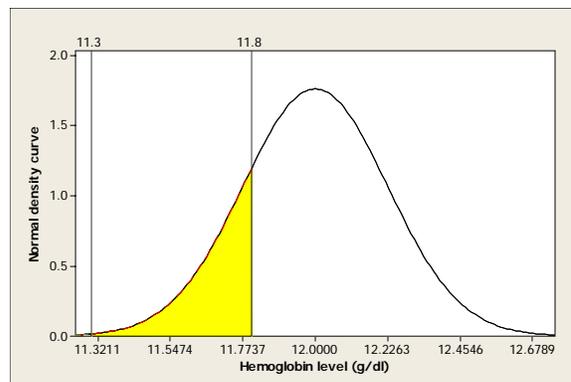
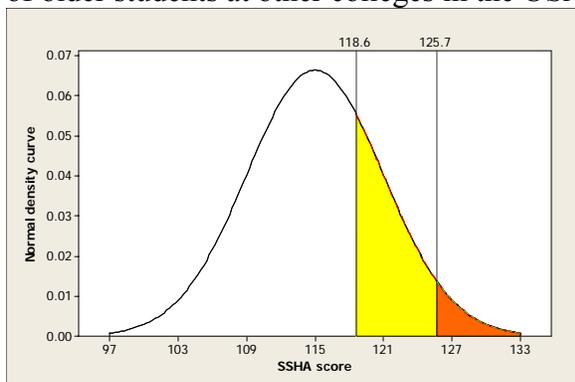


Chapter 11

11.1 (a) μ = the mean score for all older students at this college. (b) If $\mu = 115$, the sampling distribution of \bar{x} is Normal with mean 115 and standard deviation $30/\sqrt{25} = 6$ or $N(115, 6)$. See the sketch below (on the left). (c) Assuming H_0 is true, observing a mean of 118.6 or higher would not be surprising, but a mean of 125.7 or higher is less likely, and therefore provides more evidence against H_0 . (d) Yes, the sample size is not large enough ($n = 25$) to use the central limit theorem for normality. (e) No, the older students at this college may not be representative of older students at other colleges in the USA.



11.2 (a) μ = the mean hemoglobin level for all children of this age in Jordan. (b) If $\mu = 12$, the sampling distribution of \bar{x} is Normal with mean 12 g/dl and standard deviation $1.6/\sqrt{50} = 0.2263$ g/dl. See the sketch above (on the right). (c) A result like $\bar{x} = 11.3$ g/dl lies way down in the low tail of the density curve (over 3 standard deviations below the mean), while 11.8 g/dl is fairly close to the middle. If $\mu = 12$ g/dl, observing a mean of 11.8 g/dl or smaller would not be too surprising, but a mean of 11.3 g/dl or smaller is extremely unlikely, and it therefore provides strong evidence that $\mu < 12$ g/dl. (d) No, since the sample size is large ($n = 50$), the central limit theorem says that the sampling distribution of \bar{x} is approximately $N(12$ g/dl, 0.2263 g/dl). (e) No, we are told that this is sample, but we don't know if these children were randomly selected from any larger population. The only way we can generalize to a larger population is if this sample is representative of the larger population.

11.3 (a) $H_0 : \mu = 115$; $H_a : \mu > 115$. (b) $H_0 : \mu = 12$; $H_a : \mu < 12$.

11.4 (a) μ = the mean gas mileage for Larry's car on the highway.

$H_0 : \mu = 26$ mpg; $H_a : \mu > 26$ mpg. (b) p = the proportion of teens in your school who rarely or never fight with their friends. $H_0 : p = 0.72$; $H_a : p \neq 0.72$.

11.5 (a) p = the proportion of calls involving life threatening injuries where the paramedics arrived within 8 minutes. $H_0 : p = 0.78$; $H_a : p > 0.78$. (b) μ = the mean percent of local household food expenditures used for restaurant meals. $H_0 : \mu = 30$; $H_a : \mu \neq 30$.

11.6 (a) H_0 and H_a have been switched: The null hypothesis should be a statement of “no change.” (b) The null hypothesis should be a statement about μ , not \bar{x} . (c) Our hypothesis should be “some claim about the population.” Whether or not it rains tomorrow is not such a statement. Put another way, hypothesis testing—at least as described in this text—does not deal with random outcomes, but rather with statements that are either true or false. Rain (or not) is a random outcome.

11.7 (a) Because the workers were chosen without replacement, randomly sampled from the assembly workers and then randomly assigned to each group, the requirements of SRS and independence are met. The question states that the differences in job satisfaction follow a Normal distribution. (b) Yes, because the sample size ($n = 18$) is too small for the central limit theorem to apply.

11.8 (a) No, the sample size ($n = 75$) is much larger so the central limit theorem says that the sampling distribution of \bar{x} is approximately Normal. (b) For $\bar{x} = 17$ and $n = 75$, the test statistic is $z = \frac{17-0}{60/\sqrt{75}} \doteq 2.45$ and the P -value is $P(Z \leq -2.45 \text{ or } Z \geq 2.45) = 2 \times 0.0071 = 0.0142$. (c)

This is fairly strong evidence against H_0 .

11.9 See the sketch in the solution to Exercise 11.1 for parts (a) and (b). (a) The test statistic is $z = \frac{118.6-115}{30/\sqrt{25}} = 0.6$ and the P -value $= 1 - 0.7257 = 0.2743$. (b) The test statistic is $z = \frac{125.7-115}{30/\sqrt{25}} \doteq 1.78$ and the P -value $= 1 - 0.9625 = 0.0375$. (c) If $\mu = 115$, the probability of

getting a sample mean of 118.6 or something more extreme by chance is 0.2743, and the probability of getting a sample mean of 125.7 or something more extreme by chance is 0.0375—much more unlikely. A small P -value (such as 0.0375) tells us that values of \bar{x} similar to 125.7 would rarely occur when H_0 is true, while a P -value of 0.2743 indicates that results similar to 118.6 give little reason to doubt H_0 .

11.10 See the sketch in the solution to Exercise 11.2 for parts (a) and (b). (a) For $\bar{x} = 11.3$ g/dl, the test statistic is $z = \frac{11.3-12}{1.6/\sqrt{50}} \doteq -3.09$ and the P -value $= 0.0010$. (b) For $\bar{x} = 11.8$ g/dl, the test

statistic is $z = \frac{11.8-12}{1.6/\sqrt{50}} \doteq -0.88$ and the P -value $= 0.1894$. (c) The P -value of 0.0010 tells us

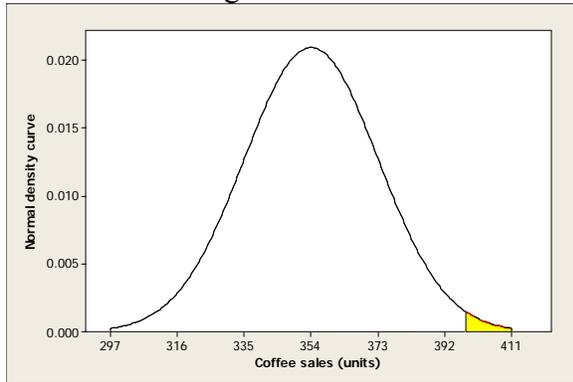
that values of \bar{x} similar to 11.3 g/dl would rarely occur when H_0 is true, while a P -value of 0.1894 indicates that results similar to 11.8 give little reason to doubt H_0 .

11.11 (a) $\bar{x} = 398$. (b) If $\mu = 354$, the sampling distribution of \bar{x} is Normal with mean 354 and standard deviation $33/\sqrt{3} \doteq 19.0526$ because weekly sales are Normal. See the sketch below. We must assume independence, and the three chosen weeks can be considered as a

representative sample of all the weeks after the price is reduced. Since three consecutive weeks have been chosen immediately after the price has been reduced, the SRS and independence

conditions are not very realistic. (c) The test statistic is $z = \frac{398 - 354}{33/\sqrt{3}} \doteq 2.31$ and the P -value = 1

– $0.9896 = 0.0104$. (d) The P -value of 0.0104 tells us that there is only about a 1.04% chance of getting values of \bar{x} at or above 398 units when H_0 is true, so this is convincing evidence that mean sales are higher.



11.12 (a) $P(Z \geq 1.6) = 1 - 0.9452 = 0.0548$. (b) $P(Z \leq 1.6) = 0.9452$. (c) $P(Z \leq -1.6 \text{ or } Z \geq 1.6) = 2 \times 0.0548 = 0.1096$.

11.13 Significance at the 1% level means that the P -value for the test is less than 0.01. If the P -value is less than 0.01, then it must also be less than 0.05. If a test is significant at the 5% level, then we know that the P -value is less than 0.05. However, we don't know how much smaller than 0.05 it is, so it may or may not be less than 0.01. In short, knowing that a test is significant at the 5% level does not tell you anything about its significance at the 1% level.

11.14 (a) The P -value is $P(Z \geq 2.42) = 1 - 0.9922 = 0.0078$. Since the P -value is less than 0.05, we say that the result is statistically significant at the 5% level. (b) Since the P -value is less than 0.01, we also say that the result is statistically significant at the 1% level. (c) For both significance levels, we would reject H_0 and conclude that the mean nicotine content is greater than 1.4 mg for this brand of cigarettes.

11.15 (a) Reject H_0 if $z > 1.645$. (b) Reject H_0 if $|z| > 1.96$. In other words, we would reject H_0 when $z \leq -1.96$ or $z \geq 1.96$. (c) For tests at a fixed significance level (α), we reject H_0 when we observe values of our statistic that are so extreme (far from the mean of the sampling distribution) that they would rarely occur when H_0 is true. For a two-sided alternative, the extreme values could be small or large (i.e., in either tail), so the significance level is divided evenly in the two tails. For a one-sided alternative, all the extreme values must be in one tail, so all of the area is in that tail.

11.16 (a) The test statistic is $z = \frac{0.4365 - 0.5}{0.2887/\sqrt{100}} \doteq -2.20$ and the P -value is

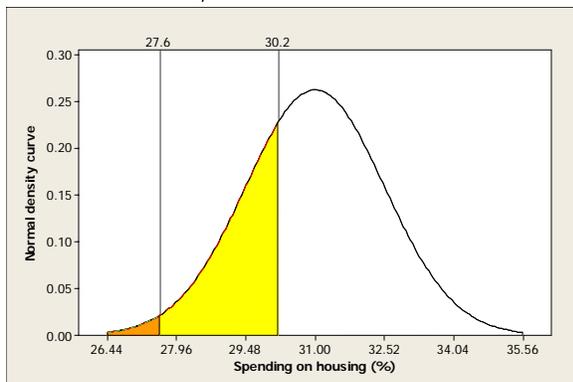
$P(Z \leq -2.20 \text{ or } Z \geq 2.20) = 2 \times 0.0139 = 0.0278$. (b) Since the P -value is less than 0.05, we say that the result is statistically significant at the 5% level. (c) Since the P -value is greater than 0.01, we say that the result is not statistically significant at the 1% level. (d) At the 5% level, we would reject H_0 and conclude that the random number generator does not produce numbers with an average of 0.5. At the 1% level, we would not reject H_0 and conclude that the observed deviation from the mean of 0.5 is something that could happen by chance. That is, we would conclude that the random number generator is working fine at the 1% level.

11.17 The command `rand(100) → L1` generates 100 random numbers in the interval (0,1) and stores them in list L_1 . The answers will vary but one simulation generated random numbers with mean $\bar{x} = 0.4851$, test statistic $z \doteq -0.52$, and P -value = 0.603. Since 0.603 is greater than 0.01 and 0.05, we do not reject H_0 at either significance level, and conclude that there is no evidence to suggest that the mean of the random numbers generated is different from 0.5.

11.18 At the 5% significance level the results of both studies would be considered statistically significant. However, the P -values convey important information about how extreme the results really are. For the first study the P -value is barely less than 0.05, so the result is barely significant. The result for the second study would be considered statistically significant at any reasonable significance level α .

11.19 (a) (1) Take a random sample of several apartments and measure the area of each. (2) $H_0 : \mu = 1250$; $H_a : \mu < 1250$. (b) (1) Take a random sample of service calls over the year and find out how long the response time was on each call. (2) $H_0 : \mu = 1.8$; $H_a : \mu \neq 1.8$. (c) (1) Take a random sample of students from his school and find the proportion of lefties. (2) $H_0 : p = 0.12$; $H_a : p \neq 0.12$.

11.20 (a) If $\mu = 31\%$, the sampling distribution of \bar{x} is Normal with mean 31% and standard deviation $9.6\%/\sqrt{40} \doteq 1.52\%$. See the sketch below.



(b) A result like $\bar{x} = 27.6\%$ g/dl lies down in the low tail of the density curve, while 30.2% is fairly close to the middle. If $\mu = 31\%$ g/dl, observing a mean of 30.2% or smaller would not be

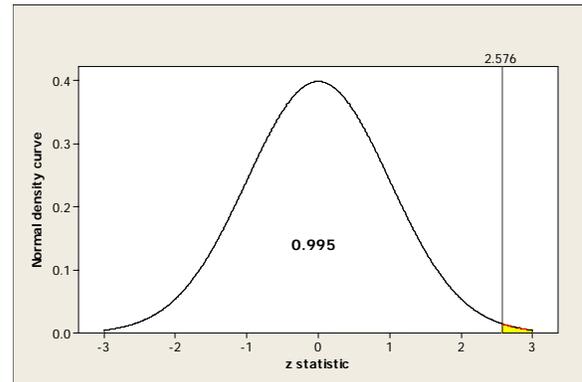
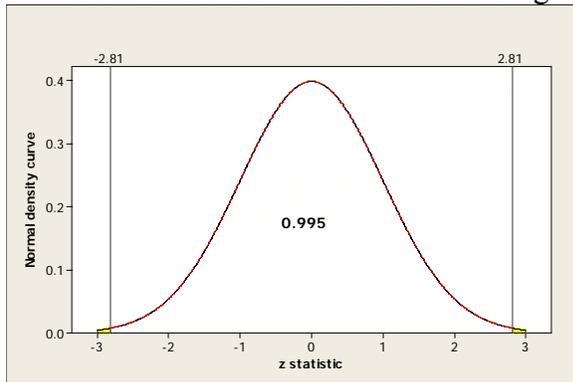
too surprising, but a mean of 27.6% or smaller is unlikely, and it therefore provides evidence that $\mu < 31\%$. (c) For $\bar{x} = 30.2\%$, the test statistic is $z = \frac{30.2 - 31}{9.6/\sqrt{40}} \doteq -0.53$ and the P -value =

0.2981. For $\bar{x} = 27.6\%$, the test statistic is $z = \frac{27.6 - 31}{9.6/\sqrt{40}} \doteq -2.24$ and the P -value = 0.0125. (d)

The P -value of 0.2981 indicates that $\bar{x} = 30.2\%$ gives little reason to doubt H_0 . At both significance levels we would conclude that an average of 31% is spent on housing. The P -value of 0.0125 tells us that $\bar{x} = 27.6\%$ is significant at the 5% level, but not at the 1% level. At the 5% level we would conclude that households spend less than 31% on average for housing, but at the 1% level we would conclude that households spend 31% on average.

11.21 (a) For a two-sided alternative, z is statistically significant at $\alpha = 0.005$ if $|z| > 2.81$. In other words, we would reject H_0 when $z \leq -2.81$ or $z \geq 2.81$. See the sketch below on the left.

(b) For a one-sided alternative (on the positive side), z is statistically significant at $\alpha = 0.005$ if $z > 2.576$. See the sketch below on the right.



11.22 The explanation is not correct. Either H_0 is true (in which case the probability that H_0 is true is 1) or H_0 is false (in which case the probability that H_0 is true is 0). Statistically significant at the $\alpha = 0.05$ level, means that if H_0 is true, then the chance of observing a test statistic of the value we obtained or something more extreme is less than 5%.

11.23 (a) If the population mean is 15, there's about an 8% chance of getting a sample mean as far from or even farther from 15 as we did in this sample. (b) We would not reject H_0 at $\alpha = 0.05$ because the P -value of 0.082 is greater than 0.05. (c) The probability that you are wrong is either 0 or 1, depending on the true value of μ .

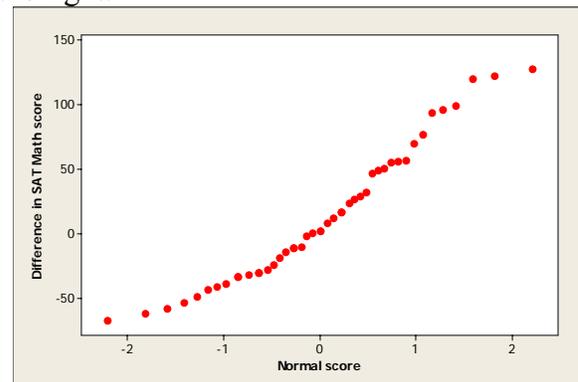
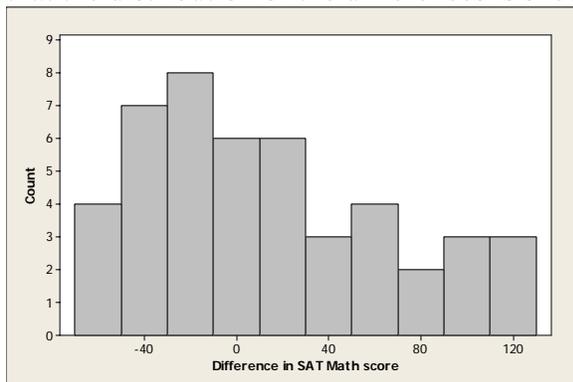
11.24 For $z^* = 2$, the P -value would be $2 \times P(Z > 2) = 2 \times 0.0228 = 0.0456$, and for $z^* = 3$, the P -value would be $2 \times P(Z > 3) = 2 \times 0.0013 = 0.0026$. *Note: In other words, the Supreme Court has chosen to use α no bigger than about 0.05.*

11.25 An SRS is necessary to ensure generalizability, Normality is required to perform calculations using z , and independence is required for the standard deviation of the sampling distribution of \bar{x} to be accurate.

11.26 $P = 0.02$ means that if there were truly no change in cholesterol—if all differences were due to random fluctuation or “chance assignment”—then only 2% of all samples would produce results as extreme or more extreme as the ones found in this sample.

11.27 We test $H_0 : \mu = 5$ versus $H_a : \mu < 5$, where μ = the mean dissolved oxygen content in the stream. Water is taken at randomly chosen locations so the SRS condition is satisfied. The observations are not independent since they are being taken from the same stream, but the number of possible collection locations along the stream (the population size) is clearly much greater than the sample size ($n = 45$). The sample size is large ($n = 45$), so the central limit theorem says the sampling distribution of the sample mean will be approximately Normal. The test statistic is $z = \frac{4.62 - 5}{0.92/\sqrt{45}} = -2.77$ and the P -value is $P(Z \leq -2.77) = 0.0028$. The P -value is less than 0.01, so there is very strong evidence to reject H_0 and conclude that the mean dissolved oxygen content in the stream is less than 5 mg.

11.28 (a) The histogram (on the left) and Normal probability plot (on the right) below suggest that the distribution of the differences is skewed to the right.



The descriptive statistics below from Minitab indicate that the distribution of the differences is centered slightly above 0 (13.11 if you use the mean and 2 if you use the median) with the smallest difference being -67 and the largest difference being 128 .

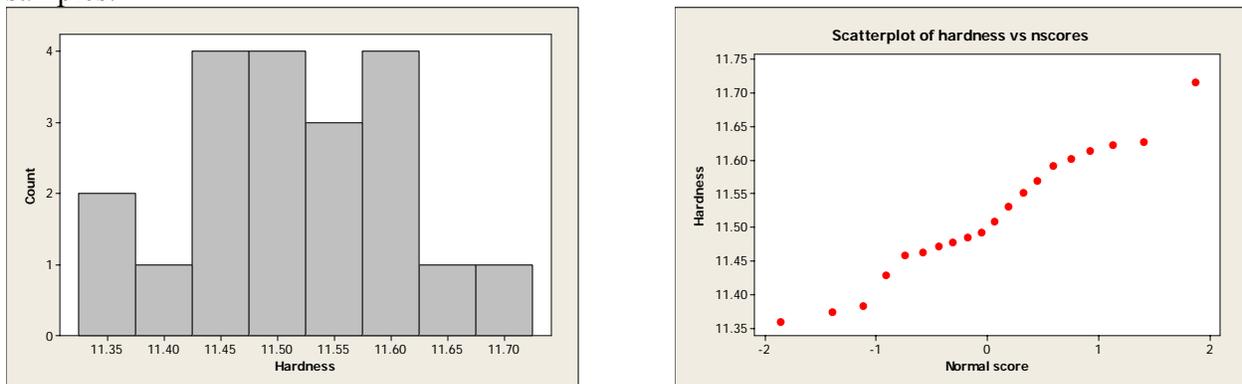
Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
scorediff	46	13.11	52.97	-67.00	-30.50	2.00	52.00	128.00

(b) No, as noted above, the distribution of the differences appears to be skewed to the right. (c) We want to test $H_0 : \mu = 0$ versus $H_a : \mu > 0$, where μ = the mean change in SAT Math scores. The students are randomly chosen high school students, so the SRS condition is satisfied. The difference in the scores for one student is independent of the differences for other students so the independence condition is also satisfied. Even though the population of difference may be skewed to the right, the sample size ($n=46$) is large enough so that the central limit theorem says the sampling distribution of \bar{x} is approximately Normal. The sample mean is approximately

13.11 points, so the test statistic is $z = \frac{13.11 - 0}{50/\sqrt{46}} \doteq 1.78$ and the P -value is $P(Z \geq 1.78) = 0.0375$.

The P -value of 0.0375 is less than 0.05, so we reject H_0 and conclude that the mean change in the SAT Math scores is greater than 0. In other words, the students significantly improve when taking the test for the second time.

11.29 (a) The histogram (on the left) and the Normal probability plot (on the right) below show that the distribution is roughly symmetric with no outliers. The Normal probability plot does show some slight departures from the overall linear trend, but this is not uncommon for small samples.



The descriptive statistics from Minitab (see output below) indicate that the mean of 11.516 is slightly larger than the median of 11.501 and the sample standard deviation of 0.095 is smaller than the population standard deviation of 0.2. 65% of the sample observations are between $\bar{x} \pm 1s$ and 95% of the observations are between $\bar{x} \pm 2s$. Although it is hard to tell based on such a small sample, the data could have come from a Normal population.

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
hardness	20	11.516	0.0950	11.360	11.459	11.501	11.600	11.715

(b) We want to test $H_0 : \mu = 11.5$ versus $H_a : \mu \neq 11.5$, where μ = the mean target value of tablet hardness. The sample was randomly selected without replacement so the SRS condition is satisfied. The number of tablets in the population must be greater than $10 \times 20 = 200$. The question of normality is discussed in part (a). The conditions for a one-sample z -test for a population mean are met. The test statistic is $z = \frac{11.5164 - 11.5}{0.2/\sqrt{20}} \doteq 0.37$ and the P -value = $2 \times 0.3557 = 0.7114$. Because the P -value is greater than any reasonable α level, we do not reject H_0 . This is reasonable variation when the null hypothesis is true, so the hardness levels appear to be on target.

11.30 We want to test $H_0 : \mu = 300$ versus $H_a : \mu < 300$, where μ = the mean amount of cola in a certain type of bottle. The SRS condition is satisfied because the bottles were randomly selected. Since the sampling is without replacement the number of bottles produced in one day must be larger than $10 \times 6 = 60$ (which should be no problem), so the independence condition is satisfied. The distribution of the mean contents for 6 bottles is Normal because the population is Normal. The test statistic is $z = \frac{299.033 - 300}{3/\sqrt{6}} \doteq -0.79$ and the P -value = 0.2148. Because the P -

value of 0.2148 is greater than any reasonable α level, we do not reject H_0 . This is reasonable variation when the null hypothesis is true, so the filling machinery appears to be working properly on this day.

11.31 (a) Yes, the P -value = 0.06 indicates that the results observed are not significant at the 5% level, so the 95% confidence interval will include 10. (b) No, because P -value $<$ 0.1, we can reject $H_0 : \mu = 10$ at the 10% level. The 90% confidence interval would include only those values a for which we could *not* reject $H_0 : \mu = a$ at the 10% level.

11.32 The 95% confidence interval for μ is (28.0, 35.0). (a) No, since the 95% confidence interval includes 34, we cannot reject $H_0 : \mu = 34$ at the 5% level. (b) Yes, since the 95% confidence interval does not include 36, we can reject $H_0 : \mu = 36$ at the 5% level. *Note: We are using a two-sided alternative for both tests.*

11.33 (a) A 90% confidence interval for the mean reading is $104.133 \pm 1.645 \frac{9}{\sqrt{12}} = (99.86, 108.41)$. With 90% confidence, we estimate the mean reading of all radon detectors exposed to 105 picocuries of radon is between 99.86 and 108.41 picocuries. (b) Because 105 falls in this 90% confidence interval, we cannot reject $H_0 : \mu = 105$ in favor of $H_a : \mu \neq 105$. The confidence interval may be wider than we would like, but we do not have evidence to suggest that the mean reading is different from 105 picocuries.

11.34 The two-sided P -value is $2 \times 0.04 = 0.08$. (a) Yes, the P -value = 0.08 indicates that the results observed are not significant at the 5% level, so the 95% confidence interval will include 30. (b) No, because P -value $<$ 0.1, we can reject $H_0 : \mu = 30$ at the 10% level. The 90% confidence interval would include only those values a for which we could *not* reject $H_0 : \mu = a$ at the 10% level.

11.35 (a) The sample may not be representative, as the women have taken themselves to the clinic. Normality should be ok due to the large sample size ($n = 160$). We are sampling without replacement, but the independence condition is satisfied because there are more than $10 \times 160 = 1600$ pregnant women in Guatemala. (b) We want to test $H_0 : \mu = 9.5$ versus $H_a : \mu \neq 9.5$. The test statistic is $z = \frac{9.57 - 9.5}{0.4/\sqrt{160}} \doteq 2.21$ and the P -value = $2 \times 0.0136 = 0.0272$. Since 0.0272 is less than 0.05, we reject H_0 at $\alpha = 0.05$ and conclude that the mean calcium level in healthy, pregnant, Guatemalan women differs from 9.5 grams per deciliter. (c) A 95% confidence interval for the mean calcium level is $9.57 \pm 1.96 \frac{0.4}{\sqrt{160}} = (9.508, 9.632)$. With 95% confidence, we estimate the mean blood calcium of all healthy pregnant women in Guatemala at their first visit to be between 9.508 and 9.632 grams per deciliter.

11.36 We want to test $H_0 : \mu = -0.545$ versus $H_a : \mu > -0.545$. The conditions for the z -test are satisfied because the containers were randomly selected, more than 50 containers are produced

by this producer, and we are told that the freezing temperatures vary according to a Normal distribution. The test statistic is $z = \frac{-0.538 - (-0.545)}{0.008/\sqrt{5}} \doteq 1.96$ and the P -value = 0.025. Since 0.025 is less than 0.05, we reject H_0 at $\alpha = 0.05$ and conclude that this producer appear to be adding water to the milk. The mean freezing point is significantly higher than -0.545°C .

11.37 (a) Yes, 30 is in the 95% confidence interval, because P -value = 0.09 means that we would not reject H_0 at $\alpha = 0.05$. (b) No, 30 is not in the 90% confidence interval, because we would reject H_0 at $\alpha = 0.10$.

11.38 (a) No, 13 is in the 90% confidence interval, so H_0 cannot be rejected at the 10% level.

(b) No, the sample mean is $\bar{x} = \frac{12+15}{2} = 13.5$ and the standard error is approximately

$\frac{1.5}{1.645} \doteq 0.91$, so 13.5 is less than one standard error from 13. (c) Yes, 10 is not in the 90%

confidence interval, so H_0 can be rejected at the 10% level. (d) Here the answer depends on the direction of the alternative. If the alternative is $H_a : \mu < 10$, the answer is no because the sample mean of 13.5 is well above 10. However, if the alternative is $H_a : \mu > 10$, the answer is yes

because the sample mean of 13.5 is approximately $\frac{3.5}{0.91} \doteq 3.85$ standard errors above 10.

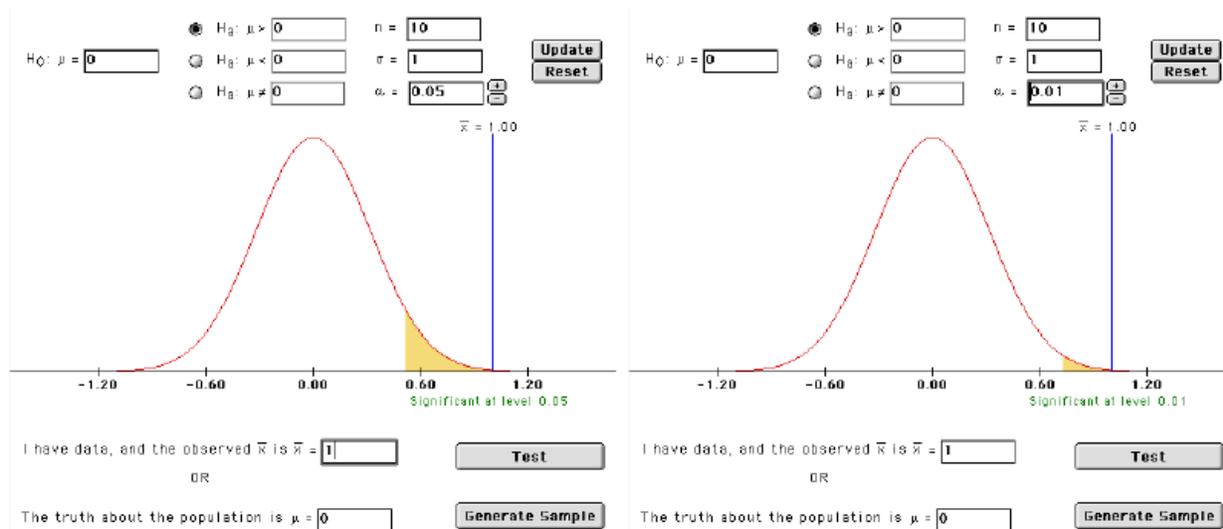
11.39 Yes. We want to test $H_0 : \mu = 450$ versus $H_a : \mu > 450$. The conditions for the z -test are satisfied because the test was given to an SRS of seniors from California, there are more than 5000 seniors in the state of California, and the central limit theorem says that the sample mean of 500 scores will be approximately Normal, even if the distribution of scores is slightly skewed.

The test statistic is $z = \frac{461 - 450}{100/\sqrt{500}} \doteq 2.46$ and the P -value = 0.0069. Since 0.0069 is less than

0.01, we reject H_0 at $\alpha = 0.01$ and conclude that mean SAT Math score is significantly higher than 450.

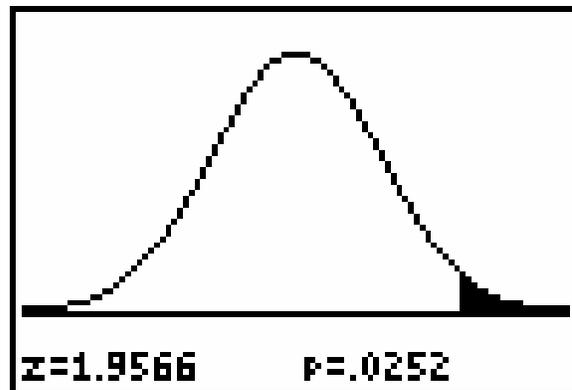
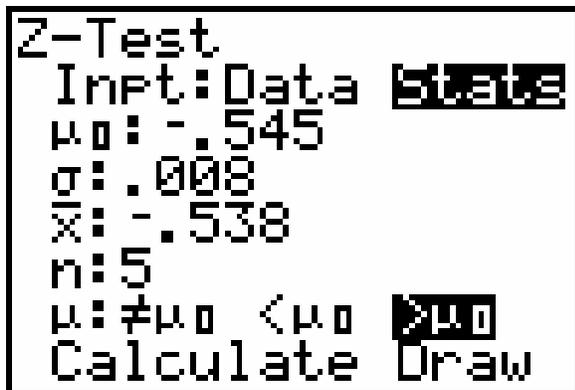
11.40 The 95% confidence interval for the mean amount of sugar in the hindgut is (1.9 mg, 6.5 mg). (a) Yes, since 7 mg is not contained in the 95% confidence interval, we reject $H_0 : \mu = 7$ in favor of the alternative $H_a : \mu \neq 7$ at the 5% level. (b) No, since 5 mg is contained in the 95% confidence interval, we cannot reject $H_0 : \mu = 5$ at the 5% level.

11.41 (a) Yes, see the screenshot below on the left. (b) As one can judge from the shading under the Normal curve, the results are not significant for $\bar{x} \leq 0.5$, and significant for $\bar{x} \geq 0.6$. In fact, the cutoff is about 0.52, which is approximately $1.645/\sqrt{10}$.



(c) Yes, see the screenshot above on the right. As one can judge from the shading under the Normal curve, the results are significant for $\bar{x} \geq 0.8$. In fact, the cutoff is about 0.7354, which is approximately $2.326/\sqrt{10}$. Smaller α means that \bar{x} must be farther away from μ_0 in order to reject H_0 .

11.42 (a) We want to test $H_0: \mu = -0.545$ versus $H_a: \mu > -0.545$. The TI calculator screens are shown below.



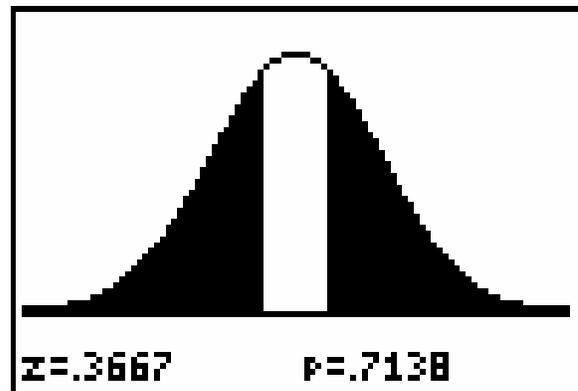
The results, except for some slight rounding differences, are the same as those in Exercise 11.36.

(b) We want to test $H_0: \mu = 11.5$ versus $H_a: \mu \neq 11.5$. The TI calculator screens are shown below.

```

Z-Test
Inpt: Data Stats
μ₀: 11.5
σ: .2
List: L1
Freq: 1
μ: 12.0 <μ₀ >μ₀
Calculate Draw

```



The results, except for some slight rounding differences are the same as those in Exercise 11.29.

11.43 (a) No, not at the 5% level. The test statistic is $z = \frac{536.7 - 518}{114/\sqrt{100}} \doteq 1.64$ and the P -value = 0.0505. (b) Yes, the test statistic is $z = \frac{536.8 - 518}{114/\sqrt{100}} \doteq 1.65$ and P -value = 0.0495 is less than 0.05.

11.44 (a) The test statistic is $z = \frac{522 - 518}{100/\sqrt{100}} \doteq 0.4$ and the P -value = 0.3446, which is not below $\alpha = 0.05$. (b) The test statistic is $z = \frac{522 - 518}{100/\sqrt{1000}} \doteq 1.26$ and the P -value = 0.1038, which is still not below $\alpha = 0.05$. (c) The test statistic is $z = \frac{522 - 518}{100/\sqrt{10000}} = 4.0$ and the P -value < 0.0001 , which is statistically significant at any reasonable α level.

11.45 A 99% confidence interval for the mean SAT Math score μ after coaching is $522 \pm 1.96 \frac{100}{\sqrt{n}}$. (a) When $n = 100$, the 99% confidence interval for μ is (496.24, 547.76). (b) When $n = 1000$, the 99% confidence interval for μ is (513.85, 530.15). (c) When $n = 10,000$, the 99% confidence interval for μ is (519.42, 524.58).

11.46 This is not information taken from an SRS, or from any kind of sample. We have information about all presidents—the whole population of interest.

11.47 (a) No, in a sample of size $n = 500$, we expect to see about 5 people who do better than random guessing, with a significance level of 0.01. These four might have ESP, or they may simply be among the “lucky” ones we expect to see. (b) The researcher should repeat the procedure on these four to see if they again perform well.

11.48 The answer (b). A test of significance is used to determine if the observed effect is due to chance.

11.49 (a) $H_0 : p = 0.75$ versus $H_a : p > 0.75$. (b) A Type I error would be committed if the manager decided that they were responding to more than 75% of the calls within 8 minutes when, in fact they were responding to less than 75% within that time. A Type II error would be committed if the manager decided that they were responding to 75% or less of the calls within 8 minutes, when in fact they were responding to more than 75% in that time. (c) The consequence of a Type I error is that city officials may be satisfied with response times and see no need to improve, when they should. The consequence of a Type II error is that officials try and make the response times even faster when there is no need to do so. (d) A Type I error is much more serious in this situation because city officials think things are better than they actually are. (e) Students can give either answer with an appropriate defense. However, most students will probably say they are more interested in testing hypotheses about the mean response time. Short rationales are: 1. Lower mean response times are better; 2. A higher proportion of successes (getting to the accident in less than 8 minutes) is better.

11.50 (a) $H_0 : \mu = 130$ versus $H_a : \mu > 130$. (b) A Type I error is committed by telling an employee that they have a high systolic blood pressure when in fact they do not. A Type II error is committed by failing to notify an employee who has high blood pressure. (c) You obviously want to make the chance of a Type II error as small as possible. While it is inconvenient to send some employees for further testing when their blood pressure is OK (a Type I error), death could result from a Type II error.

11.51 (a) A Type I error is committed if you decide the mean sales for the new catalog will be more than \$40 when in fact it is not. The consequence is that you waste company resources by changing the production process to the new cover design and it won't increase the mean sales. (b) A Type II error is committed if you decide that the mean sales for the new catalog will be \$40 (or less) when it turns out to be more than \$40. The consequence is that the company will not make the additional profits that would have been made by increasing sales with the new cover. (c) Increasing profits would be nice, but wasting money or resources is never a good idea in business—Type I is more serious. (d) The probability of a Type I error is $\alpha = 0.01$, and the probability of a Type II error is $\beta = 0.379$. (e) 44.4139 is the 99th percentile of the sampling distribution of \bar{x} when $H_0 : \mu = 40$ is true. That is, $40 + 2.32635 \times \frac{60}{\sqrt{1000}} \doteq 44.4139$.

11.52 (a) $H_0 : \mu = 10,000$ psi versus $H_a : \mu < 10,000$ psi. (b) A Type I error is committed by telling the consumer that the wood with blue stain is weaker when in fact it is not. The consequence is that you must find more wood without blue stains. A Type II error is committed by not telling the consumer that the blue-stained wood is weaker when in fact it is. A consequence is that you may lose loyal customers because they will lose their trust in the company. (c) Spending more money to find wood that does not contain blue stains (Type I error) would increase expenses for the company, but that is not as serious as being dishonest to customers. A Type II error is more serious.

11.53 (a) $H_0 : \mu = \$85,000$ versus $H_a : \mu > \$85,000$, where μ = the mean income of residents near the restaurant. (b) A Type I error is committed if you conclude that the local mean income

exceeds \$85,000 when in fact it does not. The consequence is that you will open your restaurant in a location where the residents will not be able to support it. A Type II error is committed if you conclude that the local mean income does not exceed \$85,000 when in fact it does. The consequence of this error is that you will not open your restaurant in a location where the residents would have been able to support it. (c) A Type I error is more serious. If you opened your restaurant in an inappropriate area, then you would sustain a financial loss. If you failed to open your restaurant in an appropriate area, then you would miss out on an opportunity to earn a profit, but you would not necessarily lose money (e.g., if you chose another appropriate location in its place). (d) The smallest significance level, $\alpha = 0.01$, is the most appropriate, because it would minimize your probability of committing a serious Type I error. (e) When $\mu = 87,000$, there is about a 69% chance that you will open a restaurant in that area, and the probability of committing a Type II error is 0.3078, or about 31%.

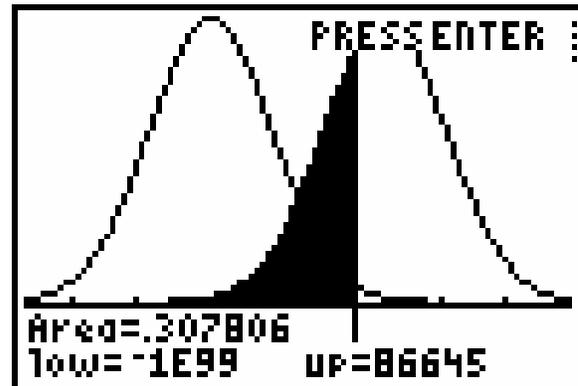
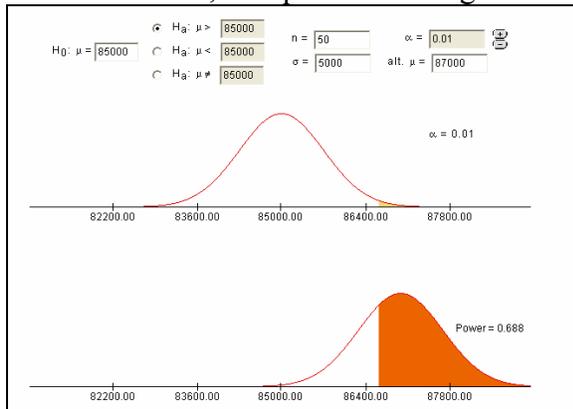
11.54 (a) $H_0 : \mu = 2 \text{ mg}$ versus $H_a : \mu \neq 2 \text{ mg}$, where μ = the mean salt content of a certain type of potato chips. (b) A Type I error is committed if the company decides that the mean salt content is different from 2 mg when in fact it is not. The probability of making a Type I error is $\alpha = 0.05$. (c) A Type II error is committed if the company sticks by its claim when the mean salt content is different from 2 mg. The probability of making a Type II error when $\mu = 2.05$ is 0.0576 or about 6%. (d) The power of the test is $1 - 0.0576 = 0.9424$ or about a 94.24% chance of detecting this difference if 0.05. (e) The power of this test is also 0.9424 because it is the same distance away from the mean specified in the null hypothesis. You should have the same power of detecting differences that are the same distance away from the mean, whether you are above the mean or below the mean does not matter for two-sided alternatives. (f) The probability of a Type I error would increase from 0.05 to 0.10, so the probability of a Type II error would decrease and the power would increase. (g) Throwing away good chips, the consequence of making a Type I error, is not a good idea, but it is better than telling consumers that the chips contain 2 mg of chips when they do not. A Type II error probability could create serious health problems for some consumers. Thus, the company should try to minimize the chance of making a Type II error, which means that the highest Type I error rate $\alpha = 0.1$ would be best in this situation.

11.55 The power of this study is far lower than what is generally desired—for example, it is well below the “80% standard” mentioned in the text. Twenty percent power for the specified effect means that, if the effect is present, we will only detect it 20% of the time. With such a small chance of detecting an important difference, the study should probably not be run (unless the sample size is increased to give sufficiently high power).

11.56 The power for $\mu = 80$ will be higher than 0.5, because larger differences are easier to detect.

11.57 (a) For both p_0 and p_1 , each probability is between 0 and 1 and the sum of the probabilities for each distribution is 1. (b) The probability of a Type I error is $P(X = 0 \text{ or } X = 1 \text{ when the distribution is } p_0) = 0.1 + 0.1 = 0.2$. (c) The probability of a Type II error is $P(X > 1 \text{ when the distribution is } p_1) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) = 5 \times 0.1 = 0.5$.

11.58 The calculator program TYPE2 overlays the two distributions on one graph, whereas the *Power* applet puts the two different distributions on two graphs. However, the calculations are almost identical, except for rounding differences. The screen shots below are for Exercise 11.53.



The upper distribution of the applet gives the value of α and shades the area corresponding to the probability of a Type I error. The lower distribution gives the value of the power and shades the corresponding area. The remaining area of the distribution corresponds to the probability of a Type II error. The TYPE2 program however, superimposes both graphs on one screen and also gives the value of β and the critical value of \bar{x} .

11.59 A larger sample gives more information and therefore gives a better chance (or larger probability) of detecting differences. That is, larger samples give more power.

11.60 (a) $H_0 : \mu = 120$ psi versus $H_a : \mu < 120$ psi, where μ = the mean water pressure for pipes from this supplier. (b) The power of the test in (a) when $n = 25$, $\alpha = 0.05$, and $\mu = 125$ is about 0.93 or 93% according to Minitab and the *Power* applet. (c) Each pipe is numbered from 0001 to 1500, and then 4 digit random numbers are examined, starting at line 140, until 40 numbers between 0001 and 1500 are identified. The pipes with these numbers would be the 40 pipes in our sample. The first two pipes in the sample are 1297 and 484. Notice that even though many students will suggest this method, it is not very efficient because a considerable number of digits will need to be examined to get your sample. It would be much more efficient to use the *Simple Random Sample* applet or a random number generator. (d) A Type I error is committed if the construction manager tells the supplier that their pipes do not meet the specification, when in fact they do. The consequence of this error will be a strained business relationship. A Type II error is committed when the manager says that the pipes meet the specification, but they don't. The consequence of this error may be leaky pipes, major water damage, or pipes that won't work and need to be replaced. (e) Type II error is obviously the most serious in this situation, so we should minimize the probability of making a Type II error by using the largest significance level of 0.10. (f) It only takes one weak pipe to create major problems, so it would be best to test hypotheses about the proportion of pipes that have a breaking strength of less than 120 psi.

11.61 (a) $H_0 : \mu = \frac{5000}{50} = 100$ versus $H_a : \mu > 100$. (b) A Type I is committed if Captain Ben concludes that the mean weight of the checked baggage is greater than 100 pounds, when in fact it is not. The consequence of this error is for Captain Ben to keep the plane on the ground, even

though it is safe to fly. Obviously, the passengers will not be happy with this decision. A Type II error is committed if Captain Ben concludes that the mean weight of the checked baggage is equal to 100 pounds, when in fact it is heavier. The consequence of this error is that Captain Ben will take off with luggage that exceeds the safety standards and the jet may experience mechanical problems. (c) We want to minimize the chance of making a Type II error, so set the chance of making a Type I error at the maximum of the values provided— $\alpha = 0.10$. (d) The sample is 20% of the population. The weight of checked bags for passengers may not be normally distributed. In short, two of our three conditions, Independence and Normality, may not be satisfied.

11.62 (a) The probability of a Type II error is $1 - 0.82 = 0.18$. (b) The power of the test will be the same for this alternative, because it is the same distance from 0. The symmetry of two-sided tests with the Normal distribution means that we only need to consider the size of the difference, not the direction. (c) The power when $\mu = -6$ would be smaller because it is closer to 0, and hence harder to detect, than the difference in part (b).

11.63 Finding something to be “statistically significant” is not really useful unless the significance level is sufficiently small. While there is some freedom to decide what “sufficiently small” means, $\alpha = 0.5$ would lead your team to incorrectly rejecting H_0 half the time, so it is clearly a bad choice. (This approach would be essentially equivalent to flipping a coin to make your decision!)

11.64 (a) A Type I error would be committed if the inspector concluded the mean contents were below the target of 300 ml, when in fact they are not. A Type II error would be committed if the inspector concluded that the filling machines were working properly, when in fact they were putting less cola in the bottles than specified. The power of the test is the probability of detecting mean content below 300 ml. (b) According to the *Power* applet, the power of this test against the alternative $\mu = 299$ is 0.198 or about 20%. (Minitab gives 0.2037 or about 20%.) (c) According to the *Power* applet, the power of this test against the alternative $\mu = 295$ is 0.992 or 99.2%. (Minitab gives 0.9926 or 99.26%.) (d) Answers will vary. Students may increase n , increase α , or decrease σ . For example increasing the sample size to $n = 7$ gives a power of 0.997 and increasing the significance level to $\alpha = 0.10$ gives a power of 0.997.

CASE CLOSED!

1. Yes, in order to use the inference methods from this chapter the sample should be an SRS. If the sample is not representative of all tablets produced in this batch, then it will not make sense to use them for inferences about the mean contents of acetylsalicylic acid for all tablets in this batch. 2. We want to test $H_0: \mu = 320\text{mg}$ versus $H_a: \mu \neq 320\text{mg}$, where μ = the mean content of the active ingredient (acetylsalicylic acid). The tablets may contain too much or too little of the active ingredient, so the alternative should be two sided. 3. A Type I error would occur if we conclude that the company is not putting the correct amount of the active ingredient in the tablets, when in fact they are. The consequence of this error is that the company will dispose of a good batch of tablets. A Type II error would occur if we conclude that the tablets contain the correct amount of the active ingredient when in fact they have too much or not enough. The consequences of this error could be death (in the most severe situation) from an overdose of the

active ingredient or unhappy customers because the pills do not have enough of the active ingredient to relieve a headache. 4. Since Type II error is the most serious error in this situation, we should use the highest reasonable significance level (0.1) to minimize the probability of a

Type II error. 5. The test statistic is $z = \frac{321.028 - 320}{3/\sqrt{36}} \doteq 2.06$ and the P -value =

$2 \times 0.0197 = 0.0394$. Since 0.04 is less than 0.1, we reject H_0 and conclude that the mean content of the active ingredient is significantly different from 320 mg. 6. A 90% confidence interval for the mean content of acetylsalicylic acid is $321.028 \pm 1.645 \frac{3}{\sqrt{36}} = (320.21, 321.85)$. 7.

The power of the test when $\mu = 321$ mg is 0.633 or about 63%. (Minitab gives a power of 0.6389.) This power could be increased by increasing the sample size or increasing the significance level (although it is rare to conduct hypothesis tests with a significance level higher than 0.1), or decreasing σ . 8. Answers will vary, but the report should contain graphical displays, numerical statistics, and a clear conclusion for the executives. In short, this batch of tablets should not be distributed to drugstores because the mean amount of the active ingredient is significantly different from the specified amount of 320 mg. Delivering disappointing news like this to executives is never easy! Medical experts may argue that the 90% confidence interval suggests that the mean contents are only slightly off target, and this difference is not of any practical significance. If that is the case, then the company may decide to send the shipment with a warning label and reexamine their production process.

11.65 (a) $H_0 : \mu = \$72,500$; $H_a : \mu > \$72,500$. (b) $H_0 : p = 0.75$; $H_a : p < 0.75$. (c) $H_0 : \mu = 20$ seconds; $H_a : \mu < 20$ seconds.

11.66 We want to test $H_0 : \mu = 120$ versus $H_a : \mu \neq 120$. The test statistic is

$z = \frac{123.8 - 120}{10/\sqrt{40}} \doteq 2.40$ and the P -value = $2 \times 0.0082 = 0.0164$. Yes, since 0.0164 is less than 0.05,

we reject H_0 and conclude that the mean yield of corn in the United States is not 120 bushels per acre. This conclusion holds even if the distribution of corn yields is slightly non-Normal, because the sample size ($n = 40$) is reasonably large.

11.67 The two-sided P -value is $2 \times 0.02 = 0.04$. (a) Yes, the P -value = 0.04 indicates that the results observed are not significant at the 1% level, so the 99% confidence interval will include 15. (b) No, because the P -value < 0.05 , we can reject $H_0 : \mu = 15$ at the 5% level. The 95% confidence interval would include only those values a for which we could *not* reject $H_0 : \mu = a$ at the 5% level.

11.68 We expect more variation with small sample sizes, so even a large difference between \bar{x} and μ_0 (or whatever measures are appropriate in our hypothesis test) might not turn out to be significant. If we were to repeat the test with a larger sample, the decrease in the standard error might give us a small enough P -value to reject H_0 .

11.69 We want to test $H_0: \mu = 150$ versus $H_a: \mu < 150$. The test statistic is $z = \frac{137 - 150}{65/\sqrt{269}} \doteq -3.28$ and the P -value = 0.0005. Since 0.0005 is less than any reasonable significance level, this is very strong evidence that students study less than an average of 2.5 hours per night.

11.70 We want to test $H_0: \mu = 0$ versus $H_a: \mu > 0$. The test statistic is $z = \frac{6.9 - 0}{55/\sqrt{104}} \doteq 1.28$ and the P -value = 0.1003. Since 0.1003 is greater than any reasonable significance level, we cannot reject H_0 and conclude that this is not good evidence that the mean real compensation of all CEOs increased. In other words, the sample mean appears to indicate a positive increase, but this increase is not statistically significant.

11.71 (a) The margin of error decreases. (b) The P -value decreases. (c) The power increases. Note: All of these changes would be viewed favorably by statisticians conducting the analysis and clients who are interested in making inferences.

11.72 (a) The 95% confidence interval would be wider. To be more confident that our interval includes the true population parameter, we must allow a larger margin of error. So the margin of error for 95% confidence is larger than for 90% confidence. (b) We would not reject $H_0: \mu = \$16$ because \$16 falls within the 90% confidence interval, indicating that the two-sided P -value is at least 0.10. However, we could reject $H_0: \mu = \$15$ at $\alpha = 0.10$ because the 90% confidence interval does not include \$15.

11.73 (a) $H_0: \mu = 300$ (the company's claim is true) versus $H_a: \mu < 300$ (the mean breaking strength is less than the company claim). (b) A Type I error would be committed if we conclude that the company's claim is incorrect ($\mu < 300$) when in fact it is legitimate ($\mu = 300$). A Type II error would be committed if we conclude that the company's claim is legitimate when in fact it is invalid. A Type II error is more serious in this case, because allowing the company to continue the "false advertising" of its chairs' strength could lead to injuries, lawsuits, and other serious consequences. (c) If the null hypothesis is true, then the sampling distribution of \bar{x} is approximately Normal with mean 300 pounds and standard deviation $\sigma_{\bar{x}} = 15/\sqrt{30} \doteq 2.7386$. The 5th percentile of the $N(300, 2.7386)$ distribution is 295.495, so all values at or below 295.495 pounds would cause us to reject H_0 . (d) The probability of a Type II error is 0.022 or about 2%. (e) Increase the sample size or increase the significance level.

11.74 The study may have rejected $\mu = \mu_0$ (or some other null hypothesis), but with such a large sample size, such a rejection might occur even if the actual mean (or other parameter) differs only slightly from μ_0 . For example, there might be no practical importance to the difference between $\mu = 10$ and $\mu = 10.5$.