Chapter 12

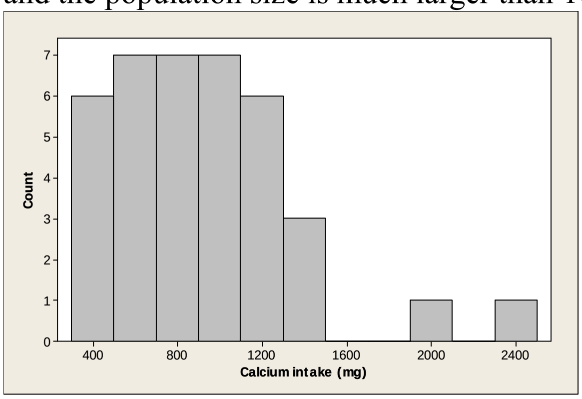
12.1 (a) 2.015. (b) 2.518.

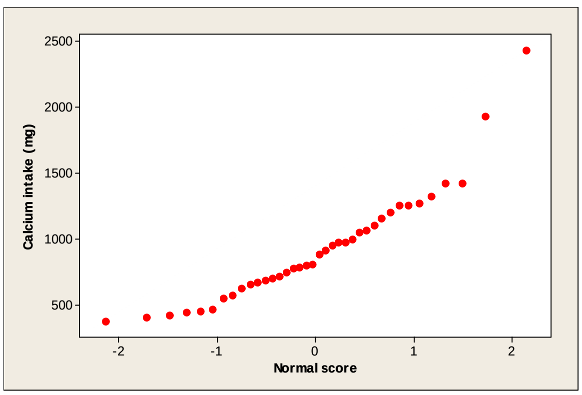
12.2 (a) 2.145. (b) 0.688.

12.3 (a) 14. (b) 1.82 is between 1.761 $(p = 0.05)$ and 2.145 $(p = 0.025)$. (c) The *P*-value is between 0.025 and 0.05. (In fact, the *P*-value is 0.0451.) (d) $t = 1.82$ is significant at $\alpha = 0.05$ but not at $\alpha = 0.01$.

12.4 (a) 24. (b) 1.12 is between 1.059 $(p = 0.15)$ and 1.318 $(p = 0.10)$. (c) The *P*-value is between 0.30 and 0.20. (In fact, the *P*-value is 0.2738.) (d) No, $t = 1.12$ is not significant at either $\alpha = 0.10$ or at $\alpha = 0.05$.

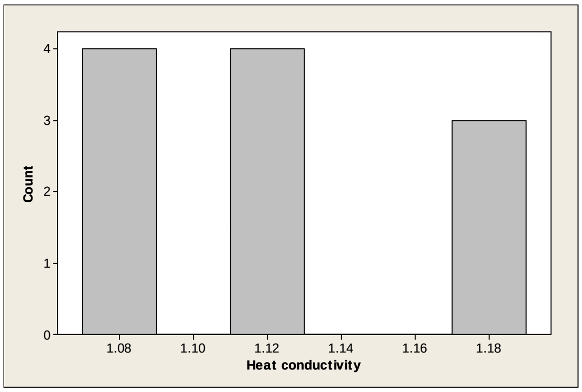
12.5 (a) $H_0 : \mu = 1200$ mg versus $H_a : \mu < 1200$ mg, where μ = the mean daily calcium intake for women between the ages of 18 and 24 years. We know that these women participated in the study, but we do not know if they were randomly selected from some larger population. These women are most likely volunteers, so we must be willing to treat them as an SRS from a larger population of women in this age group. The histogram (on the left) and the Normal probability plot (on the right) below show that the distribution of calcium intake is skewed to the right, with two outliers. There is a clear nonlinear pattern in the Normal probability plot which should create some concern. Use of the *t*-procedure is justified because the sample size is reasonably large (*n* = 38) and thus the distribution of $\bar{x}$ will be approximately Normal by the central limit theorem. The independence condition is satisfied because we are sampling without replacement and the population size is much larger than $10 \times 38 = 380$.

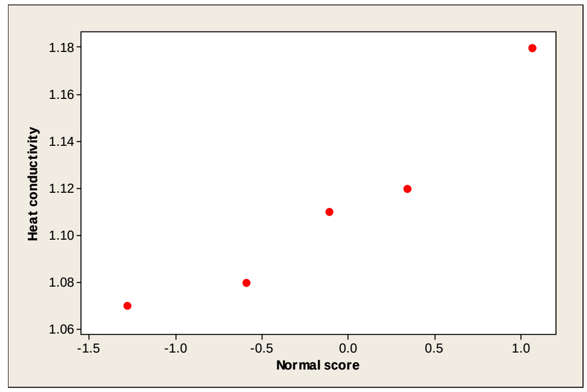




The test statistic is $t = \dfrac{926.026 - 1200}{427.230/\sqrt{38}} \doteq -3.95$, with df = 37, and *P*-value = 0.00017. Because the *P*-value is less than α =0.05, we reject H_0 and conclude that the mean daily intake is significantly less than the RDA recommendation. (b) Without the two high outliers (1933 and 2433), $t = -6.73$ and the *P*-value $\doteq$ 0. (Minitab shows one high outlier (2433)—without the outlier, $t = -5.46$ and the *P*-value $\doteq$ 0.) Our conclusion does not change.

12.6 We want to test $H_0: \mu = 1$ versus $H_a: \mu > 1$, where μ = the mean heat conductivity measured in watts of heat power transmitted per square meter of surface per degree Celsius of temperature difference on the two sides of this particular type of glass. We must be willing to treat these 11 measurements as an SRS from a larger population of this type of glass. The histogram (below on the left) and Normal probability plot (below on the right) show no serious departures from Normality or outliers so the Normal condition appears to be satisfied. The independence condition is also satisfied since we are sampling without replacement and the number of windows (or other products) made from this type of glass is clearly larger than $10 \times 11 = 110$.



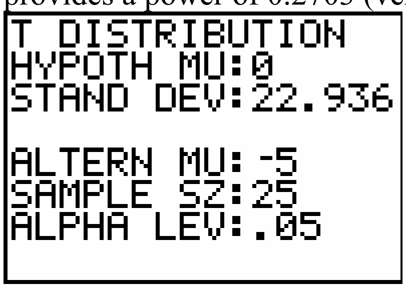


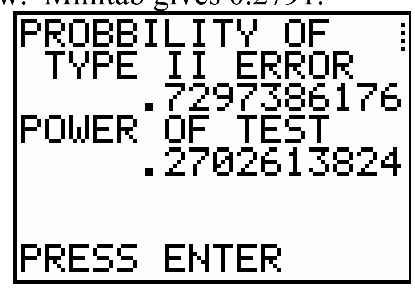
The test statistic is $t = \frac{1.1182 - 1}{0.0438/\sqrt{11}} \doteq 8.95$, with df = 10, and P-value < 0.0001. Because the P-value is less than any reasonable significance level, we reject H_0 and conclude that the mean heat conductivity for this type of glass is greater than 1. A 95% confidence interval for μ is $1.1182 \pm 2.228\left(0.0438/\sqrt{11}\right)$ = (1.089, 1.148). We are 95% confident that the mean conductivity for this type of glass is between 1.089 and 1.148 units.

12.7 We want to test $H_0: \mu = 16$ versus $H_a: \mu \neq 16$. The two-sided alternative is used because we want to see if the mean weight gain is different than what is expected. The test statistic is $t = \frac{10.4088 - 16}{3.8406/\sqrt{16}} = -5.82$ with df = 15 and a P-value < 0.0001. Since the P-value is less than any reasonable significance level, we reject H_0 and conclude that the mean weight gain is significantly different than expected. This is the same conclusion we made in Exercise 10.68.

12.8 We want to test $H_0: \mu = 0$ versus $H_a: \mu \neq 0$. The test statistic is $t = \frac{328 - 0}{256/\sqrt{16}} = 5.125$ with df = 15 and a P-value $\doteq 0.0012$. Since 0.0012 is less than $\alpha = 0.01$, we reject H_0 and conclude that there is a significant 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.

12.9 (a) Randomly assign 12 (or 13) into a group that will use the right-hand knob first; the rest should use the left-hand knob first. Alternatively, for each student, randomly select which knob he or she should use first. (b) Let μ_{R-L} denote the mean of difference (right thread − left thread) in times. We want to test $H_0: \mu_{R-L} = 0$ (no difference) versus $H_a: \mu_{R-L} < 0$. The test statistic is $t = \frac{-13.32 - 0}{22.936/\sqrt{25}} \doteq -2.9037$, with df = 24 and a *P*-value = 0.0039. Since 0.0039 is less than the significance level $\alpha = 0.05$, we reject H_0 and conclude that there is statistically significant evidence to support the hypothesis that right-handed people find right-hand threads easier to use. (Note: A student may take the difference in the other direction (left − right). The parameter should then be defined as μ_{L-R} = the mean of difference (left thread − right thread) in times, the hypotheses are $H_0: \mu_{L-R} = 0$ versus $H_a: \mu_{L-R} > 0$, and the test statistic is $t \doteq 2.9037$. The df, *P*-value, and conclusion will not change.) (c) A Type I error is committed when the designers conclude that right-handed people find right-handed threads easier to use when in fact there no difference in the times. The consequence of this error is that the designers would create two different instruments when it is unnecessary. A Type II error is committed when the designers conclude that there is no difference in the times, when in fact there is. The consequence of this error is that designers will create one instrument when two are needed. (d) The calculator provides a power of 0.2703 (very low!). See the screen shots below. Minitab gives 0.2791.





12.10 A 90% confidence for μ_{R-L} is $-13.32 \pm 1.711\left(22.936/\sqrt{25}\right)$ = (−21.169, −5.471). We are 90% confident that the mean difference in times to complete this task is between 5.5 and 21.2 seconds faster with right-hand threads than with left-hand threads. (Note: a 90% confidence interval for μ_{L-R} is (5.471, 21.169).) (b) Yes, the time saved is practically important, especially if the task must be repeated many times. The ratio of the sample means is $\frac{104.12}{117.44} \doteq 0.8866$, so right-handed students working with right-hand threads would be able to complete their task in about 89% of the time of the time it would take them to complete the same task with left-hand threads.

12.11 Let μ_{V-M} denote the mean difference (Vanguard − Managed) in annual percent returns. We want to test $H_0: \mu_{V-M} = 0$ versus $H_a: \mu_{V-M} > 0$. The test statistic is $t = \frac{2.83}{11.65/\sqrt{24}} \doteq 1.19$ with df = 23 and 0.10 < *P*-value < 0.15. (Minitab gives a *P*-value of 0.123.) Since the *P*-value is greater than any reasonable significance level, say $\alpha = 0.10$, we cannot reject H_0. The difference is not significant; it could have arisen by chance.

12.12 (a) For each subject, randomly choose which test to administer first. Alternatively, randomly assign 11 subjects to the "ARSMA first" group, and the rest to the "BI first" group. (b) Let μ_{A-B} denote the mean difference (ARSMA − BI) in scores. We want to test $H_0: \mu_{A-B} = 0$ versus $H_a: \mu_{A-B} \neq 0$. The test statistic is $t = \dfrac{0.2519}{0.2767/\sqrt{22}} \doteq 4.27$ with df = 21 and *P*-value = 0.00034 < 0.0004. Since the *P*-value is less than any reasonable significance level, say $\alpha = 0.01$, we reject H_0 and conclude that there is a significant difference in these two measures. (c) A 95% confidence interval for μ_{A-B} is $0.2519 \pm 2.08\left(0.2767/\sqrt{22}\right)$=(0.1292, 0.3746). We are 95% confident that the mean difference in the two measures is between 0.13 and 0.37 points.

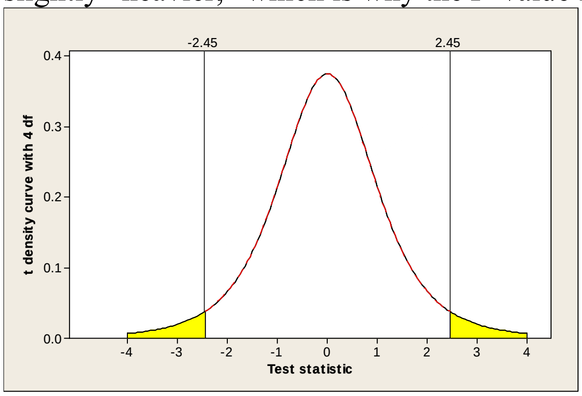
12.13 (a) Yes. Let μ_d denote the mean difference in the amount of money charged on credit cards for these two policies. We want to test $H_0: \mu_d = 0$ versus $H_a: \mu_d > 0$. The test statistic is $t = \dfrac{332}{108/\sqrt{200}} \doteq 43.47$ with df = 199 and a *P*-value of almost 0. Since the *P*-value is less than 0.01, we reject H_0 and conclude that the amount charged would significantly increase under the no-fee offer. (b) A 99% confidence interval for μ_d is $332 \pm 2.6008\left(108/\sqrt{200}\right)$ = ($312.14, $351.86). (c) The sample size is very large, and we are told that we have an SRS. This means that outliers are the only potential snag, and there are none. (d) Make the offer to an SRS of 200 customers, and choose another SRS of 200 as a control group. Compare the mean increase for the two groups.

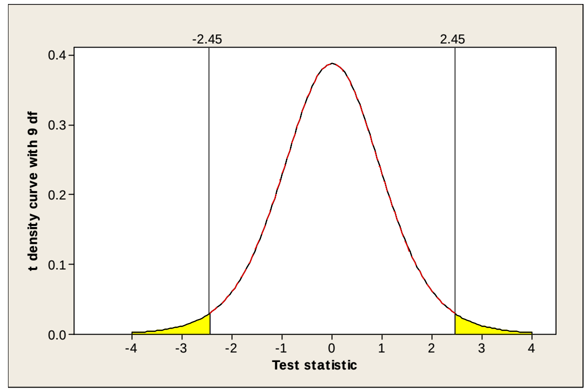
12.14 (a) A Type I error is committed when the bank concludes that the mean amount charged will increase under the no-fee offer when in fact it will not. A Type II error is committed when the bank concludes that the mean amount charged will not increase under the no-fee offer, when in fact it would. A Type II error is more serious because the bank would not collect the increased revenue that would have resulted from increased charges—that is why they want to be quite certain of detecting a mean increase of $100. (b) Minitab gives a power of 0.999973 for detecting the $100 mean difference in the amount charged. Since the power is almost 1, the bank should conduct the test using 50 customers. Additional customers are not needed to detect this difference.

12.15 (a) df = 9. (b) The *P*-value is between 0.025 and 0.05.

12.16 We want $|t| > t^*$, where t^* is the upper $\alpha/2$ critical value for the *t* distribution with 20 −1 = 19 df. From Table C, $\alpha/2 = 0.0025$ so values of the test statistic above $t^*_{19,0.0025} = 3.174$ and below $-t^*_{19,0.0025} = -3.174$ are considered statistically significant at the $\alpha = 0.005$ level.

12.17 (a) For a *t* distribution with df = 4, the *P*-value is 0.0704—not significant at the 5% level. See the sketch below (on the left). (b) For a *t* distribution with df = 9, the *P*-value is 0.0368, which is significant at the 5% level. See the sketch below (on the right). A larger sample size means that there is less variability in the sample mean, so the *t* statistic is less likely to be large when H_0 is true. Note that even with these computer produced graphs of these t distributions, it is difficult to see the subtle difference between them: The "tails" of the *t* (4) distribution are slightly "heavier," which is why the *P*-value is larger.



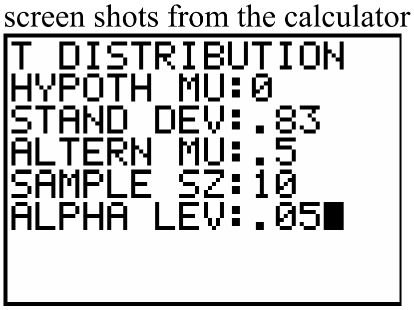


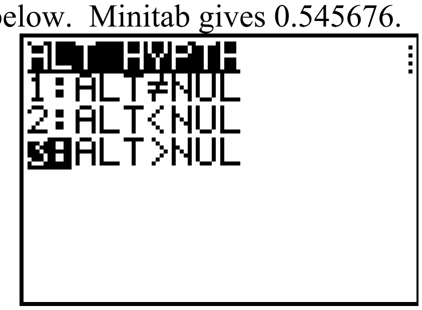
12.18 (a) The parameter μ_d is the mean difference in the yields for the two varieties of plants. (b) We want to test $H_0 : \mu_d = 0$ versus $H_a : \mu_d > 0$. It is reasonable to assume that the differences in the yields are independent and follow a Normal distribution. Nothing is mentioned about random selection, but we must also assume that these differences represent an SRS from the population of differences in the yields for these two varieties. The test statistic is

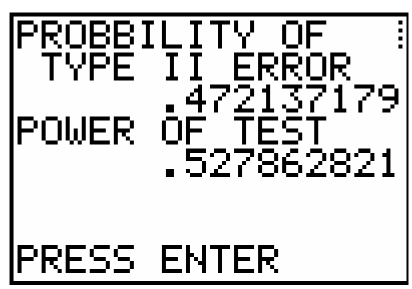
$t = \dfrac{0.34}{0.83/\sqrt{10}} \doteq 1.295$ with df = 9 and *P*-value = 0.114. Since the *P*-value is greater than 0.05,

we cannot reject H_0 and conclude that the yields for the two varieties are not significantly different. The observed difference appears to be due to chance variation.

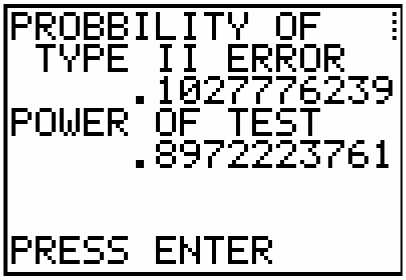
12.19 (a) A Type I error is committed when the experts conclude that there is a mean difference in the yields when in fact there is none. A Type II error is committed when the experts conclude that there is no mean difference in yields when in fact one does exist. A Type II error is more serious because the experts would like to increase the yield (and hence make more money) whenever possible. (b) The power is 0.5278. (Reject H_0 if $t > 1.833$ i.e., if $\bar{x} > 0.4811$.) See the screen shots from the calculator below. Minitab gives 0.545676.







(c) The power is 0.8972. (Reject H_0 if $t > 1.711$, i.e., if $\bar{x} > 0.2840$.) See the screen shot below. Minitab gives 0.899833.



(d) Any two of the following: Increasing the significance level, decreasing the standard deviation σ, or moving the particular alternative value for μ farther away from 0.

12.20 Let μ = the mean percent of purchases for which an alternative supplier offered a lower price than the original supplier. The conditions for inference are satisfied. The invoices were randomly selected so it is reasonable to view these differences as an SRS. The differences are independent from one invoice to another. The graphical displays suggest that the differences are skewed to the left, but there are no outliers. With n = 25, we can appeal to the robustness of the t procedures, since the distribution of the differences is not Normal. The summary statistics provided indicate that the mean is 77.76%, the standard deviation is 32.6768%, and the standard error is about 6.5354%. Using Table C with df = 24, the critical value is $t^* = 2.064$, so the 95% confidence interval for μ is $77.76\% \pm 2.064 \times 6.5353603\% = (64.27\%, 91.25\%)$. The data support the retailer's claim: 64% to 91% of the time the original supplier's price was higher.

12.21 (a) Let $\mu_{\text{H}-\text{F}}$ = the mean difference in vitamin C content (Haiti −Factory) at the two locations. We want to test $H_0: \mu_{H-F} = 0$ versus $H_a: \mu_{H-F} < 0$. The test statistic is $t = \dfrac{-5.3333}{5.5885/\sqrt{27}} \doteq -4.96$, with df = 26 and a P-value < 0.0005. (b) A 95% confidence interval for $\mu_{\text{H}-\text{F}}$ is $-5.3333 \pm 2.056\left(5.5885/\sqrt{27}\right) = (-7.54, -3.12)$. With 95% confidence, we estimate the mean loss in vitamin C content over the 50 month period to be between 3.12 and 7.54 mg/100g. (c) Yes. Let μ_F denote the mean vitamin C content of the specially marked bags of WSB at the factory. We want to test $H_0: \mu_F = 40$ versus $H_a: \mu_F \neq 40$. The test statistic is $t = \dfrac{42.852 - 40}{4.793/\sqrt{27}} \doteq 3.09$, with df=26 and 0.002 < P-value < 0.005. Since the P-value is below 0.01, we have strong evidence that the mean vitamin C content differs from the target value of 40 mg/100g for specially marked bags at the factory. (Note: Some students may identify the parameter of interest as the vitamin C content of the bags when they arrive in Haiti. The correct solution for these students is: Let μ_H denote the mean vitamin C content of the bags of WSB shipped to Haiti. We want to test $H_0: \mu_H = 40$ versus $H_a: \mu_H \neq 40$. The test statistic is $t = \dfrac{37.5185 - 40}{2.4396/\sqrt{27}} \doteq -5.29$, with df=26 and P-value < 0.0001. Since the P-value is below 0.01, we have strong evidence that the mean vitamin C content differs from the target value of 40 mg/100g.)

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

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

For $\alpha = 0.05$, the first two columns of the table below show the power for a variety of sample sizes, and we see that $n \geq 156$ achieves the desired 80% power. The power for a variety of sample sizes when $\alpha = 0.01$ is shown in the last two columns of the table below, and we see that $n \geq 254$ achieves the desired 80% power. As expected, more observations are needed with the smaller significance level. Since the significance level was not provided, students may use other values, but these will be the two most common responses.

Sample size	Power when $\alpha = 0.05$	Sample size	Power when $\alpha = 0.01$
25	0.250485	25	0.083564
50	0.401222	50	0.170827
75	0.528434	75	0.265700
100	0.633618	100	0.361801
125	0.718701	125	0.454352
150	0.786254	150	0.540181
151	0.788631	175	0.617459
152	0.790984	200	0.685391
153	0.793314	225	0.743930
154	0.795621	250	0.793528
155	0.797906	251	0.795336
156	0.800167	252	0.797131
157	0.802406	253	0.798913
158	0.804623	254	0.800682
159	0.806818	255	0.802438

12.23 (a) No, the expected number of successes np_0 and the expected number of failures $n(1-p_0)$ are both less than 10 (they both equal 5). (b) No, the expected number of failures is less than 10; $n(1-p_0) = 2$. (c) Yes, we have a SRS, the population is more than 10 times as large as the sample, and $np_0 = n(1-p_0) = 10$.

12.24 (a) We want to test $H_0: p = 0.73$ versus $H_a: p \neq 0.73$. The conditions for inference are met since this is an SRS and $np_0 = 200 \times 0.73 = 146$ and $n(1-p_0) = 200 \times 0.27 = 54$ are both at least 10. It is also reasonable to assume that the student body at this university is larger than 10×200 = 2000. The sample proportion is $\hat{p} = \frac{132}{200} = 0.66$ and the test statistic is

$z = \frac{0.66 - 0.73}{\sqrt{\frac{0.73 \times 0.27}{200}}} \doteq -2.23$, with a *P*-value of 0.026. Since 0.026 < 0.05, we reject H_0 and conclude that we do have statistically significant evidence that the proportion of all first-year students at this university who think being very well-off is important differs from the national value. (b) A 95% confidence interval for *p* is $0.66 \pm 1.96\sqrt{\frac{0.66 \times 0.34}{200}}$ = (0.5943, 0.7257). The confidence interval gives us information about the plausible values of *p*. We are 95% confident that the proportion of students at this university who would like to be well-off is between 59.4% and 72.6%.

12.25 (a) Yes. Let *p* = Shaq's free-throw percentage during the season following his off-season training. We certainly do not have an SRS of all free-throws by Shaq, but we will proceed to see if the observed difference could be due to chance. The other two conditions (expected number of successes and failures are at least 10 and large population) are both satisfied. We want to test $H_0: p = 0.533$ versus $H_a: p > 0.533$. The test statistic is $z = \frac{0.6667 - 0.533}{\sqrt{\frac{0.533 \times 0.467}{39}}} \doteq 1.67$ and the *P*-value = 0.0475. Notice that the *P*-value is just under 0.05, so we would say that this increase would not be explained by chance. Although we found a statistically significant increase in Shaq's free-throw shooting percentage for the first two games, we would not suggest making an inference about *p* based on these two games. (b) A Type I error would be committed by concluding that Shaq has improved his free-throwing when in fact he has not. A Type II error would be committed by concluding that Shaq has not improved his free-throwing when in fact he has. (c) The power is 0.2058. (d) The probability of a Type I error is $\alpha = 0.05$. The probability of a Type II error is 1 − 0.2058 = 0.7942.

12.26 (a) We want to test $H_0: p = 0.1$ versus $H_a: p < 0.1$. The conditions for inference are met. We must assume these patients are an SRS of all patients who would take this pain reliever. Both $np_0 = 440 \times 0.1 = 44$ and $n(1 - p_0) = 440 \times 0.9 = 396$ are at least 10. It is also reasonable to assume that the number of patients who would take this pain reliever is larger than 10×440 = 4400. The sample proportion is $\hat{p} = \frac{23}{440} \doteq 0.0523$ and the test statistic is $z = \frac{0.0523 - 0.1}{\sqrt{\frac{0.1 \times 0.9}{440}}} \doteq -3.34$, with a *P*-value = 0.0004. (b) A Type I error would be committed if the researchers conclude that the proportion of "adverse symptoms" is less than 0.1, when in fact it is not. A Type II error would be committed if the researchers conclude that the proportion of "adverse symptoms" is equal to 0.1, when in fact it is less than 0.1. A Type I error is more serious because the researchers do not want to mislead consumers.

12.27 (a) A Type I error would be committed by deciding that the proportion differs from the national proportion when in fact it doesn't. This may lead to the restaurant manager investigating the reason for the difference, which could waste time and money. A Type II error would be committed by deciding that the proportion is the same as the national proportion when in fact, it isn't. This may lead the manager to conclude that no action is needed, which may result in disgruntled employees. (b) Power = 0.0368. (c) When n = 200, power = 0.1019. Doubling the sample size increases the power by about 176.9%. (d) When $\alpha = 0.01$, power = 0.062. When $\alpha = 0.10$, power = 0.299.

12.28 Results will vary. (a) Suppose one student obtained 17 heads. The sample proportion is $\hat{p} = \frac{17}{20} = 0.85$ and the test statistic is $z = \frac{0.85 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{20}}} \doteq 3.13$, with a P-value = 0.0018. This student would conclude that the proportion of heads obtained from tipping U.S. pennies is significantly different from 0.5. (b) Suppose a class of 20 obtained 340 heads in 400 tips. The sample proportion is $\hat{p} = \frac{340}{400} = 0.85$ and the test statistic is $z = \frac{0.85 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{400}}} \doteq 14.00$, with a P-value very close to 0. At any reasonable significance level, the class would conclude that the proportion of heads obtained from tipping U.S. pennies is significantly different from 0.5.

12.29 We want to test $H_0 : p = 1/3$ versus $H_a : p > 1/3$. The test statistic is $z = \frac{304/803 - 1/3}{\sqrt{\frac{1/3 \times 2/3}{803}}} \doteq 2.72$ with a P-value = 0.0033. Yes, because 0.0033 is less than 0.01, this is strong evidence that more than one-third of this population never use condoms.

12.30 The table below shows that Tonya, Frank, and Sarah all recorded the same sample proportion, $\hat{p} = 0.28$, but the P-values were all quite different. Our conclusion is that the same value of the sample proportion provides different information about the strength of the evidence against the null hypothesis because the sample sizes are different. As the sample size increases, the P-value deceases, so the observed difference (or something more extreme) is less likely to be due to chance.

X	n	$\hat{p}$	z	P-value
14	50	0.28	−0.80	0.212
98	350	0.28	−2.12	0.017
140	500	0.28	−2.53	0.006

CASE CLOSED!

(1) Let μ = mean body temperature in the population of healthy 18 to 40 year olds. We want to test $H_0 : \mu = 98.6$ versus $H_a : \mu \neq 98.6$. The test statistic is $t = \frac{98.25 - 98.6}{0.73/\sqrt{700}} \doteq -12.69$, with df = 699 and a P-value very close to 0. Since the P-value is less than any reasonable significance level, say $\alpha = 0.01$, we have very strong evidence that the mean body temperature is different

from 98.6. (2) A 95% confidence interval for μ is $98.25 \pm 1.96336\left(\frac{0.73}{\sqrt{700}}\right)$ = (98.1958, 98.3042). We are 95% confident that the mean body temperature is between 98.20°F and 98.30°F. The confidence interval provides an estimate for plausible values of "normal" body temperature. (3) Now, we want to test $H_0: p = 0.5$ versus $H_a: p \neq 0.5$. The test statistic is $z = \frac{0.623 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{700}}} \doteq 6.51$ with a *P*-value very close to 0. We have statistically significant evidence that the proportion of all healthy adults in this age group with a temperature less than 98.6 is not equal to 0.5. (4) A 95% confidence interval for *p* is $0.623 \pm 1.96\sqrt{\frac{0.623 \times 0.377}{700}} \doteq$ (0.59, 0.66). We are 95% confident that the proportion of all healthy adults in this age group with a body temperature below 98.6 is between 0.59 and 0.66. (5) Repeated measurements were taken on 140 healthy adults, so these 700 temperature readings are clearly not independent. There is also no indication that these individuals were randomly selected from a larger group, so without additional information it is risky to assume they represent an SRS from some larger population. The population is much larger than 10×700=7000, so this should not be a concern. Finally, the distribution of $\bar{x}$ will be approximately Normal, even if the distribution of temperatures is slightly skewed because the sample size is reasonably large, and the expected number of successes (350) and failures (350) are both at least 10.

12.31 (a) Standard error should be replaced by margin of error. The margin of error equals the critical value z^* times the standard error. For 95% confidence, the critical value is $z^* = 1.96$. (b) H_0 should refer to *p* (the population proportion), not $\hat{p}$ (the sample proportion). (c) The Normal distribution (and a *z* test statistic) should be used for significance tests involving proportions.

12.32 Let *p* = the proportion of adults who favor an increase in the use of nuclear power as a major source of energy. We want to test $H_0: p = 0.5$ versus $H_a: p < 0.5$. The expected number of successes ($np_0 = 512 \times 0.5 = 256$) and the expected number of failures (also 256) are both at least 10, so use of the *z* test is appropriate for the SRS of adults. The sample proportion is $\hat{p} = \frac{225}{512} \doteq 0.4395$ and the test statistic is $z = \frac{0.4395 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{512}}} \doteq -2.74$ with a *P*-value = 0.0031. Yes, because 0.0031 is less than 0.01, this is strong evidence that less than one-half of all adults favor an increase in the use of nuclear power.

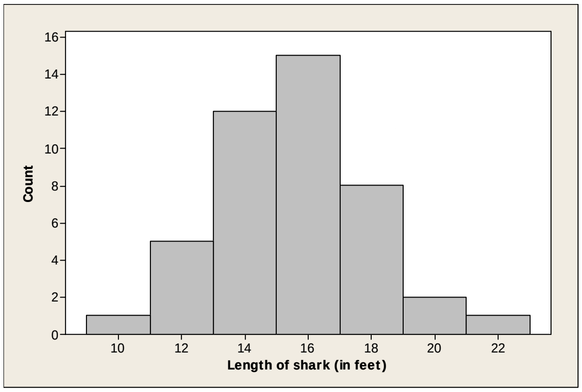
12.33 (a) There is borderline evidence. We want to test $H_0: \mu = 0\%$ versus $H_a: \mu \neq 0\%$, μ = the mean percent change (month to month) in sales. The test statistic is $t = \frac{3.8}{12/\sqrt{40}} \doteq 2.0028$, with df = 39 and *P*-value = 0.0522. (The best we can say using Table C with df = 30 is that the *P*-value is greater than 0.05.) This is not quite significant at the 5% level. Since 0.0522 is slightly larger than 0.05, we cannot reject H_0 at the $\alpha = 0.05$ significance level. However, we would reject H_0 at the $\alpha = 0.055$ significance level, since 0.0522 is less than 0.055. (b) Even if we had rejected H_0, this would only mean that the *average* change is nonzero. This does not guarantee that each *individual* store increased sales.

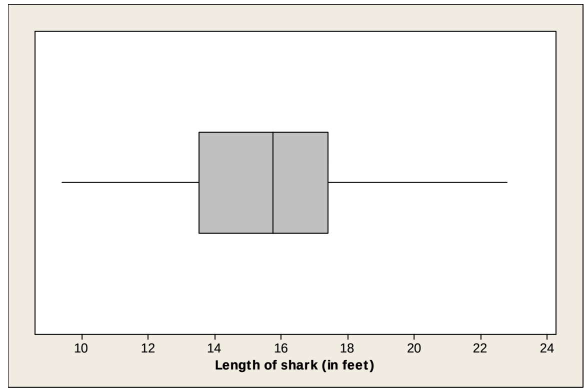
12.34 (a) (a) A subject's responses to the two treatments would not be independent. (b) We want to test $H_0: \mu_d = 0$ versus $H_a: \mu_d \neq 0$, where μ_d = the mean difference in the two chemical measurements from the brain of patients with Parkinson's disease. Since the sample size $n = 6$ is small we must assume that the differences of these measurements follow the Normal distribution. We must also assume that these 6 patients are an SRS. The independence condition is met and the population size is much larger than 60. The test statistic is $t = \frac{-0.326}{0.181/\sqrt{6}} \doteq -4.4118$, with df = 5and a P-value = 0.0069. Since 0.0069 < 0.01, we reject H_0 and conclude that there is significant evidence of a difference in the two chemical measurements from the brain.

12.35 (a) We want to test $H_0: p = 0.5$ versus $H_a: p > 0.5$. The expected number of successes ($np_0 = 50 \times 0.5 = 25$) and the expected number of failures (25) are both at least 10, so use of the z test for these subjects who must be viewed as an SRS of all coffee drinkers is appropriate. The sample proportion is $\hat{p} = \frac{31}{50} = 0.62$ and the test statistic is $z = \frac{0.62 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{50}}} \doteq 1.70$, with a P-value = 0.0446. Since 0.0446 < 0.05, we reject H_0 at the 5% level and conclude that a majority of people prefer the taste of fresh-brewed coffee. Some students may argue that the P-value is just barely below 0.05, so this result may not be practically significant. However, most students will point out that the results are significant and that this conclusion matches their personal experiences with coffee drinkers—a majority of people prefer fresh-brewed coffee. (b) A 90% confidence interval for p is $0.62 \pm 1.645\sqrt{\frac{0.62 \times 0.38}{50}}$ = (0.5071, 0.7329). We are 90% confident that between 51% and 73% of coffee drinkers prefer fresh-brewed coffee. (c) The coffee should be presented in random order. Some subjects should get the instant coffee first, and others should get the fresh-brewed coffee first.

12.36 Let μ_M = the mean masculinity score of all hotel managers. We want to test $H_0: \mu_M = 4.88$ versus $H_a: \mu_M > 4.88$. The test statistic is $t = \frac{5.91 - 4.88}{0.57/\sqrt{148}} \doteq 21.98$, with df = 147 and a P-value of 0 to many decimal places. Since the P-value is much smaller than 0.01, there is overwhelming evidence that hotel managers scored higher on the average than males in general. Turning to femininity scores, let μ_F = the mean femininity score of all hotel managers. We want to test $H_0: \mu_F = 5.19$ versus $H_a: \mu_F > 5.19$. The test statistic is $t = \frac{5.29 - 5.19}{0.75/\sqrt{148}} \doteq 1.62$, with df = 147 and a P-value of 0.053. (To use Table C, look at the df = 100 row and find that 0.05 < P-value < 0.10.) There is some evidence that hotel managers exceed males in general, but not convincing evidence (particularly because the sample size $n = 148$ is quite large).

12.37 (a) A histogram (on the left) and a boxplot (on the right) are shown below. The distribution looks reasonably symmetric with a sample mean of $\bar{x} = 15.59$ ft and a standard deviation of $s = 2.550$ ft. Notice that the two extreme values are not classified as outliers by Minitab—recall that this is because of the difference in the way the quartiles are computed with software and with the calculator.





(b) A 95% confidence interval for the mean shark length is $15.5864 \pm 2.021\left(\frac{2.5499}{\sqrt{44}}\right) \doteq$ (14.81, 16.36). (Note: Some students may use df=43 and the critical value $t^* = 2.01669$ from software of the calculator.) Yes, since 20 feet does not fall in the 95% confidence interval, we reject the claim that great white sharks average 20 feet in length at the 5% level. (c) We need to know what population these sharks were sampled from: Were these all full-grown sharks? Were they all male? (i.e., is μ the mean adult male shark length or something else?)

12.38 We want to test $H_0: p = 0.5$ versus $H_a: p \neq 0.5$, where p = the proportion of heads obtained from spinning a Belgian euro coin. The expected number of successes ($np_0 = 250 \times 0.5 = 125$) and the expected number of failures (125) are both at least 10, so use of the z test is appropriate. The sample proportion is $\hat{p} = \frac{140}{250} = 0.56$ and the test statistic is $z = \frac{0.56 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{250}}} \doteq 1.90$, with a P-value = 0.0574. Since 0.0574 > 0.05, we cannot reject H_0 at the 5% level and conclude that the observed difference could be due to chance. An interval of plausible values for p is provided by a 95% confidence interval, $0.56 \pm 1.96\sqrt{\frac{0.56 \times 0.44}{250}}$ = (0.4985, 0.6215). Notice that the 95% confidence interval includes 0.5, which would indicate that the coin is "fair" or balanced. (Note: Some students will look at the data and then conduct a one-sided test—this is not good statistical practice.)