Exercise 10-7

(b) Test the null hypothesis that the population mean equals 105 μg/100 ml versus the two-sided alternative. Using a t test with 19 degrees of freedom, the test statistic equals .49 with a p-value of .6. Using a Wilcoxon signed rank test for sample size 20, the test statistic equals 98.5 (or 72.5), with a p-value of .6. Using a sign test for sample size 20, the p-value equals .8. These experimental results are consistent with the null hypothesis.

(c) A 90% confidence interval for the population mean based on a t distribution with 19 degrees of freedom is (103.61, 107.49) μg/100 ml. An approximate 90% confidence interval based on the Wilcoxon signed rank distribution for sample size 20 is (103.5, 107.5) μg/100ml. An approximate 90% confidence interval based on a binomial distribution for sample size 20 is ((103.9, 108.0) μg/100 ml.

Exercise 10-8 A large sample approximate 95% confidence interval for the proportion of experimentally treated livers expected to last at least 9.5 hours is .366 to .509.

Exercise 10-9

(b) The test statistic equals −4.02 with 4 degrees of freedom, with a two-sided p-value of .016. A 90% confidence interval for the population mean is 19.046 to 20.954 ounces.

(c) The test statistic equals 0, with a two-sided p-value of .059. An approximate 90% confidence interval for the population mean is 19.25 to 21.00 ounces.

(d) The test statistic equals 0, with a two-sided p-value of .0625. An approximate 90% confidence interval for the population median is 19.18 to 21.15 ounces.

Exercise 10-10 The large sample test statistic equals 4.35, with a two-sided p-value of .0001. Using the standard Gaussian distribution, an approximate large sample 95% confidence interval for the carrier population mean is 114.6 to 237.2 units.

Exercise 10-11

(a) The Wilcoxon signed rank test statistic equals 1, with a two-sided p-value of .06. Interval estimates for the mean weight of twins born to exercised goats: (1018, 1512) grams with approximate confidence level 90.7%, (960, 1529) grams with approximate confidence level 94.1%, (745.5, 1660) grams with approximate confidence level 96.4%.

(b) Using the sign test, the test statistic equals 0, with a p-value of .2. Interval estimates for the population mean: (1065, 1465) grams with approximate confidence level 90.0%, (899, 1566) grams with approximate confidence level 95.0%, (745, 1660) grams with approximate confidence level 96.9%.

Exercise 10-12

(a) The test statistic equals 0.96 with 4 degrees of freedom, two-sided p-value=.4. A 90% confidence interval for the population mean bomb base height is (.827, .838) inches; a 95% interval is (.825, .839) inches; a 99% confidence interval is (.821, .844) inches.

(b) Using the sign test, the test statistic equals 0, with a two sided p-value of 1. An approximate 90% confidence interval for the population median is (.827, .839) inches; an approximate 94% confidence interval is (.826, .840) inches.

Exercise 10-13 The large sample test statistic equals 1.16, with a two-sided p-value of .2.

Exercise 10-14 The student can find the distribution of T− using Table 1 on page 632 of the text and then compare with the distribution of T+ shown in Table 2 on page 632.

Exercise 10-15 There are 16 possible assignments of + and − signs to ranks 1, 2, 3 and 4. T+ is the sum of the positive signed ranks. For values of c equal to 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10, the corresponding probabilities P(T+ = c) are, respectively: 1/16, 1/16, 1/16, 2/16, 2/16, 2/16, 2/16, 2/16, 1/16, 1/16 and 1/16. The corresponding cumulative probabilities P(T+ $\leq$ c) are, respectively: 1/16, 2/16, 3/16, 5/16, 7/16, 9/16, 11/16, 13/16, 14/16, 15/16, 16/16=1.

Exercise 10-16
(b) Let p denote the proportion of children who guess correctly. The null hypothesis is that p equals 1/2, versus the two-sided alternative. The two-sided p-value based on the Binomial(24, .5) distribution equals .006. These results are inconsistent with the null hypothesis of chance guessing.

Exercise 10-17
(a) Let p denote the proportion of parents who guess correctly. The null hypothesis is that p equals 1/2, versus the two-sided alternative. The two-sided p-value based on the Binomial(18, .5) distribution equals .001. These results are inconsistent with the null hypothesis of chance guessing.

Exercise 10-18 A large sample approximate 95% confidence interval for the proportion of interest is (.029, .071).

Exercise 10-19 An approximate 90% confidence interval for the proportion of interest, based on large or small samples, is (.3, .8).

Exercise 10-20 Sampling situations appropriate to the three tests listed correspond to situations when necessary assumptions are met.

Chapter 11

Exercise 11-1
(b) For a paired t test, the test statistic equals −3.60 with 9 degrees of freedom, with a two-sided p-value of .006. For a Wilcoxon signed rank test, the test statistic equals 0, with a two-sided p-value of .006. For a sign test, the test statistic equals 0, with a two-sided p-value of .002.
(c) A 95% confidence interval for the mean difference in numbers of mosquitos captured, based on a t distribution is −97.8 to −22.2. An approximate 95% confidence interval for the mean (median) difference in numbers of mosquitos captured, based on a Wilcoxon signed rank distribution is −105.0 to −28.0. An approximate 95% confidence interval for the median difference in numbers of mosquitoes captured, based on a binomial distribution is −77.67 to −25.6 mosquitos.

Exercise 11-2
(a) The large sample test statistic equals 3.76, with a two-sided p-value of .0002.
(b) An approximate 99% confidence interval for the difference between the two proportions is .010 to .055.

Exercise 11-3
(b) For a two-sample t test, the test statistic equals 1.72 with 14 degrees of freedom, with a two-sided p-value of .1. For a Wilcoxon-Mann-Whitney test, the test statistic equals 83 (or 53), with a two-sided p-value of .1. For the median test, the overall median is 43. For Magnum, 2 values are below and 5 above this median; for Long Rifle, 5 values are below and 2 above. (We ignore the 2 data values equal to the overall median of 43.) The p-value is:

$$\text{p-value} = 2*\left(\binom{7}{2}\binom{7}{5} + \binom{7}{1}\binom{7}{6} + \binom{7}{0}\binom{7}{7}\right)/\binom{14}{7} = .3.$$

(c) A 95% confidence interval for the difference between the two mean scores, based on a t distribution, is −.6 to 5.35. An approximate 95-96% confidence interval for the difference between the two mean (median) scores based on a Wilcoxon-Mann-Whitney distribution is −1.001 to 6.000.

Exercise 11-4

(b) Two sample t test statistic = 2.46 with 14 degrees of freedom, p-value = .03. Wilcoxon-Mann-Whitney test statistic = 89.5 (or 46.5), p-value = .03. For the median test, the overall median is 173. For Analyst 1, 2 values are below 173 and 5 above. For Analyst 2, 5 values are below 173 and 0 above. The p-value is:

$$\text{p-value} = \left(\binom{7}{2}\binom{5}{5} + \binom{7}{7}\binom{5}{0}\right)\Big/\binom{12}{7} = .0265 + .0013 = .03.$$

Each test is inconsistent with the null hypothesis of equal population medians.

(c) Using the t distribution with 14 degrees of freedom, confidence intervals for the difference between the means for analyst 1 and analyst 2 are: 90%, (.28, 1.72); 95%, (.13, 1.87); 99%, (−.21, 2.21). Using the Wilcoxon-Mann-Whitney distribution for sample sizes 8 and 8, confidence intervals for the difference between the medians for analyst 1 and analyst 2 are: 91.7%, (.50, 1.50); 95.9%, (.50, 2.0); 99.3%, (−.50, 2.50).

Exercise 11-5

(b) The overall median is 4.615. In the alcohol group, 7 are below and 3 are above this median. In the control group, 3 are below and 7 are above this overall median. For a two-sided alternative, the p-value is

$$\text{p-value} = 2*\left(\binom{10}{3}\binom{10}{7}+\binom{10}{2}\binom{10}{8}+\binom{10}{1}\binom{10}{9}+\binom{10}{0}\binom{10}{10}\right)\Big/\binom{20}{10} = .18,$$

borderline but suggesting a greater mean for the control group.

(c) Wilcoxon-Mann-Whitney test statistic = 70.0, p-value = .009, inconsistent with the null hypothesis of equal population medians, suggesting greater median plasma estrogen levels for control monkeys. A 95.5% confidence interval for the difference between the two population medians is −2.60 to −.44 nanograms/deciliter.

(d) Two sample t test statistic = −3.70 with 18 degrees of freedom, p-value = .002. A 95% confidence interval for the difference between the two population means is −2.52 to −.70 nanograms/deciliter.

Exercise 11-6

(b) The overall median is between 251.29 and 262.16. In the alcohol group, 7 are below and 3 above, while in the control group 3 are below and 7 above this overall median. The two-sided p-value is .18, as in Exercise 11-5(b).

(c) Wilcoxon-Mann-Whitney test statistic = 81, p-value = .08, borderline. A 95.5% confidence interval for the difference between the two population medians is (−578.2, 9.5) nanograms/deciliter.

(d) Two sample t test statistic = −2.36 with 18 degrees of freedom, p-value = .03, inconsistent with the null hypothesis of equal population means, suggesting greater mean plasma testosterone levels for control monkeys. A 95% confidence interval for the difference between the two population means is (−564, −32) nanograms/deciliter.

(e) Variation in the two groups looks more similar after taking logarithms.

(f) Using the logarithm of response, two sample t test statistic = −2.21 with 18 degrees of freedom, p-value = .04.

Exercise 11-7

(b) The overall median is between 39.6 and 41.7. For allergics, 2 are below and 7 above, while for nonallergics 9 are below and 4 above the overall median. The two-sided p-value is:

$$\text{p-value} = 2*\left(\binom{9}{2}\binom{13}{9}+\binom{9}{1}\binom{13}{10}+\binom{9}{0}\binom{13}{11}\right)/\binom{22}{11} = .08,$$

borderline but suggesting allergics have greater median sputum histamine levels.

(c) Wilcoxon-Mann-Whitney test statistic = 151 (or 102), p-value = .002, again suggesting greater median for allergics. A 95.5% confidence interval for the difference between the two population medians is (22.3, 95.6) μg/g dry weight sputum.

(d) Using the logarithm of response, two sample t test statistic = 3.51 with 20 degrees of freedom, p-value = .002. (Without the transformation, test statistic = 2.02, p-value = .06.)

Exercise 11-8

(b) Two sample t test statistic = 1.17 with 22 degrees of freedom, p-value = .3, consistent with the null hypothesis of equal population means. Wilcoxon-Mann-Whitney test statistic = 177 (or 123), p-value = .13. For the median test, the overall median is 11.6. For hand values, 4 are below and 8 above this overall median, while 7 automatic values are below and 3 above. The two-sided p-value is:

$$\text{p-value} = 2*\left(\binom{12}{8}\binom{10}{3}+\binom{12}{9}\binom{10}{2}+\binom{12}{10}\binom{10}{1}+\binom{12}{11}\binom{10}{0}\right)/\binom{22}{11} = .2.$$

(c) Using the t distribution with 22 degrees of freedom, a 95% confidence interval for the difference between hand and automatic means is (−4.2, 15.1). Using the Wilcoxon-Mann-Whitney distribution for sample sizes 12 and 12, a 95.4% confidence interval for the difference between hand and automatic medians is (−1.13, 11.40).

Exercise 11-9

The large sample test statistic is 9.2, with a two-sided p-value less than .0001. An approximate 99% confidence interval for the difference between the two mean lifetimes is 206 to 367 days.

Exercise 11-10

(a) We want to test the null hypothesis that the proportion of men with byssinosis complaints equals the proportion of women with such complaints, against the two-sided alternative. The sample proportion of men with complaints is 128/2916=.0439; the sample proportion of women with complaints is 37/2503=.0148; the sample proportion of complaints for men and women combined is 165/5419=.0304. The test statistic is

$$\frac{.0439-.0148}{\sqrt{(.0304)(.9696)(1/2916 + 1/2503)}} = .0291/.0047 = 6.2$$

and the p-value equals 0 to four decimal places. The results are strongly inconsistent with the null hypothesis. The proportion of men with byssinosis complaints seems to be significantly greater than the proportion of women with such complaints.

(b) An approximate, large sample, confidence interval for the difference between the proportion of men with complaints and the proportion of women with complaints has the form

$$.0291 \pm c\sqrt{(.0439)(.9561)/2916 + (.0148)(.9852)/2503}$$

where c comes from the standard Gaussian distribution. For a 90% interval, c=1.65 and the interval is (.0217, .0365).

For a 95% interval, c=1.96 and the interval is (.0203, .0379). For a 99% interval, c=2.58 and the interval is (.0175, .0407). These intervals estimate the difference in proportions for men and women.
(c) An approximate, large sample, confidence interval for the proportion of men with byssinosis complaints is

$$.0439 \pm c \sqrt{(.0439)(.9561)/2916}$$

where c comes from the standard Gaussian distribution. Intervals with approximate confidence levels 90%, 95% and 99%, respectively, are (.0376, .0502), (.0365, .0513) and (.0341, .0537).
(d) An approximate, large sample, confidence interval for the proportion of women with byssinosis complaints is

$$.0148 \pm c \sqrt{(.0148)(.9852)/2503}$$

where c comes from the standard Gaussian distribution. Intervals with approximate confidence levels 90%, 95% and 99%, respectively, are (.0108, .0188), (.0101, .0195) and (.0086, .021).

Exercise 11-11
(a) The student should plot the observations in a way that allows comparisons of before and after surgery results for each group, as well as comparisons across groups. Plots should be on the same scale for easy visual comparisons.
(b) A two sample comparison tests the null hypothesis that average before surgery liver function is the same for the two treatment groups. The test statistic for the two sample t test is .64 with 19 degrees of freedom, with a p-value of .5. The test statistic for the Wilcoxon-Mann-Whitney test is 100, with a p-value of .4. The results are consistent with the null hypothesis that on average, before surgery liver function is the same for the two treatment groups.
(c) A paired sample comparison tests the null hypothesis that the mean difference between before and after surgery liver function is 0 for the new operation. The test statistic for the paired t test with 7 degrees of freedom is −.22, with a p-value of .8. The test statistic for the Wilcoxon signed rank test is 21 (or 15), with a p-value of .7. The test statistic for the sign test is 2, with a p-value of .3. These results are consistent with the null hypothesis of no mean difference in liver function before and after surgery for the new operation.

Using the t distribution with 7 degrees of freedom, a 95% confidence interval for the mean difference between before and after surgery liver function for the new operation is (−8.89, 7.39). Using the Wilcoxon signed rank distribution for sample size 8, a 96% confidence interval for this mean difference is (−8.5, 5.5). Using a binomial distribution for sample size 8, a 95% confidence interval for the median difference is (−12.5, 6).
(d) A paired sample comparison tests the null hypothesis that the mean difference between before and after surgery liver function is 0 for the standard operation. The test statistic for the paired t test with 12 degrees of freedom is 5.71, with a p-value less than .0001. The test statistic for the Wilcoxon signed rank test is 89.5 (or 1.5), with a p-value of .002. The test statistic for the sign test is 1, with a p-value of .003. These results are strongly inconsistent with the null hypothesis, suggesting there is a difference between before and after liver function for the standard operation.

Using the t distribution with 12 degrees of freedom, a 95% confidence interval for the mean (median) difference between before and after surgery liver function for the standard operation is (7.46, 16.69). Using the Wilcoxon signed rank distribution for sample size 13, a 95% confidence interval for this median difference is (7.0, 17.5). Using a binomial distribution for sample size 13, a 95% confidence interval for the median difference is (6.68, 18).
(e) A two sample comparison tests the null hypothesis that the mean change in liver function is the same for the two treatments. The test statistic is −3.37 for a two sample t test with 19 degrees of freedom, for a p-value of .003. The test statistic for the Wilcoxon-Mann-Whitney test is 46 (or 185),

with a p-value of .003. For the median test, the overall median of before–after values is 6. For the new operation, 6 values are below and none above this overall median. For the standard operation, 2 values are below and 10 above this overall median. The p-value is $\binom{6}{6}\binom{12}{2}/\binom{18}{8} = .002$. These results are inconsistent with the null hypothesis, suggesting that the mean change in liver function is not the same for the two operations.

Exercise 11-12
(a) Students should make plots on the same scale for easy visual comparison.
A two sample comparison tests the null hypothesis that the mean change in pupil diameter is the same for the two drugs.
(b) The test statistic is 2.65 for the two sample t test with 9 degrees of freedom, p-value=.03.
(c) The test statistic is 47.5 (or 18.5) for the Wilcoxon-Mann-Whitney test, p-value=.04.
(d) The overall median is .8. In the morphine group, 1 value is below and 4 above this overall median. In the nalbuphine group, 4 values are below and 0 above. The p-value is

$$\text{p-value} = \binom{5}{1}\binom{4}{4}/\binom{9}{5} + \binom{5}{5}\binom{4}{0}/\binom{9}{5} = .05.$$

(e) The three tests give similar results, inconsistent with the null hypothesis, suggesting that the mean change in pupil diameter differs for the two drugs.

Exercise 11-13
(a) The two plots look symmetrical.
(b) This is a two sample comparison. The two sample t test has test statistic .37, with 38 degrees of freedom, p-value=.7. The Wilcoxon-Mann-Whitney test has test statistic 424.5 (or 395.5), p-value=.7.
(c) Using the t distribution with 38 degrees of freedom, a 95% confidence interval for the difference in means for the new and old methods is (−2.9, 4.2). Using the Wilcoxon-Mann-Whitney distribution for sample sizes 20 and 20, a 95% confidence interval for the difference between the two means is (−3, 5).
(d) A 95% confidence interval for the new method mean is (103.2, 107.9) using the t distribution with 19 degrees of freedom, (103, 108) using the Wilcoxon signed rank distribution for sample size 20.
(e) A 95% confidence interval for the old method mean is (102.12, 107.68) using the t distribution with 19 degrees of freedom, (102, 107.5) using the Wilcoxon signed rank distribution for sample size 20.
(f) The two methods do not have significantly different results. Both methods get results surrounding the target value of 105 μg/100 ml.

Exercise 11-14
(b) Test statistic = 2.79 with 11 degrees of freedom, p-value = .018, inconsistent with the null hypothesis of no average change in PVC count before and after treatment. Confidence intervals for the average decrease in PVC count: 90%, (4.09, 18.91); 95%, (2.42, 20.58); 99% (−1.31, 24.31).
(c) Wilcoxon signed rank test statistic = 75.5 (or 2.5), p-value = .005, inconsistent with the null hypothesis of 0 median change in PVC count before and after treatment. Confidence intervals for the average decrease in PVC count: 89.2%, (5.0, 14.5); 94.5%, (4.0, 18.0); 98.9%, (1.5, 29.0).
(d) For the sign test, one difference is less than 0 and 11 are greater than 0, p-value = .0063, inconsistent with the null hypothesis of 0 median change in PVC count. Confidence intervals for the median decrease in PVC count: 90%, (4.8, 14.6); 95%, (4.26, 16.21); 99%, (2.17, 20.05).
(f) Repeat previous analyses without patient 12. For the paired t test, test statistic = 3.56 with 10 degrees of freedom, p-value = .0052; 95% confidence interval for average change is (2.96, 12.86). For the Wilcoxon signed rank test with sample size 11, test statistic = 63.5 (or 2.5), p-value = .008; 95.5% confidence interval for median change is (3.0, 13.5). For the sign test with sample size 11, 1 difference below 0 and 10 differences above 0, p-value = .01; 95% confidence interval for median change is (3.75, 14.25).

Exercise 11-15

(b) The null hypothesis states that the median difference between survival times for closely and poorly matched grafts is 0. For the sign test, 2 differences are below 0 and 9 differences are greater than 0, p-value = .065. (If you use the paired t test or the Wilcoxon signed rank test ignoring the fact that two values are censored, p-values are smaller.)

(c) Using a binomial distribution for sample size 11, a 93.5% confidence interval for the median difference between survival times for closely and poorly matched grafts is (5, 20).

(d) Using a binomial distribution for sample size 11, a 93.5% confidence interval for the median survival of closely matched grafts is (19, 60).

(e) Using a binomial distribution for sample size 11, a 93.5% confidence interval for the median survival of poorly match grafts is (15, 29).

Exercise 11-16

(c) The null hypothesis for this paired problem states that the median difference between before and after training scores is 0. For the paired t test, test statistic = −2.14 with 5 degrees of freedom, p-value = .085. For the Wilcoxon signed rank test, test statistic = 2.5, p-value = .116. For the sign test, 5 differences are below 0 and 1 difference is above 0, p-value = .2. These results are borderline; training may make a difference but this experiment is too small to suggest any clear effect.

(d) Confidence intervals for the median difference between before and after training scores are shown. Using the t distribution with 5 degrees of freedom: 90%, (−5.66, −.17); 95%, (−6.42, .59); 99%, (−8.41, 2.58). Using the Wilcoxon signed rank distribution for sample size 6: 90.7%, (−5.7, 0); 94.1%, (−6.5, .25); 96.4%, (−8, 1.5). Using a binomial distribution for sample size 6: 90%, (−5.77, −.358); 95%, (−6.929, .607); 96.87%, (−8, 1.5).

Exercise 11-17 The proportions with high antibody levels are 26/60 = .4333 for infected volunteers and 9/23 = .3913 for uninfected volunteers. The overall proportion is 35/83 = .4217. The large sample test statistic equals .35, p-value = .7, consistent with the null hypothesis of no difference between infected and uninfected volunteers with respect of antibody levels (high versus low). A 90% confidence interval for the difference in proportions with high antibody levels for infected and uninfected volunteers is (−.16, .24).

Exercise 11-18

(b) Using the paired t test, test statistic = 1.66 with 9 degrees of freedom, p-value = .13. Using the Wilcoxon signed rank test, test statistic = 34.5 (or 10.5), p-value = .173. These two tests are borderline, suggesting possible small decrease in shortening fraction at low dose enflurane. Using the sign test, 3 differences are below 0, 1 difference equals 0, 6 differences are above 0, p-value = .5, consistent with the null hypothesis of 0 median difference in shortening fraction before and at low dose enflurane.

(c) Using the paired t test, test statistic = 1.43 with 8 degrees of freedom, p-value = .19. Using the Wilcoxon signed rank test, test statistic = 32.5 (or 12.5), p-value = .26. Using the sign test, 3 differences are below 0, 6 differences are above 0, p-value = .5, consistent with the null hypothesis of 0 median difference in shortening fraction before and at low dose halothane.

(d) Using the two sample t test, test statistic = .24 with 17 degrees of freedom, p-value = .8. Using the Wilcoxon-Mann-Whitney test for sample sizes 10 and 9, test statistic = 103 (or 87), p-value = .8. These results are consistent with the null hypothesis that the median change in shortening fraction is the same for low dose enflurane and low dose halothane.

(e) Using the paired t test, test statistic = 4.22 with 9 degrees of freedom, p-value = .002. Using the Wilcoxon signed rank test, test statistic = 52 (or 3), p-value = .014. Using the sign test, 1 difference is below 0, 9 differences are above 0, p-value = .02. These results are inconsistent with the null hypothesis of 0 median difference in blood pressure before and at low dose enflurane, suggesting a decrease in blood pressure with low dose enflurane.

(f) Using the paired t test, test statistic = 4.12 with 8 degrees of freedom, p-value = .003. Using the Wilcoxon signed rank test, test statistic = 45 (or 0), p-value = .009. Using the sign test, 9 differences are above 0, none are below 0, p-value = .0039. These results are inconsistent with the null hypothesis of 0 median difference in blood pressure fraction before and at low dose halothane, suggesting a decrease in blood pressure with low dose halothane.
(g) Using the two sample t test, test statistic = 1.5 with 17 degrees of freedom, p-value = .1. Using the Wilcoxon-Mann-Whitney test for sample sizes 10 and 9, test statistic = 122 (or 68), p-value = .08. These results are borderline, suggesting that there may be somewhat greater decrease in blood pressure on low dose enflurane than on low dose halothane.

Exercise 11-19
(b) The null hypothesis states that the median difference in transport times before and after jogging equals 0. The alternative hypothesis states that this median difference is not 0.
(c) For the sign test, 10 differences are below 0 and 1 difference is above 0, p-value = .01, suggesting that transport times are longer after jogging. Confidence intervals for the median difference in transport times before and after jogging: 90%, (−22.55, −3.632); 95%, (−25.5, −2.836); 99%, (−25.575, −.925).
(d) Using the Wilcoxon signed rank test (with the >45 value taken to be 45), test statistic = 2.5, p-value = .01, suggesting longer transport times after jogging. Confidence intervals for the median difference in transport times before and after jogging: 90%, (−17.5, −4.25); 95.5%, (−18.5, −3.25); 98.9%, (−25.5, −1).
(e) Using the paired t test (with the >45 value taken to be 45), test statistic = −3.38 with 10 degrees of freedom, p-value = .007, suggesting longer transport times after jogging. Confidence intervals for the mean difference in transport times before and after jogging: 90%, (−17.67, −5.33); 95%, (−19.08, −3.92); 99%, (−22.29, −.71).
(h) These analyses exclude volunteer 3. Paired t test statistic = −2.95 with 9 degrees of freedom, p-value = .016; Wilcoxon signed rank test statistic = 2.5, p-value = .013; for sign test, 9 differences below 0 and 1 difference above 0, p-value = .02. Confidence intervals (95%) for the median difference in transport times before and after jogging: using t distribution with 9 degrees of freedom, (−17.86, −2.34); using Wilcoxon signed rank distribution, (−17.75, −2.75); using binomial distribution, (−20.896, −2.315).

Exercise 11-20
(b) Paired t test statistic = .17, p-value = .87; Wilcoxon signed rank test statistic = 21.5 (or 14.5), p-value = .7; for sign test, 3 differences are below 0 and 5 above, p-value = .7. These results are consistent with the null hypothesis of median difference between first and second measurements equal to 0.
(c) Confidence intervals for the median difference between first and second measurements are: using t distribution with 7 degrees of freedom (95%), (−3.15, 3.65); using Wilcoxon signed rank distribution (95.8%), (−3.75, 3.50); using binomial distribution (95%), (−2.386, 4.064).

Exercise 11-21 The large sample test statistic equals 5.15, p-value < .0001, inconsistent with the null hypothesis of equal mean calcium concentrations in men with and without crystals. Confidence intervals for the difference between the means for men with crystals and men without crystals: 90%, (2.4, 4.6); 95%, (2.2, 4.9); 99%, (1.8, 5.3). Men with crystals tend to have greater calcium concentrations, on average, than men without crystals.

Exercise 11-22 The null hypothesis states that the probability p that a judge selects sample A equals 1/2. Ignore the 14 no preference responses. We have $\hat{p} = 98/(98+88) = .527$. The large sample test statistic equals .7, with an approximate p-value of about .5. A large sample approximate 95% confidence interval for the proportion p of judges who would prefer A is (.455, .599).

Exercise 11-23 Using the original units, two sample t test statistic = 11.17 with 30 degrees of freedom, p-value < .0001; Wilcoxon-Mann-Whitney test statistic = 392 (or 136), p-value < .0001; median test statistic = 0, p-value < .0001. The mean measurement for the first sample is significantly larger than for the second sample. Taking logs does not change the results for the two nonparametric tests; t test statistic = 15 with 30 degrees of freedom, p-value < .0001.

Exercise 11-24
(a) Wilcoxon-Mann-Whitney test statistic = 7 (or 3), p-value = .2.
(b) Two sample t test statistic = 4.71 with 2 degrees of freedom, p-value = .04.
(c) Note that even with such small samples, the t test shows a significant difference since the actual distances pedaled were so different under the two conditions. The nonparametric procedures do not take the actual distances into account.

Exercise 11-25
(a) The Wilcoxon signed rank test statistic = 174 (or 36), p-value = .011. Confidence intervals for the median difference between self and rival competition times: 90.3%, (.065, .315); 95%, (.05, .335); 99.1%, (−.005, .415).
(b) Using the sign test, 6 differences are below 0 and 14 differences are above 0, p-value = .1, a borderline result. Confidence intervals for the median difference between self and rival competition times: 90%, (.026, .330); 95.86%, (−.01, .33); 99%, (−.041, .355).

Exercise 11-26 The proof follows the same steps as shown in the appendix on the Wilcoxon-Mann-Whitney distributions. Use the values for T_2 shown in Table 1 on page 635 of the text. Compare the distribution of T_2 with the distribution of T_1 shown in Table 2 on page 635 of the text.

Exercise 11-27 Shown below are the $\binom{6}{3} = 20$ rank combinations for group 1, group 2, value of W_1, value of $T_1=W_1-6$, value of W_2, value of $T_2=W_2-6$:
1 2 3, 4 5 6, 6, 0, 15, 9
1 2 4, 3 5 6, 7, 1, 14, 8
1 2 5, 3 4 6, 8, 2, 13, 7
1 2 6, 3 4 5, 9, 3, 12, 6
1 3 4, 2 5 6, 8, 2, 13, 7
1 3 5, 2 4 6, 9, 3, 12, 6
1 3 6, 2 4 5, 10, 4, 11, 5
1 4 5, 2 3 6, 10, 4, 11, 5
1 4 6, 2 3 5, 11, 5, 10, 4
1 5 6, 2 3 4, 12, 6, 9, 3
2 3 4, 1 5 6, 9, 3, 12, 6
2 3 5, 1 4 6, 10, 4, 11, 5
2 3 6, 1 4 5, 11, 5, 10, 4
2 4 5, 1 3 6, 11, 5, 10, 4
2 4 6, 1 3 5, 12, 6, 9, 3
2 5 6, 1 3 4, 13, 7, 8, 2
3 4 5, 1 2 6, 12, 6, 9, 3
3 4 6, 1 2 5, 13, 7, 8, 2
3 5 6, 1 2 4, 14, 8, 7, 1
4 5 6, 1 2 3, 15, 9, 6, 0

The Wilcoxon-Mann-Whitney distribution for sample sizes 3 and 3. Shown are values of c, $P(T_1=c)$ and $P(T_1 \leq c)$:

0, 1/20, 1/20
1, 1/20, 2/20
2, 2/20, 4/20
3, 3/20, 7/20
4, 3/20, 10/20
5, 3/20, 13/20
6, 3/20, 16/20
7, 2/20, 18/20
8, 1/20, 19/20
9, 1/20, 20/20=1

Exercise 11-28
(a) The overall median is between 21 and 21.5. For the morning, 2 values are below and 3 above this overall median. For the afternoon, 3 values are below and 2 above. The p-value is:

$$\text{p-value} = 2*\left(\binom{5}{2}\binom{5}{3}+\binom{5}{1}\binom{5}{4}+\binom{5}{0}\binom{5}{5}\right)/\binom{10}{5} = 2*(100+25+1)/252 = 1.$$

(b) Wilcoxon-Mann-Whitney test statistic = 34 (or 21), p-value = .2. The tests in (a) and (b) are consistent with the null hypothesis of no difference between median weights in morning and afternoon.

Exercise 11-29
(a) The overall median is 1.59. In the treatment group, 2 values are below and 0 above this overall median, while in the control group 2 values are below and 4 above. The p-value is:

$$\text{p-value} = 2*\binom{2}{2}\binom{6}{2}/\binom{8}{4} = .4.$$

(b) The two sample t test statistic = 1.86 with 7 degrees of freedom, p-value = .1.

Exercise 11-30
(a) Wilcoxon-Mann-Whitney test statistic = 87 (or 49), p-value = .052.
(b) Two sample t test statistic = 2.52 with 14 degrees of freedom, p-value = .025. The tests in (a) and (b) suggest that microwaving may increase average yield. These two test have smaller p-values than the median test in Example 11-5.

Chapter 12

Exercise 12-1
(b) For devices: degrees of freedom = 2, sum of squares = 92, mean square = 46. For residuals: degrees of freedom = 33, sum of squares = 12888, mean square = 391. To test the null hypothesis of no difference between devices in mean mosquito response, test statistic = .12 with 2 and 33 degrees of freedom, p-value = .9, consistent with the null hypothesis.
(c) H = .06, p-value = .97, consistent with the null hypothesis of no device differences.

Exercise 12-2
(b) First analysis uses original units. For treatments: degrees of freedom = 3, sum of squares = 216, mean square = 72. For residuals: degrees of freedom = 8, sum of squares = 800, mean square = 100. To test the null hypothesis of no difference between treatments in mean stimulation index, test statistic = .72 with 3 and 8 degrees of freedom, p-value = .6, consistent with the null hypothesis. Separate 99%

confidence intervals for differences between group means, based on the t distribution with 8 degrees of freedom (c=3.355), for overall confidence level at least 94%:
group1 − group2 (−31.8, 23.0)
group1 − group3 (−36.3, 24.9)
group1 − group4 (−36.6, 14.6)
group2 − group3 (−32.0, 29.3)
group2 − group4 (−32.3, 19.0)
group3 − group4 (−34.4, 23.8)
(c) Second analysis uses logarithm of stimulation index as response. For treatments: degrees of freedom = 3, sum of squares = .656, mean square = .219. For residuals: degrees of freedom = 8, sum of squares = 2.18, mean square = .272. To test the null hypothesis of no differences between treatments in mean response, test statistic = .8, p-value = .5, consistent with the null hypothesis. Separate 99% confidence intervals for differences between group means (in log units):
group1 − group2 (−1.9, 1.0)
group1 − group3 (−2.2, .96)
group1 − group4 (−1.9, .81)
group2 − group3 (−1.8, 1.4)
group2 − group4 (−1.4, 1.2)
group3 − group4 (−1.4, 1.6)
(d) Using the original units of stimulation index, test statistic = 2.7, p-value = .4, consistent with the null hypothesis of no treatment differences. Separate confidence intervals for differences between group medians, with individual confidence levels shown:
group1 − group2, 92% (−11.40, 2.50)
group1 − group3, 85% (−8.10, −2.599)
group1 − group4, 95% (−34.11, 3.21)
group2 − group3, 85% (−7.60, 5.80)
group2 − group4, 95% (−33.60, 11.60)
group3 − group4, 90% (−28.49, 8.31)

Exercise 12-3
(b) For diets: degrees of freedom = 3, sum of squares = 4.232, mean square = 1.411. For residuals, degrees of freedom = 32, sum of squares = 14.057, mean square = .439. To test the null hypothesis of no difference between diets in mean expired nitrogen, test statistic = 3.21, p-value = .036, inconsistent with the null hypothesis. Separate 99% confidence intervals for differences between group means, based on the t distribution with 32 degrees of freedom (c=2.7385), for overall confidence level at least 94%:
group1 − group2 (−1.4, .33)
group1 − group3 (−1.5, .20)
group1 − group4 (−1.8, −.09) Note that 0 is not in this interval.
group2 − group3 (−.98, .73)
group2 − group4 (−1.3, .44)
group3 − group4 (−1.2, .56)
(c) Test statistic = 7.6, p-value = .055, borderline but suggesting possible differences between diets with respect to mean expired nitrogen. Separate confidence intervals for differences between group medians, with each individual confidence level equal to 99.2%:
group1 − group2 (−1.853, .877)
group1 − group3 (−1.761, .479)
group1 − group4 (−1.936, −.046) Note that 0 is not in this interval.
group2 − group3 (−1.298, 1.224)
group2 − group4 (−1.362, .728)
group3 − group4 (−.9149, .6711)

Exercise 12-4

(c) For emotions: degrees of freedom = 3, sum of squares = 101.43, mean square = 33.81. For volunteers(blocks): degrees of freedom = 7, sum of squares = 7021.86, mean square = 1003.12. For residuals, degrees of freedom = 21, sum of squares = 204.80, mean square = 9.75. To test the null hypothesis of no difference between emotions in mean response, test statistic = 3.47 with 3 and 21 degrees of freedom, p-value = .03, suggesting there are differences between emotions with respect to mean skin potential. To test the null hypothesis of no difference between volunteers in mean response, test statistic = 102.88 with 7 and 21 degrees of freedom, p-value < .0001, inconsistent with the null hypothesis.

(d) Test statistic = 6.45, p-value = .09, borderline but suggesting possible differences between emotions with respect to skin potential.

Exercise 12-5

(b) For flour types: degrees of freedom = 2, sum of squares = 8.6717, mean square = 4.3358. For residuals, degrees of freedom = 9, sum of squares = .6150, mean square = .0683. To test the null hypothesis of no difference between flour types in mean volume increase, test statistic = 63.45 with 2 and 9 degrees of freedom, p-value < .0001, inconsistent with the null hypothesis. Separate 99% confidence intervals based on the t distribution with 9 degrees of freedom (c=3.25) are (−2.25, −1.05) for $\mu_1-\mu_2$, (−2.53, −1.32) for $\mu_1-\mu_3$, (−.876, .326) for $\mu_2-\mu_3$ (overall confidence level at least 97%). The mean volume increase for flour type 1 seems to be less than the mean for flour type 2 and the mean for flour type 3; we cannot distinguish between flour types 2 and 3.

(c) Test statistic = 8.2, p-value = .02, inconsistent with the null hypothesis of no difference between flour types with respect to volume increase. Separate 97% confidence intervals are (−2.3, −.9) for $\mu_1-\mu_2$, (−2.3, −1.3) for $\mu_1-\mu_3$, (−.6, .2) for $\mu_2-\mu_3$ (overall confidence level at least 91%). Conclusions are the same as for part (b).

Exercise 12-6

(b) For types: degrees of freedom = 2, sum of squares = 447, mean square = 223. For residuals, degrees of freedom = 17, sum of squares = 7632, mean square = 449. To test the null hypothesis of no difference between types in mean survival, test statistic = .5, p-value = .6, consistent with the null hypothesis. Separate 99% confidence intervals based on the t distribution with 17 degrees of freedom (c=2.898) are (−43.4, 21.3) for $\mu_1-\mu_2$, (−37.1, 31.4) for $\mu_1-\mu_3$, (−29.0, 45.3) for $\mu_2-\mu_3$ (overall confidence level at least 97%).

(c) Test statistic = 2.15, p-value = .3, consistent with the null hypothesis of no difference between stopwatch types with respect to survival. Separate 99.2% confidence intervals are (−44.60, 36.29) for $\mu_1-\mu_2$, (−40.79, 57.41) for $\mu_1-\mu_3$, (−33.31, 47.69) for $\mu_2-\mu_3$ (overall confidence level at least 97%).

Exercise 12-7

(c) For methods: degrees of freedom = 2, sum of squares = .09371, mean square = .04686. For runners, degrees of freedom = 21, sum of squares = 4.21864, mean square = .20089. For residuals, degrees of freedom = 42, sum of squares = .31295, mean square = .00745. To test the null hypothesis of no difference between methods in mean time, test statistic = 6.29 with 2 and 42 degrees of freedom, p-value = .004, inconsistent with the null hypothesis. To test the null hypothesis of no difference between runners in mean time, test statistic = 26.97 with 21 and 42 degrees of freedom, p-value < .0001, inconsistent with the null hypothesis.

(d) Test statistic = 10.64, p-value = .005, inconsistent with the null hypothesis of no difference between methods with respect to running time.

Exercise 12-8

(c) For drying method: degrees of freedom = 4, sum of squares = 14.962, mean square = 3.741. For fabrics: degrees of freedom = 8, sum of squares = 9.696, mean square = 1.212. For residuals: degrees of freedom = 32, sum of squares = 3.262, mean square = .102. To test the null hypothesis of no difference between drying methods in mean smoothness, test statistic = 36.7 with 4 and 32 degrees of freedom, p-value < .0001, inconsistent with the null hypothesis. To test the null hypothesis of no difference between fabrics in mean smoothness, test statistic = 11.9 with 8 and 32 degrees of freedom, p-value < .0001, inconsistent with the null hypothesis.

(d) Friedman's test statistic = 28.2, p-value < .0001, inconsistent with the null hypothesis of no difference between drying methods with respect to smoothness.

Exercise 12-9

(b) For treatments: degrees of freedom = 3, sum of squares = 2278, mean square = 759. For residuals: degrees of freedom = 36, sum of squares = 13812, mean square = 384. To test the null hypothesis of no difference between treatments in mean biomass/m^2, test statistic = 1.98 with 3 and 36 degrees of freedom, p-value = .135. Separate 99% confidence intervals for differences between group means based on the t distribution with 36 degrees of freedom (c=2.72) are shown below (overall confidence level at least 94%):

group1 – group2 (–29.9, 17.8)
group1 – group3 (–38.7, 9.0)
group1 – group4 (–43.2, 4.4)
group2 – group3 (–32.6, 15.0)
group2 – group4 (–37.2, 10.5)
group3 – group4 (–28.4, 19.3)

(c) Test statistic = 5.04, p-value = .17, consistent with the null hypothesis of no difference between treatments with respect to biomass/m^2. Separate 99.1% confidence intervals for differences between group means:

group1 – group2 (–33.18, 16.55)
group1 – group3 (–47.01, 9.92)
group1 – group4 (–43.67, 4.46)
group2 – group3 (–43.74, 21.19)
group2 – group4 (–44.01, 14.59)
group3 – group4 (–38.19, 28.90)

Exercise 12-10

(c) For manufacturing method: degrees of freedom = 3, sum of squares = 70, mean square = 23.3. For blend: degrees of freedom = 4, sum of squares = 264, mean square = 66.0. For residuals: degrees of freedom = 12, sum of squares = 226, mean square = 18.8. To test the null hypothesis of no difference between manufacturing methods in mean yield, test statistic = 1.24 with 3 and 12 degrees of freedom, p-value=.34, consistent with the null hypothesis. To test the null hypothesis of no difference between blends(blocks) in mean yield, test statistic = 3.51 with 4 and 12 degrees of freedom, p-value=.04, inconsistent with the null hypothesis.

(e) Friedman's test statistic = 3.4, p-value=.3, consistent with the null hypothesis of no difference between manufacturing methods with respect to yield.

Exercise 12-11

(b) For species: degrees of freedom = 2, sum of squares = 836131, mean square = 418066. For residuals, degrees of freedom = 131, sum of squares = 446758, mean square = 3410. To test the null hypothesis of no difference between species in mean number of squares crossed, test statistic = 122.59, p-value < .0001, inconsistent with the null hypothesis. Separate 99% confidence intervals based on the t distribution with 131 degrees of freedom (c=2.61) are (147.63, 206.85) for $\mu_1 - \mu_2$, (65.78, 137) for

$\mu_1-\mu_3$, (−111.46, −40.24) for $\mu_2-\mu_3$ (overall confidence level at least 97%). The mean for species 1 is greater than the mean for species 3, which is greater than the mean for species 2.
(c) Kruskal-Wallis test statistic = 87.08, p-value < .0001, inconsistent with the null hypothesis of no difference between species with respect to mean number of squares crossed. Separate 99% confidence intervals are (150.99, 209) for $\mu_1-\mu_2$, (58, 147.99) for $\mu_1-\mu_3$, (−110.98, −40.99) for $\mu_2-\mu_3$ (overall confidence level at least 97%).

Exercise 12-12
(c) This analysis includes all volunteers. For electrodes: degrees of freedom = 4, sum of squares = 281575, mean square = 70394. For volunteers: degrees of freedom = 15, sum of squares = 1399155, mean square = 93277. For residuals: degrees of freedom = 60, sum of squares = 1342723, mean square = 22379. To test the null hypothesis of no difference between electrodes in mean resistance, test statistic = 3.15 with 4 and 60 degrees of freedom, p-value = .02, inconsistent with the null hypothesis. To test the null hypothesis of no difference between volunteers in mean resistance, test statistic = 4.17 with 15 and 60 degrees of freedom, p-value < .0001, inconsistent with the null hypothesis.
(d) Including all volunteers, Friedman's test statistic = 5.35, p-value = .25, consistent with the null hypothesis of no difference between electrodes with respect to resistance.
(e) These analyses exclude volunteer 15. Parametric analysis, for electrodes: degrees of freedom = 4, sum of squares = 120225, mean square = 30056. For volunteers, degrees of freedom = 14, sum of squares = 852510, mean square = 60894. For residuals, degrees of freedom = 56, sum of squares = 639109, mean square = 11413. To test the null hypothesis of no difference between electrodes in mean resistance, test statistic = 2.63 with 4 and 56 degrees of freedom, p-value = .04, inconsistent with the null hypothesis. To test the null hypothesis of no difference between volunteers in mean resistance, test statistic = 5.34 with 14 and 56 degrees of freedom, p-value < .0001, inconsistent with the null hypothesis. Friedman's test statistic = 4.51, p-value = .3, consistent with the null hypothesis of no difference between electrodes with respect to resistance.

Exercise 12-13
(b) For treatments: degrees of freedom = 2, sum of squares = 52.78, mean square = 26.39. For residuals, degrees of freedom = 15, sum of squares = 145.00, mean square = 9.67. To test the null hypothesis of no difference between treatments in mean performance, test statistic = 2.73, p-value = .097, borderline but suggesting possible effect of information feedback. Separate 99% confidence intervals based on the t distribution with 15 degrees of freedom (c=2.947) are (−6.96, 3.62) for $\mu_1-\mu_2$, (−9.46, 1.12) for $\mu_1-\mu_3$, (−7.79, 2.79) for $\mu_2-\mu_3$ (overall confidence level at least 97%). The experiment is not large enough to reveal possible differences.
(c) Kruskal-Wallis test statistic = 4.29, p-value = .1, again borderline. Separate 99.2% confidence intervals are (−9, 4.998) for $\mu_1-\mu_2$, (−11.002, 3.001) for $\mu_1-\mu_3$, (−8.001, 3.999) for $\mu_2-\mu_3$ (overall confidence level at least 97%).

Exercise 12-14
(b) No, this assumption needed for (c) and (d) is violated.
(c) For treatments: degrees of freedom = 2, sum of squares = 46857, mean square = 23429. For patients, degrees of freedom = 11, sum of squares = 92801, mean square = 8436. For residuals, degrees of freedom = 22, sum of squares = 179515, mean square = 8160. To test the null hypothesis of no difference between treatments in mean response, test statistic = 2.87 with 2 and 22 degrees of freedom, p-value = .078, borderline but suggesting possible differences between treatments. To test the null hypothesis of no difference between patients in mean response, test statistic = 1.03 with 11 and 22 degrees of freedom, p-value = .45, consistent with the null hypothesis.
(d) Friedman's test statistic = 7.88, p-value = .02, inconsistent with the null hypothesis of no difference between treatments with respect to number of premature ventricular contractions per hour.

Exercise 12-15

(b) For pH: degrees of freedom = 2, sum of squares = 1.0824, mean square = .5412. For residuals, degrees of freedom = 8, sum of squares = .2467, mean square = .0308. To test the null hypothesis of no difference between pH levels in mean height, test statistic = 17.55, p-value = .001, inconsistent with the null hypothesis. Separate 99% confidence intervals based on the t distribution with 8 degrees of freedom (c=3.355) are (−.72, .18) for $\mu_1 - \mu_2$, (−1.22, −.32) for $\mu_1 - \mu_3$, (−.92, −.08) for $\mu_2 - \mu_3$ (overall confidence level at least 97%). Average height seems to increase with pH.

(c) Kruskal-Wallis test statistic = 8.07, p-value = .018, inconsistent with the null hypothesis of no difference between pH treatments in median height. Separate confidence intervals are (−.6001, .0999) for $\mu_1 - \mu_2$ (94.8%), (−1.2000, −.4000) for $\mu_1 - \mu_3$ (94.8%), (−.9001, −.2001) for $\mu_2 - \mu_3$ (97%), overall confidence level at least 87%).

Exercise 12-16 Let E be the event that at least one of the m criteria is satisfied, and α=P(E). Let E_i be the event that criterion i is satisfied, and $\alpha_i = P(E_i)$. Then E is the union of E_1 through E_m and the result follows from the Chapter 6 hint.

Exercise 12-17 Listed are, respectively, ranks for groups 1, 2 and 3; sum of group ranks R_1, R_2 and R_3; and the Kruskal-Wallis test statistic KW:

1, 2 3, 4 5; 1, 5, 9; 3.6
1, 4 5, 2 3; 1, 9, 5; 3.6
1, 2 4, 3 5; 1, 6, 8; 2.4
1, 3 5, 2 4; 1, 8, 6; 2.4
1, 2 5, 3 4; 1, 7, 7; 2
1, 3 4, 2 5; 1, 7, 7; 2
2, 1 3, 4 5; 2, 4, 9; 3
2, 4 5, 1 3; 2, 9, 4; 3
2, 1 4, 3 5; 2, 5, 8; 1.4
2, 3 5, 1 4; 2, 8, 5; 1.4
2, 1 5, 3 4; 2, 6, 7; 0.6
2, 3 4, 1 5; 2, 7, 6; 0.6
3, 1 2, 4 5; 3, 3, 9; 3.6
3, 4 5, 1 2; 3, 9, 3; 3.6
3, 1 4, 2 5; 3, 5, 7; 0.4
3, 2 5, 1 4; 3, 7, 5; 0.4
3, 1 5, 2 4; 3, 6, 6; 0
3, 2 4, 1 5; 3, 6, 6; 0
4, 1 2, 3 5; 4, 3, 8; 3
4, 3 5, 1 2; 4, 8, 3; 3
4, 1 3, 2 5; 4, 4, 7; 1.4
4, 2 5, 1 3; 4, 7, 4; 1.4
4, 1 5, 2 3; 4, 6, 5; 0.6
4, 2 3, 1 5; 4, 5, 6; 0.6
5, 1 2, 3 4; 5, 3, 7; 3.6
5, 3 4, 1 2; 5, 7, 3; 3.6
5, 1 3, 2 4; 5, 4, 6; 2.4
5, 2 4, 1 3; 5, 6, 4; 2.4
5, 1 4, 2 3; 5, 5, 5; 2
5, 2 3, 1 4; 5, 5, 5; 2

The Kruskal-Wallis distribution for sample sizes 1, 2 and 2:
P(KW = 0) = 2/30, P(KW $\geq$ 0) = 30/30 = 1
P(KW = .4) = 2/30, P(KW $\geq$ .4) = 28/30
P(KW = .6) = 4/30, P(KW $\geq$ .6) = 26/30
P(KW = 1.4) = 4/30, P(KW $\geq$ 1.4) = 22/30
P(KW = 2) = 4/30, P(KW $\geq$ 2) = 18/30
P(KW = 2.4) = 4/30, P(KW $\geq$ 2.4) = 14/30
P(KW = 3) = 4/30, P(KW $\geq$ 3) = 10/30
P(KW = 3.6) = 6/30, P(KW $\geq$ 3.6) = 6/30

Exercise 12-18 For thermometers: degrees of freedom = 3, sum of squares = 4.417, mean square = 1.472. For residuals: degrees of freedom = 8, sum of squares = 6.500, mean square = .812. To test the null hypothesis of no difference between thermometers in mean measurement, test statistic = 1.81 with 3 and 8 degrees of freedom, p-value = .2, consistent with the null hypothesis.

Exercise 12-19 Friedman's test statistic = 6.10, p-value = .1, borderline and similar to the parametric result in Example 12-3.

Exercises 12-20 and 12-21 Refer to the assumptions for procedures discussed in the text.

Chapter 13

Exercise 13-1
(c) Listed are elements of the analysis of variance table for, respectively: brand, material, interaction, residual. Degrees of freedom: 1, 1, 1, 12. Sum of squares: 3136, 5184, 49, 950. Mean square: 3136, 5184, 49, 79.2. To test the null hypothesis of no difference in mean distance for the two brands, test statistic = 39.6 with 1 and 12 degrees of freedom, p-value < .0001. To test the null hypothesis of no difference in mean distance for the two bat materials, test statistic = 65.45 with 1 and 12 degrees of freedom, p-value < .0001. To test the null hypothesis that there is no nonadditive effect of ball brand and bat material on mean distance, test statistic = .62 with 1 and 12 degrees of freedom, p-value=.4. Average distances are greater with the Worth Red Dot ball and with the aluminum bat. There appear to be no nonadditive effects of ball brand and bat material.

Exercise 13-2
(c) For the analysis using original units, listed are elements of the analysis of variance table for, respectively: pH, temperature, interaction, residual. Degrees of freedom: 1, 1, 1, 4. Sum of squares: 366, 1582, 694, 729. Mean square: 366, 1582, 694, 182. To test the null hypothesis of no difference in mean optical density for the two levels of pH, test statistic = 2.01 with 1 and 4 degrees of freedom, p-value = .2. To test the null hypothesis of no difference in mean optical density for the two levels of temperature, test statistic = 8.69 with 1 and 4 degrees of freedom, p-value = .04. To test the null hypothesis that there is no nonadditive effect of pH and temperature on mean optical density, test statistic = 3.81 with 1 and 4 degrees of freedom, p-value = .1.
(f) For the analysis using reciprocal of optical density as response, listed are elements of the analysis of variance table for, respectively: pH, temperature, interaction, residual. Degrees of freedom: 1, 1, 1, 4. Sum of squares: .0000105, .0002318, .0000794, .0000482. Mean square: .0000105, .0002318, .0000794, .0000121. To test the null hypothesis of no difference in mean response for the two levels of pH, test statistic = .87 with 1 and 4 degrees of freedom, p-value = .4. To test the null hypothesis of no difference in mean response across the two temperatures, test statistic = 19.16 with 1 and 4 degrees of freedom, p-value = .01. To test the null hypothesis that there is no nonadditive effect of pH and temperature on mean response, test statistic = 6.56 with 1 and 4 degrees of freedom, p-value = .06.

Exercise 13-3
(c) Listed are elements of the analysis of variance table for, respectively: niacin enrichment level, laboratory, interaction, residual. Degrees of freedom: 2, 5, 10, 36−2=34 (use 34 to get residual mean square). Sum of squares: 626.664, 13.339, 2.062, 6.168. Mean square: 313.332, 2.668, .206, .181. To test the null hypothesis of no difference in mean measurement across the three niacin enrichment levels, test statistic = 1731 with 2 and 34 degrees of freedom, p-value $<$ .0001. To test the null hypothesis of no difference in mean measurement across the six laboratories, test statistic = 14.7 with 5 and 34 degrees of freedom, p-value $<$ .0001. To test the null hypothesis that there is no nonadditive effect of niacin enrichment level and laboratory on mean response, test statistic = 1.14 with 10 and 34 degrees of freedom, p-value = .36. Mean niacin measurements depend on the niacin enrichment level and on the laboratory. There does not appear to be any two way interaction effect on response.

Exercise 13-4
(c) Listed are elements of the analysis of variance table for, respectively: time, temperature, interaction, residual. Degrees of freedom: 1, 1, 1, 16. Sum of squares: 1540125, 10585125, 2145125, 739000. Mean square: 1540125, 10585125, 2145125, 46188. To test the null hypothesis of no difference between times on mean strength, test statistic = 33.34 with 1 and 16 degrees of freedom, p-value $<$ .0001. To test the null hypothesis of no difference between temperatures on mean strength, test statistic = 229.2 with 1 and 16 degrees of freedom, p-value $<$ .0001. To test the null hypothesis of no nonadditive effect of time and temperature on mean strength, test statistic = 46.4 with 1 and 16 degrees of freedom, p-value $<$ .0001. There appear to be significant effects of time and temperature and significant nonadditive effects of time and temperature on mean strength.

Exercise 13-5
(c) Using the original units, listed are elements of the analysis of variance table for, respectively: form, concentration, interaction, residual. Degrees of freedom: 1, 2, 2, 102. Sum of squares: 62.3, 983.6, 8.3, 2938.2. Mean square: 62.3, 491.8, 4.1, 28.8. To test the null hypothesis of no difference between forms on mean percentage retained, test statistic = 2.16 with 1 and 102 degrees of freedom, p-value = .14. To test the null hypothesis of no difference between concentrations on mean percentage retained, test statistic = 17.08 with 2 and 102 degrees of freedom, p-value $<$ .0001. To test the null hypothesis of no nonadditive effect of form and concentration on mean percentage retained, test statistic = .14 with 2 and 102 degrees of freedom, p-value = .87.
(e) Using the logarithm of percentage iron retained as the response, listed are elements of the analysis of variance table for, respectively: form, concentration, interaction, residual. Degrees of freedom: 1, 2, 2, 102. Sum of squares: .3912, 2.9402, .1528, 6.6572. Mean square: .3912, 1.4701, .0764, .0653. To test the null hypothesis of no difference between forms on mean response, test statistic = 5.99 with 1 and 102 degrees of freedom, p-value = .016. To test the null hypothesis of no difference between concentrations on mean response, test statistic = 22.51 with 2 and 102 degrees of freedom, p-value $<$ .0001. To test the null hypothesis of no nonadditive effect of form and concentration on mean response, test statistic = 1.17 with 2 and 102 degrees of freedom, p-value = .3.
(f) Using the original units, only the concentration effect appears significant. Using the logarithm of percentage retained, form also appears significant, although less so than concentration. Neither analysis indicates any two way interaction effect of form and concentration on response.

Exercise 13-6 For each analysis, elements of the analysis of variance table are listed in this order: activity, time, interaction, residual. Degrees of freedom in each case are: 1, 1, 1, 4. Each test statistic has 1 and 4 degrees of freedom.
(c) Analysis of logarithm of change in fecal coliform concentration for females. Sum of squares: .0289, .0654, 1.9697, .2172. Mean square: .0289, .0654, 1.9697, .0543. To test the null hypothesis of no difference between activity levels on mean response, test statistic = .53, p-value = .5. To test the

null hypothesis of no difference between times on mean response, test statistic = 1.20, p-value = .33. To test the null hypothesis of no nonadditive effect of activity and time on mean response, test statistic = 36.3, p-value = .004.

(e) Analysis of logarithm of change in fecal coliform concentration for males. Sum of squares: .0087, .0202, .0556, .2653. Mean square: .0087, .0202, .0556, .0663. To test the null hypothesis of no difference between activity levels on mean response, test statistic = .13, p-value = .7. To test the null hypothesis of no difference between times on mean response, test statistic = .3, p-value = .6. To test the null hypothesis of no nonadditive effect of activity and time on mean response, test statistic = .84, p-value = .4.

(i) Analysis of logarithm of change in total coliform concentration for females. Sum of squares: .324, .481, 1.524, 1.663. Mean square: .324, .481, 1.524, .416. To test the null hypothesis of no difference between activity levels on mean response, test statistic = .78, p-value = .4. To test the null hypothesis of no difference between times on mean response, test statistic = 1.16, p-value = .3. To test the null hypothesis of no nonadditive effect of activity and time on mean response, test statistic = 3.66, p-value = .1.

(k) Analysis of logarithm of change in total coliform concentration for males. Sum of squares: .094, .085, .063, .405. Mean square: .094, .085, .063, .101. To test the null hypothesis of no difference between activity levels on mean response, test statistic = .93, p-value = .4. To test the null hypothesis of no difference between times on mean response, test statistic = .84, p-value = .4. To test the null hypothesis of no nonadditive effect of activity and time on mean response, test statistic = .62, p-value = .5.

Exercise 13-7 For each analysis, elements of the analysis of variance table are listed in this order: treatment, poison, interaction, residual. Degrees of freedom in each case are: 3, 2, 6, 36.

(c) Using the original units, sum of squares: 91.90, 103.04, 24.75, 78.69. Mean square: 30.63, 51.52, 4.12, 2.19, To test the null hypothesis of no difference in mean survival across treatments, test statistic = 13.99 with 3 and 36 degrees of freedom, p-value < .0001. To test the null hypothesis of no difference in mean survival across poisons, test statistic = 23.53 with 2 and 36 degrees of freedom, p-value < .0001. To test the null hypothesis of no nonadditive effect of poison and treatment on mean survival, test statistic = 1.88 with 6 and 36 degrees of freedom, p-value = .11.

(e) Using the reciprocal of survival time (or death rate) as response, sum of squares: .20396, .34863, .01567, .08615. Mean square: .06799, .17432, .00261, .00239. To test the null hypothesis of no difference in mean death rate across treatments, test statistic = 28.45 with 3 and 36 degrees of freedom, p-value < .0001. To test the null hypothesis of no difference in mean death rate across poisons, test statistic = 72.94 with 2 and 36 degrees of freedom, p-value < .0001. To test the null hypothesis of no nonadditive effect of poison and treatment on mean death rate, test statistic = 1.09, p-value = 4.

Exercise 13-8

(c) Listed are elements of the analysis of variance table for, respectively: technician, machine, interaction, residual. Degrees of freedom: 3, 4, 12, 20. Sum of squares: 753.67, 31.40, 77.20, 94.50. Mean square: 251.22, 7.85, 6.43, 4.72. To test the null hypothesis of no technician differences on mean output, test statistic = 53.22 with 3 and 20 degrees of freedom, p-value < .0001. To test the null hypothesis of no machine difference on mean output, test statistic = 1.66 with 4 and 20 degrees of freedom, p-value = .2. To test the null hypothesis of no nonadditive effect of technician and machine on mean output, test statistic = 1.36 with 12 and 20 degrees of freedom, p-value = .3. The only significant effect on mean output or productivity is accounted for by differences between technicians.

Chapter 14

Exercise 14-1
(b) The sample variances are 31.07 $hour^2$ for low stress and 5.58 $hour^2$ for high stress. The null hypothesis states that the two population variances are equal; a one sided alternative states that the variance is greater at the low than at the high stress level. Test statistic = 5.57 with 5 and 5 degrees of freedom, p-value=.04, inconsistent with the null hypothesis, suggesting greater variance at the low stress level.
(c) A 95% confidence interval for the ratio of the two population variances (low stress level variance divided by high stress level variance) is from .78 to 39.79.

Exercise 14-2
(b) The sample variance equals 92,357.1 $gram^2$. The null hypothesis states that the population variance equals 10,000 $grams^2$, with the two sided alternative. Test statistic = 46.2 with 5 degrees of freedom, p-value $< .0001$, inconsistent with the null hypothesis.
(c) A 99% confidence interval for the population variance is 27,570 to 1,121,542 $gram^2$.

Exercise 14-3
(b) The sample variance equals .0000313 $inch^2$. The test statistic = 1.252 with 4 degrees of freedom, p-value = .3 = 2P(test statistic $\leq$ 1.252 when H_0 is true), consistent with the null hypothesis that the population variance equals .0001 $inch^2$.
(c) A 98% confidence interval for the population variance is from .00000943 to .000421 $inch^2$.

Exercise 14-4
(b) The sample variance equals 25.2 $(\mu g/100ml)^2$. Test statistic = 4.79 with 19 degrees of freedom, p-value = .0008 = 2P(test statistic $\leq$ 4.79 when H_0 is true), inconsistent with the null hypothesis that the population variance equals 100 $(\mu g/100ml)^2$.
(c) A 95% confidence interval for the population variance is 14.58 to 53.78 $(\mu g/100ml)^2$.

Exercise 14-5
(b) Sample variance equals 432.9 $(\mu mole/liter)^2$. Test statistic = 2.6 with 9 degrees of freedom, p-value = .04 = 2P(test statistic $\leq$ 2.6 when H_0 is true), inconsistent with the null hypothesis that the population variance equals 1500 $(\mu mol/liter)^2$.
(c) A 90% confidence interval for the population variance is 230.3 to 1171.8 $(\mu mol/liter)^2$.

Exercise 14-6
(b) Sample variance equals .77 mm^2 for morphine and .17 mm^2 for nalbuphine. The null hypothesis states that the two population variances are equal; a one sided alternative states that the variance is greater for morphine than for nalbuphine. Test statistic = 4.49 with 5 and 4 degrees of freedom, p-value = .085, borderline but suggesting possible greater variance with morphine.
(c) A 90% confidence interval for the ratio of the two population variances (morphine variance divided by nalbuphine variance) is .72 to 23.33.

Exercise 14-7
(b) The sample variances are 159.5 $units^2$ for hand and 102.1 for automatic. The null hypothesis states that the two population variances are equal; a one sided alternative states that the variance is greater using hand administration than using an automatic device. Test statistic = 1.56 with 11 and 11 degrees of freedom, p-value = .2, consistent with the null hypothesis.
(c) A 98% confidence interval for the ratio of the two population variances (hand variance divided by automatic variance) is .35 to 6.97.

Exercise 14-8
(b) The sample variances are 354,507 unit^2 for allergics and 250.3 for nonallergics. Test statistic = 1416 with 8 and 12 degrees of freedom, p-value $< .0001$, inconsistent with the null hypothesis that the two population variances are equal; variance is greater for allergics than for nonallergics.
(c) A 98% confidence interval for the ratio of the two population variances (allergics variance divided by nonallergics variance) is 314.8 to 8026.0.
(d) Using the logarithm of response, test statistic = 2.73 with 8 and 12 degrees of freedom, p-value = .06, borderline, with somewhat greater variance for the allergics than for nonallergics.
(e) Using the logarithm of response, a 98% confidence interval for the ratio of the two population variances (allergics variance divided by nonallergics variance) is .6 to 15.5.

Exercise 14-9
(b) The sample variances are 14,829 units^2 for alcohol and 145,497 for control. Test statistic = 9.8 with 9 and 9 degrees of freedom, p-value = .001, inconsistent with the null hypothesis that the two population variances are equal; there is greater variation for control monkeys.
(c) A 90% confidence interval for the ratio of the two population variances (control variance divided by alcohol variance) is 3.1 to 31.2.
(d) Using the logarithm of response, test statistic = 2.999 with 9 and 9 degrees of freedom, p-value = .059, still suggesting greater variance for the control monkeys.
(e) Using the logarithm of response, a 90% confidence interval for the ratio of the two population variances (control variance divided by alcohol variance) is .94 to 9.53.

Exercise 14-10
(b) The sample variances are .625 unit^2 for Analyst 1 and .696 for Analyst 2. Test statistic = 1.1 with 7 and 7 degrees of freedom, p-value = .4, consistent with the null hypothesis of equal population variances.
(c) A 95% confidence interval for the ratio of the two population variances (analyst 1 variance divided by analyst 2 variance) is .22 to 5.57.

Exercise 14-11
(b) The sample variances are 8.84 unit^2 for Magnum and 6.50 for Long Rifle. Test statistic = 1.36 with 7 and 7 degrees of freedom, p-value = .3, consistent with the null hypothesis of equal population variances.
(c) A 95% confidence interval for the ratio of the two population variances (.22 Magnum variance divided by .22 Long Rifle variance) is .27 to 6.79.

Exercise 14-12
(b) Bartlett's test statistic=1.8 with 5 degrees of freedom, p-value=.9, consistent with the null hypothesis of equal population variances.
(c) Levene's test statistic=.41, p-value=.8, again consistent with the equal population variance null hypothesis.

Exercise 14-13
(b) Bartlett's test statistic = 24.5 with 3 degrees of freedom, p-value $< .0001$, inconsistent with the null hypothesis of equal population variances. Separate 99% confidence intervals for variance ratios:
190 variance/220 variance (1.15, 49.2) Note, 1 is not in this interval.
190 variance/240 variance (3.96, 169.65) Note, 1 is not in this interval.
190 variance/260 variance (1.23, 52.47) Note, 1 is not in this interval.
220 variance/240 variance (.53, 22.55)
220 variance/260 variance (.16, 6.97)
240 variance/260 variance (.047, 2.02)

(c) Levene's test statistic = 4.39, p-value = .01. Separate Wilcoxon-Mann-Whitney 99.1% confidence intervals for the difference between medians for transformed values:
group1−group2 (−672.1, 1343.8)
group1−group3 (−119.8, 1819.0)
group1−group4 (−420.0, 1639.1)
group2−group3 (−119.8, 599.9)
group2−group4 (−420.1, 395.9)
group3−group4 (−420, 46)
(e) Using the logarithm of failure time, Bartlett's test statistic = 12.2 with 3 degrees of freedom, p-value = .007, inconsistent with the null hypothesis of equal population variances.
(f) Using the logarithm of failure time, Levene's test statistic = 6.74, p-value = .001.

Exercise 14-14
(b) Bartlett's test statistic = 2.6 with 4 degrees of freedom, p-value = .6, consistent with the null hypothesis of equal population variances.
(c) Levene's test statistic = 4.45, p-value = .04.

Exercise 14-15
(b) Bartlett's test statistic = 2.3 with 2 degrees of freedom, p-value = .3, consistent with the null hypothesis of equal population variances. Separate 99% confidence intervals for ratios of population variances are:
Type1/Type2 (.14, 15.98)
Type1/Type3 (.20, 37.0)
Type2/Type3 (.097, 33.97)
(c) Levene's test statistic = .66, p-value = .5. Separate 99.2% Wilcoxon-Mann-Whitney confidence intervals for the difference between medians for transformed values:
group1−group2 (−16.80, 43.89)
group1−group3 (−13.19, 61.41)
group2−group3 (−11.31, 25.40)

Exercise 14-16
(b) Bartlett's test statistic = 8.4 with 3 degrees of freedom, p-value = .04, inconsistent with the null hypothesis of equal population variances. Separate 99% confidence intervals for ratios of population variances:
saline/regimen1 (.00044, 17.4)
saline/regimen2 (.000042, 168)
saline/regimen3 (.00022, 2.17)
regimen1/regimen2 (.00048, 1925)
regimen1/regimen3 (.0025, 24.8)
regimen2/regimen3 (.00023, 277)
(c) Levene's test statistic = 1.72, p-value = .2. Using transformed values, Wilcoxon-Mann-Whitney confidence intervals for differences between group means are:
group1−group3, 92% (−6.8, 2.6)
group1−group4, 95% (−26.75, −2.45)
group2−group4, 95% (−26.76, 1.76)
Other intervals cannot be obtained because the two transformed values in the third (Regimen 2) group are equal. Parametric pairwise confidence intervals all contain 0 for individual confidence levels 90-99%.
(e) Using the logarithm of stimulation index, Bartlett's test statistic = 3.15 with 3 degrees of freedom, p-value = .4, consistent with the null hypothesis of equal population variances.
(f) Using the logarithm of stimulation index, Levene's test statistic = 1.92, p-value = .2.

Exercise 14-17

(b) Bartlett's test statistic = 32.9 with 5 degrees of freedom, p-value < .0001, inconsistent with the null hypothesis of equal population variances. Separate 99% confidence intervals for ratios of population variances:
Fe3+,10.2/Fe2+,10.2 (.14, 1.94)
Fe3+,10.2/Fe3+, 1.2 (.037, .52)
Fe3+,10.2/Fe2+,1.2 (.025, .34)
Fe3+,10.2/Fe3+,.3 (.023, .31)
Fe3+,10.2/Fe2+,.3 (.030, .41)
Fe2+,10.2/Fe3+,1.2 (.072, .98)
Fe2+,10.2/Fe2+,1.2 (.047, .65)
Fe2+,10.2/Fe3+,.3 (.043, .59)
Fe2+,10.2/Fe2+,.3 (.057, .79)
Fe3+,1.2/Fe2+,1.2 (.18, 2.46)
Fe3+,1.2/Fe3+,.3 (.16, 2.23)
Fe3+,1.2/Fe2+,.3 (.22, 2.97)
Fe2+,1.2/Fe3+,.3 (.24, 3.36)
Fe2+,1.2/Fe2+,.3 (.33, 4.49)
Fe3+,.3/Fe2+,.3 (.36, 4.95)
The variances seem to fall in two groups: small, similar variances for Fe3+,10.2 and Fe2+,10.2; large, similar variances for the other four combinations.

(c) Levene's test statistic = 3.1, p-value = .01.

(d) Using the logarithm of response, Bartlett's test statistic = 2.6 with 5 degrees of freedom, p-value = .8, consistent with the null hypothesis of equal population variances.

(e) Levene's test statistic = .49, p-value = .8.

Exercise 14-18

(b) Bartlett's test statistic = 8.75 with 3 degrees of freedom, p-value = .03, inconsistent with the null hypothesis of equal population variances.

(c) We can't use Levene's test with just two values per group; the two transformed values within each group will be the same, giving a residual mean square of 0.

(e) Using the reciprocal of each observation, Bartlett's test statistic = 5.42 with 3 degrees of freedom, p-value = .14.

(f) We can't use Levene's test with just two values per group.

Exercise 14-19

(b) Bartlett's test statistic = 1 with 3 degrees of freedom, p-value = .8, consistent with the null hypothesis of equal population variances.

(c) Levene's test statistic = .62, p-value = .6.

Exercise 14-20

(b) Bartlett's test statistic = 3 with 3 degrees of freedom, p-value = .4, consistent with the null hypothesis of equal population variances.

(c) Levene's test statistic = .66, p-value = .6.

Exercise 14-21

(a) Using the logarithm of rupture time, Bartlett's test statistic = 7.1 with 2 degrees of freedom, p-value = .03, inconsistent with the null hypothesis of equal population variances.

(b) Using the logarithm of rupture time, Levene's test statistic = 11.9, p-value = .001.

Exercise 14-22
(a) Levene's test statistic = 2.53, p-value = .1. Separate 95.5% Wilcoxon-Mann-Whitney confidence intervals for differences between medians for transformed values:
group1−group2 (−222.1, 391.8)
group1−group3 (7.2, 724.0)
group2−group3 (134.0, 332.0)

Chapter 15

Exercise 15-1
(b) The linear correlation coefficient equals −.22. The test statistic equals −.64. Comparing with the t distribution with 8 degrees of freedom, the p-value equals .5, consistent with the null hypothesis of a linear correlation coefficient of 0.
(c) The rank correlation coefficient equals .055. The test statistic equals 156. Comparing with Spearman's distribution for sample size 10, the p-value equals .9, consistent with the null hypothesis of independence of sodium and potassium levels in perspiration of healthy women.

Exercise 15-2
(b) The linear correlation coefficient equals .355. The test statistic equals 1.42. Comparing with the t distribution with 14 degrees of freedom, the p-value equals .2, consistent with the null hypothesis of a linear correlation coefficient of 0.
(c) The rank correlation coefficient equals .295. D equals 477.5, the large sample test statistic equals −1.15, p-value = .2, consistent with the null hypothesis of independence of levels of the two substances.

Exercise 15-3 Shown are (linear correlation coefficient, rank correlation coefficient) for each pair of variables: average drop net catch and average sweep net catch (.954, .948); average drop net catch, average height of plants (−.379, −.340); average sweep net catch, average height of plants (−.280, −.181).

Exercise 15-4
(b) The linear correlation coefficient equals −.072. The test statistic equals −.3. Comparing with the t distribution with 17 degrees of freedom, p-value = .8, consistent with the null hypothesis of 0 linear correlation coefficient for body weight and heart weight.
(c) The rank correlation coefficient equals −.066, D equals 1213.5, large sample test statistic equals .27, p-value = .8, consistent with the null hypothesis of independence of body weight and heart weight.

Exercise 15-5
(b) The linear correlation coefficient equals .308. The test statistic equals 1.02. Comparing with the t distribution with 10 degrees of freedom, p-value = .3, consistent with the null hypothesis of 0 linear correlation coefficient for ground measurement and satellite measurement.
(c) The rank correlation coefficient equals −.014, D = 290, large sample test statistic = .046, p-value = .96, consistent with the null hypothesis of independence of ground and satellite measurements.

Exercise 15-6

(b) The linear correlation coefficient equals .433, test statistic = 1.07. Comparing with the t distribution with 5 degrees of freedom, p-value = .3, consistent with the null hypothesis that the linear correlation coefficient for recipient and donor inulin clearance is 0.

(c) The rank correlation coefficient equals .14, D = 48. Using Spearman's distribution for sample size 7, the p-value equals .78, consistent with the null hypothesis of independence of recipient and donor inulin clearance.

Exercise 15-7

(b) The linear correlation coefficient equals .704, test statistic = 2.6. Comparing with the t distribution with 7 degrees of freedom, p-value = .03, inconsistent with the null hypothesis of 0 linear correlation coefficient between average times at two light levels.

(c) The rank correlation coefficient equals .82, D = 22. Comparing with Spearman's distribution for sample size 9, p-value = .01, inconsistent with the null hypothesis of independence of average times at the two light levels.

Exercise 15-8

(a) The linear correlation coefficient equals .77, compared with the rank correlation coefficient of .77.

(b) The least squares line is: stiffness = −.15 + 38.8 thickness.

(c) The test statistic equals 2.44 with 4 degrees of freedom, p-value = .07, a borderline result, but suggesting possible nonzero slope.

(d) The test statistic equals −.02 with 4 degrees of freedom, p-value = .987, consistent with the null hypothesis of 0 intercept.

(e) $R^2 = r^2 = .597$, so 59.7% of the variation in stiffness is explained by the straight line model.

Exercise 15-9

(b) The linear correlation coefficient equals −.845.

(c) The least squares line is: steam used = 13.6 − .0798 average temperature, R^2=.714.

Exercise 15-10

(b) The least squares line is: second reading = 4.68 + .887 first reading.

(c) The least squares line is: first reading = .71 + .997 second reading.

(d) The equation of the standard deviation line is: second reading = 2.07 + .943 first reading.

(e) $R^2 = .884$

Exercise 15-11

(b) The least squares line is manhours = 2.5 + 3.99 items.

(c) The test statistic equals 291.37 with 20 degrees of freedom, p-value < .0001, inconsistent with the null hypothesis of 0 slope.

(d) The test statistic equals .20 with 20 degrees of freedom, p-value = 0.8, consistent with the null hypothesis of 0 intercept.

(f) R^2= 1.0, so nearly 100% of the variation in monthly manhours is explained by the straight line model.

Exercise 15-12

(b) The least squares line is: distance = −22.3 + 43.5 hang time. R^2=.671.

(c) The least squares line is: hang time = 1.64 + .0154 distance.

(d) The standard deviation line is: distance = −60.1 + 53.1 hang time.

Exercise 15-13

(b) The linear correlation coefficient equals .819. The test statistic equals 4.7. Comparing with the t distribution with 11 degrees of freedom, p-value = .0006, inconsistent with the null hypothesis of 0 linear correlation coefficient for distance and hang time.

(c) The rank correlation coefficient equals .725, D = 100, large sample test statistic = −2.5, p-value = .01, inconsistent with the null hypothesis of independence of distance and hang time.

Exercise 15-14

(b) The linear correlation coefficient equals .961 for height and weight; .881 for height and catheter length; .894 for weight and catheter length.

(c) The least squares line is: catheter length = 12.1 + .597 height, R^2=.776.

(d) The least squares line is: catheter length = 25.6 + .277 weight, R^2=.799.

(e) The least squares line is: weight = −46.6 + 2.10 height, R^2=.924.

Exercise 15-15

(b) The linear correlation coefficient equals .896 for left and right leg strength, .744 for left leg strength and distance, .791 for right leg strength and distance.

(c) The least squares line is: distance = 15.0 + .902 right leg strength, R^2=.626.

(d) The least squares line is: distance = 27.0 + .843 left leg strength, R^2=.554.

(e) The least squares line is: left leg strength = 10.7 + .901 right leg strength, R^2=.802.

Exercise 15-16

(b) The least squares model is: instrument response = .0488 + .695 copper concentration.

(c) The test statistic equals 121.03 with 10 degrees of freedom, p-value < .001, inconsistent with the null hypothesis of 0 slope.

(d) The test statistic equals 37.83 with 10 degrees of freedom, p-value < .001, inconsistent with the null hypothesis of 0 intercept.

(e) R^2 = .999, so 99.9% of the variation in instrument response is explained by the straight line model.

Exercise 15-17

(b) The correlation coefficient for failure time and temperature equals −.898 (the plot is not linear).

(d) The correlation coefficient for the logarithm of failure time and the reciprocal of temperature equals .958 (the plot is more linear).

(e) The least squares line is: logarithm of failure time = .476 + 654 reciprocal of temperature, R^2=.919.

Exercise 15-18

(b) The least squares line is: stopping distance = −62.0 + 3.49 velocity.

(c) R^2 = .984, so 98.4% of the variation in stopping distance is explained by the model in (b).

(f) The least squares line is: square root of stopping distance = −.878 + .228 distance.

(g) R^2 = .993 for the model in (f).

Exercise 15-19

(c) The least squares line is: logarithm of rupture time = 8.40 − .182 stress level.

(d) The test statistic equals −18.07 with 22 degrees of freedom, p-value < .0001, inconsistent with the null hypothesis of 0 slope.

(e) The test statistic equals 24.95 with 22 degrees of freedom, p-value < .0001, inconsistent with the null hypothesis of 0 intercept.

(f) R^2 = .937 for the model in (c).

(i) R^2 = .945 when we take the logarithm of both variables.

Exercise 15-20
(b) The least squares line is: flow rate = −3.98 + 13.8 depth, R^2=.947.
(e) The least squares line is: logarithm of flow rate = 1.16 + 2.76 logarithm of depth, R^2=.938.

Exercise 15-21
(d) The least squares model is: recovery=54.5+.00770pressure+.554angle+.000113(pressure)(angle), R^2 = .964. The test statistic for the coefficient of the product term is .68 with 5 degrees of freedom, p-value = .5 (the largest p-value of the three nonintercept parameter estimates).
(e) The least squares model is: recovery = 52.0 + .00940 pressure + .724 angle, R^2 = .961, all parameters needed in the model.

Exercise 15-22 Fit the entire model and decide to first remove left leg strength, p-value = .89. Refit the model and decide to remove left flexibility, p-value=.574. Refit the model and decide to remove right flexibility, p-value=.346. The final estimated model is: distance = 12.8 + .556(right leg strength) + .272(overall leg strength) with both p-values less than .03, R^2 = .785.

Exercise 15-23 Fit the entire model and decide to first remove right flexibility, p-value = .989. Refit the model and decide to remove right leg strength, p-value = .868. Refit the model and decide to remove left flexibility, p-value = .398. The final estimated model is: hang time = 1.11 + .0137(left leg strength) + .00429(overall leg strength), with both p-values less than or equal to .01, R^2 = .871.

Exercise 15-24
(b) The linear correlation coefficient equals −.773. Since R^2 = .598, 59.8% of the variation in permeability is explained by a straight model of permeability as a function of asphalt content. The plot is not linear.
(d) The least squares model is: permeability = −49 + 560(asphalt content) − 64.6(asphalt content)^2.
(e) Since R^2 = .834, 83.4% of the variability in permeability is explained by the model in (d).

Chapter 16

Exercise 16-1
(a) For the first plant, the null hypothesis is that p equals 3/4, with the two sided alternative, where p is the probability that a plant expresses the 30-kD protein. Observed frequencies are 29 and 11. The test statistic equals .1333. Comparing with the chi square distribution with 1 degree of freedom, the p-value equals .7, consistent with the null hypothesis.
(b) For the second plant, the null hypothesis is that p equals 15/16, with the two sided alternative, where p again is the probability that a plant expresses the 30-kD protein. Observed frequencies are 93 and 7. The test statistic equals .096. Comparing with the chi square distribution with 1 degree of freedom, the p-value equals .8, consistent with the null hypothesis.

Exercise 16-2 The null hypothesis states that the category probabilities are 1/4, 1/2 and 1/4. Comparisons are made with the chi square distribution with 2 degrees of freedom. For the five experiments, summary results are:
(i) Test statistic = .86, p-value = .65, consistent with the null hypothesis.
(ii) Test statistic = .057, p-value = .97, consistent with the null hypothesis.
(iii) Test statistic = 3.1, p-value = .21, consistent with the null hypothesis.
(iv) Test statistic = 2.8, p-value = .25, consistent with the null hypothesis.
(v) Test statistic = 5.1, p-value = .08, borderline case.

Exercise 16-3 The null hypothesis states that the category probabilities are 9/16, 3/16, 3/16 and 1/16. The test statistic equals 1.47. Comparing with the chi square distribution with 3 degrees of freedom, the p-value equals .7, consistent with the null hypothesis.

Exercise 16-4 The null hypothesis states that the category probabilities are 9/16, 3/16 and 4/16. The test statistic equals 1.6. Comparing with the chi square distribution with 2 degrees of freedom, p-value equals .4, consistent with the null hypothesis.

Exercise 16-5 The null hypothesis states that each of 24 category probabilities equals 1/24. The test statistic equals 163. Comparing with the chi square distribution with 23 degrees of freedom, p-value is close to 0, strongly inconsistent with the null hypothesis. The student can make observations about which times are more "popular" than others.

Exercise 16-6

(a) The null hypothesis states that the proportion suffering heart attack or cardiac death is the same for the two treatments. The table below has columns representing cardiac problems (yes, no) and rows representing treatments (drug, placebo). Observed (expected) frequencies are:

33 (21)	697 (709)
9 (21)	721 (709)

The test statistic equals 14. Comparing with the chi square distribution with 1 degree of freedom, p-value = .0002, inconsistent with the null hypothesis.

(b) The large sample test statistic in 11-2a equals 3.76 and $3.76^2 = 14$ (rounded). This agrees with the result that the square of a standard Gaussian random variable has the chi square distribution with 1 degree of freedom. These two large sample tests are equivalent.

Exercise 16-7 The null hypothesis states that there is no association between the categorical variable survival time and the categorical variable percent change in antibody response. The p-value equals

$$\text{p-value} = \left(\binom{6}{4}\binom{7}{1}+\binom{6}{5}\binom{7}{0}+\binom{6}{0}\binom{7}{5}\right)/\binom{13}{5} = .1,$$

a borderline case, suggesting that greater antibody response might be associated with longer survival.

Exercise 16-8 The null hypothesis states that the probability of a response is the same for each of the four doses. The test statistic equals 21.4. Comparing with the chi square distribution with 3 degrees of freedom, p-value = .00009, inconsistent with the null hypothesis. There is greater response at higher doses.

Exercise 16-9 The null hypothesis states that obesity classification and hypertension are independent in the population samples. The test statistic equals 11.7. Comparing with the chi square distribution with 2 degrees of freedom, p-value=.003, inconsistent with the null hypothesis. Higher obesity categories seem to have higher proportions with hypertension.

Exercise 16-10
(a) To test the null hypothesis stated in the problem, the test statistic equals 61. Comparing with the chi square distribution with 1 degree of freedom, the p-value is close to 0, inconsistent with the null hypothesis. High parental encouragement is associated with a greater proportion of males planning to attend college.
(b) To test the null hypothesis stated in the problem, the test statistic equals 89. Comparing with the chi square distribution with 1 degree of freedom, the p-value is close to 0, inconsistent with the null hypothesis. High parental encouragement is associated with a greater proportion of females planning to attend college.
(c) The effect of parental encouragement on college plans seems very similar for males and females in this sample.

Exercise 16-11 The null hypothesis states that 3 year survival (yes, no) is independent of tumor appearance (malignant, benign) in the population of women under 50 years of age that was sampled. The test statistic equals .9. Comparing with the chi square distribution with 1 degree of freedom, the p-value equals .3, consistent with the null hypothesis.

Exercise 16-12 The null hypothesis states that there is no association between poor sleeping and irritability. The test statistic equals 41.5. Comparing with the chi square distribution with 1 degree of freedom, the p-value is close to 0, strongly inconsistent with the null hypothesis.

Exercise 16-13 The null hypothesis states that the distribution across termination categories is the same for the two doses. The test statistic equals 8.7. Comparing with the chi square distribution with 2 degrees of freedom, the p-value equals .01, inconsistent with the null hypothesis. A greater proportion developed tumor before end of experiment on the high dose than on the low dose. A greater proportion were alive with no tumor at end of experiment on the low dose than on the high dose.

Exercise 16-14
(a) The large sample test statistic equals 9.2. Comparing with the chi square distribution with 1 degree of freedom, the p-value equals .002, inconsistent with the null hypothesis. A smaller proportion developed an overwhelming infection in the penicillin group than in the placebo group.
(b) The smallest frequency in the 2×2 table is 0, p-value = .2:

$$\text{p-value} = \left(\binom{3}{0}\binom{212}{110}+\binom{3}{3}\binom{212}{107}\right)\Big/\binom{215}{110} = .2.$$

The observed table has the smaller probability, so the student may report the p-value as .1, the probability for the observed table. In either case, there is not strong enough evidence to say there is a difference in death rates between the two treatment groups.

Exercise 16-15
(a) The large sample test statistic equals 6.3. Comparing with the chi square distribution with 1 degree of freedom, the p-value equals .01, inconsistent with the null hypothesis.
(b) The smallest frequency in the 2×2 table is 2, p-value = .02, inconsistent with the null hypothesis:

$$\text{p-value} = \left(\binom{21}{7}\binom{31}{2}+\binom{21}{8}\binom{31}{1}+\binom{21}{9}\binom{31}{0}+\binom{21}{0}\binom{31}{9}\right)\Big/\binom{52}{9} = .02.$$

Exercise 16-16 The null hypothesis states that the proportion showing improvement is the same for the two treatments. The smallest frequency in the 2×2 table is 1, p=value=1, consistent with the null hypothesis:

$$\text{p-value} = 2\left(\binom{12}{1}\binom{12}{2}+\binom{12}{0}\binom{12}{3}\right)/\binom{24}{3} = 1$$

Exercise 16-17 The null hypothesis states that the proportion conceiving is the same in the two treatment groups. The smallest frequency in the 2×2 table is 0, p-value = .01:

$$\text{p-value} = 2\binom{12}{6}\binom{12}{0}/\binom{24}{6} = .01,$$

inconsistent with the null hypothesis. Infertility in female rats may be associated with exposure to nitrous oxide.

Exercise 16-18 The null hypothesis states that among divers with a history of decompression sickness, the proportion having the heart defect is .05. Using the Binomial(30, .05) distribution, the p-value is close to 0, inconsistent with the null hypothesis. This agrees with what we found in Exercise 10-4.

Exercise 16-19 The null hypothesis states that for 20 to 24 year old white men, the distribution across the two body weight categories is the same for Britain, Canada and United States. The test statistic equals 15. Comparing with the chi square distribution with 2 degrees of freedom, the p-value equals .0006, inconsistent with the null hypothesis. There is a greater proportion in the excessive category in the United States than in the other two countries.

Exercise 16-20 The test statistic equals 1.8. Comparing with the chi square distribution with 1 degree of freedom, p-value = .2, consistent with the null hypothesis. This agrees well with Fisher's exact test results in Example 6-5.

Exercise 16-21 The test statistic equals 5.0. Comparing with the chi square distribution with 1 degree of freedom, p-value = .03, inconsistent with the null hypothesis. This agrees well with Fisher's exact test results in Example 6-6.