Chapter 20 – Inference about a Population Mean

20.1 The standard error of the mean is $s/\sqrt{n} = 27.2/\sqrt{1000} = 0.8601$ minute.

20.2 The mean is 27 kg, and the standard error of the mean is $s/\sqrt{n} = 7.9/\sqrt{9} =$ 2.633 kg.

20.3 (a) $t^* = 6.965$. **(b)** $t^* = 1.701$.

20.4 (Here, df = 25 − 1 = 24.) **(a)** $t^* = 2.172$. **(b)** $t^* = 0.857$.

20.5 (a) df = 2 − 1 = 1, so $t^* = 6.314$. **(b)** df = 20 − 1 = 19, so $t^* = 2.093$. **(c)** df = 1001 − 1 = 1000, so $t^* = 2.581$.

20.6 (a) The provided stemplot shows a slight skew to the right, but not so strong that it would invalidate the *t* procedures.

```
0 | 67888
1 | 0004
1 | 67
2 | 012
2 | 57
3 | 00
3 | 78
```

(b) With $\bar{x}$ = 18.66, s = 10.2768, and t^*= 2.093 (df = 19), the 95% confidence interval for μ is $18.66 \pm 2.093\frac{10.2768}{\sqrt{20}}$ = 18.66 ± 4.8096 = 13.8504 to 23.4696.

20.7 STATE: What is the mean percent μ of the number of correct answers when people are told to identify the taller of two speakers by voice? PLAN: We will estimate μ with a 99% confidence interval. SOLVE: We are told to view the observations as an SRS. The provided stemplot shows some possible bimodality but no outliers. With $\bar{x}$ = 62.1667% and s = 5.806% correct, and t^* = 2.807 (df = 23), the 99% confidence interval for μ is $62.1667 \pm 2.807\frac{5.806}{\sqrt{24}}$ = 62.1667 ± 3.3267 = 58.84% to 65.49%. CONCLUDE: We are 99% confident that the mean percent of correct answers to identifying the taller of two people by voice is between 58.84% and 65.49%.

```
4 | 9
5 | 3
5 | 6668889
6 | 1123
6 | 556777889
7 | 00
```

20.8 (a) df = 20 − 1 = 19. **(b)** t = 2.10 is bracketed by $t^* = 2.093$ (with right-tail probability 0.025) and $t^* = 2.205$ (with right-tail probability 0.02). That is, $0.02 < P < 0.025$. **(c)** This test is significant at the 10% and 5% levels, since $P < 0.10$ and $P < 0.05$. It is not significant at the 1% level because $P > 0.01$. **(d)** From software, $P = 0.0247$.

20.9 **(a)** df = 3 − 1 = 2. **(b)** $t = 2.10$ is bracketed by $t^* = 1.886$ (with two-tail probability 0.20) and $t^* = 2.920$ (with two-tail probability 0.10). That is, $0.10 < P < 0.20$. **(c)** This test is not significant at the 10%, 5%, or 1% levels, since P > 0.10. **(d)** From software, $P = 0.1705$.

20.10 STATE: Can people really identify height from voice alone? Is there evidence that the mean percent of correct identifications of the taller of two individuals from voice alone is more than 50% (just guessing)? PLAN: We test H_0: $\mu = 50\%$ versus H_a: $\mu > 50\%$. SOLVE: We addressed the conditions for inference in Exercise 20.7. In that solution, we found $\bar{x} = 62.1667\%$ and $s = 5.806\%$ correct identifications, so $t = \frac{62.1667 - 50}{5.806/\sqrt{24}} = 10.27$. For df = 23, this is beyond anything shown in Table C, so $P < 0.0005$ (software gives $P = 2.32 \times 10^{-10}$). CONCLUDE: We have extremely strong evidence ($P \approx 0$) that people can identify the taller of two speakers solely from voice (at least better than just guessing).

20.11 PLAN: Take μ to be the mean difference (with eye grease minus without) in sensitivity. We test H_0: $\mu = 0$ versus H_a: $\mu > 0$, using a one-sided alternative, because if the eye grease works, it should increase sensitivity. SOLVE: We must assume that the students in the experiment can be regarded as an SRS of all students, that the treatments were randomized, and that athletes would experience a similar effect as the students. We were provided the difference for each student; the provided stemplot of these differences seems to show two outliers (in this plot $-1|8$ represents −0.18). Checking with the 1.5 × IQR rule, these are not outliers. (Using JMP, $Q_1 = -0.095$, $Q_3 = 0.27$, and the upper fence is 0.8175. Calculating these values by hand, $Q_1 = -0.08$, $Q_3 = 0.26$, and the upper fence is 0.77.) However, P-values will only be approximate due to the skew and relatively small sample size. The mean and standard deviation are $\bar{x} = 0.1013$ and $s = 0.2263$, so $t = \frac{0.1013 - 0}{0.2263/\sqrt{16}} = 1.79$ with df = 15. Using Table C, $P < 0.05$ (software gives 0.0469). CONCLUDE: We have evidence that eye grease does increase sensitivity to contrast. Due to the skew in the data, we may not want to place much emphasis on this result.

-1	8621
-0	5
0	23557
1	4
2	489
3	
4	3
5	
6	4

20.12 Using the information in Exercise 20.11, with t^* = 2.947, the 99% confidence interval is given by $0.1013 \pm 2.947 \frac{0.2263}{\sqrt{16}}$ = 0.1013 ± 0.1667 = −0.0654 to 0.2680. We note that this interval includes 0 (the difference is not significant at the 1% level).

20.13 The provided stemplot suggests that the distribution of nitrogen content is heavily skewed, with a strong outlier. Although t procedures are robust, they should not be used if the population being sampled is this heavily skewed. In this case, t procedures are not reliable.

```
0 | 00000000000111
0 | 2222233
0 | 44
0 |
0 |
1 |
1 |
1 | 4
```

20.14 The provided stemplot of carbon-13 ratios suggests no strong skew, so the use of t procedures is appropriate, assuming the sample was random. With $\bar{x}$ = −2.8825, s = 1.036, df = 24 − 1 = 23, and t^* = 1.714, a 90% confidence interval for the mean carbon-13 ratio is given by $-2.8825 \pm 1.714 \frac{1.036}{\sqrt{24}}$ = −2.8825 ± 0.3625 = -3.245 to -2.520.

```
-4 | 200
-3 | 9888765
-3 | 2
-2 | 8877
-2 | 4310
-1 | 85
-1 | 31
-0 | 8
```

20.15 (b) z requires that you know the population standard deviation σ, while t does not. We virtually never know the value of σ.

20.16 (b) -2. $t = \frac{98 - 100}{4/\sqrt{16}} = -2$.

20.17 (a) 8. df = 9 − 1 = 8.

20.18 (b) falls between 0.01 and 0.05. From software, P = 0.0345.

20.19 (b) 1.638. Here, df = 4 − 1 = 3.

20.20 (b) t < −7.453 or t > 7.453.

20.21 (b) \$58,808 to \$71,192. The interval is computed as $65{,}000 \pm 2.064 \frac{15{,}000}{\sqrt{25}}$.

20.22 (a) You notice that there is a clear outlier in the data. Both the t and z procedures are constructed for bell-shaped data; the t procedures are used when σ is unknown.

20.23 (b) You interview a sample of 15 instructors and another sample of 15 students and ask each how many hours per week homework assignments require. If you separately sample 15 instructors and 15 students, no matching is present.

20.24 (c) the data can be regarded as an SRS from the population.

20.25 For the student group: $t = \frac{0.08 - 0}{0.37/\sqrt{12}}$ = 0.749 (not 0.49, as stated). For the nonstudent group: $t = \frac{0.35 - 0}{0.37/\sqrt{12}}$ = 3.277 (instead of 3.25, a difference that might be due to roundoff error). From Table C, the first P-value (assuming a two-sided alternate hypothesis) is between 0.4 and 0.5 (software gives 0.47), and the second P-value is between 0.005 and 0.01 (software gives 0.007).

20.26 With $\bar{x}$ = 26.8 and s = 7.42, and using either t^* = 1.984 (using df = 100 from Table C) or t^* = 1.9636 (using df = 653 with software), the 95% confidence interval for mean BMI is $26.8 \pm t^* \frac{7.42}{\sqrt{654}}$, which is computed as 26.8 ± 0.5756 = 26.2244 to 27.3756 (using t^* = 1.984), or 26.8 ± 0.5697 = 26.2303 to 27.3697 (using t^* = 1.9636).

20.27 (a) The sample size is very large, so the only potential hazard is extreme skewness. Because scores range only from 0 to 500, there is a limit to how skewed the distribution could be. **(b)** From Table C, we have t^* = 2.581 (df = 1000), or using software, we have t^* = 2.580 (df = 1099). For either value of t^*, the 99% confidence interval is 271 ± t^*(1.3) = 267.6 to 274.4, when rounded to one decimal place. **(c)** Because the 99% confidence interval for μ is entirely above 262, we can believe that the mean for all Dallas eighth-graders is above the basic level (above 262).

20.28 (a) We have df = 23 − 1 = 22, so t^*= 2.074. A 95% confidence interval for the mean solution time is $11.58 \pm 2.074 \frac{4.37}{\sqrt{23}}$ = 11.58 ± 1.89 = 9.69 to 13.47 seconds. **(b)** We must assume that the 23 individuals in the neutral group can be regarded as an SRS from the population. Since the sample size is at least 15, we don't need to assume that the population is Normal. Indeed, t procedures can be used as long as the distribution of solution times for the neutral group is not heavily skewed, and as long as there are no strong outliers in the sample.

20.29 (a) A person's weight before and after wearing the device would not be independent; thus, the data from this before-and-after study should be analyzed using a matched pairs t test. **(b)** Let μ be the mean weight difference (weight after 24 months minus weight before using the wearable technology). We test $H_0: \mu = 0$

versus $H_a: \mu < 0$. The alternative hypothesis says that the weight after wearing the technology is less than the weight before wearing the technology. The problem gives $\bar{x} = -3.5$, s = 7.8, and $n = 237$, so $t = \frac{-3.5 - 0}{7.8/\sqrt{237}} = -6.908$. With df = 236, P < 0.0001 (using software). There is an extreme amount of evidence that there is a decrease in average weight after wearing the device.

20.30 (a) A stemplot is provided. Because the sample size is 29, we look to see if the data have strong outliers; none are seen. The stemplot is roughly symmetric (possibly somewhat right-skewed) with one peak.

Stem	Leaf	Count
5	7	1
5		
5	3	1
5	00	2
4	9	1
4	66	2
4	5	1
4	22	2
4	1	1
3	89999	5
3	6777	4
3	4	1
3	2233	4
3	0	1
2	8	1
2	66	2

2|6 represents 2.6

(b) With $\bar{x}$ = 3.9172, s = 0.796, df = 28, and t^* = 2.048, a 95% confidence interval for the mean trustworthiness of Mitt Romney's face is $3.9172 \pm 2.048\frac{0.796}{\sqrt{29}}$ = 3.6145 to 4.2199. **(c)** Because 3.5 is not contained in the 95% confidence interval computed in part (b), we would reject $H_0: \mu = 3.5$ at the 5% significance level in favor of the two-sided alternative. Because the interval is entirely above 3.5, we have significant evidence that the mean trustworthiness rating is greater than 3.5.

20.31 (a) A stemplot is provided; it suggests the presence of outliers. The sample is small and the stemplot is skewed, so the use of t procedures is not appropriate.

2	5
3	3358
4	00
5	
6	
7	
8	
9	
10	
11	5
12	
13	5

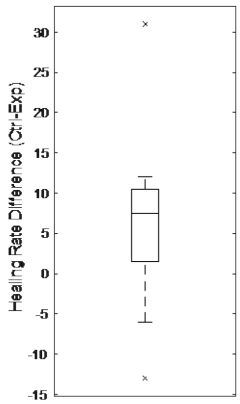
(b) We will compute two confidence intervals, as called for. In the first interval, using all nine observations, we have df = 8 and t^* = 1.860. For the second interval, removing the two outliers (1.15 and 1.35), df = 6 and t^* = 1.943. The two 90% confidence intervals are 0.549 $\pm$ 1.860(0.403/$\sqrt{9}$) = 0.299 to 0.799 gram, and 0.349 $\pm$ 1.943(0.053/$\sqrt{7}$) = 0.310 to 0.388 gram. **(c)** The confidence interval computed without the two outliers is much narrower and has a much lower center. Using fewer data values reduces degrees of freedom (yielding a larger value of t^*). Typically, smaller sample sizes yield larger margins of error. However, both of these effects are offset by removing two values far from the others, and s reduces from 0.403 to 0.053.

20.32 SOLVE: The mean is $\bar{x}$ = 25.42 degrees, the standard deviation is s = 7.47 degrees, and t^* = 2.042 (using df = 30 with Table C) or t^* = 2.0262 (using df = 37 with software). The confidence interval is nearly identical in both cases: 25.42 $\pm$ $t^*(7.47/\sqrt{38}) = 22.95$ to 27.89 degrees ($t^* = 2.042$), or 22.96 to 27.88 degrees ($t^* = 2.0262$). CONCLUDE: We are 95% confident that the mean HAV angle among such patients is between 22.95 degrees and 27.89 degrees.

20.33 (a) The control and experimental limbs are matched by newt because a newt's ability to heal in one leg is not independent of the ability to heal in the other leg. **(b)** The provided stemplot clearly shows the high outlier mentioned in the text.

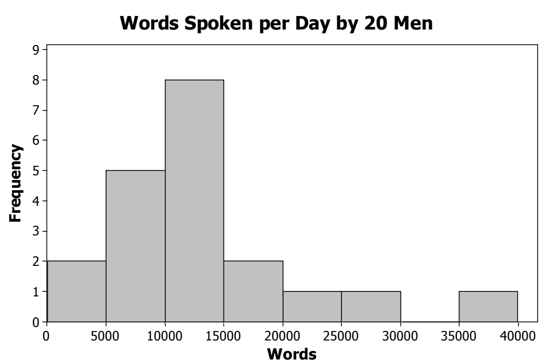
−1	3
−0	6
−0	
0	12
0	5789
1	012
1	
2	
2	
3	1

(c) Let μ be the mean difference (control minus experimental) in healing rates. We test $H_0: \mu = 0$ versus $H_a: \mu > 0$. The alternative hypothesis says that the control limb heals faster; that is, the healing rate is greater for the control limb than for the experimental limb. With all 12 differences: $\bar{x}$ = 6.417 and s = 10.7065, so $t = \frac{6.417 - 0}{10.7065/\sqrt{12}} = 2.08$. With df = 11, P = 0.0311 (using software). Omitting the outlier: $\bar{x}$ = 4.182 and s = 7.7565, so $t = \frac{4.182 - 0}{7.7565/\sqrt{11}} = 1.79$. With df = 10, P = 0.052. Hence, with all 12 differences, there is greater evidence that the mean healing time is greater for the control limb. When we omit the outlier, the evidence is weaker. **(d)** The provided modified boxplot shows the anticipated two outliers. Omitting both outliers, $\bar{x}$ = 5.9 and s = 5.5468, so $t = \frac{5.9 - 0}{5.5468/\sqrt{10}} = 3.36$. With df = 9, P = 0.0042. Hence, by removing both outliers, there is very strong evidence that the mean healing time is greater for the control limb—much stronger evidence than was found in part (c) when using the data from all 12 newts.



20.34 (a) Without the outlier, the mean is $\bar{x}$ = 24.76 degrees, and the standard deviation decreases to s = 6.34 degrees. Using df = 36 with software, t^*= 2.0281, so a 95% confidence interval for the population mean becomes 24.76 $\pm$ 2.0281(6.34/ $\sqrt{37}$) = 22.65 to 26.87 degrees. **(b)** In Exercise 20.32, using all of the data, the 95% confidence interval was 22.96 to 27.88 degrees. The confidence interval in Exercise 20.32 is wider because the presence of an outlier increases s.

20.35 (a) A histogram of the sample is provided. The sample has a significant outlier and indicates skew. We might consider applying t procedures to the sample after removing the most extreme observation (37,786).



(b) If we remove the largest observation, the remaining sample is not heavily skewed and has no outliers. Now, we test $H_0\colon \mu = 7000$ versus $H_{\text{a}}\colon \mu \neq 7000$. With the outlier removed, $\bar{x}$ = 11,555.16 and s = 6095.015. Hence, $t = \frac{11{,}555.16 - 7000}{6095.015/\sqrt{19}} =$ 3.258. With df = 18 (with software), P = 0.0044 (this is a two-sided test). There is overwhelming evidence that the mean number of words per day of men at this university differs from 7000 (the sample mean indicates they speak more than 7000 words per day).

20.36 (a) The provided stemplot does not show any severe evidence of non-Normality (the two largest observations are not outliers by the 1.5 × IQR criterion, and they are reasonable given the other data values), so t procedures should be safe.

```
0 | 67
0 | 89
1 | 00
1 | 33
1 | 4
1 |
1 | 9
2 | 0
```

(b) With $\bar{x}$ = 1.1727 days, s = 0.4606 day, and t^* = 1.812 (df = 10), the 90% confidence interval is 1.1727 $\pm$ 1.812(0.4606/$\sqrt{11}$) = 1.1727 ± 0.2516 = 0.9211 to 1.4243 days. There is no indication that the sample represents an SRS of all patients whose melanoma has not responded to existing treatments, so inferring to the population of similar patients may not be reasonable.

20.37 (a) Each patient was measured before and after treatment. **(b)** The provided stemplot of differences shows an extreme right-skew, and one or two high outliers. The t procedures should not be used.

```
0 | 0012238
1 | 0
2 | 1
3 |
4 |
5 | 1
6 |
7 | 0
```

(c) Some students might perform the test (H_0: $\mu = 0$ versus H_a: $\mu > 0$) using t procedures, despite the presence of strong skew and outliers in the sample. If so, they should find $\bar{x}$ = 156.36, s = 234.2952, and t = 2.213, yielding P = 0.0256.

20.38 [The design described is matched pairs, and we are interested in the differences (helium filled minus air filled) in punt distance.] **(a)** The provided stemplot of the 39 differences suggests a roughly symmetric, single-peaked distribution. Note that the randomization described means that we can treat the observations as a simple random sample from a population of all differences (for this kicker). Hence, t procedures seem to be appropriate here.

```
-1 | 4322
-0 | 975
-0 | 43222221110
 0 | 012222233344
 0 | 6677889
 1 | 4
 1 | 7
```

(b) Let μ denote the mean difference (helium filled minus air filled). We are interested in whether helium increases average distance traveled, so we test H_0: $\mu =$

0 versus $H_a: \mu > 0$. We have $\bar{x}$ = 0.462 foot and s = 6.867 feet. Hence, $t = \frac{0.462 - 0}{6.867/\sqrt{39}}$ = 0.42. With df = 38, P = 0.3384 (using software). There is virtually no evidence that the mean distance for helium-filled footballs is greater than that of air-filled footballs.

20.39 (a) We test $H_0: \mu = 0$ versus $H_a: \mu > 0$, where μ is the mean difference (treated minus control). This is a one-sided test because the researchers have reason to believe that CO_2 will increase growth rate. **(b)** We have $\bar{x}$ = 1.916 and s = 1.05, so $t = \frac{1.916 - 0}{1.05/\sqrt{3}}$ = 3.16 with df = 2. P = 0.0436. This is significant at the 5% significance level. **(c)** For very small samples, t procedures should only be used when we can assume that the population is Normal. We have no way to assess the Normality of the population based on these three observations. Thus, the validity of the analysis in part (b) is dubious.

20.40 (a) Weather conditions that change day to day can affect spore counts. So, the two measurements made on the same day form a matched pair. **(b)** Take the differences (kill room counts minus processing counts). For these differences, $\bar{x}$ = 1824.5 and s = 834.1 CFU/m^3. For the population mean difference, the 90% confidence interval for μ is 1824.5 $\pm$ 2.353(834.1/$\sqrt{4}$) = 1824.5 ± 981.3 = 843.2 to 2805.8 CFU/m^3. The interval is so wide because the sample size is very small, but we are confident that the mean counts in the kill room are higher. (c) The data are counts, which are, at best, only approximately Normal, and we have only a small sample.

20.41 The provided stemplot (not asked for) reveals that these data contain two extreme high outliers (5973 and 8015). Hence, t procedures are not appropriate.

```
0 | 1123788
1 | 00115677899
2 | 01112458
3 |
4 |
5 | 9
6 |
7 |
8 | 0
```

20.42 (a) We test $H_0: \mu = 0$ versus $H_a: \mu > 0$, where μ is the mean loss in sweetness (sweetness before storage minus sweetness after storage). This is a one-sided test because the researchers have reason to believe that storage reduces sweetness, and μ > 0 represents this change. We have $\bar{x}$ = 0.3 and s = 1.191, so $t = \frac{0.3 - 0}{1.191/\sqrt{10}}$ = 0.797 with df = 9. Hence, P = 0.2231. There is not evidence that storage reduces sweetness for this cola. **(b)** We have only 10 observations; it isn't possible to accurately assess Normality of the distribution of score differences. In fact, the provided stemplot of these data reveals possible skew in this distribution.

```
-2 | 0
-1 |
-1 | 1
-0 |
-0 | 31
 0 | 4
 0 | 5
 1 | 23
 1 | 56
```

20.43 (a) The mean and standard deviation are $\bar{x}$ = 48.25, and s = 40.24 thousand barrels. From Table C, t^* = 2.000 (df = 60). Using software, with df = 63, t^*= 1.998. The 95% confidence interval for μ is 48.25 $\pm$ 2.000(40.24/√64) = 48.25 ± 10.06 = 38.19 to 58.31 thousand barrels. (Using software, the confidence interval is almost identical: 38.2 to 58.3 thousand barrels.) **(b)** The provided stemplot confirms the skewness and outliers described in the exercise. The two intervals have similar widths, but the new interval (using a computer-intensive method) is shifted higher by about 2000 barrels. Although t procedures are fairly robust, we should be cautious about trusting the result in part (a) because of the strong skew and outliers. The computer-intensive method may produce a more reliable interval.

```
0 | 0000111111111
0 | 2222222333333333333
0 | 4444444555555
0 | 6666667
0 | 8899
1 | 01
1 |
1 | 5
1 |
1 | 9
2 | 0
```

20.44 (a) Let μ represent the mean *E. coli* counts for all possible 100-mL samples taken from all swimming areas in Erie County. We test $H_0: \mu = 400$ versus $H_a: \mu < 400$ because the researchers are interested in whether the average *E. coli* levels in these areas are safe. For our sample, $\bar{x}$ = 316.538 and s = 420.906 bacteria. We have $t = \frac{316.538 - 400}{420.906/\sqrt{24}} = -0.971$ with df = 23, so P = 0.1707. There is not good evidence to conclude that swimming areas in Erie County have mean *E. coli* counts less than 400 bacteria per 100 mL (that is, that the *E. coli* levels were safe). **(b)** A stemplot is provided. Note that stems are in units of 1000, and data were rounded to the nearest 100. For example, "0 | 5" represents 500, which corresponds to the original sample value of 517.2, while "2 | 0" represents 2000, which corresponds to the original sample value of 1986.3. Due to extreme skew and the presence of outliers, t procedures should not be used here. The two tests provide starkly different results; while the one-sample t test does not provide evidence that the swimming areas in Erie County are safe on average, the method that does not assume a specific shape for the distribution provides very strong evidence that these swimming areas are safe on average.

```
0 | 0000111112222222233344
0 | 56
1 | 0
1 |
2 | 0
```

20.45 STATE: Can velvetleaf seed beetles be helpful in controlling the velvetleaf plant infestations? PLAN: We will construct a 90% confidence interval for μ, the mean percent of beetle-infected seeds. SOLVE: The provided stemplot shows a single-peaked and roughly symmetric distribution. We assume that the 28 plants can be viewed as an SRS of the population, so t procedures are appropriate. We have $\bar{x}$ = 4.0786 and s = 2.0135%. Using df = 27, the 90% confidence interval for μ is $4.0786 \pm 1.703(2.0135/\sqrt{28})$ = 4.0786 ± 0.648 = 3.43% to 4.73%. CONCLUDE: The beetle infects less than 5% of seeds, so it is unlikely to be effective in controlling velvetleaf.

```
0 | 07
1 | 9
2 | 24689
3 | 666778
4 | 0000336
5 | 157
6 |
7 | 00
8 | 57
```

20.46 STATE: Will blinatumomab help cancer patients recruit T cells? PLAN: We will test $H_0\colon \mu = 0$ versus $H_a\colon \mu > 0$, where μ represents the mean increase in T cell counts after 20 days on blinatumomab. SOLVE: The provided stemplot suggests that t procedures are reasonable, with no gross evidence of non-Normality (with n = 6 observations, this is difficult to assess) and no outliers. We have $\bar{x}$ = 0.5283 and s = 0.4574 thousand cells; df = 5, $t = \frac{0.5283 - 0}{0.4574/\sqrt{6}}$ = 2.829, and P = 0.0184. We would reject H_0 at the 5% significance level. CONCLUDE: The data give convincing evidence that the mean count of T cells is higher after 20 days on blinatumomab.

```
0 | 1224
0 | 8
1 | 3
```

20.47 From Exercise 20.46, we have $\bar{x}$ = 0.5283, s = 0.4574, and df = 5. A 95% confidence interval for the mean difference in T cell counts after 20 days on blinatumomab is $0.5283 \pm 2.571(0.4574/\sqrt{6})$ = 0.5283 ± 0.4801 = 0.0482 to 1.0084 thousand cells.

20.48 **(a)** Fund and index performances are certainly not independent within a particular year; for example, a good year for one is likely to be a good year for the other. **(b)** PLAN: Let μ be the mean difference (fund minus EAFE). We test $H_0\colon \mu = 0$

versus $H_a: \mu \neq 0$, taking a two-sided alternative, because the VIG Fund could outperform or underperform the benchmark. SOLVE: The provided stemplot shows no serious deviations from Normality. We must assume that the data we have can be viewed as an SRS. We find $\bar{x}$ = 0.8784% and s = 7.4045%, so $t = \frac{0.8784 - 0}{7.4045/\sqrt{32}} = 0.6711$, for which P = 0.5071. CONCLUDE: We do not have evidence that this fund's performance differs from its benchmark.

```
-1 | 7
-1 | 32
-0 | 877
-0 | 432221100
 0 | 112334
 0 | 5566889
 1 | 0124
```

20.49 (a) For each subject, randomly select which knob (right or left) that subject should use first. **(b)** STATE: Do right-handed people find right-handed threads easier to use? PLAN: We test $H_0: \mu = 0$ versus $H_a: \mu < 0$, where μ denotes the mean difference in time (right-thread time minus left-thread time), so that μ < 0 means "right-hand time is less than left-hand time on average." SOLVE: The provided stemplot of the differences gives no reason that t procedures are not appropriate. We assume our sample can be viewed as an SRS. We have $\bar{x}$ = −13.32 seconds and s = 22.936 seconds, so $t = \frac{-13.32 - 0}{22.936/\sqrt{25}}$ = −2.9. With df = 24 we find P = 0.0039. CONCLUDE: We have good evidence (significant at the 1% level) that the mean difference really is negative—that is, the mean time for right-hand-thread knobs is less than the mean time for left-hand-thread knobs.

Stem	Leaf	Count
3	8	1
2	03	2
1	1	1
0	2	1
-0	743310	6
-1	66621	5
-2	94	2
-3	511	3
-4	853	3
-5	2	1
-6		

-5|2 represents -52

20.50 STATE: Does a generic differ significantly from the brand-name (reference) drug it is supposed to duplicate in terms of absorption in the blood? PLAN: We test $H_0: \mu = 0$ versus $H_a: \mu \neq 0$, where μ denotes the mean difference in absorption (generic minus reference). The alternative is two-sided because we have no prior expectation of a direction for the difference. SOLVE: We assume that the subjects can be considered an SRS. The provided stemplot of the differences looks reasonably Normal with no outliers, so the t procedures should be safe. We find $\bar{x}$ =

37 and s = 1070.6, so $t = \frac{37-0}{1070.6/\sqrt{20}}$ = 0.15. With df = 19, we see that P > 0.5 (software gives P = 0.8788). CONCLUDE: There is not enough evidence to conclude that the two drugs differ in mean absorption level.

```
-2 | 3
-1 | 5
-1 | 31
-0 | 5
-0 | 321
 0 | 012334
 0 | 577
 1 | 1
 1 | 6
 2 | 0
```

20.51 Refer to the solution in Exercise 20.49. With df = 24, t^* = 1.711, so the confidence interval for μ is given by $-13.32 \pm 1.711(22.936/\sqrt{25})$ = −13.32 ± 7.85 = -21.2 to -5.5 seconds. $\bar{x}_{\text{RH}}/\bar{x}_{\text{LH}} = 104.12/117.44$ = 0.887. Right-handers working with right-handed knobs can accomplish the task in about 89% of the time needed by those working with left-handed knobs.

20.52 STATE: Does receiving a bad "weather forecast" from a restaurant server influence the tip? PLAN: Let μ denote the average tip percent for all patrons receiving a bad weather forecast. We test $H_0: \mu = 20\%$ versus $H_a: \mu < 20\%$. SOLVE: We assume we may consider the sample to be an SRS taken from the population of all patrons receiving such a weather report. A stemplot of these data reveal no reason to suspect that t procedures are not appropriate. There are no outliers, and the data are roughly symmetric with one peak. We find $\bar{x}$ = 18.19% and s = 2.105%, so $t = \frac{18.19-20}{2.105/\sqrt{20}}$ = −3.845. With df = 19, P = 0.0005 (using software). CONCLUDE: There is overwhelming evidence that the mean tip percent for patrons receiving a bad weather report is less than 20%.
Note: *This does not imply that patrons who receive bad weather forecasts tip less than those who do not receive bad weather forecasts, because we do not know for a fact that the mean percent tip in the control group is 20%. No comparison was made here. A comparison will be examined in later chapters.*

20.53 (a) Starting with Table C values, for 90% confidence, $t(100)$ = 1.660 and z^* = 1.645. For 95% confidence, $t(100)$ = 1.984 and z^* = 1.96. Similarly, for 99% confidence, $t(100)$ = 2.626 and z^* = 2.576. The differences are 0.015, 0.024, and 0.05. Larger confidence levels will need more observations. We note that $t(1000)$ is within 0.01 of z^* for all these confidence levels. Using software, we find that $t(150)$ = 1.655 for 90% confidence, $t(240)$ = 1.9699 for 95% confidence, and $t(485)$ = 2.586 for 99% confidence. **(b)** Answers will vary. We'll note that the effect of the standard deviation difference multiplies the margin of error in the calculation by 100, which implies that "similar" takes more observations with $\sigma = 100$ than for $\sigma = 1$. Using n = 485 with $\sigma = 100$, the 99% t margin of error is 11.74, compared with a 99% z

margin of error of 11.70. Using $\sigma = 1$, the margins of error are both 0.117, rounding to three decimal places.

20.54 and **20.55** are Web-based exercises.