These are equal when Z = 0 (i.e. for the median) or in the unusual case when $s_1 + s_2 = \sqrt{s_1^2 + s_2^2}$, which happens when $s_1 = 0$ or $s_2 = 0$ or both s_1 and $s_2 = 0$.

77.



a. $$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = \int_0^{20}\int_{20-x}^{30-x} kxydydx + \int_{20}^{30}\int_0^{30-x} kxydydx$$
$$= \frac{81{,}250}{3} \cdot k \Rightarrow k = \frac{3}{81{,}250}$$

b. $$f_X(x) = \begin{cases} \int_{20-x}^{30-x} kxydy = k(250x - 10x^2) & 0 \le x \le 20 \\ \int_0^{30-x} kxydy = k(450x - 30x^2 + \frac{1}{2}x^3) & 20 \le x \le 30 \end{cases}$$

and by symmetry $f_Y(y)$ is obtained by substituting y for x in $f_X(x)$. Since $f_X(25) > 0$, and $f_Y(25) > 0$, but $f(25, 25) = 0$, $f_X(x) \cdot f_Y(y) \ne f(x,y)$ for all x,y so X and Y are not independent.

c. $P(X+Y \le 25) = \int_0^{20}\int_{20-x}^{25-x} kxydydx + \int_{20}^{25}\int_0^{25-x} kxydydx$

$$= \frac{3}{81{,}250} \cdot \frac{230{,}625}{24} = .355$$

d. $E(X+Y) = E(X) + E(Y) = 2\left\{\int_0^{20} x \cdot k\left(250x - 10x^2\right)dx\right.$

$\left. + \int_{20}^{30} x \cdot k\left(450x - 30x^2 + \tfrac{1}{2}x^3\right)dx\right\} \quad = 2k(351{,}666.67) = 25.969$

e. $E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \cdot f(x,y)dxdy = \int_0^{20}\int_{20-x}^{30-x} kx^2y^2dydx$

$+ \int_{20}^{30}\int_0^{30-x} kx^2y^2dydx = \frac{k}{3} \cdot \frac{33{,}250{,}000}{3} = 136.4103$, so

Cov(X,Y) = 136.4103 – $(12.9845)^2$ = -32.19, and $E(X^2) = E(Y^2)$ = 204.6154, so

$s_x^2 = s_y^2 = 204.6154 - (12.9845)^2 = 36.0182$ and $r = \frac{-32.19}{36.0182} = -.894$

f. Var (X + Y) = Var(X) + Var(Y) + 2Cov(X,Y) = 7.66

78. $F_Y(y) = P(\max(X_1, \ldots, X_n) \le y) = P(X_1 \le y, \ldots, X_n \le y) = [P(X_1 \le y)]^n = \left(\frac{y-100}{100}\right)^n$ for $100 \le y \le 200$.

Thus $f_Y(y) = \frac{n}{100^n}(y-100)^{n-1}$ for $100 \le y \le 200$.

$$E(Y) = \int_{100}^{200} y \cdot \frac{n}{100^n}(y-100)^{n-1}dy = \frac{n}{100^n}\int_0^{100}(u+100)u^{n-1}du$$

$$= 100 + \frac{n}{100^n}\int_0^{100} u^n du = 100 + 100\frac{n}{n+1} = \frac{2n+1}{n+1} \cdot 100$$

79. $E(\overline{X} + \overline{Y} + \overline{Z}) = 500 + 900 + 2000 = 3400$

$Var(\overline{X} + \overline{Y} + \overline{Z}) = \frac{50^2}{365} + \frac{100^2}{365} + \frac{180^2}{365} = 123.014$, and the std dev = 11.09.

$P(\overline{X} + \overline{Y} + \overline{Z} \le 3500) = P(Z \le 9.0) \approx 1$

80.

a. Let $X_1, \ldots, X_{12}$ denote the weights for the business-class passengers and $Y_1, \ldots, Y_{50}$ denote the tourist-class weights. Then T = total weight
$= X_1 + \ldots + X_{12} + Y_1 + \ldots + Y_{50} = X + Y$
$E(X) = 12E(X_1) = 12(30) = 360$; $V(X) = 12V(X_1) = 12(36) = 432$.
$E(Y) = 50E(Y_1) = 50(40) = 2000$; $V(Y) = 50V(Y_1) = 50(100) = 5000$.
Thus $E(T) = E(X) + E(Y) = 360 + 2000 = 2360$
And $V(T) = V(X) + V(Y) = 432 + 5000 = 5432$, std dev = 73.7021

b. $$P(T \le 2500) = P\left(Z \le \frac{2500 - 2360}{73.7021}\right) = P(Z \le 1.90) = .9713$$

81.

a. $E(N) \cdot \mu = (10)(40) = 400$ minutes

b. We expect 20 components to come in for repair during a 4 hour period,
so $E(N) \cdot \mu = (20)(3.5) = 70$

82. X ~ Bin (200, .45) and Y ~ Bin (300, .6). Because both n's are large, both X and Y are approximately normal, so X + Y is approximately normal with mean (200)(.45) + (300)(.6) = 270, variance 200(.45)(.55) + 300(.6)(.4) = 121.40, and standard deviation 11.02. Thus, $P(X + Y \ge 250) = P\left(Z \ge \frac{249.5 - 270}{11.02}\right) = P(Z \ge -1.86) = .9686$

83. $0.95 = P(m - .02 \le \overline{X} \le m + .02) \approx P\left(\frac{-.02}{.01/\sqrt{n}} \le Z \le \frac{.02}{.01/\sqrt{n}}\right)$
$= P(-.2\sqrt{n} \le Z \le .2\sqrt{n})$, but $P(-1.96 \le Z \le 1.96) = .95$ so
$.2\sqrt{n} = 1.96 \Rightarrow n = 97.$ The C.L.T.

84. I have 192 oz. The amount which I would consume if there were no limit is $T_o = X_1 + \ldots + X_{14}$ where each X_I is normally distributed with $\mu = 13$ and $\sigma = 2$. Thus T_o is normal with $m_{T_o} = 182$ and $s_{T_o} = 7.483$, so $P(T_o < 192) = P(Z < 1.34) = .9099$.

85. The expected value and standard deviation of volume are 87,850 and 4370.37, respectively, so
$$P(volume \le 100{,}000) = P\left(Z \le \frac{100{,}000 - 87{,}850}{4370.37}\right) = P(Z \le 2.78) = .9973$$

86. The student will not be late if $X_1 + X_3 \le X_2$, i.e. if $X_1 - X_2 + X_3 \le 0$. This linear combination has mean –2, variance 4.25, and standard deviation 2.06, so
$$P(X_1 - X_2 + X_3 \le 0) = P\left(Z \le \frac{0 - (-2)}{2.06}\right) = P(Z \le .97) = .8340$$

87.

a. $Var(aX+Y)=a^2s_x^2+2aCov(X,Y)+s_y^2=a^2s_x^2+2as_Xs_Yr+s_y^2.$

Substituting $a=\frac{s_Y}{s_X}$ yields $s_Y^2+2s_Y^2r+s_Y^2=2s_Y^2(1-r)\geq 0$, so $r\geq -1$

b. Same argument as in **a**

c. Suppose $r=1$. Then $Var(aX-Y)=2s_Y^2(1-r)=0$, which implies that $aX-Y=k$ (a constant), so $aX-Y=aX-k$, which is of the form $aX+b$.

88. $E(X+Y-t)^2=\int_0^1\int_0^1(x+y-t)^2\cdot f(x,y)dxdy.$ To find the minimizing value of t, take the derivative with respect to t and equate it to 0:

$$0=\int_0^1\int_0^1 2(x+y-t)(-1)f(x,y)=0\Rightarrow\int_0^1\int_0^1 tf(x,y)dxdy=t$$

$=\int_0^1\int_0^1(x+y)\cdot f(x,y)dxdy=E(X+Y)$, so the best prediction is the individual's expected score (= 1.167).

89.

a. With Y = X_1 + X_2,

$$F_Y(y)=\int_0^y\left\{\int_0^{y-x_1}\frac{1}{2^{n_1/2}\Gamma(n_1/2)}\cdot\frac{1}{2^{n_21/2}\Gamma(n_2/2)}\cdot x_1^{\frac{n_1}{2}-1}x_2^{\frac{n_2}{2}-1}e^{-\frac{x_1+x_2}{2}}dx_2\right\}dx_1.$$

But the inner integral can be shown to be equal to

$\frac{1}{2^{(n_1+n_2)/2}\Gamma((n_1+n_2)/2)}y^{[(n_1+n_2)/2]-1}e^{-y/2}$, from which the result follows.

b. By **a,** $Z_1^2+Z_2^2$ is chi-squared with $n=2$, so $(Z_1^2+Z_2^2)+Z_3^2$ is chi-squared with $n=3$, etc, until $Z_1^2+...+Z_n^2$ 9s chi-squared with $n=n$

c. $\frac{X_i-m}{s}$ is standard normal, so $\left[\frac{X_i-m}{s}\right]^2$ is chi-squared with $n=1$, so the sum is chi-squared with $n=n$.

90.

a. Cov(X, Y + Z) = E[X(Y + Z)] – E(X) · E(Y + Z)
= E(XY) + E(XZ) – E(X) · E(Y) – E(X) · E(Z)
= E(XY) – E(X) · E(Y) + E(XZ) – E(X) · E(Z)
= Cov(X,Y) + Cov(X,Z).

b. $Cov(X_1 + X_2, Y_1 + Y_2) = Cov(X_1, Y_1) + Cov(X_1, Y_2) + Cov(X_2, Y_1) + Cov(X_2, Y_2)$ (apply **a** twice) = 16.

91.

a. $V(X_1) = V(W + E_1) = s_W^2 + s_E^2 = V(W + E_2) = V(X_2)$ and
$Cov(X_1, X_2) = Cov(W + E_1, W + E_2) = Cov(W, W) + Cov(W, E_2) +$
$Cov(E_1, W) + Cov(E_1, E_2) = Cov(W, W) = V(W) = s_w^2$.

Thus, $r = \dfrac{s_W^2}{\sqrt{s_W^2 + s_E^2} \cdot \sqrt{s_W^2 + s_E^2}} = \dfrac{s_W^2}{s_W^2 + s_E^2}$

b. $r = \dfrac{1}{1 + .0001} = .9999$

92.

a. Cov(X,Y) = Cov(A+D, B+E)
= Cov(A,B) + Cov(D,B) + Cov(A,E) + Cov(D,E)= Cov(A,B). Thus

$$Corr(X, Y) = \frac{Cov(A, B)}{\sqrt{s_A^2 + s_D^2} \cdot \sqrt{s_B^2 + s_E^2}}$$

$$= \frac{Cov(A, B)}{s_A s_B} \cdot \frac{s_A}{\sqrt{s_A^2 + s_D^2}} \cdot \frac{s_B}{\sqrt{s_B^2 + s_E^2}}$$

The first factor in this expression is Corr(A,B), and (by the result of exercise 70**a**) the second and third factors are the square roots of $Corr(X_1, X_2)$ and $Corr(Y_1, Y_2)$, respectively. Clearly, measurement error reduces the correlation, since both square-root factors are between 0 and 1.

b. $\sqrt{.8100} \cdot \sqrt{.9025} = .855$. This is disturbing, because measurement error substantially reduces the correlation.

93. $E(Y) \cong h(m_1, m_2, m_3, m_4) = 120\left[\frac{1}{10} + \frac{1}{15} + \frac{1}{20}\right] = 26$

The partial derivatives of $h(m_1, m_2, m_3, m_4)$ with respect to x_1, x_2, x_3, and x_4 are $-\dfrac{x_4}{x_1^2}$, $-\dfrac{x_4}{x_2^2}$, $-\dfrac{x_4}{x_3^2}$, and $\dfrac{1}{x_1} + \dfrac{1}{x_2} + \dfrac{1}{x_3}$, respectively. Substituting $x_1 = 10$, $x_2 = 15$, $x_3 = 20$, and $x_4 = 120$ gives –1.2, -.5333, -.3000, and .2167, respectively, so $V(Y) = (1)(-1.2)^2 + (1)(-.5333)^2 + (1.5)(-.3000)^2 + (4.0)(.2167)^2 = 2.6783$, and the approximate sd of y is 1.64.

94. The four second order partials are $\dfrac{2x_4}{x_1^3}$, $\dfrac{2x_4}{x_2^3}$, $\dfrac{2x_4}{x_3^3}$, and 0 respectively. Substitution gives $E(Y) = 26 + .1200 + .0356 + .0338 = 26.1894$.

CHAPTER 6

Section 6.1

1.

a. We use the sample mean, $\bar{x}$ to estimate the population mean m.

$$\hat{m} = \bar{x} = \frac{\Sigma x_i}{n} = \frac{219.80}{27} = 8.1407$$

b. We use the sample median, $\tilde{x} = 7.7$ (the middle observation when arranged in ascending order).

c. We use the sample standard deviation, $s = \sqrt{s^2} = \sqrt{\dfrac{1860.94 - \frac{(219.8)^2}{27}}{26}} = 1.660$

d. With "success" = observation greater than 10, x = # of successes = 4, and $\hat{p} = \frac{x}{n} = \frac{4}{27} = .1481$

e. We use the sample (std dev)/(mean), or $\dfrac{s}{\bar{x}} = \dfrac{1.660}{8.1407} = .2039$

2.

a. With X = # of T's in the sample, the estimator is $\hat{p} = \frac{X}{n}; x = 10,$ so $\hat{p} = \dfrac{10}{20}, = .50$.

b. Here, X = # in sample without TI graphing calculator, and x = 16, so $\hat{p} = \dfrac{16}{20} = .80$

3.

a. We use the sample mean, $\bar{x} = 1.3481$

b. Because we assume normality, the mean = median, so we also use the sample mean $\bar{x} = 1.3481$. We could also easily use the sample median.

c. We use the 90th percentile of the sample:
$\hat{m} + (1.28)\hat{s} = \bar{x} + 1.28s = 1.3481 + (1.28)(.3385) = 1.7814$.

d. Since we can assume normality,
$$P(X < 1.5) \approx P\left(Z < \frac{1.5 - \bar{x}}{s}\right) = P\left(Z < \frac{1.5 - 1.3481}{.3385}\right) = P(Z < .45) = .6736$$

e. The estimated standard error of $\bar{x} = \frac{\hat{s}}{\sqrt{n}} = \frac{s}{\sqrt{n}} = \frac{.3385}{\sqrt{16}} = .0846$

4.

a. $E(\bar{X} - \bar{Y}) = E(\bar{X}) - E(\bar{Y}) = m_1 - m_2$; $\bar{x} - \bar{y} = 8.141 - 8.575 = .434$

b. $V(\bar{X} - \bar{Y}) = V(\bar{X}) + V(\bar{Y}) = s_{\bar{X}}^2 + s_{\bar{Y}}^2 = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$

$s_{\bar{X}-\bar{Y}} = \sqrt{V(\bar{X} - \bar{Y})} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$; The estimate would be

$s_{\bar{X}-\bar{Y}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{1.66^2}{27} + \frac{2.104^2}{20}} = .5687$.

c. $\frac{s_1}{s_2} = \frac{1.660}{2.104} = .7890$

d. $V(X - Y) = V(X) + V(Y) = s_1^2 + s_2^2 = 1.66^2 + 2.104^2 = 7.1824$

5. N = 5,000 T = 1,761,300

$\bar{y} = 374.6$ $\bar{x} = 340.6$ $\bar{d} = 34.0$

$\hat{q}_1 = N\bar{x} = (5{,}000)(340.6) = 1{,}703{,}000$

$\hat{q}_2 = T - N\bar{d} = 1{,}761{,}300 - (5{,}000)(34.0) = 1{,}591{,}300$

$\hat{q}_3 = T\left(\frac{\bar{x}}{\bar{y}}\right) = 1{,}761{,}300\left(\frac{340.6}{374.6}\right) = 1{,}601{,}438.281$

6.

a. Let $y_i = \ln(x_i)$ for I = 1, .., 31. It is easily verified that the sample mean and sample sd of the $y_i's$ are $\bar{y} = 5.102$ and $s_y = .4961$. Using the sample mean and sample sd to estimate m and s , respectively, gives $\hat{m} = 5.102$ and $\hat{s} = .4961$ (whence $\hat{s}^2 = s_y^2 = .2461$).

b. $E(X) \equiv \exp\left[m + \frac{s^2}{2}\right]$. It is natural to estimate E(X) by using $\hat{m}$ and $\hat{s}^2$ in place of m and s^2 in this expression:

$$E(\hat{X}) = \exp\left[5.102 + \frac{.2461}{2}\right] = \exp(5.225) = 185.87$$

7.

a. $\hat{m} = \bar{x} = \frac{\sum x_i}{n} = \frac{1206}{10} = 120.6$

b. $\hat{t} = 10{,}000 \qquad \hat{m} = 1{,}206{,}000$

c. 8 of 10 houses in the sample used at least 100 therms (the "successes"), so $\hat{p} = \frac{8}{10} = .80.$

d. The ordered sample values are 89, 99, 103, 109, 118, 122, 125, 138, 147, 156, from which the two middle values are 118 and 122, so $\hat{\tilde{m}} = \tilde{x} = \frac{118 + 122}{2} = 120.0$

8.

a. With q denoting the true proportion of defective components, $\hat{q} = \frac{(\#defective in.sample)}{sample.size} = \frac{12}{80} = .150$

b. P(system works) = p^2 , so an estimate of this probability is $\hat{p}^2 = \left(\frac{68}{80}\right)^2 = .723$

9.

a. $E(\overline{X}) = m = E(X) = l$, so $\overline{X}$ is an unbiased estimator for the Poisson parameter l; $\sum x_i = (0)(18) + (1)(37) + ... + (7)(1) = 317,$ since n = 150,
$\hat{l} = \overline{x} = \dfrac{317}{150} = 2.11$.

b. $s_{\overline{x}} = \dfrac{s}{\sqrt{n}} = \dfrac{\sqrt{l}}{\sqrt{n}}$, so the estimated standard error is $\sqrt{\dfrac{\hat{l}}{n}} = \dfrac{\sqrt{2.11}}{\sqrt{150}} = .119$

10.

a. $E(\overline{X}^2) = Var(\overline{X}) + [E(\overline{X})]^2 = \dfrac{s^2}{n} + m^2$, so the bias of the estimator $\overline{X}^2$ is $\dfrac{s^2}{n}$; thus $\overline{X}^2$ tends to overestimate m^2.

b. $E(\overline{X}^2 - kS^2) = E(\overline{X}^2) - kE(S^2) = m^2 + \dfrac{s^2}{n} - ks^2$, so with $k = \dfrac{1}{n}$,
$E(\overline{X}^2 - kS^2) = m^2$.

11.

a. $E\left(\dfrac{X_1}{n_1} - \dfrac{X_2}{n_2}\right) = \dfrac{1}{n_1}E(X_1) - \dfrac{1}{n_2}E(X_2) = \dfrac{1}{n_1}(n_1 p_1) - \dfrac{1}{n_2}(n_2 p_2) = p_1 - p_2$.

b. $Var\left(\dfrac{X_1}{n_1} - \dfrac{X_2}{n_2}\right) = Var\left(\dfrac{X_1}{n_1}\right) + Var\left(\dfrac{X_2}{n_2}\right) = \left(\dfrac{1}{n_1}\right)^2 Var(X_1) + \left(\dfrac{1}{n_2}\right)^2 Var(X_2)$
$\dfrac{1}{n_1^2}(n_1 p_1 q_1) + \dfrac{1}{n_2^2}(n_2 p_2 q_2) = \dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}$, and the standard error is the square root of this quantity.

c. With $\hat{p}_1 = \dfrac{x_1}{n_1}$, $\hat{q}_1 = 1 - \hat{p}_1$, $\hat{p}_2 = \dfrac{x_2}{n_2}$, $\hat{q}_2 = 1 - \hat{p}_2$, the estimated standard error is
$\sqrt{\dfrac{\hat{p}_1\hat{q}_1}{n_1} + \dfrac{\hat{p}_2\hat{q}_2}{n_2}}$.

d. $(\hat{p}_1 - \hat{p}_2) = \dfrac{127}{200} - \dfrac{176}{200} = .635 - .880 = -.245$

e. $\sqrt{\dfrac{(.635)(.365)}{200}+\dfrac{(.880)(.120)}{200}}=.041$

12. $E\left[\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\right]=\dfrac{(n_1-1)}{n_1+n_2-2}E(S_1^2)+\dfrac{(n_2-1)}{n_1+n_2-2}E(S_2^2)$

$=\dfrac{(n_1-1)}{n_1+n_2-2}s^{\,2}+\dfrac{(n_2-1)}{n_1+n_2-2}s^{\,2}=s^{\,2}.$

13. $E(X)=\int_{-1}^{1}x\cdot\tfrac{1}{2}(1+qx)dx=\dfrac{x^2}{4}+\dfrac{qx^3}{6}\Big|_{-1}^{1}=\dfrac{1}{3}q \qquad E(X)=\dfrac{1}{3}q$

$E(\overline{X})=\dfrac{1}{3}q \qquad \hat{q}=3\overline{X}\Rightarrow E(\hat{q})=E(3\overline{X})=3E(\overline{X})=3\left(\dfrac{1}{3}\right)q=q$

14.

a. $\min(x_i)=202$ and $\max(x_i)=525$, so the estimate of the number of planes manufactured is $\max(x_i)-\min(x_i)+1=525-202+1=324$.

b. The estimate will equal the true number of planes manufactured iff $\min(x_i)=\alpha$ and $\max(x_i)=\beta$, i.e., iff the smallest serial number in the population and the largest serial number in the population both appear in the sample. The estimator is not unbiased. This is because $\max(x_i)$ never overestimates β and will usually underestimate it (unless $\max(x_i)=\beta$) , so that $E[\max(x_i)]<\beta$. Similarly, $E[\min(x_i)]>\alpha$,so $E[\max(x_i)-\min(x_i)]<\beta-\alpha+1$; The estimate will usually be smaller than $\beta-\alpha+1$, and can never exceed it.

15.

a. $E(X^2)=2q$ implies that $E\left(\dfrac{X^2}{2}\right)=q$. Consider $\hat{q}=\dfrac{\sum X_i^2}{2n}$. Then

$E(\hat{q})=E\left(\dfrac{\sum X_i^2}{2n}\right)=\dfrac{\sum E(X_i^2)}{2n}=\dfrac{\sum 2q}{2n}=\dfrac{2nq}{2n}=q$, implying that $\hat{q}$ is an unbiased estimator for q .

b. $\sum x_i^2=1490.1058$, so $\hat{q}=\dfrac{1490.1058}{20}=74.505$

16.

a. $E\left[d\overline{X}+(1-d)\overline{Y}\right]=dE(\overline{X})+(1-d)E(\overline{Y})=dm+(1-d)m=m$

b. $Var\left[d\overline{X}+(1-d)\overline{Y}\right]=d^2Var(\overline{X})+(1-d)^2Var(\overline{Y})=\dfrac{d^2s^2}{m}+\dfrac{4(1-d)^2s^2}{n}$.

Setting the derivative with respect to d equal to 0 yields $\dfrac{2ds^2}{m}+\dfrac{8(1-d)s^2}{n}=0$,

from which $d=\dfrac{4m}{4m+n}$.

17.

a.
$$E(\hat{p})=\sum_{x=0}^{\infty}\frac{r-1}{x+r-1}\cdot\binom{x+r-1}{x}\cdot p^r\cdot(1-p)^x$$
$$=p\sum_{x=0}^{\infty}\frac{(x+r-2)!}{x!(r-2)!}\cdot p^{r-1}\cdot(1-p)^x=p\sum_{x=0}^{\infty}\binom{x+r-2}{x}p^{r-1}(1-p)^x$$
$$=p\sum_{x=0}^{\infty}nb(x;r-1,p)=p\,.$$

b. For the given sequence, x = 5, so $\hat{p}=\dfrac{5-1}{5+5-1}=\dfrac{4}{9}=.444$

18.

a. $f(x;m,s^2)=\dfrac{1}{\sqrt{2p}\,s}e^{-\left((x-m)^2/2s^2\right)}$, so $f(m;m,s^2)=\dfrac{1}{\sqrt{2p}\,s}$ and

$\dfrac{1}{4n[[f(m)]^2}=\dfrac{2ps^2}{4n}=\dfrac{p}{2}\cdot\dfrac{s^2}{n}$; since $\dfrac{p}{2}>1,\ Var(\tilde{X})>Var(\overline{X})$.

b. $f(m)=\dfrac{1}{p}$, so $Var(\tilde{X})\approx\dfrac{p^2}{4n}=\dfrac{2.467}{n}$.

19.

a. $l = .5p + .15 \Rightarrow 2l = p + .3$, so $p = 2l - .3$ and $\hat{p} = 2\hat{l} - .3 = 2\left(\frac{Y}{n}\right) - .3$; the estimate is $2\left(\frac{20}{80}\right) - .3 = .2$.

b. $E(\hat{p}) = E\left(2\hat{l} - .3\right) = 2E\left(\hat{l}\right) - .3 = 2l - .3 = p$, as desired.

c. Here $l = .7p + (.3)(.3)$, so $p = \frac{10}{7}l - \frac{9}{70}$ and $\hat{p} = \frac{10}{7}\left(\frac{Y}{n}\right) - \frac{9}{70}$.

Section 6.2

20.

a. We wish to take the derivative of $\ln\left[\binom{n}{x}p^x(1-p)^{n-x}\right]$, set it equal to zero and solve for p. $\frac{d}{dp}\left[\ln\binom{n}{x} + x\ln(p) + (n-x)\ln(1-p)\right] = \frac{x}{p} - \frac{n-x}{1-p}$; setting this equal to zero and solving for p yields $\hat{p} = \frac{x}{n}$. For n = 20 and x = 3, $\hat{p} = \frac{3}{20} = .15$

b. $E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n}(np) = p$; thus $\hat{p}$ is an unbiased estimator of p.

c. $(1 - .15)^5 = .4437$

21.

a. $E(X) = b \cdot \Gamma\left(1+\frac{1}{a}\right)$ and $E(X^2) = Var(X) + [E(X)]^2 = b^2\Gamma\left(1+\frac{2}{a}\right)$, so the moment estimators $\hat{a}$ and $\hat{b}$ are the solution to $\overline{X} = \hat{b} \cdot \Gamma\left(1+\frac{1}{\hat{a}}\right)$, $\frac{1}{n}\sum X_i^2 = \hat{b}^2\Gamma\left(1+\frac{2}{\hat{a}}\right)$. Thus $\hat{b} = \frac{\overline{X}}{\Gamma\left(1+\frac{1}{\hat{a}}\right)}$, so once $\hat{a}$ has been determined $\Gamma\left(1+\frac{1}{\hat{a}}\right)$ is evaluated and $\hat{b}$ then computed. Since $\overline{X}^2 = \hat{b}^2 \cdot \Gamma^2\left(1+\frac{1}{\hat{a}}\right)$,

$$\frac{1}{n}\sum \frac{X_i^2}{\overline{X}^2} = \frac{\Gamma\left(1+\frac{2}{\hat{a}}\right)}{\Gamma^2\left(1+\frac{1}{\hat{a}}\right)}$$, so this equation must be solved to obtain $\hat{a}$.

b. From **a**, $\frac{1}{20}\left(\frac{16{,}500}{28.0^2}\right) = 1.05 = \frac{\Gamma\left(1+\frac{2}{\hat{a}}\right)}{\Gamma^2\left(1+\frac{1}{\hat{a}}\right)}$, so $\frac{1}{1.05} = .95 = \frac{\Gamma^2\left(1+\frac{1}{\hat{a}}\right)}{\Gamma\left(1+\frac{2}{\hat{a}}\right)}$, and from the hint, $\frac{1}{\hat{a}} = .2 \Rightarrow \hat{a} = 5$. Then $\hat{b} = \frac{\overline{x}}{\Gamma(1.2)} = \frac{28.0}{\Gamma(1.2)}$.

22.

a. $E(X) = \int_0^1 x(q+1)x^q\,dx = \frac{q+1}{q+2} = 1 - \frac{1}{q+2}$, so the moment estimator $\hat{q}$ is the solution to $\overline{X} = 1 - \frac{1}{\hat{q}+2}$, yielding $\hat{q} = \frac{1}{1-\overline{X}} - 2$. Since $\overline{x} = .80, \hat{q} = 5 - 2 = 3.$

b. $f(x_1,\ldots,x_n;q) = (q+1)^n (x_1 x_2 \ldots x_n)^q$, so the log likelihood is $n\ln(q+1) + q\sum\ln(x_i)$. Taking $\frac{d}{dq}$ and equating to 0 yields $\frac{n}{q+1} = -\sum \ln(x_i)$, so $\hat{q} = -\frac{n}{\sum \ln(X_i)} - 1$. Taking $\ln(x_i)$ for each given x_i yields ultimately $\hat{q} = 3.12$.

23. For a single sample from a Poisson distribution,

$$f(x_1,\ldots,x_n;\, l) = \frac{e^{-l}\, l^{x_1}}{x_1!}\cdots\frac{e^{-l}\, l^{x_n}}{x_n!} = \frac{e^{-nl}\, l^{\sum x_1}}{x_1!\ldots x_n!}\text{, so}$$

$$\ln[f(x_1,\ldots,x_n;\, l)] = -nl + \sum x_i \ln(l) - \sum \ln(x_i!)\text{. Thus}$$

$$\frac{d}{dl}[\ln[f(x_1,\ldots,x_n;\, l)]] = -n + \frac{\sum x_i}{l} = 0 \Rightarrow \hat{l} = \frac{\sum x_i}{n} = \bar{x}\text{.}$$

For our problem, $f(x_1,\ldots,x_n, y_1\ldots y_n;\, l_1, l_2)$ is a product of the x sample likelihood and the y sample likelihood, implying that $\hat{l}_1 = \bar{x},\ \hat{l}_2 = \bar{y}$, and (by the invariance principle)

$$\widehat{(l_1 - l_2)} = \bar{x} - \bar{y}\text{.}$$

24. We wish to take the derivative of $\ln\left[\binom{x+r-1}{x} p^r (1-p)^x\right]$ with respect to p, set it equal to zero, and solve for p: $\frac{d}{dp}\left[\ln\binom{x+r-1}{x} + r\ln(p) + x\ln(1-p)\right] = \frac{r}{p} - \frac{x}{1-p}$.

Setting this equal to zero and solving for p yields $\hat{p} = \frac{r}{r+x}$. This is the number of successes over the total number of trials, which is the same estimator for the binomial in exercise 6.20. The unbiased estimator from exercise 6.17 is $\hat{p} = \frac{r-1}{r+x-1}$, which is not the same as the maximum likelihood estimator.

25.

a. $\hat{m} = \bar{x} = 384.4;\ s^2 = 395.16$, so $\frac{1}{n}\sum (x_i - \bar{x})^2 = \hat{s}^2 = \frac{9}{10}(395.16) = 355.64$ and $\hat{s} = \sqrt{355.64} = 18.86$ (this is not s).

b. The 95th percentile is $m + 1.645s$, so the mle of this is (by the invariance principle) $\hat{m} + 1.645\hat{s} = 415.42$.

26. The mle of $P(X \le 400)$ is (by the invariance principle)

$$\Phi\left(\frac{400-\hat{m}}{\hat{s}}\right) = \Phi\left(\frac{400-384.4}{18.86}\right) = \Phi(.80) = .7881$$

27.

a. $f(x_1,\ldots,x_n;a,b) = \dfrac{(x_1x_2\ldots x_n)^{a-1}e^{-\Sigma x_i/b}}{b^{na}\Gamma^n(a)}$, so the log likelihood is $(a-1)\sum \ln(x_i) - \dfrac{\sum x_i}{b} - na\ \ln(b) - n\ln\Gamma(a)$. Equating both $\dfrac{d}{da}$ and $\dfrac{d}{db}$ to 0 yields $\sum \ln(x_i) - n\ln(b) - n\dfrac{d}{da}\Gamma(a) = 0$ and $\dfrac{\sum x_i}{b^2} = \dfrac{na}{b} = 0$, a very difficult system of equations to solve.

b. From the second equation in **a**, $\dfrac{\sum x_i}{b} = na \Rightarrow \bar{x} = ab = m$, so the mle of m is $\hat{m} = \bar{X}$.

28.

a. $\left(\dfrac{x_1}{q}\exp\left[-x_1^2/2q\right]\right)\ldots\left(\dfrac{x_n}{q}\exp\left[-x_n^2/2q\right]\right) = (x_1\ldots x_n)\dfrac{\exp\left[-\Sigma x_i^2/2q\right]}{q^n}$. The natural log of the likelihood function is $\ln(x_i\ldots x_n) - n\ln(q) - \dfrac{\Sigma x_i^2}{2q}$. Taking the derivative wrt q and equating to 0 gives $-\dfrac{n}{q} + \dfrac{\Sigma x_i^2}{2q^2} = 0$, so $nq = \dfrac{\Sigma x_i^2}{2}$ and $q = \dfrac{\Sigma x_i^2}{2n}$. The mle is therefore $\hat{q} = \dfrac{\Sigma X_i^2}{2n}$, which is identical to the unbiased estimator suggested in Exercise 15.

b. For x > 0 the cdf of X if $F(x;q) = P(X \le x)$ is equal to $1 - \exp\left[\dfrac{-x^2}{2q}\right]$. Equating this to .5 and solving for x gives the median in terms of q: $.5 = \exp\left[\dfrac{-x^2}{2q}\right]$ implies that $\ln(.5) = \dfrac{-x^2}{2q}$, so $x = \tilde{m} = \sqrt{1.38630}$. The mle of $\tilde{m}$ is therefore $(1.38630\hat{q})^{\frac{1}{2}}$.

29.

a. The joint pdf (likelihood function) is

$$f\left(x_1,...,x_n;\,l,q\right)=\begin{cases} l^n e^{-l\Sigma(x_i-q)} & x_1\ge q,...,x_n\ge q \\ 0 & otherwise \end{cases}$$

Notice that $x_1 \ge q,...,x_n \ge q$ iff $\min\left(x_i\right)\ge q$,

and that $-\,l\Sigma\left(x_i-q\right)=-l\Sigma x_i+n\,l\,q$.

Thus likelihood = $\begin{cases} l^n\exp\left(-\,l\Sigma x_i\right)\exp\left(n\,l\,q\right) & \min\left(x_i\right)\ge q \\ 0 & \min\left(x_i\right)<q \end{cases}$

Consider maximization wrt q . Because the exponent $n\,l\,q$ is positive, increasing q will increase the likelihood provided that $\min\left(x_i\right)\ge q$; if we make q larger than $\min\left(x_i\right)$, the likelihood drops to 0. This implies that the mle of q is $\hat{q}=\min\left(x_i\right)$. The log likelihood is now $n\ln(\,l\,)-l\Sigma\left(x_i-\hat{q}\right)$. Equating the derivative wrt l to 0 and solving yields $\hat{l}=\dfrac{n}{\Sigma\left(x_i-\hat{q}\right)}=\dfrac{n}{\Sigma x_i-n\hat{q}}$.

b. $\hat{q}=\min\left(x_i\right)=.64,$ and $\Sigma x_i=55.80$, so $\hat{l}=\dfrac{10}{55.80-6.4}=.202$

30. The likelihood is $f\left(y;n,p\right)=\binom{n}{y}p^y\left(1-p\right)^{n-y}$ where

$p=P\left(X\ge 24\right)=1-\int_0^{24}l e^{-lx}dx=e^{-24l}$. We know $\hat{p}=\dfrac{y}{n}$, so by the invariance principle $e^{-24l}=\dfrac{y}{n}\Rightarrow \hat{l}=-\dfrac{\left[\ln\left(\frac{y}{n}\right)\right]}{24}=.0120$ for n = 20, y = 15.

Supplementary Exercises

31. $P\left(\left|\overline{X}-m\right|>e\right)=P\left(\overline{X}-m>e\right)+P\left(\overline{X}-m<-e\right)=P\left(\dfrac{\overline{X}-m}{s/\sqrt{n}}>\dfrac{e}{s/\sqrt{n}}\right)+P\left(\dfrac{\overline{X}-m}{s/\sqrt{n}}<\dfrac{-e}{s/\sqrt{n}}\right)$

$$=P\left(Z>\frac{\sqrt{n}e}{s}\right)+P\left(Z<\frac{-\sqrt{n}e}{s}\right)=\int_{\sqrt{n}e/s}^{\infty}\frac{1}{\sqrt{2p}}e^{-z^2/2}dz+\int_{-\infty}^{-\sqrt{n}e/s}\frac{1}{\sqrt{2p}}e^{-z^2/2}dz\,.$$

As $n\to\infty$, both integrals $\to 0$ since $\lim\limits_{c\to\infty}\int_c^{\infty}\dfrac{1}{\sqrt{2p}}e^{-z^2/2}dz=0$.

32. sp

a. $F_Y(y) = P(Y \le y) = P(X_1 \le y, ..., X_n \le y) = P(X_1 \le y)...P(X_n \le y) = \left(\frac{y}{q}\right)^n$

for $0 \le y \le q$, so $f_Y(y) = \frac{ny^{n-1}}{q^n}$.

b. $E(Y) = \int_0^q y \cdot \frac{ny^{n-1}}{n} dy = \frac{n}{n+1} q$. While $\hat{q} = Y$ is not unbiased, $\frac{n+1}{n} Y$ is, since $E\left[\frac{n+1}{n} Y\right] = \frac{n+1}{n} E(Y) = \frac{n+1}{n} \cdot \frac{n}{n+1} q = q$, so $K = \frac{n+1}{n}$ does the trick.

33. Let x_1 = the time until the first birth, x_2 = the elapsed time between the first and second births, and so on. Then $f(x_1, ..., x_n; l) = l e^{-l x_1} \cdot (2l) e^{-2l x_2} ... (nl) e^{-nl x_n} = n! \, l^n e^{-l \Sigma k x_k}$. Thus the log likelihood is $\ln(n!) + n \ln(l) - l \Sigma k x_k$. Taking $\frac{d}{dl}$ and equating to 0 yields $\hat{l} = \frac{n}{\sum_{k=1}^{n} k x_k}$. For the given sample, n = 6, x_1 = 25.2, x_2 = 41.7 – 25.2 = 16.5, x_3 = 9.5, x_4 = 4.3, x_5 = 4.0, x_6 = 2.3; so $\sum_{k=1}^{6} k x_k = (1)(25.2) + (2)(16.5) + ... + (6)(2.3) = 137.7$ and $\hat{l} = \frac{6}{137.7} = .0436$.

34. $MSE(KS^2) = Var(KS^2) + Bias(KS^2)$.

$Bias(KS^2) = E(KS^2) - s^2 = Ks^2 - s^2 = s^2(K-1)$, and

$Var(KS^2) = K^2 Var(S^2) = K^2\left(E[(S^2)^2] - [E(S^2)]^2\right) = K^2\left(\frac{(n+1)s^4}{n-1} - (s^2)^2\right)$

$= \left[\frac{2K^2}{n-1} + (k-1)^2\right] s^4$. To find the minimizing value of K, take $\frac{d}{dK}$ and equate to 0; the result is $K = \frac{n-1}{n+1}$; thus the estimator which minimizes MSE is neither the unbiased estimator (K = 1) nor the mle $K = \frac{n-1}{n}$.

35.

$x_i + x_j$	23.5	26.3	28.0	28.2	29.4	29.5	30.6	31.6	33.9	49.3
23.5	23.5	24.9	25.75	25.85	26.45	26.5	27.05	27.55	28.7	36.4
26.3		26.3	27.15	27.25	27.85	27.9	28.45	28.95	30.1	37.8
28.0			28.0	28.1	28.7	28.75	29.3	29.8	30.95	38.65
28.2				28.2	28.8	28.85	29.4	29.9	31.05	38.75
29.4					29.4	29.45	30.0	30.5	30.65	39.35
29.5						29.5	30.05	30.55	31.7	39.4
30.6							30.6	31.1	32.25	39.95
31.6								31.6	32.75	40.45
33.9									33.9	41.6
49.3										49.3

There are 55 averages, so the median is the 28th in order of increasing magnitude. Therefore, $\hat{m} = 29.5$

36. With $\sum x = 555.86$ and $\sum x^2 = 15{,}490$, $s = \sqrt{s^2} = \sqrt{2.1570} = 1.4687$. The $\left|x_i - \tilde{x}\right|$'s are, in increasing order, .02, .02, .08, .22, .32, .42, .53, .54, .65, .81, .91, 1.15, 1.17, 1.30, 1.54, 1.54, 1.71, 2.35, 2.92, 3.50. The median of these values is $\frac{(.81 + .91)}{2} = .86$. The estimate based on the resistant estimator is then $\frac{.86}{.6745} = 1.275$. This estimate is in reasonably close agreement with s.

37. Let $c = \dfrac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\cdot\sqrt{\frac{2}{n-1}}}$. Then E(cS) = cE(S), and c cancels with the two Γ factors and the square root in E(S), leaving just s. When n = 20, $c = \dfrac{\Gamma(9.5)}{\Gamma(10)\cdot\sqrt{\frac{2}{19}}}$. $\Gamma(10) = 9!$ and $\Gamma(9.5) = (8.5)(7.5)\ldots(1.5)(.5)\Gamma(.5)$, but $\Gamma(.5) = \sqrt{p}$. Straightforward calculation gives c = 1.0132.

38.

a. The likelihood is

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2p s^2}} e^{-\frac{(x_i - m_i)}{2s^2}} \cdot \frac{1}{\sqrt{2p s^2}} e^{-\frac{(y_i - m_i)}{2s^2}} = \frac{1}{\left(2p s^2\right)^n} e^{-\frac{\left(\Sigma(x_i - m_i)^2 + \Sigma(y_i - m_i)^2\right)}{2s^2}}$$

. The log likelihood is thus $-n\ln\left(2p s^2\right) - \frac{\left(\Sigma(x_i - m_i)^2 + \Sigma(y_i - m_i)^2\right)}{2s^2}$. Taking $\frac{d}{dm_i}$ and equating to zero gives $\hat{m}_i = \frac{x_i + y_i}{2}$. Substituting these estimates of the $\hat{m}_i$'s into the log likelihood gives

$$-n\ln\left(2p s^2\right) - \frac{1}{2s^2}\left(\sum\left(x_i - \frac{x_i + y_i}{2}\right)^2 + \sum\left(y_i - \frac{x_i + y_i}{2}\right)^2\right)$$

$$= -n\ln\left(2p s^2\right) - \frac{1}{2s^2}\left(\tfrac{1}{2}\Sigma(x_i - y_i)^2\right).$$

Now taking $\frac{d}{ds^2}$, equating to zero, and solving for s^2 gives the desired result.

b. $E(\hat{s}) = \frac{1}{4n}E\left(\Sigma(X_i - Y_i)^2\right) = \frac{1}{4n} \cdot \Sigma E(X_i - Y)^2$, but

$E(X_i - Y)^2 = V(X_i - Y) + \left[E(X_i - Y)\right]^2 = 2s^2 + 0 = 2s^2$. Thus

$E\left(\hat{s}^2\right) = \frac{1}{4n}\Sigma\left(2s^2\right) = \frac{1}{4n}2ns^2 = \frac{s^2}{2}$, so the mle is definitely not unbiased; the expected value of the estimator is only half the value of what is being estimated!

CHAPTER 7

Section 7.1

1.

a. $z_{\alpha/2} = 2.81$ implies that $\alpha/2 = 1 - \Phi(2.81) = .0025$, so $\alpha = .005$ and the confidence level is $100(1-\alpha)\% = 99.5\%$.

b. $z_{\alpha/2} = 1.44$ for $\alpha = 2[1 - \Phi(1.44)] = .15$, and $100(1-\alpha)\% = 85\%$.

c. 99.7% implies that $\alpha = .003$, $\alpha/2 = .0015$, and $z_{.0015} = 2.96$. (Look for cumulative area .9985 in the main body of table A.3, the Z table.)

d. 75% implies $\alpha = .25$, $\alpha/2 = .125$, and $z_{.125} = 1.15$.

2.

a. The sample mean is the center of the interval, so $\bar{x} = \dfrac{114.4 + 115.6}{2} = 115$.

b. The interval (114.4, 115.6) has the 90% confidence level. The higher confidence level will produce a wider interval.

3.

a. A 90% confidence interval will be narrower (See 2b, above) Also, the z critical value for a 90% confidence level is 1.645, smaller than the z of 1.96 for the 95% confidence level, thus producing a narrower interval.

b. Not a correct statement. Once and interval has been created from a sample, the mean μ is either enclosed by it, or not. The 95% confidence is in the general procedure, for repeated sampling.

c. Not a correct statement. The interval is an estimate for the population mean, not a boundary for population values.

d. Not a correct statement. In theory, if the process were repeated an infinite number of times, 95% of the intervals would contain the population mean μ.

4.

a. $58.3 \pm \frac{1.96(3)}{\sqrt{25}} = 58.3 \pm 1.18 = (57.1, 59.5)$

b. $58.3 \pm \frac{1.96(3)}{\sqrt{100}} = 58.3 \pm .59 = (57.7, 58.9)$

c. $58.3 \pm \frac{2.58(3)}{\sqrt{100}} = 58.3 \pm .77 = (57.5, 59.1)$

d. 82% confidence $\Rightarrow 1 - \alpha = .82 \Rightarrow \alpha = .18 \Rightarrow \alpha/2 = .09$, so $z_{\alpha/2} = z_{.09} = 1.34$ and the interval is $58.3 \pm \frac{1.34(3)}{\sqrt{100}} = (57.9, 58.7)$.

e. $n = \left[\frac{2(2.58)3}{1}\right]^2 = 239.62$ so n = 240.

5.

a. $4.85 \pm \frac{(1.96)(.75)}{\sqrt{20}} = 4.85 \pm .33 =$ (4.52, 5.18).

b. $z_{\alpha/2} = z_{.02/2} = z_{.01} = 2.33$, so the interval is $4.56 \pm \frac{(2.33)(.75)}{\sqrt{16}} =$ (4.12, 5.00).

c. $n = \left[\frac{2(1.96)(.75)}{.40}\right]^2 = 54.02$, so n = 55.

d. $n = \left[\frac{2(2.58)(.75)}{.2}\right]^2 = 93.61$, so n = 94.

6.

a. $8439 \pm \frac{(1.645)(100)}{\sqrt{25}} = 8439 \pm 32.9 =$ (8406.1, 8471.9).

b. $1 - \alpha = .92 \Rightarrow \alpha = .08 \Rightarrow \alpha/2 = .04$ so $z_{\alpha/2} = z_{.04} = 1.75$

7. If $L = 2z_{a/2}\dfrac{s}{\sqrt{n}}$ and we increase the sample size by a factor of 4, the new length is

$L' = 2z_{a/2}\dfrac{s}{\sqrt{4n}} = \left[2z_{a/2}\dfrac{s}{\sqrt{n}}\right]\left(\dfrac{1}{2}\right) = \dfrac{L}{2}$. Thus halving the length requires n to be increased fourfold. If $n' = 25n$, then $L' = \dfrac{L}{5}$, so the length is decreased by a factor of 5.

8.

a. With probability $1-a$, $z_{a_1} \le \left(\overline{X} - m\right)\left(\dfrac{s}{\sqrt{n}}\right) \le z_{a_2}$. These inequalities can be manipulated exactly as was done in the text to isolate m; the result is

$\overline{X} - z_{a_2}\dfrac{s}{\sqrt{n}} \le m \le \overline{X} + z_{a_1}\dfrac{s}{\sqrt{n}}$, so a $100(1-a)\%$ interval is

$$\left(\overline{X} - z_{a_2}\frac{s}{\sqrt{n}}, \overline{X} + z_{a_1}\frac{s}{\sqrt{n}}\right)$$

b. The usual 95% interval has length $3.92\dfrac{s}{\sqrt{n}}$, while this interval will have length

$\left(z_{a_1} + z_{a_2}\right)\dfrac{s}{\sqrt{n}}$. With $z_{a_1} = z_{.0125} = 2.24$ and $z_{a_2} = z_{.0375} = 1.78$, the length is

$(2.24 + 1.78)\dfrac{s}{\sqrt{n}} = 4.02\dfrac{s}{\sqrt{n}}$, which is longer.

9.

a. $\left(\overline{x} - 1.645\dfrac{s}{\sqrt{n}}, \infty\right)$. From 5**a**, $\overline{x} = 4.85$, $s = .75$, n = 20;

$4.85 - 1.645\dfrac{.75}{\sqrt{20}} = 4.5741$, so the interval is $(4.5741, \infty)$.

b. $\left(\overline{x} - z_a\dfrac{s}{\sqrt{n}}, \infty\right)$

c. $\left(-\infty, \overline{x} + z_a\dfrac{s}{\sqrt{n}}\right)$; From 4**a**, $\overline{x} = 58.3$, $s = 3.0$, n = 25;

$58.3 + 2.33\dfrac{3}{\sqrt{25}} = (-\infty, 59.70)$

10.

a. When n = 15, $2l\sum X_i$ has a chi-squared distribution with 30 d.f. From the 30 d.f. row of Table A.6, the critical values that capture lower and upper tail areas of .025 (and thus a central area of .95) are 16.791 and 46.979. An argument parallel to that given in Example 7.5 gives $\left(\frac{2\sum x_i}{46.979}, \frac{2\sum x_i}{16.791}\right)$ as a 95% C. I. for $m = \frac{1}{l}$. Since $\sum x_i = 63.2$ the interval is (2.69, 7.53).

b. A 99% confidence level requires using critical values that capture area .005 in each tail of the chi-squared curve with 30 d.f.; these are 13.787 and 53.672, which replace 16.791 and 46.979 in **a**.

c. $V(X) = \frac{1}{l^2}$ when X has an exponential distribution, so the standard deviation is $\frac{1}{l}$, the same as the mean. Thus the interval of **a** is also a 95% C.I. for the standard deviation of the lifetime distribution.

11. Y is a binomial r.v. with n = 1000 and p = .95, so E(Y) = np = 950, the expected number of intervals that capture m, and $s_Y = \sqrt{npq} = 6.892$. Using the normal approximation to the binomial distribution, $P(940 \le Y \le 960) = P(939.5 \le Y_{normal} \le 960.5) = P(-1.52 \le Z \le 1.52)$ $= .9357 - .0643 = .8714$.

Section 7.2

12. $$\bar{x} \pm 2.58\frac{s}{\sqrt{n}} = .81 \pm 2.58\frac{.34}{\sqrt{110}} = .81 \pm .08 = (.73, .89)$$

13.

a. $$\bar{x} \pm z_{.025}\frac{s}{\sqrt{n}} = 1.028 \pm 1.96\frac{.163}{\sqrt{69}} = 1.028 \pm .038 = (.990, 1.066)$$

b. $$w = .05 = \frac{2(1.96)(.16)}{\sqrt{n}} \Rightarrow \sqrt{n} = \frac{2(1.96)(.16)}{.05} = 12.544 \Rightarrow n = (12.544)^2 \approx 158$$

14.

a. $89.10 \pm 1.96\frac{3.73}{\sqrt{169}} = 89.10 \pm .56 = (88.54,.89.66)$. Yes, this is a very narrow interval. It appears quite precise.

b. $n = \left[\frac{(1.96)(.16)}{.5}\right]^2 = 245.86 \Rightarrow n = 246$.

15.

a. $z_a = .84$, and $\Phi(.84) = .7995 \approx .80$, so the confidence level is 80%.

b. $z_a = 2.05$, and $\Phi(2.05) = .9798 \approx .98$, so the confidence level is 98%.

c. $z_a = .67$, and $\Phi(.67) = .7486 \approx .75$, so the confidence level is 75%.

16. n = 46, $\bar{x} = 382.1$, s = 31.5; The 95% upper confidence bound =
$$\bar{x} + z_a \frac{s}{\sqrt{n}} = 382.1 + 1.645\frac{31.5}{\sqrt{46}} = 382.1 + 7.64 = 389.74$$

17. $\bar{x} - z_{.01}\frac{s}{\sqrt{n}} = 135.39 - 2.33\frac{4.59}{\sqrt{153}} = 135.39 - .865 = 134.53$ With a confidence level of 99%, the true average ultimate tensile strength is between (134.53, ∞).

18. 90% lower bound: $\bar{x} - z_{.10}\frac{s}{\sqrt{n}} = 4.25 - 1.28\frac{1.30}{\sqrt{75}} = 4.06$

19. $\hat{p} = \frac{201}{356} = .5646$; We calculate a 95% confidence interval for the proportion of all dies that pass the probe:
$$\frac{.5646 + \frac{(1.96)^2}{2(356)} \pm 1.96\sqrt{\frac{(.5646)(.4354)}{356} + \frac{(1.96)^2}{4(356)^2}}}{1 + \frac{(1.96)^2}{356}} = \frac{.5700 \pm .0518}{1.01079} = (.513,.615)$$

20. Because the sample size is so large, the simpler formula (7.11) for the confidence interval for p is sufficient.

$$.15 \pm 2.58\sqrt{\frac{(.15)(.85)}{4722}} = .15 \pm .013 = (.137,.163)$$

21. $\hat{p} = \frac{133}{539} = .2468$; the 95% lower confidence bound is:

$$\frac{.2468 + \frac{(1.645)^2}{2(539)} - 1.645\sqrt{\frac{(.2468)(.7532)}{539} + \frac{(1.645)^2}{4(539)^2}}}{1 + \frac{(1.645)^2}{539}} = \frac{.2493 - .0307}{1.005} = .218$$

22. $\hat{p} = .072$; the 99% upper confidence bound is:

$$\frac{.072 + \frac{(2.33)^2}{2(487)} + 2.33\sqrt{\frac{(.072)(.928)}{487} + \frac{(2.33)^2}{4(487)^2}}}{1 + \frac{(2.33)^2}{487}} = \frac{.0776 + .0279}{1.0111} = .1043$$

23.

a. $\hat{p} = \frac{24}{37} = .6486$; The 99% confidence interval for p is

$$\frac{.6486 + \frac{(2.58)^2}{2(37)} \pm 2.58\sqrt{\frac{(.6486)(.3514)}{37} + \frac{(2.58)^2}{4(37)^2}}}{1 + \frac{(2.58)^2}{37}} = \frac{.7386 \pm .2216}{1.1799} = (.438,.814)$$

b. $$n = \frac{2(2.58)^2(.25) - (2.58)^2(.01) \pm \sqrt{4(2.58)^4(.25)(.25 - .01) + .01(2.58)^4}}{.01}$$

$$= \frac{3.261636 \pm 3.3282}{.01} \approx 659$$

24. n = 56, $\bar{x} = 8.17$, s = 1.42; For a 95% C.I., $z_{\alpha/2} = 1.96$. The interval is

$8.17 \pm 1.96\left(\frac{1.42}{\sqrt{56}}\right) = (7.798,8.542)$. We make no assumptions about the distribution if percentage elongation.

25.

a. $n = \dfrac{2(1.96)^2(.25) - (1.96)^2(.01) \pm \sqrt{4(1.96)^4(.25)(.25-.01) + .01(1.96)^4}}{.01} \approx 381$

b. $n = \dfrac{2(1.96)^2\left(\frac{1}{3}\cdot\frac{2}{3}\right) - (1.96)^2(.01) \pm \sqrt{4(1.96)^4\left(\frac{1}{3}\cdot\frac{2}{3}\right)\left(\frac{1}{3}\cdot\frac{2}{3} - .01\right) + .01(1.96)^4}}{.01} \approx 339$

26. With $q = l$, $\hat{q} = \overline{X}$ and $s_{\hat{q}} = \sqrt{\dfrac{l}{n}}$ so $\hat{s}_{\hat{q}} = \sqrt{\dfrac{\overline{X}}{n}}$. The large sample C.I. is then $\overline{x} \pm z_{a/2}\sqrt{\dfrac{\overline{x}}{n}}$. We calculate $\sum x_i = 203$, so $\overline{x} = 4.06$, and a 95% interval for l is

$4.06 \pm 1.96\sqrt{\dfrac{4.06}{50}} = 4.06 \pm .56 = (3.50, 4.62)$

27. Note that the midpoint of the new interval is $\dfrac{x + \frac{z^2}{2}}{n + z^2}$, which is roughly $\dfrac{x+2}{n+4}$ with a confidence level of 95% and approximating $1.96 \approx 2$. The variance of this quantity is $\dfrac{np(1-p)}{(n+z^2)^2}$, or roughly $\dfrac{p(1-p)}{n+4}$. Now replacing p with $\dfrac{x+2}{n+4}$, we have

$\left(\dfrac{x+2}{n+4}\right) \pm z_{a/2}\sqrt{\dfrac{\left(\dfrac{x+2}{n+4}\right)\left(1 - \dfrac{x+2}{n+4}\right)}{n+4}}$; For clarity, let $x^* = x+2$ and $n^* = n+4$, then $\hat{p}^* = \dfrac{x^*}{n^*}$ and the formula reduces to $\hat{p}^* \pm z_{a/2}\sqrt{\dfrac{\hat{p}^*\hat{q}^*}{n^*}}$, the desired conclusion. For further discussion, see the Agresti article.

Section 7.3

28.

a. 1.341

b. 1.753

c. 1.708

d. 1.684

e. 2.704

29.

a. $t_{.025,10} = 2.228$

b. $t_{.025,20} = 2.086$

c. $t_{.005,20} = 2.845$

d. $t_{.005,50} = 2.678$

e. $t_{.01,25} = 2.485$

f. $-t_{.025,5} = -2.571$

30.

a. $t_{.025,10} = 2.228$

b. $t_{.025,15} = 2.131$

c. $t_{.005,15} = 2.947$

d. $t_{.005,4} = 4.604$

e. $t_{.01,24} = 2.492$

f. $t_{.005,37} \approx 2.712$

31.

a. $t_{05,10} = 1.812$

b. $t_{.05,15} = 1.753$

c. $t_{.01,15} = 2.602$

d. $t_{.01,4} = 3.747$

e. $\approx t_{.025,24} = 2.064$

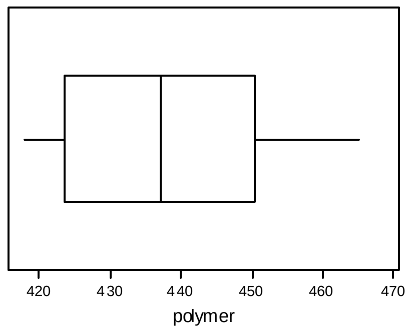
f. $t_{.01,37} \approx 2.429$

32. d.f. = n – 1 = 7, so the critical value for a 95% C.I. is $t_{.025,7} = 2.365$. The interval is

$$30.2 \pm (2.365)\left(\frac{3.1}{\sqrt{8}}\right) = 30.2 \pm 2.6 = (27.6, 32.8).$$

33.

a. The boxplot indicates a very slight positive skew, with no outliers. The data appears to center near 438.



b. Based on a normal probability plot, it is reasonable to assume the sample observations came from a normal distribution.

c. With d.f. = n – 1 = 16, the critical value for a 95% C.I. is $t_{.025,16} = 2.120$, and the interval is $438.29 \pm (2.120)\left(\frac{15.14}{\sqrt{17}}\right) = 438.29 \pm 7.785 = (430.51, 446.08)$. Since 440 is within the interval, 440 is a plausible value for the true mean. 450, however, is not, since it lies outside the interval.

34. n = 14, $\bar{x} = 8.48$, s = .79; $t_{.05,13} = 1.771$

a. A 95% lower confidence bound: $8.48 - 1.771\left(\frac{.79}{\sqrt{14}}\right) = 8.48 - .37 = 8.11$. With 95% confidence, the value of the true mean proportional limit stress of all such joints lies in the interval $(8.11, \infty)$. If this interval is calculated for sample after sample, in the long run 95% of these intervals will include the true mean proportional limit stress of all such joints. We must assume that the sample observations were taken from a normally distributed population.

b. A 95% lower prediction bound: $8.48 - 1.771(.79)\sqrt{1 + \frac{1}{14}} = 8.48 - 1.45 = 7.03$. If this bound is calculated for sample after sample, in the long run 95% of these bounds will provide a lower bound for the corresponding future values of the proportional limit stress of a single joint of this type.

35. n = 5, $\bar{x} = 2887.6$, s = .84.0; $t_{.025,4} = 2.776$

a. A 95% C.I. for the mean: $2887.6 \pm (2.776)\left(\frac{84}{\sqrt{5}}\right) \Rightarrow (2783.3, 2991.9)$

b. A 95% Prediction Interval: $2887.6 \pm 2.776(84)\sqrt{1+\frac{1}{5}} \Rightarrow (2632.1, 3143.1)$. The P.I. is considerably larger than the C.I., about 2.5 times larger.

36. n = 26, $\bar{x} = 370.69$, s = 24.36; $t_{.05,25} = 1.708$

a. A 95% upper confidence bound:

$$370.69 + (1.708)\left(\frac{24.36}{\sqrt{26}}\right) = 370.69 + 8.16 = 378.85$$

b. A 95% upper prediction bound:

$$370.69 + 1.708(24.36)\sqrt{1+\frac{1}{26}} = 370.69 + 42.45 = 413.14$$

c. Following a similar argument as that on p. 300 of the text, we need to find the variance of $\bar{X} - \bar{X}_{new}$: $V(\bar{X} - \bar{X}_{new}) = V(\bar{X}) + V(\bar{X}_{new}) = V(\bar{X}) + V\left(\tfrac{1}{2}(X_{27} + X_{28})\right)$
$= V(\bar{X}) + V\left(\tfrac{1}{2}X_{27}\right) + V\left(\tfrac{1}{2}X_{28}\right) = V(\bar{X}) + \tfrac{1}{4}V(X_{27}) + \tfrac{1}{4}V(X_{28})$
$= \frac{s^2}{n} + \frac{1}{4}s^2 + \frac{1}{4}s^2 = s^2\left(\frac{1}{2} + \frac{1}{n}\right)$. We eventually arrive at $T = \frac{\bar{X} - \bar{X}_{new}}{s\sqrt{\frac{1}{2}+\frac{1}{n}}} \sim$ t distribution with n – 1 d.f., so the new prediction interval is $\bar{x} \pm t_{\alpha/2,n-1} \cdot s\sqrt{\frac{1}{2}+\frac{1}{n}}$. For this situation, we have

$$370.69 \pm 1.708(24.36)\sqrt{\frac{1}{2}+\frac{1}{26}} = 370.69 \pm 30.53 = (39.47, 400.53)$$

37.

a. A 95% C.I. : $.9255 \pm 2.093(.0181) = .9255 \pm .0379 \Rightarrow (.8876, .9634)$

b. A 95% P.I. : $.9255 \pm 2.093(.0809)\sqrt{1+\frac{1}{20}} = .9255 \pm .1735 \Rightarrow (.7520, 1.0990)$

c. A tolerance interval is requested, with k = 99, confidence level 95%, and n = 20. The tolerance critical value, from Table A.6, is 3.615. The interval is $.9255 \pm 3.615(.0809) \Rightarrow (.6330, 1.2180)$.

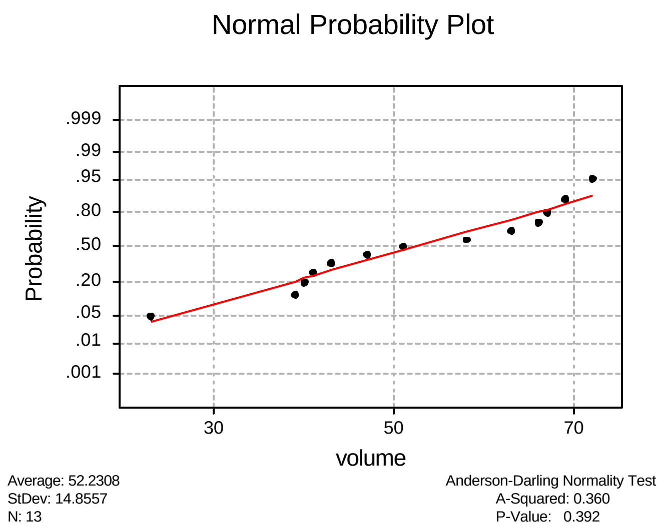
38. N = 25, $\bar{x} = .0635$, s = .0065

a. 95% P.I. : $.0635 \pm 2.064(.0065)\sqrt{1+\frac{1}{25}} = .0635 \pm .0137 \Rightarrow (.0498,.0772)$.

b. 99% Tolerance Interval, with k = 95, critical value 2.972 (table A.6):
$.0635 \pm 2.972(.0065) \Rightarrow (.0442,.0828)$.

39.

a.



Based on the above plot, generated by Minitab, it is plausible that the population distribution is normal.

b. We require a tolerance interval. (from table A6, with 95% confidence, k = 95, and n=13, the tcv = 3.081.

$$\bar{x} \pm (tcv)s = 52.231 \pm 3.081(14.856) = 52.231 \pm 45.771 \Rightarrow (6.460,98.002)$$

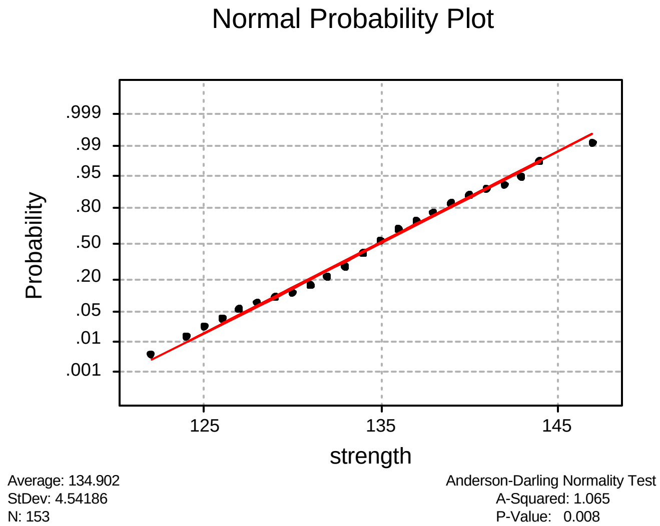
c. A prediction interval, with $t_{.025,12} = 2.179$:

$$52.231 \pm 2.179(14.856)\sqrt{1+\frac{1}{13}} = 52.231 \pm 33.593 \Rightarrow (18.638,85.824)$$

40.

a. We need to assume the samples came from a normally distributed population.

b. A Normal Probability plot, generated by Minitab:



The very small p-value indicates that the population distribution from which this data was taken is most likely not normal.

c. 95% lower prediction bound:

$$52.231 \pm 2.179(14.856)\sqrt{1+\tfrac{1}{13}} = 52.231 \pm 33.593 \Rightarrow (18.638, 85.824)$$

41. The 20 d.f. row of Table A.5 shows that 1.725 captures upper tail area .05 and 1.325 captures uppertail area .10 The confidence level for each interval is 100(central area)%. For the first interval, central area = 1 – sum of tail areas = 1 – (.25 + .05) = .70, and for the second and third intervals the central areas are 1 – (.20 + .10) = .70 and 1 – (.15 + .15) = 70. Thus each interval has confidence level 70%. The width of the first interval is

$$\frac{s(.687+1.725)}{\sqrt{n}} = \frac{.2412s}{\sqrt{n}}$$

, whereas the widths of the second and third intervals are 2.185 and 2.128 respectively. The third interval, with symmetrically placed critical values, is the shortest, so it should be used. This will always be true for a t interval.

Section 7.2

42.

a. $c^{2}_{.1,15} = 22.307$ (.1 column, 15 d.f. row)

b. $c^{2}_{.1,25} = 34.381$

c. $c^{2}_{.01,25} = 44.313$

d. $c^{2}_{.005,25} = 46.925$

e. $c^{2}_{.99,25} = 11.523$ (from .99 column, 25 d.f. row)

f. $c^{2}_{.995,25} = 10.519$

43.

a. $c^{2}_{.05,10} = 18.307$

b. $c^{2}_{.95,10} = 3.940$

c. Since $10.987 = c^{2}_{.975,22}$ and $36.78 = c^{2}_{.025,22}$, $P\left(c^{2}_{.975,22} \le c^{2} \le c^{2}_{.025,22}\right) = .95$.

d. Since $14.61 = c^{2}_{.95,25}$ and $37.65 = c^{2}_{.05,25}$, $P\left(c^{2}_{.95,25} \le c^{2} \le c^{2}_{.05,25}\right) = .90$.

44. n – 1 = 8, $c^{2}_{.025,8} = 17.543$, $c^{2}_{.975,8} = 2.180$, so the 95% interval for s^{2} is $\left(\frac{8(7.90)}{17.543}, \frac{8(7.90)}{2.180}\right) = (3.60, 28.98)$. The 95% interval for s is $\left(\sqrt{3.60}, \sqrt{28.98}\right) = (1.90, 5.38)$.

45. n = 22 implies that d.f. = n – 1 = 21, so the .995 and .005 columns of Table A.7 give the necessary chi-squared critical values as 8.033 and 41.399. $\Sigma x_i = 1701.3$ and $\Sigma x_i^2 = 132{,}097.35$, so $s^2 = 25.368$. The interval for s^{2} is $\left(\frac{21(25.368)}{41.399}, \frac{21(25.368)}{8.033}\right) = (12.868, 66.317)$ and that for s is $(3.6, 8.1)$ Validity of this interval requires that fracture toughness be (at least approximately) normally distributed.

46.

a. Using a normal probability plot, we ascertain that it is plausible that this sample was taken from a normal population distribution.

b. With s = 1.579 , n = 15, and $c^{2}_{.05,14} = 23.685$ the 95% upper confidence bound for s is $\sqrt{\frac{14(1.579)^2}{23.685}} = 1.214$

Supplementary Exercises

47.

a. n = 48, $\bar{x} = 8.079$, s^2 = 23.7017, and s = 4.868.
A 95% C.I. for m = the true average strength is

$$\bar{x} \pm 1.96\frac{s}{\sqrt{n}} = 8.079 \pm 1.96\frac{4.868}{\sqrt{48}} = 8.079 \pm 1.377 = (6.702, 9.456)$$

b. $\hat{p} = \frac{13}{48} = .2708$. A 95% C.I. is

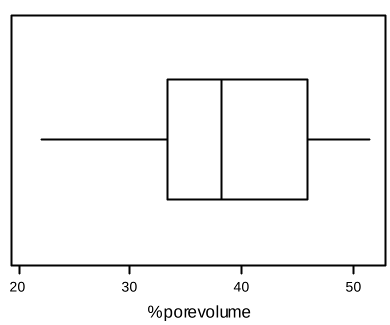
$$\frac{.2708 + \frac{1.96^2}{2(48)} \pm 1.96\sqrt{\frac{(.2708)(.7292)}{48} + \frac{1.96^2}{4(48)^2}}}{1 + \frac{1.96^2}{48}} = \frac{.3108 \pm .1319}{1.0800} = (.166, .410)$$

48. A 98% t C.I. requires $t_{a/2,n-1} = t_{.01,8} = 2.896$. The interval is

$$188.0 \pm 2.896\frac{7.2}{\sqrt{9}} = 188.0 \pm 7.0 = (181.0, 195.0).$$

49.

a. There appears to be a slight positive skew in the middle half of the sample, but the lower whisker is much longer than the upper whisker. The extent of variability is rather substantial, although there are no outliers.



b. The pattern of points in a normal probability plot is reasonably linear, so, yes, normality is plausible.

c. n = 18, $\bar{x} = 38.66$, s = 8.473, and $t_{.01,17} = 2.586$. The 98% confidence interval is

$$38.66 \pm 2.586\frac{8.473}{\sqrt{18}} = 38.66 \pm 5.13 = (33.53, 43.79).$$

50. $\bar{x} =$ the middle of the interval = $\dfrac{229.764+233.502}{2} = 231.633.$ To find *s* we use $width = 2(t_{.025,4})\left(\dfrac{s}{\sqrt{n}}\right)$, and solve for *s*. Here, n = 5, $t_{.025,4} = 2.776$, and width = upper limit – lower limit = 3.738. $3.738 = 2(2776)\dfrac{s}{\sqrt{5}} \Rightarrow s = \dfrac{\sqrt{5}(3.738)}{2(2.776)} = 1.5055$. So for a 99% C.I., $t_{.005,4} = 4.604$, and the interval is

$$231.633 \pm 4.604\frac{1.5055}{\sqrt{5}} = 213.633 \pm 3.100 = (228.533, 234.733).$$

51.

a. $\hat{p} = \dfrac{136}{200} = .680 \Rightarrow$ a 90% C.I. is

$$\frac{.680 + \frac{1.645^2}{2(200)} \pm 1.645\sqrt{\frac{(.680)(.320)}{200} + \frac{1.645^2}{4(200)^2}}}{1 + \frac{1.645^2}{200}} = \frac{.6868 \pm .0547}{1.01353} = (.624, .732)$$

b. $$n = \frac{2(1.645)^2(.25) - (1.645)^2(.05)^2 \pm \sqrt{4(1.645)^4(.25)(.25 - .0025) + .05^2(1.645)^4}}{.0025}$$

$$= \frac{1.3462 \pm 1.3530}{.0025} = 1079.7 \Rightarrow \text{ use n = 1080}$$

c. No, it gives a 95% upper bound.

52.

a. Assuming normality, $t_{.05,15} = 1.753$, do s 95% C.I. for m is

$$.214 \pm 1.753\frac{.036}{\sqrt{16}} = .214 \pm .016 = (.198, .230)$$

b. A 90% upper bound for s, with $c^2_{.10,15} = 1.341$, is $\sqrt{\dfrac{15(.036)^2}{1.341}} = \sqrt{.0145} = .120$

c. A 95% prediction interval, with $t_{.025,15} = 2.131$, is

$$.214 \pm 2.131(.036)\sqrt{1 + \tfrac{1}{16}} = .214 \pm .0791 = (.1349, .2931).$$

53. With $\hat{q} = \frac{1}{3}\left(\overline{X}_1 + \overline{X}_2 + \overline{X}_3\right) - \overline{X}_4$, $s_{\hat{q}}^2 = \frac{1}{9}Var\left(\overline{X}_1 + \overline{X}_2 + \overline{X}_3\right) + Var\left(\overline{X}_4\right) =$ $\frac{1}{9}\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} + \frac{s_3^2}{n_3}\right) + \frac{s_4^2}{n_4}$; $\hat{s}_{\hat{q}}$ is obtained by replacing each $\hat{s}_i^2$ by s_i^2 and taking the square root. The large-sample interval for q is then $\frac{1}{3}\left(\bar{x}_1 + \bar{x}_2 + \bar{x}_3\right) - \bar{x}_4 \pm z_{a/2}\sqrt{\frac{1}{9}\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} + \frac{s_3^2}{n_3}\right) + \frac{s_4^2}{n_4}}$. For the given data, $\hat{q} = -.50$, $\hat{s}_{\hat{q}} = .1718$, so the interval is $-.50 \pm 1.96(.1718) = (-.84, -.16)$.

54. $\hat{p} = \frac{11}{55} = .2 \Rightarrow$ a 90% C.I. is

$$\frac{.2 + \frac{1.645^2}{2(55)} \pm 1.645\sqrt{\frac{(.2)(.8)}{55} + \frac{1.645^2}{4(55)^2}}}{1 + \frac{1.645^2}{55}} = \frac{.2246 \pm .0887}{1.0492} = (.1295, .2986).$$

55. The specified condition is that the interval be length .2, so $n = \left[\frac{2(1.96)(.8)}{.2}\right]^2 = 245.86$, so n = 246 should be used.

56.

a. A normal probability plot lends support to the assumption that pulmonary compliance is normally distributed. Note also that the lower and upper fourths are 192.3 and 228,1, so the fourth spread is 35.8, and the sample contains no outliers.

b. $t_{.025,15} = 2.131$, so the C.I. is

$$209.75 \pm 2.131\frac{24.156}{\sqrt{16}} = 209.75 \pm 12.87 = (196.88, 222.62).$$

c. K = 95, n = 16, and the tolerance critical value is 2.903, so the 95% tolerance interval is $209.75 \pm 2.903(24.156) = 209.75 \pm 70.125 = (139.625, 279.875)$.

57. Proceeding as in Example 7.5 with T_r replacing ΣX_i, the C.I. for $\frac{1}{l}$ is $\left(\frac{2t_r}{c^2_{1-\alpha/2,2r}}, \frac{2t_r}{c^2_{\alpha/2,2r}}\right)$ where $t_r = y_1 + ... + y_r + (n-r)y_r$. In Example 6.7, n = 20, r = 10, and t_r = 1115. With d.f. = 20, the necessary critical values are 9.591 and 34.170, giving the interval (65.3, 232.5). This is obviously an extremely wide interval. The censored experiment provides less information about $\frac{1}{l}$ than would an uncensored experiment with n = 20.

58.

a. $P(\min(X_i) \le \tilde{m} \le \max(X_i)) = 1 - P(\tilde{m}, < \min(X_i) \, or \, \max(X_i) < \tilde{m})$
$= 1 - P(\tilde{m}, < \min(X_i)) - P(\max(X_i) < \tilde{m})$
$= 1 - P(\tilde{m} < X_1, ..., \tilde{m} < X_n) - P(X_1 < \tilde{m}, ..., X_n < \tilde{m})$
$= 1 - (.5)^n - (.5)^n = 1 - 2(.5)^{n-1}$, from which the confidence interval follows.

b. Since $\min(x_i) = 1.44$ and $\max(x_i) = 3.54,$ the C.I. is (1.44, 3.54).

c. $P(X_{(2)} \le \tilde{m} \le X_{(n-1)}) = 1 - P(\tilde{m}, < X_{(2)}) - P(X_{(n-1)} < \tilde{m})$
= 1 – P(at most one X_I is below $\tilde{m}$) – P(at most one X_I exceeds $\tilde{m}$)
$1 - (.5)^n - \binom{n}{1}(.5)^1 (.5)^{n-1} - (.5)^n - \binom{n}{1}(.5)^{n-1}(.5).$
$= 1 - 2(n+1)(.5)^n = 1 - (n+1)(.5)^{n-1}$
Thus the confidence coefficient is $1 - (n+1)(.5)^{n-1}$, or in another way, a $100\left(1 - (n+1)(.5)^{n-1}\right)\%$ confidence interval.

59.

a. $\int_{(a/2)^{1/n}}^{(1-a/2)^{1/n}} nu^{n-1} du = u^n \Big]_{(a/2)^{1/n}}^{(1-a/2)^{1/n}} = 1 - \frac{a}{2} - \frac{a}{2} = 1 - a$. From the probability statement, $\frac{(a/2)^{1/n}}{\max(X_i)} \le \frac{1}{q} \le \frac{(1 - a/2)^{1/n}}{\max(X_i)}$ with probability $1 - a$, so taking the reciprocal of each endpoint and interchanging gives the C.I. $\left(\frac{\max(X_i)}{(1 - a/2)^{1/n}}, \frac{\max(X_i)}{(a/2)^{1/n}} \right)$ for q.

b. $a^{1/n} \le \frac{\max(X_i)}{q} \le 1$ with probability $1 - a$, so $1 \le \frac{q}{\max(X_i)} \le \frac{1}{a^{1/n}}$ with probability $1 - a$, which yields the interval $\left(\max(X_i), \frac{\max(X_i)}{a^{1/n}} \right)$.

c. It is easily verified that the interval of **b** is shorter – draw a graph of $f_U(u)$ and verify that the shortest interval which captures area $1 - a$ under the curve is the rightmost such interval, which leads to the C.I. of **b**. With $a = .05,$ n = 5, max(x_I)=4.2; this yields (4.2, 7.65).

60. The length of the interval is $\left(z_g + z_{a-g}\right)\frac{s}{\sqrt{n}}$, which is minimized when $z_g + z_{a-g}$ is minimized, i.e. when $\Phi^{-1}(1-g) + \Phi^{-1}(1-a+g)$ is minimized. Taking $\frac{d}{dg}$ and equating to 0 yields $\frac{1}{\Phi(1-g)} = \frac{1}{\Phi(1-a+g)}$ where $\Phi(\bullet)$ is the standard normal p.d.f., whence $g = \frac{a}{2}$.

61. $\tilde{x} = 76.2,$ the lower and upper fourths are 73.5 and 79.7, respectively, and $f_s = 6.2.$ The robust interval is $76.2 \pm (1.93)\left(\frac{6.2}{\sqrt{22}}\right) = 76.2 \pm 2.6 = (73.6, 78.8)$.
$\bar{x} = 77.33$, s = 5.037, and $t_{.025,21} = 2.080$, so the t interval is
$77.33 \pm (2.080)\left(\frac{5.037}{\sqrt{22}}\right) = 77.33 \pm 2.23 = (75.1, 79.6)$. The t interval is centered at $\bar{x}$, which is pulled out to the right of $\tilde{x}$ by the single mild outlier 93.7; the interval widths are comparable.

62.

a. Since $2l\Sigma X_i$ has a chi-squared distribution with 2n d.f. and the area under this chi-squared curve to the right of $c^2_{.95,2n}$ is .95, $P\left(c^2_{.95,2n} < 2l\Sigma X_i\right) = .95$. This implies that $\frac{c^2_{.95,2n}}{2\Sigma X_i}$ is a lower confidence bound for l with confidence coefficient 95%. Table A.7 gives the chi-squared critical value for 20 d.f. as 10.851, so the bound is $\frac{10.851}{2(550.87)} = .0098$. We can be 95% confident that l exceeds .0098.

b. Arguing as in **a**, $P\left(2l\Sigma X_i < c^2_{.05,2n}\right) = .95$. The following inequalities are equivalent to the one in parentheses:

$$l < \frac{c^2_{.05,2n}}{2\Sigma X_i} \Rightarrow -lt < \frac{-tc^2_{.05,2n}}{2\Sigma X_i} \Rightarrow e^{-lt} < \exp\left[\frac{-tc^2_{.05,2n}}{2\Sigma X_i}\right].$$

Replacing the ΣX_i by Σx_i in the expression on the right hand side of the last inequality gives a 95% lower confidence bound for e^{-lt}. Substituting t = 100, $c^2_{.05,20} = 31.410$ and $\Sigma x_i = 550.87$ gives .058 as the lower bound for the probability that time until breakdown exceeds 100 minutes.

CHAPTER 8

Section 8.1

1.

a. Yes. It is an assertion about the value of a parameter.

b. No. The sample median $\tilde{X}$ is not a parameter.

c. No. The sample standard deviation s is not a parameter.

d. Yes. The assertion is that the standard deviation of population #2 exceeds that of population #1

e. No. $\overline{X}$ and $\overline{Y}$ are statistics rather than parameters, so cannot appear in a hypothesis.

f. Yes. H is an assertion about the value of a parameter.

2.

a. These hypotheses comply with our rules.

b. H_o is not an equality claim (e.g. $s = 20$), so these hypotheses are not in compliance.

c. H_o should contain the equality claim, whereas H_a does here, so these are not legitimate.

d. The asserted value of $m_1 - m_2$ in H_o should also appear in H_a. It does not here, so our conditions are not met.

e. Each S^2 is a statistic, so does not belong in a hypothesis.

f. We are not allowing both H_o and H_a to be equality claims (though this is allowed in more comprehensive treatments of hypothesis testing).

g. These hypotheses comply with our rules.

h. These hypotheses are in compliance.

3. In this formulation, H_o states the welds do not conform to specification. This assertion will not be rejected unless there is strong evidence to the contrary. Thus the burden of proof is on those who wish to assert that the specification is satisfied. Using H_a: $m < 100$ results in the welds being believed in conformance unless provided otherwise, so the burden of proof is on the non-conformance claim.

4. When the alternative is H_a: $m < 5$, the formulation is such that the water is believed unsafe until proved otherwise. A type I error involved deciding that the water is safe (rejecting H_o) when it isn't (H_o is true). This is a very serious error, so a test which ensures that this error is highly unlikely is desirable. A type II error involves judging the water unsafe when it is actually safe. Though a serious error, this is less so than the type I error. It is generally desirable to formulate so that the type 1 error is more serious, so that the probability of this error can be explicitly controlled. Using H_a: $m > 5$, the type II error (now stating that the water is safe when it isn't) is the more serious of the two errors.

5. Let s denote the population standard deviation. The appropriate hypotheses are $H_o: s = .05$ vs $H_a: s < .05$. With this formulation, the burden of proof is on the data to show that the requirement has been met (the sheaths will not be used unless H_o can be rejected in favor of H_a. Type I error: Conclude that the standard deviation is < .05 mm when it is really equal to .05 mm. Type II error: Conclude that the standard deviation is .05 mm when it is really < .05.

6. $H_o: m = 40$ vs $H_a: m \neq 40$, where m is the true average burn-out amperage for this type of fuse. The alternative reflects the fact that a departure from $m = 40$ in either direction is of concern. Notice that in this formulation, it is initially believed that the value of m is the design value of 40.

7. A type I error here involves saying that the plant is not in compliance when in fact it is. A type II error occurs when we conclude that the plant is in compliance when in fact it isn't. Reasonable people may disagree as to which of the two errors is more serious. If in your judgement it is the type II error, then the reformulation $H_o: m = 150$ vs $H_a: m < 150$ makes the type I error more serious.

8. Let $m_1 =$ the average amount of warpage for the regular laminate, and $m_2 =$ the analogous value for the special laminate. Then the hypotheses are $H_o: m_1 = m_2$ vs $H_o: m_1 > m_2$. Type I error: Conclude that the special laminate produces less warpage than the regular, when it really does not. Type II error: Conclude that there is no difference in the two laminates when in reality, the special one produces less warpage.

9.

a. R_1 is most appropriate, because x either too large or too small contradicts p = .5 and supports p $\neq$ .5.

b. A type I error consists of judging one of the tow candidates favored over the other when in fact there is a 50-50 split in the population. A type II error involves judging the split to be 50-50 when it is not.

c. X has a binomial distribution with n = 25 and p = 0.5. $a =$ P(type I error) = $P(X \leq 7 or X \geq 18$ when X ~ Bin(25, .5)) = B(7; 25,.5) + 1 – B(17; 25,.5) = .044

d. $b(.4) = P(8 \leq X \leq 17$ when p = .4) = B(17; 25,.5) – B(7, 25,.4) = 0.845, and $b(.6) = 0.845$ also. $b(.3) = B(17;25,.3) - B(7;25,.3) = .488 = b(.7)$

e. x = 6 is in the rejection region R_1, so H_o is rejected in favor of H_a.

10.

a. $H_o : m = 1300$ vs $H_a : m > 1300$

b. $\bar{x}$ is normally distributed with mean $E(\bar{x}) = m$ and standard deviation $\frac{s}{\sqrt{n}} = \frac{60}{\sqrt{20}} = 13.416.$ When H_o is true, $E(\bar{x}) = 1300$. Thus $a = P(\bar{x} \geq 1331.26$ when H_o is true) =

$$P\left(z \geq \frac{1331.26 - 1300}{13.416}\right) = P(z \geq 2.33) = .01$$

c. When $m = 1350$, $\bar{x}$ has a normal distribution with mean 1350 and standard deviation 13.416, so $b(1350) = P(\bar{x} < 1331.26$ when $m = 1350$) =

$$P\left(z \leq \frac{1331.26 - 1350}{13.416}\right) = P(z \leq -1.40) = .0808$$

d. Replace 1331.26 by c, where c satisfies $\frac{c - 1300}{13.416} = 1.645$ (since $P(z \geq 1.645) = .05)$. Thus c = 1322.07. Increasing a gives a decrease in b; now $b(1350) = P(z \leq -2.08) = .0188$.

e. $\bar{x} \geq 1331.26$ iff $z \geq \frac{1331.26 - 1300}{13.416}$ i.e. iff $z \geq 2.33$.

11.

a. $H_o : m = 10$ vs $H_a : m \neq 10$

b. $a = P($ rejecting H$_o$ when H$_o$ is true) = $P(\bar{x} \geq 10.1032$ or $\leq 9.8968 when m = 10)$.
Since $\bar{x}$ is normally distributed with standard deviation
$\frac{s}{\sqrt{n}} = \frac{.2}{5} = .04,\ a = P(z \geq 2.58 or \leq -2.58) = .005 + .005 = .01$

c. When $m = 10.1$, $E(\bar{x}) = 10.1$, so $b(10.1) = P(9.8968 < \bar{x} < 10.1032$ when $m = 10.1) = P(-5.08 < z < .08) = .5319$. Similarly,
$b(9.8) = P(2.42 < z < 7.58) = .0078$

d. $c = \pm 2.58$

e. Now $\frac{s}{\sqrt{n}} = \frac{.2}{3.162} = .0632$. Thus 10.1032 is replaced by c, where $\frac{c - 10}{.0632} = 1.96$ and so c = 10.124. Similarly, 9.8968 is replaced by 9.876.

f. $\bar{x} = 10.020$. Since $\bar{x}$ is neither ≥ 10.124 nor ≤ 9.876, it is not in the rejection region. H$_o$ is not rejected; it is still plausible that $m = 10$.

g. $\bar{x} \geq 10.1032$ or ≤ 9.8968 iff $z \geq 2.58$ or ≤ -2.58.

12.

a. Let $m =$ true average braking distance for the new design at 40 mph. The hypotheses are $H_o : m = 120$ vs $H_a : m < 120$.

b. R_2 should be used, since support for H_a is provided only by an $\overline{x}$ value substantially smaller than 120. ($E(\overline{x}) = 120$ when H_o is true and , 120 when H_a is true).

c. $s_{\overline{x}} = \frac{s}{\sqrt{n}} = \frac{10}{6} = 1.6667$, so $a = P(\overline{x} \geq 115.20$ when $m = 120) =$ $P\left(z \leq \frac{115.20 - 120}{1.6667}\right) = P(z \leq -2.88) = .002$. To obtain $a = .001$, replace 115.20 by $c = 120 - 3.08(1.6667) = 114.87$, so that $P(\overline{x} \leq 114.87$ when $m = 120) =$ $P(z \leq -3.08) = .001$.

d. $b(115) = P(\overline{x} > 115.2$ when $m = 115) = P(z > .12) = .4522$

e. $a = P(z \leq -2.33) = .01$, because when H_o is true Z has a standard normal distribution ($\overline{X}$ has been standardized using 120). Similarly $P(z \leq -2.88) = .002$, so this second rejection region is equivalent to R_2.

13.

a. $$P(\overline{x} \geq m_o + 2.33\frac{s}{\sqrt{n}}\ when m = m_o) = P\left(z \geq \frac{\left(m_o + 2.33\frac{s}{\sqrt{n}}\right)}{\frac{s}{\sqrt{n}}}\right)$$
$= P(z \geq 2.33) = .01$, where Z is a standard normal r.v.

b. P(rejecting H_o when $m = 99) = P(\overline{x} \geq 102.33$ when $m = 99)$ $= P\left(z \geq \frac{102 - 99}{1}\right) = P(z \geq 3.33) = .0004$. Similarly, $a(98) = P(\overline{x} \geq 102.33$ when $m = 98) = P(z \geq 4.33) = 0$. In general, we have P(type I error) < .01 when this probability is calculated for a value of m less than 100. The boundary value $m = 100$ yields the largest a.

14.

a. $s_{\bar{x}} = .04$, so $P(\bar{x} \geq 10.1004 or \leq 9.8940$ when $m = 10) = P(z \geq 2.51 or \leq -2.65) = .006 + .004 = .01$

b. $b(10.1) = P(9.8940 < \bar{x} < 10.1004$ when $m = 10.1) = P(-5.15 < z < .01) = .5040$, whereas $b(9.9) = P(-.15 < z < 5.01) = .5596.$ Since $m = 9.9$ and $m = 10.1$ represent equally serious departures from H_o, one would probably want to use a test procedure for which $b(9.9) = b(10.1)$. A similar result and comment apply to any other pair of alternative values symmetrically placed about 10.

Section 8.2

15.

a. $a = P(z \geq 1.88$ when z has a standard normal distribution) $= 1 - \Phi(1.88) = .0301$

b. $a = P(z \leq -2.75$ when z ~ N(0, 1) $= \Phi(-2.75) = .003$

c. $a = \Phi(-2.88) + (1 - \Phi(2.88)) = .004$

16.

a. $a = P(t \geq 3.733$ when t has a t distribution with 15 d.f.) =.001, because the 15 d.f. row of Table A.5 shows that $t_{.001,15} = .3733$

b. d.f. = n – 1 = 23, so $a = P(t \leq -2.500) = .01$

c. d.f. = 30, and $a = P(t \geq 1.697) + P(t \leq -1.697) = .05 + .05 = .10$

17.

a. $z = \dfrac{20{,}960 - 20{,}000}{1500\big/\sqrt{16}} = 2.56 > 2.33$ so reject H_o.

b. $b\left(20{,}500\right) : \Phi\left(2.33 + \dfrac{20{,}000 - 20{,}500}{1500/\sqrt{16}}\right) = \Phi(1.00) = .8413$

c. $b\left(20{,}500\right) = .05 : n = \left[\dfrac{1500(2.33 + 1.645)}{20{,}000 - 20{,}500}\right]^2 = 142.2$, so use n = 143

d. $a = 1 - \Phi(2.56) = .0052$

18.

a. $\dfrac{72.3 - 75}{1.8} = -1.5$ so 72.3 is 1.5 SD's (of $\bar{x}$) below 75.

b. H_o is rejected if $z \le -2.33$; since $z = -1.5$ is not ≤ -2.33, don't reject H_o.

c. $a =$ area under standard normal curve below –2.88 $= \Phi(-2.88) = .0020$

d. $\Phi\left(-2.88 + \dfrac{75 - 70}{9/5}\right) = \Phi(-.1) = .4602$ so $b(70) = .5398$

e. $n = \left[\dfrac{9(2.88 + 2.33)}{75 - 70}\right]^2 = 87.95$, so use n = 88

f. $a(76) = P(Z < -2.33$ when $m = 76) = P(\overline{X} < 72.9$ when $m = 76)$

$= \Phi\left(\dfrac{72.9 - 76}{.9}\right) = \Phi(-3.44) = .0003$

19.

a. Reject H_o if either $z \geq 2.58$ or $z \leq -2.58$; $\frac{s}{\sqrt{n}} = 0.3$, so

$z = \frac{94.32 - 95}{0.3} = -2.27$. Since –2.27 is not < -2.58, don't reject H_o.

b. $b(94) = \Phi\left(2.58 - \frac{1}{0.3}\right) - \Phi\left(-2.58 - \frac{1}{0.3}\right) = \Phi(-.75) - \Phi(-5.91) = .2266$

c. $n = \left[\frac{1.20(2.58 + 1.28)}{95 - 94}\right]^2 = 21.46$, so use n = 22.

20. With H_o: $m = 750$, and H_a: $m < 750$ and a significance level of .05, we reject H_o if z < -1.645; z = -2.14 < -1.645, so we reject the null hypothesis and do not continue with the purchase. At a significance level of .01, we reject H_o if z < -2.33; z = -2.14 > -2.33, so we don't reject the null hypothesis and thus continue with the purchase.

21. With H_o: $m = .5$, and H_a: $m \neq .5$ we reject H_o if $t > t_{a/2,n-1}$ or $t < -t_{a/2,n-1}$

a. 1.6 < $t_{.025,12}$ = 2.179, so don't reject H_o

b. -1.6 > $-t_{.025,12}$ = -2.179, so don't reject H_o

c. – 2.6 > $-t_{.005,24}$ = -2.797, so don't reject H_o

d. –3.9 < the negative of all t values in the df = 24 row, so we reject H_o in favor of H_a.

22.

a. It appears that the true average weight could be more than the production specification of 200 lb per pipe.

b. H_o: $m = 200$, and H_a: $m > 200$ we reject H_o if $t > t_{.05,29} = 1.699$.

$t = \frac{206.73 - 200}{6.35/\sqrt{30}} = \frac{6.73}{1.16} = 5.80 > 1.699$, so reject H_o. The test appears to substantiate the statement in part **a**.

23. H_o: $m = 360$ vs. H_a: $m > 360$; $t = \frac{\bar{x} - 360}{s/\sqrt{n}}$; reject H_o if $t > t_{.05,25} = 1.708$;

$t = \frac{370.69 - 360}{24.36/\sqrt{26}} = 2.24 > 1.708$. Thus H_o should be rejected. There appears to be a contradiction of the prior belief.

24. H_o: $m = 3000$ vs. H_a: $m \neq 3000$; $t = \dfrac{\overline{x} - 3000}{s / \sqrt{n}}$; reject H_o if $|t| > t_{.025,4} = 2.776$;
$t = \dfrac{2887.6 - 3000}{84 / \sqrt{5}} = -2.99 < -2.776$, so we reject H_o . This requirement is not satisfied.

25.

a. H_o: $m = 5.5$ vs. H_a: $m \neq 5.5$; for a level .01 test, (not specified in the problem description), reject H_o if either $z \geq 2.58$ or $z \leq -2.58$. Since $z = \dfrac{5.25 - 5.5}{.075} = -3.33 \leq -2.58$, reject H_o.

b.
$$1 - b(5.6) = 1 - \Phi\left(2.58 + \frac{(-.1)}{.075}\right) + \Phi\left(-2.58 - \frac{(-.1)}{.075}\right)$$
$$= 1 - \Phi(1.25) + \Phi(-3.91) = .105$$

c. $n = \left[\dfrac{.3(2.58 + 2.33)}{-.1}\right]^2 = 216.97$, so use n = 217.

26. Reject H_o if $z \geq 1.645$; $\dfrac{s}{\sqrt{n}} = .7155$, so $z = \dfrac{52.7 - 50}{.7155} = 3.77$. Since 3.77 is ≥ 1.645 , reject H_o at level .05 and conclude that true average penetration exceeds 50 mils.

27. We wish to test H_o: $m = 75$ vs. H_a: $m < 75$; Using $a = .01$, H_o is rejected if $t \leq -t_{.01,41} \approx -2.423$ (from the df 40 row of the t-table). Since $t = \dfrac{73.1 - 75}{5.9 / \sqrt{42}} = -2.09$, which is not ≤ -2.423 , H_o is not rejected. The alloy is not suitable.

28. With $m =$ true average recumbency time, the hypotheses are H_o: $m = 20$ vs H_a: $m < 20$. The test statistic value is $z = \dfrac{\overline{x} - 20}{s / \sqrt{n}}$, and H_o should be rejected if $z \leq -z_{.10} = -1.28$
Since $z = \dfrac{18.86 - 20}{8.6 / \sqrt{73}} = -1.13$, which is not ≤ -1.28 , H_o is not rejected. The sample data does not strongly suggest that true average time is less than 20.

29.

a. For n = 8, n – 1 = 7, and $t_{.05,7} = 1.895$, so H_o is rejected at level .05 if $t \ge 1.895$.

Since $\frac{s}{\sqrt{n}} = \frac{1.25}{\sqrt{8}} = .442$, $t = \frac{3.72 - 3.50}{.442} = .498$; this does not exceed 1.895, so H_o is not rejected.

b. $d = \frac{|m_o - m|}{s} = \frac{|3.50 - 4.00|}{1.25} = .40$, and n = 8, so from table A.17, $b(4.0) \approx .72$

30. n = 115, $\bar{x} = 11.3$, $s = 6.43$

1. Parameter of Interest: $m =$ true average dietary intake of zinc among males aged 65 – 74 years.
2. Null Hypothesis: H_o: $m = 15$
3. Alternative Hypothesis: H_a: $m < 15$
4. $z = \frac{\bar{x} - m_o}{s/\sqrt{n}} = \frac{\bar{x} - 15}{s/\sqrt{n}}$
5. Rejection Region: No value of α was given, so select a reasonable level of significance, such as α= .05. $z \le z_a$ or $z \le -1.645$
6. $z = \frac{11.3 - m_o}{6.43/\sqrt{115}} = -6.17$
7. –6.17 < -1.645, so reject H_o. The data does support the claim that average daily intake of zinc for males aged 65 - 74 years falls below the recommended daily allowance of 15 mg/day.

31. The hypotheses of interest are H_o: $m = 7$ vs H_a: $m < 7$, so a lower-tailed test is appropriate; H_o should be rejected if $t \le -t_{.1,8} = -1.397$. $t = \frac{6.32 - 7}{1.65/\sqrt{9}} = -1.24$. Because -1.24 is not ≤ -1.397, H_o (prior belief) is not rejected (contradicted) at level .01.

32. n = 12, $\bar{x} = 98.375$, $s = 6.1095$

a.

1 Parameter of Interest: $m =$ true average reading of this type of radon detector when exposed to 100 pCi/L of radon.

2 Null Hypothesis: H_o: $m = 100$

3 Alternative Hypothesis: H_a: $m \neq 100$

4 $$t = \frac{\bar{x} - m_o}{s/\sqrt{n}} = \frac{\bar{x} - 100}{s/\sqrt{n}}$$

5 $t \le -2.201$ or $t \ge 2.201$

6 $$t = \frac{98.375 - 100}{6.1095/\sqrt{12}} = -.9213$$

7 Fail to reject H_o. The data does not indicate that these readings differ significantly from 100.

b. $\sigma = 7.5$, $\beta = 0.10$. From table A.17, df ≈ 29, thus n ≈30.

33. $$b(m_o - \Delta) = \Phi\left(z_{a/2} + \Delta\sqrt{n}/s\right) - \Phi\left(-z_{a/2} - \Delta\sqrt{n}/s\right)$$
$$= 1 - \left[\Phi\left(-z_{a/2} - \Delta\sqrt{n}/s\right) + \Phi\left(z_{a/2} - \Delta\sqrt{n}/s\right)\right] = b(m_o + \Delta)$$
(since 1 - Φ(c) = Φ(-c)).

34. For an upper-tailed test, $= b(m) = \Phi\left(z_a + \sqrt{n}(m_o - m)/s\right)$. Since in this case we are considering $m > m_o$, $m_o - m$ is negative so $\sqrt{n}(m_o - m)/s \to -\infty$ as n $\to \infty$. The desired conclusion follows since $\Phi(-\infty) = 0$. The arguments for a lower-tailed and tow-tailed test are similar.

Section 8.3

35.

1 Parameter of interest: p = true proportion of cars in this particular county passing emissions testing on the first try.

2 H_o: p = .70

3 H_a: p ≠ .70

4 $$z = \frac{\hat{p} - p_o}{\sqrt{p_o(1 - p_o)/n}} = \frac{\hat{p} - .70}{\sqrt{.70(.30)/n}}$$

5 either z ≥ 1.96 or z ≤ -1.96

6 $$z = \frac{124/200 - .70}{\sqrt{.70(.30)/200}} = -2.469$$

7 Reject H_o. The data indicates that the proportion of cars passing the first time on emission testing or this county differs from the proportion of cars passing statewide.

36.

a.

1 p = true proportion of all nickel plates that blister under the given circumstances.

2 H_o: p = .10

3 H_a: p > .10

4 $$z = \frac{\hat{p} - p_o}{\sqrt{p_o(1-p_o)/n}} = \frac{\hat{p} - .10}{\sqrt{.10(.90)/n}}$$

5 Reject H_o if $z \geq 1.645$

6 $$z = \frac{14/100 - .10}{\sqrt{.10(.90)/100}} = 1.33$$

7 Fail to Reject H_o. The data does not give compelling evidence for concluding that more than 10% of all plates blister under the circumstances.

The possible error we could have made is a Type II error: Failing to reject the null hypothesis when it is actually true.

b. $b(.15) = \Phi\left[\frac{.10 - .15 + 1.645\sqrt{.10(.90)/100}}{\sqrt{.15(.85)/100}}\right] = \Phi(-.02) = .4920$. When n = 200, $b(.15) = \Phi\left[\frac{.10 - .15 + 1.645\sqrt{.10(.90)/200}}{\sqrt{.15(.85)/200}}\right] = \Phi(-.60) = .2743$

c. $n = \left[\frac{1.645\sqrt{.10(.90)} + 1.28\sqrt{.15(.85)}}{.15 - .10}\right]^2 = 19.01^2 = 361.4$, so use n = 362

37.

1 p = true proportion of the population with type A blood

2 H_o: p = .40

3 H_a: $p \neq .40$

4 $$z = \frac{\hat{p} - p_o}{\sqrt{p_o(1-p_o)/n}} = \frac{\hat{p} - .40}{\sqrt{.40(.60)/n}}$$

5 Reject H_o if $z \geq 2.58$ or $z \leq -2.58$

6 $$z = \frac{82/150 - .40}{\sqrt{.40(.60)/150}} = \frac{.147}{.04} = 3.667$$

7 Reject H_o. The data does suggest that the percentage of the population with type A blood differs from 40%. (at the .01 significance level). Since the z critical value for a significance level of .05 is less than that of .01, the conclusion would not change.

38.

a. We wish to test H_o: p = .02 vs H_a: p < .02; only if H_o can be rejected will the inventory be postponed. The lower-tailed test rejects H_o if $z \le -1.645$. With $\hat{p} = \frac{15}{1000} = .015$, z = -1.01, which is not ≤ -1.645. Thus, H_o cannot be rejected, so the inventory should be carried out.

b. $$b(.01) = \Phi\left[\frac{.02 - .01 + 1.645\sqrt{.02(.98)/1000}}{\sqrt{.01(.99)/1000}}\right] = \Phi(5.49) \approx 1$$

c. $$b(.05) = \Phi\left[\frac{.02 - .05 + 1.645\sqrt{.02(.98)/1000}}{\sqrt{.05(.95)/1000}}\right] = \Phi(-3.30) = .0005$$, so is p = .05 it is highly unlikely that H_o will be rejected and the inventory will almost surely be carried out.

39. Let p denote the true proportion of those called to appear for service who are black. We wish to test H_o: p = .25 vs H_a: p < .25. We use $z = \frac{\hat{p} - .25}{\sqrt{.25(.75)/n}}$, with the rejection region $z \le -z_{.01} = -2.33$. We calculate $\hat{p} = \frac{177}{1050} = .1686$, and $z = \frac{.1686 - .25}{.0134} = -6.1$. Because $-6.1 < -2.33$, H_o is rejected. A conclusion that discrimination exists is very compelling.

40.

a. P = true proportion of current customers who qualify. H_o: p = .05 vs H_a: p ≠ .05, $z = \frac{\hat{p} - .05}{\sqrt{.05(.95)/n}}$, reject H_o if $z \ge 2.58$ or $z \le -2.58$. $\hat{p} = .08$, so $z = \frac{.03}{.00975} = 3.07 \ge 2.58$, so H_o is rejected. The company's premise is not correct.

b. $$b(.10) = \Phi\left[\frac{.05 - .10 + 2.58\sqrt{.05(.95)/500}}{\sqrt{.10(.90)/500}}\right] = \Phi(-1.85) = .0332$$

41.

a. The alternative of interest here is H_a: $p > .50$ (which states that more than 50% of all enthusiasts prefer gut), so the rejection region should consist of large values of X (an upper-tailed test). Thus (15, 16, 17, 18, 19, 20) is the appropriate region.

b. $a = P(15 \le X \text{ when } p = .5)$ = 1 – B(14; 20, .05) = .021, so this is a level .05 test. For R = {14, 15, …, 20}, $\alpha = .058$, so this R does not specify a level .05 test and the region of **a** is the best level .05 test. ($\alpha \le .05$ along with smallest possible β).

c. $\beta(.6)$ = B(14; 20, .6) = .874, and $\beta(.8)$ = B(14; 20, .8) = .196.

d. The best level .10 test is specified by R = (14, …, 20} (with $\alpha = .052$) Since 13 is not in R, H_o is not rejected at this level.

42. The hypotheses are H_o: p = .10 vs. H_a: p > .10, so R has the form {c, …, n}. For n = 10, c = 3 (i.e. R = {3, 4, …, 10}) yields α = 1 – B(2; 10, .1) = .07 while no larger R has $\alpha \le .10$; however $\beta(.3)$ = B(2; 10, .3) = .383. For n = 20, c = 5 yields α = 1 – B(4; 20, .1) = .043, but again $\beta(.3)$ = B(4; 20, .3) = .238. For n = 25, c = 5 yields α = 1 – B(4; 25, .1) = .098 while $\beta(.7)$ = B(4; 25, .3) = .090 $\le .10$, so n = 25 should be used.

43. H_o: p = .035 vs H_a: p < .035. We use $z = \dfrac{\hat{p} - .035}{\sqrt{.035(.965)/n}}$, with the rejection region $z \le -z_{.01} = -2.33$. With $\hat{p} = \dfrac{15}{500} = .03$, $z = \dfrac{-.005}{\sqrt{.0082}} = -.61$. Because -.61 isn't $\le$ -2.33, H_o is not rejected. Robots have not demonstrated their superiority.

Section 8.4

44. Using $\alpha = .05$, H_o should be rejected whenever p-value < .05.

a. P-value = .001 < .05, so reject H_o

b. .021 < .05, so reject H_o.

c. .078 is not < .05, so don't reject H_o.

d. .047 < .05, so reject H_o (a close call).

e. .148 > .05, so H_o can't be rejected at level .05.

45.

a. p-value = .084 > .05 = α, so don't reject H_o.

b. p-value = .003 < .001 = α, so reject H_o.

c. .498 >> .05, so H_o can't be rejected at level .05

d. 084 < .10, so reject H_o at level .10

e. .039 is not < .01, so don't reject H_o.

f. p-value = .218 > .10, so H_o cannot be rejected.

46. In each case the p-value = $1-\Phi(z)$

a. .0778

b. .1841

c. .0250

d. .0066

e. .4562

47.

a. .0358

b. .0802

c. .5824

d. .1586

e. 0

48.

a. In the df = 8 row of table A.5, t = 2.0 is between 1.860 and 2.306, so the p-value is between .025 and .05: .025 < p-value < .05.

b. 2.201 < | -2.4 | < 2.718, so .01 < p-value < .025.

c. 1.341 < | -1.6 | < 1.753, so .05 < P(t < -1.6) < .10. Thus a two-tailed p-value: 2(.05 < P(t < -1.6) < .10), or .10 < p-value < .20

d. With an upper-tailed test and t = -.4, the p-value = P(t > -.4) > .50.

e. 4.032 < t=5 < 5.893, so .001 < p-value < .005

f. 3.551 < | -4.8 |, so P(t < -4.8) < .0005. A two-tailed p-value = 2[P(t < -4.8)] < 2(.0005), or p-value < .001.

49. An upper-tailed test

a. Df = 14, α=.05; $t_{.05,14} = 1.761$: 3.2 > 1.761, so reject H_o.

b. $t_{.01,18} = 2.896$; 1.8 is not > 2.896, so don't reject H_o.

c. Df = 23, p-value > .50, so fail to reject H_o at any significance level.

50. The p-value is greater than the level of significance α = .01, therefore fail to reject H_o that $m = 5.63$. The data does not indicate a difference in average serum receptor concentration between pregnant women and all other women.

51. Here we might be concerned with departures above as well as below the specified weight of 5.0, so the relevant hypotheses are H_o: $m = 5.0$ vs H_a: $m \neq 5.0$. At level .01, reject H_o if either $z \geq 2.58$ or $z \leq -2.58$. Since $\frac{s}{\sqrt{n}} = .035$, $z = \frac{-.13}{.035} = -3.71$, which is ≤ -2.58, so H_o should be rejected. Because 3.71 is "off" the z-table, p-value < 2(.0002) = .0004 (.0002 corresponds to z = -3.49).

52.

a. For testing H_o: p = .2 vs H_a: p > .2, an upper-tailed test is appropriate. The computed Z is z = .97, so p-value = $1 - \Phi(.97) = .166$. Because the p-value is rather large, H_o would not be rejected at any reasonable α (it can't be rejected for any α < .166), so no modification appears necessary.

b. With p = .5, $1 - b(.5) = 1 - \Phi[(-.3 + 2.33(.0516))/.0645] = 1 - \Phi(-2.79) = .9974$

53. p = proportion of all physicians that know the generic name for methadone.
H_o: p = .50 vs H_a: p < .50; We can use a large sample test if both $np_0 \geq 10$ and $n(1 - p_0) \geq 10$; 102(.50) = .51, so we can proceed. $\hat{p} = \frac{47}{102}$, so

$$z = \frac{\frac{47}{102} - .50}{\sqrt{\frac{(.50)(.50)}{102}}} = \frac{-.039}{.050} = -.79$$

We will reject H_0 if the p-value < .01. For this lower tailed test, the p-value = Φ(z) = Φ(-.79) =.2148, which is not < .01, so we do not reject H_0 at significance level .01.

54. $m =$ the true average percentage of organic matter in this type of soil, and the hypotheses are H_o: $m = 3$ vs H_a: $m \neq 3$. With n = 30, and assuming normality, we use the t test: $t = \frac{\bar{x} - 3}{s / \sqrt{n}} = \frac{2.481 - 3}{.295} = \frac{-.519}{.295} = -1.759$. The p-value = 2[P(t > 1.759)] = 2(.041) = .082. At significance level .10, since .082 = .10, we would reject H_0 and conclude that the true average percentage of organic matter in this type of soil is something other than 3. At significance level .05, we would not have rejected H_0.

55. The hypotheses to be tested are H_o: $m = 25$ vs H_a: $m > 25$, and H_o should be rejected if $t \geq t_{.05,12} = 1.782$. The computed summary statistics are $\bar{x} = 27.923$, $s = 5.619$, so $\frac{s}{\sqrt{n}} = 1.559$ and $t = \frac{2.923}{1.559} = 1.88$. From table A.8, P(t > 1.88) ˜ .041, which is less than .05, so H_o is rejected at level .05.

56.

a. The appropriate hypotheses are H_o: $m = 10$ vs H_a: $m < 10$

b. P-value = P(t > 2.3) = .017, which is = .05, so we would reject H_o. The data indicates that the pens do not meet the design specifications.

c. P-value = P(t > 1.8) = .045, which is not = .01, so we would not reject H_o. There is not enough evidence to say that the pens don't satisfy the design specifications.

d. P-value = P(t > 3.6) ˜ .001, which gives strong evidence to support the alternative hypothesis.

57. $m =$ true average reading, H_o: $m = 70$ vs H_a: $m \neq 70$, and $t = \frac{\bar{x} - 70}{s / \sqrt{n}} = \frac{75.5 - 70}{7 / \sqrt{6}} = \frac{5.5}{2.86} = 1.92$. From table A.8, df = 5, p-value = 2[P(t> 1.92)] ˜ 2(.058) = .116. At significance level .05, there is not enough evidence to conclude that the spectrophotometer needs recalibrating.

58. With H_o: $m = .60$ vs H_a: $m \neq .60$, and a two-tailed p-value of .0711, we fail to reject H_o at levels .01 and .05 (thus concluding that the amount of impurities need not be adjusted) , but we would reject H_o at level .10 (and conclude that the amount of impurities does need adjusting).

Section 8.5

59.

a. The formula for b is $1-\Phi\left(-2.33+\frac{\sqrt{n}}{9.4}\right)$, which gives .8980 for n = 100, .1049 for n = 900, and .0014 for n = 2500.

b. Z = -5.3, which is "off the z table," so p-value < .0002; this value of z is quite statistically significant.

c. No. Even when the departure from H_o is insignificant from a practical point of view, a statistically significant result is highly likely to appear; the test is too likely to detect small departures from H_o.

60.

a. Here $b = \Phi\left(\frac{-.01+.9320/\sqrt{n}}{.4073/\sqrt{n}}\right) = \Phi\left(\frac{\left(-.01\sqrt{n}+.9320\right)}{.4073}\right)$ = .9793, .8554, .4325, .0944, and 0 for n = 100, 2500, 10,000, 40,000, and 90,000, respectively.

b. Here $z = .025\sqrt{n}$ which equals .25, 1.25, 2.5, and 5 for the four n's, whence p-value = .4213, .1056, .0062, .0000, respectively.

c. No; the reasoning is the same as in 54 (c).

Supplementary Exercises

61. Because n = 50 is large, we use a z test here, rejecting H_o: $m = 3.2$ in favor of H_a: $m \neq 3.2$ if either $z \geq z_{.025} = 1.96$ or $z \leq -1.96$. The computed z value is $z = \frac{3.05-3.20}{.34/\sqrt{50}} = -3.12$. Since –3.12 is ≤ -1.96, H_o should be rejected in favor of H_a.

62. Here we assume that thickness is normally distributed, so that for any *n* a t test is appropriate, and use Table A.17 to determine n. We wish $p(3) = .95$ when $d = \frac{|3.2-3|}{.3} = .667$. By inspection, n = 20 satisfies this requirement, so n = 50 is too large.

63.

a. H_o: $m = 3.2$ vs H_a: $m \neq 3.2$ (Because H_a: $m > 3.2$ gives a p-value of roughly .15)

b. With a p-value of .30, we would reject the null hypothesis at any reasonable significance level, which includes both .05 and .10.

64.

a. H_o: $m = 2150$ vs H_a: $m > 2150$

b. $t = \dfrac{\bar{x} - 2150}{s/\sqrt{n}}$

c. $t = \dfrac{2160 - 2150}{30/\sqrt{16}} = \dfrac{10}{7.5} = 1.33$

d. Since $t_{.10,15} = 1.341$, p-value > .10 (actually $\approx .10$)

e. From **d**, p-value > .05, so H_o cannot be rejected at this significance level.

65.

a. The relevant hypotheses are H_o: $m = 548$ vs H_a: $m \neq 548$. At level .05, H_o will be rejected if either $t \geq t_{.025,10} = 2.228$ or $t \leq -t_{.025,10} = -2.228$. The test statistic value is $t = \dfrac{587 - 548}{10/\sqrt{11}} = \dfrac{39}{3.02} = 12.9$. This clearly falls into the upper tail of the two-tailed rejection region, so H_o should be rejected at level .05, or any other reasonable level).

b. The population sampled was normal or approximately normal.

66. $n = 8, \bar{x} = 30.7875, s = 6.5300$

1 Parameter of interest: $m =$ true average heat-flux of plots covered with coal dust

2 H_o: $m = 29.0$

3 H_a: $m > 29.0$

4 $t = \dfrac{\bar{x} - 29.0}{s/\sqrt{n}}$

5 RR: $t \geq t_{a,n-1}$ or $t \geq 1.895$

6 $t = \dfrac{30.7875 - 29.0}{6.53/\sqrt{8}} = .7742$

7 Fail to reject H_o. The data does not indicate the mean heat-flux for pots covered with coal dust is greater than for plots covered with grass.

67. N = 47, $\bar{x} = 215$ mg, s = 235 mg. Range 5 mg to 1,176 mg.

a. No, the distribution does not appear to be normal, it appears to be skewed to the right. It is not necessary to assume normality if the sample size is large enough due to the central limit theorem. This sample size is large enough so we can conduct a hypothesis test about the mean.

b.

1 Parameter of interest: $m =$ true daily caffeine consumption of adult women.

2 H_o: $m = 200$

3 H_a: $m > 200$

4 $z = \dfrac{\bar{x} - 200}{s/\sqrt{n}}$

5 RR: $z \geq 1.282$ or if p-value $\leq .10$

6 $z = \dfrac{215 - 200}{235/\sqrt{47}} = .44$; p-value $= 1 - \Phi(.44) = .33$

7 Fail to reject H_o. because .33 > .10. The data does not indicate that daily consump tion of all adult women exceeds 200 mg.

68. At the .05 significance level, reject H_o because .043 < .05. At the level .01, fail to reject H_o because .043 > .01. Thus the data contradicts the design specification that sprinkler activation is less than 25 seconds at the level .05, but not at the .01 level.

69.

a. From table A.17, when $m = 9.5$, d = .625, df = 9, and $b \approx .60$, when $m = 9.0$, d = 1.25, df = 9, and $b \approx .20$.

b. From Table A.17, $b = .25$, d = .625, n ≈ 28

70. A normality plot reveals that these observations could have come from a normally distributed population, therefore a t-test is appropriate. The relevant hypotheses are H_o: $m = 9.75$ vs H_a: $m > 9.75$. Summary statistics are n = 20, $\bar{x} = 9.8525$, and s = .0965, which leads to a test statistic $t = \dfrac{9.8525 - 9.75}{.0965/\sqrt{20}} = 4.75$, from which the p-value = .0001. (From MINITAB output). With such a small p-value, the data strongly supports the alternative hypothesis. The condition is not met.

71.

a. With H_o: p = $\frac{1}{75}$ vs H_o: p ≠ $\frac{1}{75}$, we reject H_o if either $z \geq 1.96$ or $z \leq -1.96$.

With $\hat{p} = \frac{16}{800} = .02$, $z \frac{.02 - .01333}{\sqrt{\frac{.01333(.98667)}{800}}} = 1.645$, which is not in either rejection region. Thus, we fail to reject the null hypothesis. There is not evidence that the incidence rate among prisoners differs from that of the adult population. The possible error we could have made is a type II.

b. P-value = $2[1 - \Phi(1.645)] = 2[.05] = .10$. Yes, since .10 < .20, we could reject H_o.

72. A t test is appropriate; H_o: $m = 1.75$ is rejected in favor of H_a: $m \neq 1.75$ if the p-value >.05. The computed t is $t = \frac{1.89 - 1.75}{.42 / \sqrt{26}} = 1.70$. Since $1.70 \approx 1.708 = t_{.025,25}$, $P \approx 2(.05) = .10$ (since for a two-tailed test, $.05 = a / 2$.), do not reject H_o; the data does not contradict prior research. We assume that the population from which the sample was taken was approximately normally distributed.

73. Even though the underlying distribution may not be normal, a z test can be used because n is large. H_o: $m = 3200$ should be rejected in favor of H_a: $m < 3200$ if $z \leq -z_{.001} = -3.08$. The computed z is $z = \frac{3107 - 3200}{188 / \sqrt{45}} = -3.32 \leq -3.08$, so H_o should be rejected at level .001.

74. Let p = the true proportion of mechanics who could identify the problem. Then the appropriate hypotheses are H_o: p = .75 vs H_a: p < .75, so a lower-tailed test should be used. With p_o= .75 and $\hat{p} = \frac{42}{72} = .583$, z = -3.28 and $P = \Phi(-3.28) = .0005$. Because this p-value is so small, the data argues strongly against H_o, so we reject it in favor of H_a.

75. We wish to test H_o: $l = 4$ vs H_a: $l > 4$ using the test statistic $z = \frac{\bar{x} - 4}{\sqrt{4/n}}$. For the given sample, n = 36 and $\bar{x} = \frac{160}{36} = 4.444$, so $z = \frac{4.444 - 4}{\sqrt{4/36}} = 1.33$. At level .02, we reject H_o if $z \geq z_{.02} \approx 2.05$ (since $1 - \Phi(2.05) = .0202$). Because 1.33 is not ≥ 2.05, H_o should not be rejected at this level.

76. H_o: $m = 15$ vs H_a: $m > 15$. Because the sample size is less than 40, and we can assume the distribution is approximately normal, the appropriate statistic is

$$t = \frac{\bar{x} - 15}{s/\sqrt{n}} = \frac{17.5 - 15}{2.2/\sqrt{32}} = \frac{2.5}{.390} = 6.4$$

. Thus the p-value is "off the chart" in the 20 df column of Table A.8, and so is approximately $0 < .05$, so H_o is rejected in favor of the conclusion that the true average time exceeds 15 minutes.

77. H_o: $s^2 = .25$ vs H_a: $s^2 > .25$. The chi-squared critical value for 9 d.f. that captures upper-tail area .01 is 21.665. The test statistic value is $\frac{9(.58)^2}{.25} = 12.11$. Because 12.11 is not ≥ 21.665, H_o cannot be rejected. The uniformity specification is not contradicted.

78. The 20 df row of Table A.7 shows that $c^2_{.99,20} = 8.26 < 8.58$ (H_o not rejected at level .01) and $8.58 < 9.591 = c^2_{.975,20}$ (H_o rejected at level .025). Thus .01 < p-value < .025 and H_o cannot be rejected at level .01 (the p-value is the smallest alpha at which rejection can take place, and this exceeds .01).

79.

a. $E(\bar{X} + 2.33S) = E(\bar{X}) + 2.33E(S) = m + 2.33s$, so $\hat{q} = \bar{X} + 2.33S$ is approximately unbiased.

b. $V(\bar{X} + 2.33S) = V(\bar{X}) + 2.33^2 V(S) = \frac{s^2}{n} + 5.4289\frac{s^2}{2n}$. The estimated standard error (standard deviation) is $1.927\frac{s}{\sqrt{n}}$.

c. More than 99% of all soil samples have pH less than 6.75 iff the 95th percentile is less than 6.75. Thus we wish to test H_o: $m + 2.33s = 6.75$ vs H_a: $m + 2.33s < 6.75$. H_o will be rejected at level .01 if $z \leq 2.33$. Since $z = \frac{-.047}{.0385} < 0$, H_o clearly cannot be rejected. The 95th percentile does not appear to exceed 6.75.

80.

a. When H_o is true, $2l_o \Sigma X_i = 2\sum \frac{X_i}{m_o}$ has a chi-squared distribution with df = 2n. If the alternative is H_a: $m > m_o$, large test statistic values (large Σx_i, since $\overline{x}$ is large) suggest that H_o be rejected in favor of H_a, so rejecting when $2\sum \frac{X_i}{m_o} \geq c^2_{a,2n}$ gives a test with significance level a. If the alternative is H_a: $m < m_o$, rejecting when $2\sum \frac{X_i}{m_o} \leq c^2_{1-a,2n}$ gives a level a test. The rejection region for H_a: $m \neq m_o$ is either $2\sum \frac{X_i}{m_o} \geq c^2_{a/2,2n}$ or $\leq c^2_{1-a/2,2n}$.

b. H_o: $m = 75$ vs H_a: $m < 75$. The test statistic value is $\frac{2(737)}{75} = 19.65$. At level .01, H_o is rejected if $2\sum \frac{X_i}{m_o} \leq c^2_{.99,20} = 8.260$. Clearly 19.65 is not in the rejection region, so H_o should not be rejected. The sample data does not suggest that true average lifetime is less than the previously claimed value.

81.

a. P(type I error) = P(either $Z \geq z_g$ or $Z \leq z_{a-g}$) (when Z is a standard normal r.v.) = $\Phi(-z_{a-g}) + 1 - \Phi(z_g) = a - g + g = a$.

b. $b(m) = P(\overline{X} \geq m_o + \frac{s z_g}{\sqrt{n}} \text{ or } \overline{X} \leq m_o - \frac{s z_{a-g}}{\sqrt{n}}$ when the true value is μ) =

$$\Phi\left(z_g + \frac{m_o - m}{s/\sqrt{n}}\right) - \Phi\left(-z_{a-g} + \frac{m_o - m}{s/\sqrt{n}}\right)$$

c. Let $l = \sqrt{n}\frac{\Delta}{s}$; then we wish to know when $p(m_o + \Delta) = 1 - \Phi(z_g - l) + \Phi(-z_{a-g} - l) > 1 - \Phi(z_g + l) + \Phi(-z_{a-g} + l) = p(m_o - \Delta)$. Using the fact that $\Phi(-c) = 1 - \Phi(c)$, this inequality becomes $\Phi(z_g + l) - \Phi(z_g - l) > \Phi(z_{a-g} + l) - \Phi(z_{a-g} - l)$. The l.h.s. is the area under the Z curve above the interval $(z_g + l, z_g - l)$, while the r.h.s. is the area above $(z_{a-g} - l, z_{a-g} + l)$. Both intervals have width $2l$, but when $z_g < z_{a-g}$, the first interval is closer to 0 (and thus corresponds to the large area) than is the second. This happens when $g > a - g$, i.e., when $g > a/2$.

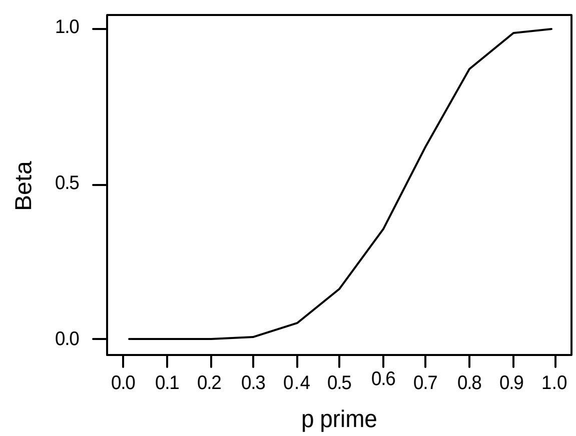
82.

a. $\alpha = P(X \le 5$ when p = .9) = B(5; 10, .9) = .002, so the region (0, 1, …, 5) does specify a level .01 test.

b. The first value to be placed in the upper-tailed part of a two tailed region would be 10, but P(X = 10 when p = .9) = .349, so whenever 10 is in the rejection region, $\alpha \ge .349$.

c. Using the two-tailed formula for ß(p') on p. 341, we calculate the value for the range of possible p' values. The values of p' we chose, as well as the associated ß(p') are in the table below, and the sketch follows. ß(p') seems to be quite large for a great range of p' values.

P'	Beta
0.01	0.0000
0.10	0.0000
0.20	0.0000
0.30	0.0071
0.40	0.0505
0.50	0.1635
0.60	0.3594
0.70	0.6206
0.80	0.8696
0.90	0.9900
0.99	1.0000



CHAPTER 9

Section 9.1

1.

a. $E(\overline{X}-\overline{Y}) = E(\overline{X}) - E(\overline{Y}) = 4.1 - 4.5 = -.4$, irrespective of sample sizes.

b. $V(\overline{X}-\overline{Y}) = V(\overline{X}) + V(\overline{Y}) = \frac{s_1^{\ 2}}{m} + \frac{s_2^{\ 2}}{n} = \frac{(1.8)^2}{100} + \frac{(2.0)^2}{100} = .0724$, and the s.d. of $\overline{X}-\overline{Y} = \sqrt{.0724} = .2691$.

c. A normal curve with mean and s.d. as given in **a** and **b** (because m = n = 100, the CLT implies that both $\overline{X}$ and $\overline{Y}$ have approximately normal distributions, so $\overline{X}-\overline{Y}$ does also). The shape is not necessarily that of a normal curve when m = n = 10, because the CLT cannot be invoked. So if the two lifetime population distributions are not normal, the distribution of $\overline{X}-\overline{Y}$ will typically be quite complicated.

2. The test statistic value is $z = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}}}$, and H_o will be rejected if either $z \ge 1.96$ or $z \le -1.96$. We compute $z = \frac{42{,}500 - 40{,}400}{\sqrt{\frac{2200^2}{45}+\frac{1900^2}{45}}} = \frac{2100}{433.33} = 4.85$. Since 4.85 > 1.96, reject H_o and conclude that the two brands differ with respect to true average tread lives.

3. The test statistic value is $z = \frac{(\bar{x}-\bar{y}) - 5000}{\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}}}$, and H_o will be rejected at level .01 if $z \ge 2.33$. We compute $z = \frac{(43{,}500 - 36{,}800) - 5000}{\sqrt{\frac{2200^2}{45}+\frac{1500^2}{45}}} = \frac{700}{396.93} = 1.76$, which is not > 2.33, so we don't reject H_o and conclude that the true average life for radials does not exceed that for economy brand by more than 500.

4.

a. From Exercise 2, the C.I. is

$$(\bar{x}-\bar{y})\pm(1.96)\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}} = 2100 \pm 1.96(433.33) = 2100 \pm 849.33$$

$=(1250.67, 2949.33)$. In the context of this problem situation, the interval is moderately wide (a consequence of the standard deviations being large), so the information about m_1 and m_2 is not as precise as might be desirable.

b. From Exercise 3, the upper bound is
$5700 + 1.645(396.93) = 5700 + 652.95 = 6352.95$.

5.

a. H_a says that the average calorie output for sufferers is more than 1 cal/cm^2/min below that for nonsufferers. $\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}} = \sqrt{\frac{(.04)^2}{10}+\frac{(.16)^2}{10}} = .1414$, so

$z = \dfrac{(.64-2.05)-(-1)}{.1414} = -2.90$. At level .01, H_o is rejected if $z \le -2.33$; since –2.90 < -2.33, reject H_o.

b. $P = \Phi(-2.90) = .0019$

c. $b = 1 - \Phi\left(-2.33 - \dfrac{-1.2+1}{.1414}\right) = 1 - \Phi(-.92) = .8212$

d. $m = n = \dfrac{.2(2.33+1.28)^2}{(-.2)^2} = 65.15$, so use 66.

6.

a. H_o should be rejected if $z \ge 2.33$. Since $z = \dfrac{(18.12 - 16.87)}{\sqrt{\dfrac{2.56}{40} + \dfrac{1.96}{32}}} = 3.53 \ge 2.33$, H_o should be rejected at level .01.

b. $b(1) = \Phi\left(2.33 - \dfrac{1-0}{.3539}\right) = \Phi(-.50) = .3085$

c. $\dfrac{2.56}{40} + \dfrac{1.96}{n} = \dfrac{1}{(1.645 + 1.28)^2} = .1169 \Rightarrow \dfrac{1.96}{n} = .0529 \Rightarrow n = 37.06$, so use n = 38.

d. Since n = 32 is not a large sample, it would no longer be appropriate to use the large sample test. A small sample t procedure should be used (section 9.2), and the appropriate conclusion would follow.

7.

1 Parameter of interest: $m_1 - m_2 =$ the true difference of means for males and females on the Boredom Proneness Rating. Let $m_1 =$ men's average and $m_2 =$ women's average.

2 H_o: $m_1 - m_2 = 0$

3 H_a: $m_1 - m_2 > 0$

4 $z = \dfrac{(\bar{x} - \bar{y}) - \Delta_o}{\sqrt{\dfrac{s_1^2}{m} + \dfrac{s_2^2}{n}}} = \dfrac{(\bar{x} - \bar{y}) - 0}{\sqrt{\dfrac{s_1^2}{m} + \dfrac{s_2^2}{n}}}$

5 RR: $z \ge 1.645$

6 $z = \dfrac{(10.40 - 9.26) - \Delta_o}{\sqrt{\dfrac{4.83^2}{97} + \dfrac{4.68^2}{148}}} = 1.83$

7 Reject H_o. The data indicates the Boredom Proneness Rating is higher for males than for females.

8.

a.

1 Parameter of interest: $m_1 - m_2 =$ the true difference of mean tensile strength of the 1064 grade and the 1078 grade wire rod. Let $m_1 =$ 1064 grade average and $m_2 =$ 1078 grade average.

2 H_o: $m_1 - m_2 = -10$

3 H_a: $m_1 - m_2 < -10$

4 $$z = \frac{(\bar{x} - \bar{y}) - \Delta_o}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} = \frac{(\bar{x} - \bar{y}) - (-10)}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$$

5 RR: $p - value < a$

6 $$z = \frac{(107.6 - 123.6) - (-10)}{\sqrt{\frac{1.3^2}{129} + \frac{2.0^2}{129}}} = \frac{-6}{.210} = -28.57$$

7 For a lower-tailed test, the p-value = $\Phi(-28.57) \approx 0$, which is less than any a, so reject H_o. There is very compelling evidence that the mean tensile strength of the 1078 grade exceeds that of the 1064 grade by more than 10.

b. The requested information can be provided by a 95% confidence interval for $m_1 - m_2$:

$$(\bar{x} - \bar{y}) \pm 1.96\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}} = (-6) \pm 1.96(.210) = (-6.412, -5.588)$$.

9.

a. point estimate $\bar{x} - \bar{y} = 19.9 - 13.7 = 6.2$. It appears that there could be a difference.

b.

H_o: $m_1 - m_2 = 0$, H_a: $m_1 - m_2 \neq 0$, $z = \frac{(19.9 - 13.7)}{\sqrt{\frac{39.1^2}{60} + \frac{15.8^2}{60}}} = \frac{6.2}{5.44} = 1.14$, and

the p-value = 2[P(z > 1.14)] = 2(.1271) = .2542. The p value is larger than any reasonable α, so we do not reject H_0. There is no significant difference.

c. No. With a normal distribution, we would expect most of the data to be within 2 standard deviations of the mean, and the distribution should be symmetric. 2 sd's above the mean is 98.1, but the distribution stops at zero on the left. The distribution is positively skewed.

d. We will calculate a 95% confidence interval for μ, the true average length of stays for patients given the treatment. $19.9 \pm 1.96\frac{39.1}{\sqrt{60}} = 19.9 \pm 9.9 = (10.0, 21.8)$

10.

a. The hypotheses are H_o: $\mu_1 - \mu_2 = 5$ and H_a: $\mu_1 - \mu_2 > 5$. At level .001, H_o should be rejected if $z \ge 3.08$. Since $z = \frac{(65.6 - 59.8) - 5}{.2272} = 2.89 < 3.08$, H_o cannot be rejected in favor of H_a at this level, so the use of the high purity steel cannot be justified.

b. $\mu_1 - \mu_2 - \Delta_o = 1$, so $\beta = \Phi\left(3.08 - \frac{1}{.2272}\right) = \Phi(-.53) = .2891$

11. $\left(\overline{X} - \overline{Y}\right) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$. Standard error $= \frac{s}{\sqrt{n}}$. Substitution yields $\left(\overline{x} - \overline{y}\right) \pm z_{\alpha/2}\sqrt{(SE_1)^2 + (SE_2)^2}$. Using $\alpha = .05$, $z_{\alpha/2} = 1.96$, so $(5.5 - 3.8) \pm 1.96\sqrt{(0.3)^2 + (0.2)^2} = (0.99, 2.41)$. Because we selected $\alpha = .05$, we can state that when using this method with repeated sampling, the interval calculated will bracket the true difference 95% of the time. The interval is fairly narrow, indicating precision of the estimate.

12. The C.I. is $\left(\overline{x} - \overline{y}\right) \pm 2.58\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}} = (-8.77) \pm 2.58\sqrt{.9104} = -8.77 \pm 2.46$ $= (-11.23, -6.31)$. With 99% confidence we may say that the true difference between the average 7-day and 28-day strengths is between -11.23 and -6.31 N/mm^2.

13. $\sigma_1 = \sigma_2 = .05$, d = .04, $\alpha = .01$, $\beta = .05$, and the test is one-tailed, so

$$n = \frac{(.0025 + .0025)(2.33 + 1.645)^2}{.0016} = 49.38$$, so use n = 50.

14. The appropriate hypotheses are H_o: $\theta = 0$ vs. H_a: $\theta < 0$, where $\theta = 2\mu_1 - \mu_2$. ($\theta < 0$ is equivalent to $2\mu_1 < \mu_2$, so normal is more than twice schizo) The estimator of θ is $\hat{\theta} = 2\overline{X} - \overline{Y}$, with $Var(\hat{\theta}) = 4Var(\overline{X}) + Var(\overline{Y}) = \frac{4\sigma_1^2}{m} + \frac{\sigma_2^2}{n}$, σ_{θ} is the square root of $Var(\hat{\theta})$, and $\hat{\sigma}_{\theta}$ is obtained by replacing each σ_i^2 with S_i^2. The test statistic is then $\frac{\hat{\theta}}{\sigma_{\hat{\theta}}}$ (since $\theta_o = 0$), and H_o is rejected if $z \le -2.33$. With $\hat{\theta} = 2(2.69) - 6.35 = -.97$ and $\hat{\sigma}_{\theta} = \sqrt{\frac{4(2.3)^2}{43} + \frac{(4.03)^2}{45}} = .9236$, $z = \frac{-.97}{.9236} = -1.05$; Because –1.05 > -2.33, H_o is not rejected.

15.

a. As either *m* or *n* increases, s decreases, so $\dfrac{m_1 - m_2 - \Delta_o}{s}$ increases (the numerator is positive), so $\left(z_a - \dfrac{m_1 - m_2 - \Delta_o}{s}\right)$ decreases, so $b = \Phi\left(z_a - \dfrac{m_1 - m_2 - \Delta_o}{s}\right)$ decreases.

b. As b decreases, z_b increases, and since z_b is the numerator of n, n increases also.

16. $z = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{s_1^2}{n} + \dfrac{s_2^2}{n}}} = \dfrac{.2}{\sqrt{\dfrac{2}{n}}}$. For n = 100, z = 1.41 and p-value = $2[1 - \Phi(1.41)] = .1586$.

For n = 400, z = 2.83 and p-value = .0046. From a practical point of view, the closeness of $\bar{x}$ and $\bar{y}$ suggests that there is essentially no difference between true average fracture toughness for type I and type I steels. The very small difference in sample averages has been magnified by the large sample sizes – statistical rather than practical significance. The p-value by itself would not have conveyed this message.

Section 9.2

17.

a. $$n = \frac{\left(\frac{5^2}{10} + \frac{6^2}{10}\right)^2}{\frac{\left(\frac{5^2}{10}\right)^2}{9} + \frac{\left(\frac{6^2}{10}\right)^2}{9}} = \frac{37.21}{.694 + 1.44} = 17.43 \approx 17$$

b. $$n = \frac{\left(\frac{5^2}{10} + \frac{6^2}{15}\right)^2}{\frac{\left(\frac{5^2}{10}\right)^2}{9} + \frac{\left(\frac{6^2}{15}\right)^2}{14}} = \frac{24.01}{.694 + .411} = 21.7 \approx 21$$

c. $$n = \frac{\left(\frac{2^2}{10} + \frac{6^2}{15}\right)^2}{\frac{\left(\frac{2^2}{10}\right)^2}{9} + \frac{\left(\frac{6^2}{15}\right)^2}{14}} = \frac{7.84}{.018 + .411} = 18.27 \approx 18$$

d. $$n = \frac{\left(\frac{5^2}{12} + \frac{6^2}{24}\right)^2}{\frac{\left(\frac{5^2}{12}\right)^2}{11} + \frac{\left(\frac{6^2}{24}\right)^2}{23}} = \frac{12.84}{.395 + .098} = 26.05 \approx 26$$

18. With H_o: $m_1 - m_2 = 0$ vs. H_a: $m_1 - m_2 \neq 0$, we will reject H_o if $p - value < a$.

$$n = \frac{\left(\frac{.164^2}{6} + \frac{.240^2}{5}\right)^2}{\frac{\left(\frac{.164^2}{6}\right)^2}{5} + \frac{\left(\frac{.240^2}{5}\right)^2}{4}} = 6.8 \approx 6$$, and the test statistic

$$t = \frac{22.73 - 21.95}{\sqrt{\frac{.164^2}{6} + \frac{.240^2}{5}}} = \frac{.78}{.1265} = 6.17$$ leads to a p-value of 2[P(t > 6.17)] < 2(.0005) =.001,

which is less than most reasonable $a's$, so we reject H_o and conclude that there is a difference in the densities of the two brick types.

19. For the given hypotheses, the test statistic $t = \frac{115.7 - 129.3 + 10}{\sqrt{\frac{5.03^2}{6} + \frac{5.38^2}{6}}} = \frac{-3.6}{3.007} = -1.20$, and

the d.f. is $n = \frac{(4.2168 + 4.8241)^2}{\frac{(4.2168)^2}{5} + \frac{(4.8241)^2}{5}} = 9.96$, so use d.f. = 9. We will reject H_o if

$t \leq -t_{.01,9} = -2.764;$ since –1.20 > -2.764, we don't reject H_o.

20. We want a 95% confidence interval for $m_1 - m_2$. $t_{.025,9} = 2.262$, so the interval is $-3.6 \pm 2.262(3.007) = (-10.40, 3.20)$. Because the interval is so wide, it does not appear that precise information is available.

21. Let $m_1 =$ the true average gap detection threshold for normal subjects, and $m_2 =$ the corresponding value for CTS subjects. The relevant hypotheses are H_o: $m_1 - m_2 = 0$ vs. H_a: $m_1 - m_2 < 0$, and the test statistic $t = \frac{1.71 - 2.53}{\sqrt{.0351125 + .07569}} = \frac{-.82}{.3329} = -2.46$.

Using d.f. $n = \frac{(.0351125 + .07569)^2}{\frac{(.0351125)^2}{7} + \frac{(.07569)^2}{9}} = 15.1$, or 15, the rejection region is

$t \leq -t_{.01,15} = -2.602$. Since –2.46 is not ≤ -2.602, we fail to reject H_o. We have insufficient evidence to claim that the true average gap detection threshold for CTS subjects exceeds that for normal subjects.

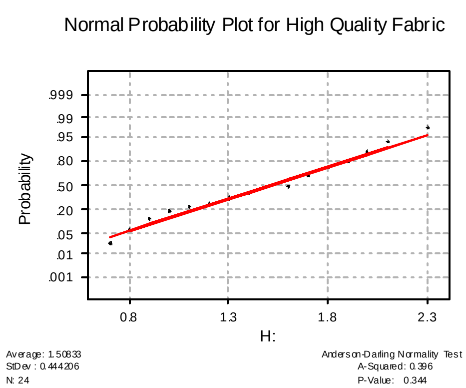
22. Let $m_1 =$ the true average strength for wire-brushing preparation and let $m_2 =$ the average strength for hand-chisel preparation. Since we are concerned about any possible difference between the two means, a two-sided test is appropriate. We test $H_0 : m_1 - m_2 = 0$ vs. $H_a : m_1 - m_2 \neq 0$. We need the degrees of freedom to find the rejection region:

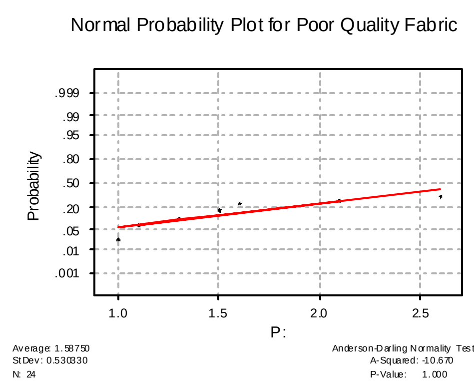
$$n = \frac{\left(\frac{1.58^2}{12} + \frac{4.01^2}{12}\right)^2}{\frac{\left(\frac{1.58^2}{12}\right)^2}{11} + \frac{\left(\frac{4.01^2}{5}\right)^2}{11}} = \frac{2.3964}{.0039 + .1632} = 14.33$$, which we round down to 14, so we reject H$_o$ if $|t| \geq t_{.025,14} = 2.145$. The test statistic is

$$t = \frac{19.20 - 23.13}{\sqrt{\left(\frac{1.58^2}{12} + \frac{4.01^2}{12}\right)}} = \frac{-3.93}{1.2442} = -3.159$$, which is ≤ -2.145, so we reject H$_o$ and conclude that there does appear to be a difference between the two population average strengths.

23.

a. Normal plots

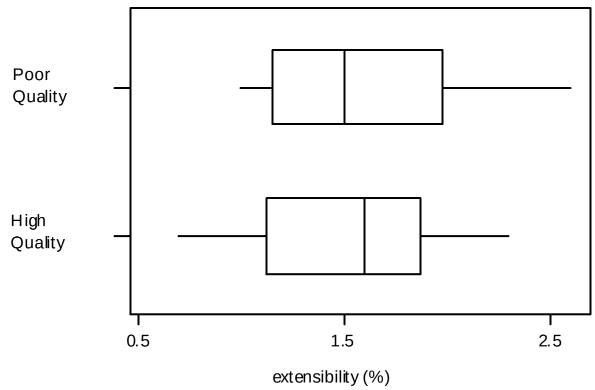




Using Minitab to generate normal probability plots, we see that both plots illustrate sufficient linearity. Therefore, it is plausible that both samples have been selected from normal population distributions.

b.

Comparative Box Plot for High Quality and Poor Quality Fabric



The comparative boxplot does not suggest a difference between average extensibility for the two types of fabrics.

c. We test $H_0 : m_1 - m_2 = 0$ vs. $H_a : m_1 - m_2 \neq 0$. With degrees of freedom $n = \frac{(.0433265)^2}{.00017906} = 10.5$, which we round down to 10, and using significance level .05 (not specified in the problem), we reject H_o if $|t| \geq t_{.025,10} = 2.228$. The test statistic is $t = \frac{-.08}{\sqrt{(.0433265)}} = -.38$, which is not ≥ 2.228 in absolute value, so we cannot reject H_o. There is insufficient evidence to claim that the true average extensibility differs for the two types of fabrics.

24. A 95% confidence interval for the difference between the true firmness of zero-day apples and the true firmness of 20-day apples is $(8.74 - 4.96) \pm t_{.025,n}\sqrt{\frac{.66^2}{20} + \frac{.39^2}{20}}$. We calculate the degrees of freedom $n = \frac{\left(\frac{.66^2}{20} + \frac{.39^2}{20}\right)^2}{\frac{\left(\frac{.66^2}{20}\right)^2}{19} + \frac{\left(\frac{.39^2}{20}\right)^2}{19}} = 30.83$, so we use 30 df, and $t_{.025,30} = 2.042$, so the interval is $3.78 \pm 2.042(.17142) = (3.43, 4.13)$. Thus, with 95% confidence, we can say that the true average firmness for zero-day apples exceeds that of 20-day apples by between 3.43 and 4.13 N.

25. We calculate the degrees of freedom $n = \dfrac{\left(\frac{5.5^2}{28}+\frac{7.8^2}{31}\right)^2}{\frac{\left(\frac{5.5^2}{28}\right)^2}{27}+\frac{\left(\frac{7.8^2}{31}\right)^2}{30}} = 53.95$, or about 54 (normally we would round down to 53, but this number is very close to 54 – of course for this large number of df, using either 53 or 54 won't make much difference in the critical t value) so the desired confidence interval is $(91.5-88.3)\pm 1.68\sqrt{\frac{5.5^2}{28}+\frac{7.8^2}{31}}$ $= 3.2 \pm 2.931 = (.269, 6.131)$. Because 0 does not lie inside this interval, we can be reasonably certain that the true difference $m_1 - m_2$ is not 0 and, therefore, that the two population means are not equal. For a 95% interval, the t value increases to about 2.01 or so, which results in the interval 3.2 ± 3.506. Since this interval does contain 0, we can no longer conclude that the means are different if we use a 95% confidence interval.

26. Let $m_1 =$ the true average potential drop for alloy connections and let $m_2 =$ the true average potential drop for EC connections. Since we are interested in whether the potential drop is higher for alloy connections, an upper tailed test is appropriate. We test $H_0: m_1 - m_2 = 0$ vs. $H_a: m_1 - m_2 > 0$. Using the SAS output provided, the test statistic, when assuming unequal variances, is t = 3.6362, the corresponding df is 37.5, and the p-value for our upper tailed test would be ½ (two-tailed p-value) = $\frac{1}{2}(.0008) = .0004$. Our p-value of .0004 is less than the significance level of .01, so we reject H_o. We have sufficient evidence to claim that the true average potential drop for alloy connections is higher than that for EC connections.

27. The approximate degrees of freedom for this estimate are

$$n = \frac{\left(\frac{11.3^2}{6}+\frac{8.3^2}{8}\right)^2}{\frac{\left(\frac{11.3^2}{6}\right)^2}{5}+\frac{\left(\frac{8.3^2}{8}\right)^2}{7}} = \frac{893.59}{101.175} = 8.83$$

, which we round down to 8, so $t_{.025,8} = 2.306$ and the desired interval is $(40.3-21.4)\pm 2.306\sqrt{\frac{11.3^2}{6}+\frac{8.3^2}{8}} = 18.9 \pm 2.306(5.4674)$ $= 18.9 \pm 12.607 = (6.3, 31.5)$. Because 0 is not contained in this interval, there is strong evidence that $m_1 - m_2$ is not 0; i.e., we can conclude that the population means are not equal. Calculating a confidence interval for $m_2 - m_1$ would change only the order of subtraction of the sample means, but the standard error calculation would give the same result as before. Therefore, the 95% interval estimate of $m_2 - m_1$ would be (-31.5, -6.3), just the negatives of the endpoints of the original interval. Since 0 is not in this interval, we reach exactly the same conclusion as before; the population means are not equal.

28. We will test the hypotheses: $H_0: \mu_1 - \mu_2 = 10$ vs. $H_a: \mu_1 - \mu_2 > 10$. The test statistic is $t = \frac{(\bar{x} - \bar{y}) - 10}{\sqrt{\left(\frac{2.75^2}{10} + \frac{4.44^2}{5}\right)}} = \frac{4.5}{2.17} = 2.08$ The degrees of freedom

$$\nu = \frac{\left(\frac{2.75^2}{10} + \frac{4.44^2}{5}\right)^2}{\frac{\left(\frac{2.75^2}{10}\right)^2}{9} + \frac{\left(\frac{4.44^2}{5}\right)^2}{4}} = \frac{22.08}{3.95} = 5.59 \approx 6$$ and the p-value from table A.8 is approx .04, which is < .10 so we reject H_0 and conclude that the true average lean angle for older females is more than 10 degrees smaller than that of younger females.

29. Let $\mu_1 =$ the true average compression strength for strawberry drink and let $\mu_2 =$ the true average compression strength for cola. A lower tailed test is appropriate. We test $H_0: \mu_1 - \mu_2 = 0$ vs. $H_a: \mu_1 - \mu_2 < 0$. The test statistic is

$$t = \frac{-14}{\sqrt{29.4 + 15}} = -2.10 . \quad \nu = \frac{(44.4)^2}{\frac{(29.4)^2}{14} + \frac{(15)^2}{14}} = \frac{1971.36}{77.8114} = 25.3$$, so use df=25.

The p-value $\approx P(t < -2.10) = .023$. This p-value indicates strong support for the alternative hypothesis. The data does suggest that the extra carbonation of cola results in a higher average compression strength.

30.

a. We desire a 99% confidence interval. First we calculate the degrees of freedom:

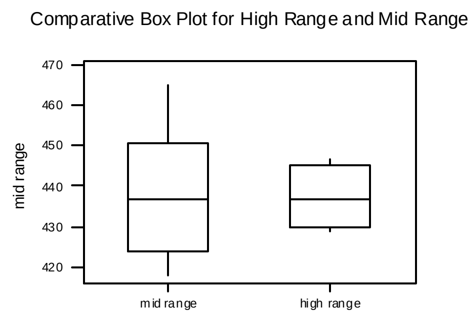
$$\nu = \frac{\left(\frac{2.2^2}{26} + \frac{4.3^2}{26}\right)^2}{\frac{\left(\frac{2.2^2}{26}\right)^2}{26} + \frac{\left(\frac{4.3^2}{26}\right)^2}{26}} = 37.24$$, which we would round down to 37, except that there is no df = 37 row in Table A.5. Using 36 degrees of freedom (a more conservative choice), $t_{.005,36} = 2.719$, and the 99% C.I. is

$$(33.4 - 42.8) \pm 2.719\sqrt{\frac{2.2^2}{26} + \frac{4.3^2}{26}} = -9.4 \pm 2.576 = (-11.98, -6.83)$$. We are very confident that the true average load for carbon beams exceeds that for fiberglass beams by between 6.83 and 11.98 kN.

b. The upper limit of the interval in part **a** does not give a 99% upper confidence bound. The 99% upper bound would be $-9.4 + 2.434(.9473) = -7.09$, meaning that the true average load for carbon beams exceeds that for fiberglass beams by at least 7.09 kN.

31.

a.



The most notable feature of these boxplots is the larger amount of variation present in the mid-range data compared to the high-range data. Otherwise, both look reasonably symmetric with no outliers present.

b. Using df = 23, a 95% confidence interval for $m_{mid-range} - m_{high-range}$ is $(438.3 - 437.45) \pm 2.069\sqrt{\frac{15.1^2}{17} + \frac{6.83^2}{11}} = .85 \pm 8.69 = (-7.84, 9.54)$. Since plausible values for $m_{mid-range} - m_{high-range}$ are both positive and negative (i.e., the interval spans zero) we would conclude that there is not sufficient evidence to suggest that the average value for mid-range and the average value for high-range differ.

32. Let $m_1 =$ the true average proportional stress limit for red oak and let $m_2 =$ the true average proportional stress limit for Douglas fir. We test $H_0 : m_1 - m_2 = 1$ vs. $H_a : m_1 - m_2 > 1$. The test statistic is $t = \frac{(8.48 - 6.65) - 1}{\sqrt{\frac{.79^2}{14} + \frac{1.28^2}{10}}} = \frac{1.83}{\sqrt{.2084}} 1.818$. With degrees of freedom

$$n = \frac{(.2084)^2}{\frac{\left(\frac{.79^2}{14}\right)^2}{13} + \frac{\left(\frac{1.28^2}{10}\right)^2}{9}} = 13.85 \approx 14$$

, the p-value $\approx P(t > 1.8) = .046$. This p-value indicates strong support for the alternative hypothesis since we would reject H_o at significance levels greater than .046. There is sufficient evidence to claim that true average proportional stress limit for red oak exceeds that of Douglas fir by more than 1 MPa.

33. Let $m_1 =$ the true average weight gain for steroid treatment and let $m_2 =$ the true average weight gain for the population not treated with steroids. The exercise asks if we can conclude that m_2 exceeds m_1 by more than 5 g., which we can restate in the equivalent form: $m_1 - m_2 < -5$. Therefore, we conduct a lower-tailed test of $H_0 : m_1 - m_2 = -5$ vs. $H_a : m_1 - m_2 < -5$. The test statistic is

$$t = \frac{(\bar{x} - \bar{y}) - (\Delta)}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} = \frac{32.8 - 40.5 - (-5)}{\sqrt{\frac{2.6^2}{8} + \frac{2.5^2}{10}}} = \frac{-2.7}{1.2124} = -2.23 \approx 2.2.$$

The approximate d.f. is

$$n = \frac{\left(\frac{2.6^2}{8} + \frac{2.5^2}{10}\right)^2}{\frac{\left(\frac{2.6^2}{8}\right)^2}{7} + \frac{\left(\frac{2.5^2}{10}\right)^2}{9}} = \frac{2.1609}{.1454} = 14.876$$

, which we round down to 14. The p-value for a lower tailed test is P(t < -2.2) = P(t > 2.2) = .022. Since this p-value is larger than the specified significance level .01, we cannot reject H_o. Therefore, this data does not support the belief that average weight gain for the control group exceeds that of the steroid group by more than 5 g.

34.

a. Following the usual format for most confidence intervals: *statistic ± (critical value)(standard error),* a pooled variance confidence interval for the difference between two means is $(\bar{x} - \bar{y}) \pm t_{a/2,m+n-2} \cdot s_p \sqrt{\frac{1}{m} + \frac{1}{n}}$.

b. The sample means and standard deviations of the two samples are $\bar{x} = 13.90$, $s_1 = 1.225$, $\bar{y} = 12.20$, $s_2 = 1.010$. The pooled variance estimate is $s_p^2 =$

$$\left(\frac{m-1}{m+n-2}\right)s_1^2 + \left(\frac{n-1}{m+n-2}\right)s_2^2 = \left(\frac{4-1}{4+4-2}\right)(1.225)^2 + \left(\frac{4-1}{4+4-2}\right)(1.010)^2$$

$= 1.260$, so $s_p = 1.1227$. With df = m+n-1 = 6 for this interval, $t_{.025,6} = 2.447$ and the desired interval is $(13.90 - 12.20) \pm (2.447)(1.1227)\sqrt{\frac{1}{4} + \frac{1}{4}}$ $= 1.7 \pm 1.943 = (-.24, 3.64)$. This interval contains 0, so it does not support the conclusion that the two population means are different.

c. Using the two-sample t interval discussed earlier, we use the CI as follows: First, we need to calculate the degrees of freedom.

$$n = \frac{\left(\frac{1.225^2}{4} + \frac{1.01^2}{4}\right)^2}{\frac{\left(\frac{1.225^2}{4}\right)^2}{3} + \frac{\left(\frac{1.01^2}{4}\right)^2}{3}} = \frac{.6302}{.0686} = 9.19 \approx 9$$

so $t_{.025,9} = 2.262$. Then the interval is

$$(13.9 - 12.2) \pm 2.262\sqrt{\frac{1.225^2}{4} + \frac{1.01^2}{4}} = 1.70 \pm 2.262(.7938) = (-.10, 3.50).$$

This interval is slightly smaller, but it still supports the same conclusion.

35. There are two changes that must be made to the procedure we currently use. First, the equation used to compute the value of the t test statistic is: $t = \dfrac{(\bar{x}-\bar{y})-(\Delta)}{s_p\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}}$ where s_p is defined as in Exercise 34 above. Second, the degrees of freedom = m + n – 2. Assuming equal variances in the situation from Exercise 33, we calculate s_p as follows:

$s_p = \sqrt{\left(\dfrac{7}{16}\right)(2.6)^2 + \left(\dfrac{9}{16}\right)(2.5)^2} = 2.544$. The value of the test statistic is, then,

$t = \dfrac{(32.8-40.5)-(-5)}{2.544\sqrt{\dfrac{1}{8}+\dfrac{1}{10}}} = -2.24 \approx -2.2$. The degrees of freedom = 16, and the p-value is P (t < -2.2) = .021. Since .021 > .01, we fail to reject H_o. This is the same conclusion reached in Exercise 33.

Section 9.3

36. $\bar{d} = 7.25$, $s_D = 11.8628$

1 Parameter of Interest: $m_D =$ true average difference of breaking load for fabric in unabraded or abraded condition.

2 $H_0 : m_D = 0$

3 $H_a : m_D > 0$

4 $t = \dfrac{\bar{d} - m_D}{s_D / \sqrt{n}} = \dfrac{\bar{d} - 0}{s_D / \sqrt{n}}$

5 RR: $t \ge t_{.01,7} = 2.998$

6 $t = \dfrac{7.25 - 0}{11.8628/\sqrt{8}} = 1.73$

7 Fail to reject H_o. The data does not indicate a difference in breaking load for the two fabric load conditions.

37.

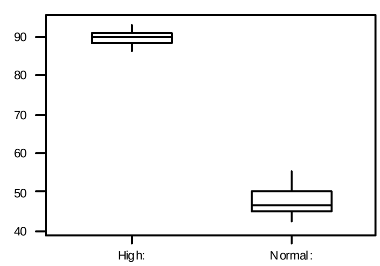
a. This exercise calls for paired analysis. First, compute the difference between indoor and outdoor concentrations of hexavalent chromium for each of the 33 houses. These 33 differences are summarized as follows: n = 33, $\bar{d} = -.4239$, $s_d = .3868$, where d = (indoor value – outdoor value). Then $t_{.025,32} = 2.037$, and a 95% confidence interval for the population mean difference between indoor and outdoor concentration is $-.4239 \pm (2.037)\left(\frac{.3868}{\sqrt{33}}\right) = -.4239 \pm .13715 = (-.5611, -.2868)$. We can be highly confident, at the 95% confidence level, that the true average concentration of hexavalent chromium outdoors exceeds the true average concentration indoors by between .2868 and .5611 nanograms/m^3.

b. A 95% prediction interval for the difference in concentration for the 34th house is $\bar{d} \pm t_{.025,32}\left(s_d\sqrt{1+\frac{1}{n}}\right) = -.4239 \pm (2.037)\left(.3868\sqrt{1+\frac{1}{33}}\right) = (-1.224, .3758)$. This prediction interval means that the indoor concentration may exceed the outdoor concentration by as much as .3758 nanograms/m^3 and that the outdoor concentration may exceed the indoor concentration by a much as 1.224 nanograms/m^3, for the 34th house. Clearly, this is a wide prediction interval, largely because of the amount of variation in the differences.

38.

a. The median of the "Normal" data is 46.80 and the upper and lower quartiles are 45.55 and 49.55, which yields an IQR of 49.55 – 45.55 = 4.00. The median of the "High" data is 90.1 and the upper and lower quartiles are 88.55 and 90.95, which yields an IQR of 90.95 – 88.55 = 2.40. The most significant feature of these boxplots is the fact that their locations (medians) are far apart.

Comparative Boxplots
for Normal and High Strength Concrete Mix



b. This data is paired because the two measurements are taken for each of 15 test conditions. Therefore, we have to work with the differences of the two samples. A quantile of the 15 differences shows that the data follows (approximately) a straight line, indicating that it is reasonable to assume that the differences follow a normal distribution. Taking differences in the order "Normal" – "High" , we find $\bar{d} = -42.23$, and $s_d = 4.34$.

With $t_{.025,14} = 2.145$, a 95% confidence interval for the difference between the population means is

$-42.23 \pm (2.145)\left(\frac{4.34}{\sqrt{15}}\right) = -42.23 \pm 2.404 = (-44.63, -39.83)$. Because 0 is not contained in this interval, we can conclude that the difference between the population means is not 0; i.e., we conclude that the two population means are not equal.

39.

a. A normal probability plot shows that the data could easily follow a normal distribution.

b. We test $H_0 : m_d = 0$ vs. $H_a : m_d \neq 0$, with test statistic

$t = \frac{\bar{d} - 0}{s_D / \sqrt{n}} = \frac{167.2 - 0}{228 / \sqrt{14}} = 2.74 \approx 2.7$. The two-tailed p-value is 2[P(t > 2.7)] = 2[.009] = .018. Since .018 < .05, we can reject H_o . There is strong evidence to support the claim that the true average difference between intake values measured by the two methods is not 0. There is a difference between them.

40.

a. H_o will be rejected in favor of H_a if either $t \geq t_{.005,15} = 2.947$ or $t \leq -2.947$. The summary quantities are $\bar{d} = -.544$, and $s_d = .714$, so $t = \frac{-.544}{.1785} = -3.05$.

Because $-3.05 \leq -2.947$, H_o is rejected in favor of H_a.

b. $s_p^2 = 7.31$, $s_p = 2.70$, and $t = \frac{-.544}{.96} = -.57$, which is clearly insignificant; the incorrect analysis yields an inappropriate conclusion.

41. We test $H_0 : m_d = 0$ vs. $H_a : m_d > 0$. With $\bar{d} = 7.600$, and $s_d = 4.178$,

$t = \frac{7.600 - 5}{4.178 / \sqrt{9}} = \frac{2.6}{1.39} = 1.87 \approx 1.9$. With degrees of freedom n – 1 = 8, the corresponding p-value is P(t > 1.9) = .047. We would reject H_o at any alpha level greater than .047. So, at the typical significance level of .05, we would (barely) reject H_o, and conclude that the data indicates that the higher level of illumination yields a decrease of more than 5 seconds in true average task completion time.

42.

1 Parameter of interest: m_d denotes the true average difference of spatial ability in brothers exposed to DES and brothers not exposed to DES. Let $m_d = m_{exp\,osed} - m_{un\,exp\,osed}$.

2 $H_0 : m_D = 0$

3 $H_a : m_D < 0$

4 $t = \dfrac{\bar{d} - m_D}{s_D / \sqrt{n}} = \dfrac{\bar{d} - 0}{s_D / \sqrt{n}}$

5 RR: P-value < .05, df = 8

6 $t = \dfrac{(12.6 - 13.7) - 0}{0.5} = -2.2$, with corresponding p-value .029 (from Table A.8)

7 Reject H_o. The data supports the idea that exposure to DES reduces spatial ability.

43.

a. Although there is a "jump" in the middle of the Normal Probability plot, the data follow a reasonably straight path, so there is no strong reason for doubting the normality of the population of differences.

b. A 95% lower confidence bound for the population mean difference is:

$$\bar{d} - t_{.05,14}\left(\frac{s_d}{\sqrt{n}}\right) = -38.60 - (1.761)\left(\frac{23.18}{\sqrt{15}}\right) = -38.60 - 10.54 = -49.14.$$

Therefore, with a confidence level of 95%, the population mean difference is above (–49.14).

c. A 95% upper confidence bound for the corresponding population mean difference is $38.60 + 10.54 = 49.14$

44. We need to check the differences to see if the assumption of normality is plausible. A probability chart will validate our use of the t distribution. A 95% confidence interval:

$$\bar{d} + t_{.05,15}\left(\frac{s_d}{\sqrt{n}}\right) = 2635.63 + (1.753)\left(\frac{508.645}{\sqrt{16}}\right) = 2635.63 + 222.91$$
$$\Rightarrow (\infty, 2858.54)$$

45. The differences (white – black) are –7.62, -8.00, -9.09, -6.06, -1.39, -16.07, -8.40, -8.89, and –2.88, from which $\bar{d} = -7.600$, and $s_d = 4.178$. The confidence level is not specified in the problem description; for 95% confidence, $t_{.025,8} = 2.306$, and the C.I. is

$$-7.600 \pm (2.306)\left(\frac{4.178}{\sqrt{9}}\right) = -7.600 \pm 3.211 = (-10.811, -4.389).$$

46. With $(x_1, y_1) = (6,5)$, $(x_2, y_2) = (15,14)$, $(x_3, y_3) = (1,0)$, and $(x_4, y_4) = (21,20)$, $\bar{d} = 1$ and $s_d = 0$ (the d_I's are 1, 1, 1, and 1), while $s_1 = s_2 = 8.96$, so $s_p = 8.96$ and t = .16.

Section 9.4

47. H_o will be rejected if $z \le -z_{.01} = -2.33$. With $\hat{p}_1 = .150$, and $\hat{p}_2 = .300$,

$\hat{p} = \frac{30+80}{200+600} = \frac{210}{800} = .263$, and $\hat{q} = .737$. The calculated test statistic is

$z = \frac{.150 - .300}{\sqrt{(.263)(.737)\left(\frac{1}{200}+\frac{1}{600}\right)}} = \frac{-.150}{.0359} = -4.18$. Because $-4.18 \le -2.33$, H_o is rejected; the proportion of those who repeat after inducement appears lower than those who repeat after no inducement.

48.

a. H_o will be rejected if $|z| \ge 1.96$. With $\hat{p}_1 = \frac{63}{300} = .2100$, and $\hat{p}_2 = \frac{75}{180} = .4167$,

$\hat{p} = \frac{63+75}{300+180} = .2875$, $z = \frac{.2100 - .4167}{\sqrt{(.2875)(.7125)\left(\frac{1}{300}+\frac{1}{180}\right)}} = \frac{-.2067}{.0427} = -4.84$.

Since $-4.84 \le -1.96$, H_o is rejected.

b. $\bar{p} = .275$ and $s = .0432$, so power =

$$1 - \left[\Phi\left(\frac{[(1.96)(.0421) + .2]}{.0432}\right) - \Phi\left(\frac{[-(1.96)(.0421) + .2]}{.0432}\right)\right] =$$

$$1 - [\Phi(6.54) - \Phi(2.72)] = .9967 .$$

49.

1 Parameter of interest: $p_1 - p_2$ = true difference in proportions of those responding to two different survey covers. Let p_1 = Plain, p_2 = Picture.

2 $H_0 : p_1 - p_2 = 0$

3 $H_a : p_1 - p_2 < 0$

4 $z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{m}+\frac{1}{n}\right)}}$

5 Reject H_o if p-value < .10

6 $z = \frac{\frac{104}{207} - \frac{109}{213}}{\sqrt{\left(\frac{213}{420}\right)\left(\frac{207}{420}\right)\left(\frac{1}{207}+\frac{1}{213}\right)}} = -.1910$; p-value = .4247

7 Fail to Reject H_o. The data does not indicate that plain cover surveys have a lower response rate.

50. Let $\alpha = .05$. A 95% confidence interval is $(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\left(\frac{\hat{p}_1\hat{q}_1}{m} + \frac{\hat{p}_2\hat{q}_2}{n}\right)}$

$$= \left(\tfrac{224}{395} - \tfrac{126}{266}\right) \pm 1.96\sqrt{\left(\frac{\left(\frac{224}{395}\right)\left(\frac{171}{395}\right)}{395} + \frac{\left(\frac{126}{266}\right)\left(\frac{140}{266}\right)}{266}\right)} = .0934 \pm .0774 = (.0160, .1708).$$

51.

a. $H_0: p_1 = p_2$ will be rejected in favor of $H_a: p_1 \neq p_2$ if either $z \geq 1.645$ or $z \leq -1.645$. With $\hat{p}_1 = .193$, and $\hat{p}_2 = .182$, $\hat{p} = .188$, $z = \dfrac{.011}{.00742} = 1.48$. Since 1.48 is not ≥ 1.645, H_o is not rejected and we conclude that no difference exists.

b. Using formula (9.7) with $p_1 = .2$, $p_2 = .18$, $\alpha = .1$, $\beta = .1$, and $z_{\alpha/2} = 1.645$,

$$n = \frac{\left(1.645\sqrt{.5(.38)(1.62)} + 1.28\sqrt{.16 + .1476}\right)^2}{.0004} = 6582$$

52. Let p_1 = true proportion of irradiated bulbs that are marketable; p_2 = true proportion of untreated bulbs that are marketable; The hypotheses are $H_0: p_1 - p_2 = 0$ vs. $H_0: p_1 - p_2 > 0$. The test statistic is $z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{m} + \frac{1}{n}\right)}}$. With $\hat{p}_1 = \dfrac{153}{180} = .850$, and $\hat{p}_2 = \dfrac{119}{180} = .661$, $\hat{p} = \dfrac{272}{360} = .756$, $z = \dfrac{.850 - .661}{\sqrt{(.756)(.244)\left(\frac{1}{180} + \frac{1}{180}\right)}} = \dfrac{.189}{.045} = 4.2$.

The p-value = $1 - \Phi(4.2) \approx 0$, so reject H_o at any reasonable level. Radiation appears to be beneficial.

53.

a. A 95% large sample confidence interval formula for $\ln(\theta)$ is $\ln(\hat{\theta}) \pm z_{\alpha/2}\sqrt{\dfrac{m-x}{mx} + \dfrac{n-y}{ny}}$. Taking the antilogs of the upper and lower bounds gives the confidence interval for θ itself.

b. $\hat{\theta} = \dfrac{\frac{189}{11,034}}{\frac{104}{11,037}} = 1.818$, $\ln(\hat{\theta}) = .598$, and the standard deviation is

$$\sqrt{\frac{10,845}{(11,034)(189)} + \frac{10,933}{(11,037)(104)}} = .1213,$$ so the CI for $\ln(\theta)$ is $.598 \pm 1.96(.1213) = (.360, .836)$. Then taking the antilogs of the two bounds gives the CI for θ to be $(1.43, 2.31)$.

54.

a. The "after" success probability is $p_1 + p_3$ while the "before" probability is $p_1 + p_2$, so $p_1 + p_3 > p_1 + p_2$ becomes $p_3 > p_2$; thus we wish to test $H_0 : p_3 = p_2$ versus $H_a : p_3 > p_2$.

b. The estimator of $(p_1 + p_3) - (p_1 + p_2)$ is $\dfrac{(X_1 + X_3) - (X_1 + X_2)}{n} = \dfrac{X_3 - X_2}{n}$.

c. When H_o is true, $p_2 = p_3$, so $Var\left(\dfrac{X_3 - X_2}{n}\right) = \dfrac{p_2 + p_3}{n}$, which is estimated by $\dfrac{\hat{p}_2 + \hat{p}_3}{n}$. The Z statistic is then $\dfrac{\frac{X_3 - X_2}{n}}{\sqrt{\frac{\hat{p}_2 + \hat{p}_3}{n}}} = \dfrac{X_3 - X_2}{\sqrt{X_2 + X_3}}$.

d. The computed value of Z is $\dfrac{200 - 150}{\sqrt{200 + 150}} = 2.68$, so $P = 1 - \Phi(2.68) = .0037$. At level .01, H_o can be rejected but at level .001 H_o would not be rejected.

55. $\hat{p}_1 = \dfrac{15 + 7}{40} = .550$, $\hat{p}_2 = \dfrac{29}{42} = .690$, and the 95% C.I. is $(.550 - .690) \pm 1.96(.106) = -.14 \pm .21 = (-.35, .07)$.

56. Using $p_1 = q_1 = p_2 = q_2 = .5$, $L = 2(1.96)\sqrt{\left(\dfrac{.25}{n} + \dfrac{.25}{n}\right)} = \dfrac{2.7719}{\sqrt{n}}$, so L=.1 requires n=769.

Section 9.5

57.

a. From Table A.9, column 5, row 8, $F_{.01,5,8} = 3.69$.

b. From column 8, row 5, $F_{.01,8,5} = 4.82$.

c. $F_{.95,5,8} = \dfrac{1}{F_{.05,8,5}} = .207$.

d. $F_{.95,8,5} = \dfrac{1}{F_{.05,5,8}} = .271$

e. $F_{.01,10,12} = 4.30$

f. $F_{.99,10,12} = \dfrac{1}{F_{.01,12,10}} = \dfrac{1}{4.71} = .212$.

g. $F_{.05,6,4} = 6.16$, so $P(F \le 6.16) = .95$.

h. Since $F_{.99,10,5} = \dfrac{1}{5.64} = .177$,
$P(.177 \le F \le 4.74) = P(F \le 4.74) - P(F \le .177) = .95 - .01 = .94$.

58.

a. Since the given f value of 4.75 falls between $F_{.05,5,10} = 3.33$ and $F_{.01,5,10} = 5.64$, we can say that the upper-tailed p-value is between .01 and .05.

b. Since the given f of 2.00 is less than $F_{.10,5,10} = 2.52$, the p-value > .10.

c. The two tailed p-value = $2P(F \ge 5.64) = 2(.01) = .02$.

d. For a lower tailed test, we must first use formula 9.9 to find the critical values:
$F_{.90,5,10} = \dfrac{1}{F_{.10,10,5}} = .3030$, $F_{.95,5,10} = \dfrac{1}{F_{.05,10,5}} = .2110$,
$F_{.99,5,10} = \dfrac{1}{F_{.01,10,5}} = .0995$. Since .0995 < f = .200 < .2110, .01 < p-value < .05 (but obviously closer to .05).

e. There is no column for numerator d.f. of 35 in Table A.9, however looking at both df = 30 and df = 40 columns, we see that for denominator df = 20, our f value is between $F_{.01}$ and $F_{.001}$. So we can say .001< p-value < .01.

59. We test $H_0 : s_1^2 = s_2^2$ vs. $H_a : s_1^2 \neq s_2^2$. The calculated test statistic is $f = \frac{(2.75)^2}{(4.44)^2} = .384$. With numerator d.f. = m – 1 = 10 – 1 = 9, and denominator d.f. = n – 1 = 5 – 1 = 4, we reject H_0 if $f \geq F_{.05,9,4} = 6.00$ or $f \leq F_{.95,9,4} = 1/F_{.05,4,9} = 1/3.63 = .275$. Since .384 is in neither rejection region, we do not reject H_0 and conclude that there is no significant difference between the two standard deviations.

60. With $s_1 =$ true standard deviation for not-fused specimens and $s_2 =$ true standard deviation for fused specimens, we test $H_0 : s_1 = s_2$ vs. $H_a : s_1 > s_2$. The calculated test statistic is $f = \frac{(277.3)^2}{(205.9)^2} = 1.814$. With numerator d.f. = m – 1 = 10 – 1 = 9, and denominator d.f. = n – 1 = 8 – 1 = 7, $f = 1.814 < 2.72 = F_{.10,9,7}$. We can say that the p-value > .10, which is obviously > .01, so we cannot reject H_o. There is not sufficient evidence that the standard deviation of the strength distribution for fused specimens is smaller than that of not-fused specimens.

61. Let $s_1^2 =$ variance in weight gain for low-dose treatment, and $s_2^2 =$ variance in weight gain for control condition. We wish to test $H_0 : s_1^2 = s_2^2$ vs. $H_a : s_1^2 > s_2^2$. The test statistic is $f = \frac{s_1^2}{s_2^2}$, and we reject H_o at level .05 if $f > F_{.05,19,22} \approx 2.08$.

$f = \frac{(54)^2}{(32)^2} = 2.85 \geq 20.8$, so reject H_o at level .05. The data does suggest that there is more variability in the low-dose weight gains.

62. $H_0 : s_1 = s_2$ will be rejected in favor of $H_a : s_1 \neq s_2$ if either $f \leq F_{.975,47,44} \approx .56$ or if $f \geq F_{.025,47,44} \approx 1.8$. Because $f = 1.22$, H_o is not rejected. The data does not suggest a difference in the two variances.

63. $P\left(F_{1-a/2,m-1,n-1} \le \dfrac{S_1^2/s_1^2}{S_2^2/s_2^2} \le F_{a/2,m-1,n-1}\right) = 1-a$. The set of inequalities inside the parentheses is clearly equivalent to $\dfrac{S_2^2 F_{1-a/2,m-1,n-1}}{S_1^2} \le \dfrac{s_2^2}{s_1^2} \le \dfrac{S_2^2 F_{a/2,m-1,n-1}}{S_1^2}$. Substituting the sample values s_1^2 and s_2^2 yields the confidence interval for $\dfrac{s_2^2}{s_1^2}$, and taking the square root of each endpoint yields the confidence interval for $\dfrac{s_2}{s_1}$. m = n = 4, so we need $F_{.05,3,3} = 9.28$ and $F_{.95,3,3} = \dfrac{1}{9.28} = .108$. Then with $s_1 = .160$ and $s_2 = .074$, the C. I. for $\dfrac{s_2^2}{s_1^2}$ is (.023, 1.99), and for $\dfrac{s_2}{s_1}$ is (.15, 1.41).

64. A 95% upper bound for $\dfrac{s_2}{s_1}$ is $\sqrt{\dfrac{s_2^2 F_{.05,9,9}}{s_1^2}} = \sqrt{\dfrac{(3.59)^2(3.18)}{(.79)^2}} = 8.10$. We are confident that the ratio of the standard deviation of triacetate porosity distribution to that of the cotton porosity distribution is at most 8.10.

Supplementary Exercises

65. We test $H_0: m_1 - m_2 = 0$ vs. $H_a: m_1 - m_2 \ne 0$. The test statistic is

$$t = \frac{(\bar{x}-\bar{y})-(\Delta)}{\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}}} = \frac{807-757}{\sqrt{\frac{27^2}{10}+\frac{41^2}{10}}} = \frac{50}{\sqrt{241}} = \frac{50}{15.524} = 3.22$$

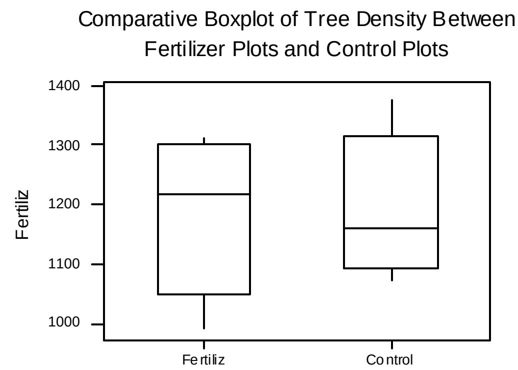
The approximate d.f. is

$$n = \frac{(241)^2}{\frac{(72.9)^2}{9}+\frac{(168.1)^2}{9}} = 15.6$$

, which we round down to 15. The p-value for a two-tailed test is approximately 2P(t > 3.22) = 2(.003) = .006. This small of a p-value gives strong support for the alternative hypothesis. The data indicates a significant difference.

66.

a.



Although the median of the fertilizer plot is higher than that of the control plots, the fertilizer plot data appears negatively skewed, while the opposite is true for the control plot data.

b. A test of $H_0 : m_1 - m_2 = 0$ vs. $H_a : m_1 - m_2 \neq 0$ yields a t value of -.20, and a two-tailed p-value of .85. (d.f. = 13). We would fail to reject H_o; the data does not indicate a significant difference in the means.

c. With 95% confidence we can say that the true average difference between the tree density of the fertilizer plots and that of the control plots is somewhere between –144 and 120. Since this interval contains 0, 0 is a plausible value for the difference, which further supports the conclusion based on the p-value.

67. Let p_1 = true proportion of returned questionnaires that included no incentive; p_2 = true proportion of returned questionnaires that included an incentive. The hypotheses are $H_0 : p_1 - p_2 = 0$ vs. $H_0 : p_1 - p_2 < 0$. The test statistic is $z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{m} + \frac{1}{n}\right)}}$.

$\hat{p}_1 = \dfrac{75}{110} = .682$, and $\hat{p}_2 = \dfrac{66}{98} = .673$. At this point we notice that since $\hat{p}_1 > \hat{p}_2$, the numerator of the z statistic will be > 0, and since we have a lower tailed test, the p-value will be > .5. We fail to reject H_o. This data does not suggest that including an incentive increases the likelihood of a response.

68. Summary quantities are m = 24, $\bar{x} = 103.66$, $s_1 = 3.74$, n = 11, $\bar{y} = 101.11$, $s_2 = 3.60$. We use the pooled t interval based on 24 + 11 – 2 = 33 d.f.; 95% confidence requires $t_{.025,33} = 2.03$. With $s_p^2 = 13.68$ and $s_p = 3.70$, the confidence interval is $2.55 \pm (2.03)(3.70)\sqrt{\frac{1}{24} + \frac{1}{11}} = 2.55 \pm 2.73 = (-.18, 5.28)$. We are confident that the difference between true average dry densities for the two sampling methods is between -.18 and 5.28. Because the interval contains 0, we cannot say that there is a significant difference between them.

69. The center of any confidence interval for $m_1 - m_2$ is always $\bar{x}_1 - \bar{x}_2$, so $\bar{x}_1 - \bar{x}_2 = \frac{-473.3 + 1691.9}{2} = 609.3$. Furthermore, half of the width of this interval is $\frac{1691.9 - (-473.3)}{2} = 1082.6$. Equating this value to the expression on the right of the 95% confidence interval formula, $1082.6 = (1.96)\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$, we find $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \frac{1082.6}{1.96} = 552.35$. For a 90% interval, the associated z value is 1.645, so the 90% confidence interval is then $609.3 \pm (1.645)(552.35) = 609.3 \pm 908.6$ $= (-299.3, 1517.9)$.

70.

a. A 95% lower confidence bound for the true average strength of joints with a side coating is $\bar{x} - t_{.025,9}\left(\frac{s}{\sqrt{n}}\right) = 63.23 - (1.833)\left(\frac{5.96}{\sqrt{10}}\right) = 63.23 - 3.45 = 59.78$. That is, with a confidence level of 95%, the average strength of joints with a side coating is at least 59.78 (Note: this bound is valid only if the distribution of joint strength is normal.)

b. A 95% lower prediction bound for the strength of a single joint with a side coating is $\bar{x} - t_{.025,9}\left(s\sqrt{1 + \frac{1}{n}}\right) = 63.23 - (1.833)\left(5.96\sqrt{1 + \frac{1}{10}}\right) = 63.23 - 11.46 = 51.77$. That is, with a confidence level of 95%, the strength of a single joint with a side coating would be at least 51.77.

c. For a confidence level of 95%, a two-sided tolerance interval for capturing at least 95% of the strength values of joints with side coating is $\bar{x} \pm$ (tolerance critical value)s. The tolerance critical value is obtained from Table A.6 with 95% confidence, k = 95%, and n = 10. Thus, the interval is $63.23 \pm (3.379)(5.96) = 63.23 \pm 20.14 = (43.09, 83.37)$. That is, we can be highly confident that at least 95% of all joints with side coatings have strength values between 43.09 and 83.37.

d. A 95% confidence interval for the difference between the true average strengths for the two types of joints is $(80.95 - 63.23) \pm t_{.025,\nu}\sqrt{\frac{(9.59)^2}{10} + \frac{(5.96)^2}{10}}$. The approximate degrees of freedom is $\nu = \dfrac{\left(\frac{91.9681}{10} + \frac{35.5216}{10}\right)^2}{\frac{\left(\frac{91.9681}{10}\right)^2}{9} + \frac{\left(\frac{35.5216}{10}\right)^2}{9}} = 15.05$, so we use 15 d.f., and $t_{.025,15} = 2.131$. The interval is , then, $17.72 \pm (2.131)(3.57) = 17.72 \pm 7.61 = (10.11, 25.33)$. With 95% confidence, we can say that the true average strength for joints without side coating exceeds that of joints with side coating by between 10.11 and 25.33 lb-in./in.

71. m = n = 40, $\bar{x} = 3975.0$, s_1 = 245.1, $\bar{y} = 2795.0$, s_2 = 293.7. The large sample 99% confidence interval for $\mu_1 - \mu_2$ is $(3975.0 - 2795.0) \pm 2.58\sqrt{\frac{245.1^2}{40} + \frac{293.7^2}{40}}$ $(1180.0) \pm 1560.5 \approx (1024, 1336)$. The value 0 is not contained in this interval so we can state that, with very high confidence, the value of $\mu_1 - \mu_2$ is not 0, which is equivalent to concluding that the population means are not equal.

72. This exercise calls for a paired analysis. First compute the difference between the amount of cone penetration for commutator and pinion bearings for each of the 17 motors. These 17 differences are summarized as follows: n = 17, $\bar{d} = -4.18$, $s_d = 35.85$, where d = (commutator value – pinion value). Then $t_{.025,16} = 2.120$, and the 95% confidence interval for the population mean difference between penetration for the commutator armature bearing and penetration for the pinion bearing is:
$-4.18 \pm (2.120)\left(\frac{35.85}{\sqrt{17}}\right) = -4.18 \pm 18.43 = (-22.61, 14.25)$. We would have to say that the population mean difference has not been precisely estimated. The bound on the error of estimation is quite large. In addition, the confidence interval spans zero. Because of this, we have insufficient evidence to claim that the population mean penetration differs for the two types of bearings.

73. Since we can assume that the distributions from which the samples were taken are normal, we use the two-sample t test. Let m_1 denote the true mean headability rating for aluminum killed steel specimens and m_2 denote the true mean headability rating for silicon killed steel. Then the hypotheses are $H_0: m_1 - m_2 = 0$ vs. $H_a: m_1 - m_2 \neq 0$. The test statistic is

$t = \dfrac{-.66}{\sqrt{.03888 + .047203}} = \dfrac{-.66}{\sqrt{.086083}} = -2.25$. The approximate degrees of freedom

$n = \dfrac{(.086083)^2}{\dfrac{(.03888)^2}{29} + \dfrac{(.047203)^2}{29}} = 57.5$, so we use 57. The two-tailed p-value

$\approx 2(.014) = .028$, which is less than the specified significance level, so we would reject H_o. The data supports the article's authors' claim.

74. Let m_1 denote the true average tear length for Brand A and let m_2 denote the true average tear length for Brand B. The relevant hypotheses are $H_0: m_1 - m_2 = 0$ vs. $H_a: m_1 - m_2 > 0$. Assuming both populations have normal distributions, the two-sample t test is appropriate. m = 16, $\bar{x} = 74.0$, s_1 = 14.8, n = 14, $\bar{y} = 61.0$, s_2 = 12.5, so the

approximate d.f. is $n = \dfrac{\left(\frac{14.8^2}{16} + \frac{12.5^2}{14}\right)^2}{\dfrac{\left(\frac{14.8^2}{16}\right)^2}{15} + \dfrac{\left(\frac{12.5^2}{14}\right)^2}{13}} = 27.97$, which we round down to 27. The test

statistic is $t = \dfrac{74.0 - 61.0}{\sqrt{\frac{14.8^2}{16} + \frac{12.5^2}{14}}} \approx 2.6$. From Table A.7, the p-value = P(t > 2.6) = .007. At a significance level of .05, H_o is rejected and we conclude that the average tear length for Brand A is larger than that of Brand B.

75.

a. The relevant hypotheses are $H_0: m_1 - m_2 = 0$ vs. $H_a: m_1 - m_2 \neq 0$. Assuming both populations have normal distributions, the two-sample t test is appropriate. m = 11, $\bar{x} = 98.1$, s_1 = 14.2, n = 15, $\bar{y} = 129.2$, s_2 = 39.1. The test statistic is

$t = \dfrac{-31.1}{\sqrt{18.3309 + 101.9207}} = \dfrac{-31.1}{\sqrt{120.252}} = -2.84$. The approximate degrees of

freedom $n = \dfrac{(120.252)^2}{\dfrac{(18.3309)^2}{10} + \dfrac{(101.9207)^2}{14}} = 18.64$, so we use 18. From Table A.7,

the two-tailed p-value $\approx 2(.006) = .012$. No, obviously, the results are different.

b. For the hypotheses $H_0: m_1 - m_2 = -25$ vs. $H_a: m_1 - m_2 < -25$, the test statistic changes to $t = \frac{-31.1 - (-25)}{\sqrt{120.252}} = -.556$. With degrees of freedom 18, the p-value $\approx P(t < -.6) = .278$. Since the p-value is greater than any sensible choice of a, we fail to reject H_o. There is insufficient evidence that the true average strength for males exceeds that for females by more than 25N.

76.

a. The relevant hypotheses are $H_0: m_1^* - m_2^* = 0$ (which is equivalent to saying $m_1 - m_2 = 0$) versus $H_a: m_1^* - m_2^* \neq 0$ (which is the same as saying $m_1 - m_2 \neq 0$). The pooled t test is based on d.f. = m + n – 2 = 8 + 9 – 2 = 15. The pooled variance is $s_p^2 = \left(\frac{m-1}{m+n-2}\right)s_1^2 + \left(\frac{n-1}{m+n-2}\right)s_2^2$

$\left(\frac{8-1}{8+9-2}\right)(4.9)^2 + \left(\frac{9-1}{8+9-2}\right)(4.6)^2 = 22.49$, so $s_p = 4.742$. The test statistic is $t = \frac{\bar{x}^* - \bar{y}^*}{s_p\sqrt{\frac{1}{m} + \frac{1}{n}}} = \frac{18.0 - 11.0}{4.742\sqrt{\frac{1}{8} + \frac{1}{9}}} = 3.04 \approx 3.0$. From Table A.7, the p-value associated with t = 3.0 is 2P(t > 3.0) = 2(.004) = .008. At significance level .05, H_o is rejected and we conclude that there is a difference between m_1^* and m_2^*, which is equivalent to saying that there is a difference between m_1 and m_2.

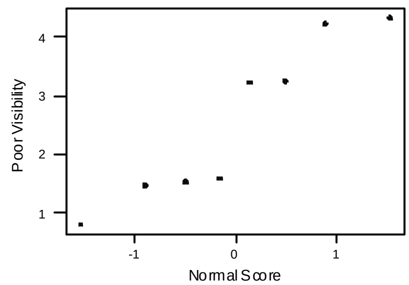
b. No. The mean of a lognormal distribution is $m = e^{m^* + (s^*)^2/2}$, where m^* and s^* are the parameters of the lognormal distribution (i.e., the mean and standard deviation of ln(x)). So when $s_1^* = s_2^*$, then $m_1^* = m_2^*$ would imply that $m_1 = m_2$. However, when $s_1^* \neq s_2^*$, then even if $m_1^* = m_2^*$, the two means m_1 and m_2 (given by the formula above) would not be equal.

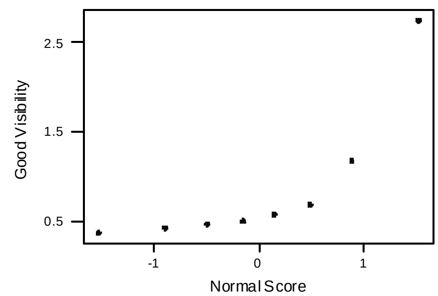
77. This is paired data, so the paired t test is employed. The relevant hypotheses are $H_0: m_d = 0$ vs. $H_a: m_d < 0$, where m_d denotes the difference between the population average control strength minus the population average heated strength. The observed differences (control – heated) are: -.06, .01, -.02, 0, and -.05. The sample mean and standard deviation of the differences are $\bar{d} = -.024$ and $s_d = .0305$. The test statistic is

$t = \frac{-.024}{.0305/\sqrt{5}} = -1.76 \approx -1.8$. From Table A.7, with d.f. = 5 – 1 = 4, the lower tailed p-value associated with t = -1.8 is P(t < -1.8) = P(t > 1.8) = .073. At significance level .05, H_o should not be rejected. Therefore, this data does not show that the heated average strength exceeds the average strength for the control population.

78. Let m_1 denote the true average ratio for young men and m_2 denote the true average ratio for elderly men. Assuming both populations from which these samples were taken are normally distributed, the relevant hypotheses are $H_0 : m_1 - m_2 = 0$ vs. $H_a : m_1 - m_2 > 0$. The value of the test statistic is $t = \dfrac{(7.47 - 6.71)}{\sqrt{\dfrac{(.22)^2}{13} + \dfrac{(.28)^2}{12}}} = 7.5$. The d.f. = 20 and the p-value is P(t > 7.5) ≈ 0. Since the p-value is $< a = .05$, we reject H$_o$. We have sufficient evidence to claim that the true average ratio for young men exceeds that for elderly men.

79.





A normal probability plot indicates the data for good visibility does not follow a normal distribution, thus a t-test is not appropriate for this small a sample size.

80. The relevant hypotheses would be $m_M = m_F$ versus $m_M \neq m_F$ for both the distress and delight indices. The reported p-value for the test of mean differences on the distress index was less than 0.001. This indicates a statistically significant difference in the mean scores, with the mean score for women being higher. The reported p-value for the test of mean differences on the delight index was > 0.05. This indicates a lack of statistical significance in the difference of delight index scores for men and women.

81. We wish to test H$_0$: $m_1 = m_2$ versus H$_a$: $m_1 \neq m_2$

Unpooled:

With H$_o$: $m_1 - m_2 = 0$ vs. H$_a$: $m_1 - m_2 \neq 0$, we will reject H$_o$ if $p - value < a$.

$$n = \frac{\left(\frac{.79^2}{14} + \frac{1.52^2}{12}\right)^2}{\frac{\left(\frac{.79^2}{14}\right)^2}{13} + \frac{\left(\frac{1.52^2}{12}\right)^2}{11}} = 15.95 \approx 16$$, and the test statistic

$$t = \frac{8.48 - 9.36}{\sqrt{\frac{.79^2}{14} + \frac{1.52^2}{12}}} = \frac{-.96}{.4869} = -1.97$$ leads to a p-value of 2[P(t > 1.97)]

$$\approx 2(.031) \approx .062$$

Pooled:

The degrees of freedom $n = m = n - 2 = 14 + 12 - 2 = 24$ and the pooled variance is $\left(\frac{13}{24}\right)(.79)^2 + \left(\frac{11}{24}\right)(1.52)^2 = 1.3970$, so $s_p = 1.181$. The test statistic is

$$t = \frac{-.96}{1.181\sqrt{\frac{1}{14} + \frac{1}{12}}} = \frac{-.96}{.465} \approx -2.1$$. The p-value = 2[P(t_{24} > 2.1)] = 2(.023) = .046.

With the pooled method, there are more degrees of freedom, and the p-value is smaller than with the unpooled method.

82. Because of the nature of the data, we will use a paired t test. We obtain the differences by subtracting intake value from expenditure value. We are testing the hypotheses H$_0$: $\mu_d = 0$ vs H$_a$: μ_d ? 0. Test statistic $t = \frac{1.757}{1.197/\sqrt{7}} = 3.88$ with df = n – 1 = 6 leads to a p-value of 2[P(t > 3.88) ˜ .004. Using either significance level .05 or .01, we would reject the null hypothesis and conclude that there is a difference between average intake and expenditure. However, at significance level .001, we would not reject.

83.

a. With n denoting the second sample size, the first is m = 3n. We then wish

$$20 = 2(2.58)\sqrt{\frac{900}{3n} + \frac{400}{n}}$$, which yields n = 47, m = 141.

b. We wish to find the n which minimizes $2(z_{a/2})\sqrt{\frac{900}{400-n} + \frac{400}{n}}$, or equivalently, the n which minimizes $\frac{900}{400-n} + \frac{400}{n}$. Taking the derivative with respect to n and equating to 0 yields $900(400-n)^{-2} - 400n^{-2} = 0$, whence $9n^2 = 4(400-n)^2$, or $5n^2 + 3200n - 640{,}000 = 0$. This yields n = 160, m = 400 – n = 240.

84. Let p_1 = true survival rate at $11^{o}C$; p_2 = true survival rate at $30^{o}C$; The hypotheses are $H_0 : p_1 - p_2 = 0$ vs. $H_a : p_1 - p_2 \neq 0$. The test statistic is $z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{m}+\frac{1}{n}\right)}}$. With $\hat{p}_1 = \frac{73}{91} = .802$, and $\hat{p}_2 = \frac{102}{110} = .927$, $\hat{p} = \frac{175}{201} = .871$, $\hat{q} = .129$.
$z = \frac{.802 - .927}{\sqrt{(.871)(.129)\left(\frac{1}{91}+\frac{1}{110}\right)}} = \frac{-.125}{.0320} = -3.91$. The p-value =
$\Phi(-3.91) < \Phi(-3.49) = .0003$, so reject H_o at any reasonable level. The two survival rates appear to differ.

85.

a. We test $H_0 : m_1 - m_2 = 0$ vs. $H_a : m_1 - m_2 \neq 0$. Assuming both populations have normal distributions, the two-sample t test is appropriate. The approximate degrees of freedom $n = \frac{(.042721)^2}{\frac{(.0325125)^2}{7} + \frac{(.0102083)^2}{11}} = 11.4$, so we use df = 11.
$t_{.0005,11} = 4.437$, so we reject H_o if $t \geq 4.437$ or $t \leq -4.437$ The test statistic is $t = \frac{.68}{\sqrt{.042721}} \approx 3.3$, which is not ≥ 4.437, so we cannot reject H_o. At significance level .001, the data does not indicate a difference in true average insulin-binding capacity due to the dosage level.

b. P-value = 2P(t > 3.3) = 2 (.004) = .008 which is > .001.

86. $$\hat{s}^2 = \frac{\left[(n_1-1)S_1^2 + (n_2-1)S_2^2 + (n_3-1)S_3^2 + (n_4-1)S_4^2\right]}{n_1+n_2+n_3+n_4-4}$$
$$E(\hat{s}^2) = \frac{\left[(n_1-1)s_1^2 + (n_2-1)s_2^2 + (n_3-1)s_3^2 + (n_4-1)s_4^2\right]}{n_1+n_2+n_3+n_4-4} = s^2$$. The estimate for the given data is $= \frac{[15(.4096)+17(.6561)+7(.2601)+11(.1225)]}{50} = .409$

87. $\Delta_0 = 0$, $s_1 = s_2 = 10$, d = 1, $s = \sqrt{\frac{200}{n}} = \frac{14.142}{\sqrt{n}}$, so $b = \Phi\left(1.645 - \frac{\sqrt{n}}{14.142}\right)$, giving $b =$ .9015, .8264, .0294, and .0000 for n = 25, 100, 2500, and 10,000 respectively. If the $m_i's$ referred to true average IQ's resulting from two different conditions, $m_1 - m_2 = 1$ would have little practical significance, yet very large sample sizes would yield statistical significance in this situation.

88. $H_0 : m_1 - m_2 = 0$ is tested against $H_a : m_1 - m_2 \neq 0$ using the two-sample t test, rejecting H_o at level .05 if either $t \geq t_{.025,15} = 2.131$ or if $t \leq -2.131$. With $\bar{x} = 11.20$, $s_1 = 2.68$, $\bar{y} = 9.79$, $s_2 = 3.21$, and m = n = 8, s_p = 2.96, and t = .95, so H_o is not rejected. In the situation described, the effect of carpeting would be mixed up with any effects due to the different types of hospitals, so no separate assessment could be made. The experiment should have been designed so that a separate assessment could be obtained (e.g., a randomized block design).

89. $H_0 : p_1 = p_2$ will be rejected at level a in favor of $H_a : p_1 > p_2$ if either $z \geq z_{.05} = 1.645$. With $\hat{p}_1 = \frac{250}{2500} = .10$, $\hat{p}_2 = \frac{167}{2500} = .0668$, and $\hat{p} = .0834$, $z = \frac{.0332}{.0079} = 4.2$, so H_o is rejected. It appears that a response is more likely for a white name than for a black name.

90. The computed value of Z is $z = \frac{34 - 46}{\sqrt{34 + 46}} = -1.34$. A lower tailed test would be appropriate, so the p-value $= \Phi(-1.34) = .0901 > .05$, so we would not judge the drug to be effective.

91.

a. Let m_1 and m_2 denote the true average weights for operations 1 and 2, respectively. The relevant hypotheses are $H_0 : m_1 - m_2 = 0$ vs. $H_a : m_1 - m_2 \neq 0$. The value of the test statistic is

$$t = \frac{(1402.24 - 1419.63)}{\sqrt{\frac{(10.97)^2}{30} + \frac{(9.96)^2}{30}}} = \frac{-17.39}{\sqrt{4.011363 + 3.30672}} = \frac{-17.39}{\sqrt{7.318083}} = -6.43.$$

The d.f. $n = \dfrac{(7.318083)^2}{\frac{(4.011363)^2}{29} + \frac{(3.30672)^2}{29}} = 57.5$, so use df = 57. $t_{.025,57} \approx 2.000$, so we can reject H_o at level .05. The data indicates that there is a significant difference between the true mean weights of the packages for the two operations.

b. $H_0 : m_1 = 1400$ will be tested against $H_a : m_1 > 1400$ using a one-sample t test with test statistic $t = \dfrac{\bar{x} - 1400}{s_1/\sqrt{m}}$. With degrees of freedom = 29, we reject H_o if $t > t_{.05,29} = 1.699$. The test statistic value is $t = \dfrac{1402.24 - 1400}{10.97/\sqrt{30}} = \dfrac{2.24}{2.00} = 1.1$.

Because 1.1 < 1.699, H_o is not rejected. True average weight does not appear to exceed 1400.

92. $Var(\bar{X} - \bar{Y}) = \dfrac{l_1}{m} + \dfrac{l_2}{n}$ and $\hat{l}_1 = \bar{X}$, $\hat{l}_2 = \bar{Y}$, $\hat{l} = \dfrac{m\bar{X} + n\bar{Y}}{m+n}$, giving $Z = \dfrac{\bar{X} - \bar{Y}}{\sqrt{\frac{\hat{l}}{m} + \frac{\hat{l}}{n}}}$. With $\bar{x} = 1.616$ and $\bar{y} = 2.557$, z = -5.3 and p-value = $2(\Phi(-5.3)) < .0006$, so we would certainly reject $H_0 : l_1 = l_2$ in favor of $H_a : l_1 \neq l_2$.

93. $\hat{l}_1 = \bar{x} = 1.62$, $\hat{l}_2 = \bar{y} = 2.56$, $\sqrt{\dfrac{\hat{l}_1}{m} + \dfrac{\hat{l}_2}{n}} = 1.77$, and the confidence interval is

$$-.94 \pm (1.96)(1.77) = -.94 \pm .35 = (-1.29, -.59)$$

CHAPTER 10

Section 10.1

1.

a. H_o will be rejected if $f \ge F_{.05,4,15} = 3.06$ (since I – 1 = 4, and I (J – 1) = (5)(3) = 15).
The computed value of F is $f = \frac{MSTr}{MSE} = \frac{2673.3}{1094.2} = 2.44$. Since 2.44 is not ≥ 3.06, H_o is not rejected. The data does not indicate a difference in the mean tensile strengths of the different types of copper wires.

b. $F_{.05,4,15} = 3.06$ and $F_{.10,4,15} = 2.36$, and our computed value of 2.44 is between those values, it can be said that .05 < p-value < .10.

2.

Type of Box	$\bar{x}$	s
1	713.00	46.55
2	756.93	40.34
3	698.07	37.20
4	682.02	39.87

Grand mean = 712.51

$$MSTr = \frac{6}{4-1}\left[(713.00-712.51)^2 + (756.93-712.51)^2 + (698.07-712.51)^2 + (682.02-712.51)^2\right] = 6{,}223.0604$$

$$MSE = \frac{1}{4}\left[(46.55)^2 + (40.34)^2 + (37.20)^2 + (39.87)^2\right] = 1{,}691.9188$$

$$f = \frac{MSTr}{MSE} = \frac{6{,}223.0604}{1{,}691.9188} = 3.678$$

$$F_{.05,3,20} = 3.10$$

3.678 > 3.10, so reject H_o. There is a difference in compression strengths among the four box types.

3. With $\mu_i =$ true average lumen output for brand i bulbs, we wish to test $H_0 : \mu_1 = \mu_2 = \mu_3$ versus $H_a :$ at least two $\mu_i's$ are unequal.

$MSTr = \hat{\sigma}_B^2 = \frac{591.2}{2} = 295.60$, $MSE = \hat{\sigma}_W^2 = \frac{4773.3}{21} = 227.30$, so

$f = \frac{MSTr}{MSE} = \frac{295.60}{227.30} = 1.30$ For finding the p-value, we need degrees of freedom I – 1 = 2 and I (J – 1) = 21. In the 2nd row and 21st column of Table A.9, we see that $1.30 < F_{.10,2,21} = 2.57$, so the p-value > .10. Since .10 is not < .05 , we cannot reject H_o. There are no differences in the average lumen outputs among the three brands of bulbs.

4. $x_{\bullet\bullet} = IJ\bar{x}_{\bullet\bullet} = 32(5.19) = 166.08$, so $SST = 911.91 - \frac{(166.08)^2}{32} = 49.95$.

$SSTr = 8\left[(4.39-5.19)^2 + ... + (6.36-5.19)^2\right] = 20.38$, so

$SSE = 49.95 - 20.38 = 29.57$. Then $f = \frac{20.38/3}{29.57/28} = 6.43$. Since

$6.43 \geq F_{.05,2,28} = 2.95$, $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ is rejected at level .05. There are differences between at least two average flight times for the four treatments.

5. $\mu_i =$ true mean modulus of elasticity for grade i (i = 1, 2, 3). We test $H_0 : \mu_1 = \mu_2 = \mu_3$ vs. $H_a :$ at least two $\mu_i's$ are unequal. Reject H_o if $f \geq F_{.01,2,27} = 5.49$. The grand mean = 1.5367,

$$MSTr = \frac{10}{2}\left[(1.63-1.5367)^2 + (1.56-1.5367)^2 + (1.42-1.5367)^2\right] = .1143$$

$MSE = \frac{1}{3}\left[(.27)^2 + (.24)^2 + (.26)^2\right] = .0660$, $f = \frac{MSTr}{MSE} = \frac{.1143}{.0660} = 1.73$. Fail to reject H_o. The three grades do not appear to differ.

6.

Source	Df	SS	MS	F
Treatments	3	509.112	169.707	10.85
Error	36	563.134	15.643	
Total	39	1,072.256		

$F_{.01,3,36} \approx F_{.01,3,30} = 4.51$. The computed test statistic value of 10.85 exceeds 4.51, so reject H_o in favor of H_a: at least two of the four means differ.

7.

Source	Df	SS	MS	F
Treatments	3	75,081.72	25,027.24	1.70
Error	16	235,419.04	14,713.69	
Total	19	310,500.76		

The hypotheses are $H_0 : m_1 = m_2 = m_3 = m_4$ vs. H_a : at least two m_i's are unequal. $1.70 < F_{.10,3,16} = 2.46$, so p-value > .10, and we fail to reject H_o.

8. The summary quantities are $x_{1\bullet} = 2332.5$, $x_{2\bullet} = 2576.4$, $x_{3\bullet} = 2625.9$, $x_{4\bullet} = 2851.5$, $x_{5\bullet} = 3060.2$, $x_{\bullet\bullet} = 13,446.5$, so CF = 5,165,953.21, SST = 75,467.58, SSTr = 43,992.55, SSE = 31,475.03, $MSTr = \frac{43,992.55}{4} = 10,998.14$, $MSE = \frac{31,475.03}{30} = 1049.17$ and $f = \frac{10,998.14}{1049.17} = 10.48$. (These values should be displayed in an ANOVA table as requested.) Since $10.48 \geq F_{.01,4,30} = 4.02$, $H_0 : m_1 = m_2 = m_3 = m_4 = m_5$ is rejected. There are differences in the true average axial stiffness for the different plate lengths.

9. The summary quantities are $x_{1\bullet} = 34.3$, $x_{2\bullet} = 39.6$, $x_{3\bullet} = 33.0$, $x_{4\bullet} = 41.9$, $x_{\bullet\bullet} = 148.8$, $\Sigma\Sigma x_{ij}^2 = 946.68$, so $CF = \frac{(148.8)^2}{24} = 922.56$,

$SST = 946.68 - 922.56 = 24.12$,

$SSTr = \frac{(34.3)^2 + ... + (41.9)^2}{6} - 922.56 = 8.98$, $SSE = 24.12 - 8.98 = 15.14$.

Source	Df	SS	MS	F
Treatments	3	8.98	2.99	3.95
Error	20	15.14	.757	
Total	23	24.12		

Since $3.10 = F_{.05,3,20} < 3.95 < 4.94 = F_{.01,3,20}$, $.01 < p-value < .05$ and H_o is rejected at level .05.

10.

a. $E\left(\overline{X}_{\bullet\bullet}\right)=\dfrac{\Sigma E\left(\overline{X}_{i\bullet}\right)}{I}=\dfrac{\Sigma m_i}{I}=m$.

b. $E\left(\overline{X}_{i\bullet}^2\right)=Var\left(\overline{X}_{i\bullet}\right)+\left[E\left(\overline{X}_{i\bullet}\right)\right]^2=\dfrac{s^2}{J}+m_i^2$.

c. $E\left(\overline{X}_{\bullet\bullet}^2\right)=Var\left(\overline{X}_{\bullet\bullet}\right)+\left[E\left(\overline{X}_{\bullet\bullet}\right)\right]^2=\dfrac{s^2}{IJ}+m^2$.

d. $E(SSTr)=E\left[J\Sigma\overline{X}_{i\bullet}^2-IJ\overline{X}_{\bullet\bullet}^2\right]=J\sum\left(\dfrac{s^2}{J+m_i^2}\right)-IJ\left(\dfrac{s^2}{IJ+m^2}\right)$

$=Is^2+J\Sigma m_i^2-s^2-IJm^2=(I-1)s^2+J\Sigma(m_i-m)^2$, so

$E(MSTr)=\dfrac{E(SSTr)}{I-1}=E\left[J\Sigma\overline{X}_{i\bullet}^2-IJ\overline{X}_{\bullet\bullet}^2\right]=s^2+J\sum\dfrac{(m_i-m)^2}{I-1}$.

e. When H_o is true, $m_1=\ldots=m_i=m$, so $\Sigma(m_i-m)^2=0$ and $E(MSTr)=s^2$. When H_o is false, $\Sigma(m_i-m)^2>0$, so $E(MSTr)>s^2$ (on average, MSTr overestimates s^2).

Section 10.2

11. $Q_{.05,5,15}=4.37$, $w=4.37\sqrt{\dfrac{272.8}{4}}=36.09$.

3	1	4	2	5
437.5	462.0	469.3	512.8	532.1

The brands seem to divide into two groups: 1, 3, and 4; and 2 and 5; with no significant differences within each group but all between group differences are significant.

12.

3	1	4	2	5
437.5	462.0	469.3	512.8	532.1

Brands 2 and 5 do not differ significantly from one another, but both differ significantly from brands 1, 3, and 4. While brands 3 and 4 do differ significantly, there is not enough evident to indicate a significant difference between 1 and 3 or 1 and 4.

13.

3	1	4	2	5
427.5	462.0	469.3	502.8	532.1

Brand 1 does not differ significantly from 3 or 4, 2 does not differ significantly from 4 or 5, 3 does not differ significantly from1, 4 does not differ significantly from 1 or 2, 5 does not differ significantly from 2, but all other differences (e.g., 1 with 2 and 5, 2 with 3, etc.) do appear to be significant.

14. I = 4, J = 8, so $Q_{.05,4,28} \approx 3.87$, $w = 3.87\sqrt{\dfrac{1.06}{8}} = 1.41$.

1	2	3	4
4.39	4.52	5.49	6.36

Treatment 4 appears to differ significantly from both 1 and 2, but there are no other significant differences.

15. $Q_{.01,4,36} = 4.75$, $w = 4.75\sqrt{\dfrac{15.64}{10}} = 5.94$.

2	1	3	4
24.69	26.08	29.95	33.84

Treatment 4 appears to differ significantly from both 1 and 2, but there are no other significant differences.

16.

a. Since the largest standard deviation (s_4 = 44.51) is only slightly more than twice the smallest (s_3 = 20.83) it is plausible that the population variances are equal (see text p. 406).

b. The relevant hypotheses are $H_0 : m_1 = m_2 = m_3 = m_4 = m_5$ vs. $H_a :$ at least two m_i's differ. With the given f of 10.48 and associated p-value of 0.000, we can reject H_o and conclude that there is a difference in axial stiffness for the different plate lengths.

c.

4	6	8	10	12
333.21	368.06	375.13	407.36	437.17

There is no significant difference in the axial stiffness for lengths 4, 6, and 8, and for lengths 6, 8, and 10, yet 4 and 10 differ significantly. Length 12 differs from 4, 6, and 8, but does not differ from 10.

17. $q = \Sigma c_i m_i$ where $c_1 = c_2 = .5$ and $c_3 = -1$, so $\hat{q} = .5\bar{x}_{1\bullet} + .5\bar{x}_{2\bullet} - \bar{x}_{3\bullet} = -.396$ and $\Sigma c_i^2 = 1.50$. With $t_{.025,6} = 2.447$ and MSE = .03106, the CI is (from 10.5 on page 418)

$$-.396 \pm (2.447)\sqrt{\frac{(.03106)(1.50)}{3}} = -.396 \pm .305 = (-.701, -.091).$$

18.

a. Let $m_i =$ true average growth when hormone #*i* is applied. $H_0 : m_1 = ... = m_5$ will be rejected in favor of $H_a :$ at least two m_i's differ if $f \ge F_{.05,4,15} = 3.06$. With

$\frac{x_{\bullet\bullet}^2}{IJ} = \frac{(278)^2}{20} = 3864.20$ and $\Sigma\Sigma x_{ij}^2 = 4280$, SST = 415.80.

$\frac{\Sigma x_{i\bullet}^2}{J} = \frac{(51)^2 + (71)^2 + (70)^2 + (46)^2 + (40)^2}{4} = 4064.50$, so SSTr = 4064.50 – 3864.20 = 200.3, and SSE = 415.80 – 200.30 = 215.50. Thus

$MSTr = \frac{200.3}{4} = 50.075$, $MSE = \frac{215.5}{15} = 14.3667$, and

$f = \frac{50.075}{14.3667} = 3.49$. Because $3.49 \ge 3.06$, reject H_o. There appears to be a difference in the average growth with the application of the different growth hormones.

b. $Q_{.05,5,15} = 4.37$, $w = 4.37\sqrt{\frac{14.3667}{4}} = 8.28$. The sample means are, in increasing order, 10.00, 11.50, 12.75, 17.50, and 17.75. The most extreme difference is 17.75 – 10.00 = 7.75 which doesn't exceed 8.28, so no differences are judged significant. Tukey's method and the F test are at odds.

19. MSTr = 140, error d.f. = 12, so $f = \frac{140}{SSE/12} = \frac{1680}{SSE}$ and $F_{.05,2,12} = 3.89$.

$w = Q_{.05,3,12}\sqrt{\frac{MSE}{J}} = 3.77\sqrt{\frac{SSE}{60}} = .4867\sqrt{SSE}$. Thus we wish $\frac{1680}{SSE} > 3.89$ (significance of f) and $.4867\sqrt{SSE} > 10$ (= 20 – 10, the difference between the extreme $\bar{x}_{i\bullet}'s$ - so no significant differences are identified). These become $431.88 > SSE$ and $SSE > 422.16$, so SSE = 425 will work.

20. Now MSTr = 125, so $f = \frac{1500}{SSE}$, $w = .4867\sqrt{SSE}$ as before, and the inequalities become $385.60 > SSE$ and $SSE > 422.16$. Clearly no value of SSE can satisfy both inequalities.

21.

a. Grand mean = 222.167, MSTr = 38,015.1333, MSE = 1,681.8333, and f = 22.6. The hypotheses are $H_0 : m_1 = ... = m_6$ vs. H_a : at least two $m_i's$ differ . Reject H_o if $f \geq F_{.01,5,78}$ (but since there is no table value for $n_2 = 78$, use $f \geq F_{.01,5,60} = 3.34$) With $22.6 \geq 3.34$, we reject H_o. The data indicates there is a dependence on injection regimen.

b. Assume $t_{.005,78} \approx 2.645$

i) Confidence interval for $m_1 - \frac{1}{5}(m_2 + m_3 + m_4 + m_5 + m_6)$:

$$\Sigma c_i \bar{x}_i \pm t_{a/2,I(J-1)}\sqrt{\frac{MSE(\Sigma c_i^2)}{J}}$$

$$= -67.4 \pm (2.645)\sqrt{\frac{1,681.8333(1.2)}{14}} = (-99.16, -35.64).$$

ii) Confidence interval for $\frac{1}{4}(m_2 + m_3 + m_4 + m_5) - m_6$:

$$= 61.75 \pm (2.645)\sqrt{\frac{1,681.8333(1.25)}{14}} = (29.34, 94.16)$$

Section 10.3

22. Summary quantities are $x_{1\bullet} = 291.4$, $x_{2\bullet} = 221.6$, $x_{3\bullet} = 203.4$, $x_{4\bullet} = 227.5$, $x_{\bullet\bullet} = 943.9$, $CF = 49{,}497.07$, $\Sigma\Sigma x_{ij}^2 = 50{,}078.07$, from which $SST = 581$,

$$SSTr = \frac{(291.4)^2}{5} + \frac{(221.6)^2}{4} + \frac{(203.4)^2}{4} + \frac{(227.5)^2}{5} - 49{,}497.07$$

$= 49{,}953.57 - 49{,}497.07 = 456.50$, and $SSE = 124.50$. Thus

$MSTr = \frac{456.50}{3} = 152.17$, $MSE = \frac{124.50}{18-4} = 8.89$, and f = 17.12. Because $17.12 \ge F_{.05,3,14} = 3.34$, $H_0 : m_1 = ... = m_4$ is rejected at level .05. There is a difference in yield of tomatoes for the four different levels of salinity.

23. $J_1 = 5$, $J_2 = 4$, $J_3 = 4$, $J_4 = 5$, $\bar{x}_{1\bullet} = 58.28$, $\bar{x}_{2\bullet} = 55.40$, $\bar{x}_{3\bullet} = 50.85$, $\bar{x}_{4\bullet} = 45.50$, MSE = 8.89. With $W_{ij} = Q_{.05,4,14} \cdot \sqrt{\frac{MSE}{2}\left(\frac{1}{J_i} + \frac{1}{J_j}\right)} = 4.11\sqrt{\frac{8.89}{2}\left(\frac{1}{J_i} + \frac{1}{J_j}\right)}$,

$\bar{x}_{1\bullet} - \bar{x}_{2\bullet} \pm W_{12} = (2.88) \pm (5.81)$; $\bar{x}_{1\bullet} - \bar{x}_{3\bullet} \pm W_{13} = (7.43) \pm (5.81)*$;
$\bar{x}_{1\bullet} - \bar{x}_{4\bullet} \pm W_{14} = (12.78) \pm (5.48)*$; $\bar{x}_{2\bullet} - \bar{x}_{3\bullet} \pm W_{23} = (4.55) \pm (6.13)$;
$\bar{x}_{2\bullet} - \bar{x}_{4\bullet} \pm W_{24} = (9.90) \pm (5.81)*$; $\bar{x}_{3\bullet} - \bar{x}_{4\bullet} \pm W_{34} = (5.35) \pm (5.81)$;

*Indicates an interval that doesn't include zero, corresponding to m's that are judged significantly different.

4 3 2 1

This underscoring pattern does not have a very straightforward interpretation.

24.

Source	Df	SS	MS	F
Groups	3-1=2	152.18	76.09	5.56
Error	74-3=71	970.96	13.68	
Total	74-1=73	1123.14		

Since $5.56 \ge F_{.01,2,71} \approx 4.94$, reject $H_0 : m_1 = m_2 = m_3$ at level .01.

25.

a. The distributions of the polyunsaturated fat percentages for each of the four regimens must be normal with equal variances.

b. We have all the $\bar{X}_{i.}$'s , and we need the grand mean:

$$\bar{X}_{..} = \frac{8(43.0)+13(42.4)+17(43.1)+14(43.5)}{52} = \frac{2236.9}{52} = 43.017$$

$$SSTr = \sum J_i(\bar{x}_{i.} - \bar{x}_{..})^2 = 8(43.0-43.017)^2 + 13(42.4-43.017)^2$$
$$+17(43.1-43.017)^2 + 13(43.5-43.017)^2 = 8.334$$

and $MSTr = \dfrac{8.334}{3} = 2.778$

$SSTr = \sum (J_i - 1)s^2 = 7(1.5)^2 + 12(1.3)^2 + 16(1.2)^2 + 13(1.2)^2 = 77.79$ and

$MSE = \dfrac{77.79}{48} = 1.621$. Then $f = \dfrac{MSTr}{MSE} = \dfrac{2.778}{1.621} = 1.714$ Since $1.714 < F_{.10,3,50} = 2.20$, we can say that the p-value is > .10. We do not reject the null hypothesis at significance level .10 (or any smaller), so we conclude that the data suggests no difference in the percentages for the different regimens.

26.

a.

i:	1	2	3	4	5	6	
J_I:	4	5	4	4	5	4	
$x_{i\bullet}$:	56.4	64.0	55.3	52.4	85.7	72.4	$x_{\bullet\bullet} = 386.2$
$\bar{x}_{i\bullet}$:	14.10	12.80	13.83	13.10	17.14	18.10	$\Sigma\Sigma x_j^2 = 5850.20$

Thus SST = 113.64, SSTr = 108.19, SSE = 5.45, MSTr = 21.64, MSE = .273, f = 79.3.

Since $79.3 \ge F_{.01,5,20} = 4.10$, $H_0 : m_1 = ... = m_6$ is rejected.

b. The modified Tukey intervals are as follows: (The first number is $\bar{x}_{i\bullet} - \bar{x}_{j\bullet}$ and the second is $W_{ij} = Q_{.01} \cdot \sqrt{\dfrac{MSE}{2}\left(\dfrac{1}{J_i} + \dfrac{1}{J_j}\right)}$.)

Pair	Interval	Pair	Interval	Pair	Interval
1,2	1.30 ± 1.37	2,3	-1.03 ± 1.37	3,5	-3.31 ± 1.37 *
1,3	$.27 \pm 1.44$	2,4	$-.30 \pm 1.37$	3,6	-4.27 ± 1.44 *
1,4	1.00 ± 1.44	2,5	-4.34 ± 1.29 *	4,5	-4.04 ± 1.37 *
1,5	-3.04 ± 1.37 *	2,6	-5.30 ± 1.37 *	4,6	-5.00 ± 1.44 *
1,6	-4.00 ± 1.44 *	3,4	$.37 \pm 1.44$	5,6	$-.96 \pm 1.37$

Asterisks identify pairs of means that are judged significantly different from one another.

c. The 99% t confidence interval is $\Sigma c_i \bar{x}_{i\bullet} \pm t_{.005,I(J-1)} \sqrt{\frac{MSE\left(\Sigma c_i^2\right)}{J_i}}$.

$\Sigma c_i \bar{x}_{i\bullet} = \frac{1}{4}\bar{x}_{1\bullet} + \frac{1}{4}\bar{x}_{2\bullet} + \frac{1}{4}\bar{x}_{3\bullet} + 14\bar{x}_{4\bullet} - 12\bar{x}_{5\bullet} - \frac{1}{2}\bar{x}_{6\bullet} = -4.16$, $\frac{\left(\Sigma c_i^2\right)}{J_i} = .1719$,

MSE = .273, $t_{.005,20} = 2.845$. The resulting interval is

$-4.16 \pm (2.845)\sqrt{(.273)(.1719)} = -4.16 \pm .62 = (-4.78, -3.54)$. The interval in the answer section is a Scheffe' interval, and is substantially wider than the t interval.

27.

a. Let $m_i =$ true average folacin content for specimens of brand I. The hypotheses to be tested are $H_0 : m_1 = m_2 = m_3 = m_4$ vs. $H_a :$ at least two $m_i's$ differ.

$\Sigma\Sigma x_{ij}^2 = 1246.88$ and $\frac{x_{\bullet\bullet}^2}{n} = \frac{(168.4)^2}{24} = 1181.61$, so SST = 65.27.

$\frac{\Sigma x_{i\bullet}^2}{J_i} = \frac{(57.9)^2}{7} + \frac{(37.5)^2}{5} + \frac{(38.1)^2}{6} + \frac{(34.9)^2}{6} = 1205.10$, so

$SSTr = 1205.10 - 1181.61 = 23.49$.

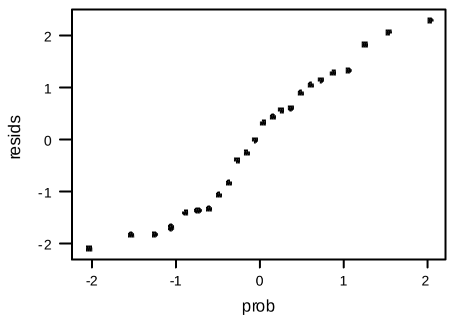
Source	Df	SS	MS	F
Treatments	3	23.49	7.83	3.75
Error	20	41.78	2.09	
Total	23	65.27		

With numerator df = 3 and denominator = 20,

$F_{.05,3,20} = 3.10 < 3.75 < F_{.01,3,20} = 4.94$, so $.01 < p-value < .05$, and since the p-value < .05, we reject H_o. At least one of the pairs of brands of green tea has different average folacin content.

b. With $\bar{x}_{i\bullet} = 8.27, 7.50, 6.35,$ and 5.82 for I = 1, 2, 3, 4, we calculate the residuals $x_{ij} - \bar{x}_{i\bullet}$ for all observations. A normal probability plot appears below, and indicates that the distribution of residuals could be normal, so the normality assumption is plausible.

Normal Probability Plot for ANOVA Residuals



c. $Q_{.05,4,20} = 3.96$ and $W_{ij} = 3.96 \cdot \sqrt{\dfrac{2.09}{2}\left(\dfrac{1}{J_i} + \dfrac{1}{J_j}\right)}$, so the Modified Tukey intervals are:

Pair	Interval	Pair	Interval
1,2	$.77 \pm 2.37$	2,3	1.15 ± 2.45
1,3	1.92 ± 2.25	2,4	1.68 ± 2.45
1,4	2.45 ± 2.25 *	3,4	$.53 \pm 2.34$

4 3 2 1

Only Brands 1 and 4 are different from each other.

28.
$$SSTr = \sum_i \left\{ \sum_j \left(\bar{X}_{i\bullet} - \bar{X}_{\bullet\bullet} \right)^2 \right\} = \sum_i J_i \left(\bar{X}_{i\bullet} - \bar{X}_{\bullet\bullet} \right)^2 = \sum_i J_i \bar{X}_{i\bullet}^2 - 2\bar{X}_{\bullet\bullet} \sum_i J_i \bar{X}_{i\bullet} + \bar{X}_{\bullet\bullet}^2 \sum_i J_i$$
$$= \sum_i J_i \bar{X}_{i\bullet}^2 - 2\bar{X}_{\bullet\bullet} X_{\bullet\bullet} + n\bar{X}_{\bullet\bullet}^2 = \sum_i J_i \bar{X}_{i\bullet}^2 - 2n\bar{X}_{\bullet\bullet}^2 + n\bar{X}_{\bullet\bullet}^2 = \sum_i J_i \bar{X}_{i\bullet}^2 - n\bar{X}_{\bullet\bullet}^2.$$

29. $E(SSTr) = E\left(\sum_i J_i \overline{X}_{i\bullet}^2 - n\overline{X}_{\bullet\bullet}^2\right) = \Sigma J_i E\left(\overline{X}_{i\bullet}^2\right) - nE\left(\overline{X}_{\bullet\bullet}^2\right)$

$= \Sigma J_i \left[Var\left(\overline{X}_{i\bullet}\right) + \left(E\left(\overline{X}_{i\bullet}\right)\right)^2\right] - n\left[Var\left(\overline{X}_{\bullet\bullet}\right) + \left(E\left(\overline{X}_{\bullet\bullet}\right)\right)^2\right]$

$= \Sigma J_i \left[\frac{s^2}{J_i} + m_i^2\right] - n\left[\frac{s^2}{n} + \frac{(\Sigma J_i m_i)^2}{n}\right]$

$= (I-1)s^2 + \Sigma J_i (m + a_i)^2 - [\Sigma J_i (m + a_i)]^2$

$= (I-1)s^2 + \Sigma J_i m^2 + 2m\Sigma J_i a_i + \Sigma J_i a_i^2 - [m\Sigma J_i]^2 \quad = (I-1)s^2 + \Sigma J_i a_i^2$, from which E(MSTr) is obtained through division by $(I-1)$.

30.

a. $a_1 = a_2 = 0$, $a_3 = -1$, $a_4 = 1$, so $\Phi^2 = \frac{2\left(0^2 + 0^2 + (-1)^2 + 1^2\right)}{1} = 4,\ \Phi = 2,$ and from figure (10.5), power $\approx .90$.

b. $\Phi^2 = .5J$, so $\Phi = .707\sqrt{J}$ and $n_2 = 4(J-1)$. By inspection of figure (10.5), J = 9 looks to be sufficient.

c. $m_1 = m_2 = m_3 = m_4, m_5 = m_1 + 1$, so $m = m_1 + \frac{1}{5}$, $a_1 = a_2 = a_3 = a_4 = -\frac{1}{5}$, $a_4 = \frac{4}{5}$, $\Phi^2 = \frac{2\left(20/25\right)}{1} = 1.60$ $\Phi = 1.26$, $n_1 = 4$, $n_2 = 45$. By inspection of figure (10.6), power $\approx .55$.

31. With $s = 1$ (any other s would yield the same Φ), $a_1 = -1,\ a_2 = a_3 = 0,\ a_4 = 1,$

$\Phi^2 = \frac{.25\left(5(-1)^2 + 5(0)^2 + 5(0)^2 + 5(1)^2\right)}{1} = 2.5$, $\Phi = 1.58$, $n_1 = 3$, $n_2 = 14$, and power $\approx .62$.

32. With Poisson data, the ANOVA should be done using $y_{ij} = \sqrt{x_{ij}}$. This gives $y_{1\bullet} = 15.43$, $y_{2\bullet} = 17.15$, $y_{3\bullet} = 19.12$, $y_{4\bullet} = 20.01$, $y_{\bullet\bullet} = 71.71$, $\Sigma\Sigma y_{ij}^2 = 263.79$, CF = 257.12, SST = 6.67, SSTr = 2.52, SSE = 4.15, MSTr = .84, MSE = .26, f = 3.23. Since $F_{.01,3,16} = 5.29$, H_o cannot be rejected. The expected number of flaws per reel does not seem to depend upon the brand of tape.

33. $g(x) = x\left(1 - \frac{x}{n}\right) = nu(1-u)$ where $u = \frac{x}{n}$, so $h(x) = \int [u(1-u)]^{-1/2} du$. From a table of integrals, this gives $h(x) = \arcsin\left(\sqrt{u}\right) = \arcsin\left(\sqrt{\frac{x}{n}}\right)$ as the appropriate transformation.

34. $E(MSTr) = s^{2} + \frac{1}{I-1}\left(n - \frac{IJ^2}{n}\right)s_A^{2} = s^{2} + \frac{n-J}{I-1}s_A^{2} = s^{2} + Js_A^{2}$

Supplementary Exercises

35.

a. $H_0 : m_1 = m_2 = m_3 = m_4$ vs. H_a : at least two m_i's differ ; 3.68 is not $\geq F_{.01,3,20} = 4.94$, thus fail to reject H_o. The means do not appear to differ.

b. We reject H_o when the p-value < alpha. Since .029 is not < .01, we still fail to reject H_o.

36.

a. $H_0 : m_1 = \ldots = m_5$ will be rejected in favor of H_a : at least two m_i's differ if $f \geq F_{.05,4,40} = 2.61$. With $\bar{x}_{\bullet\bullet} = 30.82$, straightforward calculation yields $MSTr = \frac{221.112}{4} = 55.278$, $MSE = \frac{80.4591}{5} = 16.1098$, and $f = \frac{55.278}{16.1098} = 3.43$. Because $3.43 \geq 2.61$, H_o is rejected. There is a difference among the five teaching methods with respect to true mean exam score.

b. The format of this test is identical to that of part **a**. The calculated test statistic is $f = \frac{33.12}{20.109} = 1.65$. Since $1.65 < 2.61$, H_o is not rejected. The data suggests that with respect to true average retention scores, the five methods are not different from one another.

37. Let $m_i =$ true average amount of motor vibration for each of five bearing brands. Then the hypotheses are $H_0 : m_1 = ... = m_5$ vs. $H_a :$ at least two m_i's differ. The ANOVA table follows:

Source	Df	SS	MS	F
Treatments	4	30.855	7.714	8.44
Error	25	22.838	0.914	
Total	29	53.694		

$8.44 > F_{.001,4,25} = 6.49$, so p-value < .001, which is also < .05, so we reject H$_o$. At least two of the means differ from one another. The Tukey multiple comparison is appropriate. $Q_{.05,5,25} = 4.15$ (from Minitab output. Using Table A.10, approximate with $Q_{.05,5,24} = 4.17$). $W_{ij} = 4.15\sqrt{.914/6} = 1.620$.

Pair	$\bar{x}_{i\bullet} - \bar{x}_{j\bullet}$	Pair	$\bar{x}_{i\bullet} - \bar{x}_{j\bullet}$
1,2	-2.267*	2,4	1.217
1,3	0.016	2,5	2.867*
1,4	-1.050	3,4	-1.066
1,5	0.600	3,5	0.584
2,3	2.283*	4,5	1.650*

*Indicates significant pairs.

5 3 1 4 2

38. $x_{1\bullet} = 15.48$, $x_{2\bullet} = 15.78$, $x_{3\bullet} = 12.78$, $x_{4\bullet} = 14.46$, $x_{5\bullet} = 14.94$ $x_{\bullet\bullet} = 73.44$, so $CF = 179.78$, SST = 3.62, SSTr = 180.71 – 179.78 = .93, SSE = 3.62 - .93 = 2.69.

Source	Df	SS	MS	F
Treatments	4	.93	.233	2.16
Error	25	2.69	.108	
Total	29	3.62		

$F_{.05,4,25} = 2.76$. Since 2.16 is not ≥ 2.76, do not reject H$_o$ at level .05.

39. $\hat{q} = 2.58 - \dfrac{2.63 + 2.13 + 2.41 + 2.49}{4} = .165$, $t_{.025,25} = 2.060$, MSE = .108, and $\Sigma c_i^2 = (1)^2 + (-.25)^2 + (-.25)^2 + (-.25)^2 + (-.25)^2 = 1.25$, so a 95% confidence interval for q is $.165 \pm 2.060\sqrt{\dfrac{(.108)(1.25)}{6}} = .165 \pm .309 = (-.144, .474)$. This interval does include zero, so 0 is a plausible value for q.

40. $m_1 = m_2 = m_3, m_4 = m_5 = m_1 - s$, so $m = m_1 - \frac{2}{5}s$, $a_1 = a_2 = a_3 = \frac{2}{5}s$, $a_4 = a_5 = -\frac{3}{5}s$. Then $\Phi^2 = \dfrac{J}{I}\sum \dfrac{a_i^2}{s^2}$

$$= \frac{6}{5}\left[\frac{3\left(\frac{2}{5}s\right)^2}{s^2} + \frac{2\left(-\frac{3}{5}s\right)^2}{s^2}\right] = 1.632$$ and $\Phi = 1.28$, $n_1 = 4$, $n_2 = 25$. By inspection of figure (10.6), power $\approx .48$, so $b \approx .52$.

41. This is a random effects situation. $H_0 : s_A^2 = 0$ states that variation in laboratories doesn't contribute to variation in percentage. H_o will be rejected in favor of H_a if $f \ge F_{.05,3,8} = 4.07$. SST = 86,078.9897 – 86,077.2224 = 1.7673, SSTr = 1.0559, and SSE = .7114. Thus $f = \dfrac{1.0559/3}{.7114/8} = 3.96$, which is not ≥ 4.07, so H_o cannot be rejected at level .05. Variation in laboratories does not appear to be present.

42.

a. $m_i =$ true average CFF for the three iris colors. Then the hypotheses are $H_0 : m_1 = m_2 = m_3$ vs. $H_a :$ at least two m_i's differ. SST = 13,659.67 – 13,598.36 = 61.31, $SSTR = \left(\dfrac{(204.7)^2}{8} + \dfrac{(134.6)^2}{5} + \dfrac{(169.0)^2}{6}\right) - 13{,}598.36 = 23.00$ The ANOVA table follows:

Source	Df	SS	MS	F
Treatments	2	23.00	11.50	4.803
Error	16	38.31	2.39	
Total	18	61.31		

Because $F_{.05,2,16} = 3.63 < 4.803 < F_{.01,2,16} = 6.23$, .01 < p-value < .05, so we reject H_o. There are differences in CFF based on iris color.

b. $Q_{.05,3,16} = 3.65$ and $W_{ij} = 3.65 \cdot \sqrt{\frac{2.39}{2}\left(\frac{1}{J_i} + \frac{1}{J_j}\right)}$, so the Modified Tukey intervals are:

Pair	$(\bar{x}_{i\bullet} - \bar{x}_{j\bullet}) \pm W_{ij}$
1,2	-1.33 ± 2.27
1,3	-2.58 ± 2.15 *
2,3	-1.25 ± 2.42

Brown	Green	Blue
25.59	26.92	28.17

The CFF is only significantly different for Brown and Blue iris color.

43. $\sqrt{(I-1)(MSE)(F_{.05,I-1,n-I})} = \sqrt{(2)(2.39)(3.63)} = 4.166$. For $m_1 - m_2$, $c_1 = 1$, $c_2 = -1$, and $c_3 = 0$, so $\sqrt{\sum \frac{c_i^2}{J_i}} = \sqrt{\frac{1}{8} + \frac{1}{5}} = .570$. Similarly, for $m_1 - m_3$, $\sqrt{\sum \frac{c_i^2}{J_i}} = \sqrt{\frac{1}{8} + \frac{1}{6}} = .540$; for $m_2 - m_3$, $\sqrt{\sum \frac{c_i^2}{J_i}} = \sqrt{\frac{1}{5} + \frac{1}{6}} = .606$, and for $.5m_2 + .5m_2 - m_3$, $\sqrt{\sum \frac{c_i^2}{J_i}} = \sqrt{\frac{.5^2}{8} + \frac{.5^2}{5} + \frac{(-1)^2}{6}} = .498$.

Contrast	Estimate	Interval
$m_1 - m_2$	25.59 – 26.92 = -1.33	$(-1.33) \pm (.570)(4.166) = (-3.70, 1.04)$
$m_1 - m_3$	25.59 – 28.17 = -2.58	$(-2.58) \pm (.540)(4.166) = (-4.83, -.33)$
$m_2 - m_3$	26.92 – 28.17 = -1.25	$(-1.25) \pm (.606)(4.166) = (-3.77, 1.27)$
$.5m_2 + .5m_2 - m_3$	-1.92	$(-1.92) \pm (.498)(4.166) = (-3.99, 0.15)$

The contrast between m_1 and m_3 since the calculated interval is the only one that does not contain the value (0).

44.

Source	Df	SS	MS	F	$F_{.05}$
Treatments	3	24,937.63	8312.54	1117.8	4.07
Error	8	59.49	7.44		
Total	11	24,997.12			

Because $1117.8 \geq 4.07$, $H_0 : m_1 = m_2 = m_3 = m_4$ is rejected. $Q_{.05,4,8} = 4.53$, so $w = 4.53\sqrt{\frac{7.44}{3}} = 7.13$. The four sample means are $\bar{x}_{4\bullet} = 29.92$, $\bar{x}_{1\bullet} = 33.96$, $\bar{x}_{3\bullet} = 115.84$, and $\bar{x}_{2\bullet} = 129.30$. Only $\bar{x}_{1\bullet} - \bar{x}_{4\bullet} < 7.13$, so all means are judged significantly different from one another except for m_4 and m_1 (corresponding to PCM and OCM).

45. $Y_{ij} - \bar{Y}_{\bullet\bullet} = c\left(X_{ij} - \bar{X}_{\bullet\bullet}\right)$ and $\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet} = c\left(\bar{X}_{i\bullet} - \bar{X}_{\bullet\bullet}\right)$, so each sum of squares involving Y will be the corresponding sum of squares involving X multiplied by c^2. Since F is a ratio of two sums of squares, c^2 appears in both the numerator and denominator so cancels, and F computed from Y_{ij}'s = F computed from X_{ij}'s.

46. The ordered residuals are –6.67, -5.67, -4, -2.67, -1, -1, 0, 0, 0, .33, .33, .33, 1, 1, 2.33, 4, 5.33, 6.33. The corresponding z percentiles are –1.91, -1.38, -1.09, -.86, -.67, -.51, -.36, -.21, -.07, .07, .21, .36, .51, .67, .86, 1.09, 1.38, and 1.91. The resulting plot of the respective pairs (the Normal Probability Plot) is reasonably straight, and thus there is no reason to doubt the normality assumption.

CHAPTER 11

Section 11.1

1.

a. $MSA = \frac{30.6}{4} = 7.65$, $MSE = \frac{59.2}{12} = 4.93$, $f_A = \frac{7.65}{4.93} = 1.55$. Since 1.55 is not $\geq F_{.05,4,12} = 3.26$, don't reject H_{oA}. There is no difference in true average tire lifetime due to different makes of cars.

b. $MSB = \frac{44.1}{3} = 14.70$, $f_B = \frac{14.70}{4.93} = 2.98$. Since 2.98 is not $\geq F_{.05,3,12} = 3.49$, don't reject H_{oB}. There is no difference in true average tire lifetime due to different brands of tires.

2.

a. $x_{1\bullet} = 163$, $x_{2\bullet} = 152$, $x_{3\bullet} = 142$, $x_{4\bullet} = 146$, $x_{\bullet 1} = 215$, $x_{\bullet 2} = 188$, $x_{\bullet 3} = 200$, $x_{\bullet\bullet} = 603$, $\Sigma\Sigma x_{ij}^2 = 30{,}599$, $CF = \frac{(603)^2}{12} = 30{,}300.75$, so SST = 298.25, $SSA = \frac{1}{3}\left[(163)^2 + (152)^2 + (142)^2 + (146)^2\right] - 30{,}300.75 = 83.58$,
$SSB = 30{,}392.25 - 30{,}300.75 = 91.50$,
$SSE = 298.25 - 83.58 - 91.50 = 123.17$.

Source	Df	SS	MS	F
A	3	83.58	27.86	1.36
B	2	91.50	45.75	2.23
Error	6	123.17	20.53	
Total	11	298.25		

$F_{.05,3,6} = 4.76$, $F_{.05,2,6} = 5.14$. Since neither f is greater than the appropriate critical value, neither H_{oA} nor H_{oB} is rejected.

b. $\hat{m} = \bar{x}_{\bullet\bullet} = 50.25$, $\hat{a}_1 = \bar{x}_{1\bullet} - \bar{x}_{\bullet\bullet} = 4.08$, $\hat{a}_2 = .42$, $\hat{a}_3 = -2.92$, $\hat{a}_4 = -1.58$, $\hat{b}_1 = \bar{x}_{\bullet 1} - \bar{x}_{\bullet\bullet} = 3.50$, $\hat{b}_2 = -3.25$, $\hat{b}_3 = -.25$.

3. $x_{1\bullet} = 927$, $x_{2\bullet} = 1301$, $x_{3\bullet} = 1764$, $x_{4\bullet} = 2453$, $x_{\bullet 1} = 1347$, $x_{\bullet 2} = 1529$, $x_{\bullet 3} = 1677$, $x_{\bullet 4} = 1892$, $x_{\bullet\bullet} = 6445$, $\Sigma\Sigma x_{ij}^2 = 2{,}969{,}375$,

$CF = \dfrac{(6445)^2}{16} = 2{,}596{,}126.56$, $SSA = 324{,}082.2$, $SSB = 39{,}934.2$,

$SST = 373{,}248.4$, $SSE = 9232.0$

a.

Source	Df	SS	MS	F
A	3	324,082.2	108,027.4	105.3
B	3	39,934.2	13,311.4	13.0
Error	9	9232.0	1025.8	
Total	15	373,248.4		

Since $F_{.01,3,9} = 6.99$, both H_{oA} and H_{oB} are rejected.

b. $Q_{.01,4,9} = 5.96$, $w = 5.96\sqrt{\dfrac{1025.8}{4}} = 95.4$

i:	1	2	3	4
$\bar{x}_{i\bullet}$:	231.75	325.25	441.00	613.25

All levels of Factor A (gas rate) differ significantly except for 1 and 2

c. $w = 95.4$, as in **b**

i:	1	2	3	4
$\bar{x}_{\bullet j}$:	336.75	382.25	419.25	473

Only levels 1 and 4 appear to differ significantly.

4.

a. After subtracting 400, $x_{1\bullet} = 151$, $x_{2\bullet} = 137$, $x_{3\bullet} = 125$, $x_{4\bullet} = 124$, $x_{\bullet 1} = 183$, $x_{\bullet 2} = 169$, $x_{\bullet 3} = 185$, $x_{\bullet\bullet} = 537$, $SSA = 159.98$, $SSB = 38.00$, $SST = 238.25$, $SSE = 40.67$.

Source	Df	SS	MS	f	$F_{.05}$
A	3	159.58	53.19	7.85	4.76
B	2	38.00	19.00	2.80	5.14
Error	6	40.67	6.78		
Total	11	238.25			

b. Since $7.85 \geq 4.76$, reject H_{oA}: $a_1 = a_2 = a_3 = a_4 = 0$: The amount of coverage depends on the paint brand.

c. Since 2.80 is not ≥ 5.14, do not reject H_{oA}: $b_1 = b_2 = b_3 = 0$. The amount of coverage does not depend on the roller brand.

d. Because H_{oB} was not rejected. Tukey's method is used only to identify differences in levels of factor A (brands of paint). $Q_{.05,4,6} = 4.90$, w = 7.37.

i:	4	3	2	1
$\bar{x}_{i\bullet}$:	41.3	41.7	45.7	50.3

Brand 1 differs significantly from all other brands.

5.

Source	Df	SS	MS	f
Angle	3	58.16	19.3867	2.5565
Connector	4	246.97	61.7425	8.1419
Error	12	91.00	7.5833	
Total	19	396.13		

$H_0 : a_1 = a_2 = a_3 = a_4 = 0$; H_a : at least one a_i is not zero.

$f_A = 2.5565 < F_{.01,3,12} = 5.95$, so fail to reject H_o. The data fails to indicate any effect due to the angle of pull.

6.

a. $MSA = \frac{11.7}{2} = 5.85$, $MSE = \frac{25.6}{8} = 3.20$, $f = \frac{5.85}{3.20} = 1.83$, which is not significant at level .05.

b. Otherwise extraneous variation associated with houses would tend to interfere with our ability to assess assessor effects. If there really was a difference between assessors, house variation might have hidden such a difference. Alternatively, an observed difference between assessors might have been due just to variation among houses and the manner in which assessors were allocated to homes.

7.

a. CF = 140,454, SST = 3476,

$$SSTr = \frac{(905)^2 + (913)^2 + (936)^2}{18} - 140{,}454 = 28.78,$$

$$SSBl = \frac{430{,}295}{3} - 140{,}454 = 2977.67$$, SSE = 469.55, MSTr = 14.39, MSE = 13.81, $f_{Tr} = 1.04$, which is clearly insignificant when compared to $F_{.05,2,51}$.

b. $f_{Bl} = 12.68$, which is significant, and suggests substantial variation among subjects. If we had not controlled for such variation, it might have affected the analysis and conclusions.

8.

a. $x_{1\bullet} = 4.34$, $x_{2\bullet} = 4.43$, $x_{3\bullet} = 8.53$, $x_{\bullet\bullet} = 17.30$, $SST = 3.8217$, $SSTr = 1.1458$, $SSBl = \frac{32.8906}{3} - 9.9763 = .9872$, $SSE = 1.6887$, $MSTr = .5729$, $MSE = .0938$, f = 6.1. Since $6.1 \geq F_{.05,2,18} = 3.55$, H_{oA} is rejected; there appears to be a difference between anesthetics.

b. $Q_{.05,3,18} = 3.61$, w = .35. $\bar{x}_{1\bullet} = .434$, $\bar{x}_{2\bullet} = .443$, $\bar{x}_{3\bullet} = .853$, so both anesthetic 1 and anesthetic 2 appear to be different from anesthetic 3 but not from one another.

9.

Source	Df	SS	MS	f
Treatment	3	81.1944	27.0648	22.36
Block	8	66.5000	8.3125	6.87
Error	24	29.0556	1.2106	
Total	35	176.7500		

$F_{.05,3,24} = 3.01$. Reject H_o. There is an effect due to treatments.

$Q_{.05,4,24} = 3.90$; $w = (3.90)\sqrt{\frac{1.2106}{9}} = 1.43$

1	4	3	2
8.56	9.22	10.78	12.44

(1 and 4 underlined together)

10.

Source	Df	SS	MS	f
Method	2	23.23	11.61	8.69
Batch	9	86.79	9.64	7.22
Error	18	24.04	1.34	
Total	29	134.07		

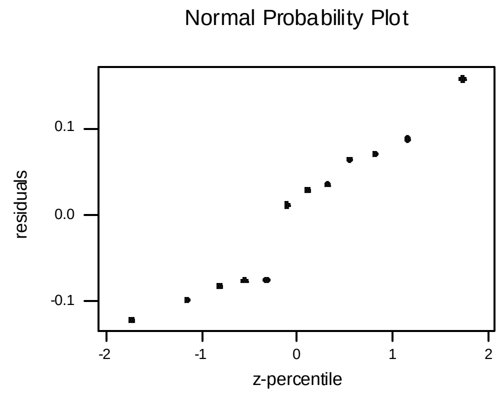
$F_{.01,2,18} = 6.01 < 8.69 < F_{.001,2,18} = 10.39$, so .001 < p-value < .01, which is significant. At least two of the curing methods produce differing average compressive strengths. (With p-value < .001, there are differences between batches as well.)

$Q_{.05,3,18} = 3.61$; $w = (3.61)\sqrt{\frac{1.34}{10}} = 1.32$

Method A	Method B	Method C
29.49	31.31	31.40

Methods B and C produce strengths that are not significantly different, but Method A produces strengths that are different (less) than those of both B and C.

11. The residual, percentile pairs are (-0.1225, -1.73), (-0.0992, -1.15), (-0.0825, -0.81), (-0.0758, -0.55), (-0.0750, -0.32), (0.0117, -0.10), (0.0283, 0.10), (0.0350, 0.32), (0.0642, 0.55), (0.0708, 0.81), (0.0875, 1.15), (0.1575, 1.73).



The pattern is sufficiently linear, so normality is plausible.

12. $MSB = \dfrac{113.5}{4} = 28.38$, $MSE = \dfrac{25.6}{8} = 3.20$, $f_B = 8.87$, $F_{.01,4,8} = 7.01$, and since $8.87 \geq 7.01$, we reject H_o and conclude that $s_B^2 > 0$.

13.

a. With $Y_{ij} = X_{ij} + d$, $\overline{Y}_{i\bullet} = \overline{X}_{i\bullet} + d$, $\overline{Y}_{\bullet j} = \overline{X}_{\bullet j} + d$, $\overline{Y}_{\bullet\bullet} = \overline{X}_{\bullet\bullet} + d$, so all quantities inside the parentheses in (11.5) remain unchanged when the Y quantities are substituted for the corresponding X's (e.g., $\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet} = \overline{X}_{i\bullet} - \overline{X}_{\bullet\bullet}$, etc.).

b. With $Y_{ij} = cX_{ij}$, each sum of squares for Y is the corresponding SS for X multiplied by c^2. However, when F ratios are formed the c^2 factors cancel, so all F ratios computed from Y are identical to those computed from X. If $Y_{ij} = cX_{ij} + d$, the conclusions reached from using the Y's will be identical to those reached using the X's.

14.

$$E\left(\overline{X}_{i\bullet} - \overline{X}_{\bullet\bullet}\right) = E\left(\overline{X}_{i\bullet}\right) - E\left(\overline{X}_{\bullet\bullet}\right) = \frac{1}{J}E\left(\sum_j X_{ij}\right) - \frac{1}{IJ}E\left(\sum_i\sum_j X_{ij}\right)$$

$$= \frac{1}{J}\sum_j\left(m + a_i + b_j\right) - \frac{1}{IJ}\sum_i\sum_j\left(m + a_i + b_j\right)$$

$$= m + a_i + \frac{1}{J}\sum_j b_j - m - \frac{1}{I}\sum_i a_i - \frac{1}{J}\sum_j b_j = a_i \text{, as desired.}$$

15.

a. $\Sigma a_i^2 = 24,$ so $\Phi^2 = \left(\frac{3}{4}\right)\left(\frac{24}{16}\right) = 1.125$, $\Phi = 1.06$, $n_1 = 3,\ n_2 = 6,$ and from figure 10.5, power $\approx .2$. For the second alternative, $\Phi = 1.59$, and power $\approx .43$.

b. $\Phi^2 = \left(\frac{1}{J}\right)\sum\frac{b_j^2}{s^2} = \left(\frac{4}{5}\right)\left(\frac{20}{16}\right) = 1.00$, so $\Phi = 1.00$, $n_1 = 4,\ n_2 = 12,$ and power $\approx .3$.

Section 11.2

16.

a.

Source	Df	SS	MS	f
A	2	30,763.0	15,381.50	3.79
B	3	34,185.6	11,395.20	2.81
AB	6	43,581.2	7263.53	1.79
Error	24	97,436.8	4059.87	
Total	35	205,966.6		

b. $f_{AB} = 1.79$ which is not ≥ $F_{.05,6,24} = 2.51$, so H_{oAB} cannot be rejected, and we conclude that no interaction is present.

c. $f_A = 3.79$ which is ≥ $F_{.05,2,24} = 3.40$, so H_{oA} is rejected at level .05.

d. $f_B = 2.81$ which is not ≥ $F_{.05,3,24} = 3.01$, so H_{oB} is not rejected.

e. $Q_{.05,3,24} = 3.53$, $w = 3.53\sqrt{\frac{4059.87}{12}} = 64.93$.

3	1	2
3960.02	4010.88	4029.10

Only times 2 and 3 yield significantly different strengths.

17.

a.

Source	Df	SS	MS	f	$F_{.05}$
Sand	2	705	352.5	3.76	4.26
Fiber	2	1,278	639.0	6.82*	4.26
Sand&Fiber	4	279	69.75	0.74	3.63
Error	9	843	93.67		
Total	17	3,105			

There appears to be an effect due to carbon fiber addition.

b.

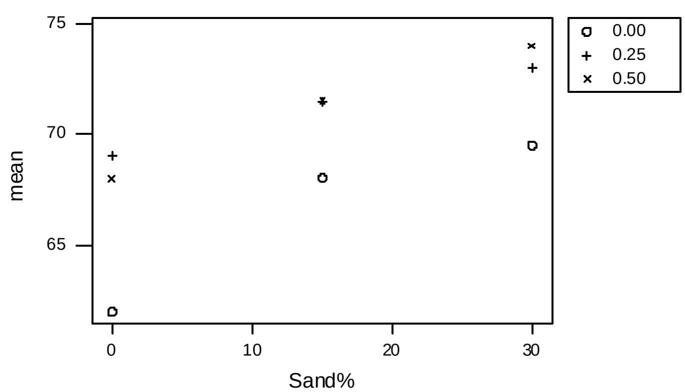
Source	Df	SS	MS	f	$F_{.05}$
Sand	2	106.78	53.39	6.54*	4.26
Fiber	2	87.11	43.56	5.33*	4.26
Sand&Fiber	4	8.89	2.22	.27	3.63
Error	9	73.50	8.17		
Total	17	276.28			

There appears to be an effect due to both sand and carbon fiber addition to casting hardness.

c.

Sand%	0	15	30	0	15	30	0	15	30
Fiber%	0	0	0	0.25	0.25	0.25	0.5	0.5	0.5
$\bar{x}$	62	68	69.5	69	71.5	73	68	71.5	74

The plot below indicates some effect due to sand and fiber addition with no significant interaction. This agrees with the statistical analysis in part **b**



18.

Source	Df	SS	MS	f	$F_{.05}$	$F_{.01}$
Formulation	1	2,253.44	2,253.44	376.2**	4.75	9.33
Speed	2	230.81	115.41	19.27**	3.89	6.93
Formulation & Speed	2	18.58	9.29	1.55	3.89	6.93
Error	12	71.87	5.99			
Total	17	2,574.7				

a. There appears to be no interaction between the two factors.

b. Both formulation and speed appear to have a highly statistically significant effect on yield.

c. Let formulation = Factor A and speed = Factor B.

For Factor A: $\hat{\mu}_{1\bullet} = 187.03 \quad \hat{\mu}_{2\bullet} = 164.66$

For Factor B: $\hat{\mu}_{\bullet 1} = 177.83 \quad \hat{\mu}_{\bullet 2} = 170.82 \quad \hat{\mu}_{\bullet 3} = 178.88$

For Interaction: $\hat{\mu}_{11} = 189.47 \quad \hat{\mu}_{12} = 180.6 \quad \hat{\mu}_{13} = 191.03$

$\hat{\mu}_{21} = 166.2 \quad \hat{\mu}_{22} = 161.03 \quad \hat{\mu}_{33} = 166.73$

overall mean: $\hat{\mu} = 175.84$

$\hat{\alpha}_i = \hat{\mu}_{i\bullet} - \hat{\mu}$: $\quad \hat{\alpha}_1 = 11.19 \quad \hat{\alpha}_2 = -11.18$

$\hat{\beta}_j = \hat{\mu}_{\bullet j} - \hat{\mu}$: $\quad \hat{\beta}_1 = 1.99 \quad \hat{\beta}_2 = -5.02 \quad \hat{\beta}_3 = 3.04$

$\hat{\gamma}_{ij} = \hat{\mu}_{ij} - \left(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j\right)$:

$\hat{\gamma}_{11} = .45 \quad \hat{\gamma}_{12} = -1.41 \quad \hat{\gamma}_{13} = .96$

$\hat{\gamma}_{21} = -.45 \quad \hat{\gamma}_{22} = 1.39 \quad \hat{\gamma}_{23} = -.97$

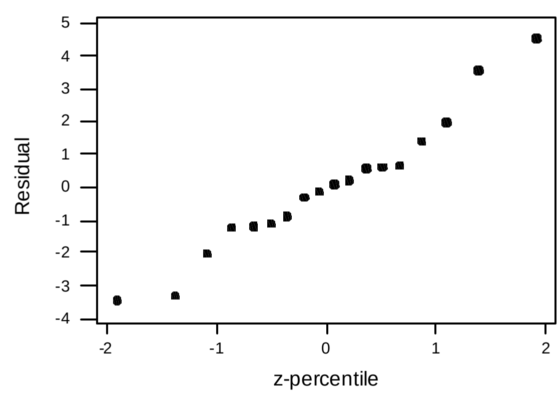
d.

Observed	Fitted	Residual	Observed	Fitted	Residual
189.7	189.47	0.23	161.7	161.03	0.67
188.6	189.47	-0.87	159.8	161.03	-1.23
190.1	189.47	0.63	161.6	161.03	0.57
165.1	166.2	-1.1	189.0	191.03	-2.03
165.9	166.2	-0.3	193.0	191.03	1.97
167.6	166.2	1.4	191.1	191.03	0.07
185.1	180.6	4.5	163.3	166.73	-3.43
179.4	180.6	-1.2	166.6	166.73	-0.13
177.3	180.6	-3.3	170.3	166.73	3.57

e.

i	**Residual**	**Percentile**	**z-percentile**
1	-3.43	2.778	-1.91
2	-3.30	8.333	-1.38
3	-2.03	13.889	-1.09
4	-1.23	19.444	-0.86
5	-1.20	25.000	-0.67
6	-1.10	30.556	-0.51
7	-0.87	36.111	-0.36
8	-0.30	41.667	-0.21
9	-0.13	47.222	-0.07
10	0.07	52.778	0.07
11	0.23	58.333	0.21
12	0.57	63.889	0.36
13	0.63	69.444	0.51
14	0.67	75.000	0.67
15	1.40	80.556	0.86
16	1.97	86.111	1.09
17	3.57	91.667	1.38
18	4.50	97.222	1.91

Normal Probability Plot of ANOVA Residuals



The residuals appear to be normally distributed.

19.

a.

$x_{ij\bullet}$	1	2	3	$x_{i\bullet\bullet}$
i = 1	16.44	17.27	16.10	49.81
i = 2	16.24	17.00	15.91	49.15
i = 3	16.80	17.37	16.20	50.37
$x_{\bullet j\bullet}$	49.48	51.64	48.21	$x_{\bullet\bullet\bullet} = 149.33$

(columns: j)

CF = 1238.8583

Thus SST = 1240.1525 – 1238.8583 = 1.2942,

$$SSE = 1240.1525 - \frac{2479.9991}{2} = .1530,$$

$$SSA = \frac{(49.81)^2 + (49.15)^2 + (50.37)^2}{6} - 1238.8583 = .1243, \; SSB = 1.0024$$

Source	**Df**	**SS**	**MS**	**f**	$\mathbf{F_{.01}}$
A	2	.1243	.0622	3.66	8.02
B	2	1.0024	.5012	29.48*	8.02
AB	4	.0145	.0036	.21	6.42
Error	9	.1530	.0170		
Total	17	1.2942			

H_{oAB} cannot be rejected, so no significant interaction; H_{oA} cannot be rejected, so varying levels of NaOH does not have a significant impact on total acidity; H_{oB} is rejected: type of coal does appear to affect total acidity.

b. $Q_{.01,3,9} = 5.43$, $w = 5.43\sqrt{\frac{.0170}{6}} = .289$

j:	3	1	2
$\bar{x}_{\bullet j\bullet}$	8.035	8.247	8.607

(8.035 and 8.247 are underlined together.)

Coal 2 is judged significantly different from both 1 and 3, but these latter two don't differ significantly from each other.

20. $x_{11\bullet} = 855$, $x_{12\bullet} = 905$, $x_{13\bullet} = 845$, $x_{21\bullet} = 705$, $x_{22\bullet} = 735$, $x_{23\bullet} = 675$, $x_{1\bullet\bullet} = 2605$, $x_{2\bullet\bullet} = 2115$, $x_{\bullet 1\bullet} = 1560$, $x_{\bullet 2\bullet} = 1640$, $x_{\bullet 3\bullet} = 1520$, $x_{\bullet\bullet\bullet} = 4720$, $\Sigma\Sigma\Sigma x_{ijk}^2 = 1{,}253{,}150$, CF = 1,237,688.89, $\Sigma\Sigma x_{ij\bullet}^2 = 3{,}756{,}950$, which yields the accompanying ANOVA table.

Source	Df	SS	MS	f	$F_{.01}$
A	1	13,338.89	13,338.89	192.09*	9.93
B	2	1244.44	622.22	8.96*	6.93
AB	2	44.45	22.23	.32	6.93
Error	12	833.33	69.44		
Total	17	15,461.11			

Clearly, $f_{AB} = .32$ is insignificant, so H_{oAB} is not rejected. Both H_{oA} and H_{oB} are both rejected, since they are both greater than the respective critical values. Both phosphor type and glass type significantly affect the current necessary to produce the desired level of brightness.

21.

a. $SST = 12{,}280{,}103 - \dfrac{(19{,}143)^2}{30} = 64{,}954.70$,

$SSE = 12{,}280{,}103 - \dfrac{(24{,}529{,}699)}{2} = 15{,}253.50$,

$SSA = \dfrac{122{,}380{,}901}{10} - \dfrac{(19{,}143)^2}{30} = 22{,}941.80$, $SSB = 22{,}765.53$,

$SSAB = 64{,}954.70 - [22{,}941.80 + 22{,}765.53 + 15{,}253.50] = 3993.87$

Source	Df	SS	MS	f
A	2	22,941.80	11,470.90	$\frac{11{,}470.90}{499.23} = 22.98$
B	4	22,765.53	5691.38	$\frac{5691.38}{499.23} = 11.40$
AB	8	3993.87	499.23	.49
Error	15	15,253.50	1016.90	
Total	29	64,954.70		

b. $f_{AB} = .49$ is clearly not significant. Since $22.98 \ge F_{.05,2,8} = 4.46$, H_{oA} is rejected; since $11.40 \ge F_{.05,4,8} = 3.84$, H_{oB} is also rejected. We conclude that the different cement factors affect flexural strength differently and that batch variability contributes to variation in flexural strength.

22. The relevant null hypotheses are $H_{0A}: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$; $H_{0B}: \sigma_B^2 = 0$; $H_{0AB}: \sigma_G^2 = 0$.

$$SST = 11{,}499{,}492 - \frac{(16{,}598)^2}{24} = 20{,}591.83,$$
$$SSE = 11{,}499{,}492 - \frac{(22{,}982{,}552)}{2} = 8216.0,$$
$$SSA = \left[\frac{(4112)^2 + (4227)^2 + (4122)^2 + (4137)^2}{6}\right] - \frac{(16{,}598)^2}{24} = 1387.5,$$
$$SSB = \left[\frac{(5413)^2 + (5621)^2 + (5564)^2}{8}\right] - \frac{(16{,}598)^2}{24} = 2888.08,$$
$$SSAB = 20{,}591.83 - [8216.0 + 1387.5 + 2888.08] = 8216.25$$

Source	Df	SS	MS	f	$F_{.05}$
A	3	1,387.5	462.5	$\frac{MSA}{MSAB} = .34$	4.76
B	2	2,888.08	1,444.04	$\frac{MSB}{MSAB} = 1.07$	5.14
AB	6	8,100.25	1,350.04	$\frac{MSAB}{MSE} = 1.97$	3.00
Error	12	8,216.0	684.67		
Total	23	20,591.83			

Interaction between brand and writing surface has no significant effect on the lifetime of the pen, and since neither f_A nor f_B is greater than its respective critical value, we can conclude that neither the surface nor the brand of pen has a significant effect on the writing lifetime.

23. Summary quantities include $x_{1\bullet\bullet} = 9410$, $x_{2\bullet\bullet} = 8835$, $x_{3\bullet\bullet} = 9234$, $x_{\bullet 1\bullet} = 5432$, $x_{\bullet 2\bullet} = 5684$, $x_{\bullet 3\bullet} = 5619$, $x_{\bullet 4\bullet} = 5567$, $x_{\bullet 3\bullet} = 5177$, $x_{\bullet\bullet\bullet} = 27{,}479$, $CF = 16{,}779{,}898.69$, $\Sigma x_{i\bullet\bullet}^2 = 251{,}872{,}081$, $\Sigma x_{\bullet j\bullet}^2 = 151{,}180{,}459$, resulting in the accompanying ANOVA table.

Source	Df	SS	MS	f
A	2	11,573.38	5786.69	$\frac{MSA}{MSAB} = 26.70$
B	4	17,930.09	4482.52	$\frac{MSB}{MSAB} = 20.68$
AB	8	1734.17	216.77	$\frac{MSAB}{MSE} = 1.38$
Error	30	4716.67	157.22	
Total	44	35,954.31		

Since $1.38 < F_{.01,8,30} = 3.17$, H_{oG} cannot be rejected, and we continue: $26.70 \ge F_{.01,2,8} = 8.65$, and $20.68 \ge F_{.01,4,8} = 7.01$, so both H_{oA} and H_{oB} are rejected. Both capping material and the different batches affect compressive strength of concrete cylinders.

24.

a. $$E\left(\overline{X}_{i..} - \overline{X}_{...}\right) = \frac{1}{JK}\sum_j\sum_k E\left(X_{ijk}\right) - \frac{1}{IJK}\sum_i\sum_j\sum_k E\left(X_{ijk}\right)$$
$$= \frac{1}{JK}\sum_j\sum_k\left(m + a_i + b_j + g_{ij}\right) - \frac{1}{IJK}\sum_i\sum_j\sum_k\left(m + a_i + b_j + g_{ij}\right) = m + a_i - m = a_i$$

b. $$E\left(\hat{g}_{ij}\right) = \frac{1}{K}\sum_k E\left(X_{ijk}\right) - \frac{1}{JK}\sum_j\sum_k E\left(X_{ijk}\right) - \frac{1}{IK}\sum_i\sum_k E\left(X_{ijk}\right) + \frac{1}{IJK}\sum_i\sum_j\sum_k E\left(X_{ijk}\right)$$
$$= m + a_i + b_j + g_{ij} - \left(m + a_i\right) - \left(m + b_j\right) + m = g_{ij}$$

25. With $q = a_i - a'_i$, $\hat{q} = \overline{X}_{i..} - \overline{X}_{i'..} = \frac{1}{JK}\sum_j \sum_k \left(X_{ijk} - X_{i'jk}\right)$, and since $i \neq i'$, $X_{ijk} and X_{i'jk}$ are independent for every j, k. Thus $Var\left(\hat{q}\right) = Var\left(\overline{X}_{i..}\right) + Var\left(\overline{X}_{i'..}\right) = \frac{s^2}{JK} + \frac{s^2}{JK} = \frac{2s^2}{JK}$ (because $Var\left(\overline{X}_{i..}\right) = Var\left(\overline{e}_{i..}\right)$ and $Var\left(e_{ijk}\right) = s^2$) so $\hat{s}_{\hat{q}} = \sqrt{\frac{2MSE}{JK}}$. The appropriate number of d.f. is IJ(K – 1), so the C.I. is $\left(\overline{x}_{i..} - \overline{x}_{i'..}\right) \pm t_{a/2, IJ(K-1)}\sqrt{\frac{2MSE}{JK}}$. For the data of exercise 19, $\overline{x}_{2..} = 49.15$, $\overline{x}_{3..} = 50.37$, MSE = .0170, $t_{.025,9} = 2.262$, J = 3, K = 2, so the C.I. is

$$\left(49.15 - 50.37\right) \pm 2.262\sqrt{\frac{.0370}{6}} = -1.22 \pm .17 = \left(-1.39, -1.05\right).$$

26.

a. $\frac{E\left(MSAB\right)}{E\left(MSE\right)} = 1 + \frac{Ks_G^2}{s^2} = 1$ if $s_G^2 = 0$ and > 1 if $s_G^2 > 0$, so $\frac{MSAB}{MSE}$ is the appropriate F ratio.

b. $\frac{E\left(MSA\right)}{E\left(MSAB\right)} = \frac{s^2 + Ks_G^2 + JKs_A^2}{s^2 + Ks_G^2} = 1 + \frac{JKs_A^2}{s^2 + Ks_G^2} = 1$ if $s_A^2 = 0$ and > 1 if $s_A^2 > 0$, so $\frac{MSA}{MSAB}$ is the appropriate F ratio.

Section 11.3

27.

a.

Source	Df	SS	MS	f	$F_{.05}$
A	2	14,144.44	7072.22	61.06	3.35
B	2	5,511.27	2755.64	23.79	3.35
C	2	244,696.39	122.348.20	1056.24	3.35
AB	4	1,069.62	267.41	2.31	2.73
AC	4	62.67	15.67	.14	2.73
BC	4	331.67	82.92	.72	2.73
ABC	8	1,080.77	135.10	1.17	2.31
Error	27	3,127.50	115.83		
Total	53	270,024.33			

b. The computed f-statistics for all four interaction terms are less than the tabled values for statistical significance at the level .05. This indicates that none of the interactions are statistically significant.

c. The computed f-statistics for all three main effects exceed the tabled value for significance at level .05. All three main effects are statistically significant.

d. $Q_{.05,3,27}$ is not tabled, use $Q_{.05,3,24} = 3.53$, $w = 3.53\sqrt{\frac{115.83}{(3)(3)(2)}} = 8.95$. All three levels differ significantly from each other.

28.

Source	Df	SS	MS	f	$F_{.01}$
A	3	19,149.73	6,383.24	2.70	4.72
B	2	2,589,047.62	1,294,523.81	546.79	5.61
C	1	157,437.52	157,437.52	66.50	7.82
AB	6	53,238.21	8,873.04	3.75	3.67
AC	3	9,033.73	3,011.24	1.27	4.72
BC	2	91,880.04	45,940.02	19.40	5.61
ABC	6	6,558.46	1,093.08	.46	3.67
Error	24	56,819.50	2,367.48		
Total	47	2,983,164.81			

The statistically significant interactions are AB and BC. Factor A appears to be the least significant of all the factors. It does not have a significant main effect and the significant interaction (AB) is only slightly greater than the tabled value at significance level .01

29. I = 3, J = 2, K = 4, L = 4; $SSA = JKL\sum(\bar{x}_{i...} - \bar{x}_{....})^2$; $SSB = IKL\sum(\bar{x}_{.j..} - \bar{x}_{....})^2$; $SSC = IJL\sum(\bar{x}_{..k.} - \bar{x}_{....})^2$.

For level A: $\bar{x}_{1...} = 3.781$ $\bar{x}_{2...} = 3.625$ $\bar{x}_{3...} = 4.469$

For level B: $\bar{x}_{.1..} = 4.979$ $\bar{x}_{.2..} = 2.938$

For level C: $\bar{x}_{..1.} = 3.417$ $\bar{x}_{..2.} = 5.875$ $\bar{x}_{..3.} = .875$ $\bar{x}_{..4.} = 5.667$

$\bar{x}_{....} = 3.958$

SSA = 12.907; SSB = 99.976; SSC = 393.436

a.

Source	Df	SS	MS	f	$F_{.05}$*
A	2	12.907	6.454	1.04	3.15
B	1	99.976	99.976	16.09	4.00
C	3	393.436	131.145	21.10	2.76
AB	2	1.646	.823	.13	3.15
AC	6	71.021	11.837	1.90	2.25
BC	3	1.542	.514	.08	2.76
ABC	6	9.805	1.634	.26	2.25
Error	72	447.500	6.215		
Total	95	1,037.833			

*use 60 df for denominator of tabled F.

b. No interaction effects are significant at level .05

c. Factor B and C main effects are significant at the level .05

d. $Q_{.05,4,72}$ is not tabled, use $Q_{.05,4,60} = 3.74$, $w = 3.74\sqrt{\dfrac{6.215}{(3)(2)(4)}} = 1.90$.

Machine:	3	1	4	2
Mean:	.875	3.417	5.667	5.875

30.

a. See ANOVA table

b.

Source	**Df**	**SS**	**MS**	**f**	**$F_{.05}$**
A	3	.22625	.075417	77.35	9.28
B	1	.000025	.000025	.03	10.13
C	1	.0036	.0036	3.69	10.13
AB	3	.004325	.0014417	1.48	9.28
AC	3	.00065	.000217	.22	9.28
BC	1	.000625	.000625	.64	10.13
ABC	3	.002925	.000975		
Error	--	--	--		
Total	15	.2384			

The only statistically significant effect at the level .05 is the factor A main effect: levels of nitrogen.

c. $Q_{.05,4,3} = 6.82$; $w = 6.82\sqrt{\frac{.002925}{(2)(2)}} = .1844$.

1	2	3	4
1.1200	1.3025	1.3875	1.4300

31.

$x_{ij.}$	B_1	B_2	B_3
A_1	210.2	224.9	218.1
A_2	224.1	229.5	221.5
A_3	217.7	230.0	202.0
$x_{.j.}$	652.0	684.4	641.6

$x_{i.k}$	A_1	A_2	A_3
C_1	213.8	222.0	205.0
C_2	225.6	226.5	223.5
C_3	213.8	226.6	221.2
$x_{i..}$	653.2	675.1	649.7

$x_{.jk}$	C_1	C_2	C_3
B_1	213.5	220.5	218.0
B_2	214.3	246.1	224.0
B_3	213.0	209.0	219.6
$x_{..k}$	640.8	675.6	661.6

$\Sigma\Sigma x_{ij.}^2 = 435{,}382.26$ $\quad \Sigma\Sigma x_{i.k}^2 = 435{,}156.74$ $\quad \Sigma\Sigma x_{.jk}^2 = 435{,}666.36$

$\Sigma x_{.j.}^2 = 1{,}305{,}157.92$ $\quad \Sigma x_{i..}^2 = 1{,}304{,}540.34$ $\quad \Sigma x_{..k}^2 = 1{,}304{,}774.56$

Also, $\Sigma\Sigma\Sigma x_{ijk}^2 = 145{,}386.40$, $x_{...} = 1978$, CF = 144,906.81, from which we obtain the ANOVA table displayed in the problem statement. $F_{.01,4,8} = 7.01$, so the AB and BC interactions are significant (as can be seen from the p-values) and tests for main effects are not appropriate.

32.

a. Since $\frac{E(MSABC)}{E(MSE)} = \frac{s^2 + Ls_{ABC}^2}{s^2} = 1$ if $s_{ABC}^2 = 0$ and > 1 if $s_{ABC}^2 > 0$, $\frac{MSABC}{MSE}$ is the appropriate F ratio for testing $H_0 : s_{ABC}^2 = 0$. Similarly, $\frac{MSC}{MSE}$ is the F ratio for testing $H_0 : s_C^2 = 0$; $\frac{MSAB}{MSABC}$ is the F ratio for testing $H_0 : all$ $g_{ij}^{AB} = 0$; and $\frac{MSA}{MSAC}$ is the F ratio for testing $H_0 : all \ \ a_i = 0$.

b.

Source	**Df**	**SS**	**MS**	**f**	$F_{.01}$
A	1	14,318.24	14,318.24	$\frac{MSA}{MSAC} = 19.85$	98.50
B	3	9656.4	3218.80	$\frac{MSB}{MSBC} = 6.24$	9.78
C	2	2270.22	1135.11	$\frac{MSC}{MSE} = 3.15$	5.61
AB	3	3408.93	1136.31	$\frac{MSAB}{MSABC} = 2.41$	9.78
AC	2	1442.58	721.29	$\frac{MSAC}{MSABC} = 2.00$	5.61
BC	6	3096.21	516.04	$\frac{MSBC}{MSE} = 1.43$	3.67
ABC	6	2832.72	472.12	$\frac{MSABC}{MSE} = 1.31$	3.67
Error	24	8655.60	360.65		
Total	47				

At level .01, no H_0's can be rejected, so there appear to be no interaction or main effects present.

33.

Source	**Df**	**SS**	**MS**	**f**
A	6	67.32	11.02	
B	6	51.06	8.51	
C	6	5.43	.91	.61
Error	30	44.26	1.48	
Total	48	168.07		

Since .61 < $F_{.05,6,30} = 2.42$, treatment was not effective.

34.

	1	**2**	**3**	**4**	**5**	**6**
$x_{i..}$	144	205	272	293	85	98
$x_{.j.}$	171	199	147	221	177	182
$x_{..k}$	180	161	186	171	169	230

Thus $x_{...} = 1097$, $CF = \frac{(1097)^2}{36} = 33{,}428.03$, $\Sigma\Sigma x_{ij(k)}^2 = 42{,}219$, $\Sigma x_{i..}^2 = 239{,}423$, $\Sigma x_{.j.}^2 = 203{,}745$, $\Sigma x_{..k}^2 = 203.619$

Source	Df	SS	MS	f
A	5	6475.80	1295.16	
B	5	529.47	105.89	
C	5	508.47	101.69	1.59
Error	20	1277.23	63.89	
Total	35	8790.97		

Since 1.59 is not $\geq F_{.05,5,20} = 2.71$, H_{oC} is not rejected; shelf space does not appear to affect sales.

35.

	1	**2**	**3**	**4**	**5**	
$x_{i..}$	40.68	30.04	44.02	32.14	33.21	$\Sigma x_{i..}^2 = 6630.91$
$x_{.j.}$	29.19	31.61	37.31	40.16	41.82	$\Sigma x_{.j.}^2 = 6605.02$
$x_{..k}$	36.59	36.67	36.03	34.50	36.30	$\Sigma x_{..k}^2 = 6489.92$

$x_{...} = 180.09$, CF = 1297.30, $\Sigma\Sigma x_{ij(k)}^2 = 1358.60$

Source	Df	SS	MS	f
A	4	28.89	7.22	10.78
B	4	23.71	5.93	8.85
C	4	0.69	0.17	0.25
Error	12	8.01	.67	
Total	24	61.30		

$F_{4,12} = 3.26$, so both factor A (plant) and B(leaf size) appear to affect moisture content, but factor C (time of weighing) does not.

36.

Source	Df	SS	MS	f	$F_{.01}$*
A (laundry treatment)	3	39.171	13.057	16.23	3.95
B (pen type)	2	.665	.3325	.41	4.79
C (Fabric type)	5	21.508	4.3016	5.35	3.17
AB	6	1.432	.2387	.30	2.96
AC	15	15.953	1.0635	1.32	2.19
BC	10	1.382	.1382	.17	2.47
ABC	30	9.016	.3005	.37	1.86
Error	144	115.820	.8043		
Total	215	204.947			

*Because denominator degrees of freedom for 144 is not tabled, use 120.

At the level .01, there are two statistically significant main effects (laundry treatment and fabric type). There are no statistically significant interactions.

37.

Source	Df	MS	f	$F_{.01}$*
A	2	2207.329	2259.29	5.39
B	1	47.255	48.37	7.56
C	2	491.783	503.36	5.39
D	1	.044	.05	7.56
AB	2	15.303	15.66	5.39
AC	4	275.446	281.93	4.02
AD	2	.470	.48	5.39
BC	2	2.141	2.19	5.39
BD	1	.273	.28	7.56
CD	2	.247	.25	5.39
ABC	4	3.714	3.80	4.02
ABD	2	4.072	4.17	5.39
ACD	4	.767	.79	4.02
BCD	2	.280	.29	5.39
ABCD	4	.347	.355	4.02
Error	36	.977		
Total	71			

*Because denominator d.f. for 36 is not tabled, use d.f. = 30

SST = (71)(93.621) = 6,647.091. Computing all other sums of squares and adding them up = 6,645.702. Thus SSABCD = 6,647.091 – 6,645.702 = 1.389 and

$$MSABCD = \frac{1.389}{4} = .347 .$$

At level .01 the statistically significant main effects are A, B, C. The interaction AB and AC are also statistically significant. No other interactions are statistically significant.

Section 11.4

38.

a.

Treatment Condition	$x_{ijk.}$	1	2	Effect Contrast	$SS = \frac{(contrast)^2}{16}$
$(1) = x_{111.}$	404.2	839.2	1991.0	3697.0	
$a = x_{211.}$	435.0	1151.8	1706.0	164.2	1685.1
$b = x_{121.}$	549.6	717.6	83.4	583.4	21,272.2
$ab = x_{221.}$	602.2	988.4	80.8	24.2	36.6
$c = x_{112.}$	339.2	30.8	312.6	-285.0	5076.6
$ac = x_{212.}$	378.4	52.6	270.8	-2.6	.4
$bc = x_{122.}$	473.4	39.2	21.8	-41.8	109.2
$abc = x_{222.}$	515.0	41.6	2.4	-19.4	23.5

$$\Sigma\Sigma\Sigma\Sigma x_{ijkl}^2 = 882{,}573.38; \quad SST = 882{,}573.38 - \frac{(3697)^2}{16} = 28{,}335.3$$

b. The important effects are those with small associated p-values, indicating statistical significance. Those effects significant at level .05 (i.e., p-value < .05) are the three main effects and the speed by distance interaction.

39.

Condition	Total	1	2	Contrast	$SS = \frac{(contrast)^2}{24}$
111	315	927	2478	5485	
211	612	1551	3007	1307	A = 71,177.04
121	584	1163	680	1305	B = 70,959.38
221	967	1844	627	199	AB = 1650.04
112	453	297	624	529	C = 11,660.04
212	710	383	681	-53	AC = 117.04
122	737	257	86	57	BC = 135.38
222	1107	370	113	27	ABC = 30.38

a. $\hat{b}_1 = \bar{x}_{.2..} - \bar{x}_{....} = \dfrac{584+967+737+1107-315-612-453-710}{24} = 54.38$

$\hat{g}_{11}^{AC} = \dfrac{315-612+584-967-453+710-737+1107}{24} = 2.21$;

$\hat{g}_{21}^{AC} = -\hat{g}_{11}^{AC} = 2.21$.

b. Factor SS's appear above. With $CF = \dfrac{5485^2}{24} = 1,253,551.04$ and $\Sigma\Sigma\Sigma\Sigma x_{ijkl}^2 = 1,411,889$, SST = 158,337.96, from which SSE = 2608.7. The ANOVA table appears in the answer section. $F_{.05,1,16} = 4.49$, from which we see that the AB interaction and al the main effects are significant.

40.

a. In the accompanying ANOVA table, effects are listed in the order implied by Yates' algorithm. $\Sigma\Sigma\Sigma\Sigma\Sigma x_{ijklm}^2 = 4783.16$, $x_{.....} = 388.14$, so $SST = 4783.16 - \frac{368.14^2}{32} = 72.56$ and SSE = 72.56 – (sum of all other SS's) = 35.85.

Source	**Df**	**SS**	**MS**	**f**
A	1	.17	.17	< 1
B	1	1.94	1.94	< 1
C	1	3.42	3.42	1.53
D	1	8.16	8.16	3.64
AB	1	.26	.26	< 1
AC	1	.74	.74	< 1
AD	1	.02	.02	< 1
BC	1	13.08	13.08	5.84
BD	1	.91	.91	< 1
CD	1	.78	.78	< 1
ABC	1	.78	.78	< 1
ABD	1	6.77	6.77	3.02
ACD	1	.62	.62	< 1
BCD	1	1.76	1.76	< 1
ABCD	1	.00	.00	< 1
Error	16	35.85	2.24	
Total	31			

b. $F_{.05,1,16} = 4.49$, so none of the interaction effects is judged significant, and only the D main effect is significant.

41. $\Sigma\Sigma\Sigma\Sigma\Sigma x_{ijklm}^2 = 3{,}308{,}143$, $x_{.....} = 11{,}956$, so $CF = \frac{(11{,}956)^2}{48} = 2{,}979{,}535.02$, and SST = 328,607.98. Each SS is $\frac{(effectcontrast)^2}{48}$ and SSE is obtained by subtraction. The ANOVA table appears in the answer section. $F_{.05,1,32} \approx 4.15$, a value exceeded by the F ratios for AB interaction and the four main effects.

42. $\Sigma\Sigma\Sigma\Sigma\Sigma x_{ijklm}^2 = 32{,}917{,}817$, $x_{.....} = 39{,}371$, $SS = \dfrac{(contrast)^2}{48}$, and error d.f. = 32.

Effect	MS	f	Effect	MS	f
A	16,170.02	3.42	**BD**	3519.19	< 1
B	332,167.69	70.17	**CD**	4700.52	< 1
C	43,140.02	9.11	**ABC**	1210.02	< 1
D	20,460.02	4.33	**ABD**	15,229.69	3.22
AB	1989.19	< 1	**ACD**	1963.52	< 1
AC	776.02	< 1	**BCD**	10,354.69	2.19
AD	16,170.02	3.42	**ABCD**	1692.19	< 1
BC	3553.52	< 1	**Error**	4733.69	

$F_{.01,1,32} \approx 7.5$, so only the B and C main effects are judged significant at the 1% level.

43.

Condition/Effect	$SS = \frac{(contrast)^2}{16}$	f	Condition/ Effect	$SS = \frac{(contrast)^2}{16}$	f
(1)	--		**D**	414.123	1067.33
A	.436	1.12	**AD**	.017	< 1
B	.099	< 1	**BD**	.456	< 1
AB	.497	1.28	**ABD**	.055	--
C	.109	< 1	**CD**	2.190	5.64
AC	.078	< 1	**ACD**	1.020	--
BC	1.404	3.62	**BCD**	.133	--
ABC	.051	--	**ABCD**	.681	--

SSE = .051 + .055 + 1.020 + .133 + .681 = 1.940, d.f. = 5, so MSE = .388. $F_{.05,1,5} = 6.61$, so only the D main effect is significant.

44.

a. The eight treatment conditions which have even number of letters in common with abcd and thus go in the first (principle) block are (1), ab, ac, bc, ad, bd, cd, and abd; the other eight conditions are placed in the second block.

b. and **c.**

$x_{....} = 1290$, $\Sigma\Sigma\Sigma\Sigma x_{ijkl}^2 = 105{,}160$, so SST = 1153.75. The two block totals are 639 and 651, so $SSBl = \frac{639^2}{8} + \frac{651^2}{8} - \frac{1290^2}{16} = 9.00$, which is identical (as it must be here) to SSABCD computed from Yates algorithm.

Condition/Effect	**Block**	$SS = \frac{(contrast)^2}{16}$	**f**
(1)	1	--	
A	2	25.00	1.93
B	2	9.00	< 1
AB	1	12.25	< 1
C	2	49.00	3.79
AC	1	2.25	< 1
BC	1	.25	< 1
ABC	2	9.00	--
D	2	930.25	71.90
AD	1	36.00	2.78
BD	1	25.00	1.93
ABD	2	20.25	--
CD	1	4.00	< 1
ACD	2	20.25	--
BCD	2	2.25	--
ABCD=Blocks	1	9.00	--
Total		1153.75	

$SSE = 9.0 + 20.25 + 20.25 + 2.25 = 51.75$; d.f. = 4, so MSE = 12.9375, $F_{.05,1,4} = 7.71$, so only the D main effect is significant.

45.

a. The allocation of treatments to blocks is as given in the answer section, with block #1 containing all treatments having an even number of letters in common with both ab and cd, etc.

b. $x_{.....} = 16{,}898$, so $SST = 9{,}035{,}054 - \frac{16{,}898^2}{32} = 111{,}853.88$. The eight $block \times replication$ totals are 2091 (= 618 + 421 + 603 + 449, the sum of the four observations in block #1 on replication #1), 2092, 2133, 2145, 2113, 2080, 2122, and 2122, so $SSBl = \frac{2091^2}{4} + ... + \frac{2122^2}{4} - \frac{16{,}898^2}{32} = 898.88$. The remaining SS's as well as all F ratios appear in the ANOVA table in the answer section. With $F_{.01,1,12} = 9.33$, only the A and B main effects are significant.

46. The result is clearly true if either defining effect is represented by either a single letter (e.g., A) or a pair of letters (e.g. AB). The only other possibilities are for both to be "triples" (e.g. ABC or ABD, all of which must have two letters in common.) or one a triple and the other ABCD. But the generalized interaction of ABC and ABD is CD, so a two-factor interaction is confounded, and the generalized interaction of ABC and ABCD is D, so a main effect is confounded.

47. See the text's answer section.

48.

a. The treatment conditions in the observed group are (in standard order) (1), ab, ac, bc, ad, bd, cd, and abcd. The alias pairs are {A, BCD}, {B, ACD}, {C, ABD}, {D, ABC}, {AB, CD}, {AC, BD}, and {AD, BC}.

b.

	A	**B**	**C**	**D**	**AB**	**AC**	**AD**
(1) = 19.09	-	-	-	-	+	+	+
Ab = 20.11	+	+	-	-	+	-	-
Ac = 21.66	+	-	+	-	-	+	-
Bc = 20.44	-	+	+	-	-	-	+
Ad = 13.72	+	-	-	+	-	-	+
Bd = 11.26	-	+	-	+	-	+	-
Cd = 11.72	-	-	+	+	+	-	-
Abcd = 12.29	+	+	+	+	+	+	+
Contrast	5.27	-2.09	1.93	-32.31	-3.87	-1.69	.79
SS	3.47	.55	.47	130.49	1.87	.36	.08
f	4.51	< 1	< 1	169.47	SSE=2.31		MSE=.770

$F_{.05,1,3} = 10.13$, so only the D main effect is judged significant.

49.

		A	B	C	D	E	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
a	70.4	+	-	-	-	-	-	-	-	-	+	+	+	+	+	+
b	72.1	-	+	-	-	-	-	+	+	+	-	-	-	+	+	+
c	70.4	-	-	+	-	-	+	-	+	+	-	+	+	-	-	+
abc	73.8	+	+	+	-	-	+	+	-	-	+	-	-	-	-	+
d	67.4	-	-	-	+	-	+	+	-	+	+	-	+	-	+	-
abd	67.0	+	+	-	+	-	+	-	+	-	-	+	-	-	+	-
acd	66.6	+	-	+	+	-	-	+	+	-	-	-	+	+	-	-
bcd	66.8	-	+	+	+	-	-	-	-	+	+	+	-	+	-	-
e	68.0	-	-	-	-	+	+	+	+	-	+	+	-	+	-	-
abe	67.8	+	+	-	-	+	+	-	-	+	-	-	+	+	-	-
ace	67.5	+	-	+	-	+	-	+	-	+	-	+	-	-	+	-
bce	70.3	-	+	+	-	+	-	-	+	-	+	-	+	-	+	-
ade	64.0	+	-	-	+	+	-	-	+	+	+	-	-	-	-	+
bde	67.9	-	+	-	+	+	-	+	-	-	-	+	+	-	-	+
cde	65.9	-	-	+	+	+	+	-	-	-	-	-	-	+	+	+
abcde	68.0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Thus $SSA = \dfrac{(70.4 - 72.1 - 70.4 + ... + 68.0)^2}{16} = 2.250$, SSB = 7.840, SSC = .360, SSD = 52.563, SSE = 10.240, SSAB = 1.563, SSAC = 7.563, SSAD = .090, SSAE = 4.203, SSBC = 2.103, SSBD = .010, SSBE = .123, SSCD = .010, SSCE = .063, SSDE = 4.840, Error SS = sum of two factor SS's = 20.568, Error MS = 2.057, $F_{.01,1,10} = 10.04$, so only the D main effect is significant.

Supplementary Exercises

50.

Source	Df	SS	MS	f
Treatment	4	14.962	3.741	36.7
Block	8	9.696		
Error	32	3.262	.102	
Total	44	27.920		

$H_0 : a_1 = a_2 = a_3 = a_4 = a_5 = 0$ will be rejected if $f = \dfrac{MSTr}{MSE} \geq F_{.05,4,32} = 2.67$.

Because $36.7 \geq 2.67$, H_o is rejected. We conclude that expected smoothness score does depend somehow on the drying method used.

51.

Source	Df	SS	MS	f
A	1	322.667	322.667	980.38
B	3	35.623	11.874	36.08
AB	3	8.557	2.852	8.67
Error	16	5.266	.329	
Total	23	372.113		

We first test the null hypothesis of no interactions ($H_0 : g_{ij} = 0$ for all I, j). H_o will be rejected at level .05 if $f_{AB} = \frac{MSAB}{MSE} \geq F_{.05,3,16} = 3.24$. Because $8.67 \geq 3.24$, H_o is rejected. Because we have concluded that interaction is present, tests for main effects are not appropriate.

52. Let $X_{ij} =$ the amount of clover accumulation when the i[th] sowing rate is used in the j[th] plot = $m + a_i + b_j + e_{ij}$. $H_0 : a_1 = a_2 = a_3 = a_4 = 0$ will be rejected if

$$f = \frac{MSTr}{MSE} \geq F_{a,I-1,(I-1)(J-1)} = F_{.05,3,9} = 3.86$$

Source	Df	SS	MS	f
Treatment	3	3,141,153.5	1,040,751.17	2.28
Block	3	19,470,550.0		
Error	9	4,141,165.5	460,129.50	
Total	15	26,752,869.0		

Because $2.28 < 3.86$, H_o is not rejected. Expected accumulation does not appear to depend on sowing rate.

53. Let A = spray volume, B = belt speed, C = brand.

Condition	Total	1	2	Contrast	$SS = \frac{(contrast)^2}{16}$
(1)	76	129	289	592	21,904.00
A	53	160	303	22	30.25
B	62	143	13	48	144.00
AB	98	160	9	134	1122.25
C	88	-23	31	14	12.25
AC	55	36	17	-4	1.00
BC	59	-33	59	-14	12.25
ABC	101	42	75	16	16.00

The ANOVA table is as follows:

Effect	**Df**	**MS**	**f**
A	1	30.25	6.72
B	1	144.00	32.00
AB	1	1122.25	249.39
C	1	12.25	2.72
AC	1	1.00	.22
BC	1	12.25	2.72
ABC	1	16.00	3.56
Error	8	4.50	
Total	15		

$F_{.05,1,8} = 5.32$, so all of the main effects are significant at level .05, but none of the interactions are significant.

54. We use Yates' method for calculating the sums of squares, and for ease of calculation, we divide each observation by 1000.

Condition	Total	1	2	Contrast	$SS = \frac{(contrast)^2}{8}$
(1)	23.1	66.1	213.5	317.2	-
A	43.0	147.4	103.7	20.2	51.005
B	71.4	70.2	24.5	44.6	248.645
AB	76.0	33.5	-4.3	-12.0	18.000
C	37.0	19.9	81.3	-109.8	1507.005
AC	33.2	4.6	-36.7	-28.8	103.68
BC	17.0	-3.8	-15.3	-118.0	1740.5
ABC	16.5	-.5	3.3	18.6	43.245

We assume that there is no three-way interaction, so the MSABC becomes the MSE for ANOVA:

Source	df	MS	f
A	1	51.005	1.179
B	1	248.645	5.750*
AB	1	18.000	< 1
C	1	1507.005	34.848*
AC	1	103.68	2.398
BC	1	1740.5	40.247*
Error	1	43.245	
Total	8		

With $F_{.05,1,8} = 5.32$, the B and C main effects are significant at the .05 level, as well as the BC interaction. We conclude that although binder type (A) is not significant, both amount of water (B) and the land disposal scenario (C) affect the leaching characteristics under study., and there is some interaction between the two factors.

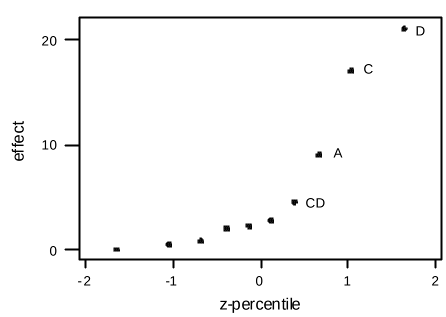
55.

a.

Effect	**%Iron**	**1**	**2**	**3**	**Effect Contrast**	**SS**
	7	18	37	174	684	
A	11	19	137	510	144	1296
B	7	62	169	50	36	81
AB	12	75	341	94	0	0
C	21	79	9	14	272	4624
AC	41	90	41	22	32	64
BC	27	165	47	2	12	9
ABC	48	176	47	-2	-4	1
D	28	4	1	100	336	7056
AD	51	5	13	172	44	121
BD	33	20	11	32	8	4
ABD	57	21	11	0	0	0
CD	70	23	1	12	72	324
ACD	95	24	1	0	-32	64
BCD	77	25	1	0	-12	9
ABCD	99	22	-3	-4	-4	1

We use $estimate = \dfrac{contrast}{2^p}$ when n = 1 (see p 472 of text) to get $\hat{a}_1 = \dfrac{144}{2^4} = \dfrac{144}{16} = 9.00$, $\hat{b}_1 = \dfrac{36}{16} = 2.25$, $\hat{d}_1 = \dfrac{272}{16} = 17.00$, $\hat{g}_1 = \dfrac{336}{16} = 21.00$. Similarly, $\left(\hat{ab}\right)_{11} = 0$, $\left(\hat{ad}\right)_{11} = 2.00$, $\left(\hat{ag}\right)_{11} = 2.75$, $\left(\hat{bd}\right)_{11} = .75$, $\left(\hat{bg}\right)_{11} = .50$, and $\left(\hat{dg}\right)_{11} = 4.50$.

b.



The plot suggests main effects A, C, and D are quite important, and perhaps the interaction CD as well. (See answer section for comment.)

56. The summary quantities are:

	$x_{ij\bullet}$	j = 1	2	3	$x_{i\bullet\bullet}$
i	1	6.2	4.0	5.8	16.0
	2	7.6	6.2	6.4	20.2
	$x_{\bullet j\bullet}$	13.8	10.2	12.2	$x_{\bullet\bullet\bullet} = 36.2$

$$CF = \frac{(36.2)^2}{30} = 43.6813,\ \Sigma\Sigma\Sigma x_{ijk}^2 = 45.560, \text{ so}$$
$$SST = 45.560 - 43.6813 = 1.8787,$$
$$SSE = 45.560 - \frac{225.24}{5} = .5120,\ SSA = \frac{(16.0)^2 + (20.2)^2}{15} - CF = .5880,$$
$$SSB = \frac{(13.8)^2 + (10.2)^2 + (12.2)^2}{10} - CF = .6507,$$

and by subtraction, SSAB = .128

```
Analysis of Variance for Average Bud Rating
Source          DF        SS        MS        F
Health           1    0.5880    0.5880    27.56
pH               2    0.6507    0.3253    15.25
Interaction      2    0.1280    0.0640     3.00
Error           24    0.5120    0.0213
Total           29    1.8787
```

Since 3.00 is not $\geq F_{.05,2,24} = 3.40$, we fail to reject the no interactions hypothesis, and we continue: $27.56 \geq F_{.05,1,24} = 4.26$, and $15.25 \geq F_{.05,2,24} = 3.40$, so we conclude that both the health of the seedling and its pH level have an effect on the average rating.

57. The ANOVA table is:

Source	df	SS	MS	f	$F_{.01}$
A	2	34,436	17,218	436.92	5.49
B	2	105,793	52,897	1342.30	5.49
C	2	516,398	258,199	6552.04	5.49
AB	4	6,868	1,717	43.57	4.11
AC	4	10,922	2,731	69.29	4.11
BC	4	10,178	2,545	64.57	4.11
ABC	8	6,713	839	21.30	3.26
Error	27	1,064	39		
Total	53	692,372			

All calculated f values are greater than their respective tabled values, so all effects, including the interaction effects, are significant at level .01.

58.

Source	**df**	**SS**	**MS**	**f**	$\mathbf{F_{.05}}$
A(pressure)	1	6.94	6.940	11.57*	4.26
B(time)	3	5.61	1.870	3.12*	3.01
C(concen.)	2	12.33	6.165	10.28*	3.40
AB	3	4.05	1.350	2.25	3.01
AC	2	7.32	3.660	6.10*	3.40
BC	6	15.80	2.633	4.39*	2.51
ABC	6	4.37	.728	1.21	2.51
Error	24	14.40	.600		
Total	47	70.82			

There appear to be no three-factor interactions. However both AC and BC two-factor interactions appear to be present.

59. Based on the p-values in the ANOVA table, statistically significant factors at the level .01 are adhesive type and cure time. The conductor material does not have a statistically significant effect on bond strength. There are no significant interactions.

60.

Source	df	SS	MS	f	$F_{.05}$
A (diet)	2	18,138	9.69.0	28.9*	≈ 3.32
B (temp.)	2	5,182	2591.0	8.3*	≈ 3.32
Interaction	4	1,737	434.3	1.4	≈ 2.69
Error	36	11,291	313.6		
Total	44	36,348			

Interaction appears to be absent. However, since both main effect f values exceed the corresponding F critical values, both diet and temperature appear to affect expected energy intake.

61. $SSA = \sum_i \sum_j \left(\overline{X}_{i...} - \overline{X}_{....}\right)^2 = \frac{1}{N}\Sigma X_{i...}^2 - \frac{X_{....}^2}{N}$, with similar expressions for SSB, SSC, and SSD, each having N – 1 df.

$SST = \sum_i \sum_j \left(X_{ij(kl)} - \overline{X}_{....}\right)^2 = \sum_i \sum_j X_{ij(kl)}^2 - \frac{X_{....}^2}{N}$ with $N^2 - 1$ df, leaving $N^2 - 1 - 4(N-1)$ df for error.

	1	2	3	4	5	Σx^2
$x_{i...}$:	482	446	464	468	434	1,053,916
$x_{.j..}$:	470	451	440	482	451	1,053,626
$x_{..k.}$:	372	429	484	528	481	1,066,826
$x_{...l}$:	340	417	466	537	534	1,080,170

Also, $\Sigma\Sigma x_{ij(kl)}^2 = 220{,}378$, $x_{....} = 2294$, and CF = 210,497.44

Source	df	SS	MS	f	$F_{.05}$
A	4	285.76	71.44	.594	3.84
B	4	227.76	56.94	.473	3.84
C	4	2867.76	716.94	5.958*	3.84
D	4	5536.56	1384.14	11.502*	3.84
Error	8	962.72	120.34		
Total	24				

H_{oA} and H_{oB} cannot be rejected, while while H_{oC} and H_{oD} are rejected.

CHAPTER 12

Section 12.1

1.

a. Stem and Leaf display of temp:

17	0	
17	23	stem = tens
17	445	leaf = ones
17	67	
17		
18	0000011	
18	2222	
18	445	
18	6	
18	8	

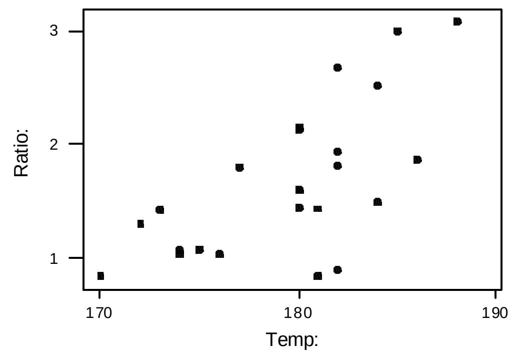
180 appears to be a typical value for this data. The distribution is reasonably symmetric in appearance and somewhat bell-shaped. The variation in the data is fairly small since the range of values (188 – 170 = 18) is fairly small compared to the typical value of 180.

0	889	
1	0000	stem = ones
1	3	leaf = tenths
1	4444	
1	66	
1	8889	
2	11	
2		
2	5	
2	6	
2		
3	00	

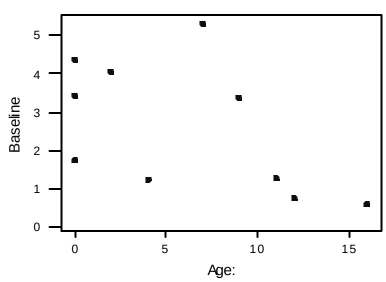
For the ratio data, a typical value is around 1.6 and the distribution appears to be positively skewed. The variation in the data is large since the range of the data (3.08 - .84 = 2.24) is very large compared to the typical value of 1.6. The two largest values could be outliers.

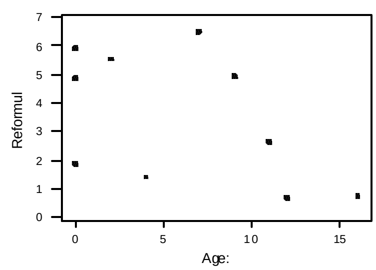
b. The efficiency ratio is not uniquely determined by temperature since there are several instances in the data of equal temperatures associated with different efficiency ratios. For example, the five observations with temperatures of 180 each have different efficiency ratios.

c. A scatter plot of the data appears below. The points exhibit quite a bit of variation and do not appear to fall close to any straight line or simple curve.



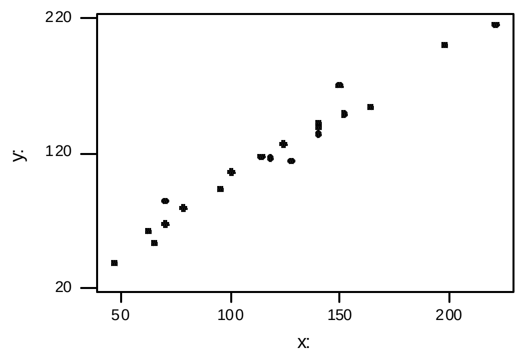
2. Scatter plots for the emissions vs age:





With this data the relationship between the age of the lawn mower and its NO_x emissions seems somewhat dubious. One might have expected to see that as the age of the lawn mower increased the emissions would also increase. We certainly do not see such a pattern. Age does not seem to be a particularly useful predictor of NO_x emission.

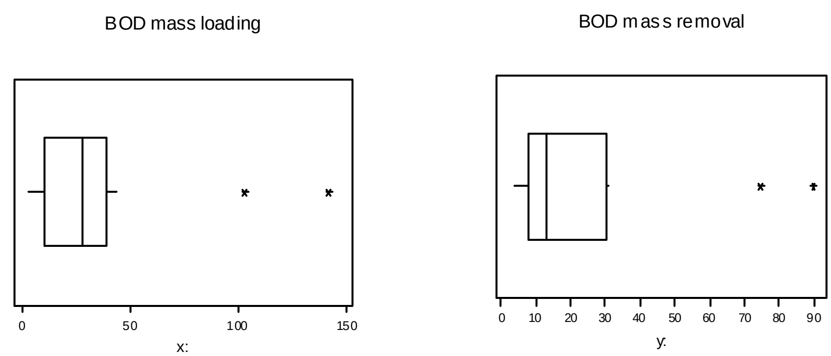
3. A scatter plot of the data appears below. The points fall very close to a straight line with an intercept of approximately 0 and a slope of about 1. This suggests that the two methods are producing substantially the same concentration measurements.



4.

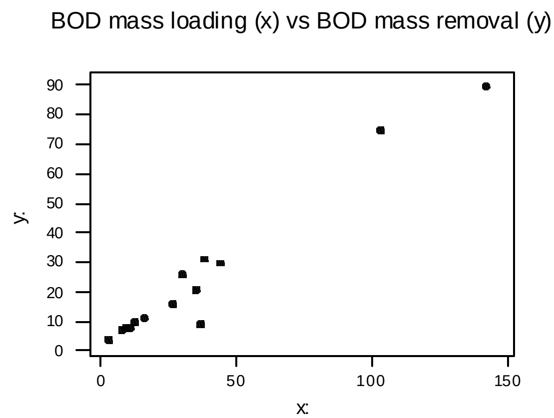
a.

Box plots of both variables:



On both the BOD mass loading boxplot and the BOD mass removal boxplot there are 2 outliers. Both variables are positively skewed.

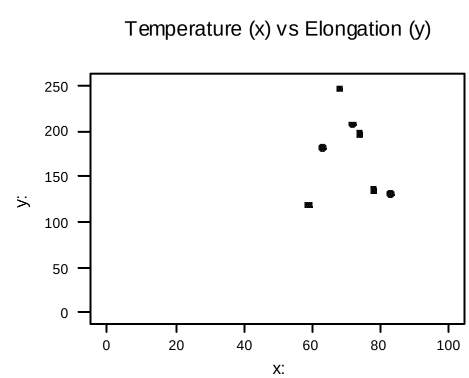
b. Scatter plot of the data:



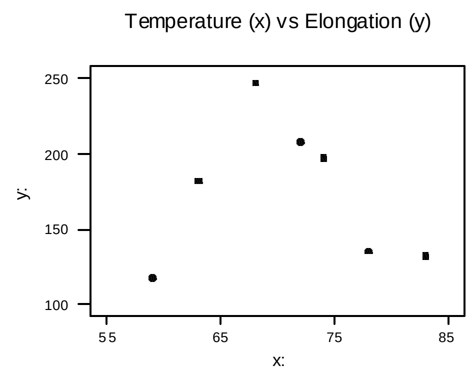
There is a strong linear relationship between BOD mass loading and BOD mass removal. As the loading increases, so does the removal. The two outliers seen on each of the boxplots are seen to be correlated here. There is one observation that appears not to match the liner pattern. This value is (37, 9). One might have expected a larger value for BOD mass removal.

5.

a. The scatter plot with axes intersecting at (0,0) is shown below.



b. The scatter plot with axes intersecting at (55, 100) is shown below.



c. A parabola appears to provide a good fit to both graphs.

6. There appears to be a linear relationship between racket resonance frequency and sum of peak-to-peak acceleration. As the resonance frequency increases the sum of peak-to-peak acceleration tends to decrease. However, there is not a perfect relationship. Variation does exist. One should also notice that there are two tennis rackets that appear to differ from the other 21 rackets. Both have very high resonance frequency values. One might investigate if these rackets differ in other ways as well.

7.

a. $m_{Y\cdot 2500} = 1800 + 1.3(2500) = 5050$

b. expected change = slope = $b_1 = 1.3$

c. expected change = $100b_1 = 130$

d. expected change = $-100b_1 = -130$

8.

a. $m_{Y\cdot 2000} = 1800 + 1.3(2000) = 4400$, and $s = 350$, so $P(Y > 5000)$
$= P\left(Z > \dfrac{5000 - 4400}{350} \right) = P(Z > 1.71) = .0436$

b. Now E(Y) = 5050, so $P(Y > 5000) = P(Z > .14) = .4443$

c. $E(Y_2 - Y_1) = E(Y_2) - E(Y_1) = 5050 - 4400 = 650$, and
$V(Y_2 - Y_1) = V(Y_2) + V(Y_1) = (350)^2 + (350)^2 = 245{,}000$, so the s.d. of
$Y_2 - Y_1 = 494.97$.
Thus $P(Y_2 - Y_1 > 0) = P\left(z > \dfrac{100 - 650}{494.97} \right) = P(Z > .71) = .2389$

d. The standard deviation of $Y_2 - Y_1 = 494.97$ (from **c**), and
$E(Y_2 - Y_1) = 1800 + 1.3x_2 - (1800 + 1.3x_1) = 1.3(x_2 - x_1)$. Thus
$P(Y_2 > Y_1) = P(Y_2 - Y_1 > 0) = P\left(z > \dfrac{-1.3(x_2 - x_1)}{494.97} \right) = .95$ implies that
$-1.645 = \dfrac{-1.3(x_2 - x_1)}{494.97}$, so $x_2 - x_1 = 626.33$.

9.

a. $b_1 =$ expected change in flow rate (y) associated with a one inch increase in pressure drop (x) = .095.

b. We expect flow rate to decrease by $5b_1 = .475$.

c. $m_{Y\cdot 10} = -.12 + .095(10) = .83,$ and $m_{Y\cdot 15} = -.12 + .095(15) = 1.305$.

d. $P(Y > .835) = P\left(Z > \dfrac{.835 - .830}{.025} \right) = P(Z > .20) = .4207$
$P(Y > .840) = P\left(Z > \dfrac{.840 - .830}{.025} \right) = P(Z > .40) = .3446$

e. Let Y_1 and Y_2 denote pressure drops for flow rates of 10 and 11, respectively. Then $m_{Y\cdot 11} = .925,$ so Y_1 - Y_2 has expected value .830 - .925 = -.095, and s.d.
$\sqrt{(.025)^2 + (.025)^2} = .035355$. Thus
$P(Y_1 > Y_2) = P(Y_1 - Y_2 > 0) = P\left(z > \dfrac{+.095}{.035355} \right) = P(Z > 2.69) = .0036$

10. Y has expected value 14,000 when x = 1000 and 24,000 when x = 2000, so the two probabilities become $P\left(z > \frac{-8500}{s}\right) = .05$ and $P\left(z > \frac{-17{,}500}{s}\right) = .10$. Thus $\frac{-8500}{s} = -1.645$ and $\frac{-17{,}500}{s} = -1.28$. This gives two different values for s, a contradiction, so the answer to the question posed is no.

11.

a. $b_1 =$ expected change for a one degree increase = -.01, and $10b_1 = -.1$ is the expected change for a 10 degree increase.

b. $m_{Y\cdot 200} = 5.00 - .01(200) = 3$, and $m_{Y\cdot 250} = 2.5$.

c. The probability that the first observation is between 2.4 and 2.6 is

$$P(2.4 \le Y \le 2.6) = P\left(\frac{2.4 - 2.5}{.075} \le Z \le \frac{2.6 - 2.5}{.075}\right)$$

$= P(-1.33 \le Z \le 1.33) = .8164$. The probability that any particular one of the other four observations is between 2.4 and 2.6 is also .8164, so the probability that all five are between 2.4 and 2.6 is $(.8164)^5 = .3627$.

d. Let Y_1 and Y_2 denote the times at the higher and lower temperatures, respectively. Then $Y_1 - Y_2$ has expected value $5.00 - .01(x+1) - (5.00 - .01x) = -.01$. The standard deviation of $Y_1 - Y_2$ is $\sqrt{(.075)^2 + (.075)^2} = .10607$. Thus

$$P(Y_1 - Y_2 > 0) = P\left(z > \frac{-(-.01)}{.10607}\right) = P(Z > .09) = .4641.$$

Section 12.2

12.

a. $S_{xx} = 39{,}095 - \frac{(517)^2}{14} = 20{,}002.929$,

$S_{xy} = 25{,}825 - \frac{(517)(346)}{14} = 13047.714$; $\hat{b}_1 = \frac{S_{xy}}{S_{xx}} = \frac{13{,}047.714}{20{,}002.929} = .652$;

$\hat{b}_0 = \frac{\Sigma y - \hat{b}_1 \Sigma x}{n} = \frac{346 - (.652)(517)}{14} = .626$, so the equation of the least squares regression line is $y = .626 + .652x$.

b. $\hat{y}_{(35)} = .626 + .652(35) = 23.456$. The residual is $y - \hat{y} = 21 - 23.456 = -2.456$.

c. $S_{yy} = 17{,}454 - \frac{(346)^2}{14} = 8902.857$, so

$SSE = 8902.857 - (.652)(13047.714) = 395.747$.

$\hat{s} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{395.747}{12}} = 5.743$.

d. $SST = S_{yy} = 8902.857$; $r^2 = 1 - \frac{SSE}{SST} = 1 - \frac{395.747}{8902.857} = .956$.

e. Without the two upper extreme observations, the new summary values are $n = 12, \Sigma x = 272, \Sigma x^2 = 8322, \Sigma y = 181, \Sigma y^2 = 3729, \Sigma xy = 5320$. The new $S_{xx} = 2156.667, S_{yy} = 998.917, S_{xy} = 1217.333$. New $\hat{b}_1 = .56445$ and $\hat{b}_0 = 2.2891$, which yields the new equation $y = 2.2891 + .56445x$. Removing the two values changes the position of the line considerably, and the slope slightly. The new $r^2 = 1 - \frac{311.79}{998.917} = .6879$, which is much worse than that of the original set of observations.

13. For this data, n = 4, $\Sigma x_i = 200$, $\Sigma y_i = 5.37$, $\Sigma x_i^2 = 12.000$, $\Sigma y_i^2 = 9.3501$,

$\Sigma x_i y_i = 333$. $S_{xx} = 12{,}000 - \dfrac{(200)^2}{4} = 2000$,

$S_{yy} = 9.3501 - \dfrac{(5.37)^2}{4} = 2.140875$, and $S_{xy} = 333 - \dfrac{(200)(5.37)}{4} = 64.5$.

$\hat{b}_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{64.5}{2000} = .03225$ and $\hat{b}_0 = \dfrac{5.37}{4} - (.03225)\dfrac{200}{4} = -.27000$.

$SSE = S_{yy} - \hat{b}_1 S_{xy} = 2.14085 - (.03225)(64.5) = .060750$.

$r^2 = 1 - \dfrac{SSE}{SST} = 1 - \dfrac{.060750}{2.14085} = .972$. This is a very high value of r^2, which confirms the authors' claim that there is a strong linear relationship between the two variables.

14.

a. n = 24, $\Sigma x_i = 4308$, $\Sigma y_i = 40.09$, $\Sigma x_i^2 = 773{,}790$, $\Sigma y_i^2 = 76.8823$,

$\Sigma x_i y_i = 7{,}243.65$. $S_{xx} = 773{,}790 - \dfrac{(4308)^2}{24} = 504.0$,

$S_{yy} = 76.8823 - \dfrac{(40.09)^2}{24} = 9.9153$, and

$S_{xy} = 7{,}243.65 - \dfrac{(4308)(40.09)}{24} = 45.8246$. $\hat{b}_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{45.8246}{504} = .09092$ and

$\hat{b}_0 = \dfrac{40.09}{24} - (.09092)\dfrac{4308}{24} = -14.6497$. The equation of the estimated regression line is $\hat{y} = -14.6497 + .09092x$.

b. When x = 182, $\hat{y} = -14.6497 + .09092(182) = 1.8997$. So when the tank temperature is 182, we would predict an efficiency ratio of 1.8997.

c. The four observations for which temperature is 182 are: (182, .90), (182, 1.81), (182, 1.94), and (182, 2.68). Their corresponding residuals are: $.90 - 1.8997 = -0.9977$, $1.81 - 1.8997 = -0.0877$, $1.94 - 1.8997 = 0.0423$, $2.68 - 1.8997 = 0.7823$. These residuals do not all have the same sign because in the cases of the first two pairs of observations, the observed efficiency ratios were smaller than the predicted value of 1.8997. Whereas, in the cases of the last two pairs of observations, the observed efficiency ratios were larger than the predicted value.

d. $SSE = S_{yy} - \hat{b}_1 S_{xy} = 9.9153 - (.09092)(45.8246) = 5.7489$.

$r^2 = 1 - \dfrac{SSE}{SST} = 1 - \dfrac{5.7489}{9.9153} = .4202$. (42.02% of the observed variation in efficiency ratio can be attributed to the approximate linear relationship between the efficiency ratio and the tank temperature.)

15.

a. The following stem and leaf display shows that: a typical value for this data is a number in the low 40's. there is some positive skew in the data. There are some potential outliers (79.5 and 80.0), and there is a reasonably large amount of variation in the data (e.g., the spread 80.0-29.8 = 50.2 is large compared with the typical values in the low 40's).

2	9	
3	33	stem = tens
3	5566677889	leaf = ones
4	1223	
4	56689	
5	1	
5		
6	2	
6	9	
7		
7	9	
8	0	

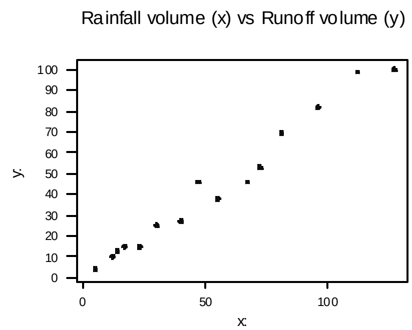
b. No, the strength values are not uniquely determined by the MoE values. For example, note that the two pairs of observations having strength values of 42.8 have different MoE values.

c. The least squares line is $\hat{y} = 3.2925 + .10748x$. For a beam whose modulus of elasticity is x = 40, the predicted strength would be $\hat{y} = 3.2925 + .10748(40) = 7.59$. The value x = 100 isfar beyond the range of the x values in the data, so it would be dangerous (i.e., potentially misleading) to extrapolated the linear relationship that far.

d. From the output, SSE = 18.736, SST = 71.605, and the coefficient of determination is r^2 = .738 (or 73.8%). The r^2 value is large, which suggests that the linear relationship is a useful approximation to the true relationship between these two variables.

16.

a.



Yes, the scatterplot shows a strong linear relationship between rainfall volume and runoff volume, thus it supports the use of the simple linear regression model.

b. $\bar{x} = 53.200$, $\bar{y} = 42.867$, $S_{xx} = 63040 - \frac{(798)^2}{15} = 20{,}586.4$,

$S_{yy} = 41{,}999 - \frac{(643)^2}{15} = 14{,}435.7$, and

$S_{xy} = 51{,}232 - \frac{(798)(643)}{15} = 17{,}024.4$. $\hat{b}_1 = \frac{S_{xy}}{S_{xx}} = \frac{17{,}024.4}{20{,}586.4} = .82697$ and

$\hat{b}_0 = 42.867 - (.82697)53.2 = -1.1278$.

c. $m_{y \cdot 50} = -1.1278 + .82697(50) = 40.2207$.

d. $SSE = S_{yy} - \hat{b}_1 S_{xy} = 14{,}435.7 - (.82697)(17{,}324.4) = 357.07$.

$s = \hat{s} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{357.07}{13}} = 5.24$.

e. $r^2 = 1 - \frac{SSE}{SST} = 1 - \frac{357.07}{14{,}435.7} = .9753$. So 97.53% of the observed variation in runoff volume can be attributed to the simple linear regression relationship between runoff and rainfall.

17. Note: n = 23 in this study.

a. For a one (mg/cm^2) increase in dissolved material, one would expect a .144 (g/l) increase in calcium content. Secondly, 86% of the observed variation in calcium content can be attributed to the simple linear regression relationship between calcium content and dissolved material.

b. $m_{y\cdot 50} = 3.678 + .144(50) = 10.878$

c. $r^2 = .86 = 1 - \dfrac{SSE}{SST}$, so $SSE = (SST)(1-.86) = (320.398)(.14) = 44.85572$.

Then $s = \sqrt{\dfrac{SSE}{n-2}} = \sqrt{\dfrac{44.85572}{21}} = 1.46$

18.

a. $\hat{b}_1 = \dfrac{15(987.645)-(1425)(10.68)}{15(139{,}037.25)-(1425)^2} = \dfrac{-404.3250}{54{,}933.7500} = -.00736023$

$\hat{b}_0 = \dfrac{10.68-(-.00736023)(1425)}{15} = 1.41122185$, $y = 1.4112 - .007360x$.

b. $\hat{b}_1 = -.00736023$

c. With x now denoting temperature in $\text{o}C$, $y = \hat{b}_0 + \hat{b}_1\left(\dfrac{9}{5}x + 32\right)$

$= \left(\hat{b}_0 + 32\hat{b}_1\right) + \dfrac{9}{5}\hat{b}_1 x = 1.175695 - .0132484x$, so the new $\hat{b}_1$ is -.0132484 and the new $\hat{b}_0 = 1.175695$.

d. Using the equation of **a**, predicted $y = \hat{b}_0 + \hat{b}_1(200) = -.0608$, but the deflection factor cannot be negative.

19. N = 14, $\Sigma x_i = 3300$, $\Sigma y_i = 5010$, $\Sigma x_i^2 = 913{,}750$, $\Sigma y_i^2 = 2{,}207{,}100$, $\Sigma x_i y_i = 1{,}413{,}500$

a. $\hat{b}_1 = \dfrac{3{,}256{,}000}{1{,}902{,}500} = 1.71143233$, $\hat{b}_0 = -45.55190543$, so we use the equation $y = -45.5519 + 1.7114x$.

b. $\hat{m}_{Y \cdot 225} = -45.5519 + 1.7114(225) = 339.51$

c. Estimated expected change $= -50\hat{b}_1 = -85.57$

d. No, the value 500 is outside the range of x values for which observations were available (the danger of extrapolation).

20.

a. $\hat{b}_0 = .3651$, $\hat{b}_1 = .9668$

b. .8485

c. $\hat{s} = .1932$

d. SST = 1.4533, 71.7% of this variation can be explained by the model. Note: $\dfrac{SSR}{SST} = \dfrac{1.0427}{1.4533} = .717$ which matches R-squared on output.

21.

a. The summary statistics can easily be verified using Minitab or Excel, etc.

b. $\hat{b}_1 = \dfrac{491.4}{744.16} = .66034186$, $\hat{b}_0 = -2.18247148$

c. predicted $y = \hat{b}_0 + \hat{b}_1(15) = 7.72$

d. $\hat{m}_{Y \cdot 15} = \hat{b}_0 + \hat{b}_1(15) = 7.72$

22.

a. $\hat{b}_1 = \dfrac{-404.325}{54.933.75} = -.00736023$, $\hat{b}_0 = 1.41122185$,
$SSE = 7.8518 - (1.41122185)(10.68) - (-.00736023)(987.654) = .049245$,
$s^2 = \dfrac{.049245}{13} = .003788$, and $\hat{s} = s = .06155$

b. $SST = 7.8518 - \dfrac{(10.68)^2}{15} = .24764$ so $r^2 = 1 - \dfrac{.049245}{.24764} = 1 - .199 = .801$

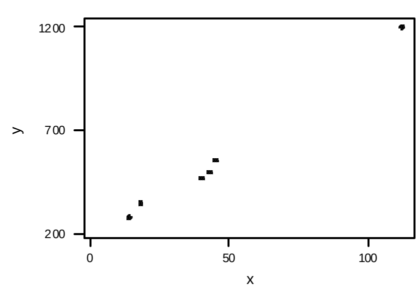
23.

a. Using the y_i's given to one decimal place accuracy is the answer to Exercise 19, $SSE = (150 - 125.6)^2 + ... + (670 - 639.0)^2 = 16,213.64$. The computation formula gives
$SSE = 2,207,100 - (-45.55190543)(5010) - (1.71143233)(1,413,500)$
$= 16,205.45$

b. $SST = 2,207,100 - \dfrac{(5010)^2}{14} = 414,235.71$ so $r^2 = 1 - \dfrac{16,205.45}{414,235.71} = .961$.

24.

a.



According to the scatter plot of the data, a simple linear regression model does appear to be plausible.

b. The regression equation is `y = 138 + 9.31 x`

c. The desired value is the coefficient of determination, $r^2 = 99.0\%$.

d. The new equation is `y* = 190 + 7.55 x*.` This new equation appears to differ significantly. If we were to predict a value of y^* for $x^* = 50$, the value would be 567.9, where using the original data, the predicted value for x = 50 would be 603.5.

25. Substitution of $\hat{b}_0 = \frac{\Sigma y_i - \hat{b}_1 \Sigma x_i}{n}$ and $\hat{b}_1$ for b_o and b_1 on the left hand side of the normal equations yields $\frac{n(\Sigma y_i - \hat{b}_1 \Sigma x_i)}{n} + \hat{b}_1 \Sigma x_i = \Sigma y_i$ from the first equation and

$$\frac{\Sigma x_i (\Sigma y_i - \hat{b}_1 \Sigma x_i)}{n} + \hat{b}_1 \Sigma x_i^2 = \frac{\Sigma x_i \Sigma y_i}{n} + \frac{\hat{b}_1 (n\Sigma x_i^2 - (\Sigma x_i)^2)}{n}$$

$\frac{\Sigma x_i \Sigma y_i}{n} + \frac{n\Sigma x_i y_i}{n} - \frac{\Sigma x_i \Sigma y_i}{n} = \Sigma x_i y_i$ from the second equation.

26. We show that when $\bar{x}$ is substituted for x in $\hat{b}_0 + \hat{b}_1 x$, $\bar{y}$ results, so that $(\bar{x}, \bar{y})$ is on the line $y = \hat{b}_0 + \hat{b}_1 x$: $\hat{b}_0 + \hat{b}_1 \bar{x} = \frac{\Sigma y_i - b_1 \Sigma x_i}{n} + \hat{b}_1 \bar{x} = \bar{y} - \hat{b}_1 \bar{x} + \hat{b}_1 \bar{x} = \bar{y}$.

27. We wish to find b_1 to minimize $\Sigma(y_i - b_1 x_i)^2 = f(b_1)$. Equating $f'(b_1)$ to 0 yields $2\Sigma(y_i - b_1 x_i)(-x_i) = 0$ so $\Sigma x_i y_i = b_1 \Sigma x_i^2$ and $b_1 = \frac{\Sigma x_i y_i}{\Sigma x_i^2}$. The least squares estimator of $\hat{b}_1$ is thus $\hat{b}_1 = \frac{\Sigma x_i Y_i}{\Sigma x_i^2}$.

28.

a. Subtracting $\bar{x}$ from each x_i shifts the plot in a rigid fashion $\bar{x}$ units to the left without otherwise altering its character. The last squares line for the new plot will thus have the same slope as the one for the old plot. Since the new line is $\bar{x}$ units to the left of the old one, the new y intercept (height at x = 0) is the height of the old line at x = $\bar{x}$, which is $\hat{b}_0 + \hat{b}_1 \bar{x} = \bar{y}$ (since from exercise 26, $(\bar{x}, \bar{y})$ is on the old line). Thus the new y intercept is $\bar{y}$.

b. We wish b_0 and b_1 to minimize $f(b_0, b_1) = \Sigma[y_i - (b_0 + b_1(x_i - \bar{x}))]^2$. Equating $\frac{\partial f}{\partial b_0}$ to $\frac{\partial f}{\partial b_1}$ to 0 yields $nb_0 + b_1 \Sigma(x_i - \bar{x}) = \Sigma y_i$, $b_0 \Sigma(x_i - \bar{x}) + b_1 \Sigma(x_i - \bar{x})^2 = \Sigma(x_i - \bar{x})^2 = \Sigma(x_i - \bar{x}) y_i$. Since $\Sigma(x_i - \bar{x}) = 0$, $b_0 = \bar{y}$, and since $\Sigma(x_i - \bar{x}) y_i = \Sigma(x_i - \bar{x})(y_i - \bar{y})$ [because $\Sigma(x_i - \bar{x})\bar{y} = \bar{y}\Sigma(x_i - \bar{x})$], $b_1 = \hat{b}_1$. Thus $\hat{b}_0^* = \bar{Y}$ and $\hat{b}_1^* = \hat{b}_1$.

29. For data set #1, r^2 = .43 and $\hat{s} = s = 4.03$; whereas these quantities are .99 and 4.03 for #2, and .99 and 1.90 for #3. In general, one hopes for both large r^2 (large % of variation explained) and small s (indicating that observations don't deviate much from the estimated line). Simple linear regression would thus seem to be most effective in the third situation.

Section 12.3

30.

a. $\Sigma(x_i - \bar{x})^2 = 7{,}000{,}000$, so $V(\hat{b}_1) = \dfrac{(350)^2}{7{,}000{,}000} = .0175$ and the standard deviation of $\hat{b}_1$ is $\sqrt{.0175} = .1323$.

b. $P(1.0 \le \hat{b}_1 \le 1.5) = P\left(\dfrac{1.0 - 1.25}{1.323} \le Z \le \dfrac{1.5 - 1.25}{1.323}\right)$
$= P(-1.89 \le Z \le 1.89) = .9412$.

c. Although n = 11 here and n = 7 in **a**, $\Sigma(x_i - \bar{x})^2 = 1{,}100{,}000$ now, which is smaller than in **a**. Because this appears in the denominator of $V(\hat{b}_1)$, the variance is smaller for the choice of x values in **a**.

31.

a. $\hat{b}_1 = -.00736023$, $\hat{b}_0 = 1.41122185$, so
$SSE = 7.8518 - (1.41122185)(10.68) - (-.00736023)(987.645) = .04925$,
$s^2 = .003788$, $s = .06155$. $\hat{s}^2_{\hat{b}_1} = \dfrac{s^2}{\Sigma x_i^2 - (\Sigma x_i)^2 / n} = \dfrac{.003788}{3662.25} = .00000103$,

$$\hat{s}_{\hat{b}_1} = s_{\hat{b}_1} = \text{estimated s.d. of } \hat{b}_1 = \sqrt{.00000103} = .001017.$$

b. $-.00736 \pm (2.160)(.001017) = -.00736 \pm .00220 = (-.00956, -.00516)$

32. Let b_1 denote the true average change in runoff for each 1 m^3 increase in rainfall. To test the hypotheses $H_o : b_1 = 0$ vs. $H_a : b_1 \neq 0$, the calculated t statistic is

$t = \frac{\hat{b}_1}{s_{\hat{b}_1}} = \frac{.82697}{.03652} = 22.64$ which (from the printout) has an associated p-value of P = 0.000. Therefore, since the p-value is so small, H_o is rejected and we conclude that there is a useful linear relationship between runoff and rainfall.

A confidence interval for b_1 is based on n – 2 = 15 – 2 = 13 degrees of freedom.

$t_{.025,13} = 2.160$, so the interval estimate is

$\hat{b}_1 \pm t_{.025,13} \cdot s_{\hat{b}_1} = .82697 \pm (2.160)(.03652) = (.748,.906)$. Therefore, we can be confident that the true average change in runoff, for each 1 m^3 increase in rainfall, is somewhere between .748 m^3 and .906 m^3.

33.

a. From the printout in Exercise 15, the error d.f. = n – 2 = 25, $t_{.025,25} = 2.060$. The confidence interval is then

$\hat{b}_1 \pm t_{.025,25} \cdot s_{\hat{b}_1} = .10748 \pm (2.060)(.01280) = (.081,.134)$. Therefore, we estimate with a high degree of confidence that the true average change in strength associated with a 1 Gpa increase in modulus of elasticity is between .081 MPa and .134 MPa.

b. We wish to test $H_o : b_1 = .1$ vs. $H_a : b_1 > .1$. The calculated t statistic is

$t = \frac{\hat{b}_1 - .1}{s_{\hat{b}_1}} = \frac{.10748 - .1}{.01280} = .58$, which yields a p-value of .277. A large p-value such as this would not lead to rejecting H_o, so there is not enough evidence to contradict the prior belief.

34.

a. $H_o : b_1 = 0$; $H_a : b_1 \neq 0$

RR: $|t| > t_{a/2,n-2}$ or $|t| > 3.106$

$t = 5.29$: Reject H_o. The slope differs significantly from 0, and the model appears to be useful.

b. At the level $a = 0.01$, reject h_o if the p-value is less than 0.01. In this case, the reported p-value was 0.000, therefore reject H_o. The conclusion is the same as that of part **a**.

c. $H_o : b_1 = 1.5$; $H_a : b_1 < 1.5$

RR: $t < -t_{a,n-2}$ or $t < -2.718$

$t = \frac{0.9668 - 1.5}{0.1829} = -2.92$: Reject H_o. The data contradict the prior belief.

35.

a. We want a 95% CI for β1: $\hat{b}_1 \pm t_{.025,15} \cdot s_{\hat{b}_1}$. First, we need our point estimate, $\hat{b}_1$.

Using the given summary statistics, $S_{xx} = 3056.69 - \frac{(222.1)^2}{17} = 155.019$,

$S_{xy} = 2759.6 - \frac{(222.1)(193)}{17} = 238.112$, and $\hat{b}_1 = \frac{S_{xy}}{S_{xx}} = \frac{238.112}{115.019} = 1.536$.

We need $\hat{b}_0 = \frac{193 - (1.536)(222.1)}{17} = -8.715$ to calculate the SSE:

$SSE = 2975 - (-8.715)(193) - (1.536)(2759.6) = 418.2494$. Then

$s = \sqrt{\frac{418.2494}{15}} = 5.28$ and $s_{\hat{b}_1} = \frac{5.28}{\sqrt{155.019}} = .424$. With $t_{.025,15} = 2.131$, our CI is $1.536 \pm 2.131 \cdot (.424)$ = (.632, 2.440). With 95% confidence, we estimate that the change in reported nausea percentage for every one-unit change in motion sickness dose is between .632 and 2.440.

b. We test the hypotheses $H_o : b_1 = 0$ vs $H_a : b_1 \neq 0$, and the test statistic is

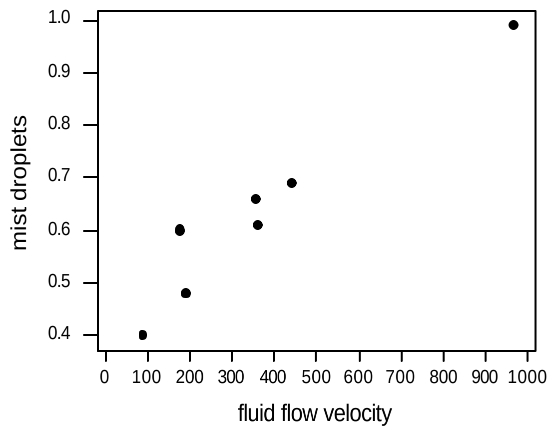
$t = \frac{1.536}{.424} = 3.6226$. With df=15, the two-tailed p-value = 2P(t > 3.6226) = 2(.001) = .002. With a p-value of .002, we would reject the null hypothesis at most reasonable significance levels. This suggests that there is a useful linear relationship between motion sickness dose and reported nausea.

c. No. A regression model is only useful for estimating values of nausea % when using dosages between 6.0 and 17.6 – the range of values sampled.

d. Removing the point (6.0, 2.50), the new summary stats are: n = 16, , $\Sigma x_i = 216.1$, $\Sigma y_i = 191.5$, $\Sigma x_i^2 = 3020.69$, $\Sigma y_i^2 = 2968.75$, $\Sigma x_i y_i = 2744.6$, and then $\hat{b}_1 = 1.561$, $\hat{b}_0 = -9.118$, SSE = 430.5264, $s = 5.55$, $s_{\hat{b}_1} = .551$, and the new CI is $1.561 \pm 2.145 \cdot (.551)$, or (.379, 2.743). The interval is a little wider. But removing the one observation did not change it that much. The observation does not seem to be exerting undue influence.

36.

a. A scatter plot, generated by Minitab, supports the decision to use linear regression analysis.



b. We are asked for the coefficient of determination, r^2. From the Minitab output, $r^2 = .931$ (which is close to the hand calculated value, the difference being accounted for by round-off error.)

c. Increasing x from 100 to 1000 means an increase of 900. If, as a result, the average y were to increase by .6, the slope would be $\frac{.6}{900} = .0006667$. We should test the hypotheses $H_o : b_1 = .0006667$ vs. $H_a : b_1 < .0006667$. The test statistic is $t = \frac{.00062108 - .0006667}{.00007579} = -.601$, which is not significant. There is not sufficient evidence that with an increase from 100 to 1000, the true average increase in y is less than .6.

d. We are asked for a confidence interval for b_1. Using the values from the Minitab output, we have $.00062108 \pm 2.776(.00007579) = (.00041069, .00083147)$

37.

a. n = 10, $\Sigma x_i = 2615$, $\Sigma y_i = 39.20$, $\Sigma x_i^2 = 860{,}675$, $\Sigma y_i^2 = 161.94$,

$\Sigma x_i y_i = 11{,}453.5$, so $\hat{b}_1 = \dfrac{12{,}027}{1{,}768{,}525} = .00680058$, $\hat{b}_0 = 2.14164770$, from which SSE = .09696713, s = .11009492 $s = .11009492 \approx .110 = \hat{s}$,

$\hat{s}_{\hat{b}_1} = \dfrac{.110}{\sqrt{176{,}852}} = .000262$

b. We wish to test $H_o : b_1 = .0060$ vs $H_a : b_1 \neq .0060$. At level .10, H_o is rejected if either $t \geq t_{.05,8} = 1.860$ or $t \leq -t_{.05,8} = -1.860$. Since

$t = \dfrac{.0068 - .0060}{.000262} = 3.06 \geq 1.1860$, H_o is rejected.

38.

a. From Exercise 23, which also refers to Exercise 19, SSE = 16.205.45, so

$s^2 = 1350.454$, $s = 36.75$, and $s_{\hat{b}_1} = \dfrac{36.75}{368.636} = .0997$. Thus

$t = \dfrac{1.711}{.0997} = 17.2 > 4.318 = t_{.0005,14}$, so p-value < .001. Because the p-value < .01,

$H_o : b_1 = 0$ is rejected at level .01 in favor of the conclusion that the model is useful

$(b_1 \neq 0)$.

b. The C.I. for b_1 is $1.711 \pm (2.179)(.0997) = 1.711 \pm .217 = (1.494, 1.928)$. Thus the C.I. for $10b_1$ is $(14.94, 19.28)$.

39. SSE = 124,039.58– (72.958547)(1574.8) – (.04103377)(222657.88) = 7.9679, and SST = 39.828

Source	**df**	**SS**	**MS**	**f**
Regr	1	31.860	31.860	18.0
Error	18	7.968	1.77	
Total	19	39.828		

Let's use α = .001. Then $F_{.001,1,18} = 15.38 < 18.0$, so $H_o : b_1 = 0$ is rejected and the model is judged useful. $s = \sqrt{1.77} = 1.33041347$, $S_{xx} = 18{,}921.8295$, so

$t = \dfrac{.04103377}{1.33041347 / \sqrt{18{,}921.8295}} = 4.2426$, and $t^2 = (4.2426)^2 = 18.0 = f$.

40. We use the fact that $\hat{b}_1$ is unbiased for b_1. $E\left(\hat{b}_0\right) = \dfrac{E\left(\Sigma y_i - \hat{b}_1 \Sigma x_i\right)}{n}$

$$= \frac{E(\Sigma y_i)}{n} - E\left(\hat{b}_1\right)\overline{x} = \frac{E(\Sigma Y_i)}{n} - b_1\overline{x}$$

$$= \frac{\Sigma(b_0 + b_1 x_i)}{n} - b_1\overline{x} = b_0 + b_1\overline{x} - b_1\overline{x} = b_0$$

41.

a. Let $c = n\Sigma x_i^2 - (\Sigma x_i)^2$. Then $E\left(\hat{b}_1\right) = \dfrac{1}{c} E\left[n\Sigma x_i Y_i \ldots Y_i - (\Sigma x_i)\ldots(\Sigma x_i)(\Sigma Y_i)\right]$

$$= \frac{n}{c}\sum x_i E(Y_i) - \frac{\Sigma x_i}{c}\sum E(Y_i) = \frac{n}{c}\sum x_i\left(b_0 + b_1 x_i\right) - \frac{\Sigma x_i}{c}\sum\left(b_0 + b_1 x_i\right)$$

$$\frac{b_1}{c}\left[n\Sigma x_i^2 - (\Sigma x_i)^2\right] = b_1.$$

b. With $c = \Sigma\left(x_i - \overline{x}\right)^2$, $\hat{b}_1 = \dfrac{1}{c}\Sigma\left(x_i - \overline{x}\right)\left(Y_i - \overline{Y}\right) = \dfrac{1}{c}\Sigma\left(x_i - \overline{x}\right)Y_i$ (since

$\Sigma\left(x_i - \overline{x}\right)\overline{Y} = \overline{Y}\Sigma\left(x_i - \overline{x}\right) = \overline{Y}\cdot 0 = 0$), so $V\left(\hat{b}_1\right) = \dfrac{1}{c^2}\Sigma\left(x_i - \overline{x}\right)^2 Var(Y_i)$

$$= \frac{1}{c^2}\Sigma\left(x_i - \overline{x}\right)^2 \cdot s^2 = \frac{s^2}{\Sigma\left(x_i - \overline{x}\right)^2} = \frac{s^2}{\Sigma x_i^2 - (\Sigma x_i)^2 / n}$$, as desired.

42. $t = \hat{b}_1 \dfrac{\sqrt{\Sigma x_i^2 - (\Sigma x_i)^2 / n}}{s}$. The numerator of $\hat{b}_1$ will be changed by the factor cd (since both $\Sigma x_i y_i$ and $(\Sigma x_i)(\Sigma y_i)$ appear) while the denominator of $\hat{b}_1$ will change by the factor c^2 (since both Σx_i^2 and $(\Sigma x_i)^2$ appear). Thus $\hat{b}_1$ will change by the factor d/c. Because $SSE = \Sigma\left(y_i - \hat{y}_i\right)^2$, SSE will change by the factor d^2, so s will change by the factor d. Since $\sqrt{\bullet}$ in t changes by the factor c, t itself will change by $\dfrac{d}{c}\cdot\dfrac{c}{d} = 1$, or not at all.

43. The numerator of d is |1 – 2| = 1, and the denominator is $\dfrac{4\sqrt{14}}{\sqrt{324.40}} = .831$, so $d = \dfrac{1}{.831} = 1.20$. The approximate power curve is for n – 2 df = 13, and b is read from Table A.17 as approximately .1.

Section 12.4

44.

a. The mean of the x data in Exercise 12.15 is $\bar{x} = 45.11$. Since x = 40 is closer to 45.11 than is x = 60, the quantity $(40 - \bar{x})^2$ must be smaller than $(60 - \bar{x})^2$. Therefore, since these quantities are the only ones that are different in the two $s_{\hat{y}}$ values, the $s_{\hat{y}}$ value for x = 40 must necessarily be smaller than the $s_{\hat{y}}$ for x = 60. Said briefly, the closer x is to $\bar{x}$, the smaller the value of $s_{\hat{y}}$.

b. From the printout in Exercise 12.15, the error degrees of freedom is df = 25. $t_{.025,25} = 2.060$, so the interval estimate when x = 40 is : $7.592 \pm (2.060)(.179)$ $7.592 \pm .369 = (7.223, 7.961)$. We estimate, with a high degree of confidence, that the true average strength for all beams whose MoE is 40 GPa is between 7.223 MPa and 7.961 MPa.

c. From the printout in Exercise 12.15, s = .8657, so the 95% prediction interval is

$$\hat{y} \pm t_{.025,25}\sqrt{s^2 + s_{\hat{y}}^2} = 7.592 \pm (2.060)\sqrt{(.8657)^2 + (.179)^2}$$

$= 7.592 \pm 1.821 = (5.771, 9.413)$. Note that the prediction interval is almost 5 times as wide as the confidence interval.

d. For two 95% intervals, the simultaneous confidence level is at least 100(1 – 2(.05)) = 90%

45.

a. We wish to find a 90% CI for $m_{y\cdot 125}$: $\hat{y}_{125} = 78.088$, $t_{.05,18} = 1.734$, and

$$s_{\hat{y}} = s\sqrt{\frac{1}{20} + \frac{(125 - 140.895)^2}{18{,}921.8295}} = .3349$$

.Putting it together, we get

$$78.088 \pm 1.734(.3349) = (77.5073, 78.6687)$$

b. We want a 90% PI: Only the standard error changes:

$$s_{\hat{y}} = s\sqrt{1 + \frac{1}{20} + \frac{(125 - 140.895)^2}{18{,}921.8295}} = 1.3719$$

, so the PI is

$$78.088 \pm 1.734(1.3719) = (75.7091, 80.4669)$$

c. Because the x^* of 115 is farther away from $\bar{x}$ than the previous value, the term $(x^* - \bar{x})^2$ will be larger, making the standard error larger, and thus the width of the interval is wider.

d. We would be testing to see if the filtration rate were 125 kg-DS/m/h, would the average moisture content of the compressed pellets be less than 80%. The test statistic is $t = \dfrac{78.088 - 80}{.3349} = -5.709$, and with 18 df the p-value is P(t<-5.709) ˜ 0.00. We would reject H_o. There is significant evidence to prove that the true average moisture content when filtration rate is 125 is less than 80%.

46.

a. A 95% CI for $m_{Y \cdot 500}$: $\hat{y}_{(500)} = -.311 + (.00143)(500) = .40$ and

$$s_{\hat{y}_{(500)}} = .131\sqrt{\frac{1}{13} + \frac{(500 - 471.54)^2}{131{,}519.23}} = .03775$$, so the interval is

$$\hat{y}_{(500)} \pm t_{.025,11} \cdot S_{\hat{y}_{(500)}} = .40 \pm 2.210(.03775) = .40 \pm .08 = (.32,.48)$$

b. The width at x = 400 will be wider than that of x = 500 because x = 400 is farther away from the mean ($\bar{x} = 471.54$).

c. A 95% CI for b_1:

$$\hat{b}_1 \pm t_{.025,11} \cdot s_{\hat{b}_1} = .00143 \pm 2.201(.0003602) = (.000637,.002223)$$

d. We wish to test $H_0 : y_{(400)} = .25$ vs. $H_0 : y_{(400)} \neq .25$. The test statistic is $t = \dfrac{\hat{y}_{(400)} - .25}{s_{\hat{y}_{(400)}}}$, and we reject H_o if $|t| \geq t_{.025,11} = 2.201$.

$\hat{y}_{(400)} = -.311 + .00143(400) = .2614$ and

$$s_{\hat{y}_{(400)}} = .131\sqrt{\frac{1}{13} + \frac{(400 - 471.54)^2}{131{,}519.23}} = .0445$$, so the calculated

$t = \dfrac{.2614 - .25}{.0445} = .2561$, which is not ≥ 2.201, so we do not reject H_o. This sample data does not contradict the prior belief.

47.

a. $\hat{y}_{(40)} = -1.128 + .82697(40) = 31.95$, $t_{.025,13} = 2.160$; a 95% PI for runoff is $31.95 \pm 2.160\sqrt{(5.24)^2 + (1.44)^2} = 31.95 \pm 11.74 = (20.21, 43.69)$. No, the resulting interval is very wide, therefore the available information is not very precise.

b. $\Sigma x = 798, \Sigma x^2 = 63{,}040$ which gives $S_{xx} = 20{,}586.4$, which in turn gives $s_{\hat{y}(50)} = 5.24\sqrt{\frac{1}{15} + \frac{(50 - 53.20)^2}{20{,}586.4}} = 1.358$, so the PI for runoff when x = 50 is $40.22 \pm 2.160\sqrt{(5.24)^2 + (1.358)^2} = 40.22 \pm 11.69 = (28.53, 51.92)$. The simultaneous prediction level for the two intervals is at least $100(1 - 2a)\% = 90\%$.

48.

a. $S_{xx} = 18.24 - \frac{(12.6)^2}{9} = .60$, $S_{xy} = 40.968 - \frac{(12.6)(27.68)}{9} = 2.216$;

$S_{yy} = 93.3448 - \frac{(27.68)^2}{9} = 8.213$ $\hat{b}_1 = \frac{S_{xy}}{S_{xx}} = \frac{2.216}{.60} = 3.693$;

$\hat{b}_0 = \frac{\Sigma y - \hat{b}_1 \Sigma x}{n} = \frac{27.68 - (3.693)(12.6)}{9} = -2.095$, so the point estimate is $\hat{y}_{(1.5)} = -2.095 + 3.693(1.5) = 3.445$. $SSE = 8.213 - 3.693(2.216) = .0293$, which yields $s = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{.0293}{7}} = .0647$. Thus

$s_{\hat{y}(1.5)} = .0647\sqrt{\frac{1}{9} + \frac{(1.5 - 1.4)^2}{.60}} = .0231$. The 95% CI for $m_{y \cdot 1.5}$ is $3.445 \pm 2.365(.0231) = 3.445 \pm .055 = (3.390, 3.50)$.

b. A 95% PI for y when x = 1.5 is similar: $3.445 \pm 2.365\sqrt{(.0647)^2 + (.0231)^2} = 3.445 \pm .162 = (3.283, 3.607)$. The prediction interval for a future y value is wider than the confidence interval for an average value of y when x is 1.5.

c. A new PI for y when x = 1.2 will be wider since x = 1.2 is farther away from the mean $\bar{x} = 1.4$.

49. 95% CI: (462.1, 597.7); midpoint = 529.9; $t_{.025,8} = 2.306$;

$529.9 + (2.306)\left(\hat{s}_{\hat{b}_0+\hat{b}_1(15)}\right) = 597.7$

$\hat{s}_{\hat{b}_0+\hat{b}_1(15)} = 29.402$

99% CI: $529.9 \pm (3.355)(29.402) = (431.3, 628.5)$

50. $\hat{b}_1 = 18.87349841$, $\hat{b}_0 = -8.77862227$, SSE = 2486.209, s = 16.6206

a. $\hat{b}_0 + \hat{b}_1(18) = 330.94$, $\bar{x} = 20.2909$, $\sqrt{\dfrac{1}{11} + \dfrac{11(18-20.2909)^2}{3834.26}} = .3255$,

$t_{.025,9} = 2.262$, so the CI is $330.94 \pm (2.262)(16.6206)(.3255)$

$= 330.94 \pm 12.24 = (318.70, 343.18)$

b. $\sqrt{1 + \dfrac{1}{11} + \dfrac{11(18-20.2909)^2}{3834.26}} = 1.0516$, so the P.I. is

$330.94 \pm (2.262)(16.6206)(1.0516) = 330.94 \pm 39.54 = (291.40, 370.48)$.

c. To obtain simultaneous confidence of at least 97% for the three intervals, we compute each one using confidence level 99%, (with $t_{.005,9} = 3.250$). For x = 15, the interval is $274.32 \pm 22.35 = (251.97, 296.67)$. For x = 18, $330.94 \pm 17.58 = (313.36, 348.52)$. For x = 20, $368.69 \pm 0.84 = (367.85, 369.53)$.

51.

a. 0.40 is closer to $\bar{x}$.

b. $\hat{b}_0 + \hat{b}_1(0.40) \pm t_{a/2,n-2} \cdot \left(\hat{s}_{\hat{b}_0+\hat{b}_1(0.40)}\right)$ or $0.8104 \pm (2.101)(0.0311)$

$= (0.745, 0.876)$

c. $\hat{b}_0 + \hat{b}_1(1.20) \pm t_{a/2,n-2} \cdot \sqrt{s^2 + s^2_{\hat{b}_0+\hat{b}_1(1.20)}}$ or

$0.2912 \pm (2.101) \cdot \sqrt{(0.1049)^2 + (0.0352)^2} = (.059, .523)$

52.

a. We wish to test $H_o : b_1 = 0$ vs $H_a : b_1 \neq 0$. The test statistic $t = \dfrac{10.6026}{.9985} = 10.62$ leads to a p-value of < .006 (2P(t > 4.0) from the 7 df row of table A.8), and H_o is rejected since the p-value is smaller than any reasonable a. The data suggests that this model does specify a useful relationship between chlorine flow and etch rate.

b. A 95% confidence interval for b_1: $10.6026 \pm (2.365)(.9985) = (8.24, 12.96)$. We can be highly confident that when the flow rate is increased by 1 SCCM, the associated expected change in etch rate will be between 824 and 1296 A/min.

c. A 95% CI for $m_{Y \cdot 3.0}$: $38.256 \pm 2.365\left(2.546\sqrt{\dfrac{1}{9} + \dfrac{9(3.0 - 2.667)^2}{58.50}}\right)$

$= 38.256 \pm 2.365(2.546)(.35805) = 38.256 \pm 2.156 = (36.100, 40.412)$, or 3610.0 to 4041.2 A/min.

d. The 95% PI is $38.256 \pm 2.365\left(2.546\sqrt{1 + \dfrac{1}{9} + \dfrac{9(3.0 - 2.667)^2}{58.50}}\right)$

$= 38.256 \pm 2.365(2.546)(1.06) = 38.256 \pm 6.398 = (31.859, 44.655)$, or 3185.9 to 4465.5 A/min.

e. The intervals for x* = 2.5 will be narrower than those above because 2.5 is closer to the mean than is 3.0.

f. No. a value of 6.0 is not in the range of observed x values, therefore predicting at that point is meaningless.

53. Choice **a** will be the smallest, with d being largest. **a** is less than **b** and **c** (obviously), and **b** and **c** are both smaller than **d**. Nothing can be said about the relationship between **b** and **c**.

54.

a. There is a linear pattern in the scatter plot, although the pot also shows a reasonable amount of variation about any straight line fit to the data. The simple linear regression model provides a sensible starting point for a formal analysis.

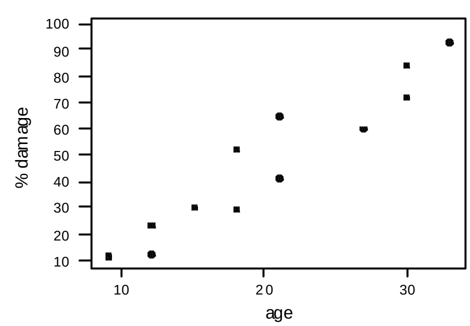
b. n = 141, $\Sigma x_i = 1185, \Sigma x_i^2 = 151{,}825, \Sigma y_i = 5960, \Sigma y_i^2 = 2{,}631{,}200$, and $\Sigma x_i y_i = 449{,}850$, from which $\hat{b}_1 = -1.060132$, $\hat{b}_0 = 515.446887$, $SSE = 36{,}036.93$, $r^2 = .616, s^2 = 3003.08, s = 54.80, s_{\hat{b}_1} = \dfrac{54.80}{\sqrt{51{,}523.21}} = .241$ $H_o : b_1 = 0$ vs $H_a : b_1 \neq 0$, $t = \dfrac{\hat{b}_1}{s_{\hat{b}_1}}$. Reject H$_o$ at level .05 if either $t \geq t_{.025,12} = 2.179$ or $t \leq -2.179$. We calculate $t = \dfrac{-1.060}{.241} = -4.39$. Since $-4.39 \leq -2.179$ H$_o$ is rejected. The simple linear regression model does appear to specify a useful relationship.

c. A confidence interval for $b_0 + b_1(75)$ is requested. The interval is centered at $\hat{b}_0 + \hat{b}_1(75) = 435.9$. $s_{\hat{b}_0 + \hat{b}_1(75)} = s\sqrt{\dfrac{1}{n} + \dfrac{n(75-\bar{x})^2}{n\Sigma x_i^2 - (\Sigma x_i)^2}} = 14.83$ (using s = 54.803). Thus a 95% CI is $435.9 \pm (2.179)(14.83) = (403.6, 559.7)$.

55.

a. $x_2 = x_3 = 12$, yet $y_2 \neq y_3$

b.



Based on a scatterplot of the data, a simple linear regression model does seem a reasonable way to describe the relationship between the two variables.

c. $\hat{b}_1 = \frac{2296}{699} = 3.284692$, $\hat{b}_0 - 19.669528$, $y = -19.67 + 3.285x$

d. $SSE = 35{,}634 - (-19.669528)(572) - (3.284692)(14{,}022) = 827.0188$,

$s^2 = 82.70188, s = 9.094$. $s_{\hat{b}_0 + \hat{b}_1(20)} = 9.094\sqrt{\frac{1}{12} + \frac{12(20-20.5)^2}{8388}} = 2.6308$,

$\hat{b}_0 + \hat{b}_1(20) = 46.03$, $t_{.025,10} = 2.228$. The PI is $46.03 \pm 2.228\sqrt{s^2 + s^2_{\hat{b}_0 + \hat{b}_1(20)}}$

$= 46.03 \pm 21.09 = (24.94, 67.12)$.

56. $\hat{b}_0 + \hat{b}_1 x = \overline{Y} - \hat{b}_1\overline{x} + \hat{b}_1 x = \overline{Y} + (x - \overline{x})\hat{b}_1 = \frac{1}{n}\sum Y_i + \frac{(x-\overline{x})\sum(x_i - \overline{x})Y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \Sigma d_i Y_i$

where $d_i = \frac{1}{n} + \frac{(x-\overline{x})(x_i - \overline{x})}{n\Sigma x_i^2 - (\Sigma x_i)^2}$. Thus $Var(\hat{b}_0 + \hat{b}_1 x) = \sum d_i^2 Var(Y_i) = s^2 \Sigma d_i^2$,

which, after some algebra, yields the desired expression.

Section 12.5

57. Most people acquire a license as soon as they become eligible. If, for example, the minimum age for obtaining a license is 16, then the time since acquiring a license, y, is usually related to age by the equation $y \approx x - 16$, which is the equation of a straight line. In other words, the majority of people in a sample will have y values that closely follow the line $y = x - 16$.

58.

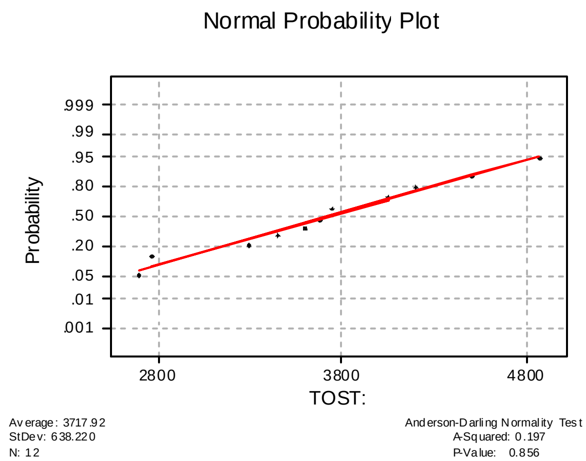
a. Summary values: $\Sigma x = 44{,}615$, $\Sigma x^2 = 170{,}355{,}425$, $\Sigma y = 3{,}860$, $\Sigma y^2 = 1{,}284{,}450$, $\Sigma xy = 14{,}755{,}500$, $n = 12$. Using these values we calculate $S_{xx} = 4{,}480{,}572.92$, $S_{yy} = 42{,}816.67$, and $S_{xy} = 404{,}391.67$. So

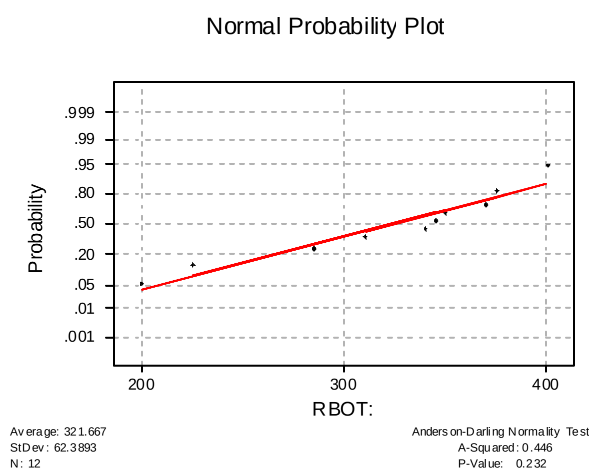
$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} = .9233.$$

b. The value of r does not depend on which of the two variables is labeled as the x variable. Thus, had we let x = RBOT time and y = TOST time, the value of r would have remained the same.

c. The value of r does no depend on the unit of measure for either variable. Thus, had we expressed RBOT time in hours instead of minutes, the value of r would have remained the same.

d.





Both TOST time and ROBT time appear to have come from normally distributed populations.

e. $H_o : r_1 = 0$ vs $H_a : r \neq 0$. $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$; Reject H_o at level .05 if either $t \geq t_{.025,10} = 2.228$ or $t \leq -2.228$. r = .923, t = 7.58, so H_o should be rejected. The model is useful.

59.

a. $S_{xx} = 251{,}970 - \frac{(1950)^2}{18} = 40{,}720$, $S_{yy} = 130.6074 - \frac{(47.92)^2}{18} = 3.033711$, and $S_{xy} = 5530.92 - \frac{(1950)(47.92)}{18} = 339.586667$, so $r = \frac{339.586667}{\sqrt{40{,}720}\sqrt{3.033711}} = .9662$. There is a very strong positive correlation between the two variables.

b. Because the association between the variables is positive, the specimen with the larger shear force will tend to have a larger percent dry fiber weight.

c. Changing the units of measurement on either (or both) variables will have no effect on the calculated value of r, because any change in units will affect both the numerator and denominator of r by exactly the same multiplicative constant.

d. $r^2 = (.966)^2 = .933$

e. $H_o : r = 0$ vs $H_a : r > 0$. $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$; Reject H_o at level .01 if $t \ge t_{.01,16} = 2.583$. $t = \frac{.966\sqrt{16}}{\sqrt{1-.966^2}} = 14.94 \ge 2.583$, so H_o should be rejected . The data indicates a positive linear relationship between the two variables.

60. $H_o : r = 0$ vs $H_a : r \ne 0$. $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$; Reject H_o at level .01 if either $t \ge t_{.005,22} = 2.819$ or $t \le -2.819$. r = .5778, t = 3.32, so H_o should be rejected. There appears to be a non-zero correlation in the population.

61.

a. We are testing $H_o : r = 0$ vs $H_a : r > 0$.
$r = \frac{7377.704}{\sqrt{36.9839}\sqrt{2{,}628{,}930.359}} = .7482$, and $t = \frac{.7482\sqrt{12}}{\sqrt{1-.7482^2}} = 3.9066$. We reject H_o since $t = 3.9066 \ge t_{.05,12} = 1.782$. There is evidence that a positive correlation exists between maximum lactate level and muscular endurance.

b. We are looking for r^2, the coefficient of determination. $r^2 = (.7482)^2 = .5598$. It is the same no matter which variable is the predictor.

62.

a. $H_o : r_1 = 0$ vs $H_a : r \neq 0$, Reject H$_o$ if; Reject H$_o$ at level .05 if either $t \geq t_{.025,12} = 2.179$ or $t \leq -2.179$. $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{(.449)\sqrt{12}}{\sqrt{1-(.449)^2}} = 1.74$. Fail to reject H$_o$, the data does not suggest that the population correlation coefficient differs from 0.

b. $(.449)^2 = .20$ so 20 percent of the observed variation in gas porosity can be accounted for by variation in hydrogen content.

63. n = 6, $\Sigma x_i = 111.71, \Sigma x_i^2 = 2{,}724.7643, \Sigma y_i = 2.9, \Sigma y_i^2 = 1.6572$, and $\Sigma x_i y_i = 63.915$.

$$r = \frac{(6)(63.915)-(111.71)(2.9)}{\sqrt{(6)(2{,}724.7943)-(111.73)^2}\cdot\sqrt{(6)(1.6572)-(2.9)^2}} = .7729$$

. $H_o : r_1 = 0$ vs $H_a : r \neq 0$; Reject H$_o$ at level .05 if $|t| \geq t_{.025,4} = 2.776$.

$$t = \frac{(.7729)\sqrt{4}}{\sqrt{1-(.7729)^2}} = 2.436$$

. Fail to reject H$_o$. The data does not indicate that the population correlation coefficient differs from 0. This result may seem surprising due to the relatively large size of r (.77), however, it can be attributed to a small sample size (6).

64. $r = \dfrac{-757.6423}{\sqrt{(3756.96)(465.34)}} = -.5730$

a. $v = .5\ln\left(\dfrac{.427}{1.573}\right) = -.652$, so (12.11) is $-.652 \pm \dfrac{(1.645)}{\sqrt{26}} = (-.976, -.3290)$, and the desired interval for r is $(-.751, -.318)$.

b. $z = (-.652 + .549)\sqrt{23} = -.49$, so H$_o$ cannot be rejected at any reasonable level.

c. $r^2 = .328$

d. Again, $r^2 = .328$

65.

a. Although the normal probability plot of the x's appears somewhat curved, such a pattern is not terribly unusual when n is small; the test of normality presented in section 14.2 (p. 625) does not reject the hypothesis of population normality. The normal probability plot of the y's is much straighter.

b. $H_o : r_1 = 0$ will be rejected in favor of $H_a : r \neq 0$ at level .01 if $|t| \geq t_{.005,8} = 3.355$. $\Sigma x_i = 864, \Sigma x_i^2 = 78{,}142, \Sigma y_i = 138.0, \Sigma y_i^2 = 1959.1$ and $\Sigma x_i y_i = 12{,}322.4$, so $r = \frac{3992}{(186.8796)(23.3880)} = .913$ and $t = \frac{.913(2.8284)}{.4080} = 6.33 \geq 3.355$, so reject H_o. There does appear to be a linear relationship.

66.

a. We used Minitab to calculate the r_i's: $r_1 = 0.192$, $r_2 = 0.382$, and $r_3 = 0.183$. It appears that the lag 2 correlation is best, but all of them are weak, based on the definitions given in the text.

b. $\frac{2}{\sqrt{100}} = .2$. We reject H_o if $|r_i| \geq .2$. For all lags, r_i does not fall in the rejection region, so we cannot reject H_o. There is not evidence of theoretical autocorrelation at the first 3 lags.

c. If we want an approximate .05 significance level for the simultaneous hypotheses, we would have to use smaller individual significance level. If the individual confidence levels were .95, then the simultaneous confidence levels would be approximately (.95)(.95)(.95) = .857.

67.

a. Because p-value = .00032 < α = .001, H_o should be rejected at this significance level.

b. Not necessarily. For this n, the test statistic *t* has approximately a standard normal distribution when $H_o : r_1 = 0$ is true, and a p-value of .00032 corresponds to $z = 3.60$ (or –3.60). Solving $3.60 = \frac{r\sqrt{498}}{\sqrt{1}} - r^2$ for r yields r = .159. This r suggests only a weak linear relationship between x and y, one that would typically have little practical import.

c. $t = 2.20 \geq t_{.025,9998} = 1.96$, so H_o is rejected in favor of H_a. The value t = 2.20 is statistically significant -- it cannot be attributed just to sampling variability in the case $r = 0$. But with this n, r = .022 implies $r = .022$, which in turn shows an extremely weak linear relationship.

Supplementary Exercises

68.

a. $n = 8$, $\Sigma x_i = 207, \Sigma x_i^2 = 6799, \Sigma y_i = 621.8, \Sigma y_i^2 = 48{,}363.76$ and $\Sigma x_i y_i = 15{,}896.8$, which gives $\hat{b}_1 = \frac{-1538.20}{11{,}543} = -.133258$, $\hat{b}_0 = 81.173051$, and $y = 81.173 - .1333x$ as the equation of the estimated line.

b. We wish to test $H_0 : b_1 = 0$ vs $H_0 : b_1 \neq 0$. At level .01, H_o will be rejected (and the model judged useful) if either $t \geq t_{.005,6} = 3.707$ or $t \leq -3.707$. SSE = 8.732664, s = 1.206, and $t = \frac{-.1333}{1.206/37.985} = \frac{-.1333}{.03175} = -4.2$, which is ≤ -3.707, so we do reject H_o and find the model useful.

c. The larger the value of $\sum (x_i - \overline{x})^2$, the smaller will be $\hat{s}_{\hat{b}_1}$ and the more accurate the estimate will tend to be. For the given x_i's, $\sum (x_i - \overline{x})^2 = 1442.88$, whereas the proposed x values $x_1 = ... = x_4 = 0$, $x_5 = ... = x_8 = 50$, $\sum (x_i - \overline{x})^2 = 5000$. Thus the second set of x values is preferable to the first set. With just 3 observations at x = 0 and 3 at x = 50, $\sum (x_i - \overline{x})^2 = 3750$, which is again preferable to the first set of x_i's.

d. $\hat{b}_0 + \hat{b}_1(25) = 77.84$, and $s_{\hat{b}_0 + \hat{b}_1(25)} = s\sqrt{\frac{1}{n} + \frac{n(25 - \overline{x})^2}{n\Sigma x_i^2 - (\Sigma x_i)^2}}$

$= 1.206\sqrt{\frac{1}{8} + \frac{8(25 - 25.875)^2}{11.543}} = .426$, so the 95% CI is

$77.84 \pm (2.447)(.426) = 77.84 \pm 1.04 = (76.80, 78.88)$. The interval is quite narrow, only 2%. This is the case because the predictive value of 25% is very close to the mean of our predictor sample.

69.

a. The test statistic value is $t = \dfrac{\hat{b}_1 - 1}{s_{\hat{b}_1}}$, and H_o will be rejected if either $t \geq t_{.025,11} = 2.201$ or $t \leq -2.201$. With $\Sigma x_i = 243, \Sigma x_i^2 = 5965, \Sigma y_i = 241, \Sigma y_i^2 = 5731$ and $\Sigma x_i y_i = 5805$, $\hat{b}_1 = .913819$, $\hat{b}_0 = 1.457072$, $SSE = 75.126$, $s = 2.613$, and $s_{\hat{b}_1} = .0693$, $t = \dfrac{.9138 - 1}{.0693} = -1.24$. Because –1.24 is neither ≤ -2.201 nor ≥ 2.201, H_o cannot be rejected. It is plausible that $b_1 = 1$.

b. $r = \dfrac{16{,}902}{(136)(128.15)} = .970$

70.

a. sample size = 8

b. $\hat{y} = 326.976038 - (8.403964)x$. When x = 35.5, $\hat{y} = 28.64$.

c. Yes, the model utility test is statistically significant at the level .01.

d. $r = \sqrt{r^2} = \sqrt{0.9134} = 0.9557$

e. First check to see if the value x = 40 falls within the range of x values used to generate the least-squares regression equation. If it does not, this equation should not be used. Furthermore, for this particular model an x value of 40 yields a g value of –9.18, which is an impossible value for y.

71.

a. $r^2 = .5073$

b. $r = +\sqrt{r^2} = \sqrt{.5073} = .7122$ (positive because $\hat{b}_1$ is positive.)

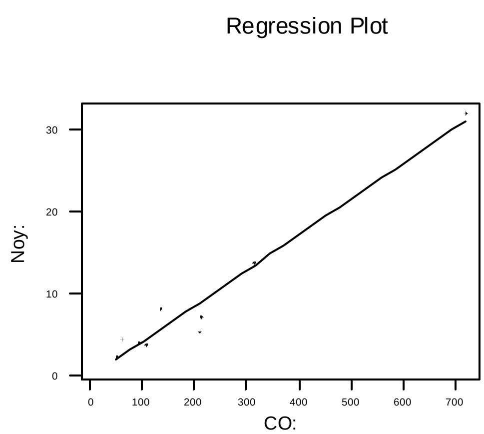
c. We test test $H_0 : b_1 = 0$ vs $H_0 : b_1 \neq 0$. The test statistic t = 3.93 gives p-value = .0013, which is < .01, the given level of significance, therefore we reject H_o and conclude that the model is useful.

d. We use a 95% CI for $m_{Y \cdot 50}$. $\hat{y}_{(50)} = .787218 + .007570(50) = 1.165718$,
$t_{.025,15} = 2.131$, s = "Root MSE" = .020308, so
$s_{\hat{y}(50)} = .20308\sqrt{\frac{1}{17} + \frac{17(50 - 42.33)^2}{17(41{,}575) - (719.60)^2}} = .051422$. The interval is , then,
$1.165718 \pm 2.131(.051422) = 1.165718 \pm .109581 = (1.056137, 1.275299)$.

e. $\hat{y}_{(30)} = .787218 + .007570(30) = 1.0143.$ The residual is
$y - \hat{y} = .80 - 1.0143 = -.2143$.

72.

a.



The above analysis was created in Minitab. A simple linear regression model seems to fit the data well. The least squares regression equation is $\hat{y} = -.220 + .0436x$. The model utility test obtained from Minitab produces a t test statistic equal to 12.72. The corresponding p-value is extremely small. So we have sufficient evidence to claim that ΔCO is a good predictor of ΔNO_y.

b. $\hat{y} = -.220 + .0436(400) = 17.228$. A 95% prediction interval produced by Minitab is (11.953, 22.503). Since this interval is so wide, it does not appear that ΔNO_y is accurately predicted.

c. While the large ΔCO value appears to be "near" the least squares regression line, the value has extremely high leverage. The least squares line that is obtained when excluding the value is $\hat{y} = 1.00 + .0346x$. The r^2 value with the value included is 96% and is reduced to 75% when the value is excluded. The value of s with the value included is 2.024, and with the value excluded is 1.96. So the large ΔCO value does appear to effect our analysis in a substantial way.

73.

a. n = 9, $\Sigma x_i = 228, \Sigma x_i^2 = 5958, \Sigma y_i = 93.76, \Sigma y_i^2 = 982.2932$ and $\Sigma x_i y_i = 2348.15$, giving $\hat{b}_1 = \frac{-243.93}{1638} = -.148919$, $\hat{b}_0 = 14.190392$, and the equation $\hat{y} = 14.19 - (.1489)x$.

b. b_1 is the expected increase in load associated with a one-day age increase (so a negative value of b_1 corresponds to a decrease). We wish to test $H_0 : b_1 = -.10$ vs. $H_0 : b_1 < -.10$ (the alternative contradicts prior belief). H_o will be rejected at level .05 if $t = \frac{\hat{b}_1 - (-.10)}{s_{\hat{b}_1}} \le -t_{.05,7} = -1.895$. With SSE = 1.4862, s = .4608, and $s_{\hat{b}_1} = \frac{.4608}{\sqrt{182}} = .0342$. Thus $t = \frac{-.1489 + 1}{.0342} = -1.43$. Because –1.43 is not ≤ -1.895, do not reject H_o.

c. $\Sigma x_i = 306, \Sigma x_i^2 = 7946,$ so $\sum (x_i - \overline{x})^2 = 7946 - \frac{(306)^2}{12} = 143$ here, as contrasted with 182 for the given 9 x_i's. Even though the sample size for the proposed x values is larger, the original set of values is preferable.

d. $(t_{.025,7})(s)\sqrt{\frac{1}{9} + \frac{9(28 - 25.33)^2}{1638}} = (2.365)(.4608)(.3877) = .42$, and $\hat{b}_0 + \hat{b}_1(28) = 10.02,$ so the 95% CI is $10.02 \pm .42 = (9.60, 10.44)$.

74.

a. $\hat{b}_1 = \frac{3.5979}{44.713} = .0805$, $\hat{b}_0 = 1.6939$, $\hat{y} = 1.69 + (.0805)x$.

b. $\hat{b}_1 = \frac{3.5979}{.2943} = 12.2254$, $\hat{b}_0 = -20.4046$, $\hat{y} = -20.40 + (12.2254)x$.

c. r = .992, so r^2 = .984 for either regression.

75.

a. The plot suggests a strong linear relationship between x and y.

b. n = 9, $\Sigma x_i = 1797, \Sigma x_i^2 = 4334.41, \Sigma y_i = 7.28, \Sigma y_i^2 = 7.4028$ and $\Sigma x_i y_i = 178.683$, so $\hat{b}_1 = \dfrac{299.931}{6717.6} = .04464854$, $\hat{b}_0 = -.08259353$, and the equation of the estimated line is $\hat{y} = -.08259 - (.044649)x$.

c. $SSE = 7.4028 - (-601281) - 7.977935 = .026146,$
$SST = 7.4028 - \dfrac{(7.28)^2}{9} = .026146, = 1.5141$, and $r^2 = 1 - \dfrac{SSE}{SST} = .983$, so 93.8% of the observed variation is "explained."

d. $\hat{y}_4 = -.08259 - (.044649)(19.1) = .7702$, and
$y_4 - \hat{y}_4 = .68 - .7702 = -.0902$.

e. s = .06112, and $s_{\hat{b}_1} = \dfrac{.06112}{\sqrt{746.4}} = .002237$, so the value of t for testing $H_0 : b_1 = 0$ vs $H_0 : b_1 \neq 0$ is $t = \dfrac{.044649}{.002237} = 19.96$. From Table A.5, $t_{.0005,7} = 5.408$, so $p - value < 2(.0005) = .001$. There is strong evidence for a useful relationship.

f. A 95% CI for b_1 is $.044649 \pm (2.365)(.002237) = .044649 \pm .005291$
$= (.0394, .0499)$.

g. A 95% CI for $b_0 + b_1(20)$ is $.810 \pm (2.365)(.002237)(.3333356)$
$= .810 \pm .048 = (.762, .858)$

76. Substituting x* = 0 gives the CI $\hat{b}_0 \pm t_{a/2,n-2} \cdot s\sqrt{\dfrac{1}{n} + \dfrac{n\bar{x}^2}{n\Sigma x_i^2 - (\Sigma x_i)^2}}$. From Example 12.8, $\hat{b}_0 = 3.621$, SSE = .262453, n = 14, $\Sigma x_i = 890, \bar{x} = 63.5714, \Sigma x_i^2 = 67{,}182,$ so with s = .1479, $t_{.025,12} = 2.179$, the CI is $3.621 \pm 2.179(.1479)\sqrt{\dfrac{1}{12} + \dfrac{56{,}578.52}{148{,}448}}$
$= 3.621 \pm 2.179(.1479)(.6815) = 3.62 \pm .22 = (3.40, 3.84).$

77. $SSE = \Sigma y^2 - \hat{b}_0 \Sigma y - \hat{b}_1 \Sigma xy$. Substituting $\hat{b}_0 = \dfrac{\Sigma y - \hat{b}_1 \Sigma x}{n}$, SSE becomes

$$SSE = \Sigma y^2 - \frac{\Sigma y\left(\Sigma y - \hat{b}_1 \Sigma x\right)}{n} - \hat{b}_1 \Sigma xy = \Sigma y^2 - \frac{(\Sigma y)^2}{n} + \frac{\hat{b}_1 \Sigma x \Sigma y}{n} - \hat{b}_1 \Sigma xy$$

$= \left[\Sigma y^2 - \dfrac{(\Sigma y)^2}{n}\right] - \hat{b}_1\left[\Sigma xy - \dfrac{\Sigma x \Sigma y}{n}\right] = S_{yy} - \hat{b}_1 S_{xy}$, as desired.

78. The value of the sample correlation coefficient using the squared y values would not necessarily be approximately 1. If the y values are greater than 1, then the squared y values would differ from each other by more than the y values differ from one another. Hence, the relationship between x and y^2 would be less like a straight line, and the resulting value of the correlation coefficient would decrease.

79.

a. With $s_{xx} = \sum (x_i - \overline{x})^2$, $s_{yy} = \sum (y_i - \overline{y})^2$, note that $\dfrac{s_y}{s_x} = \sqrt{\dfrac{s_{yy}}{s_{xx}}}$ (since the factor n-1 appears in both the numerator and denominator, so cancels). Thus

$$y = \hat{b}_0 + \hat{b}_1 x = \overline{y} + \hat{b}_1 (x - \overline{x}) = \overline{y} + \frac{s_{xy}}{s_{xx}}(x - \overline{x}) = \overline{y} + \sqrt{\frac{s_{yy}}{s_{xx}}} \cdot \frac{s_{xy}}{\sqrt{s_{xx} s_{yy}}}(x - \overline{x})$$

$= \overline{y} + \dfrac{s_y}{s_x} \cdot r \cdot (x - \overline{x})$, as desired.

b. By .573 s.d.'s above, (above, since r < 0) or (since $s_y = 4.3143$) an amount 2.4721 above.

80. With s_{xy} given in the text, $r = \dfrac{s_{xy}}{\sqrt{s_{xx}s_{yy}}}$ (where e.g. $s_{xx} = \sum (x_i - \overline{x})^2$), and

$\hat{b}_1 = \dfrac{s_{xy}}{s_{xx}}$. Also, $s = \sqrt{\dfrac{SSE}{n-2}}$ and $SSE = \Sigma y_i^2 - \hat{b}_0 \Sigma y_i - \hat{b}_1 \Sigma x_i y_i = s_{yy} - \hat{b}_1 s_{xy}$.

Thus the t statistic for $H_o : \hat{b}_1 = 0$ is

$$t = \frac{\hat{b}_1}{s / \sqrt{\sum (x_i - \overline{x})^2}} = \frac{(s_{xy} / s_{xx}) \cdot \sqrt{s_{xx}}}{\sqrt{(s_{yy} - s_{xy}^2 / s_{xx})/(n-2)}}$$

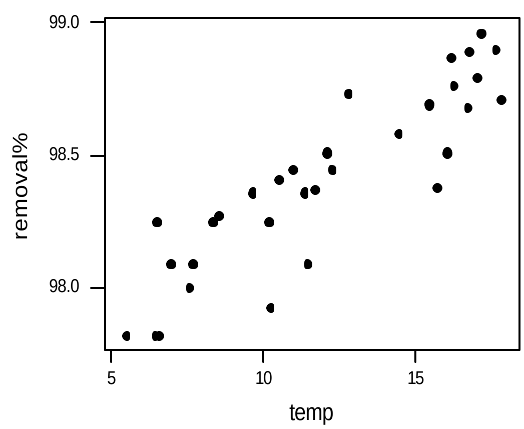
$$= \frac{s_{xy} \cdot \sqrt{n-2}}{\sqrt{(s_{xx}s_{yy} - s_{xy}^2)}} = \frac{(s_{xy} / \sqrt{s_{xx}s_{yy}})\sqrt{n-2}}{\sqrt{1 - s_{xy}^2 / s_{xx}s_{yy}}} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$ as desired.

81. Using the notation of the exercise above, $SST = s_{yy}$ and $SSE = s_{yy} - \hat{b}_1 s_{xy}$

$= s_{yy} - \dfrac{s_{xy}^2}{s_{xx}}$, so $1 - \dfrac{SSE}{SST} = 1 - \dfrac{s_{yy} - \dfrac{s_{xy}^2}{s_{xx}}}{s_{yy}} = \dfrac{s_{xy}^2}{s_{xx}s_{yy}} = r^2$, as desired.

82.

a. A Scatter Plot suggests the linear model is appropriate.



b. Minitab Output:

```
The regression equation is
removal% = 97.5 + 0.0757 temp

Predictor        Coef       StDev          T        P
Constant      97.4986      0.0889    1096.17    0.000
temp         0.075691    0.007046      10.74    0.000

S = 0.1552      R-Sq = 79.4%     R-Sq(adj) = 78.7%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1      2.7786      2.7786    115.40    0.000
Residual Error    30      0.7224      0.0241
Total             31      3.5010
```

Minitab will output all the residual information if the option is chosen, from which you can find the point prediction value $\hat{y}_{10.5} = 98.2933$, the observed value y = 98.41, so the residual = .0294.

c. Roughly .1

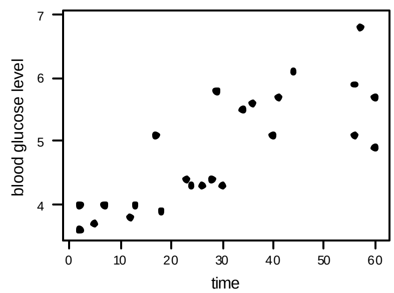
d. $R^2 = 79.4$

e. A 95% CI for β_1, using $t_{.025,30} = 2.042$:

$$.075691 \pm 2.042(.007046) = (.061303, .090079)$$

f. The slope of the regression line is steeper. The value of s is almost doubled, and the value of R^2 drops to 61.6%.

83. Using Minitab, we create a scatterplot to see if a linear regression model is appropriate.



A linear model is reasonable; although it appears that the variance in y gets larger as x increases. The Minitab output follows:

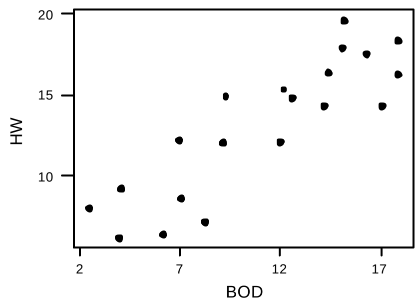
```
The regression equation is
blood glucose level = 3.70 + 0.0379 time

Predictor        Coef       StDev          T        P
Constant       3.6965      0.2159      17.12    0.000
time         0.037895    0.006137       6.17    0.000

S = 0.5525      R-Sq = 63.4%     R-Sq(adj) = 61.7%

Analysis of Variance

Source            DF          SS          MS         F         P
Regression         1      11.638      11.638     38.12     0.000
Residual Error    22       6.716       0.305
Total             23      18.353
```

The coefficient of determination of 63.4% indicates that only a moderate percentage of the variation in y can be explained by the change in x. A test of model utility indicates that time is a significant predictor of blood glucose level. (t = 6.17, p = 0.0). A point estimate for blood glucose level when time = 30 minutes is 4.833%. We would expect the average blood glucose level at 30 minutes to be between 4.599 and 5.067, with 95% confidence.

84.

a. Using the techniques from a previous chapter, we can do a t test for the difference of two means based on paired data. Minitab's paired t test for equality of means gives t = 3.54, with a p value of .002, which suggests that the average bf% reading for the two methods is not the same.

b. Using linear regression to predict HW from BOD POD seems reasonable after looking at the scatterplot, below.



The least squares linear regression equation, as well as the test statistic and p value for a model utility test, can be found in the Minitab output below. We see that we do have significance, and the coefficient of determination shows that about 75% of the variation in HW can be explained by the variation in BOD.

```
The regression equation is
HW = 4.79 + 0.743 BOD

Predictor        Coef       StDev          T        P
Constant        4.788       1.215       3.94    0.001
BOD            0.7432      0.1003       7.41    0.000

S = 2.146       R-Sq = 75.3%     R-Sq(adj) = 73.9%

Analysis of Variance

Source            DF          SS          MS        F        P
Regression         1      252.98      252.98    54.94    0.000
Residual Error    18       82.89        4.60
Total             19      335.87
```

85. For the second boiler, $n = 19$, $\Sigma x_i = 125$, $\Sigma y_i = 472.0$, $\Sigma x_i^2 = 3625$, $\Sigma y_i^2 = 37{,}140.82$, and $\Sigma x_i y_i = 9749.5$, giving $\hat{g}_1 =$ estimated slope $= \dfrac{-503}{6125} = -.0821224$, $\hat{g}_0 = 80.377551$, $SSE_2 = 3.26827$, $SSx_2 = 1020.833$. For boiler #1, n = 8, $\hat{b}_1 = -.1333$, $SSE_1 = 8.733$, and $SSx_1 = 1442.875$. Thus $\hat{s}^2 = \dfrac{8.733 + 3.286}{10} = 1.2$, $\hat{s} = 1.095$, and $t = \dfrac{-.1333 + .0821}{1.095\sqrt{\frac{1}{1442.875} + \frac{1}{1020.833}}}$ $= \dfrac{-.0512}{.0448} = -1.14$. $t_{.025,10} = 2.228$ and -1.14 is neither ≥ 2.228 nor ≤ -2.228, so H_o is not rejected. It is plausible that $b_1 = g_1$.

CHAPTER 13

Section 13.1

1.

a. $\overline{x} = 15$ and $\sum\left(x_j - \overline{x}\right)^2 = 250$, so s.d. of $Y_i - \hat{Y}_i$ is $10\sqrt{1 - \frac{1}{5} - \frac{\left(x_i - 15\right)^2}{250}} =$ 6.32, 8.37, 8.94, 8.37, and 6.32 for i = 1, 2, 3, 4, 5.

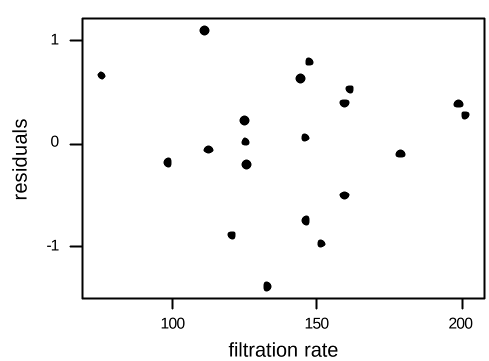
b. Now $\overline{x} = 20$ and $\sum\left(x_i - \overline{x}\right)^2 = 1250$, giving standard deviations 7.87, 8.49, 8.83, 8.94, and 2.83 for i = 1, 2, 3, 4, 5.

c. The deviation from the estimated line is likely to be much smaller for the observation made in the experiment of **b** for x = 50 than for the experiment of **a** when x = 25. That is, the observation (50, Y) is more likely to fall close to the least squares line than is (25, Y).

2. The pattern gives no cause for questioning the appropriateness of the simple linear regression model, and no observation appears unusual.

3.

a. This plot indicates there are no outliers, the variance of ε is reasonably constant, and the ε are normally distributed. A straight-line regression function is a reasonable choice for a model.

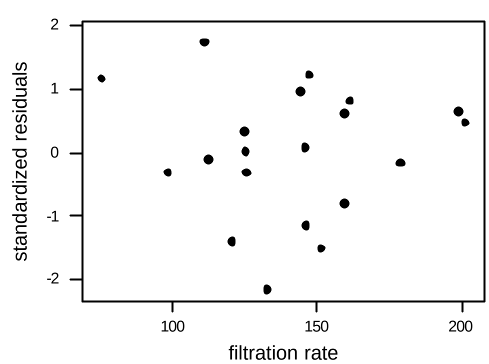


b. We need $S_{xx} = \sum(x_i - \overline{x})^2 = 415{,}914.85 - \frac{(2817.9)^2}{20} = 18{,}886.8295$. Then each e_i^* can be calculated as follows: $e_i^* = \dfrac{e_i}{.4427\sqrt{1 + \frac{1}{20} + \frac{(x_i - 140.895)^2}{18{,}886.8295}}}$. The table below shows the values:

standardized residuals	e / e_i^*	standardized residuals	e / e_i^*
-0.31064	0.644053	0.6175	0.64218
-0.30593	0.614697	0.09062	0.64802
0.4791	0.578669	1.16776	0.565003
1.2307	0.647714	-1.50205	0.646461
-1.15021	0.648002	0.96313	0.648257
0.34881	0.643706	0.019	0.643881
-0.09872	0.633428	0.65644	0.584858
-1.39034	0.640683	-2.1562	0.647182
0.82185	0.640975	-0.79038	0.642113
-0.15998	0.621857	1.73943	0.631795

Notice that if $e_i^* \approx e / s$, then $e / e_i^* \approx s$. All of the e / e_i^*'s range between .57 and .65, which are close to s.

c. This plot looks very much the same as the one in part a.

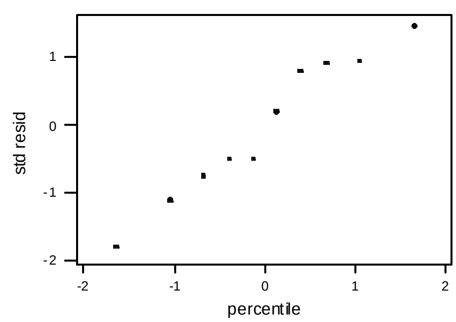


4.

a. The (x, residual) pairs for the plot are (0, -.335), (7, -.508), (17. -.341), (114, .592), (133, .679), (142, .700), (190, .142), (218, 1.051), (237, -1.262), and (285, -.719). The plot shows substantial evidence of curvature.

b. The standardized residuals (in order corresponding to increasing x) are -.50, -.75, -.50, .79, .90, .93, .19, 1.46, -1.80, and -1.12. A standardized residual plot shows the same pattern as the residual plot discussed in the previous exercise. The z percentiles for the normal probability plot are –1.645, -1.04, -.68, -.39, -.13, .13, .39, .68, 1.04, 1.645. The plot follows. The points follow a linear pattern, so the standardized residuals appear to have a normal distribution.

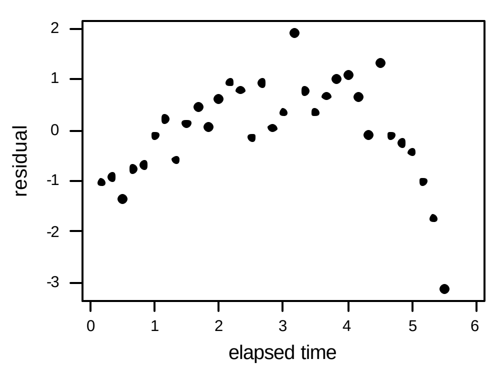
Normal Probability Plot for the Standardized Residuals



5.

a. 97.7% of the variation in ice thickness can be explained by the linear relationship between it and elapsed time. Based on this value, it appears that a linear model is reasonable.

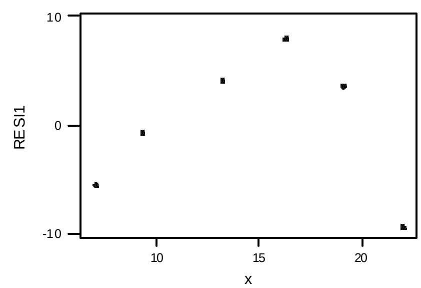
b. The residual plot shows a curve in the data, so perhaps a non-linear relationship exists. One observation (5.5, -3.14) is extreme.



6.

a. $H_o: b_1 = 0$ vs. $H_a: b_1 \neq 0$. The test statistic is $t = \dfrac{\hat{b}_1}{s_{\hat{b}_1}}$, and we will reject H_o if $t \geq t_{.025,4} = 2.776$ or if $t \leq -2.776$. $s_{\hat{b}_1} = \dfrac{s}{\sqrt{S_{xx}}} = \dfrac{7.265}{12.869} = .565$, and $t = \dfrac{6.19268}{.565} = 10.97$. Since $10.97 \geq 2.776$, we reject H_o and conclude that the model is useful.

b. $\hat{y}_{(7.0)} = 1008.14 + 6.19268(7.0) = 1051.49$, from which the residual is $y - \hat{y}_{(7.0)} = 1046 - 1051.49 = -5.49$. Similarly, the other residuals are -.73, 4.11, 7.91, 3.58, and –9.38. The plot of the residuals vs x follows:

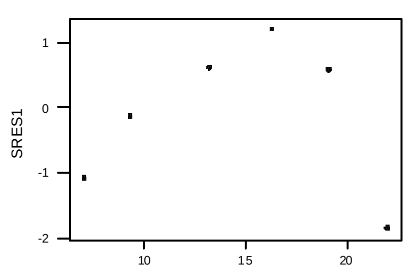


Because a curved pattern appears, a linear regression function may be inappropriate.

c. The standardized residuals are calculated as

$$e_1^* = \frac{-5.49}{7.265\sqrt{1 + \frac{1}{6} + \frac{(7.0 - 14.48)^2}{165.5983}}} = -1.074$$

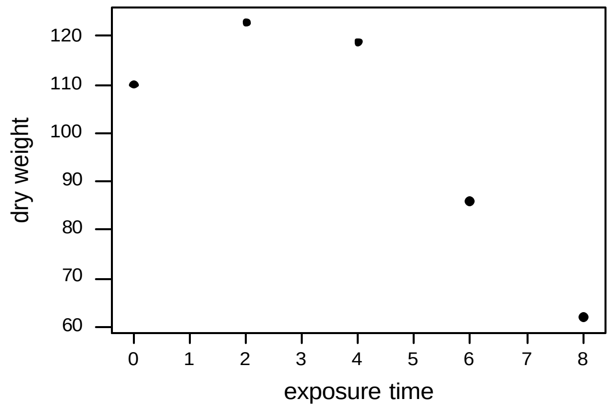
, and similarly the others are -.123, .624, 1.208, .587, and –1.841. The plot of e* vs x follows :



This plot gives the same information as the previous plot. No values are exceptionally large, but the e* of –1.841 is close to 2 std deviations away from the expected value of 0.

7.

a.

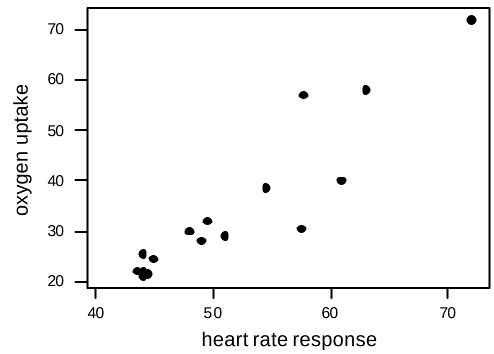


There is an obvious curved pattern in the scatter plot, which suggests that a simple linear model will not provide a good fit.

b. The $\hat{y}$'s, e's, and e*'s are given below:

x	y	$\hat{y}$	e	e*
0	110	126.6	-16.6	-1.55
2	123	113.3	9.7	.68
4	119	100.0	19.0	1.25
6	86	86.7	-.7	-.05
8	62	73.4	-11.4	-1.06

8. First, we will look at a scatter plot of the data, which is quite linear, so it seems reasonable to use linear regression.

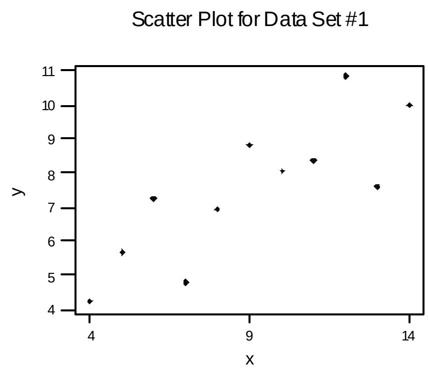


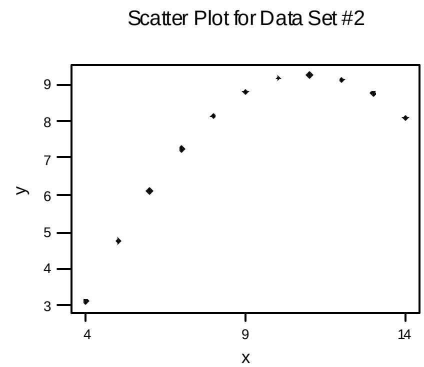
The linear regression output (Minitab) follows:

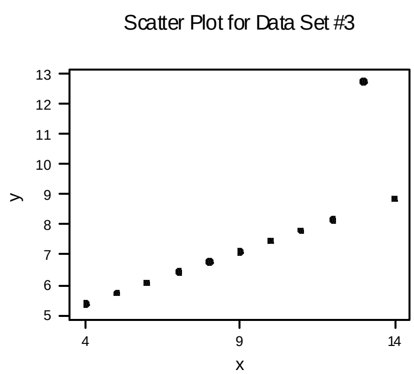
```
The regression equation is
y = - 51.4 + 1.66 x

Predictor        Coef       StDev          T        P
Constant      -51.355       9.795      -5.24    0.000
x              1.6580      0.1869       8.87    0.000

S = 6.119       R-Sq = 84.9%     R-Sq(adj) = 83.8%

Analysis of Variance

Source            DF          SS          MS         F         P
Regression         1      2946.5      2946.5     78.69     0.000
Residual Error    14       524.2        37.4
Total             15      3470.7
```

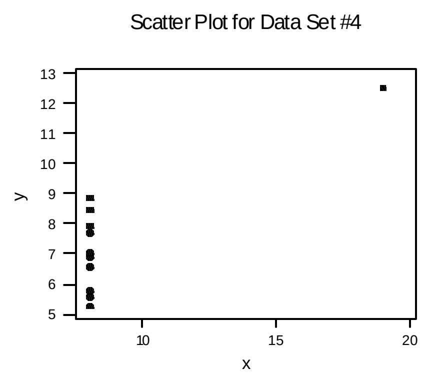
A quick look at the t and p values shows that the model is useful, and r^2 shows a strong relationship between the two variables.
The observation (72, 72) has large influence, since its x value is a distance from the others. We could run the regression again, without this value, and get the line:
`oxygen uptake = - 44.8 + 1.52 heart rate response.`

9. Both a scatter plot and residual plot (based on the simple linear regression model) for the first data set suggest that a simple linear regression model is reasonable, with no pattern or influential data points which would indicate that the model should be modified. However, scatter plots for the other three data sets reveal difficulties.









For data set #2, a quadratic function would clearly provide a much better fit. For data set #3, the relationship is perfectly linear except one outlier, which has obviously greatly influenced the fit even though its x value is not unusually large or small. The signs of the residuals here (corresponding to increasing x) are + + + + - - - - - + -, and a residual plot would reflect this pattern and suggest a careful look at the chosen model. For data set #4 it is clear that the slope of the least squares line has been determined entirely by the outlier, so this point is extremely influential (and its x value does lie far from the remaining ones).

10.

a. $e_i = y_i - \left(\hat{b}_0 - \hat{b}_1 x_i\right) = y_i - \bar{y} - \hat{b}_1(x_i - \bar{x})$, so
$\Sigma e_i = \Sigma(y_i - \bar{y}) - \hat{b}_1 \Sigma(x_i - \bar{x}) = 0 + \hat{b}_1 \cdot 0 = 0$.

b. Since $\Sigma e_i = 0$ always, the residuals cannot be independent. There is clearly a linear relationship between the residuals. If one e_I is large positive, then al least one other e_I would have to be negative to preserve $\Sigma e_i = 0$. This suggests a negative correlation between residuals (for fixed values of any n – 2, the other two obey a negative linear relationship).

c. $\Sigma x_i e_i = \Sigma x_i y_i - \Sigma x_i \bar{y} - \hat{b}_1 \Sigma x_i (x_i - \bar{x}) = \left[\Sigma x_i y_i - \frac{(\Sigma x_i)(\Sigma y_i)}{n}\right] - \hat{b}_1\left[\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}\right]$

, but the first term in brackets is the numerator of $\hat{b}_1$, while the second term is the denominator of $\hat{b}_1$, so the difference becomes (numerator of $\hat{b}_1$) – (numerator of $\hat{b}_1$) = 0.

d. The five e_i^*'s from Exercise 7 above are –1.55, .68, 1.25, -.05, and –1.06, which sum to -.73. This sum differs too much from 0 to be explained by rounding. In general it is not true that $\Sigma e_i^* = 0$.

11.

a. $Y_i - \hat{Y}_i = Y_i - \overline{Y} - \hat{b}_1(x_i - \overline{x}) = Y_i - \dfrac{1}{n}\sum_j Y_j - \dfrac{(x_i - \overline{x})\sum_j (x_j - \overline{x})Y_j}{\sum_j (x_j - \overline{x})^2} = \sum_j c_j Y_j$,

where $c_j = 1 - \dfrac{1}{n} - \dfrac{(x_i - \overline{x})^2}{n\Sigma(x_j - \overline{x})^2}$ for j = i and $c_j = 1 - \dfrac{1}{n} - \dfrac{(x_i - \overline{x})(x_j - \overline{x})}{\Sigma(x_j - \overline{x})^2}$ for

$j \neq i$. Thus $Var(Y_i - \hat{Y}_i) = \Sigma Var(c_j Y_j)$ (since the Y_j's are independent) = $s^2 \Sigma c_j^2$ which, after some algebra, gives equation (13.2).

b. $s^2 = Var(Y_i) = Var(\hat{Y}_i + (Y_i - \hat{Y}_i)) = Var(\hat{Y}_i) + Var(Y_i - \hat{Y}_i)$, so

$Var(Y_i - \hat{Y}_i) = s^2 - Var(\hat{Y}_i) = s^2 - s^2\left[\dfrac{1}{n} + \dfrac{(x_i - \overline{x})^2}{n\Sigma(x_j - \overline{x})^2}\right]$, which is exactly

(13.2).

c. As x_i moves further from $\overline{x}$, $(x_i - \overline{x})^2$ grows larger, so $Var(\hat{Y}_i)$ increases (since $(x_i - \overline{x})^2$ has a positive sign in $Var(\hat{Y}_i)$), but $Var(Y_i - \hat{Y}_i)$ decreases (since $(x_i - \overline{x})^2$ has a negative sign).

12.

a. $\Sigma e_i = 34$, which is not = 0, so these cannot be the residuals.

b. Each $x_i e_i$ is positive (since x_i and e_i have the same sign) so $\Sigma x_i e_i > 0$, which contradicts the result of exercise 10**c**, so these cannot be the residuals for the given x values.

13. The distribution of any particular standardized residual is also a t distribution with n – 2 d.f., since e_i^* is obtained by taking standard normal variable $\frac{(Y_i - \hat{Y}_i)}{(s_{Y_i - \hat{Y}})}$ and substituting the estimate of σ in the denominator (exactly as in the predicted value case). With E_i^* denoting the i[th] standardized residual as a random variable, when n = 25 E_i^* has a t distribution with 23 d.f. and $t_{.01,23} = 2.50$, so P(E_i^* outside (-2.50, 2.50)) =
$P(E_i^* \geq 2.50) + P(E_i^* \leq -2.50) = .01 + .01 = .02$.

14. space

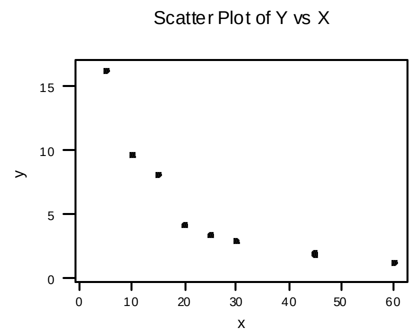
a. $n_1 = n_2 = 3$ (3 observations at 110 and 3 at 230), $n_3 = n_4 = 4$, $\bar{y}_{1.} = 202.0$, $\bar{y}_{2.} = 149.0$, $\bar{y}_{3.} = 110.5$, $\bar{y}_{4.} = 107.0$, $\Sigma\Sigma y_{ij}^2 = 288{,}013$, so
$SSPE = 288{,}013 - \left[3(202.0)^2 + 3(149.0)^2 + 4(110.5)^2 + 4(107.0)^2\right] = 4361$.
With $\Sigma x_i = 4480$, $\Sigma y_i = 1923$, $\Sigma x_i^2 = 1{,}733{,}500$, $\Sigma y_i^2 = 288{,}013$ (as above), and $\Sigma x_i y_i = 544{,}730$, SSE = 7241 so SSLF = 7241-4361=2880. With c – 2 = 2 and n – c = 10, $F_{.05,2,10} = 4.10$. $MSLF = \frac{2880}{2} = 1440$ and $SSPE = \frac{4361}{10} = 436.1$, so the computed value of F is $\frac{1440}{436.1} = 3.30$. Since 3.30 is not ≥ 4.10, we do not reject H[o]. This formal test procedure does not suggest that a linear model is inappropriate.

b. The scatter plot clearly reveals a curved pattern which suggests that a nonlinear model would be more reasonable and provide a better fit than a linear model.

Section 13.2

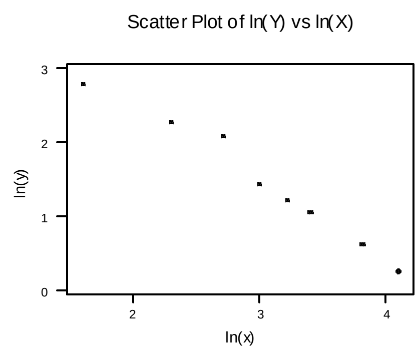
15.

a.



The points have a definite curved pattern. A linear model would not be appropriate.

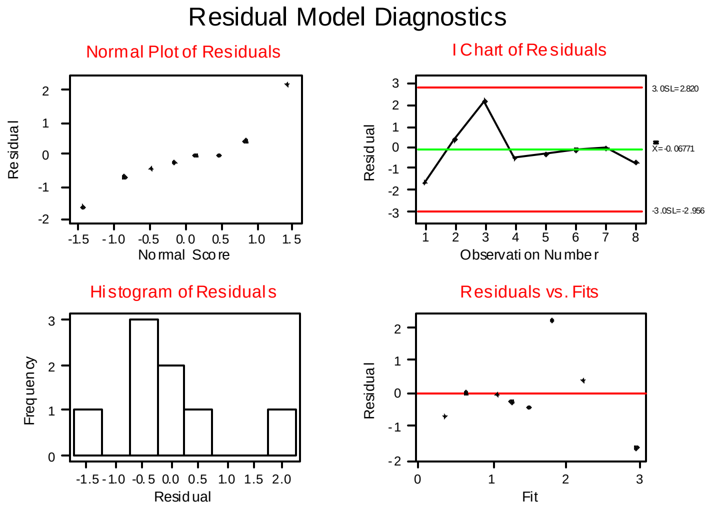
b. In this plot we have a strong linear pattern.



c. The linear pattern in **b** above would indicate that a transformed regression using the natural log of both x and y would be appropriate. The probabilistic model is then $y = ax^b \cdot e$. (The power function with an error term!)

d. A regression of ln(y) on ln(x) yields the equation $\ln(y) = 4.6384 - 1.04920\ln(x)$. Using Minitab we can get a P.I. for y when x = 20 by first transforming the x value: ln(20) = 2.996. The computer generated 95% P.I. for ln(y) when ln(x) = 2.996 is (1.1188,1.8712). We must now take the antilog to return to the original units of Y: $\left(e^{1.1188}, e^{1.8712}\right) = (3.06, 6.50)$.

e. A computer generated residual analysis:



Looking at the residual vs. fits (bottom right), one standardized residual, corresponding to the third observation, is a bit large. There are only two positive standardized residuals, but two others are essentially 0. The patterns in the residual plot and the normal probability plot (upper left) are marginally acceptable.

16.

a. $\Sigma x_i = 9.72$, $\Sigma y_i' = 313.10$, $\Sigma x_i^2 = 8.0976$, $\Sigma y_i'^2 = 288{,}013$, $\Sigma x_i y_i' = 255.11$, (all from computer printout, where $y_i' = \ln(L_{178})$), from which $\hat{b}_1 = 6.6667$ and $\hat{b}_0 = 20.6917$ (again from computer output). Thus $\hat{b} = \hat{b}_1 = 6.6667$ and $\hat{a} = e^{\hat{b}_0} = 968{,}927{,}163$.

b. We first predict y' using the linear model and then exponentiate:

$y' = 20.6917 + 6.6667(.75) = 25.6917$, so

$\hat{y} = \hat{L}_{178} = e^{25.6917} = 1.438051363 \times 10^{11}$.

c. We first compute a prediction interval for the transformed data and then exponentiate.

With $t_{.025,10} = 2.228$, s = .5946, and $\sqrt{1 + \frac{1}{12} + \frac{(.95 - \overline{x})^2}{\Sigma x^2 - (\Sigma x)^2 / 12}} = 1.082$, the prediction interval for y' is

$27.0251 \pm (2.228)(.5496)(1.082) = 27.0251 \pm 1.4334 = (25.5917, 28.4585)$.

The P.I. for y is then $\left(e^{25.5917}, e^{28.4585}\right)$.

17.

a. $\Sigma x_i' = 15.501$, $\Sigma y_i' = 13.352$, $\Sigma x_i'^2 = 20.228$, $\Sigma y_i'^2 = 16.572$, $\Sigma x_i' \ y_i' = 18.109$, from which $\hat{b}_1 = 1.254$ and $\hat{b}_0 = -.468$ so $\hat{b} = \hat{b}_1 = 1.254$ and $\hat{a} = e^{-.468} = .626$.

b. The plots give strong support to this choice of model; in addition, $r^2 = .960$ for the transformed data.

c. SSE = .11536 (computer printout), s = .1024, and the estimated sd of $\hat{b}_1$ is .0775, so $t = \frac{1.25 - 1.33}{.0775} = -1.07$. Since –1.07 is not $\le -t_{.05,11} = -1.796$, H_o cannot be rejected in favor of H_a.

d. The claim that $m_{Y \cdot 5} = 2m_{Y \cdot 2.5}$ is equivalent to $a \cdot 5^b = 2a(2.5)^b$, or that $b = 1$. Thus we wish test $H_o : b_1 = 1$ vs. $H_a : b_1 \ne 1$. With $t = \frac{1 - 1.33}{.0775} = -4.30$ and RR $-t_{.005,11} \le -3.106$, H_o is rejected at level .01 since $-4.30 \le -3.106$.

18. A scatter plot may point us in the direction of a power function, so we try $y = ax^b$. We transform $x' = \ln(x)$, so $y = a + b\ln(x)$. This transformation yields a linear regression equation $y = .0197 - .00128x'$ or $y = .0197 - .00128\ln(x)$. Minitab output follows:

```
The regression equation is
y = 0.0197 - 0.00128 x

Predictor        Coef       StDev          T        P
Constant     0.019709    0.002633       7.49    0.000
x          -0.0012805   0.0003126      -4.10    0.001

S = 0.002668    R-Sq = 49.7%     R-Sq(adj) = 46.7%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1  0.00011943  0.00011943     16.78    0.001
Residual Error    17  0.00012103  0.00000712
Total             18  0.00024046
```

The model is useful, based on a t test, with a p value of .001. But $r^2 = 49.7$, so only 49.7% of the variation in y can be explained by its relationship with ln(x).

To estimate y_{5000}, we need $x' = \ln(5000) = 8.51718$. A point estimate for y when x =5000 is y = .009906. A 95 % prediction interval for y_{5000} is $(.002257, 017555)$.

19.

a. No, there is definite curvature in the plot.

b. $Y' = b_0 + b_1(x') + e$ where $x' = \dfrac{1}{temp}$ and $y' = \ln(lifetime)$. Plotting y' vs. x' gives a plot which has a pronounced linear appearance (and in fact $r^2 = .954$ for the straight line fit).

c. $\Sigma x_i' = .082273$, $\Sigma y_i' = 123.64$, $\Sigma x_i'^2 = .00037813$, $\Sigma y_i'^2 = 879.88$, $\Sigma x_i' y_i' = .57295$, from which $\hat{b}_1 = 3735.4485$ and $\hat{b}_0 = -10.2045$ (values read from computer output). With x = 220, $x' = .00445$ so $\hat{y}' = -10.2045 + 3735.4485(.00445) = 6.7748$ and thus $\hat{y} = e^{\hat{y}'} = 875.50$.

d. For the transformed data, SSE = 1.39857, and $n_1 = n_2 = n_3 = 6,\ \bar{y}'_{1.} = 8.44695$, $\bar{y}'_{2.} = 6.83157$, $\bar{y}'_{3.} = 5.32891$, from which SSPE = 1.36594, SSLF = .02993, $f = \dfrac{.02993/1}{1.36594/15} = .33$. Comparing this to $F_{.01,1,15} = 8.68$, it is clear that H_o cannot be rejected.

20. After examining a scatter plot and a residual plot for each of the five suggested models as well as for y vs. x, I felt that the power model $Y = ax^b \cdot e$ $(y' = \ln(y)$ vs. $x' = \ln(x))$ provided the bet fit. The transformation seemed to remove most of the curvature from the scatter plot, the residual plot appeared quite random, $\left|e_i'^*\right| < 1.65$ for every i, there was no indication of any influential observations, and $r^2 = .785$ for the transformed data.

21.

a. The suggested model is $Y = b_0 + b_1(x') + e$ where $x' = \dfrac{10^4}{x}$. The summary quantities are $\Sigma x_i' = 159.01$, $\Sigma y_i = 121.50$, $\Sigma x_i'^2 = 4058.8$, $\Sigma y_i^2 = 1865.2$, $\Sigma x_i' y_i = 2281.6$, from which $\hat{b}_1 = -.1485$ and $\hat{b}_0 = 18.1391$, and the estimated regression function is $y = 18.1391 - \dfrac{1485}{x}$.

b. $x = 500 \Rightarrow \hat{y} = 18.1391 - \dfrac{1485}{500} = 15.17$.

22.

a. $\frac{1}{y} = a + bx$, so with $y' = \frac{1}{y}$, $y' = a + bx$. The corresponding probabilistic model is $\frac{1}{y} = a + bx + e$.

b. $\frac{1}{y} - 1 = e^{a+bx}$, so $\ln\left(\frac{1}{y} - 1\right) = a + bx$. Thus with $y' = \ln\left(\frac{1}{y} - 1\right)$, $y' = a + bx$. The corresponding probabilistic model is $Y' = a + bx + e'$, or equivalently $Y = \frac{1}{1 + e^{a+bx} \cdot e}$ where $e = e^{e'}$.

c. $\ln(y) = e^{a+bx} = \ln(\ln(y)) = a + bx$. Thus with $y' = \ln(\ln(y))$, $y' = a + bx$. The probabilistic model is $Y' = a + bx + e'$, or equivalently, $Y = e^{e^{a+bx}} \cdot e$ where $e = e^{e'}$.

d. This function cannot be linearized.

23. $Var(Y) = Var(ae^{bx} \cdot e) = \left[ae^{bx}\right]^2 \cdot Var(e) = a^2 e^{2bx} \cdot t^2$ where we have set $Var(e) = t^2$. If $b > 0$, this is an increasing function of x so we expect more spread in y for large x than for small x, while the situation is reversed if $b < 0$. It is important to realize that a scatter plot of data generated from this model will not spread out uniformly about the exponential regression function throughout the range of x values; the spread will only be uniform on the transformed scale. Similar results hold for the multiplicative power model.

24. $H_0 : b_1 = 0$ vs $H_a : b_1 \neq 0$. The value of the test statistic is z = .73, with a corresponding p-value of .463. Since the p-value is greater than any sensible choice of alpha we do not reject H_o. There is insufficient evidence to claim that age has a significant impact on the presence of kyphosis.

25. The point estimate of b_1 is $\hat{b}_1 = .17772$, so the estimate of the odds ratio is $e^{\hat{b}_1} = e^{.17772} \approx 1.194$. That is , when the amount of experience increases by one year (i.e. a one unit increase in x), we estimate that the odds ratio increase by about 1.194. The z value of 2.70 and its corresponding p-value of .007 imply that the null hypothesis $H_0 : b_1 = 0$ can be rejected at any of the usual significance levels (e.g., .10, .05, .025, .01). Therefore, there is clear evidence that b_1 is not zero, which means that experience does appear to affect the likelihood of successfully performing the task. This is consistent with the confidence interval (1.05, 1.36) for the odds ratio given in the printout, since this interval does not contain the value 1. A graph of $\hat{p}$ appears below.

Section 13.3

26.

a. There is a slight curve to this scatter plot. It could be consistent with a quadratic regression.

b. We desire R^2, which we find in the output: $R^2 = 93.8\%$

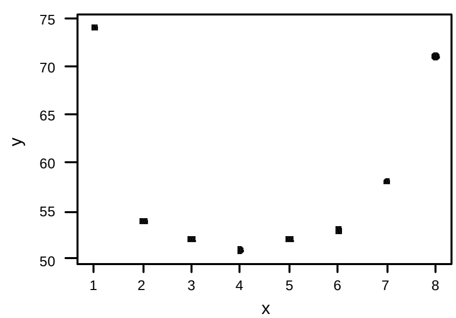
c. $H_0 : b_1 = b_2 = 0$ vs $H_a :$ at least one $b_i \neq 0$. The test statistic is $f = \dfrac{MSR}{MSE} = 22.51$, and the corresponding p-value is .016. Since the p-value < .05, we reject H_o and conclude that the model is useful.

d. We want a 99% confidence interval, but the output gives us a 95% confidence interval of (452.71, 529.48), which can be rewritten as 491.10 ± 38.38; $t_{.025,3} = 3.182$, so $s_{\hat{y}\cdot 14} = \dfrac{38.38}{3.182} = 12.06$; Now, $t_{.005,3} = 5.841$, so the 99% C.I. is $491.10 \pm 5.841(12.06) = 491.10 \pm 70.45 = (420.65, 561.55)$.

e. $H_0 : b_2 = 0$ vs $H_a : b_2 \neq 0$. The test statistic is t = -3.81, with a corresponding p-value of .032, which is < .05, so we reject H_o. the quadratic term appears to be useful in this model.

27.

a. A scatter plot of the data indicated a quadratic regression model might be appropriate.



b. $\hat{y} = 84.482 - 15.875(6) + 1.7679(6)^2 = 52.88$; residual = $y_6 - \hat{y}_6 = 53 - 52.88 = .12$;

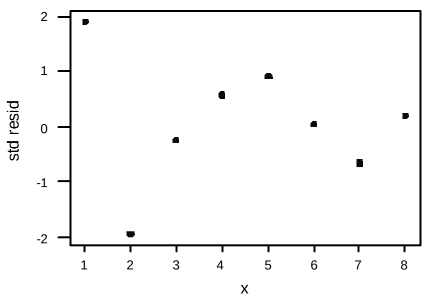
c. $SST = \Sigma y_i^2 - \dfrac{(\Sigma y_i)^2}{n} = 586.88$, so $R^2 = 1 - \dfrac{61.77}{586.88} = .895$.

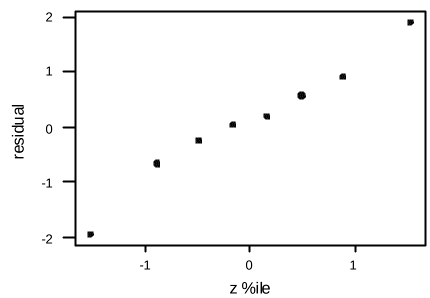
d. The first two residuals are the largest, but they are both within the interval (-2, 2). Otherwise, the standardized residual plot does not exhibit any troublesome features. For the Normal Probability Plot:

Residual	Zth percentile
-1.95	-1.53
-.66	-.89
-.25	-.49
.04	-.16
.20	.16
.58	.49
.90	.89
1.91	1.53

(continued)

The normal probability plot does not exhibit any troublesome features.





e. $\hat{m}_{Y\cdot 6} = 52.88$ (from **b**) and $t_{.025,n-3} = t_{.025,5} = 2.571$, so the C.I. is $52.88 \pm (2.571)(1.69) = 52.88 \pm 4.34 = (48.54, 57.22)$.

f. SSE = 61.77 so $s^2 = \frac{61.77}{5} = 12.35$ and $\sqrt{12.35 + (1.69)^2} = 3.90$. The P.I. is $52.88 \pm (2.571)(3.90) = 52.88 \pm 10.03 = (42.85, 62.91)$.

28.

a. $\hat{m}_{Y\cdot 75} = \hat{b}_0 + \hat{b}_1(75) + \hat{b}_2(75)^2 = -113.0937 + 3.36684(75) - .01780(75)^2 = 39.41$

b. $\hat{y} = \hat{b}_0 + \hat{b}_1(60) + \hat{b}_2(60)^2 = 24.93$.

c. $SSE = \Sigma y_i^2 - \hat{b}_0 \Sigma y_i - \hat{b}_1 \Sigma x_i y_i - \hat{b}_2 \Sigma x_i^2 y_i = 8386.43 - (-113.0937)(210.70)$
$- (3.3684)(17{,}002) - (-.0178)(1{,}419{,}780) = 217.82$,
$s^2 = \frac{SSE}{n-3} = \frac{217.82}{3} = 72.61$, s = 8.52

d. $R^2 = 1 - \frac{217.82}{987.35} = .779$

e. H_o will be rejected in favor of H_a if either $t \ge t_{.005,3} = 5.841$ or if $t \le -5.841$. The computed value of t is $t = \frac{-.01780}{.00226} = -7.88$, and since $-7.88 \le -5.841$, we reject H_o.

29.

a. From computer output:

$\hat{y}$:	111.89	120.66	114.71	94.06	58.69
$y-\hat{y}$:	-1.89	2.34	4.29	-8.06	3.31

Thus $SSE = (-1.89)^2 + \ldots + (3.31)^2 = 103.37$, $s^2 = \frac{103.37}{2} = 51.69$, $s = 7.19$.

b. $SST = \Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n} = 2630$, so $R^2 = 1 - \frac{103.37}{2630} = .961$.

c. $H_0 : b_2 = 0$ will be rejected in favor of $H_a : b_2 \neq 0$ if either $t \geq t_{.025,2} = 4.303$ or if $t \leq -4.303$. With $t = \frac{-1.84}{.480} = -3.83$, H_o cannot be rejected; the data does not argue strongly for the inclusion of the quadratic term.

d. To obtain joint confidence of at least 95%, we compute a 98% C.I. for each coefficient using $t_{.01,2} = 6.965$. For b_1 the C.I. is $8.06 \pm (6.965)(4.01) = (-19.87, 35.99)$ (an extremely wide interval), and for b_2 the C.I. is $-1.84 \pm (6.965)(.480)$ $= (-5.18, 1.50)$.

e. $t_{.05,2} = 2.920$ and $\hat{b}_0 + 4\hat{b}_1 + 16\hat{b}_2 = 114.71$, so the C.I. is $114.71 \pm (2.920)(5.01)$ $= 114.71 \pm 14.63 = (100.08, 129.34)$.

f. If we knew $\hat{b}_0, \hat{b}_1, \hat{b}_2$, the value of x which maximizes $\hat{b}_0 + \hat{b}_1 x + \hat{b}_2 x^2$ would be obtained by setting the derivative of this to 0 and solving:

$b_1 + 2b_2 x = 0 \Rightarrow x = -\frac{b_1}{2b_2}$. The estimate of this is $x = -\frac{\hat{b}_1}{2\hat{b}_2} = 2.19$.

30.

a. $R^2 = 0.853$. This means 85.3% of the variation in wheat yield is accounted for by the model.

b. $-135.44 \pm (2.201)(41.97) = (-227.82, -43.06)$

c. $H_0 : m_{y\cdot 2.5} = 1500$; $H_a : m_{y\cdot 2.5} < 1500$; $RR : t \le -t_{.01,11} = -2.718$

When x = 2.5, $\hat{y} = 140215$

$$t = \frac{1{,}402.15 - 1500}{53.5} = -1.83$$

Fail to reject H_o. The data does not indicate $m_{y\cdot 2.5}$ is less than 1500.

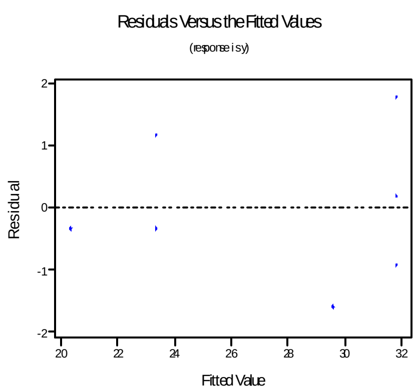
d. $140215 \pm (2.201)\sqrt{(136.5)^2 + (53.5)^2} = (1081.3, 1725.0)$

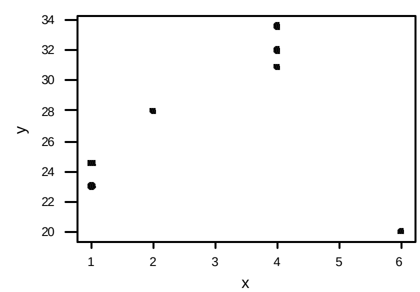
31.

a. Using Minitab, the regression equation is `y = 13.6 + 11.4x - 1.72x²`.

b. Again, using Minitab, the predicted and residual values are:

$\hat{y}$:	23.327	23.327	29.587	31.814	31.814	31.814	20.317
$y - \hat{y}$:	-.327	1.173	1.587	.914	.186	1.786	-.317





The residual plot is consistent with a quadratic model (no pattern which would suggest modification), but it is clear from the scatter plot that the point (6, 20) has had a great influence on the fit – it is the point which forced the fitted quadratic to have a maximum between 3 and 4 rather than, for example, continuing to curve slowly upward to a maximum someplace to the right of x = 6.

c. From Minitab output, s^2 = MSE = 2.040, and R^2 = 94.7%. The quadratic model thus explains 94.7% of the variation in the observed y's , which suggests that the model fits the data quite well.

d. $s^2 = Var(\hat{Y}_i) + Var\left(Y_i - \hat{Y}_i\right)$ suggests that we can estimate $Var\left(Y_i - \hat{Y}_i\right)$ by $s^2 - s_{\hat{y}}^2$ and then take the square root to obtain the estimated standard deviation of each residual. This gives $\sqrt{2.040 - (.955)^2} =$ 1.059, (and similarly for all points) 10.59, 1.236, 1.196, 1.196, 1.196, and .233 as the estimated std dev's of the residuals. The standardized residuals are then computed as $\frac{-.327}{1.059} = -.31$, (and similarly) 1.10, -1.28, -.76, .16, 1.49, and –1.28, none of which are unusually large. (Note: Minitab regression output can produce these values.) The resulting residual plot is virtually identical to the plot of **b**. $\frac{y - \hat{y}}{s} = \frac{-.327}{1.426} = -.229 \neq -.31$, so standardizing using just s would not yield the correct standardized residuals.

e. $Var(Y_f) + Var(\hat{Y}_f)$ is estimated by $2.040 + (.777)^2 = 2.638$, so $s_{y_f + \hat{y}_f} = \sqrt{2.638} = 1.624$. With $\hat{y} = 31.81$ and $t_{.05,4} = 2.132$, the desired P.I. is $31.81 \pm (2.132)(1.624) = (28.35, 35.27)$.

32.

a. $.3463 - 1.2933(x - \bar{x}) + 2.3964(x - \bar{x})^2 - 2.3968(x - \bar{x})^3$.

b. From **a**, the coefficient of x^3 is -2.3968, so $\hat{b}_3 = -2.3968$. There sill be a contribution to x^2 both from $2.3964(x - 4.3456)^2$ and from $-2.3968(x - 4.3456)^3$. Expanding these and adding yields 33.6430 as the coefficient of x^2, so $\hat{b}_2 = 33.6430$.

c. $x = 4.5 \Rightarrow x' = x - \bar{x} = .1544$; substituting into **a** yields $\hat{y} = .1949$.

d. $t = \frac{-2.3968}{2.4590} = -.97$, which is not significant ($H_0 : b_3 = 0$ cannot be rejected), so the inclusion of the cubic term is not justified.

33.

a. $\bar{x} = 20$ and s_x = 10.8012 so $x' = \dfrac{x-20}{10.8012}$. For x = 20, $x' = 0$, and $\hat{y} = \hat{b}_0^* = .9671$. For x = 25, $x' = .4629$, so $\hat{y} = .9671 - .0502(.4629) - .0176(.4629)^2 + .0062(.4629)^3 = .9407$.

b. $\hat{y} = .9671 - .0502\left(\dfrac{x-20}{10.8012}\right) - .0176\left(\dfrac{x-20}{10.8012}\right)^2 + .0062\left(\dfrac{x-20}{10.8012}\right)^3$
$.00000492x^3 - .000446058x^2 + .007290688x + .96034944$.

c. $t = \dfrac{.0062}{.0031} = 2.00$. We reject H_o if either $t \geq t_{.025,n-4} = t_{.025,3} = 3.182$ or if $t \leq -3.182$. Since 2.00 is neither ≥ 3.182 nor ≤ -3.182, we cannot reject H_o; the cubic term should be deleted.

d. $SSE = \Sigma(y_i - \hat{y}_i)$ and the $\hat{y}_i$'s are the same from the standardized as from the unstandardized model, so SSE, SST, and R^2 will be identical for the two models.

e. $\Sigma y_i^2 = 6.355538$, $\Sigma y_i = 6.664$, so SST = .011410. For the quadratic model R^2 = .987 and for the cubic mo del, R^2 = .994; The two R^2 values are very close, suggesting intuitively that the cubic term is relatively unimportant.

34.

a. $\bar{x} = 49.9231$ and s_x = 41.3652 so for x = 50, $x' = \dfrac{x-49.9231}{41.3652} = .001859$ and $\hat{m}_{Y\cdot 50} = .8733 - .3255(.001859) + .0448(.001859)^2 = .873$.

b. SST = 1.456923 and SSE = .117521, so R^2 = .919.

c. $.8733 - .3255\left(\dfrac{x-49.9231}{41.3652}\right) + .0448\left(\dfrac{x-49.9231}{41.3652}\right)^2$
$1.200887 - .01048314x + .00002618x^2$.

d. $\hat{b}_2 = \dfrac{\hat{b}_2^*}{s_x^2}$ so the estimated sd of $\hat{b}_2$ is the estimated sd of $\hat{b}_2^*$ multiplied by $\dfrac{1}{s_x}$:
$s_{\hat{b}_2} = (.0319)\left(\dfrac{1}{41.3652}\right) = .00077118$.

e. $t = \dfrac{.0448}{.0319} = 1.40$ which is not significant (compared to $\pm t_{.025,9}$ at level .05), so the quadratic term should not be retained.

35. $Y' = \ln(Y) = \ln a + bx + gx^2 + \ln(e) = b_0 + b_1 x + b_2 x^2 + e'$ where $e' = \ln(e)$, $b_0 = \ln(a)$, $b_1 = b$, and $b_2 = g$. That is, we should fit a quadratic to $(x, \ln(y))$. The resulting estimated quadratic (from computer output) is $2.00397 + .1799x - .0022x^2$, so $\hat{b} = .1799$, $\hat{g} = -.0022$, and $\hat{a} = e^{2.0397} = 7.6883$. (The ln(y)'s are 3.6136, 4.2499, 4.6977, 5.1773, and 5.4189, and the summary quantities can then be computed as before.)

Section 13.4

36.

a. Holding age, time, and heart rate constant, maximum oxygen uptake will increase by .01 L/min for each 1 kg increase in weight. Similarly, holding weight, age, and heart rate constant, the maximum oxygen uptake decreases by .13 L/min with every 1 minute increase in the time necessary to walk 1 mile.

b. $\hat{y}_{76,20,12,140} = 5.0 + .01(76) - .05(20) - .13(12) - .01(140) = 1.8$ L/min.

c. $\hat{y} = 1.8$ from **b**, and $s = .4$, so, assuming y follows a normal distribution,

$$P(1.00 < Y < 2.60) = P\left(\frac{1.00 - 1.8}{.4} < Z < \frac{2.6 - 1.8}{.4}\right) = P(-2.0 < Z < 2.0) = .9544$$

37.

a. The mean value of y when $x_1 = 50$ and $x_2 = 3$ is $m_{y \cdot 50,3} = -.800 + .060(50) + .900(3) = 4.9$ hours.

b. When the number of deliveries (x_2) is held fixed, then average change in travel time associated with a one-mile (i.e. one unit) increase in distance traveled (x_1) is .060 hours. Similarly, when distance traveled (x_1) is held fixed, then the average change in travel time associated with on extra delivery (i.e., a one unit increase in x_2) is .900 hours.

c. Under the assumption that y follows a normal distribution, the mean and standard deviation of this distribution are 4.9 (because $x_1 = 50$ and $x_2 = 3$) and $s = .5$ (since the standard deviation is assumed to be constant regardless of the values of x_1 and x_2). Therefore $P(y \le 6) = P\left(z \le \frac{6 - 4.9}{.5}\right) = P(z \le 2.20) = .9861$. That is, in the long run, about 98.6% of all days will result in a travel time of at most 6 hours.

38.

a. mean life $=125+7.75(40)+.0950(1100)-.009(40)(1100)=143.50$

b. First, the mean life when $x_1 = 30$ is equal to $125+7.75(30)+.0950x_2-.009(30)x_2 = 357.50-.175x_2$. So when the load increases by 1, the mean life decreases by .175. Second, the mean life when $x_1 = 40$ is equal to $125+7.75(40)+.0950x_2-.009(40)x_2 = 435-.265x_2$. So when the load increases by 1, the mean life decreases by .265.

39.

a. For $x_1 = 2$, $x_2 = 8$ (remember the units of x_2 are in 1000,s) and $x_3 = 1$ (since the outlet has a drive-up window) the average sales are $\hat{y}=10.00-1.2(2)+6.8(8)+15.3(1)=77.3$ (i.e., $77,300).

b. For $x_1 = 3$, $x_2 = 5$, and $x_3 = 0$ the average sales are $\hat{y}=10.00-1.2(3)+6.8(5)+15.3(0)=40.4$ (i.e., $40,400).

c. When the number of competing outlets (x_1) and the number of people within a 1-mile radius (x_2) remain fixed, the sales will increase by $15,300 when an outlet has a drive-up window.

40.

a. $\hat{m}_{Y\cdot 10,.5,50,100}=1.52+.02(10)-1.40(.5)+.02(50)-.0006(100)=1.96$

b. $\hat{m}_{Y\cdot 20,.5,50,30}=1.52+.02(20)-1.40(.5)+.02(50)-.0006(30)=1.40$

c. $\hat{b}_4=-.0006;\ 100\hat{b}_4=-.06$.

d. There are no interaction predictors – e.g., $x_5 = x_1x_4$ -- in the model. There would be dependence if interaction predictors involving x_4 had been included.

e. $R^2=1-\dfrac{20.0}{39.2}=.490$. For testing $H_0: b_1=b_2=b_3=b_4=0$ vs. H_a: at least one among $b_1,\ldots,b_4$ is not zero, the test statistic is $F=\dfrac{R^2/k}{(1-R^2)/(n-k-1)}$. H_o will be rejected if $f \ge F_{.05,4,25}=2.76$. $f=\dfrac{.490/4}{.510/25}=6.0$. Because $6.0 \ge 2.76$, H_o is rejected and the model is judged useful (this even though the value of R^2 is not all that impressive).

41. $H_0 : b_1 = b_2 = ... = b_6 = 0$ vs. H_a: at least one among $b_1, ..., b_6$ is not zero. The test statistic is $F = \dfrac{R^2/k}{(1-R^2)/(n-k-1)}$. H_o will be rejected if $f \geq F_{.05,6,30} = 2.42$.
$f = \dfrac{.83/6}{(1-.83)/30} = 24.41$. Because $24.41 \geq 2.42$, H_o is rejected and the model is judged useful.

42.

a. To test $H_0 : b_1 = b_2 = 0$ vs. $H_a :$ at least one $b_i \neq 0$, the test statistic is $f = \dfrac{MSR}{MSE} = 319.31$ (from output). The associated p-value is 0, so at any reasonable level of significance, H_o should be rejected. There does appear to be a useful linear relationship between temperature difference and at leas one of the two predictors.

b. The degrees of freedom for SSE = n – (k + 1) = 9 – (2 – 1) = 6 (which you could simply read in the DF column of the printout), and $t_{.025,6} = 2.447$, so the desired confidence interval is $3.000 \pm (2.447)(.4321) = 3.000 \pm 1.0573$, or about $(1.943, 4.057)$. Holding furnace temperature fixed, we estimate that the average change in temperature difference on the die surface will be somewhere between 1.943 and 4.057.

c. When x_1 = 1300 and x_2 = 7, the estimated average temperature difference is $\hat{y} = -199.56 + .2100x_1 + 3.000x_2 = -199.56 + .2100(1300) + 3.000(7) = 94.44$. The desired confidence interval is then $94.44 \pm (2.447)(.353) = 94.44 \pm .864$, or $(93.58, 95.30)$.

d. From the printout, s = 1.058, so the prediction interval is $94.44 \pm (2.447)\sqrt{(1.058)^2 + (.353)^2} = 94.44 \pm 2.729 = (91.71, 97.17)$.

43.

a. $x_1 = 2.6$, $x_2 = 250$, and $x_1x_2 = (2.6)(250) = 650$, so

$$\hat{y} = 185.49 - 45.97(2.6) - 0.3015(250) + 0.0888(650) = 48.313$$

b. No, it is not legitimate to interpret b_1 in this way. It is not possible to increase by 1 unit the cobalt content, x_1, while keeping the interaction predictor, x_3, fixed. When x_1 changes, so does x_3, since $x_3 = x_1x_2$.

c. Yes, there appears to be a useful linear relationship between y and the predictors. We determine this by observing that the p-value corresponding to the model utility test is < .0001 (F test statistic = 18.924).

d. We wish to test $H_0 : b_3 = 0$ vs. $H_a : b_3 \neq 0$. The test statistic is t=3.496, with a corresponding p-value of .0030. Since the p-value is < alpha = .01, we reject H_o and conclude that the interaction predictor does provide useful information about y.

e. A 95% C.I. for the mean value of surface area under the stated circumstances requires the following quantities:

$$\hat{y} = 185.49 - 45.97(2) - 0.3015(500) + 0.0888(2)(500) = 31.598$$. Next,

$t_{.025,16} = 2.120$, so the 95% confidence interval is

$$31.598 \pm (2.120)(4.69) = 31.598 \pm 9.9428 = (21.6552, 41.5408)$$

44.

a. Holding starch damage constant, for every 1% increase in flour protein, the absorption rate will increase by 1.44%. Similarly, holding flour protein percentage constant, the absorption rate will increase by .336% for every 1-unit increase in starch damage.

b. $R^2 = .96447$, so 96.447% of the observed variation in absorption can be explained by the model relationship.

c. To answer the question, we test $H_0 : b_1 = b_2 = 0$ vs $H_a :$ at least one $b_i \neq 0$. The test statistic is $f = 339.31092$, and has a corresponding p-value of zero, so at any significance level we will reject H_o. There is a useful relationship between absorption and at least one of the two predictor variables.

d. We would be testing $H_a : b_2 \neq 0$. We could calculate the test statistic $t = \dfrac{b_2}{s_{b_2}}$, or we could look at the 95% C.I. given in the output. Since the interval (.29828, 37298) does not contain the value 0, we can reject H_o and conclude that 'starch damage' should not be removed from the model.

e. The 95% C.I. is $42.253 \pm (2.060)(.350) = 42.253 \pm 0.721 = (41.532, 42.974)$.
The 95% P.I. is

$$42.253 \pm (2.060)\left(\sqrt{1.09412^2 + .350^2}\right) = 42.253 \pm 2.366 = (39.887, 44.619).$$

f. We test $H_a : b_3 \neq 0$, with $t = \dfrac{-.04304}{.01773} = -2.428$. The p-value is approximately 2(.012) = .024. At significance level .01 we do not reject H_o. The interaction term should not be retained.

45.

a. The appropriate hypotheses are $H_0 : b_1 = b_2 = b_3 = b_4 = 0$ vs. $H_a :$ at least one $b_i \neq 0$. The test statistic is $f = \dfrac{R^2/k}{(1-R^2)/(n-k-1)} = \dfrac{.946/4}{(1-.946)/20} = 87.6 \geq 7.10 = F_{.001,4,20}$ (the smallest available significance level from Table A.9), so we can reject H_o at any significance level. We conclude that at least one of the four predictor variables appears to provide useful information about tenacity.

b. The adjusted R^2 value is $1 - \dfrac{n-1}{n-(k+1)}\left(\dfrac{SSE}{SST}\right) = 1 - \dfrac{n-1}{n-(k+1)}\left(1 - R^2\right)$

$= 1 - \dfrac{24}{20}(1 - .946) = .935$, which does not differ much from $R^2 = .946$.

c. The estimated average tenacity when $x_1 = 16.5$, $x_2 = 50$, $x_3 = 3$, and $x_4 = 5$ is
$\hat{y} = 6.121 - .082x + .113x + .256x - .219x$
$\hat{y} = 6.121 - .082(16.5) + .113(50) + .256(3) - .219(5) = 10.091$. For a 99% C.I., $t_{.005,20} = 2.845$, so the interval is $10.091 \pm 2.845(.350) = (9.095, 11.087)$. Therefore, when the four predictors are as specified in this problem, the true average tenacity is estimated to be between 9.095 and 11.087.

46.

a. Yes, there does appear to be a useful linear relationship between repair time and the two model predictors. We determine this by conducting a model utility test:
$H_0 : b_1 = b_2 = 0$ vs. $H_a :$ at least one $b_i \neq 0$. We reject H_o if $f \geq F_{.05,2,9} = 4.26$.
The calculated statistic is $f = \dfrac{SSR/k}{SSE/(n-k-1)} = \dfrac{MSR}{MSE} = \dfrac{10.63/2}{(20.9)/9} = \dfrac{5.315}{.232} = 22.91$. Since $22.91 \geq 4.26$, we reject H_o and conclude that at least one of the two predictor variables is useful.

b. We will reject $H_0 : b_2 = 0$ in favor of $H_a : b_2 \neq 0$ if $|t| \geq t_{.005,9} = 3.25$. The test statistic is $t = \dfrac{1.250}{.312} = 4.01$ which is ≥ 3.25, so we reject H_o and conclude that the "type of repair" variable does provide useful information about repair time, given that the "elapsed time since the last service" variable remains in the model.

c. A 95% confidence interval for b_3 is: $1.250 \pm (2.262)(.312) = (.5443, 1.9557)$. We estimate, with a high degree of confidence, that when an electrical repair is required the repair time will be between .54 and 1.96 hours longer than when a mechanical repair is required, while the "elapsed time" predictor remains fixed.

d. $\hat{y} = .950 + .400(6) + 1.250(1) = 4.6$, $s^2 = MSE = .23222$, and $t_{.005,9} = 3.25$, so the 99% P.I. is $4.6 \pm (3.25)\sqrt{(.23222) + (.192)^2} = 4.6 \pm 1.69 = (2.91, 6.29)$ The prediction interval is quite wide, suggesting a variable estimate for repair time under these conditions.

47.

a. For a 1% increase in the percentage plastics, we would expect a 28.9 kcal/kg increase in energy content. Also, for a 1% increase in the moisture, we would expect a 37.4 kcal/kg decrease in energy content.

b. The appropriate hypotheses are $H_0 : b_1 = b_2 = b_3 = b_4 = 0$ vs. $H_a :$ at least one $b_i \neq 0$. The value of the F test statistic is 167.71, with a corresponding p-value that is extremely small. So, we reject H_o and conclude that at least one of the four predictors is useful in predicting energy content, using a linear model.

c. $H_0 : b_3 = 0$ vs. $H_a :\ b_3 \neq 0$. The value of the t test statistic is t = 2.24, with a corresponding p-value of .034, which is less than the significance level of .05. So we can reject H_o and conclude that percentage garbage provides useful information about energy consumption, given that the other three predictors remain in the model.

d. $\hat{y} = 2244.9 + 28.925(20) + 7.644(25) + 4.297(40) - 37.354(45) = 1505.5$, and $t_{.025,25} = 2.060$. (Note an error in the text: $s_{\hat{y}} = 12.47$, not 7.46). So a 95% C.I for the true average energy content under these circumstances is $1505.5 \pm (2.060)(12.47) = 1505.5 \pm 25.69 = (1479.8, 1531.1)$. Because the interval is reasonably narrow, we would conclude that the mean energy content has been precisely estimated.

e. A 95% prediction interval for the energy content of a waste sample having the specified characteristics is $1505.5 \pm (2.060)\sqrt{(31.48)^2 + (12.47)^2}$
$= 1505.5 \pm 69.75 = (1435.7, 1575.2)$.

48.

a. $H_0: b_1 = b_2 = ... = b_9 = 0$

H_a : at least one $b_i \neq 0$

RR: $f \geq F_{.01,9,5} = 10.16$

$$f = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{.938/9}{(1-.938)/(5)} = 8.41$$

Fail to reject H_o . The model does not appear to specify a useful relationship.

b. $\hat{m}_y = 21.967$, $t_{a/2,n-(k+1)} = t_{.025,5} = 2.571$, so the C.I. is

$21.967 \pm (2.571)(1.248) = (18.76, 25.18)$.

c. $s^2 = \dfrac{SSE}{n-(k+1)} = \dfrac{23.379}{5} = 4.6758$, and the C.I. is

$21.967 \pm (2.571)\sqrt{4.6758 + (1.248)^2} = (15.55, 28.39)$.

d. $SSE_k = 23.379, \ SSE_l = 203.82,$

$H_0: b_4 = b_5 = ... = b_9 = 0$

H_a : at least one of the above $b_i \neq 0$

RR: $f \geq F_{a,k-l,n-(k+1)} = F_{.05,6,5} = 4.95$

$$f = \frac{(203.82-23.379)/(9-3)}{(23.379)/(5)} = 6.43.$$

Reject H_o. At least one of the second order predictors appears useful.

49.

a. $\hat{m}_{y\cdot18.9,43} = 96.8303$; Residual = 91 – 96.8303 = -5.8303.

b. $H_0: b_1 = b_2 = 0$; H_a : at least one $b_i \neq 0$

RR: $f \geq F_{.05,2,9} = 8.02$

$$f = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{.768/2}{(1-.768)/9} = 14.90.$$ Reject H_o. The model appears useful.

c. $96.8303 \pm (2.262)(8.20) = (78.28, 115.38)$

d. $96.8303 \pm (2.262)\sqrt{24.45^2 + 8.20^2} = (38.50, 155.16)$

e. We find the center of the given 95% interval, 93.875, and half of the width, 57.845. This latter value is equal to $t_{.025,9}(s_{\hat{y}}) = 2.262(s_{\hat{y}})$, so $s_{\hat{y}} = 25.5725$. Then the 90% interval is $93.785 \pm (1.833)(25.5725) = (46.911, 140.659)$

f. With the p-value for $H_a : b_1 \neq 0$ being 0.208 (from given output), we would fail to reject H_o. This factor is not significant given x_2 is in the model.

g. With $R_k^2 = .768$ (full model) and $R_l^2 = .721$ (reduced model), we can use an alternative f statistic (compare formulas 13.19 and 13.20). $F = \dfrac{R_k^2 - R_l^2 \big/ k-l}{(1-R_k^2) \big/ n-(k+1)}$. With n=12, k=2 and l=1, we have $F = \dfrac{.768 - .721}{(1-.768)/9} = \dfrac{.047}{.0257} = 1.83$.

$t^2 = (-1.36)^2 = 1.85$. The discrepancy can be attributed to rounding error.

50.

a. Here k = 5, n – (k+1) = 6, so H_o will be rejected in favor of H_a at level .05 if either $t \geq t_{.025,6} = 2.447$ or $t \leq -2.447$. The computed value of t is $t = \dfrac{.557}{.94} = .59$, so H_o cannot be rejected and inclusion of x_1x_2 as a carrier in the model is not justified.

b. No, in the presence of the other four carriers, any particular carrier is relatively unimportant, but this is not equivalent to the statement that all carriers are unimportant.

c. $SSE_k = SST(1-R^2) = 3224.65,$ so $f = \dfrac{(5384.18 - 3224.65)/3}{(3224.65)/6} = 1.34$, and since 1.34 is not $\geq F_{.05,3,6} = 4.76$, H_o cannot be rejected; the data does not argue for the inclusion of any second order terms.

51.

a. No, there is no pattern in the plots which would indicate that a transformation or the inclusion of other terms in the model would produce a substantially better fit.

b. k = 5, n – (k+1) = 8, so $H_0 : b_1 = ... = b_5 = 0$ will be rejected if $f \geq F_{.05,5,8} = 3.69;$ $f = \dfrac{(.759)/5}{(.241)/8} = 5.04 \geq 3.69$, so we reject H_o. At least one of the coefficients is not equal to zero.

c. When $x_1 = 8.0$ and $x_2 = 33.1$ the residual is e = 2.71 and the standardized residual is e* = .44; since e* = e/(sd of the residual), sd of residual = e/e* = 6.16. Thus the estimated variance of $\hat{Y}$ is $(6.99)^2 - (6.16)^2 = 10.915$, so the estimated sd is 3.304. Since $\hat{y} = 24.29$ and $t_{.025,8} = 2.306$, the desired C.I. is $24.29 \pm 2.306(3.304) = (16.67, 31.91)$.

d. $F_{.05,3,8} = 4.07$, so $H_0 : b_3 = b_4 = b_5 = 0$ will be rejected if $f \geq 4.07$. With $SSE_k = 8, s^2 = 390.88,$ and $f = \dfrac{(894.95-390.88)/3}{(390.88)/8} = 3.44$, and since 3.44 is not ≥ 4.07, H_o cannot be rejected and the quadratic terms should all be deleted. (n.b.: this is not a modification which would be suggested by a residual plot.

52.

a. The complete 2nd order model obviously provides a better fit, so there is a need to account for interaction between the three predictors.

b. A 95% CI for y when x_1=x_2=30 and x_3=10 is $.66573 \pm 2.120(.01785) = (.6279, .7036)$

53. Some possible questions might be:

Is this model useful in predicting deposition of poly-aromatic hydrocarbons? A test of model utility gives us an F = 84.39, with a p-value of 0.000. Thus, the model is useful.

Is x_1 a significant predictor of y while holding x_2 constant? A test of $H_0 : b_1 = 0$ vs the two-tailed alternative gives us a t = 6.98 with a p-value of 0.000., so this predictor is significant.

A similar question, and solution for testing x_2 as a predictor yields a similar conclusion: With a p-value of 0.046, we would accept this predictor as significant if our significance level were anything larger than 0.046.

54.

a. For $x_1 = x_2 = x_3 = x_4 = +1$, $\hat{y} = 84.67 + .650 - .258 + ... + .050 = 85.390$. The single y corresponding to these x_i values is 85.4, so $y - \hat{y} = 85.4 - 85.390 = .010$.

b. Letting $x_1', ..., x_4'$ denote the uncoded variables, $x_1' = .1x_1 + .3,$ $x_2' = .1x_2 + .3,$ $x_3' = x_3 + 2.5,$ and $x_4' = 15x_4 + 160$; Substitution of $x_1 = 10x_1' - 3,$ $x_2 = 10x_2' - 3,$ $x_3 = x_3' - 2.5,$ and $x_4 = \dfrac{x_4' + 160}{15}$ yields the uncoded function.

c. For the full model k = 14 and for the reduced model l – 4, while n – (k + 1) = 16. Thus $H_0: b_5 = ... = b_{14} = 0$ will be rejected if $f \geq F_{.05,10,16} = 2.49$.
$SSE = (1 - R^2)SST$ so $SSE_k = 1.9845$ and $SSE_l = 4.8146$, giving
$f = \dfrac{(4.8146 - 1.9845)/10}{(1.9845)/16} = 2.28$. Since 2.28 is not ≥ 2.49, H_o cannot be rejected, so all higher order terms should be deleted.

d. $H_0: m_{Y\cdot 0,0,0,0} = 85.0$ will be rejected in favor of $H_a: m_{Y\cdot 0,0,0,0} < 85.0$ if $t \leq -t_{.05,26} = -1.706$. With $\hat{m} = \hat{b}_0 = 85.5548$, $t = \dfrac{85.5548 - 85}{.0772} = 7.19$, which is certainly not ≤ -1.706, so H_o is not rejected and prior belief is not contradicted by the data.

Section 13.5

55.

a. $\ln(Q) = Y = \ln(a) + b\ln(a) + g\ln(b) + \ln(e) = b_0 + b_1 x_1 + b_2 x_2 + e'$ where $x_1 = \ln(a), x_2 = \ln(b), b_0 = \ln(a), b_1 = b, b_2 = g$ and $e' = \ln(e)$. Thus we transform to $(y, x_1, x_2) = (\ln(Q), \ln(a), \ln(b))$ (take the natural log of the values of each variable) and do a multiple linear regression. A computer analysis gave $\hat{b}_0 = 1.5652$, $\hat{b}_1 = .9450$, and $\hat{b}_2 = .1815$. For a = 10 and b = .01, x_1 = ln(10) = 2.3026 and x_2 = ln(.01) = -4.6052, from which $\hat{y} = 2.9053$ and $\hat{Q} = e^{2.9053} = 18.27$.

b. Again taking the natural log, $Y = \ln(Q) = \ln(a) + ba + gb + \ln(e)$, so to fit this model it is necessary to take the natural log of each Q value (and not transform a or b) before using multiple regression analysis.

c. We simply exponentiate each endpoint: $(e^{.217}, e^{1.755}) = (1.24, 5.78)$.

56.

a. $n = 20, k = 5, n-(k+1) = 14$, so $H_0 : b_1 = ... = b_5 = 0$ will be rejected in favor of H_a : at least one among $b_1,..., b_5 \neq 0$, if $f \geq F_{.01,5,14} = 4.69$. With $f = \frac{(.769)/5}{(.231)/14} = 9.32 \geq 4.69$, so H_o is rejected. Wood specific gravity appears to be linearly related to at lest one of the five carriers.

b. For the full model, adjusted $R^2 = \frac{(19)(.769)-5}{14} = .687$, while for the reduced model, the adjusted $R^2 = \frac{(19)(.769)-4}{15} = .707$.

c. From **a**, $SSE_k = (.231)(.0196610) = .004542$, and $SSE_l = (.346)(.0196610) = .006803$, so $f = \frac{(.002261)/3}{(.004542)/14} = 2.32$. Since $F_{.05,3,14} = 3.34$ and 2.32 is not ≥ 3.34, we conclude that $b_1 = b_2 = b_4 = 0$.

d. $x_3' = \frac{x_3 - 52.540}{5.4447} = -.4665$ and $x_5' = \frac{x_5 - 89.195}{3.6660} = .2196$, so $\hat{y} = .5255 - (.0236)(-.4665) + (.0097)(.2196) = .5386$.

e. $t_{.025,17} = 2.110$ (error df = n – (k+1) = 20 – (2+1) = 17 for the two carrier model), so the desired C.I. is $-.0236 \pm 2.110(.0046) = (-.0333, -.0139)$.

f. $y = .5255 - .0236\left(\frac{x_3 - 52.540}{5.4447}\right) + .0097\left(\frac{x_5 - 89.195}{3.6660}\right)$, so $\hat{b}_3$ for the unstandardized model $= \frac{-.0236}{5.447} = -.004334$. The estimated sd of the unstandardized $\hat{b}_3$ is $= \frac{.0046}{5.447} = -.000845$.

g. $\hat{y} = .532$ and $\sqrt{s^2 + s_{\hat{b}_0 + \hat{b}_3 x_3' + \hat{b}_5 x_5'}} = .02058$, so the P.I. is $.532 \pm (2.110)(.02058) = .532 \pm .043 = (.489, .575)$.

57.

k	R^2	Adj. R^2	$C_k = \frac{SSE_k}{s^2} + 2(k+1) - n$
1	.676	.647	138.2
2	.979	.975	2.7
3	.9819	.976	3.2
4	.9824		4

Where $s^2 = 5.9825$

a. Clearly the model with k = 2 is recommended on all counts.

b. No. Forward selection would let x_4 enter first and would not delete it at the next stage.

58. At step #1 (in which the model with all 4 predictors was fit), t = .83 was the t ratio smallest in absolute magnitude. The corresponding predictor x_3 was then dropped from the model, and a model with predictors x_1, x_2, and x_4 was fit. The t ratio for x_4 , -1.53, was the smallest in absolute magnitude and 1.53 < 2.00, so the predictor x_4 was deleted. When the model with predictors x_1 and x_2 only was fit, both t ratios considerably exceeded 2 in absolute value, so no further deletion is necessary.

59. The choice of a "best" model seems reasonably clear–cut. The model with 4 variables including all but the summerwood fiber variable would seem bests. R^2 is as large as any of the models, including the 5 variable model. R^2 adjusted is at its maximum and CP is at its minimum . As a second choice, one might consider the model with k = 3 which excludes the summerwood fiber and springwood % variables.

60. Backwards Stepping:

Step 1: A model with all 5 variables is fit; the smallest t-ratio is t = .12, associated with variable x_2 (summerwood fiber %). Since t = .12 < 2, the variable x_2 was eliminated.
Step 2: A model with all variables except x_2 was fit. Variable x_4 (springwood light absorption) has the smallest t-ratio (t = -1.76), whose magnitude is smaller than 2. Therefore, x_4 is the next variable to be eliminated.
Step 3: A model with variables x_3 and x_5 is fit. Both t-ratios have magnitudes that exceed 2, so both variables are kept and the backwards stepping procedure stops at this step. The final model identified by the backwards stepping method is the one containing x_3 and x_5.

(continued)

Forward Stepping:

Step 1: After fitting all 5 one-variable models, the model with x_3 had the t-ratio with the largest magnitude (t = -4.82). Because the absolute value of this t-ratio exceeds 2, x_3 was the first variable to enter the model.

Step 2: All 4 two-variable models that include x_3 were fit. That is, the models $\{x_3, x_1\}$, $\{x_3, x_2\}$, $\{x_3, x_4\}$, $\{x_3, x_5\}$ were all fit. Of all 4 models, the t-ratio 2.12 (for variable x_5) was largest in absolute value. Because this t-ratio exceeds 2, x_5 is the next variable to enter the mo del.

Step 3: (not printed): All possible tree-variable models involving x_3 and x_5 and another predictor, None of the t-ratios for the added variables have absolute values that exceed 2, so no more variables are added. There is no need to print anything in this case, so the results of these tests are not shown.

Note; Both the forwards and backwards stepping methods arrived at the same final model, $\{x_3, x_5\}$, in this problem. This often happens, but not always. There are cases when the different stepwise methods will arrive at slightly different collections of predictor variables.

61. If multicollinearity were present, at least one of the four R^2 values would be very close to 1, which is not the case. Therefore, we conclude that multicollinearity is not a problem in this data.

62. Looking at the h_{ii} column and using $\frac{2(k+1)}{n} = \frac{8}{19} = .421$ as the criteria, three observations appear to have large influence. With h_{ii} values of .712933, .516298, and .513214, observations 14, 15, 16, correspond to response (y) values 22.8, 41.8, and 48.6.

63. We would need to investigate further the impact these two observations have on the equation. Removing observation #7 is reasonable, but removing #67 should be considered as well, before regressing again.

64.

a. $\frac{2(k+1)}{n} = \frac{6}{10} = .6;$ since $h_{44} > .6$, data point #4 would appear to have large influence.
(Note: Formulas involving matrix algebra appear in the first edition.)

b. For data point #2, $x'_{(2)} = (1 \quad 3.453 \quad -4.920)$, so $\hat{b} - \hat{b}_{(2)} =$

$$\frac{-.766}{1-.302}(X'X)^{-1}\begin{pmatrix}1\\3.453\\-4.920\end{pmatrix} = -1.0974\begin{pmatrix}.3032\\.1644\\.1156\end{pmatrix} = \begin{pmatrix}-.333\\-.180\\-.127\end{pmatrix}$$ and similar

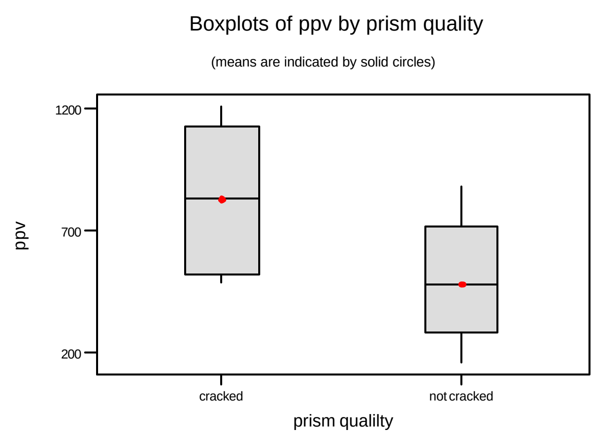
calculations yield $\hat{b} - \hat{b}_{(4)} = \begin{pmatrix}.106\\-.040\\.030\end{pmatrix}$.

c. Comparing the changes in the $\hat{b}_i$'s to the $s_{\hat{b}_i}$'s, none of the changes is all that substantial (the largest is 1.2sd's for the change in $\hat{b}_1$ when point #2 is deleted). Thus although h_{44} is large, indicating a potential high influence of point #4 on the fit, the actual influence does not appear to be great.

Supplementary Exercises

65.

a.



A two-sample t confidence interval, generated by Minitab:

```
Two sample T for ppv

prism qu      N       Mean      StDev   SE Mean
cracked      12        827        295        85
not cracke   18        483        234        55

95% CI for mu (cracked    ) - mu (not cracke): ( 132,  557)
```

b. The simple linear regression results in a significant model, r^2 is .577, but we have an extreme observation, with std resid = -4.11. Minitab output is below. Also run, but not included here was a model with an indicator for cracked/ not cracked, and for a model with the indicator and an interaction term. Neither improved the fit significantly.

```
The regression equation is
ratio = 1.00 -0.000018 ppv

Predictor        Coef       StDev          T        P
Constant      1.00161     0.00204     491.18    0.000
ppv       -0.00001827  0.00000295      -6.19    0.000

S = 0.004892    R-Sq = 57.7%     R-Sq(adj) = 56.2%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1  0.00091571  0.00091571     38.26    0.000
Residual Error    28  0.00067016  0.00002393
Total             29  0.00158587

Unusual Observations
Obs       ppv      ratio         Fit   StDev Fit    Residual    St Resid
 29      1144   0.962000    0.980704    0.001786   -0.018704      -4.11R

R denotes an observation with a large standardized residual
```

66.

a. For every 1 cm^{-1} increase in inverse foil thickness (x), we estimate that we would expect steady-state permeation flux to increase by $.26042 mA/cm^2$. Also, 98% of the observed variation in steady-state permeation flux can be explained by its relationship to inverse foil thickness.

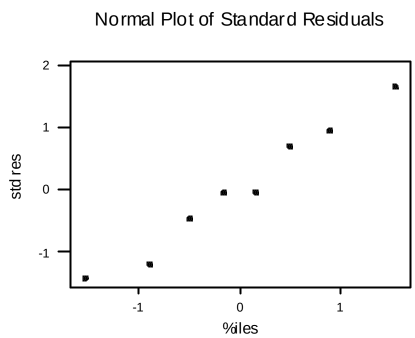
b. A point estimate of flux when inverse foil thickness is 23.5 can be found in the Observation 3 row of the Minitab output: $\hat{y} = 5.722 mA/cm^2$.

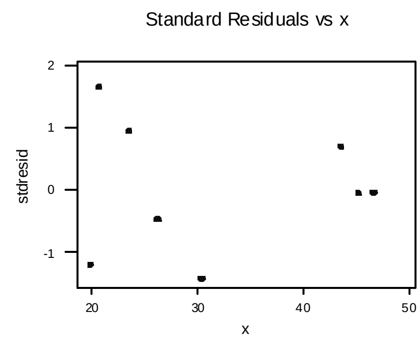
c. To test model usefulness, we test the hypotheses $H_0 : b_1 = 0$ vs. $H_a : b_1 \neq 0$. The test statistic is t = 17034, with associated p-value of .000, which is less than any significance level, so we reject H_o and conclude that the model is useful.

d. With $t_{.025,6} = 2.447$, a 95% Prediction interval for $Y_{(45)}$ is

$$11.321 \pm 2.447\sqrt{.203 + (.253)^2} = 11.321 \pm 1.264 = (10.057, 12.585).$$ That is, we are confident that when inverse foil thickness is 45 cm^{-1}, a predicted value of steady-state flux will be between 10.057 and 12.585 mA/cm^2.

e.





The normal plot gives no indication to question the normality assumption, and the residual plots against both x and y (only vs x shown) show no detectable pattern, so we judge the model adequate.

67.

a. For a one-minute increase in the 1-mile walk time, we would expect the VO_2max to decrease by .0996, while keeping the other predictor variables fixed.

b. We would expect male to have an increase of .6566 in VO_2max over females, while keeping the other predictor variables fixed.

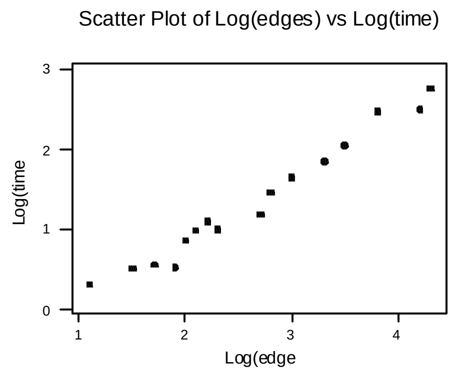
c. $\hat{y} = 3.5959 + .6566(1) + .0096(170) - .0996(11) - .0880(140) = 3.67$. The residual is $\hat{y} = (3.15 - 3.67) = -.52$.

d. $R^2 = 1 - \dfrac{SSE}{SST} = 1 - \dfrac{30.1033}{102.3922} = .706,$ or 70.6% of the observed variations in VO_2max can be attributed to the model relationship.

e. $H_0 : b_1 = b_2 = b_3 = b_4 = 0$ will be rejected in favor of H_a : at least one among $b_1, ..., b_4 \neq 0$, if $f \geq F_{.05,4,15} = 8.25$. With $f = \dfrac{(.706)/4}{(1-.706)/15} = 9.005 \geq 8.25$, so H_o is rejected. It appears that the model specifies a useful relationship between VO_2max and at least one of the other predictors.

68.

a.



Yes, the scatter plot of the two transformed variables appears quite linear, and thus suggests a linear relationship between the two.

b. Letting y denote the variable 'time', the regression model for the variables y' and x' is $\log_{10}(y) = y' = \alpha + \beta x' + \varepsilon'$. Exponentiating (taking the antilogs of) both sides gives $y = 10^{\alpha + \beta \log(x) + \varepsilon'} = (10^{\alpha})(x^{\beta})10^{\varepsilon'} = \gamma_0 x^{\gamma_1} \cdot \varepsilon$; i.e., the model is $y = \gamma_0 x^{\gamma_1} \cdot \varepsilon$ where $\gamma_0 = \alpha$ and $\gamma_1 = \beta$. This model is often called a "power function" regression model.

c. Using the transformed variables y' and x', the necessary sums of squares are

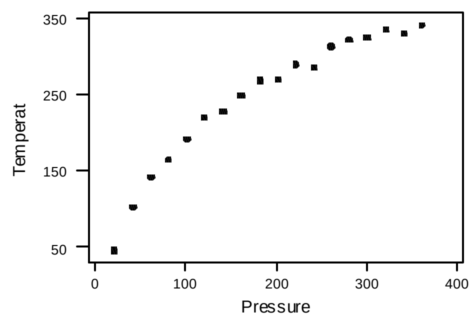
$$S_{x'y'} = 68.640 - \frac{(42.4)(21.69)}{16} = 11.1615 \text{ and}$$

$$S_{x'x'} = 126.34 - \frac{(42.4)^2}{16} = 13.98 \text{. Therefore } \hat{\beta}_1 = \frac{S_{x'y'}}{S_{x'x'}} = \frac{11.1615}{13.98} = .79839$$

and $\hat{\beta}_0 = \frac{21.69}{16} - (.79839)\left(\frac{42.4}{16}\right) = -.76011$. The estimate of γ_1 is $\hat{\gamma}_1 = .7984$ and $\gamma_0 = 10^{\alpha} = 10^{-.76011} = .1737$. The estimated power function model is then $y = .1737x^{.7984}$. For x = 300, the predicted value of y is $\hat{y} = .1737(300)^{.7984}16.502$, or about 16.5 seconds.

69.

a. Based on a scatter plot (below), a simple linear regression model would not be appropriate. Because of the slight, but obvious curvature, a quadratic model would probably be more appropriate.



b. Using a quadratic model, a Minitab generated regression equation is $\hat{y} = 35.423 + 1.7191x - .0024753x^2$, and a point estimate of temperature when pressure is 200 is $\hat{y} = 280.23$. Minitab will also generate a 95% prediction interval of (256.25, 304.22). That is, we are confident that when pressure is 200 psi, a single value of temperature will be between 256.25 and 304.22 $^{\circ}F$.

70.

a. For the model excluding the interaction term, $R^2 = 1 - \frac{5.18}{8.55} = .394$, or 39.4% of the observed variation in lift/drag ratio can be explained by the model without the interaction accounted for. However, including the interaction term increases the amount of variation in lift/drag ratio that can be explained by the model to $R^2 = 1 - \frac{3.07}{8.55} = .641$, or 64.1%.

b. Without interaction, we are testing $H_0 : b_1 = b_2 = 0$ vs. $H_a :$ either b_1 or $b_2 \neq 0$. The test statistic is $f = \frac{R^2/k}{(1-R^2)/(n-k-1)}$, The rejection region is $f \geq F_{.05,2,6} = 5.14$, and the calculated statistic is $f = \frac{.394/2}{(1-.394)/6} = 1.95$, which does not fall in the rejection region, so we fail to reject H_o. This model is not useful. With the interaction term, we are testing $H_0 : b_1 = b_2 = b_3 = 0$ vs. $H_a :$ at least one of the $b_i's \neq 0$. With rejection region $f \geq F_{.05,3,5} = 5.41$ and calculated statistic $f = \frac{.641/3}{(1-.641)/5} = 2.98$, we still fail to reject the null hypothesis. Even with the interaction term, there is not enough of a significant relationship between lift/drag ratio and the two predictor variables to make the model useful (a bit of a surprise!)

71.

a. Using Minitab to generate the first order regression model, we test the model utility (to see if any of the predictors are useful), and with $f = 21.03$ and a p-value of .000, we determine that at least one of the predictors is useful in predicting palladium content. Looking at the individual predictors, the p-value associated with the pH predictor has value .169, which would indicate that this predictor is unimportant in the presence of the others.

b. Testing $H_0 : b_1 = ... = b_{20} = 0$ vs. $H_a :$ at least one of the $b_i's \neq 0$. With calculated statistic $f = 6.29$, and p-value .002, this model is also useful at any reasonable significance level.

c. Testing $H_0 : b_6 = ... = b_{20} = 0$ vs. $H_a :$ at least one of the listed $b_i's \neq 0$, the test statistic is $f = \frac{(SSE_l - SSE_k)/(k-l)}{(SSE_k)/(n-k-1)} = \frac{(716.10-290.27)/(20-5)}{290.27/(32-20-1)} = 1.07$. Using significance level .05, the rejection region would be $f \geq F_{.05,15,11} = 2.72$. Since 1.07 < 2.72, we fail to reject H_o and conclude that all the quadratic and interaction terms should not be included in the model. They do not add enough information to make this model significantly better than the simple first order model.

d. Partial output from Minitab follows, which shows all predictors as significant at level .05:

```
The regression equation is
pdconc = - 305 + 0.405 niconc + 69.3 pH - 0.161 temp + 0.993 currdens
           + 0.355 pallcont - 4.14 pHsq

Predictor        Coef       StDev          T        P
Constant      -304.85       93.98      -3.24    0.003
niconc        0.40484     0.09432       4.29    0.000
pH              69.27       21.96       3.15    0.004
temp         -0.16134     0.07055      -2.29    0.031
currdens       0.9929      0.3570       2.78    0.010
pallcont      0.35460     0.03381      10.49    0.000
pHsq           -4.138       1.293      -3.20    0.004
```

72.

a. $R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{.80017}{16.18555} = .9506$, or 95.06% of the observed variation in weld strength can be attributed to the given model.

b. The complete second order model consists of nine predictors and nine corresponding coefficients. The hypotheses are $H_0 : b_1 = ... = b_9 = 0$ vs. $H_a :$ at least one of the b_i's $\neq 0$. The test statistic is $f = \frac{R^2/k}{(1-R^2)/(n-k-1)}$, where k = 9, and n = 37.The rejection region is $f \geq F_{.05,9,27} = 2.25$. The calculated statistic is $f = \frac{.9506/9}{(1-.9506)/27} = 57.68$ which is ≥ 2.25, so we reject the null hypothesis. The complete second order model is useful.

c. To test $H_0 : b_7 = 0$ vs $H_a : b_7 \neq 0$ (the coefficient corresponding to the wc*wt predictor), $t = \sqrt{f} = \sqrt{2.32} = 1.52$. With df = 27, the p-value $\approx 2(.073) = .146$ (from Table A.8). With such a large p-value, this predictor is not useful in the presence of all the others, so it can be eliminated.

d. The point estimate is $\hat{y} = 3.352 + .098(10) + .222(12) + .297(6) - .0102(10^2) - .037(6^2) + .0128(10)(12) = 7.962$. With $t_{.025,27} = 2.052$, the 95% P.I. would be $7.962 \pm 2.052(.0750) = 7.962 \pm .154 = (7.808, 8.116)$. Because of the narrowness of the interval, it appears that the value of strength can be accurately predicted.

73.

a. We wish to test $H_0 : b_1 = b_2 = 0$ vs. $H_a :$ either b_1 or $b_2 \neq 0$. The test statistic is $f = \frac{R^2/k}{(1-R^2)/(n-k-1)}$, where k = 2 for the quadratic model. The rejection region is $f \geq F_{\alpha,k,n-k-1} = F_{.01,2,5} = 13.27$. $R^2 = 1 - \frac{.29}{202.88} = .9986$, giving f = 1783. No doubt about it, folks – the quadratic model is useful!

b. The relevant hypotheses are $H_0 : b_2 = 0$ vs. $H_a : b_2 \neq 0$. The test statistic value is $t = \frac{\hat{b}_2}{s_{\hat{b}_2}}$, and H_o will be rejected at level .001 if either $t \geq 6.869$ or $t \leq -6.869$ (df = n – 3 = 5). Since $t = \frac{-.00163141}{.00003391} = -48.1 \leq -6.869$, H_o is rejected. The quadratic predictor should be retained.

c. No. R^2 is extremely high for the quadratic model, so the marginal benefit of including the cubic predictor would be essentially nil – and a scatter plot doesn't show the type of curvature associated with a cubic model.

d. $t_{.025,5} = 2.571$, and $\hat{b}_0 + \hat{b}_1(100) + \hat{b}_2(100)^2 = 21.36$, so the C.I. is $21.36 \pm (2.571)(.1141) = 21.36 \pm .69 = (20.67, 22.05)$

e. First, we need to figure out s^2 based on the information we have been given. $s^2 = MSE = \frac{SSE}{df} = \frac{.29}{5} = .058$. Then, the 95% P.I. is $21.36 \pm 2.571\left(\sqrt{.058 + .1141}\right) = 21.36 \pm 1.067 = (20.293, 22.427)$

74. A scatter plot of $y' = \log_{10}(y)$ vs. x shows a substantial linear pattern, suggesting the model $Y = a \cdot (10)^{bx} \cdot e$, i.e. $Y' = \log(a) + bx + \log(e) = b_0 + b_1 x + e'$. The necessary summary quantities are
$\Sigma x_i = 397,\ \Sigma x_i^2 = 14{,}263,\ \Sigma y_i' = -74.3,\ \Sigma y_i'^2 = 47{,}081,$ and $\Sigma x_i y_i' = -2358.1$,
giving $\hat{b}_1 = \dfrac{12(-2358.1) - (397)(-74.3)}{12(14{,}263) - (397)^2} = .08857312$ and $\hat{b}_0 = -9.12196058$.
Thus $\hat{b} = .08857312$ and $a = 10^{-9.12196058}$. The predicted value of y' when x = 35 is $-9.12196058 + .08857312(35) = -6.0219$, so $\hat{y} = 10^{-6.0219}$.

75.

a. $H_0 : b_1 = b_2 = 0$ will be rejected in favor of H_a : either b_1 or $b_2 \neq 0$ if
$f = \dfrac{R^2/k}{(1-R^2)/(n-k-1)} \geq F_{a,k,n-k-1} = F_{.01,2,7} = 9.55$. $SST = \Sigma y^2 - \dfrac{(\Sigma y)}{n} = 264.5$, so
$R^2 = 1 - \dfrac{26.98}{264.5} = .898$, and $f = \dfrac{.898/2}{(.102)/7} = 30.8$. Because $30.8 \geq 9.55$ H_o is rejected at significance level .01 and the quadratic model is judged useful.

b. The hypotheses are $H_0 : b_2 = 0$ vs. $H_a :\ b_2 \neq 0$. The test statistic value is
$t = \dfrac{\hat{b}_2}{s_{\hat{b}_2}} = \dfrac{-2.3621}{.3073} = -7.69$, and $t_{.0005,7} = 5.408$, so H_o is rejected at level .001 and p-value < .001. The quadratic predictor should not be eliminated.

c. x = 1 here, and $\hat{m}_{Y \cdot 1} = \hat{b}_0 + \hat{b}_1(1) + \hat{b}_2(1)^2 = 45.96$. $t_{.025,7} = 1.895$, giving the C.I. $45.96 \pm (1.895)(1.031) = (44.01, 47.91)$.

76.

a. 80.79

b. Yes, p-value = .007 which is less than .01.

c. No, p-value = .043 which is less than .05.

d. $.14167 \pm (2.447)(.03301) = (.0609, .2224)$

e. $\hat{m}_{y \cdot 9,66} = 6.3067$, using $a = .05$, the interval is
$6.3067 \pm (2.447)\sqrt{(.4851)^2 + (.162)^2} = (5.06, 7.56)$

77.

a. Estimate = $\hat{b}_0 + \hat{b}_1(15) + \hat{b}_2(3.5)^2 = 180 + (1)(15) + (10.5)(3.5) = 231.75$

b. $R^2 = 1 - \dfrac{117.4}{1210.30} = .903$

c. $H_0: b_1 = b_2 = 0$ vs. H_a : either b_1 or $b_2 \neq 0$ (or both). $f = \dfrac{.903/2}{.097/9} = 41.9$, which greatly exceeds $F_{.01,2,9}$ so there appears to be a useful linear relationship.

d. $s^2 = \dfrac{117.40}{12-3} = 13.044$, $\sqrt{s^2 + (est.st.dev)^2} = 3.806$, $t_{.025,9} = 2.262$. The P.I. is
$229.5 \pm (2.262)(3.806) = (220.9, 238.1)$

78. The second order model has predictors $x_1, x_2, x_3, x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3$ with corresponding coefficients $b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9$. We wish to test $H_0: b_4 = b_5 = b_6 = b_7 = b_8 = b_9 = 0$ vs. the alternative that at least one of these six b_i's is not zero. The test statistic value is $f = \dfrac{(821.5 - 5027.1)/(9-3)}{(5027.1)/(20-10)} = \dfrac{530.9}{502.71} = 1.1$. Since 1.1 < $F_{.05,6,10} = 3.22$, H_0 cannot be rejected. It doesn't appear as though any of the quadratic or interaction carriers should be included in the model.

79. There are obviously several reasonable choices in each case.

a. The model with 6 carriers is a defensible choice on all three grounds, as are those with 7 and 8 carriers.

b. The models with 7, 8, or 9 carriers here merit serious consideration. These models merit consideration because R_k^2, MSE_k, and CK meet the variable selection criteria given in Section 13.5.

80.

a. $f = \dfrac{(.90)/(15)}{(.10)/(4)} = 2.4$. Because 2.4 < 5.86, $H_0 : b_1 = ... = b_{15} = 0$ cannot be rejected. There does not appear to be a useful linear relationship.

b. The high R^2 value resulted from saturating the model with predictors. In general, one would be suspicious of a model yielding a high R^2 value when K is large relative to n.

c. $\dfrac{(R^2)/(15)}{(1-R^2)/(4)} \geq 5.86$ iff $\dfrac{R^2}{1-R^2} \geq 21.975$ iff $R^2 \geq \dfrac{21.975}{22.975} = .9565$

81.

a. The relevant hypotheses are $H_0 : b_1 = ... = b_5 = 0$ vs. H_a: at least one among $b_1, ..., b_5$ is not 0. $F_{.05,5,111} = 2.29$ and $f = \dfrac{(.827)/(5)}{(.173)/(111)} = 106.1$. Because $106.1 \geq 2.29$, H_o is rejected in favor of the conclusion that there is a useful linear relationship between Y and at least one of the predictors.

b. $t_{.05,111} = 1.66$, so the C.I. is $.041 \pm (1.66)(.016) = .041 \pm .027 = (.014,.068)$. b_1 is the expected change in mortality rate associated with a one-unit increase in the particle reading when the other four predictors are held fixed; we cab be 90% confident that .014 < b_1 < .068.

c. $H_0 : b_4 = 0$ will be rejected in favor of $H_a : b_4 \neq 0$ if $t = \dfrac{\hat{b}_4}{s_{\hat{b}_4}}$ is either ≥ 2.62 or ≤ -2.62. $t = \dfrac{.014}{.007} = 5.9 \geq 2.62$, so H_o is rejected and this predictor is judged important.

d. $\hat{y} = 19.607 + .041(166) + .071(60) + .001(788) + .041(68) + .687(.95) = 99.514$ and the corresponding residual is 103 – 99.514 = 3.486.

82.

a. The set $x_1, x_3, x_4, x_5, x_6, x_8$ includes both x_1, x_4, x_5, x_8 and x_1, x_3, x_5, x_6, so $R^2_{1,3,4,5,6,8} \geq \max\left(R^2_{1,4,5,8}, R^2_{1,3,5,6}\right) = .723$.

b. $R^2_{1,4} \leq R^2_{1,4,5,8} = .723$, but it is not necessarily $\leq .689$ since x_1, x_4 is not a subset of x_1, x_3, x_5, x_6.

CHAPTER 14

Section 14.1

1.

a. We reject H_o if the calculated χ^2 value is greater than or equal to the tabled value of $\chi^2_{\alpha,k-1}$ from Table A.7. Since $12.25 \ge \chi^2_{.05,4} = 9.488$, we would reject H_o.

b. Since 8.54 is not $\ge \chi^2_{.01,3} = 11.344$, we would fail to reject H_o.

c. Since 4.36 is not $\ge \chi^2_{.10,2} = 4.605$, we would fail to reject H_o.

d. Since 10.20 is not $\ge \chi^2_{.01,5} = 15.085$, we would fail to reject H_o.

2.

a. In the d.f. = 2 row of Table A.7, our χ^2 value of 7.5 falls between $\chi^2_{.025,2} = 7.378$ and $\chi^2_{.01,2} = 9.210$, so the p-value is between .01 and .025, or .01 < p-value < .025.

b. With d.f. = 6, our χ^2 value of 13.00 falls between $\chi^2_{.05,6} = 12.592$ and $\chi^2_{.025,6} = 14.440$, so .025 < p-value < .05.

c. With d.f. = 9, our χ^2 value of 18.00 falls between $\chi^2_{.05,9} = 16.919$ and $\chi^2_{.025,9} = 19.022$, so .025 < p-value < .05.

d. With k = 5, d.f. = k – 1 = 4, and our χ^2 value of 21.3 exceeds $\chi^2_{.005,4} = 14.860$, so the p-value < .005.

e. The d.f. = k – 1 = 4 – 1 = 3; $\chi^2 = 5.0$ is less than $\chi^2_{.10,3} = 6.251$, so p-value > .10.

3. Using the number 1 for business, 2 for engineering, 3 for social science, and 4 for agriculture, let $p_i =$ the true proportion of all clients from discipline i. If the Statistics department's expectations are correct, then the relevant null hypothesis is
$H_o : p_1 = .40, p_2 = .30, p_3 = .20, p_4 = .10,$ versus $H_a :$ The Statistics department's expectations are not correct. With d.f = k – 1 = 4 – 1 = 3, we reject H_o if
$c^2 \geq c^2_{.05,3} = 7.815$. Using the proportions in H_o, the expected number of clients are :

Client's Discipline	Expected Number
Business	(120)(.40) = 48
Engineering	(120)(.30) = 36
Social Science	(120)(.20) = 24
Agriculture	(120)(.10) = 12

Since all the expected counts are at least 5, the chi-squared test can be used. The value of the test statistic is $c^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} = \sum_{allcells} \frac{(observed - \exp ected)^2}{\exp ected}$

$= \left[\frac{(52-48)^2}{48} + \frac{(38-36)^2}{36} + \frac{(21-24)^2}{24} + \frac{(9-12)^2}{12} \right] = 1.57$, which is not

≥ 7.815, so we fail to reject H_o. (Alternatively, p-value = $P(c^2 \geq 1.57)$ which is > .10, and since the p-value is not < .05, we reject H_o). Thus we have no evidence to suggest that the statistics department's expectations are incorrect.

4. The uniform hypothesis implies that $p_{i0} = \frac{1}{8} = .125$ for I = 1, …, 8, so
$H_o : p_{10} = p_{20} = ... = p_{80} = .125$ will be rejected in favor of H_a if
$c^2 \geq c^2_{.10,7} = 12.017$. Each expected count is np_{i0} = 120(.125) = 15, so
$c^2 = \left[\frac{(12-15)^2}{15} + ... + \frac{(10-15)^2}{15} \right] = 4.80$. Because 4.80 is not ≥ 12.017, we fail to reject H_o. There is not enough evidence to disprove the claim.

5. We will reject H_o if the p-value < .10. The observed values, expected values, and corresponding χ^2 terms are :

Obs	4	15	23	25	38	21	32	14	10	8
Exp	6.67	13.33	20	26.67	33.33	33.33	26.67	20	13.33	6.67
χ^2	1.069	.209	.450	.105	.654	.163	1.065	1.800	.832	.265

$\chi^2 = 1.069 + ... + .265 = 6.612$. With d.f. = 10 – 1 = 9, our χ^2 value of 6.612 is less than $\chi^2_{.10,9} = 14.684$, so the p-value > .10, which is not < .10, so we cannot reject H_o. There is no evidence that the data is not consistent with the previously determined proportions.

6. A 9:3:4 ratio implies that $p_{10} = \frac{9}{16} = .5625$, $p_{20} = \frac{3}{16} = .1875$, and $p_{30} = \frac{4}{16} = .2500$. With n = 195 + 73 + 100 = 368, the expected counts are 207.000, 69.000, and 92.000, so $\chi^2 = \left[\frac{(195-207)^2}{207} + \frac{(73-69)^2}{69} + \frac{(100-92)^2}{92} \right] = 1.623$. With d.f. = 3 – 1 = 2, our χ^2 value of 1.623 is less than $\chi^2_{.10,2} = 4.605$, so the p-value > .10, which is not < .05, so we cannot reject H_o. The data does confirm the 9:3:4 theory.

7. We test $H_o : p_1 = p_2 = p_3 = p_4 = .25$ vs. $H_a :$ at least one proportion $\neq .25$, and d.f. = 3. We will reject H_o if the p-value < .01.

Cell	1	2	3	4
Observed	328	334	372	327
Expected	340.25	340.25	340.25	34.025
χ^2 term	.4410	.1148	2.9627	.5160

$\chi^2 = 4.0345$, and with 3 d.f., p-value > .10, so we fail to reject H_o. The data fails to indicate a seasonal relationship with incidence of violent crime.

8. $H_o: p_1 = \frac{15}{365}, p_2 = \frac{46}{365}, p_3 = \frac{120}{365}, p_4 = \frac{184}{365},$ versus H_a : at least one proportion is not a stated in H_o. The degrees of freedom = 3, and the rejection region is $\chi^2 \ge \chi^2_{.01,3} = 11.344$.

Cell	1	2	3	4
Observed	11	24	69	96
Expected	8.22	25.21	65.75	100.82
χ^2 term	.9402	.0581	.1606	.2304

$\chi^2 = \sum \frac{(obs - \exp)^2}{\exp} = 1.3893$, which is not ≥ 11.344, so H_o is not rejected. The data does not indicate a relationship between patients' admission date and birthday.

9.

a. Denoting the 5 intervals by $[0, c_1), [c_1, c_2), \ldots, [c_4, \infty)$, we wish c_1 for which $.2 = P(0 \le X \le c_1) = \int_0^{c_1} e^{-x}dx = 1 - e^{-c_1}$, so c_1 = -ln(.8) = .2231. Then $.2 = P(c_1 \le X \le c_2) \Rightarrow .4 = P(0 \le X_1 \le c_2) = 1 - e^{-c_2}$, so c_2 = -ln(.6) = .5108. Similarly, c_3 = -ln(.4) = .0163 and c_4 = -ln(.2) = 1.6094. the resulting intervals are [0, .2231), [.2231, .5108), [.5108, .9163), [.9163, 1.6094), and [1.6094, ∞).

b. Each expected cell count is 40(.2) = 8, and the observed cell counts are 6, 8, 10, 7, and 9, so $\chi^2 = \left[\frac{(6-8)^2}{8} + \ldots + \frac{(9-8)^2}{8}\right] = 1.25$. Because 1.25 is not $\ge \chi^2_{.10,4} = 7.779$, even at level .10 H_o cannot be rejected; the data is quite consistent with the specified exponential distribution.

10.

a. $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_{i0})^2}{np_{i0}} = \sum_i \frac{N_i^2 - 2np_{i0}N_i + n^2 p_{i0}^2}{np_{i0}} = \sum_i \frac{N_i^2}{np_{i0}} - 2\sum_i N_i + n\sum_i p_{i0}$
$= \sum_i \frac{N_i^2}{np_{i0}} - 2n + n(1) = \sum_i \frac{N_i^2}{np_{i0}} - n$ as desired. This formula involves only one subtraction, and that at the end of the calculation, so it is analogous to the shortcut formula for $s^{2.}$

b. $\chi^2 = \frac{k}{n}\sum_i N_i^2 - n$. For the pigeon data, k = 8, n = 120, and $\Sigma N_i^2 = 1872$, so $\chi^2 = \frac{8(1872)}{120} - 120 = 124.8 - 120 = 4.8$ as before.

11.

a. The six intervals must be symmetric about 0, so denote the 4th, 5th and 6th intervals by [0, a0, [a, b), [b, ∞). a must be such that $\Phi(a) = .6667\left(\frac{1}{2}+\frac{1}{6}\right)$, which from Table A.3 gives $a \approx .43$. Similarly $\Phi(b) = .8333$ implies $b \approx .97$, so the six intervals are ($-\infty$, -.97), [-.97, -.43), [-.43, 0), [0, .43), [.43, .97), and [.97, ∞).

b. The six intervals are symmetric about the mean of .5. From **a**, the fourth interval should extend from the mean to .43 standard deviations above the mean, i.e., from .5 to .5 + .43(.002), which gives [.5, .50086). Thus the third interval is [.5 - .00086, .5) = [.49914, .5). Similarly, the upper endpoint of the fifth interval is .5 + .97(.002) = .50194, and the lower endpoint of the second interval is .5 - .00194 = .49806. The resulting intervals are ($-\infty$, .49806), [.49806, .49914), [.49914, .5), [.5, .50086), [.50086, .50194), and [.50194, ∞).

c. Each expected count is $45\left(\frac{1}{6}\right) = 7.5$, and the observed counts are 13, 6, 6, 8, 7, and 5, so $c^2 = 5.53$. With 5 d.f., the p-value > .10, so we would fail to reject H_o at any of the usual levels of significance. There is no evidence to suggest that the bolt diameters are not normally distributed.

Section 14.2

12.

a. Let q denote the probability of a male (as opposed to female) birth under the binomial model. The four cell probabilities (corresponding to x = 0, 1, 2, 3) are $p_1(q) = (1-q)^3$, $p_2(q) = 3q(1-q)^2$, $p_3(q) = 3q^2(1-q)$, and $p_4(q) = q^3$.

The likelihood is $3^{n_2+n_3} \cdot (1-q)^{3n_1+2n_2+n_3} \cdot q^{n_2+2n_3+3n_4}$. Forming the log likelihood, taking the derivative with respect to q, equating to 0, and solving yields

$$\hat{q} = \frac{n_2+2n_3+3n_4}{3n} = \frac{66+128+48}{480} = .504$$

. The estimated expected counts are $160(1-.504)^3 = 19.52$, $480(.504)(.496)^2 = 59.52$, 60.48, and 20.48, so

$$c^2 = \left[\frac{(14-19.52)^2}{19.52} + \ldots + \frac{(16-20.48)^2}{20.48}\right] = 1.56+.71+.20+.98 = 3.45.$$

The number of degrees of freedom for the test is 4 – 1 – 1 = 2. H_o of a binomial distribution will be rejected using significance level .05 if $c^2 \geq c^2_{.05,2} = 5.992$. Because 3.45 < 5.992, H_o is not rejected, and the binomial model is judged to be quite plausible.

b. Now $\hat{q} = \frac{53}{150} = .353$ and the estimated expected counts are 13.54, 22.17, 12.09, and 2.20. The last estimated expected count is much less than 5, so the chi-squared test based on 2 d.f. should not be used.

13. According to the stated model, the three cell probabilities are $(1 – p)^2$, $2p(1 – p)$, and p^2, so we wish the value of p which maximizes $(1-p)^{2n_1}\left[2p(1-p)\right]^{n_2}p^{2n_3}$. Proceeding as in example 14.6 gives $\hat{p}=\dfrac{n_2+2n_3}{2n}=\dfrac{234}{2776}=.0843$. The estimated expected cell counts are then $n(1-\hat{p})^2=1163.85$, $n\left[2\hat{p}(1-\hat{p})\right]^2=214.29$, $n\hat{p}^2=9.86$. This gives

$$c^2=\left[\frac{(1212-1163.85)^2}{1163.85}+\frac{(118-214.29)^2}{214.29}+\frac{(58-9.86)^2}{9.86}\right]=280.3.$$

According to (14.15), H_o will be rejected if $c^2\ge c^2_{a,2}$, and since $c^2_{.01,2}=9.210$, H_o is soundly rejected; the stated model is strongly contradicted by the data.

14.

a. We wish to maximize $p^{\Sigma x_i-n}(1-p)^n$, or equivalently $(\Sigma x_i-n)\ln p+n\ln(1-p)$. Equating $\dfrac{d}{dp}$ to 0 yields $\dfrac{(\Sigma x_i-n)}{p}=\dfrac{n}{(1-p)}$, whence $p=\dfrac{(\Sigma x_i-n)}{\Sigma x_i}$. For the given data, $\Sigma x_i=(1)(1)+(2)(31)+\ldots+(12)(1)=363$, so

$$\hat{p}=\frac{(363-130)}{363}=.642, \text{ and } \hat{q}=.358.$$

b. Each estimated expected cell count is $\hat{p}$ times the previous count, giving $n\hat{q}=130(.358)=46.54$, $n\hat{q}\hat{p}=46.54(.642)=29.88$, 19.18, 12.31, 17.91, 5.08, 3.26, … . Grouping all values ≥ 7 into a single category gives 7 cells with estimated expected counts 46.54, 29.88, 19.18, 12.31, 7.91, 5.08 (sum = 120.9), and 130 – 120.9 = 9.1. The corresponding observed counts are 48, 31, 20, 9, 6, 5, and 11, giving $c^2=1.87$. With k = 7 and m = 1 (p was estimated), from (14.15) we need $c^2_{.10,5}=9.236$. Since 1.87 is not ≥ 9.236, we don't reject H_o.

15. The part of the likelihood involving q is $\left[(1-q)^4\right]^{n_1} \cdot \left[q(1-q)^3\right]^{n_2} \cdot \left[q^2(1-q)^2\right]^{n_3} \cdot$ $\left[q^3(1-q)\right]^{n_4} \cdot \left[q^4\right]^{n_5} = q^{n_2+2n_3+3n_4+4n_5}(1-q)^{4n_1+3n_2+2n_3+n_4} = q^{233}(1-q)^{367}$, so $\ln(likelihood) = 233\ln q + 367\ln(1-q)$. Differentiating and equating to 0 yields $\hat{q} = \frac{233}{600} = .3883,$ and $(1-\hat{q}) = .6117$ [note that the exponent on q is simply the total # of successes (defectives here) in the n = 4(150) = 600 trials.] Substituting this q' into the formula for p_i yields estimated cell probabilities .1400, .3555, .3385, .1433, and .0227. Multiplication by 150 yields the estimated expected cell counts are 21.00, 53.33, 50.78, 21.50, and 3.41. the last estimated expected cell count is less than 5, so we combine the last two categories into a single one (≥ 3 defectives), yielding estimated counts 21.00, 53.33, 50.78, 24.91, observed counts 26, 51, 47, 26, and $c^2 = 1.62$. With d.f. = 4 – 1 – 1 = 2, since $1.62 < c^2_{.10,2} = 4.605$, the p-value > .10, and we do not reject H$_o$. The data suggests that the stated binomial distribution is plausible.

16. $\hat{l} = \bar{x} = \frac{(0)(6)+(1)(24)+(2)(42)+...+(8)(6)+(9)(2)}{300} = \frac{1163}{300} = 3.88$, so the estimated cell probabilities are computed from $\hat{p} = e^{-3.88}\frac{(3.88)^x}{x!}$.

x	0	1	2	3	4	5	6	7	≥ 8
np(x)	6.2	24.0	46.6	60.3	58.5	45.4	29.4	16.3	13.3
obs	6	24	42	59	62	44	41	14	8

This gives $c^2 = 7.789$. To see whether the Poisson model provides a good fit, we need $c^2_{.10,9-1-1} = c^2_{.10,7} = 12.017$. Since $7.789 < 12.017$, the Poisson model does provide a good fit.

17. $\hat{l} = \frac{380}{120} = 3.167$, so $\hat{p} = e^{-3.167}\frac{(3.167)^x}{x!}$.

x	0	1	2	3	4	5	6	≥ 7
$\hat{p}$	.0421	.1334	.2113	.2230	.1766	.1119	.0590	.0427
$n\hat{p}$	5.05	16.00	25.36	26.76	21.19	13.43	7.08	5.12
obs	24	16	16	18	15	9	6	16

The resulting value of $c^2 = 103.98$, and when compared to $c^2_{.01,7} = 18.474$, it is obvious that the Poisson model fits very poorly.

18. $\hat{p}_1 = P(X < .100) = P\left(Z < \frac{.100 - .173}{.066}\right) = \Phi(-1.11) = .1335$,
$\hat{p}_2 = P(.100 \le X \le .150) = P(-1.11 \le Z \le -.35) = .2297$,
$\hat{p}_3 = P(-.35 \le Z \le .41) = .2959$, $\hat{p}_4 = P(.41 \le Z \le 1.17) = .2199$, and
$\hat{p}_5 = .1210$. The estimated expected counts are then (multiply $\hat{p}_i$ by n = 83) 11.08, 19.07, 24.56, 18.25, and 10.04, from which $c^2 = 1.67$. Comparing this with
$c^2_{.05,5-1-2} = c^2_{.05,2} = 5.992$, the hypothesis of normality cannot be rejected.

19. With $A = 2n_1 + n_4 + n_5$, $B = 2n_2 + n_4 + n_6$, and $C = 2n_3 + n_5 + n_6$, the likelihood is proportional to $q_1^A q_2^B (1 - q_1 - q_2)^C$, where A + B + C = 2n. Taking the natural log and equating both $\frac{\partial}{\partial q_1}$ and $\frac{\partial}{\partial q_2}$ to zero gives $\frac{A}{q_1} = \frac{C}{1 - q_1 - q_2}$ and $\frac{B}{q_2} = \frac{C}{1 - q_1 - q_2}$, whence
$q_2 = \frac{Bq_1}{A}$. Substituting this into the first equation gives $q_1 = \frac{A}{A + B + C}$, and then
$q_2 = \frac{B}{A + B + C}$. Thus $\hat{q}_1 = \frac{2n_1 + n_4 + n_5}{2n}$, $\hat{q}_2 = \frac{2n_2 + n_4 + n_6}{2n}$, and
$\left(1 - \hat{q}_1 - \hat{q}_2\right) = \frac{2n_3 + n_5 + n_6}{2n}$. Substituting the observed n_I's yields
$\hat{q}_1 = \frac{2(49) + 20 + 53}{400} = .4275$, $\hat{q}_2 = \frac{110}{400} = .2750$, and $\left(1 - \hat{q}_1 - \hat{q}_2\right) = .2975$, from which $\hat{p}_1 = (.4275)^2 = .183$, $\hat{p}_2 = .076$, $\hat{p}_3 = .089$, $\hat{p}_4 = 2(.4275)(.275) = .235$,
$\hat{p}_5 = .254$, $\hat{p}_6 = .164$.

Category	**1**	**2**	**3**	**4**	**5**	**6**
np	36.6	15.2	17.8	47.0	50.8	32.8
observed	49	26	14	20	53	38

This gives $c^2 = 29.1$. With $c^2_{.01,6-1-2} = c^2_{.01,3} = 11.344$, and
$c^2_{.01,6-1} = c^2_{.01,5} = 15.085$, according to (14.15) H_o must be rejected since
$29.1 \ge 15.085$.

20. The pattern of points in the plot appear to deviate from a straight line, a conclusion that is also supported by the small p-value (< .01000) of the Ryan-Joiner test. Therefore, it is implausible that this data came from a normal population. In particular, the observation 116.7 is a clear outlier. It would be dangerous to use the one-sample t interval as a basis for inference.

21. The Ryan-Joiner test p-value is larger than .10, so we conclude that the null hypothesis of normality cannot be rejected. This data could reasonably have come from a normal population. This means that it would be legitimate to use a one-sample t test to test hypotheses about the true average ratio.

22.

x_i	y_i	x_i	y_i	x_i	y_i
69.5	-1.967	75.5	-.301	79.6	.634
71.9	-1.520	75.7	-.199	79.7	.761
72.6	-1.259	75.8	-.099	79.9	.901
73.1	-1.063	76.1	.000	80.1	1.063
73.3	-.901	76.2	.099	82.2	1.259
73.5	-.761	76.9	.199	83.7	1.520
74.1	-.634	77.0	.301	93.7	1.967
74.2	-.517	77.9	.407		
75.3	-.407	78.1	.517		

n.b.: Minitab was used to calculate the y_I's. $\Sigma x_{(i)} = 1925.6$, $\Sigma x_{(i)}^2 = 148{,}871$, $\Sigma y_i = 0$, $\Sigma y_i^2 = 22.523$, $\Sigma x_{(i)} y_i = 103.03$, so

$$r = \frac{25(103.03)}{\sqrt{25(148{,}871) - (1925.6)^2}\sqrt{25(25.523)}} = .923$$. Since $c_{.01} = .9408$, and $.923 < .9408$, even at the very smallest significance level of .01, the null hypothesis of population normality must be rejected (the largest observation appears to be the primary culprit).

23. Minitab gives r = .967, though the hand calculated value may be slightly different because when there are ties among the $x_{(i)}$'s, Minitab uses the same y_I for each $x_{(i)}$ in a group of tied values. $C_{10} = .9707$, and $c_{.05} = 9639$, so $.05 < \text{p-value} < .10$. At the 5% significance level, one would have to consider population normality plausible.

Section 14.3

24. H_o: TV watching and physical fitness are independent of each other
H_a: the two variables are not independent
Df = (4 – 1)(2 – 1) = 3
With $\alpha = .05$, RR: $\chi^2 \geq 7.815$
Computed $\chi^2 = 6.161$
Fail to reject H_o. The data fail to indicate an association between daily TV viewing habits and physical fitness.

25. Let P_{ij} = the proportion of white clover in area of type i which has a type j mark (i = 1, 2; j = 1, 2, 3, 4, 5). The hypothesis H_o: $p_{1j} = p_{2j}$ for j = 1, …, 5 will be rejected at level .01 if $c^2 \geq c^2_{.01,(2-1)(5-1)} = c^2_{.01,4} = 13.277$.

$\hat{E}_{ij}$	1	2	3	4	5		
1	449.66	7.32	17.58	8.79	242.65	726	$c^2 = 23.18$
2	471.34	7.68	18.42	9.21	254.35	761	
	921	15	36	18	497	1487	

Since $23.18 \geq 13.277$, H_o is rejected.

26. Let p_{i1} = the probability that a fruit given treatment i matures and p_{i2} = the probability that a fruit given treatment i aborts. Then H_o: $p_{i1} = p_{i2}$ for i = 1, 2, 3, 4, 5 will be rejected if $c^2 \geq c^2_{.01,4} = 13.277$.

Observed		Estimated Expected		
Matured	Aborted	Matured	Aborted	n_i
141	206	110.7	236.3	347
28	69	30.9	66.1	97
25	73	31.3	66.7	98
24	78	32.5	69.5	102
20	82	32.5	69.5	102
		238	508	746

Thus $c^2 = \frac{(141-110.7)^2}{110.7} + \ldots + \frac{(82-69.5)^2}{69.5} = 24.82$, which is ≥ 13.277, so H_o is rejected at level .01.

27. With i = 1 identified with men and i = 2 identified with women, and j = 1, 2, 3 denoting the 3 categories L>R, L=R, L<R, we wish to test H_o: $p_{1j} = p_{2j}$ for j = 1, 2, 3 vs. H_a: p_{1j} not equal to p_{2j} for at least one j. The estimated cell counts for men are 17.95, 8.82, and 13.23 and for women are 39.05, 19.18, 28.77, resulting in $c^2 = 44.98$. With (2 – 1)(3 – 1) = 2 degrees of freedom, since $44.98 > c^2_{.005,2} = 10.597$, p-value < .005, which strongly suggests that H_o should be rejected.

28. With p_{ij} denoting the probability of a type j response when treatment i is applied, H_o: $p_{1j} = p_{2j} = p_{3j} = p_{4j}$ for j = 1, 2, 3, 4 will be rejected at level .005 if $\chi^2 \ge \chi^2_{.005,9} = 23.587$.

$\hat{E}_{ij}$	1	2	3	4
1	24.1	10.0	21.6	40.4
2	25.8	10.7	23.1	43.3
3	26.1	10.8	23.4	43.8
4	30.1	12.5	27.0	50.5

$\chi^2 = 27.66 \ge 23.587$, so reject H_o at level .005

29. H_o: $p_{1j} = \ldots = p_{6j}$ for j = 1, 2, 3 is the hypothesis of interest, where p_{ij} is the proportion of the j^{th} sex combination resulting from the i^{th} genotype. H_o will be rejected at level .10 if $\chi^2 \ge \chi^2_{.10,10} = 15.987$.

$\hat{E}_{ij}$	1	2	3	
1	35.8	83.1	35.1	154
2	39.5	91.8	38.7	170
3	35.1	81.5	34.4	151
4	9.8	22.7	9.6	42
5	5.1	11.9	5.0	22
6	26.7	62.1	26.2	115
	152	353	149	654

χ^2	1	2	3	
	.02	.12	.44	
	.06	.66	1.01	
	.13	.37	.34	
	.32	.49	.26	
	.00	.06	.19	
	.40	.14	1.47	
				6.46

(carrying 2 decimal places in $\hat{E}_{ij}$ yields $\chi^2 = 6.49$). Since 6.46 < 15.987, H_o cannot be rejected at level .10.

30. H_o: the design configurations are homogeneous with respect to type of failure vs. H_a: the design configurations are not homogeneous with respect to type of failure.

$\hat{E}_{ij}$	1	2	3	4	
1	16.11	43.58	18.00	12.32	90
2	7.16	19.37	8.00	5.47	40
3	10.74	29.05	12.00	8.21	60
	34	92	38	26	190

$\chi^2 = \dfrac{(20-16.11)^2}{16.11} + \ldots + \dfrac{(5-8.21)^2}{8.21} = 13.253$. With 6 df, $\chi^2_{.05,6} = 12.592 < 13.253 < \chi^2_{.025,6} = 14.440$, so .025 < p-value < .05. Since the p-value is < .05, we reject H_o. (If a smaller significance level were chosen, a different conclusion would be reached.) Configuration appears to have an effect on type of failure.

31. With I denoting the Ith type of car (I = 1, 2, 3, 4) and j the jth category of commuting distance, H_o: $p_{ij} = p_{i.}\,p_{.j}$ (type of car and commuting distance are independent) will be rejected at level .05 if $\chi^2 \ge \chi^2_{.05,6} = 12.592$.

$\hat{E}_{ij}$	1	2	3	
1	10.19	26.21	15.60	52
2	11.96	30.74	18.30	61
3	19.40	49.90	29.70	99
4	7.45	19.15	11.40	38
	49	126	75	250

$\chi^2 = 14.15 \ge 12.592$, so the independence hypothesis H_o is rejected at level .05 (but not at level .025!)

32.
$$\chi^2 = \frac{(479-494.4)^2}{494.4} + \frac{(173-151.5)^2}{151.5} + \frac{(119-125.2)^2}{125.2} + \frac{(214-177.0)^2}{177.0} + \frac{(47-54.2)^2}{54.2}$$
$$= \frac{(15-44.8)^2}{44.8} + \frac{(172-193.6)^2}{193.6} + \frac{(45-59.3)^2}{59.3} + \frac{(85-49.0)^2}{49.0} = 64.65 \ge \chi^2_{.01,4} = 13.277$$
so the independence hypothesis is rejected in favor of the conclusion that political views and level of marijuana usage are related.

33. $c^2 = \Sigma\Sigma \dfrac{\left(N_{ij} - \hat{E}_{ij}\right)^2}{\hat{E}_{ij}} = \Sigma\Sigma \dfrac{N_{ij}^2 - 2\hat{E}_{ij}N_{ij} + \hat{E}_{ij}^2}{\hat{E}_{ij}} = \dfrac{\Sigma\Sigma N_{ij}^2}{\hat{E}_{ij}} - 2\Sigma\Sigma N_{ij} + \Sigma\Sigma\hat{E}_{ij}$, but

$\Sigma\Sigma\hat{E}_{ij} = \Sigma\Sigma N_{ij} = n$, so $c^2 = \Sigma\Sigma \dfrac{N_{ij}^2}{\hat{E}_{ij}} - n$. This formula is computationally efficient because there is only one subtraction to be performed, which can be done as the last step in the calculation.

34. This is a $3 \times 3 \times 3$ situation, so there are 27 cells. Only the total sample size n is fixed in advance of the experiment, so there are 26 freely determined cell counts. We must estimate $p_{..1}$, $p_{..2}$, $p_{..3}$, $p_{.1.}$, $p_{.2.}$, $p_{.3.}$, $p_{1..}$, $p_{2..}$, and $p_{3..}$, but $\Sigma p_{i..} = \Sigma p_{.j.} = \Sigma p_{..k} = 1$ so only 6 independent parameters are estimated. The rule for d.f. now gives c^2 df = 26 – 6 = 20.

35. With p_{ij} denoting the common value of p_{ij1}, p_{ij2}, p_{ij3}, p_{ij4} (under H_o), $\hat{p}_{ij} = \dfrac{N_{ij}}{n}$ and

$\hat{E}_{ijk} = \dfrac{n_k N_{ij}}{n}$. With four different tables (one for each region), there are 8 + 8 + 8 + 8 = 32 freely determined cell counts. Under H_o, $p_{11}, \ldots, p_{33}$ must be estimated but $\Sigma\Sigma p_{ij} = 1$ so only 8 independent parameters are estimated, giving c^2 df = 32 – 8 = 24.

36.

a.

Observed

13	19	28	60
7	11	22	40
20	30	50	100

Estimated Expected

12	18	30
8	12	20

$c^2 = \dfrac{(13-12)^2}{12} + \ldots + \dfrac{(22-20)^2}{20} = .6806$. Because $.6806 < c^2_{.10,2} = 4.605$, H_o is not rejected.

b. Each observation count here is 10 times what it was in **a**, and the same is true of the estimated expected counts so now $c^2 = 6.806 \geq 4.605$, and H_o is rejected. With the much larger sample size, the departure from what is expected under H_o, the independence hypothesis, is statistically significant – it cannot be explained just by random variation.

c. The observed counts are .13n, .19n, .28n, .07n, .11n, .22n, whereas the estimated expected $\dfrac{(.60n)(.20n)}{n} =$ .12n, .18n, .30n, .08n, .12n, .20n, yielding $c^2 = .006806n$. H_o will be rejected at level .10 iff $.006806n \geq 4.605$, i.e., iff $n \geq 676.6$, so the minimum n = 677.

Supplementary Exercises

37. There are 3 categories here – firstborn, middleborn, (2nd or 3rd born), and lastborn. With p_1, p_2, and p_3 denoting the category probabilities, we wish to test H_o: $p_1 = .25$, $p_2 = .50$ ($p_2 = P$(2nd or 3rd born) $= .25 + .25 = .50$), $p_3 = .25$. H_o will be rejected at significance level .05 if $c^2 \geq c^2_{.05,2} = 5.992$. The expected counts are (31)(.25) = 7.75, (31)(.50) = 15.5, and 7.75, so $c^2 = \frac{(12-7.75)^2}{7.75} + \frac{(11-15.5)^2}{15.5} + \frac{(8-7.75)^2}{7.75} = 3.65$. Because 3.65 < 5.992, H_o is not rejected. The hypothesis of equiprobable birth order appears quite plausible.

38. Let p_{i1} = the proportion of fish receiving treatment i (i = 1, 2, 3) who are parasitized. We wish to test H_o: $p_{1j} = p_{2j} = p_{3j}$ for j = 1, 2. With df = (2 – 1)(3 – 1) = 2, H_o will be rejected at level .01 if $c^2 \geq c^2_{.01,2} = 9.210$.

Observed		
30	3	33
16	8	24
16	16	32
62	27	89

Estimated Expected	
22.99	10.01
16.72	7.28
22.29	9.71

This gives $c^2 = 13.1$. Because $13.1 \geq 9.210$, H_o should be rejected. The proportion of fish that are parasitized does appear to depend on which treatment is used.

39. H_o: gender and years of experience are independent; H_a: gender and years of experience are not independent. Df = 4, and we reject H_o if $c^2 \geq c^2_{.01,4} = 13.277$.

	Years of Experience				
Gender	1 – 3	4 – 6	7 – 9	10 – 12	13 +
Male Observed	202	369	482	361	811
Expected	285.56	409.83	475.94	347.04	706.63
$\frac{(O-E)^2}{E}$	24.451	4.068	.077	.562	15.415
Female Observed	230	251	238	164	258
Expected	146.44	210.17	244.06	177.96	362.37
$\frac{(O-E)^2}{E}$	47.680	7.932	.151	1.095	30.061

$c^2 = \Sigma \frac{(O-E)^2}{E} = 131.492$. Reject H_o. The two variables do not appear to be independent. In particular, women have higher than expected counts in the beginning category (1 – 3 years) and lower than expected counts in the more experienced category (13+ years).

40.

a. H_o: The probability of a late-game leader winning is independent of the sport played; H_a: The two variables are not independent. With 3 df, the computed $\chi^2 = 10.518$, and the p-value < .015 is also < .05, so we would reject H_o. There appears to be a relationship between the late-game leader winning and the sport played.

b. Quite possibly: Baseball had many fewer than expected late-game leader losses.

41. The null hypothesis H_o: $p_{ij} = p_{i.}\,p_{.j}$ states that level of parental use and level of student use are independent in the population of interest. The test is based on (3 – 1)(3 – 1) = 4 df.

Estimated Expected

119.3	57.6	58.1	235
82.8	33.9	40.3	163
23.9	11.5	11.6	47
226	109	110	445

The calculated value of $\chi^2 = 22.4$. Since $22.4 > \chi^2_{.005,4} = 14.860$, p-value < .005, so H_o should be rejected at any significance level greater than .005. Parental and student use level do not appear to be independent.

42. The estimated expected counts are displayed below, from which $\chi^2 = 197.70$. A glance at the 6 df row of Table A.7 shows that this test statistic value is highly significant – the hypothesis of independence is clearly implausible.

Estimated Expected

	Home	Acute	Chronic	
15 – 54	90.2	372.5	72.3	535
55 – 64	113.6	469.3	91.1	674
65 – 74	142.7	589.0	114.3	846
> 74	157.5	650.3	126.2	934
	504	2081	404	2989

43. This is a test of homogeneity: H_o: $p_{1j} = p_{2j} = p_{3j}$ for j = 1, 2, 3, 4, 5. The given SPSS output reports the calculated $c^2 = 70.64156$ and accompanying p-value (significance) of .0000. We reject H_o at any significance level. The data strongly supports that there are differences in perception of odors among the three areas.

44. The accompanying table contains both observed and estimated expected counts, the latter in parentheses.

	Age					
Want	127 (131.1)	118 (123.3)	77 (71.7)	61 (55.1)	41 (42.8)	424
Don't	23 (18.9)	23 (17.7)	5 (10.3)	2 (7.9)	8 (6.2)	61
	150	141	82	63	49	485

This gives $c^2 = 11.60 \geq c^2_{.05,4} = 9.488$. At level .05, the null hypothesis of independence is rejected, though it would not be rejected at the level .01 (.01 < p-value < .025).

45. $(n_1 - np_{10})^2 = (np_{10} - n_1)^2 = (n - n_1 - n(1 - p_{10}))^2 = (n_2 - np_{20})^2$. Therefore

$$c^2 = \frac{(n_1 - np_{10})^2}{np_{10}} + \frac{(n_2 - np_{20})^2}{np_{20}} = \frac{(n_1 - np_{10})^2}{n_2}\left(\frac{n}{p_{10}} + \frac{n}{p_{20}}\right)$$

$$= \left(\frac{n_1}{n} - p_{10}\right)^2 \cdot \left(\frac{n}{p_{10}p_{20}}\right) = \frac{(\hat{p}_1 - p_{10})^2}{p_{10}p_{20}/n} = z^2.$$

46.

a.

obsv	22	10	5	11
exp	13.189	10	7.406	17.405

H_0: probabilities are as specified.
H_a: probabilities are not as specified.

Test Statistic: $c^2 = \frac{(22 - 13.189)^2}{13.189} + \frac{(10 - 10)^2}{10} + \frac{(5 - 7.406)^2}{7.406} + \frac{(11 - 17.405)^2}{17.405}$

$= 5.886 + 0 + 0.782 + 2.357 = 9.025$. Rejection Region: $c^2 > c^2_{.05,2} = 5.99$

Since 9.025 > 5.99, we reject H_0. The model postulated in the exercise is not a good fit.

b.

p_i	0.45883	0.18813	0.11032	0.24272
exp	22.024	9.03	5.295	11.651

$$c^2 = \frac{(22-22.024)^2}{22.024} + \frac{(10-9.03)^2}{9.03} + \frac{(5-5.295)^2}{5.295} + \frac{(11-11.651)^2}{11.651}$$
$$= .0000262 + .1041971 + .0164353 + .0363746 = .1570332$$

With the same rejection region as in a, we do not reject the null hypothesis. This model does provide a good fit.

47.

a. Our hypotheses are H_0: no difference in proportion of concussions among the three groups. Vs H_a: there is a difference …

Observed	Concussion	No Concussion	Total
Soccer	45	46	91
Non Soccer	28	68	96
Control	8	45	53
Total	81	159	240

Expected	Concussion	No Concussion	Total
Soccer	30.7125	60.2875	91
Non Soccer	32.4	63.6	96
Control	17.8875	37.1125	53
Total	81	159	240

$$c^2 = \frac{(45-30.7125)^2}{30.7125} + \frac{(46-60.2875)^2}{60.2875} + \frac{(28-32.4)^2}{32.4} + \frac{(68-63.6)^2}{63.6}$$

$+ \dfrac{(8-17.8875)^2}{17.8875} + \dfrac{(45-37.1125)^2}{37.1125} = 19.1842$. The df for this test is (I – 1)(J – 1) = 2, so we reject H_o if $c^2 > c^2_{.05,2} = 5.99$. 19.1842 > 5.99, so we reject H_0. There is a difference in the proportion of concussions based on whether a person plays soccer.

b. We are testing the hypothesis H_0: ρ = 0 vs H_a: ρ ? 0. The test statistic is $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{-.22\sqrt{89}}{\sqrt{1-.22^2}} = -2.13$. At significance level α = .01, we would fail to reject and conclude that there is no evidence of non-zero correlation in the population. If we were willing to accept a higher significance level, our decision could change. At best, there is evidence of only weak correlation.

c. We will test to see if the average score on a controlled word association test is the same for soccer and non-soccer athletes. H_0: $\mu_1 = \mu_2$ vs H_a: μ_1 ? μ_2 . We'll use test statistic $t = \dfrac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\dfrac{s_1^2}{m} + \dfrac{s_2^2}{n}}}$. With $\dfrac{s_1^2}{m} = 3.206$ and $\dfrac{s_2^2}{n} = 1.854$,

$t = \dfrac{(37.50 - 39.63)}{\sqrt{3.206 + 1.854}} = -.95$. The df = $\dfrac{(3.206 + 1.854)^2}{\dfrac{3.206^2}{25} + \dfrac{1.854^2}{55}} \approx 56$. The p-value will be > .10, so we do not reject H_0 and conclude that there is no difference in the average score on the test for the two groups of athletes.

d. Our hypotheses for ANOVA are H_0: all means are equal vs H_a: not all means are equal. The test statistic is $f = \dfrac{MSTr}{MSE}$.

$$SSTr = 91(.30 - .35)^2 + 96(.49 - .35)^2 + 53(.19 - .35)^2 = 3.4659$$

$$MSTr = \frac{3.4659}{2} = 1.73295$$

$SSE = 90(.67)^2 + 95(.87)^2 + 52(.48)^2 = 124.2873$ and

$MSE = \dfrac{124.2873}{237} = .5244$. Now, $f = \dfrac{1.73295}{.5244} = 3.30$. Using df 2,200 from table A.9, the p value is between .01 and .05. At significance level .05, we reject the null hypothesis. There is sufficient evidence to conclude that there is a difference in the average number of prior non-soccer concussions between the three groups.

48.

a. H_o: $p_0 = p_1 = \ldots = p_9 = .10$ vs H_a: at least one p_i ? .10, with df = 9.

b. H_o: $p_{ij} = .01$ for I and j= 1,2,…,9 vs H_a: at least one p_{ij} ? 0, with df = 99.

c. For this test, the number of p's in the Hypothesis would be 10^5 = 100,000 (the number of possible combinations of 5 digits). Using only the first 100,000 digits in the expansion, the number of non-overlapping groups of 5 is only 20,000. We need a much larger sample size!

d. Based on these p-values, we could conclude that the digits of p behave as though they were randomly generated.

CHAPTER 15

Section 15.1

1. We test $H_0 : m = 100$ vs. $H_a : m \neq 100$. The test statistic is s_+ = sum of the ranks associated with the positive values of $(x_i - 100)$, and we reject H_o at significance level .05 if $s_+ \geq 64$. (from Table A.13, n = 12, with $a/2 = .026$, which is close to the desired value of . 025), or if $s_+ \leq \frac{12(13)}{2} - 64 = 78 - 64 = 14$.

x_i	$(x_i - 100)$	ranks
105.6	5.6	7*
90.9	-9.1	12
91.2	-8.8	11
96.9	-3.1	3
96.5	-3.5	5
91.3	-8.7	10
100.1	0.1	1*
105	5	6*
99.6	-0.4	2
107.7	7.7	9*
103.3	3.3	4*
92.4	-7.6	8

S_+ = 27, and since 27 is neither ≥ 64 nor ≤ 14, we do not reject H_o. There is not enough evidence to suggest that the mean is something other than 100.

2. We test $H_0 : m = 25$ vs. $H_a : m > 25$. With n = 5 and $a \approx .03$, reject H_o if $s_+ \geq 15$. From the table below we arrive at s_+ =1+5+2+3 = 11, which is not ≥ 15, so do not reject H_o. It is still plausible that the mean = 25.

x_i	$(x_i - 25)$	ranks
25.8	0.8	1*
36.6	11.6	5*
26.3	1.3	2*
21.8	-3.2	4
27.2	2.2	3*

3. We test $H_0: m = 7.39$ vs. $H_a: m \neq 7.39$, so a two tailed test is appropriate. With n = 14 and $\alpha/2 = .025$, Table A.13 indicates that H_o should be rejected if either $s_+ \geq 84 or \leq 21$. The $(x_i - 7.39)$'s are -.37, -.04, -.05, -.22, -.11, .38, -.30, -.17, .06, -.44, .01, -.29, -.07, and -.25, from which the ranks of the three positive differences are 1, 4, and 13. Since $s_+ = 18 \leq 21$, H_o is rejected at level .05.

4. The appropriate test is $H_0: m = 30$ vs. $H_a: m < 30$. With n = 15, and $\alpha = .10$, reject H_o if $s_+ \leq \frac{15(16)}{2} - 83 = 37$.

x_i	$(x_i - 30)$	ranks	x_i	$(x_i - 30)$	ranks
30.6	0.6	3*	31.9	1.9	5*
30.1	0.1	1*	53.2	23.2	15*
15.6	-14.4	12	12.5	-17.5	13
26.7	-3.3	7	23.2	-6.8	11
27.1	-2.9	6	8.8	-21.2	14
25.4	-4.6	8	24.9	-5.1	10
35	5	9*	30.2	0.2	2*
30.8	0.8	4*			

$S_+ = 39$, which is not ≤ 37, so H_o cannot be rejected. There is not enough evidence to prove that diagnostic time is less than 30 minutes at the 10% significance level.

5. The data is paired, and we wish to test $H_0: m_D = 0$ vs. $H_a: m_D \neq 0$. With n = 12 and $\alpha = .05$, H_o should be rejected if either $s_+ \geq 64$ or if $s_+ \leq 14$.

d_i	-.3	2.8	3.9	.6	1.2	-1.1	2.9	1.8	.5	2.3	.9	2.5
rank	1	10*	12*	3*	6*	5	11*	7*	2*	8*	4*	9*

$s_+ = 72$, and $72 \geq 64$, so H_o is rejected at level .05. In fact for $\alpha = .01$, the critical value is c = 71, so even at level .01 H_o would be rejected.

6. We wish to test $H_0 : m_D = 5$ vs. $H_a : m_D > 5$, where $m_D = m_{black} - m_{white}$. With n = 9 and $a \approx .05$, H_o will be rejected if $s_+ \geq 37$. As given in the table below, $s_+ = 37$, which is ≥ 37, so we can (barely) reject H_o at level approximately .05, and we conclude that the greater illumination does decrease task completion time by more than 5 seconds.

d_i	$d_i - 5$	rank	d_i	$d_i - 5$	rank
7.62	2.62	3*	16.07	11.07	9*
8	3	4*	8.4	3.4	5*
9.09	4.09	8*	8.89	3.89	7*
6.06	1.06	1*	2.88	-2.12	2
1.39	-3.61	6			

7. $H_0 : m_D = .20$ vs. $H_a : m_D > .20$, where $m_D = m_{outdoor} - m_{indoor}$. $a = .05$, and because n = 33, we can use the large sample test. The test statistic is $Z = \dfrac{s_+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$, and we reject H_o if $z \geq 1.96$.

d_i	$d_i - .2$	rank	d_i	$d_i - .2$	rank	d_i	$d_i - .2$	rank
0.22	0.02	2	0.15	-0.05	5.5	0.63	0.43	23
0.01	-0.19	17	1.37	1.17	32	0.23	0.03	4
0.38	0.18	16	0.48	0.28	21	0.96	0.76	31
0.42	0.22	19	0.11	-0.09	8	0.2	0	1
0.85	0.65	29	0.03	-0.17	15	-0.02	-0.22	18
0.23	0.03	3	0.83	0.63	28	0.03	-0.17	14
0.36	0.16	13	1.39	1.19	33	0.87	0.67	30
0.7	0.5	26	0.68	0.48	25	0.3	0.1	9.5
0.71	0.51	27	0.3	0.1	9.5	0.31	0.11	11
0.13	-0.07	7	-0.11	-0.31	22	0.45	0.25	20
0.15	-0.05	5.5	0.31	0.11	12	-0.26	-0.46	24

$s_+ = 434$, so $z = \dfrac{424 - 280.5}{\sqrt{3132.25}} = \dfrac{143.5}{55.9665} = 2.56$. Since $2.56 \geq 1.96$, we reject H_o at significance level .05.

8. We wish to test $H_0 : m = 75$ vs. $H_a : m > 75$. Since n = 25 the large sample approximation is used, so H_o will be rejected at level .05 if $z \geq 1.645$. The $(x_i - 75)'s$ are –5.5, -3.1, -2.4, -1.9, -1.7, 1.5, -.9, -.8, .3, .5, .7, .8, 1.1, 1.2, 1.2, 1.9, 2.0, 2.9, 3.1, 4.6, 4.7, 5.1, 7.2, 8.7, and 18.7. The ranks of the positive differences are 1, 2, 3, 4.5, 7, 8.5, 8.5, 12.5, 14, 16, 17.5, 19, 20, 21, 23, 24, and 25, so s_+ = 226.5 and $\frac{n(n+1)}{4} = 162.5$. Expression (15.2) for s^2 should be used (because of the ties): $t_1 = t_2 = t_3 = t_4 = 2$, so

$$s^2_{s_+} = \frac{25(26)(51)}{24} - \frac{4(1)(2)(3)}{48} = 1381.25 - .50 = 1380.75 \text{ and } s = 37.16. \text{ Thus}$$

$$z = \frac{226.5 - 162.5}{37.16} = 1.72.$$ Since $1.72 \geq 1.645$, H_o is rejected.

$p - value \approx 1 - \Phi(1.72) = .0427$. The data indicates that true average toughness of the steel does exceed 75.

9.

r_1	1	1	1	1	1	1	2	2	2	2	2	2
r_2	2	2	3	3	4	4	1	1	3	3	4	4
r_3	3	4	2	4	2	3	3	4	1	4	1	3
r_4	4	3	4	2	3	2	4	3	4	1	3	1
D	0	2	2	6	6	8	2	4	6	12	10	14

r_1	3	3	3	3	3	3	4	4	4	4	4	4
r_2	1	1	2	2	4	4	1	1	2	2	3	3
r_3	2	4	1	4	1	2	2	3	1	3	1	2
r_4	4	2	4	1	2	1	3	2	3	1	2	1
D	6	10	8	14	16	18	12	14	14	18	18	20

When H_o is true, each of the above 24 rank sequences is equally likely, which yields the distribution of D when H_o is true as described in the answer section (e.g., P(D = 2) = P(1243 or 1324 or 2134) = 3/24). Then c = 0 yields $a = \frac{1}{24} = .042$ while c = 2 implies $a = \frac{4}{24} = .167$.

Section 15.2

10. The ordered combined sample is 163(y), 179(y), 213(y), 225(y), 229(x), 245(x), 247(y), 250(x), 286(x), and 299(x), so w = 5 + 6 + 8 + 9 + 10 = 38. With m = n = 5, Table A.14 gives the upper tail critical value for a level .05 test as 36 (reject H_o if W ≥ 36). Since $38 \geq 36$, H_o is rejected in favor of H_a.

11. With X identified with pine (corresponding to the smaller sample size) and Y with oak, we wish to test $H_0: m_1 - m_2 = 0$ vs. $H_a: m_1 - m_2 \neq 0$. From Table A.14 with m = 6 and n = 8, H_o is rejected in favor of H_a at level .05 if either $w \geq 61$ or if $w \leq 90 - 61 = 29$ (the actual α is 2(.021) = .042). The X ranks are 3 (for .73), 4 (for .98), 5 (for 1.20), 7 (for 1.33), 8 (for 1.40), and 10 (for 1.52), so w = 37. Since 37 is neither ≥ 61 nor ≤ 29, H_o cannot be rejected.

12. The hypotheses of interest are $H_0: m_1 - m_2 = 1$ vs. $H_a: m_1 - m_2 > 1$, where 1(X) refers to the original process and 2 (Y) to the new process. Thus 1 must be subtracted from each x_I before pooling and ranking. At level .05, H_o should be rejected in favor of H_a if $w \geq 84$.

x – 1	3.5	4.1	4.4	4.7	5.3	5.6	7.5	7.6
rank	1	4	5	6	8	10	15	16
y	3.8	4.0	4.9	5.5	5.7	5.8	6.0	7.0
rank	2	3	7	9	11	12	13	14

Since w = 65, H_o is not rejected.

13. Here m = n = 10 > 8, so we use the large-sample test statistic from p. 663. $H_0: m_1 - m_2 = 0$ will be rejected at level .01 in favor of $H_a: m_1 - m_2 \neq 0$ if either $z \geq 2.58$ or $z \leq -2.58$. Identifying X with orange juice, the X ranks are 7, 8, 9, 10, 11, 16, 17, 18, 19, and 20, so w = 135. With $\frac{m(m+n+1)}{2} = 105$ and $\sqrt{\frac{mn(m+n+1)}{12}} = \sqrt{175} = 13.22$, $z = \frac{135-105}{13.22} = 2.27$. Because 2.27 is neither ≥ 2.58 nor ≤ -2.58, H_o is not rejected. $p-value \approx 2(1-\Phi(2.27)) = .0232.$

14.

x	8.2	9.5	9.5	9.7	10.0	14.5	15.2	16.1	17.6	21.5
rank	7	9	9	11	12.5	16	17	18	19	20
y	4.2	5.2	5.8	6.4	7.0	7.3	9.5	10.0	11.5	11.5
rank	1	2	3	4	5	6	9	12.5	14.5	14.5

The denominator of z must now be computed according to (15.6). With $t_1 = 3$, $t_2 = 2$, $t_3 = 2$, $s^2 = 175 - .0219[2(3)(4) + 1(2)(3) + 1(2)(3)] = 174.21$, so $z = \frac{138.5-105}{\sqrt{174.21}} = 2.54$. Because 2.54 is neither ≥ 2.58 nor ≤ -2.58, H_o is not rejected.

15. Let m_1 and m_2 denote true average cotanine levels in unexposed and exposed infants, respectively. The hypotheses of interest are $H_0 : m_1 - m_2 = -25$ vs. $H_a : m_1 - m_2 < -25$. With m = 7, n = 8, H_o will be rejected at level .05 if $w \le 7(7+8+1) - 71 = 41$. Before ranking, -25 is subtracted from each x_I (i.e. 25 is added to each), giving 33, 36, 37, 39, 45, 68, and 136. The corresponding ranks in the combined set of 15 observations are 1, 3, 4, 5, 6, 8, and 12, from which w = 1 + 3 + … + 12 = 39. Because $39 \le 41$, H_o is rejected. The true average level for exposed infants appears to exceed that for unexposed infants by more than 25 (note that H_o would not be rejected using level .01).

16.

a.

X	rank	Y	rank
0.43	2	1.47	9
1.17	8	0.8	7
0.37	1	1.58	11
0.47	3	1.53	10
0.68	6	4.33	16
0.58	5	4.23	15
0.5	4	3.25	14
2.75	12	3.22	13

We verify that w = sum of the ranks of the x's = 41.

b. We are testing $H_0 : m_1 - m_2 = 0$ vs. $H_a : m_1 - m_2 < 0$. The reported p-value (significance) is .0027, which is < .01 so we reject H_o. There is evidence that the distribution of good visibility response time is to the left (or lower than) that response time with poor visibility.

Section 15.3

17. n = 8, so from Table A.15, a 95% C.I. (actually 94.5%) has the form $\left(\overline{x}_{(36-32+1)}, \overline{x}_{(32)}\right) = \left(\overline{x}_{(5)}, \overline{x}_{(32)}\right)$. It is easily verified that the 5 smallest pairwise averages are $\frac{5.0+5.0}{2} = 5.00$, $\frac{5.0+11.8}{2} = 8.40$, $\frac{5.0+12.2}{2} = 8.60$, $\frac{5.0+17.0}{2} = 11.00$, and $\frac{5.0+17.3}{2} = 11.15$ (the smallest average not involving 5.0 is $\overline{x}_{(6)} = \frac{11.8+11.8}{2} = 11.8$), and the 5 largest averages are 30.6, 26.0, 24.7, 23.95, and 23.80, so the confidence interval is (11.15, 23.80).

18. With n = 14 and $\frac{n(n+1)}{2} = 105$, from Table A.15 we se that c = 93 and the 99% interval is $\left(\bar{x}_{(13)}, \bar{x}_{(93)}\right)$. Subtracting 7 from each x_I and multiplying by 100 (to simplify the arithmetic) yields the ordered values –5, 2, 9, 10, 14, 17, 22, 28, 32, 34, 35, 40, 45, and 77. The 13 smallest *sums* are –10, -3, 4, 4, 5, 9, 11, 12, 12, 16, 17, 18, and 19 (so $\bar{x}_{(13)} = \frac{14.19}{2} = 7.095)$ while the 13 largest sums are 154, 122, 117, 112, 111, 109, 99, 91, 87, and 86 (so $\bar{x}_{(93)} = \frac{14.86}{2} = 7.430)$. The desired C.I. is thus (7.095, 7.430).

19. The ordered d_i's are –13, -12, -11, -7, -6; with n = 5 and $\frac{n(n+1)}{2} = 15$, Table A.15 shows the 94% C.I. as (since c = 1) $\left(\bar{d}_{(1)}, \bar{d}_{(15)}\right)$. The smallest average is clearly $\frac{-13-13}{2} = -13$ while the largest is $\frac{-6-6}{2} = -6$, so the C.I. is (-13, -6).

20. For n = 4 Table A.13 shows that a two tailed test can be carried out at level .124 or at level .250 (or, of course even higher levels), so we can obtain either an 87.6% C.I. or a 75% C.I. With $\frac{n(n+1)}{2} = 10$, the 87.6% interval is $\left(\bar{x}_{(1)}, \bar{x}_{(10)}\right) = (.045, .177)$.

21. m = n = 5 and from Table A.16, c = 21 and the 90% (actually 90.5%) interval is $\left(d_{ij(5)}, d_{ij(21)}\right)$. The five smallest $x_i - y_j$ differences are –18, -2, 3, 4, 16 while the five largest differences are 136, 123, 120, 107, 86 (construct a table like Table 15.5), so the desired interval is $(16, 86)$.

22. m = 6, n = 8, mn = 48, and from Table A.16 a 99% interval (actually 99.2%) requires c = 44 and the interval is $\left(d_{ij(5)}, d_{ij(44)}\right)$. The five largest $x_i - y_j$'s are 1.52 - .48 = 1.04, 1.40 - .48 = .92, 1.52 - .67 = .85, 1.33 - .48 = .85, and 1.40 - .67 = .73, while the five smallest are –1.04, -.99, -.83, -.82, and -.79, so the confidence interval for $m_1 - m_2$ (where m_1 refers to pine and m_2 refers to oak) is (-.79, .73).

Section 15.4

23. Below we record in parentheses beside each observation the rank of that observation in the combined sample.

1:	5.8(3)	6.1(5)	6.4(6)	6.5(7)	7.7(10)	$r_{1.} = 31$
2:	7.1(9)	8.8(12)	9.9(14)	10.5(16)	11.2(17)	$r_{2.} = 68$
3:	5.191)	5.7(2)	5.9(4)	6.6(8)	8.2(11)	$r_{3.} = 26$
4:	9.5(13)	1.0.3(15)	11.7(18)	12.1(19)	12.4(20)	$r_{4.} = 85$

H_o will be rejected at level .10 if $k \ge \chi^2_{.10,3} = 6.251$. The computed value of k is $k = \frac{12}{20(21)}\left[\frac{31^2 + 68^2 + 26^2 + 85^2}{5}\right] - 3(21) = 14.06$. Since $14.06 \ge 6.251$, reject H_o.

24. After ordering the 9 observation within each sample, the ranks in the combined sample are

1:	1	2	3	7	8	16	18	22	27	$r_{1.} = 104$
2:	4	5	6	11	12	21	31	34	36	$r_{2.} = 160$
3:	9	10	13	14	15	19	28	33	35	$r_{3.} = 176$
4:	17	20	23	24	25	26	29	30	32	$r_{4.} = 226$

At level .05, $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ will be rejected if $k \ge \chi^2_{.05,3} = 7.815$. The computed k is $k = \frac{12}{36(37)}\left[\frac{104^2 + 160^2 + 176^2 + 226^2}{5}\right] - 3(37) = 7.587$. Since 7.587 is not ≥ 7.815, H_o cannot be rejected.

25. $H_0 : \mu_1 = \mu_2 = \mu_3$ will be rejected at level .05 if $k \ge \chi^2_{.05,2} = 5.992$. The ranks are 1, 3, 4, 5, 6, 7, 8, 9, 12, 14 for the first sample; 11, 13, 15, 16, 17, 18 for the second; 2, 10, 19, 20, 21, 22 for the third; so the rank totals are 69, 90, and 94.

$k = \frac{12}{22(23)}\left[\frac{69^2}{10} + \frac{90^2}{6} + \frac{94^2}{5}\right] - 3(23) = 9.23$. Since $9.23 \ge 5.992$, we reject H_o.

26.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	r_i	r_i^2
A	2	2	2	2	2	2	2	2	2	2	20	400
B	1	1	1	1	1	1	1	1	1	1	10	100
C	4	4	4	4	3	4	4	4	4	4	39	1521
D	3	3	3	3	4	3	3	3	3	3	31	961
												2982

The computed value of F_r is $\frac{12}{4(10)(5)}(2982)-3(10)(5)=28.92$, which is $\geq \chi^2_{.01,3}=11.344$, so H_o is rejected.

27.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	r_i	r_i^2
I	1	2	3	3	2	1	1	3	1	2	19	361
H	2	1	1	2	1	2	2	1	2	3	17	289
C	3	3	2	1	3	3	3	2	3	1	24	576
												1226

The computed value of F_r is $\frac{12}{10(3)(4)}(1226)-3(10)(4)=2.60$, which is not $\geq \chi^2_{.05,2}=5.992$, so don't reject H_o.

Supplementary Exercises

28. The Wilcoxon signed-rank test will be used to test $H_0: \mu_D = 0$ vs. $H_0: \mu_D \neq 0$, where $\mu_D =$ the difference between expected rate for a potato diet and a rice diet. From Table A.11 with n = 8, H_o will be rejected if either $s_+ \geq 32$ or $s_+ \leq \frac{8(9)}{2}-32=4$. The $d_i's$ are (in order of magnitude) .16, .18, .25, -.56, .60, .96, 1.01, and –1.24, so $s_+ = 1+2+3+5+6+7=24$. Because 24 is not in the rejection region, H_o is not rejected.

29. Friedman's test is appropriate here. At level .05, H_o will be rejected if $f_r \geq \chi^2_{.05,3} = 7.815$. It is easily verified that $r_{1.} = 28$, $r_{2.} = 29$, $r_{3.} = 16$, $r_{4.} = 17$, from which the defining formula gives $f_r = 9.62$ and the computing formula gives $f_r = 9.67$. Because $f_r \geq 7.815$, $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ is rejected, and we conclude that there are effects due to different years.

30. The Kruskal-Wallis test is appropriate for testing $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$. H_o will be rejected at significance level .01 if $k \geq \chi^2_{.01,3} = 11.344$

Treatment			**ranks**			r_i
I	4	1	2	3	5	15
II	8	7	10	6	9	40
III	11	15	14	12	13	65
IV	16	20	19	17	18	90

$k = \frac{12}{420}\left[\frac{225 + 1600 + 4225 + 8100}{5}\right] - 63 = 17.86$. Because $17.86 \geq 11.344$, reject H_o.

31. From Table A.16, m = n = 5 implies that c = 22 for a confidence level of 95%, so $mn - c + 1 = 25 - 22 = 1 = 4$. Thus the confidence interval extends from the 4th smallest difference to the 4th largest difference. The 4 smallest differences are –7.1, -6.5, -6.1, -5.9, and the 4 largest are –3.8, -3.7, -3.4, -3.2, so the C.I. is (-5.9, -3.8).

32.

a. $H_0: m_1 - m_2 = 0$ will be rejected in favor of $H_a: m_1 - m_2 \neq 0$ if either $w \geq 56$ or $w \leq 6(6+7+1) - 56 = 28$.

Gait	D	L	L	D	D	L	L
Obs	.85	.86	1.09	1.24	1.27	1.31	1.39
Gait	D	L	L	L	D	D	
obs	1.45	1.51	1.53	1.64	1.66	1.82	

$w = 1+4+5+8+12+13 = 43$. Because 43 is neither ≥ 56 nor ≤ 28, we don't reject H_o. There appears to be no difference between m_1 and m_2.

b.

Differences

		Lateral Gait						
		.86	1.09	1.31	1.39	1.51	1.53	1.64
	.85	.01	.24	.46	.54	.66	.68	.79
Diagonal	1.24	-.38	-.15	.07	.15	.27	.29	.40
gait	1.27	-.41	-.18	.04	.12	.24	.26	.37
	1.45	-.59	-.36	-.14	-.06	.06	.08	.19
	1.66	-.80	-.57	-.35	-.27	-.15	-.13	-.02
	1.82	-.96	-.73	-.51	-.43	-.31	-.29	-.18

From Table A.16, c = 35 and $mn - c + 1 = 8$, giving (-.41, .29) as the C.I.

33.

a. With "success" as defined, then Y is a binomial with n = 20. To determine the binomial proportion "p" we realize that since 25 is the hypothesized median, 50% of the distribution should be above 25, thus p = .50. From the Binomial Tables (Table A.1) with n = 20 and p = .50, we see that

$$a = P(Y \geq 15) = 1 - P(Y \leq 14) = 1 - .979 = .021.$$

b. From the same binomial table as in **a**, we find that $P(Y \geq 14) = 1 - P(Y \leq 13) = 1 - .942 = .058$ (a close as we can get to .05), so c = 14. For this data, we would reject H_o at level .058 if $Y \geq 14$. Y = (the number of observations in the sample that exceed 25) = 12, and since 12 is not ≥ 14, we fail to reject H_o.

34.

a. Using the same logic as in Exercise 33, $P(Y \le 5) = .021$, and $P(Y \ge 15) = .021$, so the significance level is $\alpha = .042$.

b. The null hypothesis will not be rejected if the median is between the 6th smallest observation in the data set and the 6th largest, exclusive. (If the median is less than or equal to 14.4, then there are at least 15 observations above, and we reject H_o. Similarly, if any value at least 41.5 is chosen, we have 5 or less observations above.) Thus with a confidence level of 95.8% the median will fall between 14.4 and 41.5.

35.

Sample:	y	x	y	y	x	x	x	y	y
Observations:	3.7	4.0	4.1	4.3	4.4	4.8	4.9	5.1	5.6
Rank:	1	3	5	7	9	8	6	4	2

The value of W' for this data is $w' = 3+6+8+9 = 26$. At level .05, the critical value for the upper-tailed test is (Table A.14, m = 4, n = 5) c = 27 ($\alpha = .056$). Since 26 is not ≥ 27, H_o cannot be rejected at level .05.

36. The only possible ranks now are 1, 2, 3, and 4. Each rank triple is obtained from the corresponding X ordering by the "code" 1 = 1, 2 = 2, 3 = 3, 4 = 4, 5 = 3, 6 = 2, 7 = 1 (so e.g. the X ordering 256 corresponds to ranks 2, 3, 2).

X ordering	ranks	w'	X ordering	ranks	w'	X ordering	ranks	w'
123	123	6	156	132	66	267	221	5
124	124	7	157	131	5	345	343	10
125	123	6	167	121	4	346	342	9
126	122	5	234	234	9	347	341	8
127	121	4	235	233	8	356	332	8
134	134	8	236	232	7	357	331	7
135	133	7	237	231	6	367	321	6
136	132	6	245	243	9	456	432	9
137	131	5	246	242	8	457	431	8
145	143	8	247	241	7	467	421	7
146	142	7	256	232	7	567	321	6
147	141	6	257	231	6			

Since when H_o is true the probability of any particular ordering is 1/35, we easily obtain the null distribution and critical values given in the answer section.

CHAPTER 16

Section 16.1

1. All ten values of the quality statistic are between the two control limits, so no out-of-control signal is generated.

2. All ten values are between the two control limits. However, it is readily verified that all but one plotted point fall below the center line (at height .04975). Thus even though no single point generates an out-of-control signal, taken together, the observed values do suggest that there may be a decrease in the average value of the quality statistic. Such a "small" change is more easily detected by a CUSUM procedure (see section 16.5) than by an ordinary chart.

3. P(10 successive points inside the limits) = P(1st inside) x P(2nd inside) x…x P(10th inside) = $(.998)^{10}$ = .9802. P(25 successive points inside the limits) = $(.998)^{25}$ = .9512. $(.998)^{52}$ = .9011, but $(.998)^{53}$ = .8993, so for 53 successive points the probability that at least one will fall outside the control limits when the process is in control is 1 - .8993 = .1007 > .10.

Section 16.2

4. For Z, a standard normal random variable, $P(-c \le Z \le c) = .995$ implies that $\Phi(c) = P(Z \le c) = .995 + \frac{.005}{2} = .9975$. Table A.3 then gives c = 2.81. The appropriate control limits are therefore $m \pm 2.81s$.

5.

a. P(point falls outside the limits when $m = m_0 + .5s$)

$$= 1 - P\left(m_0 - \frac{3s}{\sqrt{n}} < \overline{X} < m_0 + \frac{3s}{\sqrt{n}} \text{ } when m = m_0 + .5s \right)$$
$$= 1 - P\left(-3 - .5\sqrt{n} < Z < 3 - .5\sqrt{n}\right)$$
$$= 1 - P(-4.12 < Z < 1.882) = 1 - .9699 = .0301.$$

b. $$1 - P\left(m_0 - \frac{3s}{\sqrt{n}} < \overline{X} < m_0 + \frac{3s}{\sqrt{n}} \text{ } when m = m_0 - s \right)$$
$$= 1 - P\left(-3 + \sqrt{n} < Z < 3 + \sqrt{n}\right) = 1 - P(-.76 < Z < 5.24) = .2236$$

c. $$1 - P\left(-3 - 2\sqrt{n} < Z < 3 - 2\sqrt{n}\right) = 1 - P(-7.47 < Z < -1.47) = .6808$$

6. The limits are $13.00 \pm \frac{(3)(.6)}{\sqrt{5}} = 13.00 \pm .80$, from which LCL = 12.20 and UCL = 13.80. Every one of the 22 $\bar{x}$ values is well within these limits, so the process appears to be in control with respect to location.

7. $\bar{\bar{x}} = 12.95$ and $\bar{s} = .526$, so with $a_5 = .940$, the control limits are $12.95 \pm 3\frac{.526}{.940\sqrt{5}} = 12.95 \pm .75 = 12.20, 13.70$. Again, every point $(\bar{x})$ is between these limits, so there is no evidence of an out-of-control process.

8. $\bar{r} = 1.336$ and $b_5 = 2.325$, yielding the control limits $12.95 \pm 3\frac{1.336}{2.325\sqrt{5}} = 12.95 \pm .77 = 12.18, 13.72$. All points are between these limits, so the process again appears to be in control with respect to location.

9. $\bar{\bar{x}} = \frac{2317.07}{24} = 96.54$, $\bar{s} = 1.264$, and $a_6 = .952$, giving the control limits $96.54 \pm 3\frac{1.264}{.952\sqrt{6}} = 96.54 \pm 1.63 = 94.91, 98.17$. The value of $\bar{x}$ on the 22nd day lies above the UCL, so the process appears to be out of control at that time.

10. Now $\bar{\bar{x}} = \frac{2317.07 - 98.34}{23} = 96.47$ and $\bar{s} = \frac{30.34 - 1.60}{23} = 1.250$, giving the limits $96.47 \pm 3\frac{1.250}{.952\sqrt{6}} = 96.47 \pm 1.61 = 94.86, 98.08$. All 23 remaining $\bar{x}$ values are between these limits, so no further out-of-control signals are generated.

11.

a. $P\left(m_0 - \frac{2.81s}{\sqrt{n}} < \bar{X} < m_0 + \frac{2.81s}{\sqrt{n}} \; whenm = m_0\right)$ $= P(-2.81 < Z < 2.81) = .995$, so the probability that a point falls outside the limits is .005 and $ARL = \frac{1}{.005} = 200$.

b. P = P(a point is outside the limits)

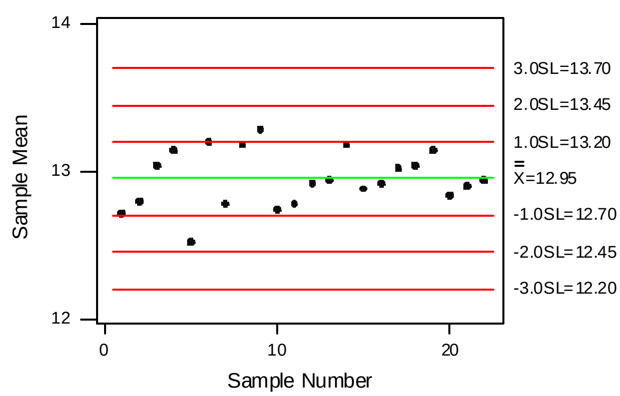
$$= 1 - P\left(m_0 - \frac{2.81s}{\sqrt{n}} < \overline{X} < m_0 + \frac{2.81s}{\sqrt{n}} \, when m = m_0 + s \right)$$

$$= 1 - P\left(-2.81 - \sqrt{n} < Z < 2.81 - \sqrt{n}\right)$$

$$= 1 - P(-4.81 < Z < .81) = 1 - .791 = .209$$. Thus $ARL = \frac{1}{.209} = 4.78$

c. 1 - .9974 = .0026 so $ARL = \frac{1}{.0026} = 385$ for an in-control process, and when $m = m_0 + s$, the probability of an out-of-control point is $1 - P(-3 - 2 < Z < 1)$ $= 1 - P(Z < 1) = .1587$, so $ARL = \frac{1}{.1587} = 6.30$.

12.



The 3-sigma control limits are from problem 7. The 2-sigma limits are $12.95 \pm .50 = 12.45, 13.45$, and the 1-sigma limits are $12.95 \pm .25 = 12.70, 13.20$. No points fall outside the 2-sigma limits, and only two points fall outside the 1-sigma limits. There are also no runs of eight on the same side of the center line – the longest run on the same side of the center line is four (the points at times 10, 11, 12, 13). No out-of-control signals result from application of the supplemental rules.

13. $\overline{\overline{x}} = 12.95$, IQR = .4273, $k_5 = .990$. The control limits are

$$12.95 \pm 3 \frac{.4273}{.990\sqrt{5}} = 12.45, 13.45 = 12.37, 13.53 .$$

Section 16.3

14. $\Sigma s_i = 4.895$ and $\bar{s} = \frac{4.895}{24} = .2040$. With $a_5 = .940$, the lower control limit is zero and the upper limit is $.2040 + \frac{3(.2040)\sqrt{1-(.940)^2}}{.940} = .2040 + .2221 = .4261$. Every s_I is between these limits, so the process appears to be in control with respect to variability.

15.

a. $\bar{r} = \frac{85.2}{30} = 2.84$, $b_4 = 2.058$, and $c_4 = .880$. Since n = 4, LCL = 0 and UCL
$= 2.84 + \frac{3(.880)(2.84)}{2.058} = 2.84 + 3.64 = 6.48$.

b. $\bar{r} = 3.54$, $b_8 = 2.844$, and $c_8 = .820$, and the control limits are
$= 3.54 \pm \frac{3(.820)(3.54)}{2.844} = 3.54 \pm 3.06 = .48, 6.60$.

16. $\bar{s} = .5172$, $a_5 = .940$, LCL = 0 (since n = 5) and UCL =
$.5172 + \frac{3(.5172)\sqrt{1-(.940)^2}}{.940} = .5172 + .5632 = 1.0804$. The largest s_I is s_9 = .963, so all points fall between the control limits.

17. $\bar{s} = 1.2642$, $a_6 = .952$, and the control limits are
$1.2642 \pm \frac{3(1.2642)\sqrt{1-(.952)^2}}{.952} = 1.2642 \pm 1.2194 = .045, 2.484$. The smallest s_I is s_{20} = .75, and the largest is s_{12} = 1.65, so every value is between .045 and 2.434. The process appears to be in control with respect to variability.

18. $\Sigma s_i^2 = 39.9944$ and $\bar{s}^2 = \frac{39.9944}{24} = 1.6664$, so LCL = $\frac{(1.6664)(.210)}{5} = .070$, and UCL = $\frac{(1.6664)(20.515)}{5} = 6.837$. The smallest s^2 value is $s_{20}^2 = (.75)^2 = .5625$ and the largest is $s_{12}^2 = (1.65)^2 = 2.723$, so all s_i^2's are between the control limits.

Section 16.4

19. $\overline{p} = \Sigma \dfrac{\hat{p}_i}{k}$ where $\Sigma \hat{p}_i = \dfrac{x_1}{n} + ... + \dfrac{x_k}{n} = \dfrac{x_1 + ... + x_k}{n} = \dfrac{578}{100} = 5.78$. Thus $\overline{p} = \dfrac{5.78}{25} = .231$.

a. The control limits are $.231 \pm 3\sqrt{\dfrac{(.231)(.769)}{100}} = .231 \pm .126 = .105, .357$.

b. $\dfrac{13}{100} = .130$, which is between the limits, but $\dfrac{39}{100} = .390$, which exceeds the upper control limit and therefore generates an out-of-control signal.

20. $\Sigma x_i = 567$, from which $\overline{p} = \dfrac{\Sigma x_i}{nk} = \dfrac{567}{(200)(30)} = .0945$. The control limits are $.0945 \pm 3\sqrt{\dfrac{(.0945)(.9055)}{200}} = .0945 \pm .0621 = .0324, .1566$. The smallest x_i is $x_7 = 7$, with $\hat{p}_7 = \dfrac{7}{200} = .0350$. This (barely) exceeds the LCL. The largest x_i is $x_5 = 37$, with $\hat{p}_5 = \dfrac{37}{200} = .185$. Thus $\hat{p}_5 > UCL = .1566$, so an out-of-control signal is generated. This is the only such signal, since the next largest x_i is $x_{25} = 30$, with $\hat{p}_{25} = \dfrac{30}{200} = .1500 < UCL$.

21. LCL > 0 when $\overline{p} > 3\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}}$, i.e. (after squaring both sides) $50\overline{p}^2 > 3\overline{p}(1-\overline{p})$, i.e. $50\overline{p} > 3(1-\overline{p})$, i.e. $53\overline{p} > 3 \Rightarrow \overline{p} = \dfrac{3}{53} = .0566$.

22. The suggested transformation is $Y = h(X) = \sin^{-1}\left(\sqrt{X/n}\right)$, with approximate mean value $\sin^{-1}\left(\sqrt{p}\right)$ and approximate variance $\dfrac{1}{4n}$. $\sin^{-1}\left(\sqrt{x_1/n}\right) = \sin^{-1}\left(\sqrt{.050}\right) = .2255$ (in radians), and the values of $y_i = \sin^{-1}\left(\sqrt{x_i/n}\right)$ for i = 1, 2, 3, …, 30 are

0.2255	0.2367	0.2774	0.3977
0.3047	0.3537	0.3381	0.2868
0.3537	0.3906	0.2475	0.2367
0.2958	0.2774	0.3218	0.3218
0.4446	0.2868	0.2958	0.2678
0.3133	0.3300	0.3047	0.3835
0.1882	0.3047	0.2475	
0.3614	0.2958	0.3537	

These give $\Sigma y_i = 9.2437$ and $\bar{y} = .3081$. The control limits are $\bar{y} \pm 3\sqrt{\frac{1}{4n}} = .3081 \pm 3\sqrt{\frac{1}{800}} = .3081 \pm .1091 = .2020, .4142$. In contrast ot the result of exercise 20, there I snow one point below the LCL (.1882 < .2020) as well as one point above the UCL.

23. $\Sigma x_i = 102$, $\bar{x} = 4.08$, and $\bar{x} \pm 3\sqrt{\bar{x}} = 4.08 \pm 6.06 \approx (-2.0, 10.1)$. Thus LCL = 0 and UCL = 10.1. Because no x_i exceeds 10.1, the process is judged to be in control.

24. $\bar{x} - 3\sqrt{\bar{x}} < 0$ is equivalent to $\sqrt{\bar{x}} < 3$, i.e. $\bar{x} < 9$.

25. With $u_i = \dfrac{x_i}{g_i}$, the u_i's are 3.75, 3.33, 3.75, 2.50, 5.00, 5.00, 12.50, 12.00, 6.67, 3.33, 1.67, 3.75, 6.25, 4.00, 6.00, 12.00, 3.75, 5.00, 8.33, and 1.67 for I = 1, …, 20, giving $\bar{u} = 5.5125$. For $g_i = .6$, $\bar{u} \pm 3\sqrt{\dfrac{\bar{u}}{g_i}} = 5.5125 \pm 9.0933$, LCL = 0, UCL = 14.6. For $g_i = .8$, $\bar{u} \pm 3\sqrt{\dfrac{\bar{u}}{g_i}} = 5.5125 \pm 7.857$, LCL = 0, UCL = 13.4. For $g_i = 1.0$, $\bar{u} \pm 3\sqrt{\dfrac{\bar{u}}{g_i}} = 5.5125 \pm 7.0436$, LCL = 0, UCL = 12.6. Several u_i's are close to the corresponding UCL's but none exceed them, so the process is judged to be in control.

26. $y_i = 2\sqrt{x_i}$ and the $y_i's$ are 3/46, 5.29, 4.47, 4.00, 2.83, 5.66, 4.00, 3.46, 3.46, 4.90, 5.29, 2.83, 3.46, 2.83, 4.00, 5.29, 3.46, 2.83, 4.00, 4.00, 2.00, 4.47, 4.00, and 4.90 for I = 1, …, 25, from which $\Sigma y_i = 98.35$ and $\bar{y} = 3.934$. Thus $\bar{y} \pm 3 = 3.934 \pm 3 = .934, 6.934$. Since every y_i is well within these limits it appears that the process is in control.

Section 16.5

27. $m_0 = 16$, $k = \frac{\Delta}{2} = 0.05$, $h = .20$, $d_i = \max(0, d_{i-1} + (\bar{x}_i - 16.05))$, $e_i = \max(0, e_{i-1} + (\bar{x}_i - 15.95))$.

i	$\bar{x}_i - 16.05$	d_i	$\bar{x}_i - 15.95$	e_i
1	-0.058	0	0.024	0
2	0.001	0.001	0.101	0
3	0.016	0.017	0.116	0
4	-0.138	0	-0.038	0.038
5	-0.020	0	0.080	0
6	0.010	0.010	0.110	0
7	-0.068	0	0.032	0
8	-0.151	0	-0.054	0.054
9	-0.012	0	0.088	0
10	0.024	0.024	0.124	0
11	-0.021	0.003	0.079	0
12	-0.115	0	-0.015	0.015
13	-0.018	0	0.082	0
14	-0.090	0	0.010	0
15	0.005	0.005	0.105	0

For no time r is it the case that $d_r > .20$ or that $e_r > .20$, so no out-of-control signals are generated.

28. $m_0 = .75$, $k = \frac{\Delta}{2} = 0.001$, $h = .003$, $d_i = \max\left(0, d_{i-1} + \left(\bar{x}_i - .751\right)\right)$,
$e_i = \max\left(0, e_{i-1} + \left(\bar{x}_i - .749\right)\right)$.

i	$\bar{x}_i - .751$	d_i	$\bar{x}_i - .749$	e_i
1	-.0003	0	.0017	0
2	-.0006	0	.0014	0
3	-.0018	0	.0002	0
4	-.0009	0	.0011	0
5	-.0007	0	.0013	0
6	.0000	0	.0020	0
7	-.0020	0	.0000	0
8	-.0013	0	.0007	0
9	-.0022	0	-.0002	.0002
10	-.0006	0	.0014	0
11	.0006	.0006	.0026	0
12	-.0038	0	-.0018	.0018
13	-.0021	0	-.0001	.0019
14	-.0027	0	-.0007	.0026
15	-.0039	0	-.0019	.0045*
16	-.0012	0	.0008	.0037
17	-.0050	0	-.0030	.0067
18	-.0028	0	-.0008	.0075
19	-.0040	0	-.0020	.0095
20	-.0017	0	.0003	.0092
21	-.0048	0	-.0028	.0120
22	-.0029	0	-.0009	.0129

Clearly $e_{15} = .0045 > .003 = h$, suggesting that the process mean has shifted to a value smaller than the target of .75.

29. Connecting 600 on the in-control ARL scale to 4 on the out-of-control scale and extending to the k' scale gives k' = .87. Thus $k' = \frac{\Delta/2}{s/\sqrt{n}} = \frac{.002}{.005/\sqrt{n}}$ from which $\sqrt{n} = 2.175 \Rightarrow n = 4.73 = s$. Then connecting .87 on the k' scale to 600 on the out-of-control ARL scale and extending to h' gives h' = 2.8, so
$h = \left(\frac{s}{\sqrt{n}}\right)(2.8) = \left(\frac{.005}{\sqrt{5}}\right)(2.8) = .00626$.

30. In control ARL = 250, out-of-control ARL = 4.8, from which $k' = .7 = \dfrac{\Delta/2}{s/\sqrt{n}} = \dfrac{s\ /2}{s/\sqrt{n}} = \dfrac{\sqrt{n}}{2}$. So $\sqrt{n} = 1.4 \Rightarrow n = 1.96 \approx 2$. Then h' = 2.85, giving $h = \left(\dfrac{s}{\sqrt{n}}\right)(2.85) = 2.0153s$.

Section 16.6

31. For the binomial calculation, n = 50 and we wish

$$P(X \le 2) = \binom{50}{0} p^0 (1-p)^{50} + \binom{50}{1} p^1 (1-p)^{49} + \binom{50}{2} p^2 (1-p)^{48}$$

$= (1-p)^{50} + 50p(1-p)^{49} + 1225p^2(1-p)^{48}$ when p = .01, .02, …, .10. For the hypergeometric calculation

$$P(X \le 2) = \frac{\binom{M}{0}\binom{500-M}{50}}{\binom{500}{50}} + \frac{\binom{M}{1}\binom{500-M}{49}}{\binom{500}{50}} + \frac{\binom{M}{2}\binom{500-M}{48}}{\binom{500}{50}}$$, to be

calculated for M = 5, 10, 15, …, 50. The resulting probabilities appear in the answer section in the text.

32. $P(X \le 1) = \binom{50}{0} p^0 (1-p)^{50} + \binom{50}{1} p^1 (1-p)^{49} = (1-p)^{50} + 50p(1-p)^{49}$

p	.01	.02	.03	.04	.05	.06	.07	.08	.09	.10
$P(X \le 1)$	.9106	.7358	.5553	.4005	.2794	.1900	.1265	.0827	.0532	.0338

33. $P(X \le 2) = \binom{100}{0} p^0 (1-p)^{100} + \binom{100}{1} p^1 (1-p)^{99} + \binom{100}{2} p^2 (1-p)^{98}$

p	.01	.02	.03	.04	.05	.06	.07	.08	.09	.10
$P(X \le 2)$	.9206	.6767	.4198	.2321	.1183	.0566	.0258	.0113	.0048	.0019

For values of p quite close to 0, the probability of lot acceptance using this plan is larger than that for the previous plan, whereas for larger p this plan is less likely to result in an “accept the lot” decision (the dividing point between “close to zero” and “larger p” is someplace between .01 and .02). In this sense, the current plan is better.

34. $\dfrac{LTPD}{AQL} = \dfrac{.07}{.02} = 3.5 \approx 3.55$, which appears in the $\dfrac{p_1}{p_2}$ column in the c = 5 row. Then

$$n = \frac{np_1}{p_1} = \frac{2.613}{.02} = 130.65 \approx 131.$$

$$P(X > 5 \text{ when } p = .02) = 1 - \sum_{x=0}^{5} \binom{131}{x} (.02)^x (.98)^{131-x} = .0487 \approx .05$$

$$P(X \le 5 \text{ when } p = .07) = \sum_{x=0}^{5} \binom{131}{x} (.07)^x (.93)^{131-x} = .0974 \approx .10$$

35. P(accepting the lot) = P(X_1 = 0 or 1) + P(X_1 = 2, X_2 = 0, 1, 2, or 3) + P(X_1 = 3, X_2 = 0, 1, or 2) = P(X_1 = 0 or 1) + P(X_1 = 2)P(X_2 = 0, 1, 2, or 3) + P(X_1 = 3)P(X_2 = 0, 1, or 2).

p = .01: $= .9106 + (.0756)(.9984) + (.0122)(.9862) = .9981$

p = .05: $= .2794 + (.2611)(.7604) + (.2199)(.5405) = .5968$

p = .10: $= .0338 + (.0779)(.2503) + (.1386)(.1117) = .0688$

36. P(accepting the lot) = P(X_1 = 0 or 1) + P(X_1 = 2, X_2 = 0 or 1) + P(X_1 = 3, X_2 = 0) [since $c_2 = r_1 - 1 = 3$] = P(X_1 = 0 or 1) + P(X_1 = 2)P(X_2 = 0 or 1) + P(X_1 = 3)P(X_2 = 0)

$$= \sum_{x=0}^{1} \binom{50}{x} p^x (1-p)^{50-x} + \binom{50}{2} p^2 (1-p)^{48} \cdot \sum_{x=0}^{1} \binom{100}{x} p^x (1-p)^{100-x}$$

$$= \binom{50}{3} p^3 (1-p)^{47} \cdot \binom{100}{0} p^0 (1-p)^{100}.$$

p = .02: $= .7358 + (.1858)(.4033) + (.0607)(.1326) = .8188$

p = .05: $= .2794 + (.2611)(.0371) + (.2199)(.0059) = .2904$

p = .10: $= .0338 + (.0779)(.0003) + (.1386)(.0000) = .0038$

37.

a. $AOQ = pP(A) = p[(1-p)^{50} + 50p(1-p)^{49} + 1225p^2(1-p)^{48}]$

p	.01	.02	.03	.04	.05	.06	.07	.08	.09	.10
AOQ	.010	.018	.024	.027	.027	.025	.022	.018	.014	.011

b. p = .0447, AOQL = .0447P(A) = .0274

c. ATI = 50P(A) + 2000(1 – P(A))

p	.01	.02	.03	.04	.05	.06	.07	.08	.09	.10
ATI	77.3	202.1	418.6	679.9	945.1	1188.8	1393.6	1559.3	1686.1	1781.6

38. $AOQ = pP(A) = p[(1-p)^{50} + 50p(1-p)^{49}]$. Exercise 32 gives P(A), so multiplying each entry in the second row by the corresponding entry in the first row gives AOQ:

p	.01	.02	.03	.04	.05	.06	.07	.08	.09	.10
AOQ	.0091	.0147	.0167	.0160	.0140	.0114	.0089	.0066	.0048	.0034

ATI = 50P(A) + 2000(1 – P(A))

p	.01	.02	.03	.04	.05	.06	.07	.08	.09	.10
ATI	224.3	565.2	917.2	1219.0	1455.2	1629.5	1753.3	1838.7	1896.3	1934.1

$\frac{d}{dp}AOQ = \frac{d}{dp}\left[pP(A) = p[(1-p)^{50} + 50p(1-p)^{49}]\right] = 0$ gives the quadratic equation $2499p^2 - 48p - 1 = 0$, from which $p = \frac{48 + 110.91}{4998} = .0318$, and $AOQL = .0318P(A) \approx .0167$.

Supplementary Exercises

39. n = 6, k = 26, $\Sigma\bar{x}_i = 10{,}980$, $\bar{\bar{x}} = 422.31$, $\Sigma s_i = 402$, $\bar{s} = 15.4615$, $\Sigma r_i = 1074$, $\bar{r} = 41.3077$

S chart: $15.4615 \pm \frac{3(15.4615)\sqrt{1-(.952)^2}}{.952} = 15.4615 \pm 14.9141 \approx .55, 30.37$

R chart: $41.31 \pm \frac{3(.848)(41.31)}{2.536} = 41.31 \pm 41.44$, so LCL = 0, UCL = 82.75

$\overline{X}$ chart based on $\bar{s}$: $422.31 \pm \frac{3(15.4615)}{.952\sqrt{6}} = 402.42, 442.20$

$\overline{X}$ chart based on $\bar{r}$: $422.31 \pm \frac{3(41.3077)}{2.536\sqrt{6}} = 402.36, 442.26$

40. A c chart is appropriate here. $\Sigma x_i = 92$ so $\bar{x} = \frac{92}{24} = 3.833$, and

$\bar{x} \pm 3\sqrt{\bar{x}} = 3.833 \pm 5.874$, giving LCL = 0 and UCL = 9.7. Because $x_{22} = 10 >$ UCL, the process appears to have been out of control at the time that the 22nd plate was obtained.

41.

i	$\bar{x}_i$	s_i	r_i
1	50.83	1.172	2.2
2	50.10	.854	1.7
3	50.30	1.136	2.1
4	50.23	1.097	2.1
5	50.33	.666	1.3
6	51.20	.854	1.7
7	50.17	.416	.8
8	50.70	.964	1.8
9	49.93	1.159	2.1
10	49.97	.473	.9
11	50.13	.698	.9
12	49.33	.833	1.6
13	50.23	.839	1.5
14	50.33	.404	.8
15	49.30	.265	.5
16	49.90	.854	1.7
17	50.40	.781	1.4
18	49.37	.902	1.8
19	49.87	.643	1.2
20	50.00	.794	1.5
21	50.80	2.931	5.6
22	50.43	.971	1.9

$\Sigma s_i = 19.706$, $\bar{s} = .8957$, $\Sigma \bar{x}_i = 1103.85$, $\bar{\bar{x}} = 50.175$, $a_3 = .886$, from which an s chart has LCL = 0 and UCL = $.8957 + \dfrac{3(.8957)\sqrt{1-(.886)^2}}{.886} = 2.3020$, and $s_{21} = 2.931 > UCL$. Since an assignable cause is assumed to have been identified we eliminate the 21st group. Then $\Sigma s_i = 16.775$, $\bar{s} = .7998$, $\bar{\bar{x}} = 50.145$. The resulting UCL for an s chart is 2.0529, and $s_i < 2.0529$ for every remaining i. The $\bar{x}$ chart based on $\bar{s}$ has limits $50.145 \pm \dfrac{3(.7988)}{.886\sqrt{3}} = 48.58, 51.71$. All $\bar{x}_i$ values are between these limits.

42. $\overline{p} = .0608$, n = 100, so $UCL = n\overline{p} + 3\sqrt{n\overline{p}(1-\overline{p})} = 6.08 + 3\sqrt{6.08(.9392)}$ $= 6.08 + 7.17 = 13.25$ and LCL = 0. All points are between these limits, as was the case for the p-chart. The p-chart and np-chart will always give identical results since

$$\overline{p} - 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} < \hat{p}_i < \overline{p} + 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} \quad \textbf{iff}$$

$$n\overline{p} - 3\sqrt{n\overline{p}(1-\overline{p})} < n\hat{p}_i = x_i < n\overline{p} + 3\sqrt{n\overline{p}(1-\overline{p})}$$

43. $\Sigma n_i = 4(16) + (3)(4) = 76$, $\Sigma n_i \overline{x}_i = 32{,}729.4$, $\overline{\overline{x}} = 430.65$,

$s^2 = \dfrac{\Sigma(n_i - 1)s_i^2}{\Sigma(n_i - 1)} = \dfrac{27{,}380.16 - 5661.4}{76 - 20} = 590.0279$, so s = 24.2905. For variation:

when n = 3, $UCL = 24.2905 + \dfrac{3(24.2905)\sqrt{1-(.886)^2}}{.886} = 24.29 + 38.14 = 62.43$,

when n = 4, $UCL = 24.2905 + \dfrac{3(24.2905)\sqrt{1-(.921)^2}}{.921} = 24.29 + 30.82 = 55.11$.

For location: when n = 3, $430.65 \pm 47.49 = 383.16, 478.14$, and when n = 4, $430.65 \pm 39.56 = 391.09, 470.21$.

44.

a. Provided the $E(\overline{X}_i) = m$ for each i,

$$\begin{aligned}
E(W_t) &= aE(\overline{X}_t) + a(1-a)E(\overline{X}_{t-1}) + \dots + a(1-a)^{t-1}E(\overline{X}_1) + (1-a)^t m \\
&= m\left[a + a(1-a) + \dots + a(1-a)^{t-1} + (1-a)^t\right] \\
&= m\left[a\left(1 + (1-a) + \dots + (1-a)^{t-1}\right) + (1-a)^t\right] \\
&= m\left[a\sum_{i=0}^{\infty}(1-a)^i - a\sum_{i=t}^{\infty}(1-a)^i + (1-a)^t\right] \\
&= m\left[\frac{a}{1-(1-a)} - a(1-a)^t \cdot \frac{1}{1-(1-a)} + (1-a)^t\right] = m
\end{aligned}$$

b. $V(W_t) = a^2 V(\overline{X}_t) + a^2(1-a)^2 V(\overline{X}_{t-1}) + \dots + a^2(1-a)^{2(t-1)}V(\overline{X}_1)$

$= a^2\left[1 + (1-a)^2 + \dots + (1-a)^{2(t-1)}\right] \cdot V(\overline{X}_1)$

$= a^2\left[1 + C + \dots + C^{t-1}\right] \cdot \dfrac{s^2}{n}$ (where $C = (1-a)^2$.)

$= a^2 \dfrac{1-C^t}{1-C} \cdot \dfrac{s^2}{n}$, which gives the desired expression.

c. From Example 16.8, $\sigma = .5$ (or $\bar{s}$ can be used instead). Suppose that we use $\alpha = .6$ (not specified in the problem). Then

$$w_0 = \mu_0 = 40$$
$$w_1 = .6\bar{x}_1 + .4\mu_0 = .6(40.20) + .4(40) = 40.12$$
$$w_2 = .6\bar{x}_2 + .4w_1 = .6(39.72) + .4(40.12) = 39.88$$
$$w_3 = .6\bar{x}_3 + .4w_2 = .6(40.42) + .4(39.88) = 40.20$$
$$w_4 = 40.07,\ w_5 = 40.06,\ w_6 = 39.88,\ w_7 = 39.74,\ w_8 = 40.14,$$
$$w_9 = 40.25,\ w_{10} = 40.00,\ w_{11} = 40.29,\ w_{12} = 40.36,\ w_{13} = 40.51,$$
$$w_{14} = 40.19,\ w_{15} = 40.21,\ w_{16} = 40.29$$

$$\sigma_1^2 = \frac{.6\left[1-(1-.6)^2\right]}{2-.6}\cdot\frac{.25}{4} = .0225,\ \sigma_1 = .1500,$$
$$\sigma_2^2 = \frac{.6\left[1-(1-.6)^4\right]}{2-.6}\cdot\frac{.25}{4} = .0261,\ \sigma_2 = .1616,$$
$$\sigma_3 = .1633,\ \sigma_4 = .1636,\ \sigma_5 = .1637 = \sigma_6 \ldots \sigma_{16}$$

Control limits are:

For t = 1, $40 \pm 3(.1500) = 39.55, 40.45$
For t = 2, $40 \pm 3(.1616) = 39.52, 40.48$
For t = 3, $40 \pm 3(.1633) = 39.51, 40.49$.
These last limits are also the limits for t = 4, …, 16.

Because $w_{13} = 40.51 > 40.49 = \text{UCL}$, an out-of-control signal is generated.