# Chapter 15 Solutions

**15.1.** The rankings are shown on the right. Group A ranks are 1, 2, 4, 6, and 8; Group B ranks are 3, 5, 7, 9, and 10.

| Group | Rooms | Ranks |
|---|---|---|
| A | 30 | 1 |
| A | 68 | 2 |
| B | 240 | 3 |
| A | 243 | 4 |
| B | 329 | 5 |
| A | 448 | 6 |
| B | 540 | 7 |
| A | 552 | 8 |
| B | 560 | 9 |
| B | 780 | 10 |

**15.2.** The list of ranks is not shown because it is nearly identical to the one shown in the previous solution; the only change needed is to change 780 to 4003 in the last line. The ranks assigned to each group are exactly the same.

**15.3.** The null hypothesis is $H_0$: no difference in distribution of number of rooms. The alternative might be two-sided ("there is a difference") or one-sided if we had a prior suspicion that one group had more rooms than the other. The test statistic is $W = 1 + 2 + 4 + 6 + 8 = 21$.

**15.4.** Changing the data does not change the hypotheses, so they are the same as in the previous solution. Additionally, because the assigned ranks did not change, the test statistic is still $W = 21$.

**15.5.** Under $H_0$, $\mu_W = \frac{(5)(11)}{2} = 27.5$ and $\sigma_W = \sqrt{\frac{(5)(5)(11)}{12}} \doteq 4.7871$. We found $W = 21$, so $z = \frac{21-27.5}{4.7871} \doteq -1.36$, for which the two-sided $P$-value is $2P(Z \leq -1.36) \doteq 0.1738$. With the continuity correction, we find $z = \frac{21.5-27.5}{4.787} \doteq -1.25$, which gives $P = 2P(Z \leq -1.25) = 0.2112$. The Minitab output below gives a similar $P$-value to that found with the continuity correction; the difference is due to the rounding of $z$. Regardless of the $P$-value used, we do not reject $H_0$.

**Note:** *If a one-sided alternative was specified in Exercise 15.3, the P-value would be half as big: $P \doteq 0.0869$, or 0.1056 with the continuity correction.*

**Minitab output**

```
Mann-Whitney Confidence Interval and Test

GrpA       N =   5     Median =       243.0
GrpB       N =   5     Median =       540.0
Point estimate for ETA1-ETA2 is      -228.0
96.3 Percent C.I. for ETA1-ETA2 is (-537.0,208.0)
W = 21.0
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.2101
```

**15.6.** Because the ranks and test statistic are unchanged, all answers are the same as those given in the previous solution.

**15.7. (a)** Normal quantile plots are not shown. The score 0.00 for child 8 seems to be a low outlier (although with only five observations, such judgments are questionable). **(b)** For testing $H_0$: $\mu_1 = \mu_2$ vs. $H_a$: $\mu_1 > \mu_2$, we have $\bar{x}_1 = 0.676$, $s_1 \doteq 0.1189$, $\bar{x}_2 = 0.406$, and $s_2 \doteq 0.2675$. Then, $t = 2.062$, which gives $P = 0.0447$ (df $= 5.5$). We have some evidence that high-progress readers have higher mean scores. **(c)** We test:

$H_0$: Scores for both groups are identically distributed
vs. $H_a$: High-progress children systematically score higher

for which we find $W = 36$ and $P \doteq 0.0473$ or $0.0463$—significant evidence (at $\alpha = 0.05$) against the hypothesis of identical distributions. This is equivalent to the conclusion reached in (b).

**Note:** *Minitab refers to the medians as ETA1 and ETA2 ("eta" is the Greek letter $\eta$). Minitab also reports an estimate of 0.21 for the difference $\eta_1 - \eta_2$; note that this is not the same as the difference between the two sample medians ($0.7 - 0.4 = 0.3$). This estimate, called the Hodges-Lehmann estimate, is not discussed in the text and has been removed from the Minitab outputs accompanying other solutions for this chapter. Briefly, this estimate is found by taking every response from the first group and subtracting every response from the second group, yielding (in this case) a total of 25 differences. The median of this set of differences is the Hodges-Lehmann estimate.*

**Minitab output**

```
Mann-Whitney Confidence Interval and Test

HiProg1    N =   5     Median =      0.7000
LoProg1    N =   5     Median =      0.4000
Point estimate for ETA1-ETA2 is      0.2100
96.3 Percent C.I. for ETA1-ETA2 is (-0.0199,0.7001)
W = 36.0
Test of ETA1 = ETA2  vs.  ETA1 > ETA2 is significant at 0.0473
The test is significant at 0.0463 (adjusted for ties)
```

**15.8. (a)** Normal quantile plots are not shown. The score 0.54 for child 3 seems to be a low outlier. **(b)** For testing $H_0$: $\mu_1 = \mu_2$ vs. $H_a$: $\mu_1 > \mu_2$, we have $\bar{x}_1 = 0.768$, $s_1 \doteq 0.1333$, $\bar{x}_2 = 0.516$, $s_2 \doteq 0.2001$. Then, $t = 2.344$, which gives $P = 0.0259$ (df $= 6.97$). We have fairly strong evidence that high-progress readers have higher mean scores. **(c)** We test:

$H_0$: Scores for both groups are identically distributed
vs. $H_a$: High-progress children systematically score higher

for which we find $W = 38$ and $P \doteq 0.0184$. This is evidence against $H_0$, slightly stronger than that found in (b).

**Minitab output**

```
Mann-Whitney Confidence Interval and Test

HiProg2    N =   5     Median =      0.8000
LoProg2    N =   5     Median =      0.4900
W = 38.0
Test of ETA1 = ETA2  vs.  ETA1 > ETA2 is significant at 0.0184
```

**15.9.** **(a)** See table. **(b)** For Story 2, $W = 8 + 9 + 4 + 7 + 10 = 38$. Under $H_0$:

$$\mu_W = \frac{(5)(11)}{2} = 27.5$$

$$\sigma_W = \sqrt{\frac{(5)(5)(11)}{12}} \doteq 4.7871$$

**(c)** $z = \frac{38 - 27.5}{4.787} \doteq 2.19$; with the continuity correction, we compute $\frac{37.5 - 27.5}{4.787} \doteq 2.09$, which gives $P = P(Z > 2.09) = 0.0183$. **(d)** See the table.

| | | Story 1 | | Story 2 | |
|---|---|---|---|---|---|
| Child | Progress | Score | Rank | Score | Rank |
| 1 | high | 0.55 | 4.5 | 0.80 | 8 |
| 2 | high | 0.57 | 6 | 0.82 | 9 |
| 3 | high | 0.72 | 8.5 | 0.54 | 4 |
| 4 | high | 0.70 | 7 | 0.79 | 7 |
| 5 | high | 0.84 | 10 | 0.89 | 10 |
| 6 | low | 0.40 | 3 | 0.77 | 6 |
| 7 | low | 0.72 | 8.5 | 0.49 | 3 |
| 8 | low | 0.00 | 1 | 0.66 | 5 |
| 9 | low | 0.36 | 2 | 0.28 | 1 |
| 10 | low | 0.55 | 4.5 | 0.38 | 2 |

**15.10.** **(a)** We find $W = 26$ and $P \doteq 0.0152$. We have strong evidence against the hypothesis of identical distributions; we conclude that the weed-free yield is higher. **(b)** For testing $H_0$: $\mu_0 = \mu_9$ vs. $H_a$: $\mu_0 > \mu_9$, we find $\bar{x}_0 = 170.2$, $s_0 \doteq 5.4216$, $\bar{x}_9 = 157.575$, $s_9 \doteq 10.1181$, and $t = 2.20$, which gives $P = 0.0423$ (df $= 4.6$). We have fairly strong evidence that the mean yield is higher with no weeds—but the evidence is not quite as strong as in (a). **(c)** Both tests still reach the same conclusion, so there is no "practically important impact" on our conclusions. The Wilcoxon evidence is slightly weaker: $W = 22$, $P \doteq 0.0259$. The $t$-test evidence is slightly stronger: $t = 2.79$, df $= 3$, $P = 0.0341$. The new statistics for the 9-weeds-per-meter group are $\bar{x}_9 = 162.633$ and $s_9 \doteq 0.2082$; these are substantial changes for each value.

**Minitab output**

```
- - - - - - - - - - - - - - - - - - -  All points  - - - - - - - - - - - - - - - - - - -
Weed0       N =    4     Median =       169.45
Weed9       N =    4     Median =       162.55
W = 26.0
Test of ETA1 = ETA2  vs.  ETA1 > ETA2 is significant at 0.0152
- - - - - - - - - - - - - - - - -  With outlier removed  - - - - - - - - - - - - - - - - -
C2          N =    4     Median =       169.45
C4          N =    3     Median =       162.70
W = 22.0
Test of ETA1 = ETA2  vs.  ETA1 > ETA2 is significant at 0.0259
```

**15.11.** **(a)** The distributions do not appear to have the same shape; in particular, the 16-week distribution is much more spread out. **(b)** The Wilcoxon rank-sum test gives $W = 33$ and $P \doteq 0.1481$ (0.1467, adjusted for ties). We do not have enough evidence to conclude that burial lowers breaking strength.

| 2 weeks | | 16 weeks |
|---|---|---|
| | 9 | 8 |
| | 10 | |
| 8 | 11 | 00 |
| 9660 | 12 | 4 |
| | 13 | |
| | 14 | 0 |

**Minitab output**

```
2weeks      N =    5     Median =       126.00
16weeks     N =    5     Median =       110.00
W = 33.0
Test of ETA1 = ETA2  vs.  ETA1 > ETA2 is significant at 0.1481
The test is significant at 0.1467 (adjusted for ties)
```

**15.12.** **(a)** The outline is shown below. **(b)** We consider score improvements (posttest minus pretest). The means, medians, and standard deviations are:

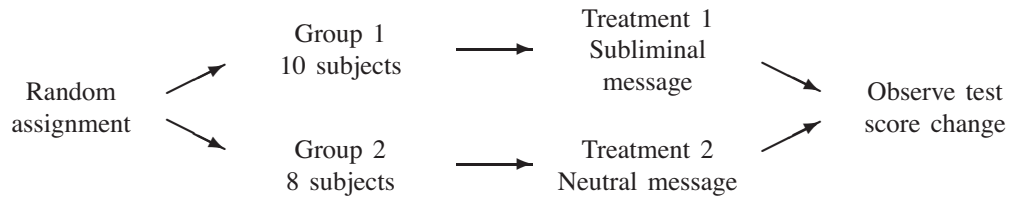
| | | | |
|---|---|---|---|
| Treatment: | $\bar{x} = 11.4$ | $M = 11.5$ | $s \doteq 3.1693$ |
| Control: | $\bar{x} = 8.25$ | $M = 7.5$ | $s \doteq 3.6936$ |

| Treatment | | Control |
|---|---|---|
| | 0 | 455 |
| 76 | 0 | 7 |
| | 0 | 8 |
| 110 | 1 | 1 |
| 332 | 1 | 2 |
| 5 | 1 | 4 |
| 6 | 1 | |

A back-to-back stemplot is one way to compare the distributions graphically. Both of these comparisons support the idea that the positive subliminal message resulted in higher test scores. **(c)** We have $W = 114$, for which $P = 0.0501$ (or 0.0494, adjusted for ties). This is just about significant at $\alpha = 0.05$, and at least warrants further study.

**Minitab output**

```
Trtmt      N =  10     Median =      11.500
Ctrl       N =   8     Median =       7.500
W = 114.0
Test of ETA1 = ETA2  vs.  ETA1 > ETA2 is significant at 0.0501
The test is significant at 0.0494 (adjusted for ties)
```




**15.13.** **(a)** At right. Unlogged plots appear to have a greater number of species. **(b)** $H_0$: There is no difference in the number of species on logged and unlogged plots vs. $H_a$: Unlogged plots have a greater variety of species. The Wilcoxon test gives $W = 159$ and $P \doteq 0.0298$ (0.0290, adjusted for ties). We conclude that the observed difference is significant; unlogged plots really do have a greater number of species.

| Unlogged | | Logged |
|---|---|---|
| | 0 | 4 |
| | 0 | |
| | 0 | |
| | 1 | 0 |
| 333 | 1 | 2 |
| 55 | 1 | 455 |
| | 1 | 7 |
| 998 | 1 | 88 |
| 10 | 2 | |
| 22 | 2 | |

**Minitab output**

```
Unlogged   N =  12     Median =      18.500
Logged     N =   9     Median =      15.000
W = 159.0
Test of ETA1 = ETA2  vs.  ETA1 > ETA2 is significant at 0.0298
The test is significant at 0.0290 (adjusted for ties)
```

**15.14.** For the Wilcoxon test, we have $W = 579$, for which $P = 0.0064$ (0.0063, adjusted for ties). The evidence is slightly stronger with the Wilcoxon test than for the $t$ and permutation tests.

**Minitab output**

```
Trtmt      N =  21     Median =       53.00
Ctrl       N =  23     Median =       42.00
W = 579.0
Test of ETA1 = ETA2  vs.  ETA1 > ETA2 is significant at 0.0064
The test is significant at 0.0063 (adjusted for ties)
```

**15.15.** A table of counts and percents suggests that women do give higher ratings. One could do a chi-square test (which yields $X^2 = 21.940$ and $P$ tiny), but this is not advisable because there are so many small counts. The Wilcoxon test is better; it gives $W = 32{,}267.5$ and $P = 0.0003$ (or $P = 0.0001$ adjusted for ties)—strong evidence that women really do rate restaurant food as less safe than men do.

| | Women | | Men | |
|---|---|---|---|---|
| Rating | Count | % | Count | % |
| 1 | 48 | 24.49 | 52 | 48.60 |
| 2 | 123 | 62.76 | 45 | 42.06 |
| 3 | 19 | 9.69 | 8 | 7.48 |
| 4 | 5 | 2.55 | 0 | 0 |
| 5 | 1 | 0.51 | 2 | 1.87 |

**Minitab output**

```
women       N = 196     Median =      2.0000
men         N = 107     Median =      2.0000
W = 32267.5
Test of ETA1 = ETA2  vs.  ETA1 > ETA2 is significant at 0.0003
The test is significant at 0.0001 (adjusted for ties)
```

**15.16.** We do not have independent samples from two populations; rather, we have dependent samples (each person answered both questions).

**15.17. (a)** We find $X^2 = 3.955$ with df $= (5-1)(2-1) = 4$, giving $P = 0.413$. There is little evidence to make us believe that there is a relationship between city and income.
**(b)** Minitab reports $W = 56{,}370$, with $P \doteq 0.5$; again, there is no evidence that incomes are systematically higher in one city.

**Minitab output**

```
City1      N = 241     Median =      2.0000
City2      N = 218     Median =      2.0000
W = 56370.0
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.5080
The test is significant at 0.4949 (adjusted for ties)
```

**15.18.** We test:

$H_0$: Food scores and activities scores have the same distribution
vs. $H_a$: Food scores are higher

The differences, and their ranks, are:

| Spa | Food score | Activities score | Difference | Rank |
|---|---|---|---|---|
| 1 | 90.9 | 93.8 | −2.9 | 4 |
| 2 | 92.3 | 92.3 | 0.0 | – |
| 3 | 88.6 | 91.4 | −2.8 | 3 |
| 4 | 81.8 | 95.0 | −13.2 | 6 |
| 5 | 85.7 | 89.2 | −3.5 | 5 |
| 6 | 88.9 | 88.2 | 0.7 | 1 |
| 7 | 81.0 | 81.8 | −0.8 | 2 |

In fact, it is not necessary to give the complete set of rankings; the only observations we need to make are (1) there is only one positive difference and (2) it is the smallest (in absolute value) of all the nonzero differences. Therefore, $W^+ = 1$.

**Note:** *In assigning ranks, differences of 0 are ignored; see the comment in the text toward the bottom of page 750. If a student mistakenly assigns a rank of 1 to 0, they would find $W^+ = 2$ (or perhaps 3 if they erroneously count 0 as a "positive difference").*

**15.19.** We test:

$H_0$: Food scores and activities scores have the same distribution
vs. $H_a$: Food scores are higher

The differences, and their ranks, are:

| Spa | Food score | Activities score | Difference | Rank |
|---|---|---|---|---|
| 1 | 77.3 | 95.7 | −18.4 | 6 |
| 2 | 85.7 | 78.0 | 7.7 | 2 |
| 3 | 84.2 | 87.2 | −3.0 | 1 |
| 4 | 85.3 | 85.3 | 0.0 | – |
| 5 | 83.7 | 93.6 | −9.9 | 5 |
| 6 | 84.6 | 76.0 | 8.6 | 4 |
| 7 | 78.5 | 86.3 | −7.8 | 3 |

The two positive differences have ranks 2 and 4, so $W^+ = 6$.

**Note:** *In assigning ranks, differences of 0 are ignored; see the comment in the text toward the bottom of page 750. If a student mistakenly assigns a rank of 1 to 0, they would find $W^+ = 3 + 5 = 8$ (or perhaps 9 if they erroneously count 0 as a "positive difference").*

**15.20.** Because one difference was 0, we ignore it and take $n = 6$, so that:

$$\mu_{W^+} = \frac{6(6+1)}{4} \doteq 10.5, \quad \sigma_{W^+} = \sqrt{\frac{6(6+1)(12+1)}{24}} = \sqrt{22.75} \doteq 4.7697,$$

and the approximate $P$-value is $P(W^+ \geq 0.5) \doteq P(Z \geq -2.10) \doteq 0.9821$. This agrees with the Minitab output below; see the note in the solution to Exercise 15.22 for an explanation of the estimated median reported by Minitab.

**Note:** *If a student does not see the instruction about discarding differences of 0 at the bottom of page 750, they might compute the mean and standard deviation using $n = 7$: $\mu_{W^+} = \frac{7(7+1)}{4} \doteq 14$ and $\sigma_{W^+} = \sqrt{\frac{7(7+1)(14+1)}{24}} = \sqrt{35} \doteq 5.9161$. Such a student would presumably take $W^+ = 2$ (or 3), so they would compute the approximate P-value as $P(W^+ \geq 1.5) \doteq P(Z \geq -2.11) \doteq 0.9826$ or $P(W^+ \geq 2.5) \doteq P(Z \geq -1.94) \doteq 0.9738$. These are quite close to the right answer (and of course lead to the same conclusion) but are not quite correct. In other situations, failing to ignore differences of 0 may lead to the wrong conclusion.*

**Minitab output**

```
Wilcoxon Signed Rank Test

TEST OF MEDIAN = 0.000000 VERSUS MEDIAN G.T. 0.000000

               N FOR   WILCOXON           ESTIMATED
          N     TEST  STATISTIC  P-VALUE     MEDIAN
Diff      7        6        1.0    0.982     -2.000
```

**15.21.** Because one difference was 0, we ignore it and take $n = 6$, so that:

$$\mu_{W^+} = \frac{6(6+1)}{4} \doteq 10.5, \quad \sigma_{W^+} = \sqrt{\frac{6(6+1)(12+1)}{24}} = \sqrt{22.75} \doteq 4.7697,$$

and the approximate $P$-value is $P(W^+ \geq 5.5) \doteq P(Z \geq -1.05) \doteq 0.8531$. This agrees with the Minitab output below; see the note in the next solution for an explanation of the Minitab's estimated median.

**Note:** *If a student does not see the instruction about discarding differences of 0 at the bottom of page 750, they might compute the mean and standard deviation using $n = 7$: $\mu_{W^+} = \frac{7(7+1)}{4} \doteq 14$ and $\sigma_{W^+} = \sqrt{\frac{7(7+1)(14+1)}{24}} = \sqrt{35} \doteq 5.9161$. Such a student would presumably take $W^+ = 8$ (or 9), so they would compute the approximate P-value as $P(W^+ \geq 7.5) \doteq P(Z \geq -1.10) \doteq 0.8643$ or $P(W^+ \geq 8.5) \doteq P(Z \geq -0.93) \doteq 0.8238$. These are close to the right answer (and of course lead to the same conclusion) but are not quite correct. In other situations, failing to ignore differences of 0 may lead to the wrong conclusion.*

**Minitab output**

```
Wilcoxon Signed Rank Test

TEST OF MEDIAN = 0.000000 VERSUS MEDIAN G.T. 0.000000

               N FOR   WILCOXON           ESTIMATED
          N     TEST  STATISTIC  P-VALUE     MEDIAN
Diff      7        6        6.0    0.853     -3.450
```

**15.22.** **(a)** The differences (treatment minus control) were 0.01622, 0.01102, and 0.01607. The mean difference was $\bar{x} \doteq 0.01444$, and $s \doteq 0.002960$. The fact that all are positive supports the idea that there was more growth in the treated plots. **(b)** For testing $H_0$: $\mu = 0$ vs. $H_a$: $\mu > 0$, with $\mu$ the mean (treatment minus control) difference, we have $t = \frac{\bar{x}}{s/\sqrt{3}} \doteq 8.45$, df $= 2$, and $P = 0.0069$. We conclude that growth was greater in treated plots. **(c)** The Wilcoxon statistic is $W^+ = 6$, for which $P = 0.091$. We would not reject $H_0$ (which states that there is no difference among pairs). **(d)** A low-power test has a low probability of rejecting $H_0$ when it is false.

**Minitab output**

```
TEST OF MEDIAN = 0.000000 VERSUS MEDIAN G.T. 0.000000

               N FOR   WILCOXON           ESTIMATED
          N     TEST  STATISTIC  P-VALUE     MEDIAN
Diff      3        3        6.0    0.091    0.01485
```

**Note:** *With only three pairs, the Wilcoxon signed rank test can* never *give a P-value smaller than 0.091. This is one difference between some nonparametric tests and parametric tests like the t test: With the t test, the power improves when we consider alternatives that are farther from the null hypothesis; for example, if $H_0$ says $\mu = 0$, we have higher power for the alternative $\mu = 10$ than for $\mu = 5$. With the Wilcoxon signed rank test, all alternatives look the same; the values of $W^+$ and P would be the same if the three differences had been 100, 200, and 300.*

*Also, note that the "estimated median" in the Minitab output (0.01485) is not the same as the median of the three differences (0.01607). The process of computing this point estimate is not discussed in the text, but we will illustrate it for this simple case: The Wilcoxon estimated median is the median of the set of* Walsh averages *of the differences.*

*This set consists of every possible pairwise average $(x_i + x_j)/2$ for $i \le j$; note that this includes $i = j$, in which case the average is $x_i$. In general, there are $n(n + 1)/2$ such averages, so with $n = 3$ differences, we have 6 Walsh averages: the three differences (0.01622, 0.01102, and 0.01607) and the averages of each pair of distinct differences (0.013545, 0.01362, and 0.016145). The median of 0.01102, 0.013545, 0.01362, 0.01607, 0.016145, and 0.01622 is 0.014845.*

**15.23.** We examine the heart-rate increase (final minus resting) from low-rate exercise; our hypotheses are $H_0$: median $= 0$ vs. $H_a$: median $> 0$. The statistic is $W^+ = 10$ (the first four differences are positive, and the fifth is 0, so we drop it). We compute $P = P(W^+ \ge 9.5) = P\left(\frac{W^+ - 5}{2.739} \ge \frac{9.5 - 5}{2.739}\right) \doteq P(Z \ge 1.64) = 0.0505$. This is right on the borderline of significance: It is fairly strong evidence that heart rate increases, but (barely) not significant at 5%. (See the note in the solution to the previous exercise for an explanation of the estimated median reported by Minitab.)

**Minitab output**

```
TEST OF MEDIAN = 0.000000 VERSUS MEDIAN G.T. 0.000000

               N FOR   WILCOXON           ESTIMATED
            N   TEST  STATISTIC  P-VALUE     MEDIAN
LowDiff     5      4       10.0    0.050      7.500
```

**15.24.** **(a)** For this exercise, we must first find the Final − Resting differences for both exercise rates (Low: 15, 9, 6, 9, 0; Medium: 21, 24, 15, 15, 18), then compute the differences of these differences (6, 15, 9, 6, 18). To this last list of differences, we apply the Wilcoxon signed rank test. The hypotheses are $H_0$: median $= 0$ vs. $H_a$: median $> 0$. (The rank sum test is not appropriate because we do not have two independent samples.) **(b)** The statistic is $W^+ = 15$ (all five differences were positive), and the reported $P$-value is 0.030—fairly strong evidence that medium-rate exercise increases are greater than low-rate exercise increases. (See the note in the solution to Exercise 15.22 for an explanation of the estimated median reported by Minitab.)

**Minitab output**

```
TEST OF MEDIAN = 0.000000 VERSUS MEDIAN G.T. 0.000000

               N FOR   WILCOXON           ESTIMATED
            N   TEST  STATISTIC  P-VALUE     MEDIAN
LowMed      5      5       15.0    0.030      10.50
```

**15.25.** For testing $H_0$: median $= 0$ vs. $H_a$: median $> 0$, the Wilcoxon statistic is $W^+ = 119$ (14 of the 15 differences were positive, and the one negative difference was the smallest in absolute value), and $P < 0.0005$—very strong evidence that there are more aggressive incidents during moon days. This agrees with the results of the $t$ and permutation tests. (See the note in the solution to Exercise 15.22 for an explanation of the estimated median reported by Minitab.)

**Minitab output**

```
TEST OF MEDIAN = 0.000000 VERSUS MEDIAN G.T. 0.000000

               N FOR   WILCOXON           ESTIMATED
            N   TEST  STATISTIC  P-VALUE     MEDIAN
diff       15     15      119.0    0.000      2.570
```

**15.26.** There are 17 nonzero differences; only one is negative (the boldface 6 in the list below).

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Diff: | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | **6** | 6 | 6 | 6 | 6 | 6 |
| Rank: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| Value: | 1.5 | | | 4 | | | | | 8.5 | | | | | | 14.5 | | | |

This gives $W^+ = 138.5$. (Note that the only tie we really need to worry about is the last group; all other ties involve only positive differences.)

**15.27.** For the differences sfair − srest, we find $\bar{x} = 0.5149$ and $M = 0.5$. Applying the Wilcoxon signed rank test to these differences, with the one-sided alternative—"food at fairs is systematically rated higher (less safe) than restaurant food"—we obtain $W^+ = 10{,}850.5$ ($P < 0.0005$), so we conclude that restaurant food is viewed as being safer. However, there were 146 ties; the text cautions us that, when there are many ties, the test may be biased in favor of $H_a$, so we must be cautious about our conclusion. (See the note in the solution to Exercise 15.22 for an explanation of the estimated median reported by Minitab.)

**Minitab output**

```
TEST OF MEDIAN = 0.000000 VERSUS MEDIAN G.T. 0.000000

               N FOR   WILCOXON           ESTIMATED
          N    TEST  STATISTIC  P-VALUE     MEDIAN
Diffs   303     157    10850.5    0.000     0.5000
```

**15.28.** Considering the differences (after − before), we test $H_0$: median = 0 vs. $H_a$: median > 0. The Wilcoxon statistic is $W^+ = 66$ (11 of the 14 differences were positive, and the other 3 were 0), and $P = 0.002$—very strong evidence that glove use increased after the presentation. (See the note in the solution to Exercise 15.22 for an explanation of the estimated median reported by Minitab.)

**Minitab output**

```
TEST OF MEDIAN = 0.000000 VERSUS MEDIAN G.T. 0.000000

               N FOR   WILCOXON           ESTIMATED
          N    TEST  STATISTIC  P-VALUE     MEDIAN
Diff     14      11       66.0    0.002     0.6665
```

**15.29.** **(a)** At right. The distribution is clearly right-skewed but has no outliers. **(b)** $W^+ = 31$ (only 4 of 12 differences were positive) and $P = 0.556$—there is no evidence that the median is other than 105. (See the note in the solution to Exercise 15.22 for an explanation of the estimated median reported by Minitab.)

```
 9 | 1
 9 | 5679
10 | 134
10 | 5
11 | 1
11 | 9
12 | 2
```

**Minitab output**

```
TEST OF MEDIAN = 105.0 VERSUS MEDIAN N.E. 105.0

               N FOR   WILCOXON           ESTIMATED
          N    TEST  STATISTIC  P-VALUE     MEDIAN
Radon    12      12       31.0    0.556      103.2
```

**15.30.** If we compute Haiti content minus factory content (so that a negative difference means that the amount of vitamin C decreased), we find that the mean change is $-5.33$ and the median is $-6$. (See the note in the solution to Exercise 15.22 for an explanation of the estimated median reported by Minitab.) The stemplot is right-skewed. There are five positive differences; the Wilcoxon statistic is $W^+ = 37$, for which $P < 0.0005$. The differences are systematically negative, so vitamin C content is lower in Haiti.

| | |
|---|---|
| $-1$ | 4 |
| $-1$ | 3322 |
| $-1$ | |
| $-0$ | 9988 |
| $-0$ | 7776666 |
| $-0$ | 5444 |
| $-0$ | 2 |
| $-0$ | 1 |
| 0 | 1 |
| 0 | 33 |
| 0 | 4 |
| 0 | |
| 0 | 8 |

**Minitab output**

```
TEST OF MEDIAN = 0.000000 VERSUS MEDIAN L.T. 0.000000

                N FOR   WILCOXON           ESTIMATED
             N   TEST  STATISTIC  P-VALUE     MEDIAN
Change      27     27       37.0    0.000     -5.500
```

**15.31.** **(a)** The Wilcoxon statistic is $W^+ = 0$ (all of the differences were less than 16), for which $P = 0$—very strong evidence against $H_0$. We conclude that the median weight gain is less than 16 pounds. **(b)** Minitab gives the interval 3.75 to 5.90 kg for the median weight gain. (For comparison, in the solution to Exercise 7.32, the 95% confidence interval for the mean $\mu$ was about 3.80 to 5.66 kg. See the note in the solution to Exercise 15.22 for an explanation of the estimated median reported by Minitab.)

**Minitab output**

```
TEST OF MEDIAN = 16.00 VERSUS MEDIAN N.E. 16.00

                N FOR   WILCOXON           ESTIMATED
             N   TEST  STATISTIC  P-VALUE     MEDIAN
Diff        16     16        0.0    0.000      4.800
 - - - - - - - - - - - - - - - -  Confidence interval  - - - - - - - - - - - - - - - -
Wilcoxon Signed Rank Confidence Interval

                ESTIMATED    ACHIEVED
             N     MEDIAN  CONFIDENCE  CONFIDENCE INTERVAL
Diff        16       4.80        94.8  (    3.75,    5.90)
```

**15.32.** **(a)** The Kruskal-Wallis test (Minitab output below) gives $H = 8.73$, df $= 4$, and $P = 0.069$—not significant at $\alpha = 0.05$. Note, however, that the rankings clearly suggest that vitamin C content decreases over time; the samples are simply too small to achieve significance even with such seemingly strong evidence. (See also a related comment in the solution to Exercise 15.22.) **(b)** The more accurate $P$-value is more in line with the apparent strength of the evidence, and does change our conclusion. With it, we reject $H_0$ and conclude that the distribution changes over time.

**Minitab output**

```
Kruskal-Wallis Test

LEVEL    NOBS    MEDIAN  AVE. RANK   Z VALUE
    0       2    48.705        9.5      2.09
    1       2    41.955        7.5      1.04
    3       2    21.795        5.5      0.00
    5       2    12.415        3.5     -1.04
    7       2     8.320        1.5     -2.09
OVERALL    10                  5.5

H = 8.73  d.f. = 4  p = 0.069
```

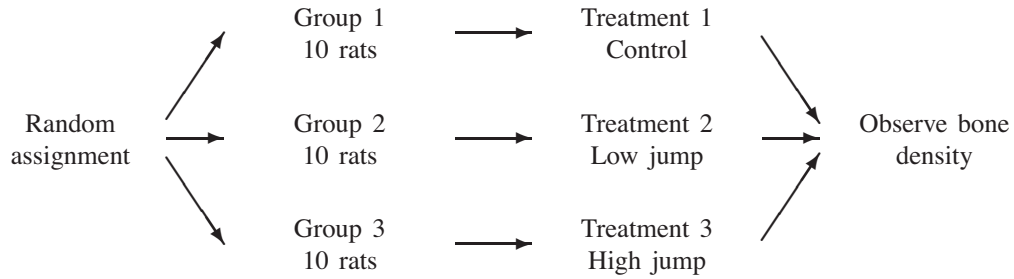
**15.33. (a)** Diagram below. **(b)** The stemplots (right) suggest greater density for high-jump rats and a greater spread for the control group. **(c)** $H = 10.66$ with $P = 0.005$. We conclude that bone density differs among the groups. ANOVA tests $H_0$: all means are equal, assuming Normal distributions with the same standard deviation. For Kruskal-Wallis, the null hypothesis is that the distributions are the same (but not necessarily Normal). **(d)** There is strong evidence that the three groups have different bone densities; specifically, the high-jump group has the highest average rank (and the highest density), the low-jump group is in the middle, and the control group is lowest.

| Control | | Low jump | | High jump | |
|---|---|---|---|---|---|
| 55 | 4 | 55 | | 55 | |
| 56 | 9 | 56 | | 56 | |
| 57 | | 57 | | 57 | |
| 58 | | 58 | 8 | 58 | |
| 59 | 33 | 59 | 469 | 59 | |
| 60 | 03 | 60 | 57 | 60 | |
| 61 | 14 | 61 | | 61 | |
| 62 | 1 | 62 | | 62 | 2266 |
| 63 | | 63 | 1258 | 63 | 1 |
| 64 | | 64 | | 64 | 33 |
| 65 | 3 | 65 | | 65 | 00 |
| 66 | | 66 | | 66 | |
| 67 | | 67 | | 67 | 4 |

**Minitab output**

```
LEVEL    NOBS    MEDIAN  AVE. RANK   Z VALUE
Ctrl       10     601.5       10.2     -2.33
Low        10     606.0       13.6     -0.81
High       10     637.0       22.6      3.15
OVERALL    30                 15.5

H = 10.66  d.f. = 2  p = 0.005
H = 10.68  d.f. = 2  p = 0.005 (adjusted for ties)
```




**15.34. (a)** For ANOVA, $H_0$: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ vs. $H_a$: Not all $\mu_i$ are equal. For Kruskal-Wallis, $H_0$ says that the distribution of the trapped insect count is the same for all board colors; the alternative is that the count is systematically higher for some colors. **(b)** In the order given, the medians are 46.5, 15.5, 34.5, and 15 insects; it appears that yellow is most effective, green is in the middle, and white and blue are least effective. The Kruskal-Wallis test statistic is $H = 16.95$, with df $= 3$; the $P$-value is 0.001, so we have strong evidence that color affects the insect count (that is, the difference we observed is statistically significant).

**Minitab output**

```
LEVEL    NOBS    MEDIAN  AVE. RANK   Z VALUE
Lemon       6     46.50       21.2      3.47
White       6     15.50        7.3     -2.07
Green       6     34.50       14.8      0.93
Blue        6     15.00        6.7     -2.33
OVERALL    24                 12.5

H = 16.95  d.f. = 3  p = 0.001
H = 16.98  d.f. = 3  p = 0.001 (adjusted for ties)
```

**15.35. (a)** $I = 4,\ n_i = 6,\ N = 24$. **(b)** The table below lists color, insect count, and rank. There are only three ties (and the second could be ignored, as both of those counts are for white boards). The $R_i$ (rank sums) are:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Yellow | 17 | + 20 | + 21 | + 22 | + 23 | + 24 | = 127 |
| White | 3 | + 4 | + 5.5 | + 9.5 | + 9.5 | + 12.5 | = 44 |
| Green | 7 | + 14 | + 15 | + 16 | + 18 | + 19 | = 89 |
| Blue | 1 | + 2 | + 5.5 | + 8 | + 11 | + 12.5 | = 40 |

**(c)** $H = \dfrac{12}{24(25)}\left(\dfrac{127^2 + 44^2 + 89^2 + 40^2}{6}\right) - 3(25) = 91.95\overline{3} - 75 = 16.95\overline{3}$.
Under $H_0$, this has approximately the chi-squared distribution with df $= I - 1 = 3$; comparing to this distribution tells us that $0.0005 < P < 0.001$.

| B | B | W | W | W | B | G | B | W | W | B | W | B | G | G | G | Y | G | G | Y | Y | Y | Y | Y |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | 11 | 12 | 13 | 14 | 14 | 15 | 16 | 17 | 17 | 20 | 21 | 21 | 25 | 32 | 37 | 38 | 39 | 41 | 45 | 46 | 47 | 48 | 59 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| | | | | 5.5 | | | | 9.5 | | | 12.5 | | | | | | | | | | | | |

**15.36. (a)** The standard deviations are (in order) 4.6043, 6.5422, 9.0443, and 16.0873 pounds. **(b)** The four medians are 126, 126, 131, and 110 pounds. The Kruskal-Wallis hypotheses are $H_0$: all medians are equal vs. $H_a$: at least one median is different. (Recall that the hypotheses can be stated in terms of the distributions or in terms of the medians.) **(c)** With $H = 5.35,\ \text{df} = 3$, and $P \doteq 0.149$, we do not have enough evidence to conclude that breaking strength differs for varying lengths of burial.

**Minitab output**

```
LEVEL     NOBS    MEDIAN  AVE. RANK   Z VALUE
    2        5     126.0        9.7     -0.35
    4        5     126.0       10.2     -0.13
    8        5     131.0       15.3      2.09
   16        5     110.0        6.8     -1.61
OVERALL     20                 10.5

H = 5.35  d.f. = 3  p = 0.149
H = 5.38  d.f. = 3  p = 0.147 (adjusted for ties)
```

**15.37.** For the Kruskal-Wallis test, we need two or more independent samples. Because these data come from different questions being asked of the same people, the responses are not independent. (We have several variables measured from a single group, rather than a single variable measured for several different groups.)

**15.38.** **(a)** The stemplots (right) appear to suggest that logging reduces the number of species per plot and that recovery is slow (the 1-year-after and 8-years-after stemplots are similar). The logged stemplots have some outliers and appear to have more spread than the unlogged stemplot. The medians are 18.5, 12.5, and 15. **(b)** For testing $H_0$: all medians are equal vs. $H_a$: at least one median is different, we have $H = 9.31$, df $= 2$, and $P = 0.010$ (or 0.009, adjusted for ties). This is good evidence of a difference among the groups.

| Unlogged | | 1 year ago | | 8 years ago | |
|---|---|---|---|---|---|
| 0 | | 0 | 2 | 0 | |
| 0 | | 0 | | 0 | 4 |
| 0 | | 0 | 7 | 0 | |
| 0 | | 0 | 8 | 0 | |
| 1 | | 1 | 11 | 1 | 0 |
| 1 | 333 | 1 | 23 | 1 | 2 |
| 1 | 55 | 1 | 4555 | 1 | 455 |
| 1 | | 1 | | 1 | 7 |
| 1 | 899 | 1 | 8 | 1 | 88 |
| 2 | 01 | 2 | | 2 | |
| 2 | 22 | 2 | | 2 | |

**Minitab output**

```
LEVEL    NOBS    MEDIAN  AVE. RANK   Z VALUE
    1      12     18.50       23.4      2.88
    2      12     12.50       11.5     -2.47
    3       9     15.00       15.8     -0.44
OVERALL    33                 17.0

H = 9.31  d.f. = 2  p = 0.010
H = 9.44  d.f. = 2  p = 0.009 (adjusted for ties)
```

**15.39.** **(a)** Yes, the data support this statement: $\frac{68}{211} \doteq 32.2\%$ of high-SES subjects have never smoked, compared to 17.3% and 23.7% of middle- and low-SES subjects (respectively). Also, only $\frac{51}{211} \doteq 24.2\%$ of high-SES subjects are current smokers, versus 42.3% and 46.2% of those in the middle- and low-SES groups. **(b)** $X^2 = 18.510$ with df $= 4$; this has $P = 0.001$ (Minitab output on the right). There is a significant relationship. **(c)** $H = 12.72$ with df $= 2$, so $P = 0.002$—or, after adjusting for ties, $H = 14.43$ and $P = 0.001$ (Minitab output below). The observed differences are significant; some SES groups smoke systematically more.

**Minitab output**

```
          Never    Former      Curr     Total
 High        68        92        51       211
          58.68     83.57     68.75

  Mid         9        21        22        52
          14.46     20.60     16.94

  Low        22        28        43        93
          25.86     36.83     30.30

Total        99       141       116       356

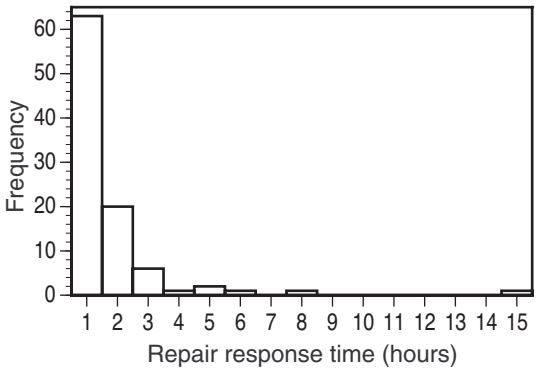
ChiSq =  1.481 +  0.850 +  4.584 +
         2.062 +  0.008 +  1.509 +
         0.577 +  2.119 +  5.320 = 18.510
df = 4, p = 0.001
```

**Minitab output**

```
LEVEL    NOBS    MEDIAN  AVE. RANK   Z VALUE
 High     211     2.000       162.4     -3.56
  Mid      52     2.000       203.6      1.90
  Low      93     2.000       201.0      2.46
OVERALL   356                 178.5

H = 12.72  d.f. = 2  p = 0.002
H = 14.43  d.f. = 2  p = 0.001 (adjusted for ties)
```

**15.40.** **(a)** On the right is a histogram of service times for Verizon customers. With only 10 CLEC service calls, it is hardly necessary to make such a graph for them; we can simply observe that 7 of those 10 calls took 5 hours, which is quite different from the distribution for Verizon customers. The means and medians tell the same story: For Verizon customers, $\bar{x}_V \doteq 1.7263$ hr and $M_V = 1$ hr, while for CLEC customers, $\bar{x}_C = 3.8$ hr and $M_C = 5$ hr. **(b)** The distributions are sharply skewed, and the sample sizes are quite different; the $t$ test is not reliable in situations like this. The Wilcoxon rank-sum test gives $W = 4778.5$, which is highly significant ($P = 0.0026$ or 0.0006). We have strong evidence that response times for Verizon customers are shorter. It is also possible to apply the Kruskal-Wallis test (with two groups). While the $P$-values are slightly different ($P = 0.005$, or 0.001 adjusted for ties), the conclusion is the same: We have strong evidence of a difference in response times.

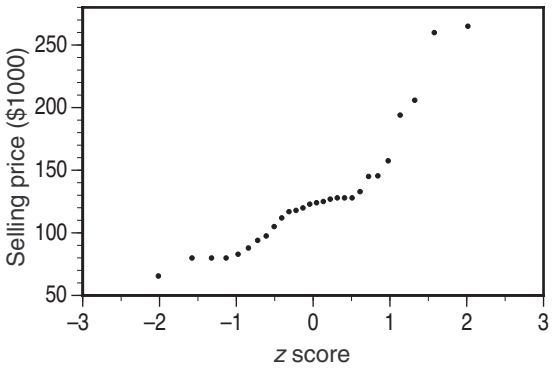



**Minitab output**

```
- - - - - - - - - - - - - - -   Wilcoxon rank-sum test   - - - - - - - - - - - - - - -
Verizon    N =  95     Median =       1.000
CLEC       N =  10     Median =       5.000
W = 4778.5
Test of ETA1 = ETA2  vs.  ETA1 < ETA2 is significant at 0.0026
The test is significant at 0.0006 (adjusted for ties)
 - - - - - - - - - - - - - - - -   Kruskal-Wallis test   - - - - - - - - - - - - - - - -
LEVEL    NOBS    MEDIAN  AVE. RANK   Z VALUE
    1      95     1.000       50.3     -2.80
    2      10     5.000       78.7      2.80
OVERALL   105                 53.0

H =  7.84  d.f. = 1  p = 0.005
H = 10.54  d.f. = 1  p = 0.001 (adjusted for ties)
```

**15.41.** See also the solution to Exercise 7.141. **(a)** The distribution of prices for three-bedroom houses is clearly right-skewed, with high outliers. **(b)** For testing $H_0$: $\mu_3 = \mu_4$ vs. $H_a$: $\mu_3 \neq \mu_4$, we have $t \doteq -3.08$ with either df $= 12.1$ ($P = 0.0095$) or df $= 8$ ($P = 0.0151$). We conclude that the mean prices are different (specifically, that 4BR houses are more expensive). Summary statistics are below. **(c)** We test $H_0$: medians are equal vs. $H_a$: medians are different, for which $W = 447$ and $P \doteq 0.0028$—significant evidence that prices differ. This is equivalent to the conclusion reached in (b).



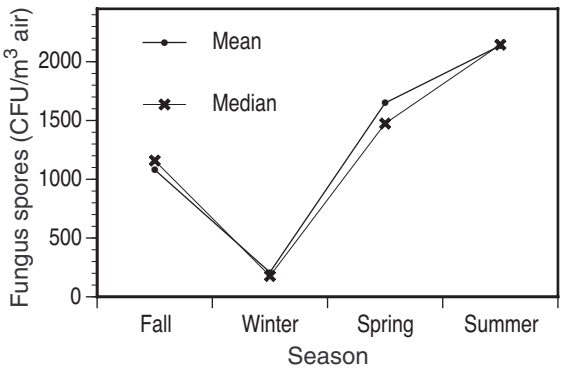
**Minitab output**

```
BR3        N =  28     Median =      123450
BR4        N =   9     Median =      176900
W = 447.0
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.0028
The test is significant at 0.0028 (adjusted for ties)
```

| | $n$ | $\bar{x}$ | $s$ | Min | $Q_1$ | $M$ | $Q_3$ | Max |
|---|---|---|---|---|---|---|---|---|
| 3BR | 28 | \$129,546 | \$49,336 | \$65,500 | \$95,750 | \$123,450 | \$138,950 | \$265,000 |
| 4BR | 9 | \$194,944 | \$57,204 | \$121,900 | \$148,450 | \$176,900 | \$240,000 | \$294,000 |

**15.42.** **(a)** Means and standard deviations are in the table below; medians are in the Minitab output.

| | $\bar{x}$ | $s$ |
|---|---|---|
| Fall | 1081.25 | 231.46 |
| Winter | 209 | 136.57 |
| Spring | 1648.75 | 1071.24 |
| Summer | 2138.5 | 906.43 |




Graphical displays may vary; one approach is the plot on the right, which shows how means and medians change over the seasons.
**(b)** Using the Kruskal-Wallis test, we find $H = 9.95$, df $= 3$, $P = 0.019$—significant evidence at $\alpha = 0.05$ that spore count varies over the seasons.

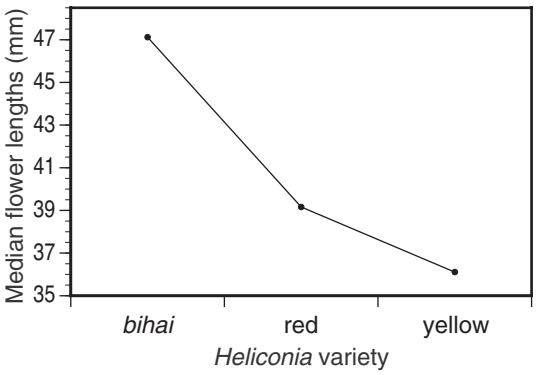
**Minitab output**

```
LEVEL     NOBS    MEDIAN  AVE. RANK   Z VALUE
  Fall       4    1159.5        8.8      0.12
Winter       4     177.5        2.5     -2.91
Spring       4    1473.5       10.0      0.73
Summer       4    2144.5       12.8      2.06
OVERALL     16                 8.5

H = 9.95  d.f. = 3  p = 0.019
```

**15.43.** See also the solutions to Exercises 1.79 and 12.35; the latter exercise requests the same analysis for ANOVA. Medians are given in the Minitab output (on the next page) and in the plot on the right. The means and standard deviations (all in millimeters) are:

| Variety | $n$ | $\bar{x}$ | $s$ |
|---|---|---|---|
| *bihai* | 16 | 47.5975 | 1.2129 |
| red | 23 | 39.7113 | 1.7988 |
| yellow | 15 | 36.1800 | 0.9753 |




We reject $H_0$ and conclude that at least one species has different lengths ($H = 45.35$, df $= 2$, $P < 0.0005$).

**Minitab output**

```
LEVEL     NOBS     MEDIAN  AVE. RANK   Z VALUE
    1       16      47.12        46.5      5.76
    2       23      39.16        26.7     -0.32
    3       15      36.11         8.5     -5.51
OVERALL     54                   27.5

H = 45.35  d.f. = 2  p = 0.000
H = 45.36  d.f. = 2  p = 0.000 (adjusted for ties)
```

**15.44.** **(a)** The mean and median suggest that iron content is least for aluminum pots and greatest for iron pots. ANOVA requires Normal data with equal standard deviations; the former is difficult to assess with such small samples, and for the latter, the largest-to-smallest ratio is 1.99—just within our guidelines for pooling. **(b)** The Kruskal-Wallis test gives $H = 8.00$, df $= 2$, $P = 0.019$. We conclude that vegetable iron content differs by pot type.

| | $n$ | $M$ | $\bar{x}$ | $s$ |
|---|---|---|---|---|
| Aluminum | 4 | 1.185 | 1.2325 | 0.2313 |
| Clay | 4 | 1.615 | 1.46 | 0.4601 |
| Iron | 4 | 2.79 | 2.79 | 0.2399 |

**Minitab output**

```
LEVEL     NOBS     MEDIAN  AVE. RANK   Z VALUE
 Alum        4      1.185        3.5     -2.04
 Clay        4      1.615        5.5     -0.68
 Iron        4      2.860       10.5      2.72
OVERALL     12                   6.5

H = 8.00  d.f. = 2  p = 0.019
```

**15.45.** Use the Wilcoxon rank sum test with a two-sided alternative. For meat, $W = 15$ and $P = 0.4705$, and for legumes, $W = 10.5$ and $P = 0.0433$ (or 0.0421). There is no evidence of a difference in iron content for meat, but for legumes the evidence is significant at $\alpha = 0.05$.

**Minitab output**

```
- - - - - - - - - - - - - - - - - - -  Meat  - - - - - - - - - - - - - - - - - - -
Alum       N =   4     Median =       2.050
Clay       N =   4     Median =       2.375
W = 15.0
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.4705
Cannot reject at alpha = 0.05
- - - - - - - - - - - - - - - - - - -  Legumes  - - - - - - - - - - - - - - - - - -
Alum       N =   4     Median =      2.3700
Clay       N =   4     Median =      2.4550
W = 10.5
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.0433
The test is significant at 0.0421 (adjusted for ties)
```

**15.46.** Using a Kruskal-Wallis test, we find $H = 9.85$, df $= 2$, and $P = 0.007$. We conclude that there is a difference in iron content for foods cooked in iron pots.

**Minitab output**

```
LEVEL     NOBS     MEDIAN  AVE. RANK   Z VALUE
 Meat        4      4.695       10.5      2.72
  Leg        4      3.705        6.5      0.00
  Veg        4      2.860        2.5     -2.72
OVERALL     12                   6.5

H = 9.85  d.f. = 2  p = 0.007
```