Chapter 7 Solutions

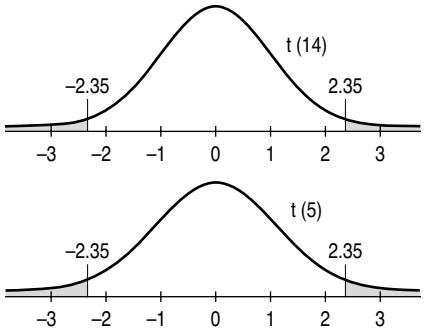
7.1. (a) The standard error of the mean is $\frac{s}{\sqrt{n}} = \frac{\$570}{\sqrt{15}} \doteq \$27.1109$. **(b)** The degrees of freedom are $\text{df} = n - 1 = 14$.

7.2. In each case, use $\text{df} = n - 1$; if that number is not in Table D, drop to the *lower* degrees of freedom. **(a)** For 95% confidence and $\text{df} = 11$, use $t^* = 2.201$. **(b)** For 99% confidence and $\text{df} = 23$, use $t^* = 2.807$. **(c)** For 90% confidence and $\text{df} = 199$, we drop to $\text{df} = 100$ and use $t^* = 1.660$. (Software gives $t^* = 1.6525$ for $\text{df} = 199$.)

7.3. For the mean monthly rent, the 95% confidence interval is $\$570 \pm (2.145)(\$105/\sqrt{15}) = \$570 \pm \$58.15 = \$511.85$ to $\$628.15$.

7.4. The margin of error for 90% confidence would be smaller—and the interval would be narrower—because we are taking a greater risk that the interval does *not* include the true mean μ.

7.5. (a) Yes, $t = 2.35$ is significant when $n = 15$. This can be determined either by comparing to the $\text{df} = 14$ line in Table D (where we see that $t > 2.264$, the 2% critical value) or by computing the two-sided P-value (which is $P = 0.0340$). **(b)** No, $t = 2.35$ is not significant when $n = 6$, as can be seen by comparing to the $\text{df} = 5$ line in Table D (where we see that $t < 2.571$, the 2.5% critical value) or by computing the two-sided P-value (which is $P = 0.0656$). **(c)** Student sketches will likely be indistinguishable from Normal distributions; careful students may try to show that the $t(5)$ distribution is shorter in the center and heavier to the left and right than the $t(14)$ distribution (as is the case here), but in reality, the difference is nearly imperceptible.



7.6. For the hypotheses H_0: $\mu = \$550$ vs. H_a: $\mu > \$550$, we find $t = \frac{570 - 550}{105/\sqrt{15}} \doteq 0.7377$ with $\text{df} = 14$, for which $P \doteq 0.2364$. We do not reject H_0; we do not have sufficient evidence to conclude that the mean rent is greater than \$550.

Note: *Computing a P-value is not really necessary; a t(14) distribution is similar enough to a Normal distribution that we can quickly determine that $t \doteq 0.74$ gives no reason to doubt the null hypothesis.*

7.7. Software will typically give a more accurate value for t^* than that given in Table D, but otherwise the details of this computation are the same as what is shown in the textbook: $\text{df} = 7$, $t^* = 2.3646$, $5 \pm t^*(3.63/\sqrt{8}) = 5 \pm 3.0348 = 1.9652$ to 8.0348, or about 2.0 to 8.0 hours per month.

7.8. We wish to test H_0: $\mu = 0$ vs. H_a: $\mu \neq 0$, where μ is the mean of (drink A rating) minus (drink B rating). Compute the differences for each subject (-2, 1, 5, 11, and 8), then find the mean and standard deviation of these differences: $\bar{x} \doteq 4.6$ and $s \doteq 5.2249$. Therefore, $t = \frac{4.6-0}{5.2249/\sqrt{5}} \doteq 1.9686$ with df $= 4$, for which $P = 0.1204$. We do not have enough evidence to conclude that there is a difference in preference.

7.9. Using the mean and standard deviation from the previous exercise, the 95% confidence interval is $4.6 \pm (2.7764)(5.2249/\sqrt{5}) = 4.6 \pm 6.4876 =$ -1.8876 to 11.0876.

7.10. See also the solutions to Exercises 1.30, 1.64 and 1.144. The CO_2 data are sharply right-skewed (clearly non-Normal). However, the robustness of the t procedures should make them safe for this situation because the sample size is large ($n = 48$). The bigger question is whether we can treat the data as an SRS; we have recorded CO_2 emissions for every country with a population over 20 million, rather than a random sample.

7.11. The sample size ($n = 90$) should be sufficient to overcome any non-Normality, especially in the absence of strong skewness. One might question the independence of 90 consecutive measurements.

7.12. The power would be lower because it is easier to detect larger differences (like $\mu = 1$) than smaller differences (like $\mu < 1$).

7.13. As was found in Example 7.9, we reject H_0 if $t = \frac{\bar{x}}{1.5/\sqrt{20}} \geq 1.729$, which is equivalent to $\bar{x} \geq 0.580$. The power we seek is $P(\bar{x} \geq 0.580$ when $\mu = 0.75)$, which is:

$$P\left(\frac{\bar{x} - 0.75}{1.5/\sqrt{20}} \geq \frac{0.580 - 0.75}{1.5/\sqrt{20}}\right) = P(Z \geq -0.51) = 0.6950$$

7.14. We test H_0: median $= 0$ vs. H_a: median $\neq 0$—or equivalently, H_0: $p = 1/2$ vs. H_a: $p \neq 1/2$, where p is the probability that the rating for drink A is higher. We note that four of the five differences are positive (and none are 0). The P-value is $2P(X \geq 4) = 0.375$ from a $B(5, 0.5)$ distribution; there is not enough evidence to conclude that the median ratings are different.

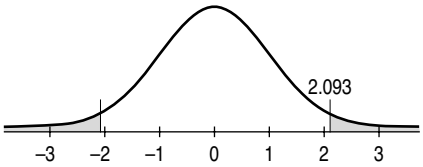
Minitab output

```
Sign test of median = 0.00000 versus  N.E.  0.00000

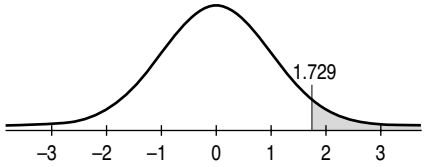
              N  BELOW  EQUAL  ABOVE   P-VALUE     MEDIAN
Diff          5      1      0      4    0.3750      5.000
```

7.15. **(a)** df $= 14$, $t^* = 2.145$. **(b)** df $= 24$, $t^* = 2.064$. **(c)** df $= 24$, $t^* = 1.711$. **(d)** For a given confidence level, t^* (and therefore the margin of error) decreases with increasing sample size. For a given sample size, t^* increases with increasing confidence.

7.16. This t distribution has df $= 19$. The 2.5% critical value is 2.093, so we reject H_0 when $t < -2.093$ or $t > 2.093$.

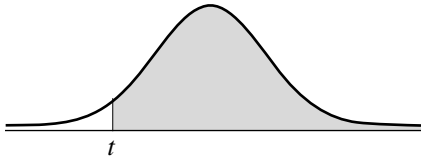


7.17. The 5% critical value for a t distribution with df $= 19$ is 1.729. Only one of the one-sided options (reject H_0 when $t > 1.729$) is shown; the other is simply the mirror image of this sketch (shade the area to the left of -1.729, and reject when $t < -1.729$).



7.18. Because the value of $\bar{x}$ is positive, which supports the direction of the alternative ($\mu > 0$), the P-value for the one-sided test is half as big as that for the two-sided test: $P = 0.02$.

7.19. $\bar{x} = -15.3$ would support the alternative $\mu < 0$, and for that alternative, the P-value would still be 0.02. For the alternative $\mu > 0$ given in Exercise 7.18, the P-value is 0.98. Note that in the sketch shown, no scale has been given, because in the absence of a sample size, we do not know the degrees of freedom. Nevertheless, the P-value for the alternative $\mu > 0$ is the area above the computed value of the test statistic t, which will be the opposite of that found when $\bar{x} = 15.3$. As the area below t is 0.02, the area above this point must be 0.98.



7.20. **(a)** df $= 19$. **(b)** $2.093 < t < 2.205$. **(c)** $0.02 < P < 0.025$. **(d)** $t = 2.10$ is significant at 5%, but not at 1%. **(e)** From software, $P \doteq 0.0247$.

7.21. **(a)** df $= 23$. **(b)** $2.177 < t < 2.500$. **(c)** Because the alternative is two-sided, we double the upper-tail probabilities to find the P-value: $0.02 < P < 0.04$. **(d)** $t = 2.40$ is significant at 5%, but not at 1%. **(e)** From software, $P \doteq 0.0248$.

7.22. **(a)** df $= 114$. **(b)** Using Table D, we refer to df $= 100$. Because $1.290 < |t| < 1.660$, the P-value is between $0.05 < P < 0.10$. **(c)** From software, $P \doteq 0.0619$.

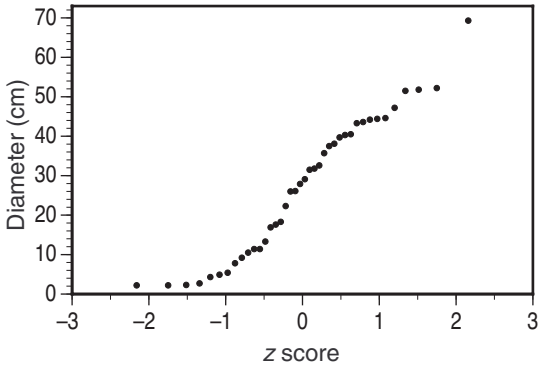
7.23. Let P be the given (two-sided) P-value, and suppose that the alternative is $\mu > \mu_0$. If $\bar{x}$ is greater than μ_0, this supports the alternative over H_0. However, if $\bar{x} < \mu_0$, we would not take this as evidence against H_0 because $\bar{x}$ is on the "wrong" side of μ_0. So, if the value of $\bar{x}$ is on the "correct" side of μ_0, the one-sided P-value is simply $P/2$. However, if the value of $\bar{x}$ is on the "wrong" side of μ_0, the one-sided P-value is $1 - P/2$ (which will always be at least 0.5, so it will never indicate significant evidence against H_0).

7.24. **(a)** A stemplot (shown) or a histogram shows no outliers and no particular skewness. (In fact, for such a small sample, it suggests no striking deviations from Normality.) The use of t methods seems to be safe. **(b)** The mean is $\bar{x} \doteq 43.17$ mpg, the standard deviation is $s \doteq 4.4149$ mpg, and the standard error is $s/\sqrt{20} \doteq 0.9872$ mpg. For df $= 19$, the 2.5% critical value is $t^* \doteq 2.093$, so the margin of error is $t^* s/\sqrt{20} \doteq 2.0662$ mpg. **(c)** The 95% confidence interval is 41.1038 to 45.2362 mpg.

3	4
3	677
3	9
4	1
4	23333
4	445
4	667
4	88
5	0

7.25. **(a)** A stemplot (right) reveals that the distribution has two peaks and a high value (not quite an outlier). Both the stemplot and quantile plot show that the distribution is not Normal. The five-number summary is 2.2, 10.95, 28.5, 41.9, 69.3 (all in cm); a boxplot is not shown, but the long "whisker" between Q_3 and the maximum is an indication of the skewness. **(b)** Maybe: We have a large enough sample to overcome the non-Normal distribution, but we are sampling from a small population. **(c)** The mean is $\bar{x} = 27.29$ cm, $s \doteq 17.7058$ cm, and the margin of error is $t^* \cdot s/\sqrt{40}$:

0	222244
0	579
1	0113
1	678
2	2
2	6679
3	112
3	5789
4	0033444
4	7
5	112
5	
6	
6	9



	df	t^*	Interval
Table D	30	2.042	$27.29 \pm 5.7167 = 21.57$ to 33.01 cm
Software	39	2.0227	$27.29 \pm 5.6626 = 21.63$ to 32.95 cm

(d) One could argue for either answer. We chose a random sample from this tract, so the main question is, can we view trees in this tract as being representative of trees elsewhere?

7.26. **(a)** The distribution is sharply right-skewed, with two or three high outliers. **(b)** Means are typically *not* the best measure of center for skewed distributions. **(c)** The mean CRP is $\bar{x} \doteq 10.0323$ mg/l, $s \doteq 16.5632$ mg/l, and the margin of error is $t^* \cdot s/\sqrt{40}$:

	df	t^*	Interval
Table D	30	2.042	$\bar{x} \pm 5.3477 = 4.68$ to 15.38 mg/l
Software	39	2.0227	$\bar{x} \pm 5.2972 = 4.74$ to 15.33 mg/l

The skewness of the distribution makes this methodology somewhat suspect.

0	00000000000000003334
0	55555677899
1	2
1	5
2	02
2	6
3	0
3	
4	
4	6
5	
5	9
6	
6	
7	3

7.27. Shown are the results for taking *natural* logarithms; common (base 10) logarithms would differ by a factor of about 0.434. **(a)** The distribution is still skewed to the right, but considerably less so than the original data. There do not appear to be any outliers in the transformed data (although the group of 0s stands out). **(b)** Means are typically *not* the best measure of center for skewed distributions. **(c)** The mean log-CRP is $\bar{x} \doteq 1.4952$, $s \doteq 1.3912$, and the margin of error is $t^* \cdot s/\sqrt{40}$:

0	0000000000000000	
0		
1		
1	555788888	
2	001233	
2	58	
3	0134	
3	8	
4	13	

	df	t^*	Interval
Table D	30	2.042	$1.4952 \pm 0.4492 = 1.0460$ to 1.9443
Software	39	2.0227	$1.4952 \pm 0.4449 = 1.0502$ to 1.9401

After undoing the transformation (that is, exponentiating these limits and then subtracting 1), this is about 1.85 to 5.99 mg/l.

7.28. The distribution is skewed to the right, with two peaks—clearly not Normal. However, the sample size of 40 should be sufficient to overcome this, so the t methods should be fairly reliable. The mean is $\bar{x} = 0.76475$ μmol/l, $s \doteq 0.3949$ μmol/l, and the margin of error is $t^* \cdot s/\sqrt{40}$:

0	2333333333333	
0	455	
0	6667	
0	88889999	
1	00011111	
1	23	
1	4	
1		
1	9	

	df	t^*	Interval
Table D	30	2.042	$0.76475 \pm 0.1275 = 0.6372$ to 0.8923 μmol/l
Software	39	2.0227	$0.76475 \pm 0.1263 = 0.6384$ to 0.8911 μmol/l

7.29. **(a)** The distribution is not Normal—there were lots of 1s and 10s—but the nature of the scale means that there can be no extreme outliers, so with a sample of size 60, the t methods should be acceptable. **(b)** The mean is $\bar{x} \doteq 5.9$, $s \doteq 3.7719$, and the margin of error is $t^* \cdot s/\sqrt{60}$:

1	0000000000000000	
2	0000	
3	0	
4	0	
5	00000	
6	000	
7	0	
8	000000	
9	00000	
10	000000000000000000	

	df	t^*	Interval
Table D	50	2.009	$5.9 \pm 0.9783 = 4.9217$ to 6.8783
Software	59	2.0010	$5.9 \pm 0.9744 = 4.9256$ to 6.8744

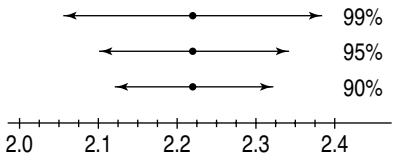
(c) Because this is not a random sample, it may not represent other children well.

7.30. **(a)** The distribution cannot be Normal because all values must be (presumably) integers between 0 and 4. **(b)** The sample size (282) should make the t methods appropriate because the distribution of ratings can have no outliers. **(c)** The margin of error is $t^* \cdot s/\sqrt{282}$, which is either 0.1611 (Table D) or 0.1591 (software):

	df	t^*	Interval
Table D	100	2.626	$2.22 \pm 0.1611 = 2.0589$ to 2.3811
Software	281	2.5934	$2.22 \pm 0.1591 = 2.0609$ to 2.3791

(d) The sample might not represent children from other locations well (or perhaps more accurately, it might not represent well the opinions of parents of children from other locations).

7.31. These intervals are constructed as in the previous exercise, except for the choice of t^*. We see that the width of the interval increases with confidence level.



		df	t^*	Interval
90% confidence	Table D	100	1.660	$2.22 \pm 0.1018 = 2.1182$ to 2.3218
	Software	281	1.6503	$2.22 \pm 0.1012 = 2.1188$ to 2.3212
95% confidence	Table D	100	1.984	$2.22 \pm 0.1217 = 2.0983$ to 2.3417
	Software	281	1.9684	$2.22 \pm 0.1207 = 2.0993$ to 2.3407

7.32. (a) & **(b)** For example, the weight change for Subject 1 is $61.7 - 55.7 = 6$ kg. The mean change is $\bar{x} = 4.73125$ kg and the standard deviation is $s \doteq 1.7457$ kg. **(c)** $\text{SE}_{\bar{x}} = s/\sqrt{16} \doteq 0.4364$ kg; for df $= 15$, $t^* = 2.131$, so the margin of error for 95% confidence is ± 0.9300 (software: ± 0.9302). Based on a method that gives correct results 95% of the time, the mean weight change is 3.8012 to 5.6613 kg (software: 3.8010 to 5.6615 kg). **(d)** $\bar{x} = 10.40875$ lb, $s \doteq 3.8406$ lb, and the 95% confidence interval is 8.3626 to 12.4549 lb (software: 8.3622 to 12.4553 lb). **(e)** H_0 is $\mu = 16$ lb. The test statistic is $t \doteq -5.823$ with df $= 15$, which is highly significant evidence ($P < 0.0001$) against H_0 (unless H_a is $\mu > 16$ lb). **(f)** The data suggest that the excess calories were not converted into weight; the subjects must have used this energy some other way. (See the next exercise for more information.)

7.33. (a) $t = \frac{328 - 0}{256/\sqrt{16}} \doteq 5.1250$ with df $= 15$, for which $P \doteq 0.0012$. There is strong evidence of a change in NEAT. **(b)** With $t^* = 2.131$, the 95% confidence interval is 191.6 to 464.4 cal/day. This tells us how much of the additional calories might have been burned by the increase in NEAT: It consumed 19% to 46% of the extra 1000 cal/day.

7.34. This is a matched-pairs test of H_0: $\mu = 0$ vs. H_a: $\mu > 0$. (The alternative hypothesis is one-sided because the insurance adjusters suspect that Jocko's estimates may be too high.) For this test, we find $t \doteq 2.91$ with df $= 9$, for which $P = 0.0086$ (Minitab output below). This is significant evidence against H_0—that is, we have good reason to believe that Jocko's estimates are higher.

Minitab output

```
Test of mu =   0.0 vs mu >   0.0

Variable     N      Mean    StDev   SE Mean        T    P-Value
diff        10     115.0    124.8      39.5     2.91     0.0086
```

7.35. (a) We wish to test H_0: $\mu_c = \mu_d$ vs. H_a: $\mu_c \neq \mu_d$, where μ_c is the mean computer-calculated mpg and μ_d is the mean mpg computed by the driver. Equivalently, we can state the hypotheses in terms of μ, the mean difference between computer- and driver-calculated mpgs, testing H_0: $\mu = 0$ vs. H_a: $\mu \neq 0$. **(b)** With mean difference $\bar{x} \doteq 2.73$ and standard deviation $s \doteq 2.8015$, the test statistic is $t = \frac{2.73 - 0}{2.8015/\sqrt{20}} \doteq 4.3580$ with df $= 19$, for which $P \doteq 0.0003$. We have strong evidence that the results of the two computations are different.

7.36. (a) $\bar{x} = 5.3\overline{6}$ mg/dl and $s \doteq 0.6653$ mg/dl, so $\text{SE}_{\bar{x}} \doteq 0.2716$ mg/dl. **(b)** For df $= 5$, we have $t^* = 2.015$, so the interval is 4.819 to 5.914 mg/dl.

7.37. To test H_0: $\mu = 4.8$ mg/dl vs. H_a: $\mu > 4.8$ mg/dl, we find $t = \frac{5.3\overline{6} - 4.8}{0.2716} \doteq 2.086$. For df $= 5$, Table D shows that $0.025 < P < 0.05$ (software gives $P = 0.0457$). This is fairly strong, though not overwhelming, evidence that the patient's phosphate level is above normal.

7.38. The 90% confidence interval is $3.9 \pm t^*(0.98/\sqrt{1406})$.
With Table D, take df $= 1000$ and $t^* = 1.646$; with software, take df $= 1405$ and $t^* = 1.6459$. Either way, the confidence interval is 3.8570 to 3.9430.

7.39. (a) The differences are spread from -0.018 to 0.020 g, with mean $\bar{x} = -0.0015$ and standard deviation $s \doteq 0.0122$ g. A stemplot is shown on the right; the sample is too small to make judgments about skewness or symmetry. **(b)** For H_0: $\mu = 0$ vs. H_a: $\mu \neq 0$, we find $t = \frac{-0.0015 - 0}{s/\sqrt{8}} \doteq -0.347$ with df $= 7$, for which $P = 0.7388$. We cannot reject H_0 based on this sample. **(c)** The 95% confidence interval is $-0.0015 \pm (2.365)(0.0122/\sqrt{8}) = -0.0015 \pm 0.0102 = -0.0117$ to 0.0087 g. **(d)** The subjects from this sample may be representative of future subjects, but the test results and confidence interval are suspect because this is not a random sample.

−1	85
−1	
−0	65
−0	
0	2
0	55
1	
1	
2	0

7.40. (a) The differences are spread from -31 to 45 g/cm^2, with mean $\bar{x} = 4.625$ and standard deviation $s \doteq 26.8485$ g/cm^2. A stemplot is shown on the right; the sample is too small to make judgments about skewness or symmetry. **(b)** For H_0: $\mu = 0$ vs. H_a: $\mu \neq 0$, we find $t = \frac{4.625 - 0}{s/\sqrt{8}} \doteq 0.487$ with df $= 7$, for which $P = 0.6410$. We cannot reject H_0 based on this sample. **(c)** The 95% confidence interval is $4.625 \pm (2.365)(26.8485/\sqrt{8}) = 4.625 \pm 22.4494 = -17.8244$ to 27.0744 g/cm^2. **(d)** See the answer to part (d) of the previous exercise.

−3	1
−2	8
−1	
−0	4
0	16
1	3
2	
3	5
4	5

7.41. (a) We test H_0: $\mu = 0$ vs. H_a: $\mu > 0$, where μ is the mean change in score (that is, the mean improvement). **(b)** The distribution is slightly left-skewed, with mean $\bar{x} = 2.5$ and $s \doteq 2.8928$. **(c)** $t = \frac{2.5 - 0}{s/\sqrt{20}} \doteq 3.8649$, df $= 19$, and $P = 0.00052$; there is strong evidence of improvement in listening test scores. **(d)** With df $= 19$, we have $t^* = 2.093$, so the 95% confidence interval is 1.1461 to 3.8539.

−0	6
−0	
−0	
−0	0
0	0011
0	222333333
0	
0	66666

7.42. A stemplot of these call lengths can be seen in Figure 1.6 of the text (page 12). The distribution is sharply right-skewed, with one clear outlier (2631 minutes). In fact, based on the $1.5 \times IQR$ criterion, the top eight call lengths are outliers. The mean and standard deviation are $\bar{x} = 196.575$ and $s \doteq 342.0215$ min. The five-number summary (all in minutes) is Min $= 1$, $Q_1 = 54.5$, $M = 103.5$, $Q_3 = 200$, Max $= 2631$.

With Table D (df $= 70$ and $t^* = 1.994$), the 95% confidence interval for μ is 120.3261 to 272.8239 min. With software (df $= 79$ and $t^* = 1.9905$), it is 120.4618 to 272.6882 min. Given the sharp skewness of the data and the large number of outliers, we should be quite cautious in using this interval for inference.

7.43. The distribution is fairly symmetrical with no outliers. The mean IQ is $\bar{x} = 114.98\overline{3}$ and the standard deviation is $s \doteq 14.8009$. The 95% confidence interval is $\bar{x} \pm t^*(s/\sqrt{60})$. With Table D (df $= 50$ and $t^* = 2.009$), this is 111.1446 to 118.8221. With software (df $= 59$ and $t^* = 2.0010$), this is 111.1598 to 118.8068. Because all students in the sample came from the same school, this *might* adequately describe the mean IQ at this school, but the sample could not be considered representative of all fifth graders.

8	12
8	9
9	04
9	67
10	01112223
10	568999
11	0002233444
11	5677788
12	223444
12	56778
13	01344
13	6799
14	2
14	5

7.44. With all 50 states and Puerto Rico listed in the table, we have information about the entire population in question; no statistical procedures are needed (or meaningful).

7.45. We test H_0: median $= 0$ vs. H_a: median > 0—or equivalently, H_0: $p = 1/2$ vs. H_a: $p > 1/2$, where p is the probability that Jocko's estimate is higher. One difference is 0; of the nine non-zero differences, seven are positive. The P-value is $P(X \geq 7) = 0.0898$ from a $B(9, 0.5)$ distribution; there is not quite enough evidence to conclude that Jocko's estimates are higher. In Exercise 7.34 we were able to reject H_0; here we cannot.

Note: *The failure to reject H_0 in this case is because with the sign test, we pay attention only to the* sign *of each difference, not the* size. *In particular, the negative differences are each given the same "weight" as each positive difference, in spite of the fact that the negative differences are only −$50 each, while most of the positive differences are larger. See the "Caution" about the sign test on page 440 of the text.*

Minitab output

```
Sign test of median = 0.00000 versus  G.T.  0.00000

               N  BELOW  EQUAL  ABOVE   P-VALUE     MEDIAN
diff          10      2      1      7    0.0898      125.0
```

7.46. We test H_0: median $= 0$ vs. H_a: median $\neq 0$. The Minitab output below gives $P = 1$ because there were four positive and four negative differences, giving us no reason to doubt H_0. The t test P-value was 0.7388. (This is the same conclusion we reached with the t test.)

Minitab output

```
Sign test of median = 0.00000 versus  N.E.  0.00000

               N  BELOW  EQUAL  ABOVE   P-VALUE     MEDIAN
opdiff         8      4      0      4    1.0000   -0.00150
```

7.47. We test H_0: median $= 0$ vs. H_a: median $\neq 0$. There were three negative and five positive differences, so the P-value is $2P(X \geq 5)$ for a binomial distribution with parameters $n = 8$ and $p = 0.5$. From Table C or software (Minitab output below), we have $P = 0.7266$, which gives no reason to doubt H_0. The t test P-value was 0.6410.

Minitab output

```
Sign test of median = 0.00000 versus  N.E.  0.00000

               N  BELOW  EQUAL  ABOVE   P-VALUE     MEDIAN
opdiff         8      3      0      5    0.7266      3.500
```

7.48. We test H_0: median $= 0$ vs. H_a: median > 0, or H_0: $p = 1/2$ vs. H_a: $p > 1/2$. Three of the 20 differences are zero; of the other 17, 16 are positive. The P-value is $P(X \geq 16)$ for a $B(17, 0.5)$ distribution. While Table C cannot give us the exact value of this probability, if we weaken the evidence by pretending that the three zero differences were negative and look at the $B(20, 0.5)$ distribution, we can estimate that $P < 0.0059$—enough information to reject the null hypothesis. In fact, the P-value is 0.0001.

Minitab output

```
Sign test of median = 0.00000 versus  G.T.  0.00000

               N  BELOW  EQUAL  ABOVE   P-VALUE     MEDIAN
gain          20      1      3     16    0.0001      3.000
```

7.49. We test H_0: median $= 0$ vs. H_a: median > 0, or H_0: $p = 1/2$ vs. H_a: $p > 1/2$. Out of the 20 differences, 17 are positive (and none equal 0). The P-value is $P(X \geq 17)$ for a $B(20, 0.5)$ distribution. From Table C or software (Minitab output below), we have $P = 0.0013$, so we reject H_0 and conclude that the results of the two computations are different. (Using a t test, we found $P \doteq 0.0003$, which led to the same conclusion.)

Minitab output

```
Sign test of median = 0.00000 versus  G.T.  0.00000

               N  BELOW  EQUAL  ABOVE   P-VALUE     MEDIAN
diff          20      3      0     17    0.0013      3.000
```

7.50. After taking logarithms, the 90% confidence interval is $\bar{x} \pm t^*(s/\sqrt{5})$. For df $= 4$, $t^* = 2.132$, and the confidence intervals are as shown in the table. (As we would expect, after exponentiating to undo the logarithms, both intervals are equivalent except for rounding differences: 311.2 to 414.5 hours.)

Log	$\bar{x}$	s	Confidence interval
Common	2.5552	0.0653	2.4930 to 2.6175
Natural	5.8836	0.1504	5.7403 to 6.0270

7.51. The standard deviation for the given data was $s \doteq 0.012224$. With $t = \dfrac{\bar{x}}{s/\sqrt{15}}$, $\alpha = 0.05$, and df $= 14$, we reject H_0 if $|t| \geq 2.145$, which means $|\bar{x}| \geq (2.145)(s/\sqrt{15})$, or $|\bar{x}| \geq 0.00677$. Assuming $\mu = 0.002$:

$$\begin{aligned} P(|\bar{x}| \geq 0.00677) &= 1 - P\,(-0.00677 \leq \bar{x} \leq 0.00677) \\ &= 1 - P\left(\frac{-0.00677 - 0.002}{s/\sqrt{15}} \leq \frac{\bar{x} - 0.002}{s/\sqrt{15}} \leq \frac{0.00677 - 0.002}{s/\sqrt{15}}\right) \\ &= 1 - P\,(-2.78 \leq Z \leq 1.51) \\ &= 1 - (0.9345 - 0.0027) \doteq 0.07 \end{aligned}$$

The power is about 7% against this alternative—not surprising, given the small sample size, and the fact that the difference (0.002) is small relative to the standard deviation.

Note: *You can find a collection of online power calculators for various types of statistical tests at* `http://calculators.stat.ucla.edu/powercalc/`. *At the time this solutions manual was prepared, these online tools were not available, but the message at that site ("Please check back later") suggests that they may return.*

Note that for one-sample tests, this page gives slightly different answers from those found by the method described in the text; for this problem, it gives the power as about 0.09. The online calculator, like other statistical software, uses a "noncentral t distribution" (used in the text for two-sample power problems) rather than a Normal distribution, resulting in more accurate answers. In most situations, the practical conclusions drawn from the power computations are the same regardless of the method used.

7.52. We will reject H_0 when $t = \frac{\bar{x}}{s/\sqrt{n}} \geq t^*$, where t^* is the appropriate critical value for the chosen sample size. This corresponds to $\bar{x} \geq 10t^*/\sqrt{n}$, so the power against $\mu = 2$ is:

$$\begin{aligned} P(\bar{x} \geq 10t^*/\sqrt{n}) &= P\left(\frac{\bar{x} - 2}{10/\sqrt{n}} \geq \frac{10t^*/\sqrt{n} - 2}{10/\sqrt{n}}\right) \\ &= P\left(Z \geq t^* - 0.2\sqrt{n}\right) \end{aligned}$$

For $\alpha = 0.05$, the table on the right shows the power for a variety of sample sizes, and we see that $n \geq 156$ achieves the desired 80% power.

n	t^*	$t^* - 0.2\sqrt{n}$	Power
50	1.6766	0.2623	0.3965
75	1.6657	-0.0663	0.5264
100	1.6604	-0.3396	0.6329
125	1.6572	-0.5788	0.7186
150	1.6551	-0.7943	0.7865
151	1.6551	-0.8026	0.7889
152	1.6550	-0.8108	0.7912
153	1.6549	-0.8189	0.7936
154	1.6549	-0.8271	0.7959
155	1.6548	-0.8352	0.7982
156	1.6547	-0.8433	0.8005

Note: *Some power analysis software will determine the required sample size to achieve a specified level of power. For example, the online calculator noted above reports that the appropriate sample size is 155.926.*

7.53. Taking $s = 1.5$ as in Example 7.9, the power for the alternative $\mu = 1.0$ is:

$$P\left(\bar{x} \geq \frac{t^* s}{\sqrt{n}} \text{ when } \mu = 1.0\right) = P\left(\frac{\bar{x} - 1.0}{s/\sqrt{n}} \geq \frac{t^* s/\sqrt{n} - 1.0}{s/\sqrt{n}}\right) = P\left(Z \geq t^* - \frac{\sqrt{n}}{1.5}\right)$$

Starting from the result of Example 7.9 (power $\doteq 0.89$ when $n = 20$) and using trial-and-error, we find that with $n = 15$, power $\doteq 0.7941$, and with $n = 16$, power $\doteq 0.8195$. Therefore, we need $n \geq 16$.

7.54. **(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 ($n_1 = n_2 = 25$), so the appropriate degrees of freedom would be $\text{df} = 25 - 1 = 24$. **(c)** For a two-sided test at $\alpha = 0.05$, we need $|t| > t^*$, where $t^* = 2.064$ is the 0.025 critical value for a t distribution with $\text{df} = 24$.

7.55. Because $2.492 < t < 2.797$ and the alternative is two-sided, Table D tells us that the P-value is $0.01 < P < 0.02$. (Software gives $P = 0.0112$.)

7.56. We find $\text{SE}_D \doteq 2.2091$. The options for the 95% confidence interval for $\mu_1 - \mu_2$ are shown on the right. This interval includes fewer values than a 99% confidence interval would (that is, a 99% confidence interval would be wider) because increasing our confidence level means that we need a larger margin of error.

df	t^*	Confidence interval
94.9	1.9853	−24.3856 to −15.6144
49	2.0096	−24.4393 to −15.5607
40	2.021	−24.4645 to −15.5355

7.57. We find $\text{SE}_D \doteq 4.9396$. The options for the 95% confidence interval for $\mu_1 - \mu_2$ are shown on the right. Because 0 does not fall in the 95% confidence interval, we would reject H_0 at $\alpha = 0.05$ against a two-sided alternative. (In fact, the two-sided P-value is about 0.0008.)

df	t^*	Confidence interval
17.4	2.1060	−30.4027 to −9.5973
9	2.262	−31.1735 to −8.8265

7.58. $$\text{df} = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}} = \frac{(10^2/20 + 12^2/18)^2}{\dfrac{(10^2/20)^2}{19} + \dfrac{(12^2/18)^2}{17}} = \frac{(5 + 8)^2}{\dfrac{5^2}{19} + \dfrac{8^2}{17}} \doteq 33.2645.$$

7.59. SPSS and SAS give both results, while Minitab and Excel show only the unpooled procedures. The pooled t statistic is -56.99, which has a tiny P-value.

Note: *When the sample sizes are equal—as in this case—the pooled and unpooled t statistics are equal. (See the next exercise.)*

The title of the Excel output is not quite accurate: It says, "...Assuming Unequal Variances," but in fact, unpooled procedures make no *assumptions about the variances.*

Finally, note that both Excel and Minitab can do pooled procedures as well as the unpooled procedures that are shown.

7.60. If $n_1 = n_2 = n$, the pooled estimate of the variance is:

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(n-1)s_1^2 + (n-1)s_2^2}{2(n-1)} = \frac{s_1^2+s_2^2}{2}$$

The pooled standard error is therefore $s_p\sqrt{\frac{1}{n}+\frac{1}{n}} = \sqrt{s_p^2 \cdot \frac{2}{n}} = \sqrt{\frac{s_1^2+s_2^2}{n}} = \sqrt{\frac{s_1^2}{n}+\frac{s_2^2}{n}}$, which is the same as the unpooled standard error.

7.61. (a) Assuming we have SRSs from each population, use of two-sample t procedures seems reasonable. **(b)** We wish to test H_0: $\mu_f = \mu_m$ vs. H_a: $\mu_f \neq \mu_m$. **(c)** We find $\text{SE}_D \doteq 6.8490$ mg/dl. The test statistic is $t \doteq 0.276$, with df $\doteq$ 76.1 (or 36—use 30 for Table D), for which $P \doteq 0.78$. We have no reason to believe that male and female cholesterol levels are different. **(d)** The options for the 95% confidence interval for $\mu_f - \mu_m$ are shown on the right. **(e)** It might not be appropriate to treat these students as SRSs from larger populations.

df	t^*	Confidence interval
76.1	1.9916	−11.7508 to 15.5308
36	2.0281	−12.0005 to 15.7805
30	2.042	−12.0957 to 15.8757

Note: *Because t distributions are more spread out than Normal distributions, a t-value that would not be significant for a Normal distribution cannot possibly be significant when compared to a t distribution.*

7.62. Considering the LDL cholesterol levels, we test H_0: $\mu_f = \mu_m$ vs. H_a: $\mu_f < \mu_m$. We find $\text{SE}_D \doteq 6.2087$ mg/dl, so the test statistic is $t \doteq -2.104$, with df $\doteq$ 70.5 (or 36—use 30 for Table D), for which $P \doteq 0.0195$. We have enough evidence to conclude LDL cholesterol levels are higher for males than for females.

Note: *This t statistic was computed by subtracting the male mean from the female mean. Reversing the order of subtraction would make t positive, but would not change the P-value or the conclusion.*

7.63. (a) The distribution cannot be Normal because all numbers are integers. **(b)** The t procedures should be appropriate because we have two large samples with no outliers. **(c)** H_0: $\mu_I = \mu_C$; H_a: $\mu_I > \mu_C$ (or $\mu_I \neq \mu_C$). The one-sided alternative reflects the researchers' (presumed) belief that the intervention would increase scores on the test. The two-sided alternative allows for the possibility that the intervention might have had a negative effect. **(d)** $\text{SE}_D = \sqrt{s_I^2/n_I + s_C^2/n_C} \doteq 0.1198$ and $t = (\bar{x}_I - \bar{x}_C)/\text{SE}_D \doteq 6.258$. Regardless of how we compute degrees of freedom (df $\doteq$ 354 or 164), the P-value is very small: $P < 0.0001$. We reject H_0 and conclude that the intervention increased test scores. **(e)** The interval is $\bar{x}_I - \bar{x}_C \pm t^*\text{SE}_D$; the value of t^* depends on the df (see the table), but note that in every case the interval rounds to 0.51 to 0.99. **(f)** The results for this sample may not generalize well to other areas of the country.

df	t^*	Confidence interval
354.0	1.9667	0.5143 to 0.9857
164	1.9745	0.5134 to 0.9866
100	1.984	0.5122 to 0.9878

7.64. **(a)** The distribution cannot be Normal, because all numbers are integers. **(b)** The t procedures should be appropriate because we have two large samples with no outliers. **(c)** H_0: $\mu_I = \mu_C$; H_a: $\mu_I > \mu_C$ (or $\mu_I \neq \mu_C$). The one-sided alternative reflects the researchers' (presumed) belief that the intervention would increase self-efficacy scores. The two-sided alternative allows for the possibility that the intervention might have had a negative effect. **(d)** $\text{SE}_D = \sqrt{s_I^2/n_I + s_C^2/n_C} \doteq 0.1204$ and $t = (\bar{x}_I - \bar{x}_C)/\text{SE}_D \doteq 3.571$. Regardless of how we compute degrees of freedom (df $\doteq$ 341.8 or 164), the (one-sided) P-value is about 0.0002. We reject H_0 and conclude that the intervention increased self-efficacy scores. **(e)** The interval is $\bar{x}_I - \bar{x}_C \pm t^*\text{SE}_D$; the value of t^* depends on the df (see the table), but note that in every case the interval rounds to 0.19 to 0.67. **(f)** As in the previous exercise, the results for this sample may not generalize well to other areas of the country.

df	t^*	Confidence interval
341.8	1.9669	0.1932 to 0.6668
164	1.9745	0.1922 to 0.6678
100	1.984	0.1911 to 0.6689

7.65. **(a)** This may be near enough to an SRS, if this company's working conditions were similar to that of other workers. **(b)** $\text{SE}_D \doteq 0.7626$; regardless of how we choose df, the interval rounds to 9.99 to 13.01 mg.y/m^3. **(c)** A one-sided alternative would seem to be reasonable here; specifically, we would likely expect that the mean exposure for outdoor workers would be lower. For testing H_0, we find $t = 15.08$, for which $P < 0.0001$ with either df = 137 or 114 (and for either a one- or a two-sided alternative). We have strong evidence that outdoor concrete workers have lower dust exposure than the indoor workers. **(d)** The sample sizes are large enough that skewness should not matter.

df	t^*	Confidence interval
137.1	1.9774	9.9920 to 13.0080
114	1.9810	9.9893 to 13.0107
100	1.984	9.9870 to 13.0130

7.66. With the given standard deviations, $\text{SE}_D \doteq 0.2653$; regardless of how we choose df, a 95% confidence interval for the difference in means rounds to 4.37 to 5.43 mg.y/m^3. With the null hypothesis H_0: $\mu_i = \mu_o$ (and either a one- or two-sided alternative, as in the previous exercise), we find $t = 18.47$, for which $P < 0.0001$ regardless of df and the chosen alternative. We have strong evidence that outdoor concrete workers have lower respirable dust exposure than the indoor workers.

df	t^*	Confidence interval
121.5	1.9797	4.3747 to 5.4253
114	1.9810	4.3744 to 5.4256
100	1.984	4.3736 to 5.4264

7.67. To find a confidence interval $(\bar{x}_1 - \bar{x}_2) \pm t^*\text{SE}_D$, 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 $t = (\bar{x}_1 - \bar{x}_2)/\text{SE}_D$, so we could compute $\text{SE}_D = (\bar{x}_1 - \bar{x}_2)/t$ 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).

7.68. **(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 (unless some children are regurgitating sweetened drinks), the distributions must be right-skewed. **(b)** We find $SE_D \doteq 4.3786$ and $t \doteq -1.439$, with either df $\doteq 7.8$ ($P = 0.1890$) or df $= 4$ ($P = 0.2236$). We do not have enough evidence to reject H_0. **(c)** The possible 95% confidence intervals are given in the table. (The two different intervals for df $= 4$ are using t^* from software, and from Table D.) **(d)** Because the distributions are not Normal and the samples are small, the t procedures are 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 very cautious about extending these results to other children.

df	t^*	Confidence interval
7.8	2.3159	−16.4404 to 3.8404
4	2.776	−18.4551 to 5.8551

7.69. **(a)** Hypotheses should involve μ_1 and μ_2 (population means) rather than $\bar{x}_1$ and $\bar{x}_2$ (sample means). **(b)** The samples are not independent; we would need to compare the 24 males to the 26 females. **(c)** We need P to be small (for example, less than 0.10) to reject H_0. A large P-value like this gives no reason to doubt H_0. **(d)** Assuming the researcher computed the t statistic using $\bar{x}_1 - \bar{x}_2$, a positive value of t does not support H_a.

7.70. **(a)** Because 0 is in the confidence interval, we would not reject H_0 at the 5% level. **(b)** Larger samples generally give smaller margins of error (at the same confidence level, and assuming that the standard deviations for the large and small samples are about the same). One (conceptual) explanation for this is that larger samples give more information and therefore offer more precise results. A more mathematical explanation: In looking at the formula for a two-sample confidence interval, we see that $SE_D = \sqrt{s_1^2/n_1 + s_2^2/n_2}$, so that if n_1 and n_2 are increased, the standard error decreases.

Note: *For (a), we can even make some specific statements about t and its P-value: The confidence interval tells us that $\bar{x}_1 - \bar{x}_2 = 0.7$ and the margin of error is 0.8. As t^* for a 95% confidence interval is at least 1.96, $SE_D = \frac{0.8}{t^*}$ is less than about 0.408 and the t-statistic $t = 0.7/SE_D$ is at least 1.715. (The largest possible value of t, for df = 1, is about 11.1.) A little experimentation with different df reveals that the P-value is at least 0.057 for all df, and if df > 3, then $P > 0.07$.*

7.71. **(a)** We can reject H_0: $\mu_1 = \mu_2$ in favor of the two-sided alternative at the 5% level because $P \doteq 0.005$. **(b)** We would also 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 $\mu_1 - \mu_2 < 0$. In fact, $P = 0.0025$ (half as big as before).

7.72. We find $SE_D \doteq 2.2091$. The options for the 95% confidence interval for $\mu_1 - \mu_2$ are shown on the right. A 99% confidence interval would include more values (it would be wider) because increasing our confidence level means that we need a larger margin of error.

df	t^*	Confidence interval
94.9	1.9853	−24.3856 to −15.6144
49	2.0096	−24.4393 to −15.5607
40	2.021	−24.4645 to −15.5355

7.73. 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).

7.74. **(a)** Results for this randomization will depend on the technique used. **(b)** $SE_D \doteq 0.5235$, and the options for the 95% confidence interval are given on the right. **(c)** Because 0 falls outside the 95% confidence interval, the P-value is less than 0.05, so we would reject H_0. (For reference, $t \doteq 3.439$ and the actual P-value is either 0.0045 or 0.0074, depending on which df we use.)

df	t^*	Confidence interval
12.7	2.1651	0.6667 to 2.9333
9	2.2622	0.6159 to 2.9841
9	2.262	0.6160 to 2.9840

7.75. The next 10 employees who need screens might not be an independent group—perhaps they all come from the same department, for example. Randomization reduces the chance that we end up with such unwanted groupings.

7.76. **(a)** We test H_0: $\mu_0 = \mu_3$ vs. H_a: $\mu_0 > \mu_3$. For the two samples, we find $\bar{x}_0 = 48.705$, $s_0 = 1.5344$, $\bar{x}_3 = 21.795$, and $s_3 = 0.7707$ mg/100g; thus $SE_D \doteq 1.2142$ and $t = 22.16$. This gives either $P = 0.0080$ (df = 1.47) or $P = 0.0143$ (df = 1); either way, this is fairly strong evidence that vitamin C is lost in storage. **(b)** Options for the 90% confidence interval for $\mu_0 - \mu_3$ are given in the table.

df	t^*	Confidence interval
1.47	4.7037	21.1988 to 32.6212
1	6.3137	19.2439 to 34.5761
1	6.314	19.2436 to 34.5764

7.77. **(a)** This is now a matched-pairs design: The immediate and three-days-later numbers are not independent. **(b)** The vitamin C content changed by -26.37 and -27.45 mg/100g, for which $\bar{x} = -26.91$ and $s \doteq 0.7637$ mg/100g. Then $t = \frac{\bar{x}}{s/\sqrt{2}} \doteq -49.83$ with df = 1, for which $P = 0.0064$. The 90% confidence interval for the change in vitamin C content is -30.3196 to -23.5004 mg/100g.

7.78. **(a)** We test H_0: $\mu_0 = \mu_3$ vs. H_a: $\mu_0 > \mu_3$. For the two samples, we find $\bar{x}_0 = 95.3$, $s_0 = 0.9899$, $\bar{x}_3 = 95.85$, and $s_3 = 2.1920$ mg/100g. Because $\bar{x}_3 > \bar{x}_0$, we have no evidence against H_0; further analysis is not necessary. (However, just for reference, $t = -0.323$.) **(b)** Options for the 90% confidence interval for $\mu_0 - \mu_3$ are given in the table.

df	t^*	Confidence interval
1.39	4.9847	-9.0276 to 7.9276
1	6.3137	-11.2880 to 10.1880
1	6.314	-11.2884 to 10.1884

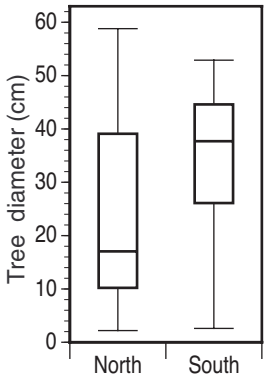
7.79. Small samples may lead to rejection of H_0, if (as in Exercise 7.76) the evidence is very strong. (The weakness of small samples is that they are not very powerful; the rejection in 7.76 occurred because the evidence suggests that the true means are quite different.)

7.80. **(a)** The null hypothesis is $\mu_1 = \mu_2$; the alternative can be either two- or one-sided. (It might be a reasonable expectation that $\mu_1 > \mu_2$.) We find $\text{SE}_D \doteq 0.2796$ and $t = 8.369$. Regardless of df and H_a, the conclusion is the same: P is very small, and we conclude that *WSJ* ads are more trustworthy. **(b)** Possible 95% confidence intervals are given in the table; all place the difference in trustworthiness at between about 1.8 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.

df	t^*	Confidence interval
121.5	1.9797	1.7865 to 2.8935
60	2.0003	1.7807 to 2.8993
60	2.000	1.7808 to 2.8992

7.81. **(a)** The north distribution (with five-number summary 2.2, 10.2, 17.05, 39.1, 58.8 cm) is right-skewed, while the south distribution (2.6, 26.1, 37.70, 44.6, 52.9 cm) is left-skewed. Stemplots and boxplots are shown on the right. **(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)** H_0: $\mu_n = \mu_s$; H_a: $\mu_n \neq \mu_s$; 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 $\bar{x}_n = 23.7$, $s_n \doteq 17.5001$, $\bar{x}_s = 34.5\overline{3}$, and $s_s \doteq 14.2583$ cm. Then $\text{SE}_D \doteq 4.1213$, so $t = -2.629$ with df $= 55.7$ ($P = 0.011$) or df $= 29$ ($P = 0.014$). We conclude that the means are different (specifically, the south mean is greater than the north mean). **(e)** See the table for possible 95% confidence intervals.

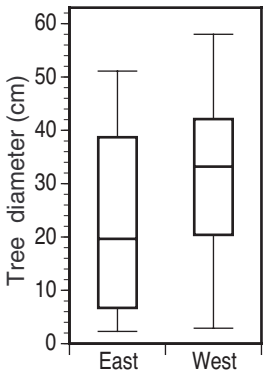
North		South
43322	0	2
65	0	57
443310	1	2
955	1	8
	2	13
8755	2	689
0	3	2
996	3	566789
43	4	003444
6	4	578
4	5	0112
85	5	



df	t^*	Confidence interval
55.7	2.0035	−19.0902 to −2.5765
29	2.0452	−19.2624 to −2.4043
29	2.045	−19.2614 to −2.4053

7.82. (a) The east distribution (with five-number summary 2.3, 6.7, 19.65, 38.7, 51.1 cm) is right-skewed, while the west distribution (2.9, 20.4, 33.20, 42.1, 58.0 cm) is left-skewed. Stemplots and boxplots are shown on the right. **(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)** H_0: $\mu_e = \mu_w$; 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 $\bar{x}_e = 21.71\overline{6}$, $s_e \doteq 16.0743$, $\bar{x}_w = 30.28\overline{3}$, and $s_w \doteq 15.3314$ cm. Then $\text{SE}_D \doteq 4.0556$, so $t = -2.112$ with df $= 57.8$ ($P = 0.0390$) or df $= 29$ ($P = 0.0434$). We conclude that the means are different at $\alpha = 0.05$ (specifically, the west mean is greater than the east mean). **(e)** See the table for possible 95% confidence intervals.

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



df	t^*	Confidence interval
57.8	2.0018	−16.6852 to −0.4481
29	2.0452	−16.8613 to −0.2720
29	2.045	−16.8604 to −0.2730

7.83. (a) $\text{SE}_D \doteq 3.0175$. Answers will vary with the df used; see the table. **(b)** Because of random fluctuations between stores, we might (just by chance) have seen a rise in the average number of units sold even if actual mean sales had remained unchanged—or even if they dropped slightly.

df	t^*	Confidence interval
104.6	1.9829	−2.9834 to 8.9834
54	2.0049	−3.0497 to 9.0497
50	2.009	−3.0621 to 9.0621

7.84. (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 for the one-tailed test: $P = 0.12$.

7.85. (a) We test H_0: $\mu_b = \mu_f$; H_a: $\mu_b > \mu_f$. $\text{SE}_D \doteq 0.5442$ and $t = 1.654$, for which $P = 0.0532$ (df $= 37.6$) or 0.0577 (df $= 18$); there is not quite enough evidence to reject H_0 at $\alpha = 0.05$. **(b)** The confidence interval depends on the degrees of freedom used; see the table. **(c)** We need two independent SRSs from Normal populations.

df	t^*	Confidence interval
37.6	2.0251	−0.2021 to 2.0021
18	2.1009	−0.2434 to 2.0434
18	2.101	−0.2434 to 2.0434

7.86. The standard deviations $s_1 = 29.8$ and $s_2 = 31.1$ are very similar, and the sample sizes are reasonably large, so the pooled procedures should be fairly reliable. $s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}} \doteq 30.2173$, so the standard error is $\text{SE}_D = s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \doteq 6.1269$. For the hypothesis test, we find $t = -2.132$, df $= 106$, and $P \doteq 0.0177$ (so we conclude that mean male LDL levels are lower). The 95% confidence interval is $\bar{x}_1 - \bar{x}_2 \pm t^*\text{SE}_D$, for which the two possible intervals (with software, or with Table D) are shown in the table. The test results are nearly the same as in Exercise 7.62. That exercise did not call for a confidence interval, but the unpooled interval is also similar to the pooled interval.

df	t^*	Confidence interval
106	1.9826	−25.2071 to −0.9129
100	1.984	−25.2157 to −0.9043

7.87. The standard deviations $s_1 = 1.15$ and $s_2 = 1.16$ are very similar, and the sample sizes are quite large, so the pooled procedures should be fairly reliable. $s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}} \doteq 1.1556$, so the standard error is $\text{SE}_D = s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \doteq 0.1200$. For the hypothesis test, we find $t = 6.251$, df $= 375$, and $P < 0.0001$ (so we conclude that the intervention increased test scores). The 95% confidence interval is $\bar{x}_1 - \bar{x}_2 \pm t^*\text{SE}_D$, for which the two possible intervals (with software, or with Table D) are shown in the table. These results are nearly the same as in Exercise 7.63.

df	t^*	Confidence interval
375	1.9663	0.5141 to 0.9859
100	1.984	0.5120 to 0.9880

7.88. The standard deviations $s_1 = 1.19$ and $s_2 = 1.12$ are fairly similar, and the sample sizes are quite large, so the pooled procedures should be fairly reliable. $s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}} \doteq 1.1511$, so the standard error is $\text{SE}_D = s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \doteq 0.1195$. For the hypothesis test, we find $t = 3.598$, df $= 375$, and $P = 0.0002$ (for a one-sided alternative); we conclude that the intervention increased self-efficacy scores. The 95% confidence interval is $\bar{x}_1 - \bar{x}_2 \pm t^*\text{SE}_D$, for which the two possible intervals (with software, or with Table D) are shown in the table. These results are nearly the same as in Exercise 7.64.

df	t^*	Confidence interval
375	1.9663	0.1950 to 0.6650
100	1.984	0.1929 to 0.6671

7.89. The pooled standard deviation is $s_p \doteq 15.9617$, and the standard error is $\text{SE}_D \doteq 4.1213$. For the significance test, $t = -2.629$, df $= 58$, and $P = 0.0110$, so we have fairly strong evidence (though not quite significant at $\alpha = 0.01$) that the south mean is greater than the north mean. Possible answers for the confidence interval (with software, and with Table D) are given in the table. All results are similar to those found in Exercise 7.81.

Note: *If $n_1 = n_2$ (as in this case), the standard error and t statistic are the same for the usual and pooled procedures. The degrees of freedom will usually be different (specifically, df is larger for the pooled procedure, unless $s_1 = s_2$ and $n_1 = n_2$).*

df	t^*	Confidence interval
58	2.0017	−19.0830 to −2.5837
50	2.009	−19.1130 to −2.5536

7.90. Testing the same hypotheses as in that example (H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 \neq \mu_2$), we have pooled standard deviation $s_p \doteq 0.2105$ so that $\text{SE}_D \doteq 0.0384$ and $t \doteq 17.169$. With either df $= 133$ or 100, we find that $P < 0.001$, so we have very strong evidence that mean wheat prices were higher in September than they were in July. This is nearly identical to the results of the unpooled analysis ($t = 18.03$, $P < 0.001$).

7.91. With $s_n \doteq 17.5001$, $s_s \doteq 14.2583$, and $n_n = n_s = 30$, we have $s_n^2/n_n \doteq 10.2085$ and $s_s^2/n_s \doteq 6.7767$, so:

$$\text{df} = \frac{\left(\frac{s_n^2}{n_n} + \frac{s_s^2}{n_s}\right)^2}{\frac{1}{n_n - 1}\left(\frac{s_n^2}{n_n}\right)^2 + \frac{1}{n_s - 1}\left(\frac{s_s^2}{n_s}\right)^2} \doteq \frac{(10.2085 + 6.7767)^2}{\frac{1}{29}(10.2085^2 + 6.7767^2)} \doteq 55.7251$$

7.92. With $s_e \doteq 16.0743$, $s_w \doteq 15.3314$, and $n_e = n_w = 30$, we have $s_e^2/n_e \doteq 8.6128$ and $s_w^2/n_w \doteq 7.8351$, so:

$$\text{df} = \frac{\left(\frac{s_e^2}{n_e} + \frac{s_w^2}{n_w}\right)^2}{\frac{1}{n_e - 1}\left(\frac{s_e^2}{n_e}\right)^2 + \frac{1}{n_w - 1}\left(\frac{s_w^2}{n_w}\right)^2} \doteq \frac{(8.6128 + 7.8351)^2}{\frac{1}{29}(8.6128^2 + 7.8351^2)} \doteq 57.8706$$

7.93. (a) With $s_i \doteq 7.8$, $n_i = 115$, $s_o \doteq 3.4$, and $n_o = 220$, we have $s_i^2/n_i \doteq 0.5290$ and $s_o^2/n_o \doteq 0.05455$, so:

$$\text{df} \doteq \frac{(0.5290 + 0.05455)^2}{\frac{0.5290^2}{114} + \frac{0.05455^2}{219}} \doteq 137.0661$$

	df	t^*	Confidence interval
Part (d)	333	1.9671	10.2931 to 12.7069
	100	1.984	10.2827 to 12.7173
Part (e)	333	1.9671	4.5075 to 5.2925
	100	1.984	4.5042 to 5.2958

(b) $s_p = \sqrt{\frac{(n_i - 1)s_i^2 + (n_o - 1)s_o^2}{n_i + n_o - 2}} \doteq 5.3320$, which is slightly closer to s_o (the standard deviation from the larger sample). **(c)** With no assumption of equality, $\text{SE}_1 = \sqrt{s_i^2/n_i + s_o^2/n_o} \doteq 0.7626$. With the pooled method, $\text{SE}_2 = s_p\sqrt{1/n_i + 1/n_o} \doteq 0.6136$. **(d)** With the pooled standard deviation, $t \doteq 18.74$ and df $= 333$, for which $P < 0.0001$, and the 95% confidence interval is as shown in the table. With the smaller standard error, the t value is larger (it had been 15.08), and the confidence interval is narrower. The P-value is also smaller (although both are less than 0.0001). **(e)** With $s_i \doteq 2.8$, $n_i = 115$, $s_o \doteq 0.7$, and $n_o = 220$, we have $s_i^2/n_i \doteq 0.06817$ and $s_o^2/n_o \doteq 0.002227$, so:

$$\text{df} \doteq \frac{(0.06817 + 0.002227)^2}{\frac{0.06817^2}{114} + \frac{0.002227^2}{219}} \doteq 121.5030$$

The pooled standard deviation is $s_p \doteq 1.7338$; the standard errors are $\text{SE}_1 = 0.2653$ (with no assumptions) and $\text{SE}_2 = 0.1995$ (assuming equal standard deviations). The pooled t is 24.56 (df $= 333$, $P < 0.0001$), and the 95% confidence intervals are shown in the table. The pooled and usual t procedures compare similarly to the results for part (d): t is larger, and the interval is narrower.

7.94. (a) For a two-sided test with df $= 1$, the critical value is $t^* = 12.71$. **(b)** With the pooled procedures, df $= 2$ and the critical value is $t^* = 4.303$. **(c)** The smaller critical value with the pooled approach means that a smaller t-value (that is, weaker evidence) is needed to reject H_0.

7.95. (a) From an $F(15, 20)$ distribution, $F^* = 2.20$. **(b)** P is between $2(0.025) = 0.05$ and $2(0.05) = 0.10$; $F = 2.45$ is significant at the 10% level but not at the 5% level.

7.96. The power would be higher. Larger differences are easier to detect; that is, when $\mu_1 - \mu_2$ is more than 5, there is a greater chance that the test statistic will be significant. (In fact, if we repeat the computations of Example 7.23 with $\mu_1 - \mu_2 = 6$, for example, we find that the power increases to about 93%. When $\mu_1 - \mu_2 = 10$, the power is nearly 100%.)

7.97. The power would be higher. A smaller value of σ means that large differences between the sample means would arise less often by chance so that, if we observe such a difference, it gives greater evidence of a difference in the population means. (In fact, software gives the 0.8436, and the Normal approximation gives 0.8458.)

G•Power output

```
- - - - - - - - - - - - - - -  Standard deviation 7.4  - - - - - - - - - - - - - - -
Post-hoc analysis for "t-Test (means)", one-tailed:
Alpha: 0.0100
Power (1-beta): 0.7965
Effect size "d": 0.6757
Total sample size: 90 (n 1:45, n 2: 45)
Critical value: t(88) = 2.3695
Delta: 3.2051
- - - - - - - - - - - - - - - -  Standard deviation 7  - - - - - - - - - - - - - - - -
Power (1-beta): 0.8436
Effect size "d": 0.7143
Total sample size: 90 (n 1:45, n 2: 45)
Critical value: t(88) = 2.3695
Delta: 3.3882
```

7.98. (a) $F = \frac{9.3}{3.1} = 3$. **(b)** For 15 and 9 degrees of freedom, we need $F > 3.77$ to reject H_0 at the 5% level (with a two-sided alternative). **(c)** We do not have enough evidence to conclude that the standard deviations are different. (Software gives $P = 0.1006$.)

7.99. The test statistic is $F = \left(\frac{34.79}{33.24}\right)^2 \doteq 1.0954$, with df 70 and 36. The two-sided P-value is 0.7794, so we do not have enough evidence to conclude that the standard deviations are different. We do not know if the distributions are Normal, so this test may not be reliable. However, with s_1 and s_2 so close together, it seems likely that the conclusion ("we do not have enough evidence. . . ") is appropriate; the reliability of the test would be a more crucial issue if we had rejected H_0.

7.100. The test statistic is $F = \left(\frac{13.75}{7.94}\right)^2 \doteq 2.9989$, with df 70 and 36. The two-sided P-value is 0.0005, so we have strong evidence that the standard deviations are different. However, this test assumes that the underlying distributions are Normal; if this is not true, then the conclusion may not be reliable.

7.101. The test statistic is $F = \frac{1.16^2}{1.15^2} \doteq 1.0175$ with df 211 and 164. Table E tells us that $P > 0.20$, while software gives $P = 0.9114$. The distributions are not Normal ("total score was an integer between 0 and 6"), so the test may not be reliable (although with s_1 and s_2 so close, the conclusion is probably correct). To reject at the 5% level, we would need $F > F^*$, where $F^* = 1.46$ (using df 120 and 100 from Table E) or $F^* = 1.3392$ (using software). As $F = s_2^2/s_1^2$, we would need $s_2^2 > s_1^2 F^*$, or $s_2 > 1.15\sqrt{F^*}$, which is about 1.3896 (Table E) or 1.3308 (software).

7.102. The test statistic is $F = \frac{1.19^2}{1.12^2} \doteq 1.1289$ with df 164 and 211. Table E tells us that $P > 0.2$, while software gives $P = 0.4063$. We cannot conclude that the standard deviations are different. The distributions are not Normal (because all responses are integers from 1 to 5), so the test may not be reliable.

7.103. The test statistic is $F = \frac{7.8^2}{3.7^2} \doteq 5.2630$ with df 114 and 219. Table E tells us that $P < 0.002$, while software gives $P < 0.0001$; we have strong evidence that the standard deviations differ. The authors described the distributions as somewhat skewed, so the Normality assumption may be violated.

7.104. The test statistic is $F = \frac{2.8^2}{0.7^2} = 16$ with df 114 and 219. Table E tells us that $P < 0.002$, while software gives $P < 0.0001$; we have strong evidence that the standard deviations differ. We have no information about the Normality of the distributions, so it is difficult to determine how reliable these conclusions are. (We can observe that for Exercise 7.65, $\bar{x}_1 - 3s_1$ and $\bar{x}_2 - 3s_2$ were both negative, hinting at the skewness of those distributions. For Exercise 7.66, this is not the case, suggesting that these distributions might not be as skewed.)

7.105. The test statistic is $F \doteq \frac{17.5001^2}{14.2583^2} \doteq 1.5064$ with df 29 and 29. Table E tells us that $P > 0.2$, while software gives $P = 0.2757$; we cannot conclude that the standard deviations differ. The stemplots and boxplots of the north/south distributions in Exercise 7.81 do not appear to be Normal (north was right-skewed, south was left-skewed), so the results may not be reliable.

7.106. The test statistic is $F \doteq \frac{16.0743^2}{15.3314^2} \doteq 1.0993$ with df 29 and 29. Table E tells us that $P > 0.2$, while software gives $P = 0.8006$; we cannot conclude that the standard deviations differ. The stemplots and boxplots of the east/west distributions in Exercise 7.82 do not appear to be Normal (east was right-skewed, west was left-skewed), so the results may not be reliable.

7.107. **(a)** With an $F(1, 1)$ distribution with a two-sided alternative, we need the critical value for $p = 0.025$: $F^* = 647.79$. This is a very low-power test, since large differences between σ_1 and σ_2 would rarely be detected. **(b)** To test H_0: $\sigma_1 = \sigma_2$ vs. H_a: $\sigma_1 \neq \sigma_2$, we find $F \doteq \frac{1.5344^2}{0.7707^2} \doteq 3.9634$. Not surprisingly, we do not reject H_0.

7.108. **(a)** For two samples of size 20, we have $\delta = \frac{10}{20\sqrt{2/20}} \doteq 1.5811$. With df $= 38$, $t^* = 2.0244$ (or 2.042 for df $= 30$ from Table D). The approximate power is $P(z > t^* - \delta) = 0.3288$ (0.3228 using Tables A and D). Software computation of the power (using a noncentral t distribution) gives 0.3379. **(b)** With $n_1 = n_2 = 60$, we have $\delta = \frac{10}{20\sqrt{2/60}} \doteq 2.7386$, df $= 118$, and $t^* = 1.9803$ (or 1.984 for df $= 100$ from Table D). The approximate power is $P(z > t^* - \delta) = 0.7759$ (0.7734 using Tables A and D). Software computation of the power (using a noncentral t distribution) gives 0.7753. **(c)** Samples of size 60 would give a reasonably good chance of detecting a difference of 20 cm.

G•Power output

```
- - - - - - - - - - - - - - - -  Sample size: 20 each  - - - - - - - - - - - - - - - -
Post-hoc analysis for "t-Test (means)", two-tailed:
Alpha: 0.0500
Power (1-beta): 0.3379
Effect size "d": 0.5000
Total sample size: 40 (n 1:20, n 2: 20)
Critical value: t(38) = 2.0244
Delta: 1.5811
- - - - - - - - - - - - - - - -  Sample size: 60 each  - - - - - - - - - - - - - - - -
Power (1-beta): 0.7753
Effect size "d": 0.5000
Total sample size: 120 (n 1:60, n 2: 60)
Critical value: t(118) = 1.9803
Delta: 2.7386
```

7.109. The four standard deviations from Exercises 7.81 and 7.82 are $s_n \doteq 17.5001$, $s_s \doteq 14.2583$, $s_e \doteq 16.0743$, and $s_w \doteq 15.3314$ cm. Using a larger σ for planning the study is advisable because it provides a conservative (safe) estimate of the power. For example, if we choose a sample size to provide 80% power and the true σ is smaller than that used for planning, the actual power of the test is greater than the desired 80%.

	Power with $n =$	
σ	20	60
15	0.5334	0.9527
16	0.4809	0.9255
17	0.4348	0.8928
18	0.3945	0.8560

Results of additional power computations will depend on what students consider to be "other reasonable values of σ." Shown in the table are some possible answers using the Normal approximation. (Powers computed using the noncentral t distribution are slightly greater.)

7.110. **(a)** The noncentrality parameter is $\delta = \frac{1.5}{1.6\sqrt{2/65}} \doteq 5.3446$. With such a large value of δ, the value of t^* (1.9787 for df $= 128$, or 1.984 for df $= 100$ from Table D) does not matter very much. The Normal approximation for the power is $P(Z > t^* - \delta) \doteq 0.9996$ for either choice of t^*. Software (see output below) gives the same result. **(b)** For samples of size 100, $\delta \doteq 6.6291$, and once again the value of t^* makes little difference; the power is very close to 1 (using the Normal approximation or software). **(c)** Because the effect is large relative to the standard deviation, small samples are sufficient. (Even samples of size 20 will detect this difference with probability 0.8236.)

G•Power output

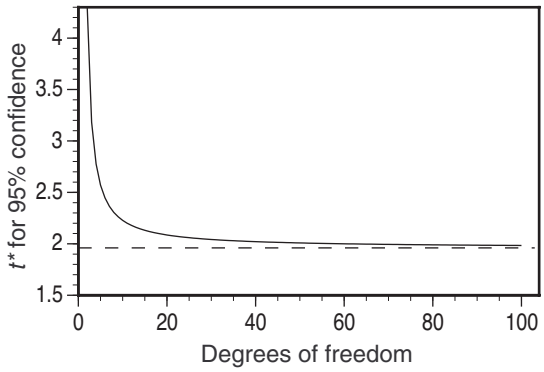
```
- - - - - - - - - - - - - - - -  Sample size: 65 each  - - - - - - - - - - - - - - - -
Post-hoc analysis for "t-Test (means)", two-tailed:
Alpha: 0.0500
Power (1-beta): 0.9996
Effect size "d": 0.9375
Total sample size: 130 (n 1:65, n 2: 65)
Critical value: t(128) = 1.9787
Delta: 5.3446
- - - - - - - - - - - - - - - -  Sample size: 100 each  - - - - - - - - - - - - - - - -
Power (1-beta): 1.0000
Effect size "d": 0.9375
Total sample size: 200 (n 1:100, n 2: 100)
Critical value: t(198) = 1.9720
Delta: 6.6291
```

7.111. The mean is $\bar{x} = 156$, the standard deviation is $s \doteq 10.30$, and the standard error of the mean is $s_{\bar{x}} \doteq 5.15$. It would not be appropriate to construct a confidence interval because we cannot consider these four scores to be an SRS.

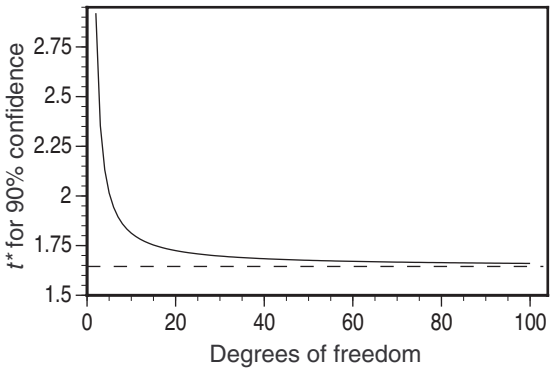
7.112. To support the alternative $\mu_1 > \mu_2$, we need to see $\bar{x}_1 > \bar{x}_2$, so that $t = (\bar{x}_1 - \bar{x}_2)/\mathrm{SE}_D$ must be positive. **(a)** If $t = 1.81$, the one-sided P-value is half of the reported two-sided value: $P = 0.035$. We therefore reject H_0 at $\alpha = 0.05$. **(b)** $t = -1.81$ does not support H_a; the one-sided P-value is 0.965. We do not reject H_0 at $\alpha = 0.05$ (or any other reasonable choice of α).

7.113. The plot (below, left) shows that t^* approaches 1.96 as df increases, at first very rapidly and then more slowly.

For 7.113



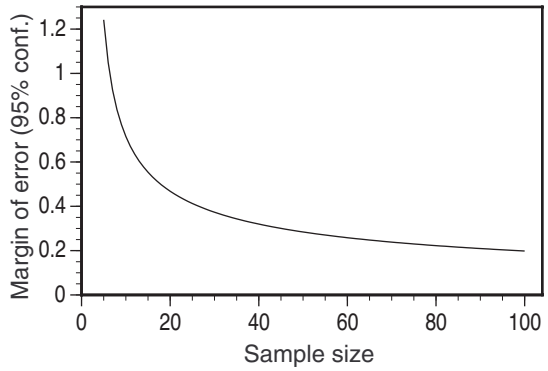
For 7.114



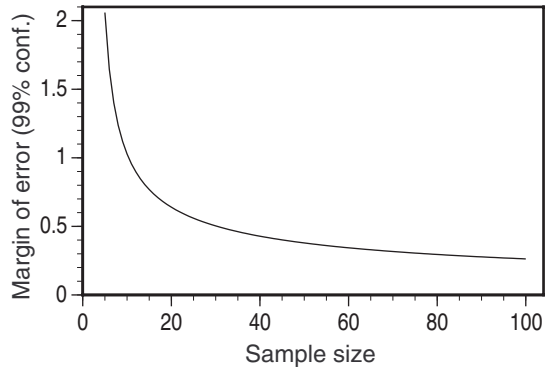
7.114. The plot (above, right) shows that t^* approaches 1.645 as df increases, at first very rapidly, and then more slowly.

7.115. The margin of error is $t^*/\sqrt{n}$, using t^* for df $= n - 1$ and 95% confidence. For example, when $n = 5$, the margin of error is 1.2417, and when $n = 10$, it is 0.7154, and for $n = 100$, it is 0.1984. As the plot (below, left) shows, as sample size increases, margin of error decreases (toward 0, although it gets there very slowly).

For 7.115



For 7.116

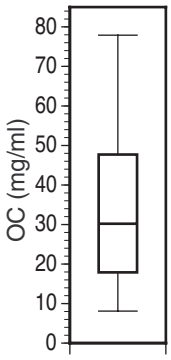


7.116. The margin of error is $t^*/\sqrt{n}$, using t^* for df $= n - 1$ and 99% confidence. For example, when $n = 5$, the margin of error is 2.0590, and when $n = 10$, it is 1.0277, and for $n = 100$, it is 0.2626. As the plot (above, right) shows, as sample size increases, margin of error decreases (toward 0, although it gets there very slowly).

7.117. **(a)** The mean difference in body weight change (with wine minus without wine) was $\bar{x}_1 = 0.4 - 1.1 = -0.7$ kg, with standard error $\text{SE}_1 = 8.6/\sqrt{14} \doteq 2.2984$ kg. The mean difference in caloric intake was $\bar{x}_2 = 2589 - 2575 = 14$ cal, with $\text{SE}_2 = 210/\sqrt{14} \doteq 56.1249$ cal. **(b)** The t statistics $t_i = \bar{x}_i/\text{SE}_i$, both with df $= 13$, are $t_1 = -0.3046$ ($P_1 = 0.7655$) and $t_2 = 0.2494$ ($P_2 = 0.8069$). **(c)** For df $= 13$, $t^* = 2.160$, so the 95% confidence intervals $\bar{x}_i \pm t^*\text{SE}_i$ are -5.6646 to 4.2646 kg (-5.6655 to 4.2655 with software) and -107.2297 to 135.2297 cal (-107.2504 to 135.2504 with software). **(d)** Students might note a number of factors in their discussions; for example, all subjects were males, weighing 68 to 91 kg (about 150 to 200 lb), which may limit how widely we can extend these conclusions.

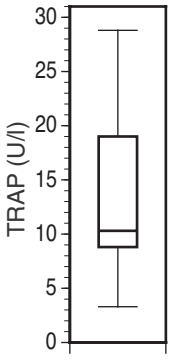
7.118. **(a)** A stemplot and boxplot are shown on the right; the distribution is right-skewed, with $n = 31$ and $\bar{x} \doteq 33.4161$, $s = 19.6097$, Min $= 8.1$, $Q_1 = 17.9$, $M = 30.2$, $Q_3 = 47.7$, Max $= 77.9$ mg/ml. **(b)** Using $t^* = 2.042$ from Table D, the 95% confidence interval is 26.2242 to 40.6081 mg/ml. With $t^* = 2.0423$ from software, the interval is 26.2232 to 40.6090 mg/ml. Because the sample size is fairly large, and there are no extreme outliers, the t procedures should be safe, provided the 31 women in the study can be considered an SRS.

```
0 | 89
1 | 0
1 | 5677799
2 | 00004
2 |
3 | 011
3 | 568
4 | 04
4 | 7
5 | 244
5 | 6
6 |
6 | 8
7 |
7 | 67
```



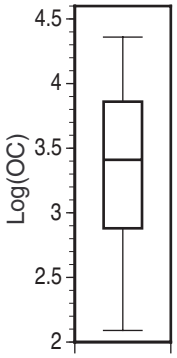
7.119. **(a)** A stemplot and boxplot are shown on the right; the distribution is somewhat right-skewed with no extreme outliers, with $n = 31$ and $\bar{x} \doteq 13.2484$, $s = 6.5282$, Min $= 3.3$, $Q_1 = 8.8$, $M = 10.3$, $Q_3 = 19$, Max $= 28.8$ U/l. **(b)** Using $t^* = 2.042$ from Table D, the 95% confidence interval is 10.8541 to 15.6426 U/l. With $t^* = 2.0423$ from software, the interval is 10.8538 to 15.6430 U/l. Because the sample size is fairly large and there are no extreme outliers, the t procedures should be safe, provided the 31 women in the study can be considered an SRS.

```
0 | 3
0 | 5
0 | 66
0 | 888899999
1 | 00000
1 |
1 | 4444
1 |
1 | 89999
2 |
2 | 3
2 | 55
2 |
2 | 8
```



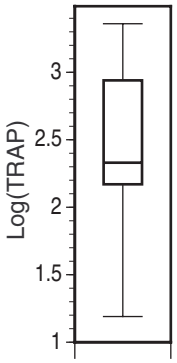
7.120. **(a)** A stemplot and boxplot are shown on the right. The distribution is now roughly symmetrical (perhaps slightly left-skewed), with $n = 31$ and $\bar{x} \doteq 3.3379$, $s = 0.6085$, Min $= 2.09$, $Q_1 = 2.88$, $M = 3.41$, $Q_3 = 3.86$, Max $= 4.36$. **(b)** Using either $t^* = 2.042$ (Table D) or $t^* = 2.0423$ (software), the 95% confidence interval is 3.1147 to 3.5611. The t procedures should be safe, provided the 31 women can be considered an SRS. **(c)** All three numbers are lower than in Exercise 7.118: $e^{\bar{x}} \doteq 28.1605$ mg/ml, while exponentiating the confidence interval yields 22.5271 to 35.2027 mg/ml.

```
2 | 0
2 | 23
2 |
2 | 7
2 | 888899
3 | 00001
3 |
3 | 44455
3 | 667
3 | 89
4 | 000
4 | 233
```



7.121. **(a)** A stemplot and boxplot are shown on the right. The distribution is now slightly left-skewed, with $n = 31$ and $\bar{x} \doteq 2.4674$, $s = 0.4979$, Min $= 1.19$, $Q_1 = 2.17$, $M = 2.33$, $Q_3 = 2.94$, Max $= 3.36$. **(b)** Using either $t^* = 2.042$ (Table D) or $t^* = 2.0423$ (software), the 95% confidence interval is 2.2848 to 2.6501. The t procedures should be safe, provided the 31 women can be considered an SRS. **(c)** All three numbers are lower than in Exercise 7.119: $e^{\bar{x}} \doteq 11.7921$ U/l, while exponentiating the confidence interval yields 9.8236 to 14.1549 U/l.

```
1 | 1
1 |
1 |
1 | 7
1 | 89
2 | 0011
2 | 2222233333
2 |
2 | 6667
2 | 99999
3 | 1
3 | 223
```



7.122. The mean is $\bar{x} \doteq 26.8437$ cm, $s \doteq 18.3311$ cm, and the margin of error is $t^* \cdot s/\sqrt{584}$:

	df	t^*	Interval
Table D	100	1.984	$26.8437 \pm 1.5050 = 25.3387$ to 28.3486 cm
Software	583	1.9640	$26.8437 \pm 1.4898 = 25.3538$ to 28.3335 cm

The confidence interval is much narrower with the whole data set, largely because the standard error is about one-fourth what it was with a sample of size 40. The distribution of the 584 measurements is right-skewed (although not as much as the smaller sample). If we can view these trees as an SRS of similar stands—a fairly questionable assumption—the t procedures should be fairly reliable because of the large n. See the solution to Exercise 7.124 for an examination of the distribution.

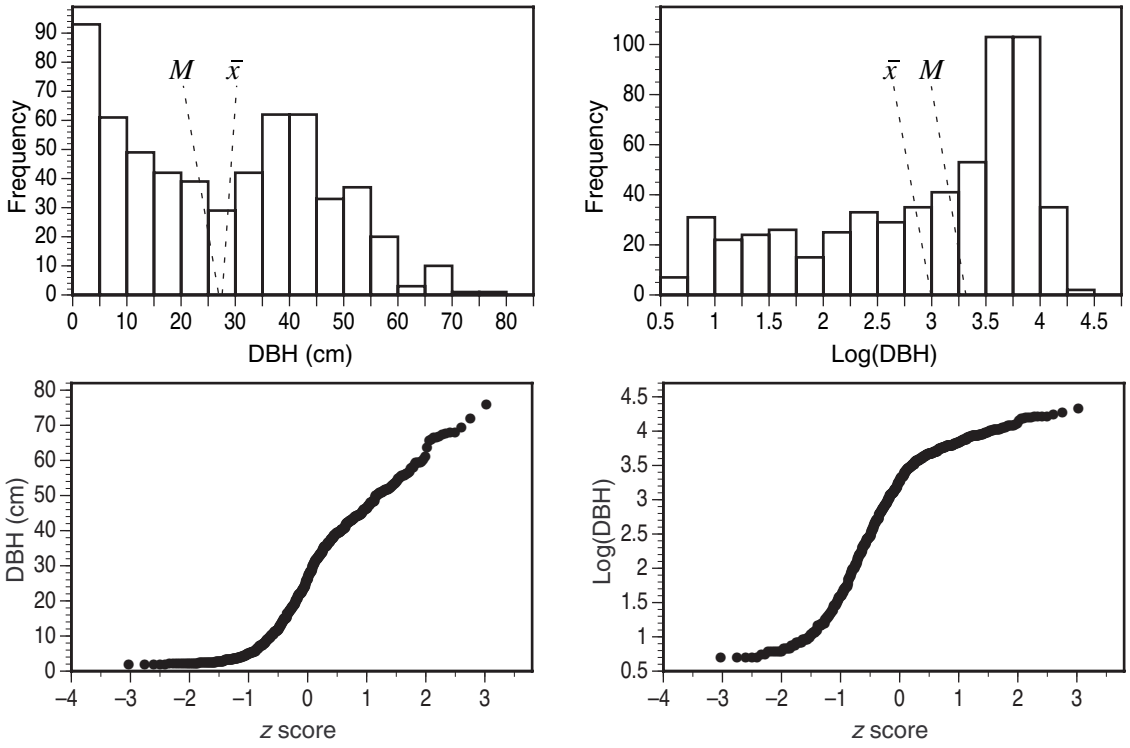
7.123. For north/south differences, the test of H_0: $\mu_n = \mu_s$ gives $t = -7.15$ with df $= 575.4$ or 283; either way, $P < 0.0001$, so we reject H_0. The 95% confidence interval is given in the table below. For east/west differences, $t = -3.69$ with df $= 472.7$ or 230; either way, $P \doteq 0.0003$, so we reject H_0. See the table on the next page for the 95% confidence interval. The larger data set results in smaller standard errors (both are near 1.5, compared to about 4 in Exercises 7.81 and 7.82), meaning that t is larger and the margin of error is smaller.

	$\bar{x}$	s	n
North	21.7990	18.9230	300
South	32.1725	16.0763	284
East	24.5785	17.7315	353
West	30.3052	18.7264	231

	df	t^*	Confidence interval
N–S	575.4	1.9641	−13.2222 to −7.5248
	283	1.9684	−13.2285 to −7.5186
	100	1.984	−13.2511 to −7.4960
E–W	472.7	1.9650	−8.7764 to −2.6770
	230	1.9703	−8.7847 to −2.6687
	100	1.984	−8.8059 to −2.6475

7.124. The histograms and quantile plots are shown below, and the means and medians are given in the table on the right and are marked on the histograms. (The plots were created using natural logarithms; for common logs, the appearance would be roughly the same except for scale.) The transformed data does not look notably more Normal; it is left-skewed instead of right-skewed. The t procedures should be fairly dependable anyway because of the large sample size, but only if we can view the data as an SRS from some population.

	$\bar{x}$	M
Original	26.8437	26.15
Natural log	2.9138	3.2638
Common log	1.2654	1.4175



7.125. The mean and standard deviation of the 25 numbers are $\bar{x} = 78.32\%$ and $s \doteq 33.3563\%$, so the standard error is $\text{SE}_{\bar{x}} \doteq 6.6713\%$. For df $= 24$, Table D gives $t^* = 2.064$, so the 95% confidence interval is $\bar{x} \pm 13.7695\% = 64.5505\%$ to 92.0895% (with software, $t^* = 2.0639$ and the interval is $\bar{x} \pm 13.7688\% = 64.5512\%$ to 92.0888%). This seems to support the retailer's claim: 65% to 93% of the time, the original supplier's price was higher.

7.126. (a) We are interested in weight change; the pairs are the "before" and "after" measurements. **(b)** The mean weight change was a loss. The exact amount lost is not specified, but it was large enough so that it would rarely happen by chance for an ineffective weight-loss program. **(c)** Comparing to a $t(40)$ distribution in Table D, we find $P < 0.0005$ for a one-sided alternative ($P < 0.0010$ for a two-sided alternative). Software reveals that it is even smaller than that: about 0.000013 (or 0.000026 for a two-sided alternative).

7.127. Back-to-back stemplot below. The distributions appear similar; the most striking difference is the relatively large number of boys with low GPAs. Testing the difference in GPAs (H_0: $\mu_b = \mu_g$; H_a: $\mu_b < \mu_g$), we obtain $\text{SE}_D \doteq 0.4582$ and $t = -0.91$, which is not significant, regardless of whether we use df = 74.9 ($P = 0.1839$) or 30 ($0.15 < P < 0.20$). The confidence interval for the difference $\mu_b - \mu_g$ in GPAs is -1.3278 to 0.4978 (df = 74.9) or -1.3506 to 0.5206 (df = 30).

		GPA		IQ	
	n	$\bar{x}$	s	$\bar{x}$	s
Boys	47	7.2816	2.3190	110.96	12.121
Girls	31	7.6966	1.7208	105.84	14.271

For the difference in IQs, we find $\text{SE}_D \doteq 3.1138$. For testing H_0: $\mu_b = \mu_g$ vs. H_a: $\mu_b > \mu_g$, we find $t = 1.64$, which is fairly strong evidence, although it is not quite significant at the 5% level: $P = 0.0528$ (df = 56.9) or $0.05 < P < 0.10$ (df = 30). The confidence interval for the difference $\mu_b - \mu_g$ in IQs is -1.1175 to 11.3535 (df = 56.9) or -1.2404 to 11.4764 (df = 30).

GPA: Girls		Boys
	0	5
	1	7
	2	4
4	3	689
7	4	068
952	5	0
4200	6	019
988855432	7	1124556666899
998731	8	001112238
95530	9	1113445567
17	10	57

IQ: Girls		Boys
42	7	
	7	79
	8	
96	8	
31	9	03
86	9	77
433320	10	0234
875	10	556667779
44422211	11	00001123334
98	11	556899
0	12	03344
8	12	67788
20	13	
	13	6

7.128. The median self-concept score is 59.5. Back-to-back stemplot (below) suggests that high self-concept students have a higher mean GPA and IQ. Testing H_0: $\mu_{\text{low}} = \mu_{\text{high}}$ vs. H_a: $\mu_{\text{low}} \neq \mu_{\text{high}}$ for GPAs leads to $\text{SE}_D \doteq 0.4167$ and $t = 4.92$, which is quite significant ($P < 0.0005$ regardless of df). The confidence interval for the difference $\mu_{\text{high}} - \mu_{\text{low}}$ in GPAs is shown in the table.

		GPA		IQ	
	n	$\bar{x}$	s	$\bar{x}$	s
High SC	39	8.4723	1.3576	114.23	10.381
Low SC	39	6.4208	2.2203	103.62	13.636

	df	t^*	Confidence interval
GPA	62.9	1.9984	1.2188 to 2.8843
	38	2.0244	1.2079 to 2.8951
	30	2.042	1.2006 to 2.9025
IQ	70.9	1.9940	5.1436 to 16.0871
	38	2.0244	5.0601 to 16.1707
	30	2.042	5.0118 to 16.2190

For the difference in IQs, we find $\text{SE}_D \doteq 2.7442$. For testing H_0: $\mu_{\text{low}} = \mu_{\text{high}}$ vs. H_a: $\mu_{\text{low}} \neq \mu_{\text{high}}$, we find $t = 3.87$, which is quite significant ($P < 0.0006$ regardless of df). The confidence interval for the difference $\mu_{\text{high}} - \mu_{\text{low}}$ in IQs is shown in the table.

In summary, both differences are significant; with 95% confidence, high self-concept students have a mean GPA that is 1.2 to 2.9 points higher, and their mean IQ is 5 to 16 points higher.

GPA: Low SC		High SC
5	0	
7	1	
4	2	
864	3	9
8760	4	
9520	5	
94000	6	12
986655421	7	1234556688899
887321100	8	112399
97	9	0111334455556
	10	1577

IQ: Low SC		High SC
42	7	
97	7	
	8	
96	8	
310	9	3
776	9	8
44300	10	22333
9877766	10	55567
44432100	11	00111222334
99985	11	568
4	12	00334
7	12	67888
	13	02
	13	6

7.129. It is reasonable to have a prior belief that people who evacuated their pets would score higher, so we test H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 > \mu_2$. We find $\text{SE}_D \doteq 0.4630$ and $t = 3.65$, which gives $P < 0.0005$ no matter how we choose degrees of freedom (115 or 237.0). As one might suspect, people who evacuated their pets have a higher mean score.

df	t^*	Confidence interval
237.0	1.9700	0.7779 to 2.6021
115	1.9808	0.7729 to 2.6071
100	1.984	0.7714 to 2.6086

One might also compute a 95% confidence interval for the difference; these are given in the table.

7.130. (a) "se" is standard error (of the mean). To find s, multiply the standard error by $\sqrt{n}$. **(b)** No: We test H_0: $\mu_d = \mu_c$ vs. H_a: $\mu_d < \mu_c$, for which $SE_D \doteq 65.1153$ and $t \doteq -0.3532$, so $P = 0.3623$ (df = 173.9) or 0.3625 (df = 82)—in either case, there is little evidence against H_0. **(c)** The evidence is not very significant: To test H_0: $\mu_d = \mu_c$ vs. H_a: $\mu_d \neq \mu_c$, $SE_D \doteq 0.1253$, $t \doteq -1.1971$, for which $P = 0.2335$ (df = 128.4) or 0.2348 (df = 82). **(d)** The 95% confidence interval is $0.39 \pm t^*(0.11)$. With Table D, $t^* = 1.990$ (df = 80) and the interval is 0.1711 to 0.6089 g; with software, $t^* = 1.9893$ (df = 82) and the interval is 0.1712 to 0.6088 g. **(e)** The 99% confidence interval is $(0.24 - 0.39) \pm t^*\sqrt{0.06^2 + 0.11^2}$; see the table.

		Calories		Alcohol	
	n	$\bar{x}$	s	$\bar{x}$	s
Drivers	98	2821	435.58	0.24	0.59397
Conductors	83	2844	437.30	0.39	1.00215

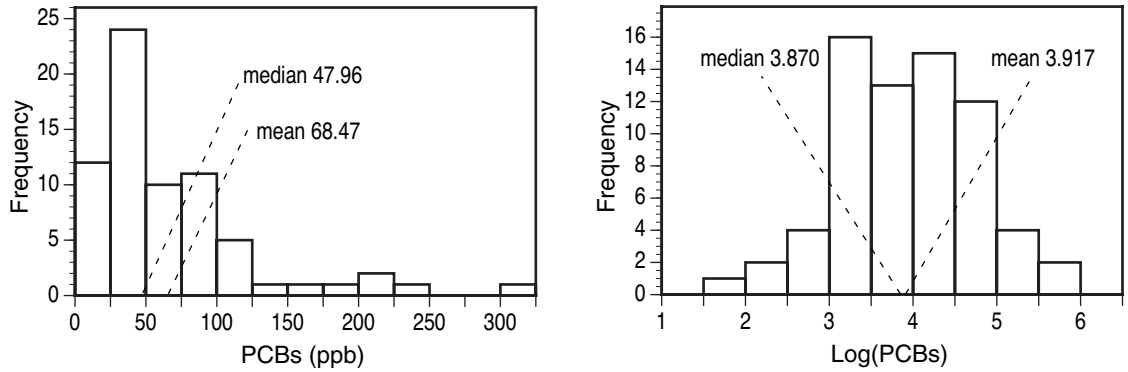
df	t^*	Confidence interval
128.4	2.6146	−0.4776 to 0.1776
82	2.6371	−0.4804 to 0.1804
80	2.639	−0.4807 to 0.1807

7.131. The similarity of the sample standard deviations suggests that the population standard deviations are likely to be similar. The pooled standard deviation is $s_p \doteq 436.368$ and $t \doteq -0.3533$, so $P = 0.3621$ (df = 179)—still not significant.

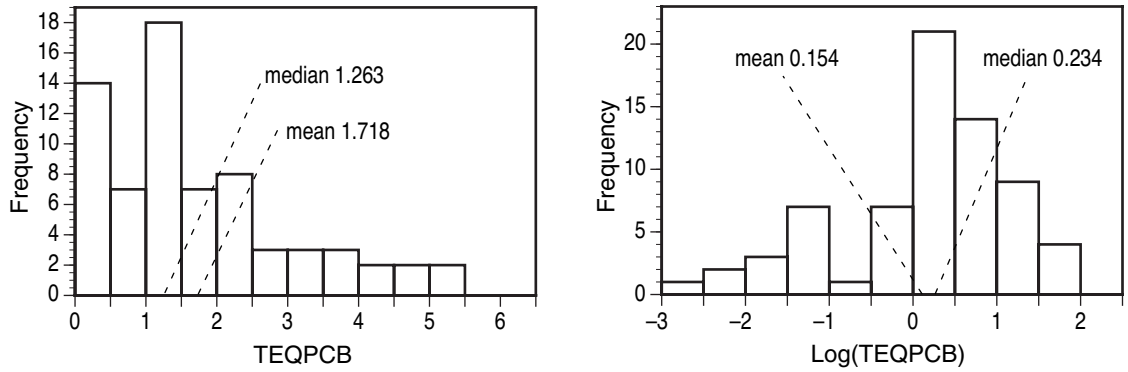
7.132. (a) The sample sizes (98 and 83) are quite large, so the t test should be reasonably safe (provided there are no extreme outliers). **(b)** Large samples do not make the F test more reliable when the underlying distributions are skewed, so it should not be used.

7.133. No: What we have is nothing like an SRS of the population of Islamic men. (Additionally, the average of the 17 numbers in Table 1.2 would not really be an appropriate measure of the literacy rate among Islamic men; it would not take into account the different population sizes in the 17 countries.)

7.134. (a) A histogram (on the next page, left) shows that the distribution is strongly right-skewed. Aside from the mean 68.4674 and median 47.96 shown in the histogram, $s \doteq 59.3906$ ppb. **(b)** The skewness, and more importantly the outliers in the distribution, make the t procedures questionable. The sample size *might* be enough to overcome this. **(c)** The 95% confidence interval is either 54.1678 to 82.7669 ($t^* = 2.000$ with df = 60 from Table D) or 54.2002 to 82.7345 ($t^* = 1.9955$ with df = 68 from software). This interval does *not* include 95% of the observations, nor should we expect it to; a confidence interval is a range of values for the *mean,* not for individual observations. **(d)** The logPCB distribution (histogram on the next page, right) appears to be much more like a Normal distribution. The mean is $\bar{x} \doteq 3.9171$, the median is $M \doteq 3.8704$, and $s \doteq 0.8020$. (This assumes natural logarithms; for common logs, multiply these numbers by 0.4343.) A 95% confidence interval for the mean is 3.7239 to 4.1101 ($t^* = 2.000$ with df = 60 from Table D) or 3.7244 to 4.1097 ($t^* = 1.9955$ with df = 68 from software). These limits translate to 41.4 to 60.9 ppb. (Note that the mean of the original data is not in this interval—an indication of the strength of the skew.) The t procedures are almost certainly more appropriate for the transformed data, as the outliers and skewness are no longer present after taking logarithms. **(e)** Visiting this web site should give your students a glimpse at the details of planning a major statistical study.



7.135. (a) A histogram (below, left) shows that the distribution is strongly right-skewed. Aside from the mean 1.7182 and median 1.2633 shown in the histogram, $s \doteq 1.3428$. **(b)** The sample size should be enough to overcome the skewness. (Unlike the PCB levels in the previous exercise, this distribution has no outliers.) **(c)** The 95% confidence interval is either 1.3949 to 2.0415 ($t^* = 2.000$ with df $= 60$ from Table D) or 1.3956 to 2.0408 ($t^* = 1.9955$ with df $= 68$ from software). This interval does *not* include 95% of the observations, nor should we expect it to; a confidence interval is a range of values for the *mean,* not for individual observations. **(d)** The logTEQPCB distribution (histogram below, right) is left-skewed, although it may be somewhat more Normal than the untransformed data. The mean is $\bar{x} \doteq 0.1542$, the median is $M \doteq 0.2337$, and $s \doteq 1.0176$. (This assumes natural logarithms; for common logs, multiply these numbers by 0.4343.) A 95% confidence interval for the mean is -0.0908 to 0.3992 ($t^* = 2.000$ with df $= 60$ from Table D) or -0.0902 to 0.3987 ($t^* = 1.9955$ with df $= 68$ from software). These limits translate to 0.914 to 1.490. (Note that the mean of the original data is not in this interval—an indication of the strength of the skew.)



7.136. For the 35 noninfected children (low CRP), $\bar{x}_1 \doteq 0.7780$ and $s_1 \doteq 0.3359$; for the 55 infected children, $\bar{x}_2 \doteq 0.6198$ and $s_2 \doteq 0.3397$. Back-to-back stemplot or other comparison of the two distributions shows that both distributions are right-skewed, and the infected group has a high outlier. We cautiously use t procedures: The 95% confidence interval for $\mu_1 - \mu_2$ is:

df	t^*	Confidence interval
73.2	1.9929	0.0128 to 0.3036
34	2.0322	0.0099 to 0.3064
30	2.042	0.0092 to 0.3072

High CRP		Low CRP
6443	2	3
9888766554433310	3	1455559
211000	4	69
976652	5	89
998877	6	2777
8	7	0
742221	8	368
77	9	0499
96	10	0248
953	11	1357
30	12	2
	13	6
	14	4
	15	
	16	
	17	
	18	
0	19	

The researchers were interested in testing H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 > \mu_2$. For this test, we find $t = 2.168$, for which $P = 0.0167$ (df $= 73.2$) or $P = 0.0186$ (df $= 34$). We have significant evidence at $\alpha = 0.05$ and conclude that retinol levels are lower in recently infected children.

7.137. The stemplot on the right shows that the distribution is slightly right-skewed with one high outlier. Only 17 of the 90 children had AGP levels above 1 g/l (although 9 more had levels above 0.9). The mean AGP level for the sample was $\bar{x} = 0.804\overline{3}$ g/l, and the standard deviation was $s \doteq 0.2765$ g/l. A 95% confidence interval for the population mean was 0.7463 to 0.8623 g/l ($t^* = 1.990$, df $= 80$ from Table D) or 0.7464 to 0.8622 g/l ($t^* = 1.9870$, df $= 89$ from software).

3	03
4	11355588899
5	45566899
6	011133557788999
7	00333366678
8	001224557999
9	00000126667999
10	256
11	11467888
12	
13	344
14	04
15	
16	
17	
18	2

7.138. **(a)** Back-to-back stemplot below; summary statistics on the right. With a pooled standard deviation, $s_p \doteq 83.6388$, $t \doteq 4.00$ with df $= 222$, so $P < 0.0001$. Without pooling, $\text{SE}_D \doteq 11.6508$, $t \doteq 4.01$ with df $= 162.2$, and again $P < 0.0001$ (or, with df $= 78$, we conclude that $P < 0.0005$). The test for equality of standard deviations gives $F \doteq 1.03$ with df 144 and 78; the P-value is 0.9114, so the pooled procedure should be appropriate. In either case, we conclude that male mean SATM scores are higher than female mean SATM scores. Options for a 95% confidence interval for the male − female difference are given in the table below. **(b)** With a pooled standard deviation, $s_p \doteq 92.6348$, $t \doteq 0.94$ with df $= 222$, so $P = 0.3485$. Without pooling, $\text{SE}_D \doteq 12.7485$, $t \doteq 0.95$ with df $= 162.2$, so $P = 0.3410$ (or, with df $= 78$, $P = 0.3426$). The test for equality of standard deviations gives $F \doteq 1.11$ with df 144 and 78; the P-value is 0.6033, so the pooled procedure should be appropriate. In either case, we cannot see a difference between male and female mean SATV scores. Options for a 95% confidence interval for the male − female difference are given in the table below. **(c)** The results may generalize fairly well to students in different years, less well to students at other schools, and probably not very well to college students in general.

		SATM		**SATV**	
	n	$\bar{x}$	s	$\bar{x}$	s
Men	145	611.772	84.0206	508.841	94.3485
Women	79	565.025	82.9294	496.671	89.3849

df	t^*	Confidence interval for SATM $\mu_M - \mu_F$
162.3	1.9747	23.7402 to 69.7538
78	1.9908	23.5521 to 69.9419
70	1.994	23.5154 to 69.9786
222	1.9707	23.6978 to 69.7962
100	1.984	23.5423 to 69.9517

df	t^*	Confidence interval for SATV $\mu_M - \mu_F$
167.9	1.9742	−12.9981 to 37.3381
78	1.9908	−13.2104 to 37.5504
70	1.994	−13.2506 to 37.5906
222	1.9707	−13.3584 to 37.6984
100	1.984	−13.5306 to 37.8706

Men's SATM		Women's SATM
	3	0
	3	5
400	4	1334
99999888776	4	56777888999
44444333322222111000	5	0111123334
999999888888777666555555555	5	55555556667777788899999
444444444433333332222222211100000000	6	00011222233334444
999999987777776666655555	6	55555789
3222211100000	7	1124
77766655555	7	
0	8	

Men's SATV		Women's SATV
98	2	9
4322	3	33
9999988766	3	55669
444444444332111100000	4	0122223333444
999998888888888877777666666555	4	566666677777788888899999
444433333222211100000000000000	5	01111122334
998888777777666666555	5	566777777889
43333332111100000	6	0000
9987775	6	668
420	7	00
6	7	5

7.139. **(a)** We test H_0: $\mu_B = \mu_D$ vs. H_a: $\mu_B < \mu_D$. Pooling might be appropriate for this problem, in which case $s_p \doteq 6.5707$. Whether or not we pool, $\text{SE}_D \doteq 1.9811$ and $t \doteq 2.87$ with df $= 42$ (pooled), 39.3, or 21, so $P = 0.0032$, or 0.0033, or 0.0046. We conclude that the mean score using DRTA is higher than the mean score with the Basal method. The difference in the average scores is 5.68; options for a 95% confidence interval for the difference $\mu_D - \mu_B$ are given in the table below. **(b)** We test H_0: $\mu_B = \mu_S$ vs. H_a: $\mu_B < \mu_S$. If we pool, $s_p \doteq 5.7015$. Whether or not we pool, $\text{SE}_D \doteq 1.7191$ and $t \doteq 1.88$ with df $= 42$, 42.0, or 21, so $P = 0.0337$, or 0.0337, or 0.0372. We conclude that the mean score using Strat is higher than the Basal mean score. The difference in the average scores is 3.23; options for a 95% confidence interval for the difference $\mu_S - \mu_B$ are given in the table below.

	n	$\bar{x}$	s
Basal	22	41.0455	5.6356
DRTA	22	46.7273	7.3884
Strat	22	44.2727	5.7668

df	t^*	Confidence interval for $\mu_D - \mu_B$
39.3	2.0223	1.6754 to 9.6882
21	2.0796	1.5618 to 9.8018
21	2.080	1.5610 to 9.8026
42	2.0181	1.6837 to 9.6799
40	2.021	1.6779 to 9.6857

df	t^*	Confidence interval for $\mu_S - \mu_B$
42.0	2.0181	−0.2420 to 6.6966
21	2.0796	−0.3477 to 6.8023
21	2.080	−0.3484 to 6.8030
42	2.0181	−0.2420 to 6.6965
40	2.021	−0.2470 to 6.7015

7.140. For testing $\mu_1 = \mu_2$ against a two-sided alternative, we would reject H_0 if $t = \frac{\bar{x}_1 - \bar{x}_2}{\text{SE}_D}$ is greater (in absolute value) than t^*, where $\text{SE}_D = 2.5\sqrt{2/n}$. (Rather than determining t^* for each considered sample size, we might use $t^* \doteq 2$.) We therefore want to choose n so that

$$P\left(\left|\frac{\bar{x}_1 - \bar{x}_2}{\text{SE}_D}\right| > t^*\right) = 1 - P\left(-t^* < \frac{\bar{x}_1 - \bar{x}_2}{\text{SE}_D} < t^*\right) = 0.90$$

when $\mu_1 - \mu_2 = 0.4$ inch. With $\delta = 0.4/\text{SE}_D \doteq 0.1131\sqrt{n}$, this means that

$$P\left(-t^* - \delta < Z < t^* - \delta\right) = 0.10$$

where Z has a $N(0, 1)$ distribution. By trial and error, we find that two samples of size 822 will do this. (This answer will vary slightly depending on what students do with t^* in the formula above.)

Note: *Determining the necessary sample size can be made a bit easier with some software, or a Web site like* `http://calculators.stat.ucla.edu/powercalc/`. *The output of the G•Power program below gives the total sample size as 1644 (i.e., two samples of size 822). The Web site tells us that, "The equal sample size for both groups is calculated to be 821.854."*

G•Power output

```
A priori analysis for "t-Test (means)", two-tailed:
Alpha: 0.0500
Power (1-beta): 0.9000
Effect size "d": 0.1600
Total sample size: 1644
Actual power: 0.9001
Critical value: t(1642) = 1.9614
Delta: 3.2437
```

3BR		4BR
7776	0	
9988	0	
11110	1	
32222222	1	23
544	1	55
	1	7
9	1	
0	2	
	2	23
5	2	4
6	2	
	2	9

7.141. **(a)** The distributions can be compared using a back-to-back stemplot (shown), or two histograms, or side-by-side boxplots. Both distributions are right-skewed; four-bedroom homes are generally more expensive. The top three prices from the three-bedroom distribution qualify as outliers using the $1.5 \times IQR$ criterion. Boxplots are probably a poor choice for displaying the distributions because they leave out so much detail, but five-number summaries do illustrate that four-bedroom prices are higher at every level. Summary statistics are given in the table below. **(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 reject H_0 and conclude that the mean prices are different (specifically, that 4BR houses are more expensive). **(c)** The one-sided alternative $\mu_3 < \mu_4$ could have been justified because it would be reasonable to expect that four-bedroom homes would be more expensive. **(d)** The 95% confidence interval for the difference $\mu_4 - \mu_3$ is about \$19,182 to \$111,614 (df $= 12.1$) or \$16,452 to \$114,344 (df $= 8$). **(e)** While the data were not gathered from an SRS, it seems that they should be a fair representation of three- and four-bedroom houses in West Lafayette. (Even so, the small sample size for four-bedroom houses, the skewness of the data, and the outliers in the three-bedroom data should make us cautious about the t procedures. Additionally, we might question independence in these data: When setting the asking price for a home, sellers are almost certainly influenced by the asking prices for similar homes on the market in the area.)

	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