7.15. (a) df = 14, t^* = 2.145. **(b)** df = 27, t^* = 2.052. **(c)** df = 27, t^* = 1.703. **(d)** As sample size increases (comparing a and b), the margin of error decreases. As confidence increases (comparing c and b), the margin of error increases.

7.20. (a) df = $n - 1 = 21$. **(b)** $2.189 < t < 2.518$. **(c)** $0.01 < P\text{-value} < 0.02$. **(d)** $t = 2.24$ is significant at the 5% level but not at the 1% level. **(e)** From software, P-value = 0.0180.

7.21. (a) df = $n - 1 = 12$. **(b)** $2.681 < t < 3.055$. **(c)** Because the alternative is two-sided, we double the upper-tail probability to find the P-value: $0.01 < P\text{-value} < 0.02$. **(d)** $t = 2.78$ is significant at the 5% level but not at the 1% level. **(e)** From software, P-value = 0.0167.

7.22. (a) df = $n - 1 = 8$. **(b)** Because $1.397 < |t| < 1.860$, the P-value is between $0.05 < P\text{-value} < 0.10$. **(c)** From software, P-value = 0.0507.

7.30. (a) df = 74, so using 60 from the table gives $t^* = 2.000$, so the 95% confidence interval is 28.5 ± $2.000(23.1/\sqrt{75}) = (23.17, 33.83)$. **(b)** Because the interval contains 32.5 hours, we cannot reject the average reported by Nielsen as a possible value of μ.

7.31. (a) If μ is the mean number of uses a person can produce in 5 minutes after witnessing rudeness, we wish to test H_0: $\mu = 10$ versus H_a: $\mu < 10$. **(b)** $t = \dfrac{7.88 - 10}{2.35 / \sqrt{34}} = -5.2603$, with df = 33, for which P-value < 0.0005. This is strong evidence that witnessing rudeness decreases performance.

7.32. (a) A normal quantile plot shows the data is roughly normally distributed; the use of t methods seems to be appropriate. **(b)** We have $\bar{x} = 43.17$ mpg, and $s = 4.4149$ mpg; the standard error is $s/\sqrt{20} =$ 0.9872. For df = 19, $t^* = 2.093$, so the margin of error is 2.0662. **(c)** The 95% confidence interval is (41.104, 45.236).

7.40. (a) For the differences, $\overline{x}$ = \$114 and s = \$114.402. **(b)** We wish to test H_0: $\mu = 0$ versus H_a: $\mu > 0$, where μ is the mean difference between Jocko's estimates and those of the other garage. (The alternative hypothesis is one-sided because the insurance adjusters suspect that Jocko's estimates may be too high.) For this test, we find $t = \frac{114 - 0}{114.4 / \sqrt{10}} = 3.15$ with df = 9, for which 0.005 < P-value < 0.001 (software gives 0.0059). This is significant evidence against H_0—that is, we have good reason to believe that Jocko's estimates are higher. **(c)** Using df = 9, $t^* = 2.262$, the 95% confidence interval is 114 ± 81.83 = (\$32.17, \$195.83). **(d)** Student answers may vary; based on the confidence interval, one could justify any answer in the range \$32.17 to \$195.83.

7.41. (a) H_0: $\mu = 0$ versus H_a: $\mu \neq 0$. **(b)** With mean difference $\overline{x} = 2.73$ and standard deviation s = 2.8015, the test statistic is $t = \frac{2.73 - 0}{2.8015 / \sqrt{20}} = 4.358$ with df = 19, for which P-value < 0.001 (software gives 0.0003). We have strong evidence that the results of the two computations are different.

7.42. (a) The histogram, boxplot, and normal quantile plot reveals that the distribution is Normal. The five-number summary is 886, 919.5, 936.5, 958, 986. **(b)** Because the data is a Normal distribution, we can use t procedures. **(c)** $\overline{x}$ = 938.2, s = 24.2971. SE = $24.2971/\sqrt{36}$ = 4.0495 **(d)** df = 35, so using 30 we have t^* = 1.697; thus, the 90% confidence interval is 938.2 ± 1.697(4.0495) = (931.328, 945.072).

7.43. (a) To test H_0: μ = 925 picks versus H_a: μ > 925 picks, we have $t = \frac{938.2 - 925}{24.2971 / \sqrt{36}} = 3.27$

with df = 35, for which P-value = 0.0012. **(b)** For H_0: μ = 935 picks versus H_a: μ > 935 picks, we have $t = \frac{938.2 - 935}{24.2971 / \sqrt{36}} = 0.80,$ again with df = 35, for which P-value = 0.2146. **(c)** The 90% confidence interval from the previous exercise was 931.3 to 945.0 picks, which includes 935 but not 925. For a test of H_0: $\mu = \mu_0$ versus H_a: $\mu \neq \mu_0$, we know that P-value < 0.10 for values of μ_0 outside the interval and P-value > 0.10 if μ_0 is inside the interval. The one-sided P-value would be half of the two-sided P-value.

7.44. We have df = 1567 (use 1000 if using the tables), $t^* = 1.646$. The 90% confidence interval is 3.83 ± $1.646(1.10/\sqrt{1568}) = (3.7843, 3.8757)$.

7.57. (a) We cannot reject H_0: $\mu_1 = \mu_2$ in favor of the two-sided alternative at the 5% level because 0.05 < P-value < 0.10 (0.0771 from software). **(b)** We could reject H_0 in favor of H_a: $\mu_1 < \mu_2$. A negative t statistic means that $\bar{x}_1 < \bar{x}_2$, which supports the claim that $\mu_1 < \mu_2$, and the one-sided P-value would be half of its value from part **(a)**: 0.025 < P-value < 0.05 (0.0386 from software).

7.58. For all tests we have H_0: $\mu_1 = \mu_2$ versus H_a: $\mu_1 \neq \mu_2$. For sprint speed: $t = \dfrac{27.3 - 26.0}{\sqrt{\dfrac{(0.7)^2}{16} + \dfrac{(1.5)^2}{13}}} = 2.88.$

df = 12, 0.01 < P-value < 0.02. There is evidence of a significant difference in sprint speed between the elite players and the university players. For peak heart rate: $t = \dfrac{192.0 - 193.0}{\sqrt{\dfrac{(6.0)^2}{16} + \dfrac{(6.0)^2}{13}}} = -0.45.$ df = 12, P-value > 0.50. There is not enough evidence to show a difference in peak heart rate between the elite players and the university players. For intermittent recovery test: $t = \dfrac{1160 - 781}{\sqrt{\dfrac{(191)^2}{16} + \dfrac{(129)^2}{13}}} = 6.35.$ df = 12, 0.01 < P-value < 0.02. There is evidence of a significant difference in the intermittent recovery test between the elite players and the university players.

7.68. (a) For testing H_0: $\mu_{\text{LC}} = \mu_{\text{LF}}$ versus H_a: $\mu_{\text{LC}} \neq \mu_{\text{LF}}$, we have $t = 4.165$, so P-value < 0.001 (<0.0001 from software), we clearly reject H_0. **(b)** It might be that the moods of subjects who dropped out differed from the moods of those who stayed; in particular, it seems reasonable to suspect that those who dropped out had higher TMDS scores (more negative moods).

7.69. (a) The problem with averages on rating is that there is no guarantee the differences between ratings are equal, so that going from a rating of 1 to 2, 2 to 3, etc., are equal. Taking averages assumes this, so it is likely not appropriate. **(b)** The data are ratings from 1–5; as such, they certainly will not be Normally distributed, but because $n_1 + n_2 \geq 40$ and outliers are not possible, the t procedures can be used. **(c)** McDonald's: $\overline{X} = 3.9937$, $s = 0.8930$; Taco Bell: $\overline{X} = 4.2208$, $s = 0.7331$. **(d)** H_0: $\mu_M = \mu_T$, H_a: $\mu_M \neq \mu_T$. $t = -3.48$. df = 307. P-value < 0.001 (0.0005 from software). The data are significant at the 5% level, and there is evidence the average customer ratings between the two chains is different.

7.70. (a) Use a two-sided alternative (H_a: $\mu_A \neq \mu_B$) because we (presumably) have no prior suspicion that one design will be better than the other. **(b)** Both sample sizes are the same ($n_1 = n_2 = 15$), so the appropriate degrees of freedom would be df = 15 − 1 = 14. **(c)** For a two-sided test at $\alpha = 0.05$, we need $|t| > t^*$, where $t^* = 2.145$ is the 0.025 critical value for a t distribution with df = 14.

7.71. (a) Assuming we have SRSs from each population, this seems reasonable. **(b)** H_0: $\mu_{\text{Early}} = \mu_{\text{Late}}$, H_a: $\mu_{Early} \neq \mu_{Late}$. **(c)** $\text{SE}_\text{D} = 1.0534$, $t = 1.614$, df = 199, P-value = 0.1081 (0.1075 from software). We fail to detect a difference in mean fat consumption for the two groups. **(d)** The 95% confidence interval for the difference in mean fat consumption is (−0.39, 3.79). Software gives (–0.372, 3.772). We note that the intervals contain 0; these support the results of the test.

7.72. For both protein and carbohydrates, we have hypotheses H_0: $\mu_{\text{Early}} = \mu_{\text{Late}}$ and H_a: $\mu_{Early} \neq \mu_{Late}$. For protein, we have $\text{SE}_\text{D} = 0.7729$, $t = 2.458$. Using df = 100, 0.01 < P-value < 0.02 (0.0144 from software). There is a difference in mean protein consumption between early and late eaters. The 95% confidence interval for the difference is (0.367, 3.433). For carbohydrates, $\text{SE}_\text{D} = 2.0848$, $t = 0.288$. Using df = 100, P-value > 0.50 (0.7737 from software). We fail to detect a difference in mean carbohydrate consumption. The 95% confidence interval for the difference is (−3.536, 4.736). Software gives (–3.499, 4.699). There was a difference in protein consumption for early and late eaters, but there was no significant difference for fat and carbohydrate consumption.

7.73. To find a confidence interval $(\overline{x}_1 - \overline{x}_2) \pm t^*\text{SE}_\text{D}$, we need one of the following:

- Sample sizes and standard deviations, in which case we could find the interval in the usual way
- t and df because $t = (\bar{x}_1 - \bar{x}_2)/\text{SE}_\text{D}$, so we could compute SE_D and use df to find t^*
- df and a more accurate *P*-value from which we could determine *t*, and then proceed as above

The confidence interval could give us useful information about the magnitude of the difference (although, with such a small *P*-value, we do know that a 95% confidence interval would not include 0).

7.74. (a) The 68–95–99.7 rule suggests that the distributions are not Normal: If they were Normal, then (for example) 95% of 7- to 10-year-olds drink between −13.2 and 29.6 oz. of sweetened drinks per day. As negative numbers do not make sense, the distributions must be right-skewed. **(b)** We find SE_D = 4.3786 and $t = -1.439$, with df = 4, 0.20 < *P*-value < 0.30 (0.1890 from software using df = 7.8). We do not have enough evidence to reject H_0. There is insufficient evidence to say that one age group on average drinks more sweetened drinks than the other. **(c)** The 95% confidence interval is (−18.4551, 5.8551). Software gives (–16.4404, 3.8404). **(d)** Because the distributions are not Normal and the samples are small, the *t* procedures are *very* questionable for these data. **(e)** Because this group is not an SRS—and indeed might not be random in any way—we would have to be cautious about extending these results to other children.

7.75. This is a matched pairs design; for example, Monday hits are (at least potentially) not independent of one another. The correct approach would be to use one-sample *t* methods on the seven differences (Monday hits for design 1 minus Monday hits for design 2, Tuesday/1 minus Tuesday/2, and so on).

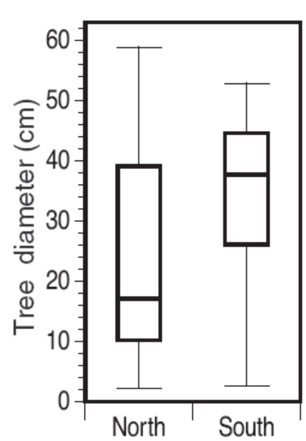
7.76. (a) $4.3 - 3.7 \pm 2.365\sqrt{\frac{(0.7)^2}{8} + \frac{(1.5)^2}{8}} = (-0.78, 1.98)$. **(b)** We fail to reject H_0 because 0 is within the confidence interval.

7.77. There could be things that are similar about the next eight employees who need new computers as well as the following eight, which could bias the results (like being from the same office or department).

7.78. (a) The null hypothesis is H_0: $\mu_1 = \mu_2$; the alternative can be either two- or one-sided. (It might be a reasonable expectation that $\mu_1 > \mu_2$.) We find SE_D = 0.2796 and $t = 8.369$. Regardless of the df and H_a, the conclusion is the same: *P*-value is very small, and we conclude that *WSJ* ads are more trustworthy. **(b)** The 95% confidence interval is (1.78, 2.90); the difference in trustworthiness is between 1.78 and 2.9 points. **(c)** Advertising in *WSJ* is seen as more reliable than advertising in the *National Enquirer*—a conclusion that probably comes as a surprise to no one.

7.79. (a) Stemplots and boxplots are shown on the right. The north distribution is right-skewed, while the south distribution is left-skewed. **(b)** The methods of this section seem to be appropriate in spite of the skewness because the sample sizes are relatively large, and there are no outliers in either distribution. **(c)** We test H_0: $\mu_\text{N} = \mu_\text{S}$ versus H_a: $\mu_\text{N} \neq \mu_\text{S}$; we should use a two-sided alternative because we have no reason (before looking at the data) to expect a difference in a particular direction. **(d)** The means and standard deviations are $\bar{x}_N = 23.7$, $s_N = 17.5001$, $\bar{x}_S = 34.53$, and $s_S = 14.2583$ cm. Then, SE_D =

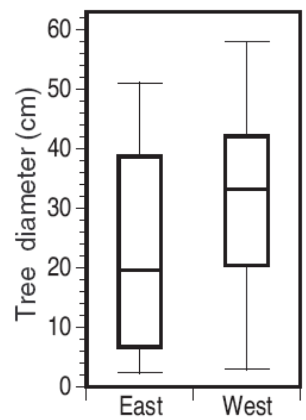
North		South
43322	0	2
65	0	57
443310	1	2
955	1	8
	2	13
8755	2	689
0	3	2
996	3	566789
43	4	003444
6	4	578
4	5	0112
85	5	



4.1213, and $t = -2.629$ with df = 29, $0.01 < P\text{-value} < 0.02$ (0.011 from software). We conclude that the means are different. **(e)** The 95% confidence interval is (−19.2614, −2.4053) [software gives (–19.0902, –2.5765)]. The interval not only tells us that a difference exists, but that the northern trees are, on average, between about 2.4 and 19 cm smaller in dbh than the trees in the southern part of the tract.

7.80. (a) Stemplots and boxplots are shown on the right. The east distribution is right-skewed, while the west distribution is left-skewed. **(b)** The methods of this section seem to be appropriate in spite of the skewness because the sample sizes are relatively large, and there are no outliers in either distribution. **(c)** We test H_0: $\mu_E = \mu_W$ versus H_a: $\mu_E \neq \mu_W$; we should use a two-sided alternative because we have no reason (before looking at the data) to expect a difference in a particular direction. **(d)** The means and standard deviations are $\bar{x}_E = 21.716$,

East		West
222	0	233
9566655	0	
3100	1	11
7	1	78
33222	2	0011
	2	55
11	3	00
98	3	555669
333	4	023444
86	4	1
1	5	78



$s_E = 16.0743$, $\bar{x}_W = 30.283$, and $s_W = 15.3314$ cm. Then, $SE_D = 4.0556$, so $t = -2.112$ with df = 29, $0.04 < P\text{-value} < 0.05$ (0.0434 from software). We conclude that the means are different. **(e)** The 95% confidence interval is (−16.8604, −0.2730) [software gives (–16.6852, –0.4481)]. The interval not only tells us that a difference exists, but that the eastern trees are, on average, between about 0.3 and 16.8 cm smaller in dbh than the trees in the western part of the tract.

7.81. (a) Using df = 50, the 95% confidence interval is (−1.07, 7.07). **(b)** With 95% confidence, the mean change in sales from last year to this year is between –1.07 and 7.07. Because the interval covers 0 and includes some negative values, it is possible sales have actually decreased.

7.82. (a) Good statistical practice dictates that the alternative hypothesis should be chosen without looking at the data; we should only choose a one-sided alternative if we have some reason to expect it *before* looking at the sample results. **(b)** The correct P-value is twice that reported, so 0.12.

7.83. (a) We test H_0: $\mu_B = \mu_F$ versus H_a: $\mu_B > \mu_F$. $SE_D = 0.5442$ and $t = 1.654$, using df = 18 we have $0.05 < P\text{-value} < 0.10$ (0.0532 from software); there is not quite enough evidence to reject H_0 at $\alpha = 0.05$. **(b)** The 95% confidence interval is (−0.2434, 2.0434). Software gives (–0.2021, 2.0021). **(c)** We need two independent SRSs from Normal populations.

7.84. See the solution to **Exercise 7.67** for a table of means and standard deviations. H_0: $\mu_N = \mu_S$, H_a: $\mu_N \neq \mu_S$. The pooled standard deviation is $s_p = 1.0454$, so the pooled standard error is $s_p\sqrt{1/14+1/17} =$ 0.3773. The test statistic is $t = -4.098$ with df = 29, for which $P = 0.0002$, and the 95% confidence interval (with $t^* = 2.045$) is −2.3178 to −0.7747. In the solution to **Exercise 7.67**, we reached the same conclusion on the significance test ($t = -4.303$ and $P = 0.0001$), and the confidence interval was quite similar (roughly −2.3 to −0.8).

7.85. The pooled standard deviation is $s_p = 27.06$, so the pooled standard error is $s_p\sqrt{1/32+1/33} =$ 6.713. The test statistic is $t = 4.17$, df = 63, $P\text{-value} < 0.001$. The 95% confidence interval is (14.57, 41.43). The results are similar.