Chapter 18 Solutions

18.1. This is a matched-pairs design: attitudes of husbands and wives may be connected, so we should keep them together.

18.2. This involves two independent samples: those students who read the *Wall Street Journal* ads, and those who read the *National Enquirer* ads.

18.3. This involves a single sample.

18.4. This involves two independent samples (because the result of, for example, the first measurement using the new method is independent of the first measurement using the old method).

18.5. **(a)** If the loggers had known that a study would be done, they might have (consciously or subconsciously) cut down fewer trees than they typically would, in order to reduce the impact of logging.
(b) STATE: Does logging significantly reduce the mean number of species in a plot after 8 years?
PLAN: We test H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 > \mu_2$, where μ_1 is the mean number of species in unlogged plots and μ_2 is the mean number of species in plots logged 8 years earlier. We use a one-sided alternative because we expect that logging reduces the number of tree species.

Unlogged		Logged	
13	000	0	4
14		0	
15	00	0	
16		1	0
17		1	2
18	0	1	455
19	00	1	7
20	0	1	88
21	0		
22	00		

SOLVE: We assume that the data come from SRSs of the two populations. Stemplots (above) suggest some deviation from Normality, and a possible low outlier for the logged-plot counts. Note that these stemplots do not allow for easy comparison of the two distributions because the software used to create them used different scales (leaf units). In spite of these concerns, we proceed with the t test. The means and standard deviations are given in the Minitab output below; we compute SE $= \sqrt{\frac{3.53^2}{12} + \frac{4.50^2}{9}} \doteq 1.813$ and $t = \frac{17.50-13.67}{1.813} \doteq 2.11$. With the conservative df $= 8$, we find that $0.025 < P < 0.05$. Note that Minitab uses the more accurate df $= 14$ (truncated from the true computed df $= 14.8$), rather than the conservative approach. For this df, we find that $P = 0.026$. If we remove the low outlier from the logged data, a few things change: $\bar{x}_2 = 14.875$, $s_2 \doteq 2.8504$, SE $= \sqrt{\frac{3.53^2}{12} + \frac{2.85^2}{8}} \doteq 1.433$, and $t \doteq 1.832$. Now we have either $0.05 < P < 0.1$ (df $= 7$) or $P = 0.042$ (df $= 17.2$).
CONCLUDE: If we use all the data, we have fairly strong evidence (significant at 5% but not at 1%) that logged plots have fewer species. If we have a reason to remove the low outlier from the logged data, the evidence is weaker, although it is still significant when we use the more accurate df.

Minitab output

```
Twosample T for Species
Code   N      Mean      StDev    SE Mean
1     12     17.50       3.53        1.0
2      9     13.67       4.50        1.5

90% C.I. for mu 1 - mu 2: ( 0.6,  7.0)
T-Test mu 1 = mu 2 (vs >): T= 2.11  P=0.026  DF=  14
```

18.6. STATE: Do lean and obese people differ in the average time they spend lying down?

Lean		Obese
9	3	
	4	1
	4	
5	4	44
6	4	6
8	4	
10	5	001
33	5	23
5	5	
6	5	6

PLAN: Let μ_1 be the mean time spent lying down by the lean group, and μ_2 be the mean time for the obese group. We test H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 \neq \mu_2$. (The question asked in the exercise suggests a two-sided alternative, but we might reasonably expect that lean people spend less time lying down, and so choose H_a: $\mu_1 < \mu_2$. As we shall see, that will not matter.)

SOLVE: See Example 18.2 for a discussion of the conditions for inference. The back-to-back stemplot on the right shows some irregularity, but no clear departures from Normality. They also suggest little difference between the two groups—an impression which is supported by the means:

	n	$\bar{x}$	s
Lean	10	501.6461	52.0449
Obese	10	491.7426	46.5932

$$\text{SE} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \doteq 22.0898$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\text{SE}} \doteq 0.448$$

The standard error and the test statistic are given above. With such a small t value, the P-value will obviously be quite large regardless of the degrees of freedom; in fact, $P = 0.6596$ (df $= 17.8$) or $P = 0.6647$ (df $= 9$).

CONCLUDE: There is no reason to reject the hypothesis that lean and moderately obese people spend (on the average) the same amount of time lying down.

18.7. Use the means and the standard error SE $\doteq 1.813$ found in the solution to Exercise 18.5, and critical value $t^* = 1.860$ (df $= 8$) or $t^* = 1.755$ (df $= 14.8$). The 90% confidence interval is $\bar{x}_1 - \bar{x}_2 \pm t^*\text{SE}$, which gives 0.46 to 7.20 species (df $= 8$) or 0.65 to 7.02 species (df $= 14.8$). Minitab's result, shown in the solution to Exercise 18.5 above, is based on the truncated df $= 14$.

18.8. To test H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 > \mu_2$ (which is equivalent to the hypotheses stated in Figure 18.5), we have $t \doteq 0.9889$, df $\doteq 4.65$, and $P = 0.1857$. This gives us little reason to doubt that fabric buried two weeks and fabric buried 16 weeks have the same mean breaking strength.

18.9. **(a)** Back-to-back stemplots of the time data are shown below on the left. They appear to be reasonably Normal, and the discussion in the exercise justifies our treating the data as independent SRSs, so we can use the t procedures.

We wish to test H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 < \mu_2$, where μ_1 is the population mean time in the restaurant with no scent, and μ_2 is the mean time with a lavender odor. The means and standard deviations (in minutes), as well as the standard error and test statistic, are

	n	$\bar{x}$	s
No scent	30	$91.2\overline{6}$	14.9296
Lavender	30	105.7	13.1048

$$\text{SE} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \doteq 3.6269$$

$$t = \frac{91.2\overline{6} - 105.7}{3.6269} \doteq -3.98$$

For this value of t, $P < 0.0005$ (df $= 29$) or $P = 0.0001$ (df $= 57.041$). This is strong evidence that customers stay longer when the lavender odor is present. **(b)** Back-to-back stemplots of the spending data are below on the right. The distributions are skewed and have many gaps. Again we test H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 < \mu_2$, where μ_1 is the population mean spending with no scent, and μ_2 is the mean spending with a lavender odor. The means and standard deviations (in euros), as well as the standard error and test statistic, are

	n	$\bar{x}$	s
No scent	30	17.5133	2.3588
Lavender	30	21.1233	2.3450

$$\text{SE} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \doteq 0.6073$$

$$t = \frac{17.5133 - 21.1233}{0.6073} \doteq -5.95$$

For this value of t, $P < 0.0005$ (df $= 29$) or P is very small (df $= 57.998$). This is strong evidence that customers spend more when the lavender odor is present.

No scent		Lavender
98	6	
322	7	
965	7	6
44	8	
7765	8	89
32221	9	234
86	9	578
31	10	1234
9776	10	5566788999
	11	4
85	11	6
1	12	14
	12	69
	13	
	13	7

No scent		Lavender
9	12	
	13	
	14	
99999999999999	15	
	16	
	17	
555555555555	18	55555555555
	19	
5	20	7
9	21	5599999999
	22	3558
	23	
	24	99
5	25	59

18.10. **(a)** The "compressed" stemplot shows no particular cause for concern. The "intermediate" stemplot shows the skewness and outlier described in the text. Note that these stemplots do not allow for easy comparison of the two distributions because the software used to create them used different scales (leaf units). **(b)** We wish to test H_0: $\mu_C = \mu_I$ vs. H_a: $\mu_C < \mu_I$. The summary statistics are:

Compressed	
26	8
27	
27	6888
28	122
28	66
29	024
29	68
30	0
30	88
31	
31	68

Intermediate	
2	99
3	0111111
3	2333
3	4445
3	6
3	8
4	
4	2

	n	$\bar{x}$	s
Compressed	20	2.9075	0.1390
Intermediate	19	3.2874	0.2397

We compute SE $= \sqrt{\frac{0.1390^2}{20} + \frac{0.2397^2}{19}} \doteq 0.06317$ and $t = \frac{2.9075-3.2874}{0.06317} \doteq -6.013$. Regardless of the chosen df (conservative 18, or the software value 28.6), the P-value is very small; this is very significant evidence that the mean penetrability for compressed soil is lower than the mean for intermediate soil.

Minitab output

```
Twosample T for Comp vs Inter
        N      Mean     StDev   SE Mean
Comp   20     2.908     0.139     0.031
Inter  19     3.287     0.240     0.055

95% C.I. for mu Comp - mu Inter: ( -0.509,  -0.250)
T-Test mu Comp = mu Inter (vs <): T= -6.01  P=0.0000  DF=  28
```

18.11. We have two small samples (each $n = 4$), so t procedures are not reliable unless both distributions are Normal.

18.12. Use the means and the standard error SE $\doteq$ 0.06317 found in the solution to Exercise 18.10, and critical value $t^* = 2.101$ (df = 18) or $t^* = 2.047$ (df = 28.6). The 95% confidence interval is $\bar{x}_1 - \bar{x}_2 \pm t^*\text{SE}$, which gives -0.5126 to -0.2471 (df = 18) or -0.5092 to -0.2506 (df = 28.6). Minitab's result, shown in the solution to Exercise 18.10, is based on the truncated df = 28.

18.13. Here are the details of the computations:

$$\text{SE}_F = \frac{12.6961}{\sqrt{31}} \doteq 2.2803$$

$$\text{SE}_M = \frac{12.2649}{\sqrt{47}} \doteq 1.7890$$

$$\text{SE} = \sqrt{\text{SE}_F^2 + \text{SE}_M^2} \doteq 2.8983$$

$$\text{df} = \frac{\text{SE}^4}{\frac{1}{30}\left(\frac{12.6961^2}{31}\right)^2 + \frac{1}{46}\left(\frac{12.2649^2}{47}\right)^2} = \frac{70.565}{1.1239} \doteq 62.8$$

$$t = \frac{55.5161 - 57.9149}{\text{SE}} \doteq -0.8276$$

18.14. The means and standard deviations are not given in Figure 18.5, but they are easily computed. They are shown in the table on the right, along with s_1^2/n_1 and s_2^2/n_2, which we need to find the standard error and degrees of freedom. Here are the details of the computations:

Group	$\bar{x}$	s	s^2/n
2 weeks	123.8	4.6043	4.24
18 week	116.4	16.0873	51.76

$$\text{SE} = \sqrt{4.24 + 51.76} \doteq 7.4833$$

$$\text{df} = \frac{\text{SE}^4}{\frac{1}{4}(4.24)^2 + \frac{1}{4}(51.76)^2} = \frac{3136}{674.2688} \doteq 4.651$$

$$t = \frac{123.8 - 116.4}{\text{SE}} \doteq 0.9889$$

18.15. Reading from the software output shown in Exercise 18.13, we find that there was no significant difference in mean Self-Concept Scale scores for men and women ($t = -0.8276$, df = 62.8, and $P = 0.4110$).

18.16. (a) We have a single sample, and only one score from each member of the sample.

18.17. (c) We consider the boys and girls as being two independent samples.

18.18. (b) Measure the amount of pollutant in each carp using both methods, and examine the differences between those pairs of measurements.

18.19. (b) This is the definition of robust.

18.20. (a) We have two independent samples, each of size $n = 8$, so df = 7.

18.21. (b) $\text{SE} = \sqrt{\frac{9^2}{8} + \frac{13^2}{8}} \doteq 5.59$, and $t = \frac{105-89}{\text{SE}} \doteq 2.86$.

18.22. (c) Random-digit dialing should produce something close to an SRS, and the samples are easily large enough to overcome non-Normality (especially because the scale 0 to 20 limits skewness and outliers in the distribution of scores).

18.23. (a) Our alternative hypothesis should express our suspicion that men are more prone to road rage than women. (We assume that a high road-rage score means a subject is *more* prone to road rage.)

18.24. (b) Using Table C, refer to df = 100, where we see that $3.174 < t < 3.390$. This means that P is between 0.0005 and 0.001.

18.25. (a) To test the belief that women talk more than men, we use a one-sided alternative: H_0: $\mu_M = \mu_F$ vs. H_a: $\mu_M < \mu_F$. (b)–(d) The t statistics, degrees of freedom, and P-values are given in the table below. The two-sample t statistics are $\frac{\bar{x}_F - \bar{x}_M}{\sqrt{s_F^2/n_F + s_M^2/n_M}}$, so positive values would tend to support H_a. The conservative df is the smaller sample size, minus 1.

Study	t	df	Table C values	P-value
1	−0.248	55	$\lvert t\rvert < 0.679$	$P > 0.25$
2	1.507	19	$1.328 < t < 1.729$	$0.05 < P < 0.10$

For Study 1, we use df = 50 in Table C. (e) The first study gives no support to the belief that women talk more than men; the second study gives only moderate (not significant) support.

18.26. (a) The two standard deviations are $s_u = 7\sqrt{9} = 21$ and $s_r = 10\sqrt{11} \doteq 33.1662$. (b) Using Option 2, with sample sizes 9 and 11, we take df = 8. (c) We find $\text{SE} = \sqrt{7^2 + 10^2} \doteq 12.2065$ g. With $t^* = 1.860$ (df = 8), the 90% confidence interval is $(59 - 32) \pm t^*\text{SE} \doteq 4.301$ to 49.699 g. (For comparison, using the software df $\doteq 17.08$, $t^* = 1.7392$ and the interval is 5.771 to 48.229 g.)

18.27. (a) If $\text{SEM} = s/\sqrt{n}$, then $s = \text{SEM} \times \sqrt{n}$. Arithmetic gives the table on the right. (b) The smaller sample has $n = 6$, so the conservative df is 5. (c) $\text{SE} = \sqrt{1.56^2 + 2.68^2} \doteq 3.1010$, so $t = \frac{26.9 - 11.9}{\text{SE}} \doteq 4.837$. (d) Because $4.773 < t < 5.893$, the two-sided P-value is between 0.002 and 0.005. (Software gives $P = 0.0047$.)

Location	n	$\bar{x}$	s
Oregon	6	26.9	3.82
California	7	11.9	7.09

18.28. (a) Parents who choose a Montessori school probably have different attitudes about education than other parents. (b) Over 72% of Montessori parents participated in the study, compared to less than 28% of the other parents. (c) To test H_0: $\mu_M = \mu_C$ vs. H_a: $\mu_M \neq \mu_C$, we find $\text{SE} = \sqrt{\frac{3.11^2}{19} + \frac{4.19^2}{17}} \doteq 1.0122$, so the two-sample test statistic is $t = \frac{19 - 17}{\text{SE}} \doteq 1.976$. This is not quite significant the 5% level: $P = 0.0545$ with df = 43.5, or $0.05 < P < 0.10$ with df = 24.

18.29. **(a)** A placebo is an inert pill that allows researchers to account for any psychological benefit (or detriment) the subject might get from taking a pill. **(b)** Neither the subjects nor the researchers who worked with them knew who was getting ginkgo extract; this prevents expectations or prejudices from affecting the evaluation of the effectiveness of the treatment. **(c)** $\text{SE} = \sqrt{\frac{0.01462^2}{21} + \frac{0.01549^2}{18}} \doteq 0.0048$, so the two-sample test statistic is $t = \frac{0.06383 - 0.05342}{\text{SE}} \doteq 2.147$. This is significant at the 5% level: $P = 0.0387$ (df $\doteq 35.35$) or $0.04 < P < 0.05$ (df $= 17$). Those who took gingko extract had significantly more misses per line.

18.30. "Do Hispanic and Anglo bank customers differ?" calls for two-sided tests: H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 \neq \mu_2$ (for both reliability and empathy). The table below gives the standard errors, t statistics, degrees of freedom, and P-values for both tests. (The conservative df would be 85, but for use with Table C, we must take df $= 80$.) Both results are very significant; there is strong evidence that Anglos value reliability more than Hispanics do, and that Hispanics value empathy more than Anglos do.

	$\text{SE} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$t = \frac{\bar{x}_1 - \bar{x}_2}{\text{SE}}$	*Conservative* df	*Conservative* P	*Software* df	*Software* P
Reliability	0.1182	3.892	80	< 0.001	143.69	0.00015
Empathy	0.1196	−3.595	80	< 0.001	171.08	0.00042

18.31. To test H_0: $\mu_A = \mu_E$ vs. H_a: $\mu_A \neq \mu_E$, we find $\text{SE} = \sqrt{\frac{0.60^2}{12} + \frac{0.57^2}{9}} \doteq 0.2571$, so the two-sample test statistic is $t = \frac{1.92 - 1.74}{\text{SE}} \doteq 0.7$. This is clearly not significant: $P > 0.5$ (df $= 8$) or $P = 0.4930$ (df $= 17.85$). There is no evidence of a difference in mean stress level between Asian and European mothers.

18.32. **(a)** The appropriate test is the matched-pairs test because a student's score on Try 1 is certainly correlated with his/her score on Try 2. **(b)** To test H_0: $\mu = 0$ vs. H_a: $\mu > 0$, we compute $t = \frac{29-0}{59/\sqrt{427}} \doteq 10.16$ with df $= 426$, which is certainly significant ($P < 0.0005$). Coached students do improve their scores. The TI-83 output screen (on the right) shows that the P-value is, in fact, *much* smaller than 0.0005! **(c)** Table C gives $t^* = 2.626$ for df $= 100$, while software gives $t^* = 2.587$ for df $= 426$. The confidence interval is $29 \pm t^* 59/\sqrt{427} = 21.50$ to 36.50 points (or 21.61 to 36.39, using the software critical value). The TI-83 output confirms the more exact interval.

```
T-Test
 μ>0
 t=10.1568707
 p=3.773541E-22
 x̄=29
 Sx=59
 n=427
```

```
TInterval
 (21.612,36.388)
 x̄=29
 Sx=59
 n=427
```

18.33. (a) The hypotheses are H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 > \mu_2$, where μ_1 is the mean gain among all coached students, and μ_2 the mean gain among uncoached students. We find SE $= \sqrt{59^2/427 + 52^2/2733} \doteq 3.0235$ and $t = \frac{29-21}{3.0235} \doteq 2.646$ with conservative df $= 426$ or software df $= 534.45$. Comparing with df $= 100$ critical values in Table C, we find $0.0025 < P < 0.005$; software gives $P \doteq 0.004$ for df $= 534.45$. There is evidence that coached students had a greater average increase. (b) The 99% confidence interval is $8 \pm 3.0235t^*$, where t^* equals 2.626 (df $= 100$, from Table C) or 2.585 (df $= 534.45$). This gives either 0.06 to 15.94 points, or 0.184 to 15.816 points. (c) Increasing one's score by 0 to 16 points is not likely to make a difference in being granted admission to, or receiving scholarships from, any colleges.

```
2-SampTTest
 μ1>μ2
 t=2.645931232
 p=.0041933105
 df=534.4495612
 x̄1=29
↓x̄2=21
```

```
2-SampTInt
 (.18405,15.816)
 df=534.4495612
 x̄1=29
 x̄2=21
 Sx1=59
↓Sx2=52
```

18.34. This was an observational study, not an experiment. The students (or their parents) chose whether or not to be coached; students who choose coaching might have other motivating factors that help them do better the second time. For example, perhaps students who choose coaching have some personality trait that also compels them to try harder the second time.

18.35. (a) The back-to-back stemplots on the right confirm the description given in the text. (b) Means and standard deviations are given in the table on the right. Testing H_0: $\mu_G = \mu_B$ vs. H_a: $\mu_G \neq \mu_B$, we find SE $\doteq 3.1138$ and $t \doteq 1.64$. This gives $0.10 < P < 0.20$ (df $= 30$) or $P = 0.1057$ (df $= 56.9$), so it is not strong enough evidence to reject H_0.

	n	$\bar{x}$	s
Girls	31	105.84	14.271
Boys	47	110.96	12.121

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

18.36. (a) Based on the stemplots, the t procedures should be safe. Both the stemplots and the means suggest that customers stayed (very slightly) longer when there was no odor. (b) Testing H_0: $\mu_N = \mu_L$ vs. H_a: $\mu_N \neq \mu_L$, we find SE $\doteq$ 3.9927 and $t \doteq 0.371$. This is not at all significant: $P > 0.5$ (df = 27) or $P = 0.7121$ (df = 55.4). We cannot conclude that mean time in the restaurant is different when the lemon odor is present.

	n	$\bar{x}$	s
No odor	30	91.2667	14.9296
Lemon	28	89.7857	15.4377

No odor		Lemon
	5	6
	6	03
98	6	
322	7	34
965	7	58
44	8	33
7765	8	8889
32221	9	0144
86	9	677
31	10	14
9776	10	5688
	11	23
85	11	
1	12	

18.37. The 95% confidence interval is $(110.96 - 105.84) \pm t^*\text{SE}$, where SE $\doteq$ 3.1138 and $t^* = 2.042$ (df = 30) or 2.003 (df = 56.9). These intervals are -1.24 to 11.48 or -1.12 to 11.35.

18.38. (a) The sample means in the table on the right suggest that the Permafresh swatches are slightly stronger. (The second Permafresh line was computed with the low outlier omitted.) (b) Shown on the right are back-to-back stemplots for the two processes. (c) With the means and standard deviations listed in the table, we find the following test statistics and P-values:

	n	$\bar{x}$	s
Permafresh	5	29.54	1.1675
Permafresh*	4	30.025	0.4992
Hylite	5	25.20	2.6693

Permafresh		Hylite
	22	1
	23	9
	24	2
	25	
	26	
6	27	0
	28	8
95	29	
70	30	

		Conservative		*Software*	
	t	df	P	df	P
All points	3.33	4	$0.02 < P < 0.04$	5.48	0.0184
Outlier removed	3.96	3	$0.02 < P < 0.04$	4.35	0.0147

The mild outlier in the Permafresh sample had almost no effect on our conclusion. Despite the small samples, there is good evidence (significant at $\alpha = 0.05$) that the mean breaking strengths of the two processes differ.

18.39. (a) The Hylite mean is greater than the Permafresh mean. (b) Shown on the right are back-to-back stemplots for the two processes, which confirm that there are no extreme outliers. (c) We find SE $\doteq$ 1.334 and $t \doteq -6.296$, for which the P-value is $0.002 < P < 0.005$ (df = 4) or 0.0003 (df = 7.779). There is very strong evidence of a difference between the population means. As we might expect, the stronger process (Permafresh) is less resistant to wrinkles.

	n	$\bar{x}$	s
Permafresh	5	134.8	1.9235
Hylite	5	143.2	2.2804

Permafresh		Hylite
2	13	
54	13	
76	13	
	13	
	14	11
	14	3
	14	5
	14	6

18.40. The 90% confidence interval is $\bar{x}_1 - \bar{x}_2 \pm t^*\text{SE}$, where $t^* = 2.132$ (df = 4) or $t^* = 1.981$ (df = 5.476). This gives either

$$4.34 \pm 2.778 = 1.562 \text{ to } 7.118 \text{ pounds (df} = 4\text{), or}$$
$$4.34 \pm 2.581 = 1.759 \text{ to } 6.921 \text{ pounds (df} = 5.476\text{).}$$

As usual, the conservative interval is wider than the more accurate software result.

18.41. The 90% confidence interval is $\bar{x}_1 - \bar{x}_2 \pm t^*\text{SE}$, where $t^* = 2.132$ (df = 4) or $t^* = 1.867$ (df = 7.779). This gives either

$$-8.4 \pm 2.844 = -11.244^\circ \text{ to } -5.556^\circ \text{ (df} = 4\text{), or}$$
$$-8.4 \pm 2.491 = -10.891^\circ \text{ to } -5.909^\circ \text{ (df} = 7.779\text{).}$$

As usual, the conservative interval is wider than the more accurate software result.

18.42. This is a two-sample t statistic, comparing two independent groups (supplemented and control). Using the conservative df = 5, $t = -1.05$ would have a P-value between 0.30 and 0.40, which (as the report said) is not significant. ($t = -1.05$ would not be significant for any df.)

18.43. To test H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 \neq \mu_2$, we find $\text{SE} = \sqrt{\frac{3.10934^2}{6} + \frac{3.92556^2}{7}} \doteq 1.95263$, and $t = \frac{4.0-11.3}{\text{SE}} \doteq -3.74$. The two-sided P-value is either $0.01 < P < 0.02$ (df = 5) or 0.0033 (df = 10.96), agreeing with the stated conclusion (a significant difference).

	n	$\bar{x}$	s
Control	6	4.0	3.10934
Supplemented	7	11.3	3.92556

18.44. These are paired t statistics: for each bird, the number of days behind the caterpillar peak was observed, and the t values were computed based on the pairwise differences between the first and second years.

For the control group, df = 5, and for the supplemented group, df = 6. The control t is not significant (so the birds in that group did *not* "advance their laying date in the second year"), while the supplemented group t is significant with one-sided $P = 0.0195$ (so those birds did change their laying date).

18.45. PLAN: We test H_0: $\mu_T = \mu_C$ vs. H_a: $\mu_T > \mu_C$, using a one-sided alternative because we suspect that the treatment group will wait longer before asking for help.

	n	$\bar{x}$	s
Treatment	17	314.0588	172.7898
Control	17	186.1176	118.0926

Treatment		Control
65	0	5689
3	1	012444
976	1	58
44	2	
5	2	79
	3	
6	3	7
44	4	01
876	4	
3	5	
	5	
0	6	

SOLVE: We must assume that the data comes from an SRS of the intended population (whatever we consider that to be); we cannot check this with the data. The back-to-back stemplot (right) shows some irregularity in the treatment times, and skewness in the control times. We hope that our equal and moderately large sample sizes will overcome any deviation from Normality.

With the means and standard deviations listed in the table, we find SE $\doteq$ 50.7602 and $t \doteq 2.521$, for which $0.01 < P < 0.02$ (df = 16) or $P = 0.0088$ (df = 28.27).

CONCLUDE: There is strong evidence that the treatment group waited longer to ask for help on the average.

18.46. PLAN: We test H_0: $\mu_A = \mu_P$ vs. H_a: $\mu_A > \mu_P$; the one-sided alternative is suggested by the statement of the exercise—and presumably by a suspicion that active learning is better.

	n	$\bar{x}$	s
Active	24	24.4167	6.31022
Passive	24	17.8750	4.02506

Active		Passive
	1	223
5	1	45555
76	1	66777
	1	889
111100	2	00111
332	2	
4444	2	5
7	2	66
9888	2	
1	3	
	3	
5	3	
	3	
	3	
	4	
	4	
4	4	

SOLVE: We must assume that the data comes from an SRS of the population of learning-impaired children; we cannot check this with the data. We also assume that the data are close to Normal. The back-to-back stemplot (right) shows a high outlier and some skewness in the "active" scores, but with reasonably large (and equal) sample sizes such as those we have here, we can allow some variation from Normality. With the means and standard deviations listed in the table, we find SE $\doteq$ 1.5278 and $t \doteq 4.28$. With either df = 23 or df = 39.1, $P < 0.0005$.

CONCLUDE: There is very strong evidence that active learning results in more correct identifications.

18.47. PLAN: Compare mean gains by testing H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 > \mu_2$, and by finding a 90% confidence interval for $\mu_1 - \mu_2$.

SOLVE: We assume that we have two SRSs, and that the distributions of score improvements are Normal. Shown on the right are stemplots of the differences ("after" minus "before") for the two groups; the samples are too small to assess Normality, but we can see that there are no outliers. With the means and standard deviations listed in the table, we find SE $\doteq$ 1.646 and $t \doteq 1.914$, for which the P-value is $0.025 < P < 0.05$ (df = 7) or 0.0382 (df = 13.92). The 90% confidence interval is $(11.40 - 8.25) \pm t^*\text{SE}$, where $t^* = 1.895$ (df = 7) or $t^* = 1.762$ (df = 13.92): either 0.03 to 6.27 points, or 0.25 to 6.05 points.

	n	$\bar{x}$	s
Treatment	10	11.4	3.1693
Control	8	8.25	3.6936

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

CONCLUDE: We have evidence (significant at 5%) that the encouraging subliminal message led to a greater improvement in math scores. We are 90% confident that this increase is between 0.03 and 6.27 points (or 0.25 and 6.05 points).

18.48. **(a)** The 90% confidence interval is $(24.4167 - 17.8750) \pm t^*\text{SE}$, where SE $\doteq$ 1.5278 and $t^* = 1.714$ (df = 23) or 1.685 (df = 39.05). These intervals are 3.92 to 9.16 Blissymbols, or 3.97 to 9.12 Blissymbols. **(b)** With $t^* = 1.714$ from a $t(23)$ distribution, the 90% confidence interval is $24.4167 \pm (1.714)(6.31022/\sqrt{24}) = 22.2$ to 26.6 Blissymbols. (Note that this is a one-sample question.)

18.49. PLAN: Compare mean length by testing H_0: $\mu_r = \mu_y$ vs. H_a: $\mu_r \neq \mu_y$, and by finding a 95% confidence interval for $\mu_r - \mu_y$.

SOLVE: We must assume that the data comes from an SRS. We also assume that the data are close to Normal. The back-to-back stemplots (right) show some skewness in the red lengths, but the t procedures should be reasonably safe. With the means and standard deviations listed in the table, we find SE $\doteq$ 0.4518 and $t \doteq 7.817$. With either df = 14 or df = 35.1, $P < 0.001$. The 95% confidence interval is $(39.711 - 36.180) \pm t^*\text{SE}$, where $t^* = 2.145$ (df = 14) or $t^* = 2.030$ (df = 35.1): either 2.562 to 4.500 mm, or 2.614 to 4.448 mm.

	n	$\bar{x}$	s
Red	23	39.7113	1.7988
Yellow	15	36.1800	0.9753

Red		Yellow
	34	56
	35	146
	36	0015678
9874	37	01
8722100	38	1
761	39	
65	40	
9964	41	
10	42	
0	43	

CONCLUDE: We have very strong evidence that the two varieties differ in mean length. We are 95% confident that the mean red length minus yellow length is between 2.562 and 4.500 mm (or 2.614 and 4.448 mm).

18.50. *Because this exercise asks for a "complete analysis," without suggesting hypotheses or confidence levels, student responses may vary. This solution gives 95% confidence intervals for the means in parts (a) and (b), and performs a hypothesis test and gives a 95% confidence interval for part (c). Note that the first two problems call for single-sample t procedures (Chapter 17), while the last uses the Chapter 18 procedures. Student answers should be formatted according to the "four-step process" of the text; these answers are not, but can be used to check student results.*

We begin (as students should) with summary statistics and a display of the distributions, using either stemplots or histograms:

	n	$\bar{x}$	s
Women	95	4.2737	2.1472
Men	81	6.5185	3.3471

Women		Men
00000000	1	000
5555555500000	2	0000
555500000000000000000	3	0000000
500000000000000000	4	0000000000555
00000000000	5	000000005
50000000	6	00000005
00000000	7	000000005
000	8	000000000
000	9	0000
00	10	00000005
	11	0
	12	05
	13	
	14	
	15	000
	16	0

(a) The stemplot shows that the distribution of claimed drinks per day for women is right-skewed, but has no particular outliers, so with such a large sample, the t procedures should be safe. Let μ_w be the mean claimed drinks per day for sophomore women. The standard error for the women's mean is $\text{SE}_w = s_w/\sqrt{95} \doteq 0.2203$, and the margin of error for 95% confidence is $t^*\text{SE}_w$, where $t^* = 1.990$ (df $= 80$) or $t^* = 1.9855$ (df $= 94$). Therefore, the 95% confidence interval for μ_w is either 4.2737 ± 0.4384 or 4.2737 ± 0.4374. With either df, we could say that we are 95% confident that, among sophomore women who drink, the mean claimed number of drinks is between 3.84 and 4.71 drinks.

(b) The stemplot shows that the distribution of claimed drinks per day for men has four high numbers that may be considered outliers (exaggerations). Apart from these four numbers, the distribution is fairly symmetric. The t procedures should be safe, even with these high numbers included (they are not too extreme). Let μ_m be the mean claimed drinks per day for sophomore men. The standard error for the men's mean is $\text{SE}_m = s_m/\sqrt{81} \doteq 0.3719$, and the margin of error for 95% confidence is $t^*\text{SE}_m = 1.990\,\text{SE}_m = 0.7401$ (df $= 80$). Therefore, we are 95% confident that, among sophomore men who drink, the mean claimed number of drinks is in the range $6.5185 \pm 0.7401 = 5.78$ to 7.26 drinks.

(c) For the two-sided test H_0: $\mu_w = \mu_m$ vs. H_a: $\mu_w \neq \mu_m$, we find

$$\text{SE} = \sqrt{\frac{2.1472^2}{95} + \frac{3.3471^2}{81}} \doteq 0.4322 \quad \text{and} \quad t = \frac{4.2737 - 6.5185}{\text{SE}} \doteq -5.193.$$

Regardless of the choice of df (80 or 132.15), this is highly significant ($P < 0.001$); we have very strong evidence that the claimed number of drinks is different for men and women. To construct a 95% confidence interval for $\mu_m - \mu_w$, we take $\bar{x}_m - \bar{x}_w \pm t^*\text{SE}$, with $t^* = 1.990$ (df $= 80$) or $t^* = 1.9781$ (df $= 132.15$). This gives either 2.2448 ± 0.8601 or 2.2448 ± 0.8549. After rounding either interval, we can report with 95% confidence that, on the average, sophomore men who drink claim an additional 1.4 to 3.1 drinks per day compared to sophomore women who drink.