Answers to Cumulative AP® Practice Test 4

AP4.1 e
AP4.2 c
AP4.3 e
AP4.4 a
AP4.5 b
AP4.6 e
AP4.7 d
AP4.8 a
AP4.9 e
AP4.10 a
AP4.11 b
AP4.12 d
AP4.13 e
AP4.14 d
AP4.15 b
AP4.16 a
AP4.17 d
AP4.18 c
AP4.19 a
AP4.20 b
AP4.21 e
AP4.22 e
AP4.23 b
AP4.24 c
AP4.25 d
AP4.26 e
AP4.27 c
AP4.28 c
AP4.29 b
AP4.30 b
AP4.31 d
AP4.32 a
AP4.33 e
AP4.34 b
AP4.35 a
AP4.36 b
AP4.37 d
AP4.38 a
AP4.39 d
AP4.40 d

AP4.41 S: $H_0: \mu_1 - \mu_2 = 0$ vs. $H_a: \mu_1 - \mu_2 \neq 0$, where μ_1 = true mean difference in electrical potential for diabetic mice and μ_2 = true mean difference in electrical potential for normal mice. P: Two-sample t test for $\mu_1 - \mu_2$. Random: Independent random samples. 10%: $n_1 = 24$ is less than 10% of all diabetic mice and $n_2 = 18$ is less than 10% of all normal mice. Normal/Large Sample Size: No outliers or strong skewness. D: $t = 2.55$. Using df = 23, the P-value is between 0.01 and 0.02. Using df = 38.46, P-value = 0.0149. C: Because the P-value of 0.0149 is less than $\alpha = 0.05$, we reject H_0. There is convincing evidence that the true mean difference in electric potential for diabetic mice is different than for normal mice.

AP4.42 (a) $H_0: p_1 - p_2 = 0$ vs. $H_a: p_1 - p_2 < 0$, where p_1 = the true proportion of women like the ones in the study who were physically active as teens that would suffer a cognitive decline and p_2 = the true proportion of women like the ones in the study who were not physically active as teens that would suffer a cognitive decline. (b) A two-sample z test for $p_1 - p_2$. (c) No. Because the participants were mostly white women from only four states, the findings may not be generalizable to women in other racial and ethnic groups or who live in other states. (d) Two variables are confounded when their effects on the response variable cannot be distinguished from one another. For example, women who were physically active as teens might have also done other things differently as well, such as eating a healthier diet. We would be unable to determine if it was their physically active youth or their healthier diet that slowed their level of cognitive decline.

AP4.43 (a) Because the first question called it a "fat tax," people may have reacted negatively because they believe this is a tax on those who are overweight. The second question provides extra information that gets people thinking about the obesity problem in the U.S. and the increased health care that could be provided as a benefit with the tax money. Better: "Would you support or oppose a tax on non-diet sugared soda?" (b) This method samples only people at fast-food restaurants. They may go to these restaurants because they like the sugary drinks and wouldn't want to pay a tax on their favorite beverages. Thus, it is likely that the proportion of those who would oppose such a tax will be overestimated with this method. Better: take a random sample of all New York State residents. (c) Use a stratified random sampling method in which each state is a stratum.

AP4.44 (a) $P(S) = (0.1)(0.3) + (0.4)(0.2) + (0.5)(0.1) = 0.16$.

(b) $P(C|S) = \dfrac{(0.5)(0.1)}{(0.1)(0.3) + (0.4)(0.2) + (0.5)(0.1)} = 0.3125.$

AP4.45 (a) No. The scatterplot exhibits a strong curved pattern. (b) B, because the scatterplot shows a much more linear pattern and its residual plot shows no leftover patterns.

(c) $\widehat{\ln(\text{weight})} = 15.491 - 1.5222 \ln(3700) = 2.984$, thus

$\widehat{\text{weight}} = e^{2.984} = 19.77$mg. (d) About 86.3% of the variation in ln(seed weight) is accounted for by the linear model relating ln(seed weight) to ln(seed count).

AP4.46 (a) Let X = diameter of a randomly selected lid. Because X follows a Normal distribution, the sampling distribution of $\bar{x}$ also follows a Normal distribution. $\mu_{\bar{x}}$ = 4 inches and $\sigma_{\bar{x}} = \dfrac{0.02}{\sqrt{25}} = 0.004$ 0.004 inches. (b) We want to find $P(\bar{x} < 3.99$ or $\bar{x} > 4.01)$ using the $N(4, 0.004)$ distribution. $z = \dfrac{3.99 - 4}{0.004} = -2.50$ and $z = \dfrac{4.01 - 4}{0.004} = 2.50$. $P(Z < -2.50$ or $Z > 2.50) = 0.0124$. Assuming that the machine is working properly, there is a 0.0124 probability that the mean diameter of a sample of 25 lids is less than 3.99 inches or greater than 4.01 inches. (c) We want to find $P(4 < \bar{x} < 4.01)$ using the $N(4, 0.004)$ distribution. $z = \dfrac{4 - 4}{0.004} = 0$ and $z = \dfrac{4.01 - 4}{0.004} = 2.50$. $P(0 < Z < 2.50) = 0.4938$. Assuming that the machine is working properly, there is a 0.4938 probability that the mean diameter of a sample of 25 lids is between 4.00 and 4.01 inches. (d) Let Y = the number of samples (out of 5) in which the sample mean is between 4.00 and 4.01. The random variable Y has a binomial distribution with $n = 5$ and $p = 0.4938$. *Using technology:* $P(X \geq 4) = 1 - P(X \leq 3)$ = `1 - binomcdf(trials:5, p:0.4938, x value:3)` = 0.1798. (e) Because the probability found in part (b) is less than the probability found in part (c), getting a sample mean below 3.99 or above 4.01 is more convincing evidence that the machine needs to be shut down. This event is much less likely to happen by chance when the machine is working correctly. (f) Answers will vary.

12.53 (a) *S*: p = true proportion of all AP® teachers attending this workshop who have tattoos. *P*: One-sample z interval for p. Random: Random sample. 10%: The sample size ($n = 98$) is less than 10% of the population of teachers at this workshop (1100). Large Counts: 23 and 75 are both ≥ 10. *D*: (0.151, 0.319). *C*: We are 95% confident that the interval from 0.151 to 0.319 captures the true proportion of AP® teachers at this workshop who have tattoos. (b) Yes. Because the value 0.14 is not included in the interval, we have convincing evidence that the true proportion of teachers at the workshop who have a tattoo is not 0.14. (c) If we had two more failures, the interval will shift to lower values and might include the value 0.14. However, the new interval is (0.148, 0.312), which does not include the value 0.14. So the answer would not change if we got responses from the 2 nonresponders.

Answers to Chapter 12 Review Exercises

R12.1 (a) There is a moderately strong, positive linear relationship between the thickness and the velocity. (b) $\hat{y} = 70.44 + 274.78x$, where y is the velocity and x is the thickness. (c) Residual = $104.8 - 180.352 = -75.552$, so the line overpredicts the velocity by 75.552 ft/sec. (d) The linear model is appropriate. The scatterplot shows a linear relationship and the residual plot has no leftover patterns. (e) Slope: For each increase of an inch in thickness, the predicted velocity increases by 274.78 feet/second. s: When using the least-squares regression line with x = thickness to predict y = velocity, we will typically be off by about 56.36 feet per second. r^2: About 49.3% of the variation in velocity is accounted for by the linear relationship relating velocity to thickness. SE_b: If we take many different random samples of 12 pistons and compute the least-squares regression line for each sample, the estimated slope will typically vary from the slope of the population regression line for predicting velocity from thickness by about 88.18.

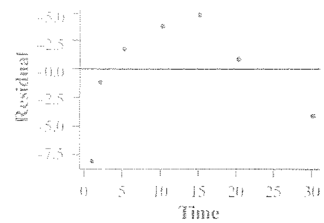
R12.2 *S*: $H_0{:}\beta = 0$ versus $H_a{:}\beta \neq 0$, where β is the slope of the population regression line relating thickness to velocity. *P*: t test for β. Linear: The residual plot shows no leftover patterns. Independent: Knowing the velocity for one piston should not help us predict the velocity for another piston. Also, the sample size ($n = 12$) is less than 10% of the pistons in the population. Normal: We are told that the Normal probability plot of the residuals is roughly linear. Equal SD: The residual plot shows roughly equal scatter for all x values. Random: The data come from a random sample. *D*: $t = 3.116$. With df = 10, the *P*-value is between 0.01 and 0.02 (0.0109). *C*: Because the *P*-value of 0.0109 is less than $\alpha = 0.05$, we reject H_0. There is convincing evidence of a linear relationship between thickness and gate velocity in the population of pistons formed from this alloy of metal.

R12.3 *D*: With df = 12 − 2 = 10, (78.315, 471.245). *C*: We are 95% confident that the interval from 78.315 to 471.245 captures the slope of the population regression line for predicting velocity from thickness for the population of pistons formed from this alloy of metal. Because 0 is not in the interval, we reject 0 as a plausible value for the slope of the population regression line, as in R12.2.

R12.4 The Linear condition is violated because there is clear curvature to the scatterplot and an obvious curved pattern in the residual plot. The Random condition may not be met because we weren't told if the sample was selected at random.

R12.5 (a) Yes, because there is no leftover pattern in the residual plot. (b) $\hat{y} = -0.000595 + 0.3\left(\frac{1}{x^2}\right)$. Here, y = intensity and x = distance. (c) $\hat{y} = -0.000595 + 0.3\left(\frac{1}{(2.1)^2}\right) = 0.0674$ candelas

R12.6 (a)



(b) There is a leftover pattern in the residual plot, so the relationship between practice time and percent of words recalled is not linear. (c) Power, because the scatterplot showing ln(recall) versus ln(time) is more linear than the scatterplot showing ln(recall) versus time. (d) Power: $\widehat{\ln y} = 3.48 + 0.293 \ln(25) = 4.423$ and $\hat{y} = e^{4.423} = 83.35$ percent of words recalled. Exponential: $\widehat{\ln y} = 3.69 + 0.0304(25) = 4.45$ and $\hat{y} = e^{4.45} = 85.63$ percent of words recalled. Based on my answer to part (c), I think the power model will give a better prediction.

Answers to Chapter 12 AP® Statistics Practice Test

T12.1 c
T12.2 b
T12.3 d
T12.4 a
T12.5 d
T12.6 d
T12.7 e
T12.8 d
T12.9 d
T12.10 c

T12.11 (a) $\hat{y} = 4.546 + 4.832x$, where y is the weight gain and x is the dose of growth hormone. (b) (i) For each 1-mg increase in growth hormone, the predicted weight gain increases by about 4.832 ounces. (ii) If a chicken is given no growth hormone ($x = 0$), the predicted weight gain is 4.546 ounces. (iii) When using the least-squares regression line with x = dose of growth hormone to predict y = weight gain, we will typically be off by about 3.135 ounces. (iv) If we repeated this experiment many times, the sample slope will typically vary by about 1.0164 from the true slope of the least-squares regression line with y = weight gain and x = dose of growth hormone. (v) About 38.4% of the variation in weight gain is accounted for by the linear model relating weight gain to the dose of growth hormone. (c) *S*: $H_0{:}\beta = 0$ versus $H_a{:}\beta \neq 0$, where β is the slope of the true regression line relating y = weight gain to x = dose of growth hormone. *P*: t test for β. *D*: $t = 4.75$, df = 13, and *P*-value = 0.0004. *C*: Because the *P*-value of 0.0004 is less than $\alpha = 0.05$, we reject H_0. There is convincing evidence of a linear relationship between the dose of growth hormone and weight gain for chickens like these. (d) *D*: With df = 13, (2.6373, 7.0273). *C*: We are 95% confident that the interval from 2.6373 to 7.0273 captures the slope of the true regression line relating y = weight gain to x = dose of growth hormone for chickens like these.

T12.12 (a) There is clear curvature evident in both the scatterplot and the residual plot. (b) 1: $\hat{y} = 2.078 + 0.0042597(30)^3 = 117.09$ board feet. 2: $\widehat{\ln y} = 1.2319 + 0.113417(30) = 4.63441$ and $\hat{y} = e^{4.63441} = 102.967$ board feet. (c) The residual plot for Option 1 is much more scattered, while the plot for Option 2 shows curvature, meaning that the model from Option 1 relating the amount of usable lumber to cube of the diameter is more appropriate.

Section 12.2

Answers to Check Your Understanding

page 782: 1. Option 1: $\widehat{\text{premium}} = -343 + 8.63(58) = \$157.54$

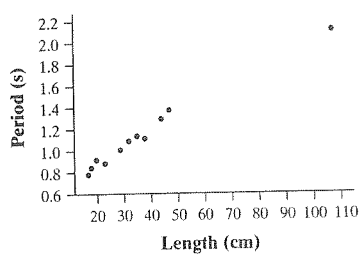
Option 2: $\widehat{\ln(\text{premium})} = -12.98 + 4.416(\ln 58) = 4.9509$
$\rightarrow \hat{y} = e^{4.9509} = \$141.30$

Option 3: $\widehat{\ln(\text{premium})} = -0.063 + 0.0859(58) = 4.9192$
$\rightarrow \hat{y} = e^{4.9192} = \$136.89$

2. Exponential (Option 3), because the scatterplot showing ln(premium) versus age was the most linear and this model had the most randomly scattered residual plot.

Answers to Odd-Numbered Section 12.2 Exercises

12.31 (a) The scatterplot shows a fairly strong, positive, slightly curved association between length and period with one very unusual point (106.5, 2.115) in the top right corner.



(b) The class used the square root of x = length. (c) The class used the square of y = period.

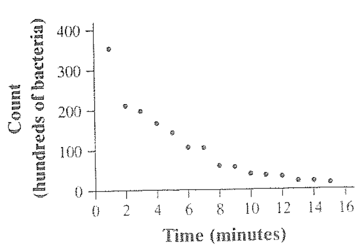
12.33 (a) 1: $\hat{y} = -0.08594 + 0.21\sqrt{x}$, where y is the period and x is the length. 2: $\widehat{y^2} = -0.15465 + 0.0428x$, where y is the period and x is the length. (b) 1: $\hat{y} = -0.08594 + 0.21\sqrt{80} = 1.792$ seconds. 2: $\widehat{y^2} = -0.15465 + 0.0428(80) = 3.269$, so $\hat{y} = \sqrt{3.269} = 1.808$ seconds.

12.35 (a) The scatterplot of log(period) versus log(length) is roughly linear and the residual plot shows no obvious leftover patterns. (b) $\widehat{\log y} = -0.73675 + 0.51701\log(x)$, where y is the period and x is the length.

12.37 $\widehat{\log y} = -0.73675 + 0.51701\log(80) = 0.24717$. Thus, $\hat{y} = 10^{0.24717} = 1.77$ seconds.

12.39 $\widehat{\log y} = 1.01 + 0.72\log(127) = 2.525$. Thus, $\hat{y} = 10^{2.525} = 334.97$ grams.

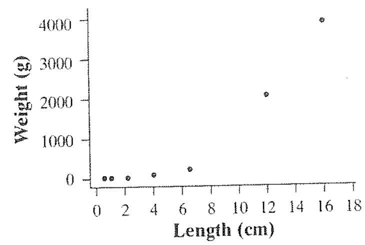
12.41 (a) The relationship between bacteria count and time is strong, negative, and curved with a possible outlier in the top left-hand corner.



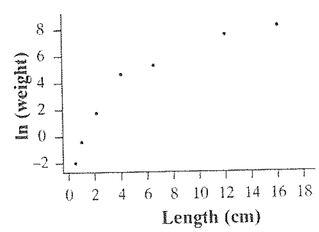
(b) Because the scatterplot of ln(count) versus time is fairly linear. (c) $\ln y = 5.97316 - 0.218425x$, where y is the count of surviving bacteria and x is time in minutes. (d) $\widehat{\ln y} = 5.97316 - 0.218425(17) = 2.26$, so $\hat{y} = e^{2.26} = 9.58$ or 958 bacteria.

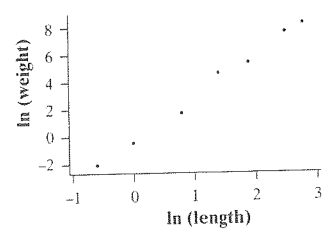
12.43 (a) Exponential, because the scatterplot of log(height) versus bounce number is more linear. (b) $\widehat{\log y} = 0.45374 - 0.11716x$, where y = height in feet and x = bounce number. (c) $\widehat{\log y} = 0.45374 - 0.11716(7) = -0.36638$, so $\hat{y} = 10^{-0.36638} = 0.43$ feet. (d) The trend in the residual plot suggests that the residual for $x = 7$ would be positive, meaning that the predicted height will be less than the actual height.

12.45 (a) There is a strong, positive curved relationship between heart weight and length of left ventricle for mammals.



(b) Two scatterplots are given below. Because the relationship between ln(weight) and ln(length) is roughly linear, heart weight and length seem to follow a power model.





(c) $\widehat{\ln y} = -0.314 + 3.1387\ln x$, where y is the weight of the heart and x is the length of the cavity of the left ventricle. (d) $\widehat{\ln y} = -0.314 + 3.1387\ln(6.8) = 5.703$, so $\hat{y} = e^{5.703} = 299.77$ grams.

12.47 c

12.49 e

12.51 (a) For Marcela, X = the length of her shower on a randomly selected day follows a Normal distribution with mean 4.5 minutes and standard deviation 0.9 minutes. We want to find $P(3 < X < 6)$. $z = \dfrac{3 - 4.5}{0.9} = -1.67$ and $z = \dfrac{6 - 4.5}{0.9} = 1.67$, so $P(3 < X < 6) = 0.9050$. *Using technology:* 0.9044. There is a 0.9044 probability that Marcela's shower lasts between 3 and 6 minutes. (b) Solving $-0.67 = \dfrac{Q_1 - 4.5}{0.9}$ gives $Q_1 = 3.897$ minutes. Solving $0.67 = \dfrac{Q_3 - 4.5}{0.9}$ gives $Q_3 = 5.103$ minutes. *Using technology:* $Q_1 = 3.893$ minutes and $Q_3 = 5.107$ minutes. Thus, an outlier is any value above $5.107 + 1.5(5.107 - 3.893) = 6.928$. Because $7 > 6.928$, a shower of 7 minutes would be considered an outlier for Marcela. (c) $P(X > 7) = 0.0027$. Let Y = the number of days that Marcela's shower is 7 minutes or higher. Y is a binomial random variable with $n = 10$ and $p = 0.0027$. $P(Y \geq 2) = 1 - P(Y \leq 1) =$ `1 − binomcdf(trials: 10, p: 0.0027, x value: 1)` $= 0.0003$. (d) $\bar{x}$ follows a $N(4.5, 0.285)$ distribution and we want to find $P(\bar{x} > 5)$. $z = \dfrac{5 - 4.5}{0.285} = 1.75$ and $P(Z > 1.75) =$. *Using technology:* 0.0397. There is a 0.0397 probability that the mean length of Marcela's showers on these 10 days exceeds 5 minutes.

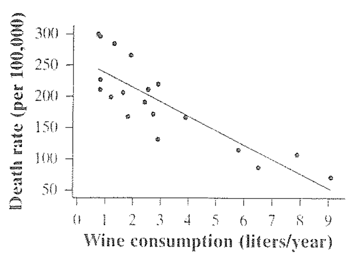
and computed a confidence interval for the slope each time, about 99% of the resulting intervals would contain the slope of the true regression line for predicting BAC from the number of beers consumed.

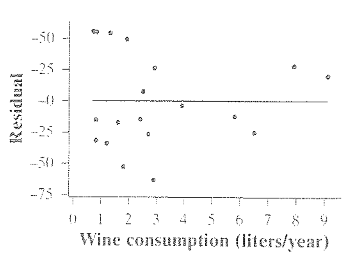
12.9 S: β = the slope of the population regression line relating number of clusters of beetle larvae to number of stumps. P: t interval for β. D: With df = 21, (8.678, 15.11). C: We are 99% confident that the interval from 8.678 to 15.11 captures the slope of the population regression line relating number of clusters of beetle larvae to number of stumps.

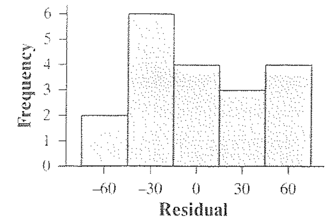
12.11 (a) $\hat{y} = -1.286 + 11.894(5) = 58.184$ clusters. (b) $s = 6.419$, so we would expect our prediction to be off from the actual number of clusters by about 6.419 clusters.

12.13 (a) $\hat{y} = 166.483 - 1.0987x$, where $\hat{y}$ is the predicted corn yield and x is the number of weeds per meter. Slope: for each additional weed per meter, the predicted corn yield will decrease by about 1.0987 bushels/acre. y intercept: if there are no weeds per meter, we would predict a corn yield of 166.483 bushels/acre. (b) When using weeds per meter to predict corn yield, the actual yield will typically vary from the predicted yield by about 7.98 bushels/acre. (c) S: H_0: $\beta = 0$ versus H_a: $\beta < 0$, where β is the slope of the true regression line relating corn yield to weeds per meter. P: t test for β. D: $t = -1.92$. P-value = 0.075/2 = 0.0375. C: Because the P-value of 0.0375 is less than $\alpha = 0.05$, we reject H_0. There is convincing evidence that the slope of the true regression line relating corn yield to weeds per meter is negative.

12.15 S: H_0: $\beta = 0$ versus H_a: $\beta < 0$, where β is the slope of the population regression line relating heart disease death rate to wine consumption in the population of countries. P: t test for β. Linear: There is no leftover pattern in the residual plot. Independent: The sample size ($n = 19$) is less than 10% of all countries. Normal: The histogram of residuals shows no strong skewness or outliers. Equal SD: The residual plot shows that the standard deviation of the death rates might be a little smaller for large values of wine consumption, x, but it is hard to tell with so few data values. Random: Random sample.







D: $t = -6.46$, df = 17, and P-value ≈ 0. C: Because the P-value of approximately 0 is less than $\alpha = 0.05$, we reject H_0. There is convincing evidence of a negative linear relationship between wine consumption and heart disease death rate in the population of countries.

12.17 (a) With df = 19, 11,630.6 ± 2.093(1249) = (9016.4, 14,244.8). (b) Because the automotive group claims that people drive 15,000 miles per year, this says that for every increase of 1 year, the mileage would increase by 15,000 miles. (c) $t = -2.70$. With df = 19, the P-value is between 0.01 and 0.02 (0.0142). Because the P-value of 0.0142 is less than $\alpha = 0.05$, we reject H_0. We have convincing evidence that the slope of the population regression line relating miles to years is not equal to 15,000. (d) Yes. Because the interval in part (a) does not include the value 15,000, the interval also provides convincing evidence that the slope of the population regression line relating miles to years is not equal to 15,000.

12.19 c

12.21 a

12.23 b

12.25 (a) The two treatments (say the color, read the word) were deliberately assigned to the students. (b) He used a randomized block design where each student was a block. He did this to help account for the different abilities of students to read the words or to say the color they were printed in. (c) To help average out the effects of the order in which people did the two treatments. If every subject said the color of the printed word first and were frustrated by this task, the times for the second treatment might be worse. Then we wouldn't know the reason the times were longer for the second treatment—because of frustration or because the second method actually takes longer.

12.27 There is a small number of differences ($n_d = 16 < 30$) and there is an outlier.

12.29 (a) (i) $\dfrac{295}{1526} = 0.1933$. (ii) $\dfrac{295 + 77 + 212}{1526} = 0.3827$. (iii) $\dfrac{212}{305} = 0.6951$. (b) No. The probability that a person is a snowmobile owner (295/1526 = 0.1933) is different from the probability that the person is a snowmobile owner given that he or she belongs to an environmental organization (16/305 = 0.0525).

(c) (i) $P(\text{both are owners}) = \left(\dfrac{295}{1526}\right)\left(\dfrac{294}{1525}\right) = 0.0373$

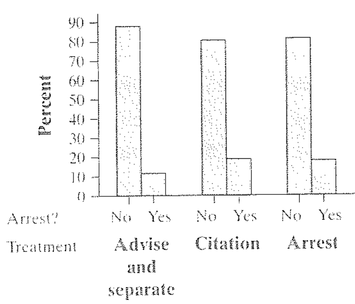
(ii) P(at least one belongs to an environmental organization)

$= 1 - P(\text{neither belong}) = 1 - \left(\dfrac{1221}{1526}\right)\left(\dfrac{1220}{1525}\right) = 0.3599$

Both groups of children have the largest percentage reporting grades as the goal. But after that, boys were more likely to pick sports, whereas girls were more likely to pick being popular.
(b) *S*: H_0: There is no association between gender and goals for 4th, 5th, and 6th grade students versus H_a: There is an association. . . . *P*: Chi-square test for independence. Random: Random sample. 10% : $n = 478 < 10\%$ of all 4th, 5th, and 6th grade students. Large Counts: 129.70, 117.30, 74.04, 66.96, 47.26, 42.74 all ≥ 5. *D*: $\chi^2 = 21.455$. With df = 2, the *P*-value < 0.0005 (0.00002). *C*: Because the *P*-value of 0.00002 < $\alpha = 0.05$, we reject H_0. There is convincing evidence that an association exists between gender and goals for 4th, 5th, and 6th grade students.

Answers to Chapter 11 AP® Statistics Practice Test

T11.1 b
T11.2 c
T11.3 e
T11.4 d
T11.5 c
T11.6 c
T11.7 b
T11.8 a
T11.9 d
T11.10 d
T11.11 *S*: H_0: The distribution of gas types is the same as the distributor's claim versus H_a: The distribution of gas types is not the same as the distributor's claim. *P*: Chi-square test for goodness of fit. Random: Random sample. 10% : $n = 400 < 10\%$ of all customers at this distributor's service stations. Large Counts: 240, 80, 80 all ≥ 5. *D*: $\chi^2 = 13.15$. With df = 2, the *P*-value is between 0.001 and 0.0025 (0.0014). *C*: Because the *P*-value of 0.0014 < $\alpha = 0.05$, we reject H_0. There is convincing evidence that the distribution of gas type is not the same as the distributor claims.
T11.12 (a) Random assignment was used to create three roughly equivalent groups at the beginning of the study.
(b)



(c) H_0: The true proportion of spouse abusers like the ones in the study who will be arrested within 6 months is the same for all three police responses versus H_a: The true proportions are not all the same. (d) *P-value*: If the true proportion of spouse abusers like the ones in the study who will be arrested within 6 months is the same for all three police responses, there is a 0.0796 probability of getting differences between the three groups as large as or larger than the ones observed by chance alone. *Conclusion*: Because the *P*-value of 0.0796 is larger than $\alpha = 0.05$, we fail to reject H_0. There is not convincing evidence that true proportion of spouse abusers like the ones in the study who will be arrested within 6 months is not the same for all three police responses.

T11.13 (a) *S*: H_0: There is no association between smoking status and educational level among French men aged 20 to 60 years versus H_a: There is an association. . . . *P*: Chi-square test for independence. Random: Random sample. 10% : $n = 459 < 10\%$ of all French men aged 20 to 60 years. Large Counts: 59.48, 44.21, 42.31, 50.93, 37.85, 36.22, 42.37, 31.49, 30.14, 34.22, 25.44, 24.34 all ≥ 5. *D*: $\chi^2 = 13.305$. With df = 6, the *P*-value is between 0.025 and 0.05 (0.0384). *C*: Because the *P*-value of 0.0384 < $\alpha = 0.05$, we reject H_0. There is convincing evidence of an association between smoking status and educational level among French men aged 20 to 60 years.

Chapter 12

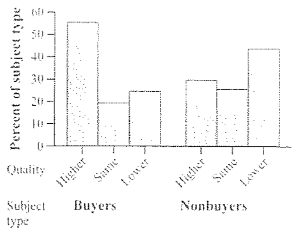
Section 12.1

Answers to Check Your Understanding

page 752: S: β = slope of the population regression line relating fat gain to change in NEA. *P*: *t* interval for the slope. Linear: There is no leftover pattern in the residual plot. Independent: The sample size ($n = 16$) is less than 10% of all healthy young adults. Normal: The histogram of the residuals shows no strong skewness or outliers. Equal SD: Other than one point with a large positive residual, the residual plot shows roughly equal scatter for all *x* values. Random: Random sample. *D*: With df = 14, (−0.005032, −0.001852). *C*: We are 95% confident that the interval from −0.005032 to −0.001852 captures the slope of the population regression line relating fat gain to change in NEA.
page 757: S: H_0: $\beta = 0$ versus H_a: $\beta < 0$, where β is the slope of the true regression line relating fat gain to NEA change. *P*: *t* test for the slope β. *D*: $t = -4.64$. *P*-value $\approx 0.000/2 \approx 0$. *C*: Because the *P*-value of approximately 0 is less than $\alpha = 0.05$, we reject H_0. There is convincing evidence that the slope of the true regression line relating fat gain to NEA change is negative.

Answers to Odd-Numbered Section 12.1 Exercises

12.1 The Equal SD condition is not met because the SD of the residuals clearly increases as the laboratory measurement (*x*) increases.
12.3 Linear: There is no leftover pattern in the residual plot. Independent: Knowing the BAC for one subject should not help us predict the BAC for another subject. Normal: The histogram of the residuals shows no strong skewness or outliers. Equal SD: The residual plot shows roughly equal scatter for all *x* values. Random: These data come from a randomized experiment.
12.5 α is the true *y* intercept, which measures the true mean BAC level if no beers had been drunk ($a = -0.012701$). β is the true slope, which measures how much the true mean BAC changes with the drinking of one additional beer ($b = 0.018$). Finally, σ is the true standard deviation of the residuals, which measures how much the observed values of BAC typically vary from the population regression line ($s = 0.0204$).
12.7 (a) $SE_b = 0.0024$. If we repeated the experiment many times, the slope of the sample regression line would typically vary by about 0.0024 from the slope of the true regression line for predicting BAC from the number of beers consumed. (b) With df = 14, $0.018 \pm 2.977(0.0024) = (0.011, 0.025)$. (c) We are 99% confident that the interval from 0.011 to 0.025 captures the slope of the true regression line for predicting BAC from the number of beers consumed. (d) If we repeated the experiment many times



11.43 (a) H_0: There is no association between beliefs about the quality of recycled products and whether or not a person buys recycled products in the population of adults versus H_a: There is an association. . . . (b)13.26, 35.74, 8.66, 23.34, 14.08, 37.92 (c) $\chi^2 = 7.64$. With df = 2, the *P*-value is between 0.02 and 0.025 (0.022). (d) Because the *P*-value of $0.022 < \alpha = 0.05$, we reject H_0. There is convincing evidence of an association between beliefs about the quality of recycled products and whether or not a person buys recycled products in the population of adults.

11.45 *S*: H_0: There is no association between education level and opinion about a handgun ban in the adult population versus H_a: There is an association. . . . *P*: Chi-square test for independence. Random: Random sample. 10%: $n = 1201 < 10\%$ of all adults. Large Counts: 46.94, 86.19, 187.36, 94.29, 71.22, 69.06, 126.81, 275.64, 138.71, 104.78 all ≥ 5. *D*: $\chi^2 = 8.525$. With df = 4, the *P*-value is between 0.05 and 0.10 (0.0741). *C*: Because the *P*-value of $0.0741 > \alpha = 0.05$, we fail to reject H_0. We do not have convincing evidence that there is an association between educational level and opinion about a handgun ban in the adult population.

11.47 (a) Independence, because the data come from a single random sample. (b) H_0: There is no association between gender and where people live in the population of young adults versus H_a: There is an association. . . . (c) Random: Random sample. $10\% : n = 4854 < 10\%$ of all young adults. Large Counts: The expected counts are all at least 5. (d) *P-value*: If no association exists between gender and where people live in the population of young adults, there is a 0.012 probability of getting a random sample of 4854 young adults with an association as strong or even stronger than the one found in this study. *Conclusion*: Because the *P*-value of $0.012 < \alpha = 0.05$, we reject H_0. There is convincing evidence that an association exists between gender and where people live in the population of young adults.

11.49 (a) *Hypotheses*: H_0: There is no difference in the improvement rates for patients like these who receive gastric freezing and those who receive the placebo versus H_a: There is a difference. . . . *P-value*: Assuming that no difference exists in the improvement rates between those receiving gastric freezing and those receiving the placebo, there is a 0.570 probability of observing a difference in improvement rates as large or larger than the difference observed in the study by chance alone. *Conclusion*: Because the *P*-value of 0.570 is larger than $\alpha = 0.05$, we fail to reject H_0. There is not convincing evidence that a difference exists in the improvement rates for patients like these who receive gastric freezing and those who receive the placebo. (b) The *P*-values are equal and $z^2 = (-0.57)^2 = 0.3249 \approx \chi^2 = 0.322$.

11.51 d

11.53 d

11.55 a

11.57 (a) One-sample *t* interval for a mean. (b) Two-sample *z* test for the difference between two proportions.

11.59 (a) Experiment, because a treatment (type of rating scale) was deliberately imposed on the students who took part in the study. (b) Several of the expected counts are less than 5.

Answers to Chapter 11 Review Exercises

R11.1 *S*: H_0: The proposed 1:2:1 genetic model is correct versus H_a: The proposed 1:2:1 genetic model is not correct. *P*: Chi-square test for goodness of fit. Random: Random sample. $10\% : n = 84 < 10\%$ of all yellow-green parent plants. Large Counts: 21, 42, 21 all ≥ 5. *D*: $\chi^2 = 6.476$. Using df = 2, the *P*-value is between 0.025 and 0.05 (0.0392). *C*: Because the *P*-value of $0.0392 > \alpha = 0.01$, we fail to reject H_0. We do not have convincing evidence that the proposed 1:2:1 genetic model is not correct.

R11.2 Several of the expected counts are less than 5.

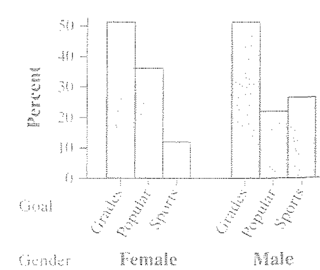
R11.3 (a)

	Stress management	Exercise	Usual care	Total
Suffered cardiac event	3	7	12	**22**
No cardiac event	30	27	28	**85**
Total	33	34	40	**107**

(b) The success rate was highest for stress management (0.909), followed by exercise (0.794) and usual care (0.70). (c) *S*: H_0: The true success rates for patients like these are the same for all three treatments versus H_a: The true success rates are not all the same. . . . *P*: Chi-square test for homogeneity. Random: 3 groups in a randomized experiment. Large Counts: 6.79, 6.99, 8.22, 26.21, 27.01, 31.78 all ≥ 5. *D*: $\chi^2 = 4.840$. With df = 2, the *P*-value is between 0.05 and 0.10 (0.0889). *C*: Because the *P*-value $> \alpha = 0.05$, we fail to reject H_0. We do not have convincing evidence that the true success rates for patients like these are not the same for all three treatments.

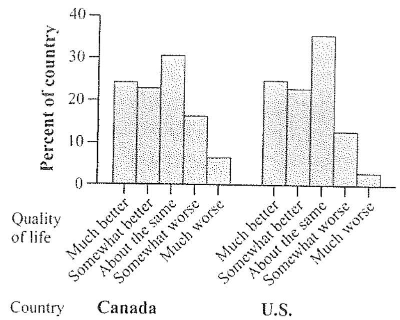
R11.4 (a) The data could have been collected from 3 independent random samples—a random sample of ads from magazines aimed at young men, a random sample of ads from magazines aimed at young women, and a random sample of ads aimed at young adults in general. In each sample, the ads would be classified as sexual or not sexual. (b) The data could have been collected from a single random sample of ads from magazines aimed at young adults. Then each ad in the sample would be classified as sexual or not sexual, and the magazine that the ad was from would be classified as aimed at young men, young women, or young adults in general. (c) $\frac{351}{576} = 0.6094 = 60.94\%$. $\frac{(1113)(576)}{1509} = 424.8$. $\frac{(351-424.8)^2}{424.8} = 12.82$. (The difference is due to rounding error.) (d) The "sexual, Women" cell. There were 225 observed ads in this cell, which was 73.8 more than expected.

R11.5 (a)



use among students at the main campus and the commonwealth campuses as large or larger than the one found in this study. 6. Because the *P*-value of $0.000059 < \alpha = 0.05$, we reject H_0. There is convincing evidence that the distribution of Facebook use is different among students at Penn State's main campus and students at Penn State's commonwealth campuses.

page 711: 1.

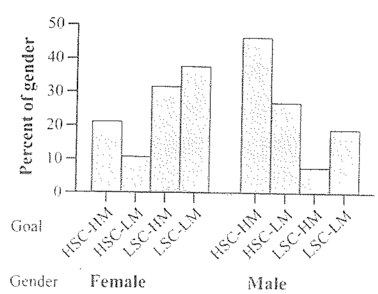


2. *S*: H_0: There is no difference in the distribution of quality of life for patients who have suffered a heart attack in Canada and the U.S. versus H_a: There is a difference. . . . *P*: Chi-square test for homogeneity. Random: Independent random samples. 10%: $n_1 = 311 < 10$ % of all Canadian heart attack patients and $n_2 = 2165 < 10\%$ of all U.S. heart attack patients. Large Counts: 77.37, 538.63, 71.47, 497.53, 109.91, 765.09, 41.70, 290.30, 10.55, 73.45 all ≥ 5. *D*: $\chi^2 = 11.725$. With df = 4, the *P*-value is between 0.01 and 0.02 (0.0195). *C*: Because the *P*-value of $0.0195 > \alpha = 0.01$, we fail to reject H_0. There is not convincing evidence that a difference exists in the distribution of quality of life for heart attack patients in Canada and the United States.

page 717: *S*: H_0: There is no association between an exclusive territory clause and business survival versus H_a: There is an association. . . . *P*: Chi-square test for independence. Random: Random sample. 10%: We assume that $n = 170 < 10\%$ of all new franchise firms. Large Counts: 102.74, 20.26, 39.26, 7.74 all ≥ 5. *D*: $\chi^2 = 5.911$. Using df = 1, the *P*-value is between 0.01 and 0.02 (0.0150). *C*: Because the *P*-value of $0.0150 > \alpha = 0.01$, we fail to reject H_0. There is not convincing evidence of an association between exclusive territory clause and business survival.

Answers to Odd-Numbered Section 11.2 Exercises

11.27 (a) Female: 0.209, 0.104, 0.313, 0.373. Male: 0.463, 0.269, 0.075, 0.194.
(b)



(c) In general, it appears that females were classified mostly as low social comparison, whereas males were classified mostly as high social comparison. However, about an equal percentage of males and females were classified as high mastery.

11.29 (a) H_0: There is no difference in the distribution of sports goals for male and female undergraduates at this university versus H_a: There is a difference. . . . (b) 22.5, 12.5, 13, 19, 22.5, 12.5, 13, 19. (c) $\chi^2 = 24.898$.

11.31 (a) Random: Independent random samples. 10%: $n_1 = 67 < 10\%$ of all males and $n_2 = 67 < 10\%$ of all females at the university. Large Counts: All expected counts ≥ 5. (b) With df = 3, the *P*-value < 0.0005 (0.000016). (c) Assuming that no difference exists in the distributions of goals for playing sports among males and females, there is a 0.000016 probability of observing independent random samples that show a difference in the distributions of goals for playing sports among males and females as large or larger than the one found in this study. (d) Because the *P*-value of $0.000016 < \alpha = 0.05$, we reject H_0. There is convincing evidence of a difference in the distribution of goals for playing sports among male and female undergraduates at this university.

11.33 (a) Cold: 0.593 hatched. Neutral: 0.679 hatched. Hot: 0.721 hatched. As the temperature warms up from cold to neutral to hot, the proportion of eggs that hatch appears to increase. (b) *S*: H_0: There is no difference in the true proportion of eggs that hatch in cold, neutral, or hot water versus H_a: There is a difference. . . . *P*: Chi-square test for homogeneity. Random: 3 groups in a randomized experiment. Large Counts: 18.63, 38.63, 71.74, 8.37, 17.37, 32.26 all ≥ 5. *D*: $\chi^2 = 1.703$. With df = 2, the *P*-value > 0.25 (0.4267). *C*: Because the *P*-value of $0.4267 > \alpha = 0.05$, we fail to reject H_0. We do not have convincing evidence that there is a difference in the true proportions of eggs that hatch in cold, neutral, or hot water.

11.35 We do not have the actual counts of the travelers in each category. We also do not know if the sample was taken randomly or if the samples are independent.

11.37 (a) The data are given in the table below. The best success rate is for the patch plus the drug (0.355), followed by the drug alone (0.303). The patch alone (0.164) is just a little better than the placebo (0.156).

	Nicotine Patch	Drug	Patch plus drug	Placebo	Total
Success	40	74	87	25	226
Failure	204	170	158	135	667
Total	244	244	245	160	893

(b) Each of the four treatments has the same probability of success for smokers like these. (c) *S*: H_0: The true proportions of smokers like these who are able to quit for a year are the same for each of the four treatments versus H_a: The true proportions are not the same. . . . *P*: Chi-square test for homogeneity. Random: 4 groups in a randomized experiment. Large Counts: 61.75, 61.75, 62, 40.49, 182.25, 182.25, 183, 119.51 all ≥ 5. *D*: $\chi^2 = 34.937$. With df = 3, the *P*-value < 0.0005. *C*: Because the *P*-value of approximately $0 < \alpha = 0.05$, we reject H_0. There is convincing evidence that the true proportions of smokers like these who are able to quit for a year are not the same for each of the four treatments.

11.39 The largest component comes from those who had success using both the patch and the drug (25 more than expected). The next largest component comes from those who had success using just the patch (21.75 less than expected).

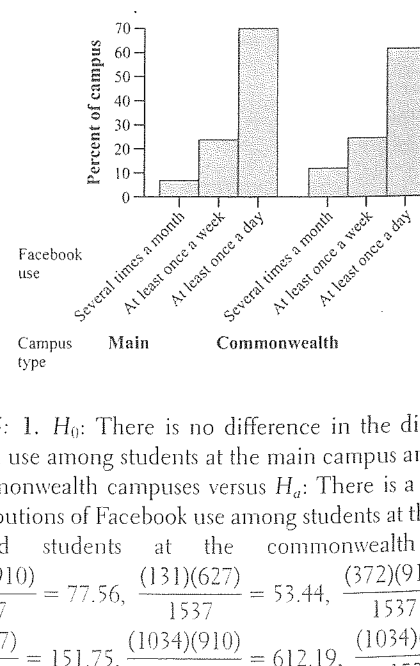
11.41 Buyers are much more likely to think the quality of recycled coffee filters is higher, while nonbuyers are more likely to think the quality is the same or lower.

11.7 *S*: H_0: Nuthatches do not prefer particular types of trees when searching for seeds and insects versus H_a: Nuthatches do prefer particular types of trees when searching for seeds and insects. *P*: Chi-square test for goodness of fit. Random: Random sample. 10%: $n = 156 < 10\%$ of all nuthatches. Large Counts: 84.24, 62.4, 9.36 all ≥ 5. *D*: $\chi^2 = 7.418$. With df = 2, the *P*-value is between 0.02 and 0.025 (0.0245). *C*: Because the *P*-value of $0.0245 < \alpha = 0.05$, we reject H_0. There is convincing evidence that nuthatches prefer particular types of trees when they are searching for seeds and insects.
11.9 Time spent doing homework is quantitative. Chi-square tests for goodness of fit should be used only for distributions of categorical data.
11.11 (a) *S*: H_0: The first digit of invoices from this company follow Benford's law versus H_a: The first digit of invoices from this company do not follow Benford's law. *P*: Chi-square test for goodness of fit. Random: Random sample. 10%: Assume $n = 250 < 10\%$ of all invoices from this company. Large Counts: 75.25, 44, 31.25, 24.25, 19.75, 16.75, 14.5, 12.75, 11.5 all ≥ 5. *D*: $\chi^2 = 21.563$. With df = 8, the *P*-value is between 0.005 and 0.01 (0.0058). *C*: Because the *P*-value of $0.0058 < \alpha = 0.05$, we reject H_0. There is convincing evidence that the first digit of invoices from this company do not follow Benford's law. *Follow-up analysis*: The largest contributors to the statistic are amounts with first digit 3, 4 and 7. There are more invoices that start with 3 or 4 than expected and fewer invoices that start with 7 than expected. (b) I: Finding convincing evidence that the company's invoices do not follow Benford's law (suggesting fraud), when in reality they are consistent with Benford's law. A consequence is falsely accusing this company of fraud. II: Not finding convincing evidence that the invoices do not follow Benford's law (suggesting fraud), when in reality they do not. A consequence is allowing this company to continue committing fraud. A Type I error would be more serious for the accountant.
11.13 (a) H_0: The true distribution of flavors for Skittles candies is the same as the company's claim versus H_a: The true distribution of flavors for Skittles candies is not the same as the company's claim. (b) Expected counts all = 12. (c) Using df = 4, χ^2 statistics greater than 9.49 would provide significant evidence at the $\alpha = 0.05$ level and χ^2 values greater than 13.28 would provide significant evidence at the $\alpha = 0.01$ level. (d) Answers will vary.
11.15 *S*: H_0: All 12 astrological signs are equally likely versus H_a: All 12 astrological signs are not equally likely. *P*: Chi-square test for goodness of fit. Random: Random sample. 10%: $n = 4344 < 10\%$ of all people in the United States. Large Counts: All expected counts = 362, which are ≥ 5. *D*: $\chi^2 = 19.76$. With df = 11, the *P*-value is between 0.025 and 0.05 (0.0487). *C*: Because the *P*-value of $0.0487 < \alpha = 0.05$, we reject H_0. There is convincing evidence that the 12 astrological signs are not equally likely. *Follow-up analysis*: The largest contributors to the statistic are Aries and Virgo. There are fewer Aries $(321 - 362 = -41)$ and more Virgos $(402 - 362 = 40)$ than we would expect.
11.17 *S*: H_0: Mendel's 3:1 genetic model is correct versus H_a: Mendel's 3:1 genetic model is not correct. *P*: Chi-square test for goodness of fit. Conditions are met. *D*: $\chi^2 = 0.3453$. With df = 1, the *P*-value > 0.25 (0.5568). *C*: Because the *P*-value of $0.5568 > \alpha = 0.05$, we fail to reject H_0. We do not have convincing evidence that Mendel's 3:1 genetic model is wrong.
11.19 d
11.21 c
11.23 The distribution of English grades for the heavy readers is skewed to the left, while the distribution of English grades for the light readers is roughly symmetric. The center of the distribution of English grades is greater for the heavy readers than for the light readers. The English grades are more variable for the light readers. There is one low outlier in the heavy reading group but no outliers in the light reading group.
11.25 (a) For each additional book read, the predicted English GPA increases by about 0.024. The predicted English grade for a student who has read 0 books is about 3.42. (b) residual $= 2.85 - 3.828 = -0.978$. This student's English GPA is 0.978 less than predicted, based on the number of books this student has read. (c) Not very strong. On the scatterplot, the points are quite spread out from the line. Also, the value of r^2 is 0.083, which means that only 8.3% of the variation in English grades is accounted for by the linear model relating English GPA to number of books read.

Section 11.2

Answers to Check Your Understanding

page 699: 1. Main: 0.060 several times a month or less, 0.236 at least once a week, 0.703 at least once a day. Commonwealth: 0.121 several times a month or less, 0.250 at least once a week, 0.628 at least once a day. 2. Because there was such a big difference in the sample size from the two different types of campuses. 3. Students on the main campus are more likely to be everyday users of Facebook. Also, those on the commonwealth campuses are more likely to use Facebook several times a month or less.



page 705: 1. H_0: There is no difference in the distributions of Facebook use among students at the main campus and students at the commonwealth campuses versus H_a: There is a difference in the distributions of Facebook use among students at the main campus and students at the commonwealth campuses.

2. $\frac{(131)(910)}{1537} = 77.56$, $\frac{(131)(627)}{1537} = 53.44$, $\frac{(372)(910)}{1537} = 220.25$, $\frac{(372)(627)}{1537} = 151.75$, $\frac{(1034)(910)}{1537} = 612.19$, $\frac{(1034)(627)}{1537} = 421.81$

3. $\chi^2 = \frac{(55 - 77.56)^2}{77.56} + \cdots + \frac{(394 - 421.81)^2}{421.81} = 19.49$

4. With df = 2, the *P*-value < 0.0005 (0.000059). 5. Assuming that no difference exists in the distributions of Facebook use between students on Penn State's main campus and students at Penn State's commonwealth campuses, there is a 0.000059 probability of observing samples that show a difference in the distributions of Facebook

age 20 is less than the true proportion of non-VLBW babies who graduate from high school by age 20.

AP3.33 (a) $\hat{y} = -73.64 + 5.7188x$, where $\hat{y}$ = predicted distance and x = temperature (degrees Celsius) (b) For each increase of 1°C in the water discharge temperature, the predicted distance from the nearest fish to the outflow pipe increases by about 5.7188 meters. (c) Yes. The residual plot shows no leftover pattern. (d) residual = 78 − 92.21 = −14.21 meters. The actual distance on this afternoon was 14.21 meters closer than expected, based on the temperature of the water.

AP3.34 (a) Define W = the weight of a randomly selected gift box. Then $\mu_W = 8(2) + 2(4) + 3 = 27$ ounces and $\sigma_W = \sqrt{8(0.5^2) + 2(1^2) + 0.2^2} = 2.01$ ounces. (b) We want to find $P(W > 30)$ using the $N(27, 2.01)$ distribution. $z = \frac{30 - 27}{2.01} = 1.49$ and $P(W > 30) = 0.0681$. *Using technology*: 0.0678. There is a 0.0678 probability of randomly selecting a box that weighs more than 30 ounces. (c) P(at least one box is greater than 30 ounces) $= 1 - P$(none of the boxes is greater than 30 ounces) $= 1 - (1 - 0.0678)^5 = 1 - (0.9322)^5 = 0.2960$. (d) Because the distribution of W is Normal, the distribution of $\overline{W}$ will also be Normal, with mean $\mu_{\overline{W}} = 27$ ounces and standard deviation $\sigma_{\overline{W}} = \frac{2.01}{\sqrt{5}} = 0.899$. We want to find $P(\overline{W} > 30)$. $z = \frac{30 - 27}{0.899} = 3.34$ and $P(Z > 3.34) = 0.0004$. There is a 0.0004 probability of randomly selecting 5 boxes that have a mean weight of more than 30 ounces.

AP3.35 (a) S: $H_0: \mu_A - \mu_B = 0$ versus $H_a: \mu_A - \mu_B \neq 0$, where μ_A is the true mean annualized return for stock A and μ_B is the true mean annualized return for stock B. P: Two-sample t test. Random: Independent random samples. 10%: $n_A = 50$ is less than 10% of all days in the past 5 years and $n_B = 50$ is less than 10% of all days in the past 5 years. Normal/Large Sample: $n_A = 50 \geq 30$ and $n_B = 50 \geq 30$. D: $t = 2.07$. Using df = 40, the P-value is between 0.04 and 0.05. Using df = 90.53, P-value = 0.0416. C: Because the P-value of 0.0416 is less than $\alpha = 0.05$, we reject H_0. We have convincing evidence that the true mean annualized return for stock A is different than the true mean annualized return for stock B. (b) $H_0: \sigma_A - \sigma_B = 0$ vs. $H_a: \sigma_A - \sigma_B > 0$, where σ_A is the true standard deviation of returns for stock A and σ_B is the true standard deviation of returns for stock B. (c) When the standard deviation of stock A is greater than the standard deviation of stock B, the variance of stock A will be bigger than the variance of stock B. Thus, values of F that are significantly greater than 1 would indicate that the price volatility for stock A is higher than that for stock B. (d) $F = \frac{(12.9)^2}{(9.6)^2} = 1.806$. (e) In the simulation, a test statistic of 1.806 or greater occurred in only 6 out of the 200 trials. Thus, the approximate P-value is 6/200 = 0.03. Because the approximate P-value of 0.03 is less than $\alpha = 0.05$, we reject H_0. There is convincing evidence that the true standard deviation of returns for stock A is greater than the true standard deviation of returns for stock B.

Chapter 11

Section 11.1

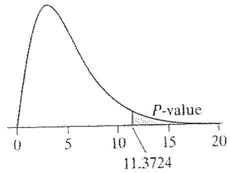
Answers to Check Your Understanding

page 684: 1. H_0: The company's claimed color distribution for its Peanut M&M'S is correct versus H_a: The company's claimed color distribution is not correct. 2. The expected count of both blue and orange candies is 46(0.23) = 10.58, for green and yellow is 46(0.15) = 6.9, and for red and brown is 46(0.12) = 5.52.

3. $$\chi^2 = \frac{(12 - 10.58)^2}{10.58} + \frac{(7 - 10.58)^2}{10.58} + \frac{(13 - 6.9)^2}{6.9} + \frac{(4 - 6.9)^2}{6.9} + \frac{(8 - 5.52)^2}{5.52} + \frac{(2 - 5.52)^2}{5.52} = 11.3724$$

page 687: 1. The expected counts are all at least 5. df = 6 − 1 = 5.

2.



Chi-square distribution with 5 df

3. The P-value is between 0.025 and 0.05 (0.0445). 4. Because the P-value of 0.0445 < α = 0.05, we reject H_0. There is convincing evidence that the color distribution of M&M'S® Peanut Chocolate Candies is different from what the company claims.

page 691: S: H_0: The distribution of eye color and wing shape is the same as what the biologists predict versus H_a: The distribution of eye color and wing shape is not what the biologists predict. P: Chi-square test for goodness of fit. Random: Random sample. 10%: $n = 200 <$ 10% of all fruit flies. Large Counts: 112.5, 37.5, 37.5, 12.5 all ≥ 5. D: $\chi^2 = 6.1867$, df = 3, the P-value is between 0.10 and 0.15 (0.1029). C: Because the P-value of 0.1029 > α = 0.01, we fail to reject H_0. We do not have convincing evidence that the distribution of eye color and wing shape is different from what the biologists predict.

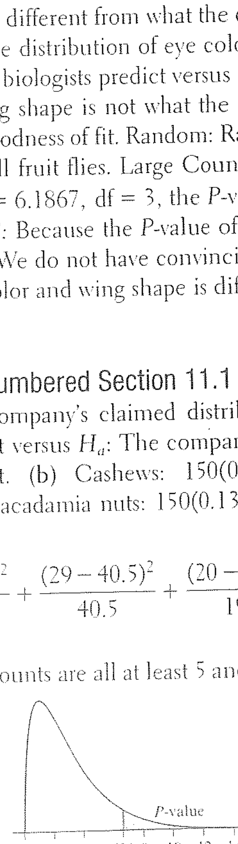
Answers to Odd-Numbered Section 11.1 Exercises

11.1 (a) H_0: The company's claimed distribution for its deluxe mixed nuts is correct versus H_a: The company's claimed distribution is not correct. (b) Cashews: 150(0.52) = 78, almonds: 150(0.27) = 40.5, macadamia nuts: 150(0.13) = 19.5, brazil nuts: 150(0.08) = 12.

11.3 $$\chi^2 = \frac{(83 - 78)^2}{78} + \frac{(29 - 40.5)^2}{40.5} + \frac{(20 - 19.5)^2}{19.5} + \frac{(18 - 12)^2}{12} = 6.599$$

11.5 (a) Expected counts are all at least 5 and df = 3.

(b)



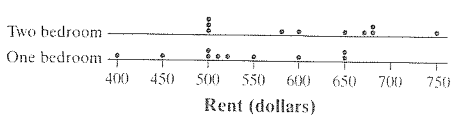
Chi-square distribution with 3 df

(c) The P-value is between 0.05 and 0.10 (0.0858). (d) Because the P-value of 0.0858 > α = 0.05, we fail to reject H_0. We do not have convincing evidence that the company's claimed distribution for its deluxe mixed nuts is incorrect.

Random: Two groups in a randomized experiment. Normal/Large Sample: $n_1 = 104 \geq 30$ and $n_2 = 96 \geq 30$. *D*: Using df = 80, $(-4.17, -1.03)$. Using df = 165.12, $(-4.16, -1.04)$. *C*: We are 95% confident that the interval from -4.16 to -1.04 captures the true difference in mean length of hospital stay for patients like these who get heating blankets during surgery and those who have their core temperatures reduced during surgery. (b) Yes. Because 0 is not in the interval, we have convincing evidence that the true mean hospital stay for patients like these who get heating blankets during surgery is different than the true mean hospital stay for patients like these who have core temperatures reduced during surgery. (c) If we were to repeat this experiment many times and calculate 95% confidence intervals for the difference in means each time, about 95% of the intervals would capture the true difference in mean hospital stay for patients like these who get heating blankets during surgery and mean hospital stay for patients like these who have core temperatures reduced during surgery.

T10.12 (a) *S*: $H_0: p_1 - p_2 = 0$ versus $H_a: p_1 - p_2 > 0$, where p_1 is the true proportion of cars that have the brake defect in last year's model and p_2 is the true proportion of cars that have the brake defect in this year's model. *P*: Two-sample z test for $p_1 - p_2$. Random: Independent random samples. 10%: $n_1 = 100 < 10\%$ of last year's model and $n_2 = 350 < 10\%$ of this year's model. Large Counts: 20, 80, 50, 300 are all ≥ 10. *D*: $z = 1.39$, *P*-value = 0.0822. *C*: Because the *P*-value of $0.0822 > \alpha = 0.05$, we fail to reject H_0. We do not have convincing evidence that the true proportion of brake defects is smaller in this year's model compared to last year's model. (b) I: Finding convincing evidence that there is a smaller proportion of brake defects in this year's car model, when in reality there is not. This might result in more accidents because people think that their brakes are safe. II: Not finding convincing evidence that there is a smaller proportion of brake defects in this year's model, when in reality there is a smaller proportion. This might result in reduced sales of this year's model.

T10.13 (a) $H_0: \mu_1 - \mu_2 = 0$ versus $H_a: \mu_1 - \mu_2 < 0$, where μ_1 = the true mean rent for one-bedroom apartments in the area of her college campus and μ_2 = the true mean rent for two-bedroom apartments in the area of her college campus. (b) Two-sample t test for $\mu_1 - \mu_2$. Random: Independent random samples. 10%: $n_1 = 10 < 10\%$ of all one-bedroom apartments in this area and $n_2 = 10 < 10\%$ of all two-bedroom apartments in this area. Normal/Large Sample: The dotplots below show no strong skewness or outliers in either distribution.

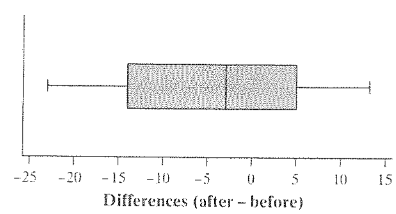


(c) Assuming the true mean rent of the two types of apartments is really the same, there is a 0.029 probability of getting an observed difference in mean rents as large as or larger than the one in this study. (d) Because the *P*-value of $0.029 < \alpha = 0.05$, Pat should reject H_0. She has convincing evidence that the true mean rent of two-bedroom apartments is greater than the true mean rent of one-bedroom apartments in the area of her college campus.

Answers to Cumulative AP® Practice Test 3

AP3.1 e
AP3.2 e
AP3.3 d
AP3.4 c
AP3.5 d
AP3.6 d
AP3.7 c
AP3.8 a
AP3.9 d
AP3.10 c
AP3.11 b
AP3.12 c
AP3.13 c
AP3.14 d
AP3.15 d
AP3.16 e
AP3.17 b
AP3.18 b
AP3.19 e
AP3.20 c
AP3.21 a
AP3.22 d
AP3.23 b
AP3.24 e
AP3.25 a
AP3.26 b
AP3.27 c
AP3.28 d
AP3.29 a
AP3.30 b

AP3.31 *S*: $H_0: \mu_d = 0$ versus $H_a: \mu_d < 0$, where μ_d is the true mean change in weight (after − before) in pounds for people like these who follow a five-week crash diet. *P*: Paired t test for μ_d. Random: Random sample. 10%: $n_d = 15$ is less than 10% of all dieters. Normal/Large Sample: There is no strong skewness or outliers.



D: $\bar{x} = -3.6$ and $s_x = 11.53$. $t = -1.21$. Using df = 14, the *P*-value is between 0.10 and 0.15 (0.1232). *C*: Because the *P*-value of 0.1232 is greater than $\alpha = 0.05$, we fail to reject H_0. We do not have convincing evidence that the true mean change in weight (after − before) for people like these who follow a five-week crash diet is less than 0.

AP3.32 (a) Observational study. No treatments were imposed on the individuals in the study. (b) $H_0: p_1 - p_2 = 0$ versus $H_a: p_1 - p_2 < 0$, where p_1 is the true proportion of VLBW babies who graduate from high school by age 20 and p_2 is the true proportion of non-VLBW babies who graduate from high school by age 20. *P*: Two-sample z test for $p_1 - p_2$. Random: Independent random samples. 10%: $n_1 = 242$ is less than 10% of all VLBW babies and $n_2 = 233$ is less than 10% of all non-VLBW babies. Large Counts: 179, 63, 193, 40 are all ≥ 10. *Do*: $z = -2.34$ and *P*-value = 0.0095. *Conclude*: Because the *P*-value of 0.0095 is less than $\alpha = 0.05$, we reject H_0. We have convincing evidence that the true proportion of VLBW babies who graduate from high school by

10.63 (a) P(at least one mean outside interval) = 1 − P(neither mean outside interval) = $1-(0.95)^2 = 1-0.9025 = 0.0975$. (b) Let X = the number of samples that must be taken to observe one falling above $\mu_{\bar{x}} + 2\sigma_{\bar{x}}$. Then X is a geometric random variable with $p = 0.025$. $P(X = 4) = (1 - 0.025)^3(0.025) = 0.0232$. (c) Let X = the number of sample means out of 5 that fall outside this interval. X is a binomial random variable with $n = 5$ and $p = 0.32$. We want $P(X \geq 4) = 1 - P(X \leq 3)$ = `1 - binomcdf (trials:5, p:0.32,x value:3)` = 1 − 0.961 = 0.039. This is a reasonable criterion because when the process is under control, we would only get a "false alarm" about 4% of the time.

10.65 (a) Perhaps the people who responded are prouder of their improvements and are more willing to share. This could lead to an overestimate of the true mean improvement. (b) This was an observational study, not an experiment. The students (or their parents) chose whether or not to be coached; students who choose coaching might have other motivating factors that help them do better the second time.

Answers to Chapter 10 Review Exercises

R10.1 (a) Paired t test for the mean difference. (b) Two-sample z interval for the difference in proportions. (c) One-sample t interval for the mean. (d) Two-sample t interval for the difference between two means.

R10.2 (a) $SE_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{0.832(1-0.832)}{220} + \frac{0.581(1-0.581)}{117}}$ = 0.0521. If we were to take many random samples of 220 Hispanic female drivers in New York and 117 Hispanic female drivers in Boston, the difference in the sample proportions who wear seatbelts will typically be 0.0521 from the true difference in proportions of all Hispanic female drivers in New York and Boston who wear seat belts. (b) S: p_1 = proportion of all Hispanic female drivers in New York who wear seat belts and p_2 = proportion of all Hispanic female drivers in Boston who wear seat belts. P: Two-sample z interval for $p_1 - p_2$. Random: Independent random samples. 10%: $n_1 = 220 < 10\%$ of all Hispanic female drivers in New York and $n_2 = 117 < 10\%$ of all Hispanic female drivers in Boston. Large Counts: 183, 37, 68, 49 are all ≥ 10. D: (0.149, 0.353). C: We are 95% confident that the interval from 0.149 to 0.353 captures the true difference in the proportions of Hispanic women drivers in New York and Boston who wear their seat belts.

R10.3 (a) The women in the study were randomly assigned to one of the two treatments. (b) Because both groups are large ($n_C = 45 \geq 30$ and $n_A = 45 \geq 30$), the sampling distribution of $\bar{x}_C - \bar{x}_A$ should be approximately Normal. (c) Assuming no difference exists in the true mean ratings of the product for women like these who read or don't read the news story, there is less than a 0.01 probability of observing a difference as large as or larger than 0.49 by chance alone.

R10.4 (a) S: μ_1 = the true mean NAEP quantitative skills test score for young men and μ_2 = the true mean NAEP quantitative skills test score for young women. P: Two-sample t interval for $\mu_1 - \mu_2$. Random: Reasonable to consider these independent random samples. 10%: $n_1 = 840 < 10\%$ of all young men and $n_2 = 1077 < 10\%$ of all young women. Normal/Large Sample: $n_1 = 840 \geq 30$ and $n_2 = 1077 \geq 30$. D: Using df = 100, (−6.80,2.14). Using df = 1777.52, (−6.76,2.10). C: We are 90% confident that the interval from −6.76 to 2.10 captures the true difference in the mean NAEP quantitative skills test score for young men and the mean NAEP quantitative skills test score for young women. (b) Because 0 is in the interval, we do not have convincing evidence of a difference in mean score for male and female young adults.

R10.5 (a) S: $H_0: p_1 - p_2 = 0$ versus $H_a: p_1 - p_2 < 0$, where p_1 is the true proportion of patients like these who take AZT and develop AIDS and p_2 is the true proportion of patients like these who take placebo and develop AIDS. P: Two-sample z test for $p_1 - p_2$. Random: Two groups in a randomized experiment. Large Counts: 17, 418, 38, 397 are all ≥ 10. D: $z = -2.91$, P-value = 0.0018. C: Because the P-value of $0.0018 < \alpha = 0.05$, we reject H_0. We have convincing evidence that taking AZT lowers the proportion of patients like these who develop AIDS compared to a placebo. (b) I: Finding convincing evidence that AZT lowers the risk of developing AIDS, when in reality it does not. Consequence: patients will pay for a drug that doesn't help. II: Not finding convincing evidence that AZT lowers the risk of developing AIDS, when in reality it does. Consequence: patients won't take the drug when it could actually delay the onset of AIDS. It is possible that we made a Type I error.

R10.6 (a) The Large Counts condition is not met because there are only 7 failures in the control area. (b) The Normal/Large Sample condition is not met because both sample sizes are small and there are outliers in the male distribution.

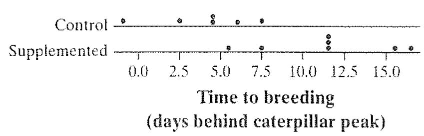
R10.7 (a) Even though each subject has two scores (before and after), the two groups of students are independent. (b) The distribution for the control group is slightly skewed to the right, while the distribution for the treatment group is roughly symmetric. The center for the treatment group is greater than the center for the control group. The differences in the control group are more variable than the differences in the treatment group. (c) S: $H_0: \mu_1 - \mu_2 = 0$ versus $H_a: \mu_1 - \mu_2 > 0$, where μ_1 = the true mean difference in test scores for students like these who get the treatment message and μ_2 = the true mean difference in test scores for students like these who get the neutral message. P: Two-sample t test for $\mu_1 - \mu_2$. Random: Two groups in a randomized experiment. Normal/Large Sample: Neither boxplot showed strong skewness or any outliers. D: Using the differences, $\bar{x}_1 = 11.4$, $s_1 = 3.169$, $\bar{x}_2 = 8.25$, $s_2 = 3.69$. $t = 1.91$. Using df = 7, the P-value is between 0.025 and 0.05. Using df = 13.92, P-value = 0.0382. C: Because the P-value of $0.0382 < \alpha = 0.05$, we reject H_0. There is convincing evidence that the true mean difference in test scores for students like these who get the treatment message is greater than the true mean difference in test scores for students like these who get the neutral message. (d) We cannot generalize to all students who failed the test because our sample was not a random sample of all students who failed the test.

Answers to Chapter 10 AP® Statistics Practice Test

T10.1 e
T10.2 b
T10.3 a
T10.4 a
T10.5 e
T10.6 e
T10.7 c
T10.8 c
T10.9 b
T10.10 a
T10.11 (a) S: μ_1 = the true mean hospital stay for patients like these who get heating blankets during surgery and μ_2 = the true mean hospital stay for patients like these who have core temperatures reduced during surgery. P: Two-sample t interval for $\mu_1 - \mu_2$.

Reasonable to consider these independent random samples. 10%: $n_1 = 675 < 10\%$ of male students at a large university and $n_2 = 621 < 10\%$ of female students at a large university. Normal/Large Sample: $n_1 = 675 \geq 30$ and $n_2 = 621 \geq 30$. *D*: Using df = 100, (412.68, 635.58). Using df = 1249.21, (413.62, 634.64). *C*: We are 90% confident that the interval from \$413.62 to \$634.64 captures the true difference in mean summer earnings of male students and female students at this large university. (c) If we took many random samples of 675 males and 621 females from this university and each time constructed a 90% confidence interval in this same way, about 90% of the resulting intervals would capture the true difference in mean earnings for males and females.

10.41 (a) *S*: $H_0: \mu_1 - \mu_2 = 0$ versus $H_a: \mu_1 - \mu_2 < 0$, where μ_1 is the true mean time to breeding for the birds relying on natural food supply and μ_2 is the true mean time to breeding for birds with food supplementation. *P*: Two-sample *t* test. Random: Two groups in a randomized experiment. Normal/Large Sample: Neither distribution displays strong skewness or outliers.



D: $\bar{x}_1 = 4.0$, $s_1 = 3.11$, $\bar{x}_2 = 11.3$, $s_2 = 3.93$. $t = -3.74$. Using df = 5, the *P*-value is between 0.005 and 0.01. Using df = 10.95, *P*-value = 0.0016. *C*: Because the *P*-value of $0.0016 < \alpha = 0.05$, we reject H_0. We have convincing evidence that the true mean time to breeding is less for birds relying on natural food supply than for birds with food supplements. (b) Assuming that the true mean time to breeding is the same for birds relying on natural food supply and birds with food supplements, there is a 0.0016 probability that we would observe a difference in sample means of −7.3 or smaller by chance alone.

10.43 *S*: $H_0: \mu_1 - \mu_2 = 0$ versus $H_a: \mu_1 - \mu_2 \neq 0$, where μ_1 is the true mean number of words spoken per day by female students and μ_2 is the true mean number of words spoken per day by male students. *P*: Two-sample *t* test. Random: Independent random samples. 10%: $n_1 = 56 < 10\%$ of females at a large university and $n_2 = 56 < 10\%$ of males at a large university. Normal/Large Sample: $n_1 = 56 \geq 30$ and $n_2 = 56 \geq 30$. *D*: $t = -0.248$. Using df = 50, *P*-value > 0.50. Using df = 106.20, *P*-value = 0.8043. *C*: Because the *P*-value of $0.8043 > \alpha = 0.05$, we fail to reject H_0. We do not have convincing evidence that the true mean number of words spoken per day by female students is different than the true mean number of words spoken per day by male students at this university.

10.45 (a) The distribution for the activities group is slightly skewed to the left, while the distribution for the control group is slightly skewed to the right. The center of the activities group is higher than the center of the control group. The scores in the activities group are less variable than the scores in the control group. (b) *S*: $H_0: \mu_1 - \mu_2 = 0$ versus $H_a: \mu_1 - \mu_2 > 0$, where μ_1 is the true mean DRP score for third-grade students like the ones in the experiment who do the activities and μ_2 is the true mean DRP score for third-grade students like the ones in the experiment who don't do the activities. *P*: Two-sample *t* test. Random: Two groups in a randomized experiment. Normal/Large Sample: No strong skewness or outliers in either boxplot. *D*: $t = 2.311$. Using df = 20, the *P*-value is between 0.01 and 0.02. Using df = 37.86, *P*-value = 0.0132. *C*: Because the *P*-value of $0.0132 < \alpha = 0.05$, we reject H_0. We have convincing evidence that the true mean DRP score for third-grade students like the ones in the experiment who do the activities is greater than the true mean DRP score for third-grade students like the ones in the experiment who don't do the activities. (c) Because this was a randomized controlled experiment, we can conclude that the activities caused the increase in the mean DRP score.

10.47 *D*: Using df = 50, (−3563, 2779). Using df = 106.2, (−3521, 2737). *C*: We are 95% confident that the interval from −3521 to 2737 captures the true difference between mean number of words spoken per day by female students and the mean number of words spoken per day by male students. This interval allows us to determine if 0 is a plausible value for the difference in means and also provides other plausible values for the difference in mean words spoken per day.

10.49 (a) *S*: $H_0: \mu_1 - \mu_2 = 10$ versus $H_a: \mu_1 - \mu_2 > 10$, where μ_1 is the true mean cholesterol reduction for people like the ones in the study when using the new drug and μ_2 is the true mean cholesterol reduction for people like the ones in the study when using the current drug. *P*: Two-sample *t* test. Random: Two groups in a randomized experiment. Normal/Large Sample: No strong skewness or outliers. *D*: $t = 0.982$. Using df = 13, the *P*-value is between 0.15 and 0.20. Using df = 26.96, *P*-value = 0.1675. *C*: Because the *P*-value of $0.1675 > \alpha = 0.05$, we fail to reject H_0. We do not have convincing evidence that the true mean cholesterol reduction is more than 10 mg/dl greater for the new drug than for the current drug. (b) Type II error. It is possible that the difference in mean cholesterol reduction is more than 10 mg/dl greater for the new drug than the current drug, but we didn't find convincing evidence that it was.

10.51 (a) The researchers randomly assigned the subjects to create two groups that were roughly equivalent at the beginning of the experiment. (b) Only about 5 out of the 1000 differences were ≥ 4.15, *P*-value ≈ 0.005. Because the *P*-value of $0.005 < \alpha = 0.05$, we have convincing evidence that the true mean rating for students like these that are provided with internal reasons is higher than the true mean rating for students like these that are provided with external reasons. (c) Because we found convincing evidence that the mean is higher for students with internal reasons when it is possible that there is no difference in the means, we could have made a Type I error.

10.53 (a) Two-sample. Two distinct groups of cars in a randomized experiment. (b) Paired. Both treatments are applied to each subject. (c) Two-sample. Two distinct groups of women.

10.55 (a) Paired, because we have two scores for each student. (b) *S*: $H_0: \mu_d = 0$ versus $H_a: \mu_d > 0$, where μ_d is the true mean increase in SAT verbal scores of students who were coached. *P*: Paired *t* test for μ_d. Random: Random sample. 10%: $n_d = 427 < 10\%$ of students who are coached. Normal/Large Sample: $427 \geq 30$. *D*: $t = 10.16$. Using df = 426, *P*-value ≈ 0. *C*: Because the *P*-value of approximately $0 < \alpha = 0.05$, we reject H_0. There is convincing evidence that students who are coached increase their scores on the SAT verbal test, on average.

10.57 a

10.59 b

10.61 (a) One-sample *z* interval for a proportion. (b) Paired *t* test for the mean difference. (c) Two-sample *z* interval for the difference in proportions. (d) Two-sample *t* test for a difference in means.

proportion of students like these who would pass the driver's license exam when taught by instructor B. *P*: Two-sample *z* test for $p_A - p_B$. Random: Two groups in a randomized experiment. Large Counts: 30, 20, 22, 28 are all ≥ 10. *D*: $z = 1.60$ and *P*-value = 0.0547. *C*: Because the *P*-value of $0.0547 > \alpha = 0.05$, we fail to reject H_0. There is not convincing evidence that the true proportion of students like these who would pass using instructor A is greater than the true proportion who would pass using instructor B. (b) I: Finding convincing evidence that instructor A is more effective than instructor B, when in reality the instructors are equally effective. II: Not finding convincing evidence that instructor A is better, when in reality instructor A is more effective. It is possible we made a Type II error.

10.23 (a) Two-sample *z* test for $p_1 - p_2$. Random: Two groups in a randomized experiment. Large Counts: 44, 44, 21, 60 are all ≥ 10. (b) If no difference exists in the true pregnancy rates of women who are being prayed for and those who are not, there is a 0.0007 probability of getting a difference in pregnancy rates as large or larger than the one observed in the experiment by chance alone. (c) Because the *P*-value of $0.0007 < \alpha = 0.05$, we reject H_0. There is convincing evidence that the pregnancy rates among women like these who are prayed for are higher than the pregnancy rates for those who are not prayed for. (d) Knowing they were being prayed for might have affected their behavior in some way that would have affected whether they became pregnant or not. Then we wouldn't know if it was the prayer or the other behaviors that caused the higher pregnancy rate.

10.25 a

10.27 c

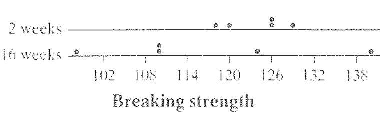
10.29 (a) $\hat{y} = -13{,}832 + 14{,}954x$, where $\hat{y}$ = the predicted mileage and x = the age in years of the cars. (b) For each year older the car is, the predicted mileage will increase by 14,954 miles. (c) Residual = −25,708. The student's car had 25,708 fewer miles than expected, based on its age.

Section 10.2

Answers to Check Your Understanding

page 644: *S*: μ_1 = the true mean price of wheat in July and μ_2 = the true mean price of wheat in September. *P*: Two-sample *t* interval for $\mu_1 - \mu_2$. Random: Independent random samples. 10%: $n_1 = 90 < 10\%$ of all wheat producers in July and $n_2 = 45 < 10\%$ of all wheat producers in September. Normal/Large Sample: $n_1 = 90 \geq 30$ and $n_2 = 45 \geq 30$. *D*: Using df = 40, (−0.759, −0.561). Using df = 100.45, (−0.756, −0.564). *C*: We are 99% confident that the interval from −0.756 to −0.564 captures the true difference in mean wheat prices in July and September.

page 649: *S*: $H_0: \mu_1 - \mu_2 = 0$ versus $H_a: \mu_1 - \mu_2 > 0$, where μ_1 is the true mean breaking strength for polyester fabric buried for 2 weeks and μ_2 is the true mean breaking strength for polyester fabric buried for 16 weeks. *P*: Two-sample *t* test. Random: Two groups in a randomized experiment. Normal/Large Sample: The dotplots below show no strong skewness or outliers in either group.



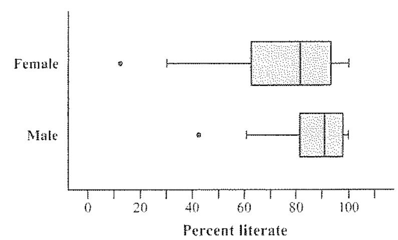
D: $\bar{x}_1 = 123.8$, $s_1 = 4.60$, $\bar{x}_2 = 116.4$, $s_2 = 16.09$. $t = 0.989$. Using df = 4, the *P*-value is between 0.15 and 0.20. Using df = 4.65, *P*-value = 0.1857. *C*: Because the *P*-value of $0.1857 > \alpha = 0.05$, we fail to reject H_0. We do not have convincing evidence that the true mean breaking strength of polyester fabric that is buried for 2 weeks is greater than the true mean breaking strength for polyester fabric that is buried for 16 weeks.

Answers to Odd-Numbered Section 10.2 Exercises

10.31 (a) Because the distributions of *M* and *B* are Normal, the distribution of $\bar{x}_M - \bar{x}_B$ is also Normal. (b) $\mu_{\bar{x}_M - \bar{x}_B} = 188 - 170 = 18$ mg/dl. (c) Because $25 < 10\%$ of all 20- to 34-year-old males and $36 < 10\%$ of all 14-year-old boys, $\sigma_{\bar{x}_M - \bar{x}_B} = \sqrt{\frac{(41)^2}{25} + \frac{(30)^2}{36}} = 9.60$ mg/dl.

10.33 Random: Two independent random samples. 10%: $20 < 10\%$ of all males at the school and $20 < 10\%$ of all females at the school. Normal/Large Sample: not met because there are fewer than 30 observations in each group and the stemplot for Males shows several outliers.

10.35 Random: not met because these data are not from two *independent* random samples. Knowing the literacy percent for females in a country helps us predict the literacy percent for males in that country. 10%: not met because 24 is more than 10% of Islamic countries. Normal/Large Sample: not met because the samples sizes are both small and both distributions are skewed to the left and have an outlier (see boxplots below).



10.37 (a) The distributions of percent change are both slightly skewed to the left. People drinking red wine generally have more polyphenols in their blood, on average. The distribution of percent change for the white wine drinkers is a little bit more variable. (b) *S*: μ_1 = the true mean change in polyphenol level in the blood of people like those in the study who drink red wine and μ_2 = the true mean polyphenol level in the blood of people like those in the study who drink white wine. *P*: Two-sample *t* interval for $\mu_1 - \mu_2$. Random: Two groups in a randomized experiment. Normal/Large Sample: The dotplots given in the problem do not show strong skewness or outliers. *D*: $\bar{x}_1 = 5.5$, $s_1 = 2.517$, $\bar{x}_2 = 0.23$, $s_2 = 3.292$. Using df = 8, (2.701, 7.839). Using df = 14.97, (2.845, 7.689). *C*: We are 90% confident that the interval from 2.845 to 7.689 captures the true difference in mean change in polyphenol level for men like these who drink red wine and men like these who drink white wine. (c) Because all of the plausible values in the interval are positive, this interval supports the researcher's belief that red wine is more effective than white wine.

10.39 (a) Earnings amounts cannot be negative, yet the standard deviation is almost as large as the distance between the mean and 0. However, the sample sizes are both very large ($675 \geq 30$ and $621 \geq 30$). (b) *S*: μ_1 = the true mean summer earnings of male students and μ_2 = the true mean summer earnings of female students. *P*: Two-sample *t* interval for $\mu_1 - \mu_2$. Random:

Sample: $n = 50 \geq 30$. *D*: $t = 2.47$; using df = 40, the *P*-value is between 0.01 and 0.02 (using df = 49, 0.0168). *C*: Because the *P*-value of $0.0168 < \alpha = 0.05$, we reject H_0. We have convincing evidence that the true mean amount spent on food per household in this city is different from the national average of \$158.

Chapter 10

Section 10.1

Answers to Check Your Understanding

page 619: *S*: p_1 = true proportion of teens who go online every day and p_2 = true proportion of adults who go online every day. *P*: Two-sample z interval for $p_1 - p_2$. Random: Independent random samples. 10%: $n_1 = 799 < 10\%$ of teens and $n_2 = 2253 < 10\%$ of adults. Large Counts: 503, 296, 1532, and 721 are all ≥ 10. *D*:

$$(0.63 - 0.68) \pm 1.645\sqrt{\frac{0.63(0.37)}{799} + \frac{0.68(0.32)}{2253}} =$$

$(-0.0824, -0.0176)$. *C*: We are 90% confident that the interval from -0.0824 to -0.0176 captures the true difference in the proportion of U.S. adults and teens who go online every day.

page 628: *S*: H_0: $p_1 - p_2 = 0$ versus H_a: $p_1 - p_2 > 0$, where p_1 is the true proportion of children like the ones in the study who do not attend preschool that use social services later and p_2 is the true proportion of children like the ones in the study who attend preschool that use social services later. *P*: Two-sample z test for $p_1 - p_2$. Random: Two groups in a randomized experiment. Large Counts: 49, 12, 38, 24 are all ≥ 10. *D*:

$$z = \frac{(0.8033 - 0.6129) - 0}{\sqrt{\frac{0.7073(0.2927)}{61} + \frac{0.7073(0.2927)}{62}}} = 2.32$$

and *P*-value = 0.0102. *C*: Because the *P*-value of $0.0102 < \alpha = 0.05$, we reject H_0. There is convincing evidence that the true proportion of children like the ones in the study who do not attend preschool that use social services later is greater than the true proportion of children like the ones in the study who attend preschool that use social services later.

Answers to Odd-Numbered Section 10.1 Exercises

10.1 (a) Approximately Normal because $100(0.25) = 25$, $100(0.75) = 75$, $100(0.35) = 35$, and $100(0.65) = 65$ are all at least 10. (b) $\mu_{\hat{p}_1 - \hat{p}_2} = 0.25 - 0.35 = -0.10$. (c) Because $n_1 = 100 < 10\%$ of the first bag and $n_2 = 100 < 10\%$ of the second bag, $\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{0.25(0.75)}{100} + \frac{0.35(0.65)}{100}} = 0.0644$.

10.3 (a) Approximately Normal because $50(0.30) = 15$, $50(0.7) = 35$, $100(0.15) = 15$, and $100(0.85) = 85$ are all at least 10. (b) $\mu_{\hat{p}_C - \hat{p}_A} = 0.30 - 0.15 = 0.15$. (c) Because $n_C = 50 < 10\%$ of the jelly beans in the Child mix and $n_A = 100 < 10\%$ of the jelly beans in the Adult mix, $\sigma_{\hat{p}_C - \hat{p}_A} = \sqrt{\frac{0.3(0.7)}{50} + \frac{0.15(0.85)}{100}} = 0.0740$.

10.5 The data do not come from independent random samples or two groups in a randomized experiment. Also, there were less than 10 successes (3) in the group from the west side of Woburn.

10.7 There were less than 10 failures (0) in the treatment group, less than 10 successes (8) in the control group, and less than 10 failures in the control group (4).

10.9 (a) $SE_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{0.26(1-0.26)}{316} + \frac{0.14(1-0.14)}{532}} = 0.0289$. If we were to take many random samples of 316 young adults and 532 older adults, the difference in the sample proportions of young adults and older adults who use Twitter will typically be 0.0289 from the true difference. (b) *S*: p_1 = true proportion of young adults who use Twitter and p_2 = true proportion of older adults who use Twitter. *P*: Two-sample z interval for $p_1 - p_2$. Random: Two independent random samples. 10%: $n_1 = 316 < 10\%$ of all young adults and $n_2 = 532 < 10\%$ of all older adults. Large Counts: 82, 234, 74, 458 are all at least 10. *D*: (0.072, 0.168). *C*: We are 90% confident that the interval from 0.072 to 0.168 captures the true difference in the proportions of young adults and older adults who use Twitter.

10.11 (a) *S*: p_1 = true proportion of young men who live in their parents' home and p_2 = true proportion of young women who live in their parents' home. *P*: Two-sample z interval for $p_1 - p_2$. Random: Reasonable to consider these independent random samples. 10%: $n_1 = 2253 < 10\%$ of the population of young men and $n_2 = 2629 < 10\%$ of the population of young women. Large Counts: 986, 1267, 923, 1706 are all at least 10. *D*: (0.051, 0.123). *C*: We are 99% confident that the interval from 0.051 to 0.123 captures the true difference in the proportions of young men and young women who live in their parents' home. (b) Because the interval does not contain 0, there is convincing evidence that the true proportion of young men who live in their parents' home is different from the true proportion of young women who live in their parents' home.

10.13 $H_0: p_1 - p_2 = 0$ versus $H_a: p_1 - p_2 \neq 0$, where p_1 is the true proportion of all teens who would say that they own an iPod or MP3 player and p_2 is the true proportion of all young adults who would say that they own an iPod or MP3 player.

10.15 *P*: Two-sample z test for $p_1 - p_2$. Random: Independent random samples. 10%: $n_1 = 800 < 10\%$ of all teens and $n_2 = 400 < 10\%$ of all young adults. Large Counts: 632, 168, 268, and 132 are all at least 10. *D*: $z = 4.53$ and *P*-value ≈ 0. *C*: Because the *P*-value of close to $0 < \alpha = 0.05$, we reject H_0. There is convincing evidence that the true proportion of teens who would say that they own an iPod or MP3 player is different from the true proportion of young adults who would say that they own an iPod or MP3 player.

10.17 *D*: (0.066, 0.174). *C*: We are 95% confident that the interval from 0.066 to 0.174 captures the true difference in proportions of teens and young adults who own iPods or MP3 players. Because 0 is not included in the interval, it is consistent with the results of Exercise 15.

10.19 *S*: $H_0: p_1 - p_2 = 0$ versus $H_a: p_1 - p_2 > 0$, where p_1 is the true proportion of 6- to 7-year-olds who would sort correctly and p_2 is the true proportion of 4- to 5-year-olds who would sort correctly. *P*: Two-sample z test for $p_1 - p_2$. Random: Independent random samples. 10%: $n_1 = 53 < 10\%$ of all 6- to 7-year olds and $n_2 = 50 < 10\%$ of all 4- to 5-year-olds. Large Counts: 28, 25, 10, 40 are all ≥ 10. *D*: $z = 3.45$ and *P*-value = 0.0003. *C*: Because the *P*-value of $0.0003 < \alpha = 0.05$, we reject H_0. We have convincing evidence that the true proportion of 6- to 7-year-olds who would sort correctly is greater than the true proportion of 4- to 5-year-olds who would sort correctly.

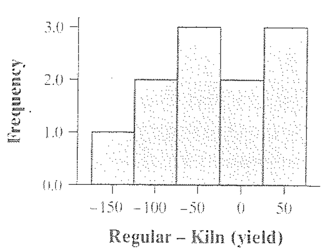
10.21 (a) *S*: $H_0: p_A - p_B = 0$ versus $H_a: p_A - p_B > 0$, where p_A is the true proportion of students like these who would pass the driver's license exam when taught by instructor A and p_B is the true evidence that the true mean breaking strength < 300 pounds. (e) Increase the sample size or increase the significance level.

R9.4 (a) *S*: H_0: $p = 0.05$ versus H_a: $p < 0.05$, where p is the true proportion of adults who will get the flu after using the vaccine. *P*: One-sample z test for p. Random: Random sample. 10%: The sample size (1000) < 10% of the population of adults. Large Counts: $1000(0.05) = 50 \geq 10$ and $1000(0.95) = 950 \geq 10$. *D*: $z = -1.02$ and *P*-value $= 0.1539$. *C*: Because the *P*-value of $0.1539 > \alpha = 0.05$, we fail to reject H_0. We do not have convincing evidence that fewer than 5% of adults who receive this vaccine will get the flu. (b) Because we failed to reject the null hypothesis, we could have made a Type II error—not finding convincing evidence that the true proportion of adults get the flu after using this vaccine <0.05, when in reality the true proportion <0.05. (c) Answers will vary.

R9.5 (a) Assuming that the roulette wheel is fair, there is a 0.0384 probability that we would get a sample proportion of reds at least this different from the expected proportion of reds (18/38) by chance alone. (b) Because the *P*-value of $0.0384 < \alpha = 0.05$, the results are statistically significant at the $\alpha = 0.05$ level. This means that we reject H_0 and have convincing evidence that the true proportion of reds is different than $p = 18/38$. (c) Because $18/38 = 0.474$ is one of the plausible values in the interval, this interval does not provide convincing evidence that the wheel is unfair. It does not, however, prove that the wheel is fair as there are many other plausible values in the interval that are not equal to 18/38. Also, the conclusion here is inconsistent with the conclusion in part (b) because the manager used a 99% confidence interval, which is equivalent to a test using $\alpha = 0.01$.

R9.6 (a) *S*: H_0:$\mu = 105$ versus H_a:$\mu \neq 105$, where μ is the true mean reading from radon detectors. *P*: One-sample t test for μ. Random: Random sample. 10%: The sample size (11) < 10% of all radon detectors. Normal/Large Sample: A graph of the data shows no strong skewness or outliers. *D*: $t = -0.06$, df = 10, and and *P*-value > 0.50 (0.9513). *C*: Because the *P*-value of $0.9513 > \alpha = 0.10$, we fail to reject H_0. We do not have convincing evidence that the true mean reading from the radon detectors is different than 105. (b) Yes. Because 105 is in the interval from 99.61 to 110.03, both the confidence interval and the significance test agree that 105 is a plausible value for the true mean reading from the radon detectors.

R9.7 (a) The random condition can be satisfied by randomly allocating which plot got the regular barley seeds and which one got the kiln-dried seeds within each pair of adjacent plots. (b) *S*: H_0: $\mu_d = 0$ versus H_a: $\mu_d < 0$, where μ_d is the true mean difference (regular − kiln) in yield between regular barley seeds and kiln-dried barley seeds. *P*: Paired t test for μ_d. Random: Assumed. Normal/Large Sample: The histogram below shows no strong skewness or outliers.



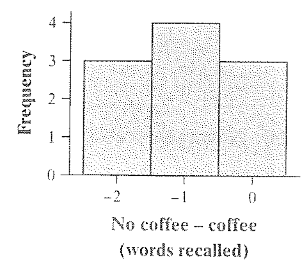
D: $\bar{x} = -33.7$ and $s_x = 66.2$. $t = -1.690$, df = 10, and the *P*-value is between 0.05 and 0.10 (0.0609). *C*: Because the *P*-value of $0.0609 > \alpha = 0.05$, we fail to reject H_0. We do not have convincing evidence that the true mean difference (regular − kiln) in yield <0.

Answers to Chapter 9 AP® Statistics Practice Test

T9.1 b
T9.2 e
T9.3 c
T9.4 e
T9.5 b
T9.6 c
T9.7 e
T9.8 d
T9.9 a
T9.10 c

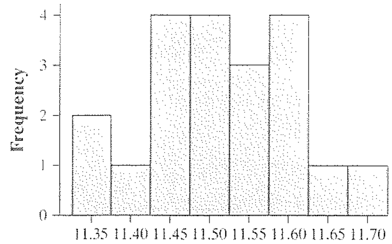
T9.11 (a) *S*: H_0: $p = 0.20$ versus H_a: $p > 0.20$, where p is the true proportion of customers who would pay \$100 for the upgrade. *P*: One-sample z test for p. Random: Random sample. 10%: The sample size (60) < 10% of this company's customers. Large Counts: $60(0.20) = 12 \geq 10$ and $60(0.8) = 48 \geq 10$. *D*: $z = 1.29$, *P*-value $= 0.0984$. *C*: Because the *P*-value of $0.0984 > \alpha = 0.05$, we fail to reject H_0. We do not have convincing evidence that more than 20% of customers would pay \$100 for the upgrade. (b) I: Finding convincing evidence that more than 20% of customers would pay for the upgrade, when in reality they would not. II: Not finding convincing evidence that more than 20% of customers would pay for the upgrade, when in reality more than 20% would. For the company, a Type I error is worse because they would go ahead with the upgrade and lose money. (c) Increase the sample size or increase the significance level.

T9.12 (a) Students may improve from Monday to Wednesday just because they have already done the task once. Then we wouldn't know if the experience with the test or the caffeine is the cause of the difference in scores. A better way to run the experiment would be to randomly assign half the students to get 1 cup of coffee on Monday and the other half to get no coffee on Monday. Then have each person do the opposite treatment on Wednesday. (b) *S*: H_0: $\mu_d = 0$ versus H_a: $\mu_d < 0$, where μ_d is the true mean difference (no coffee − coffee) in the number of words recalled without coffee and with coffee. *P*: Paired t test for μ_d. Random: The treatments were assigned at random. Normal/Large Sample: The histogram below shows a symmetric distribution with no outliers.



D: $\bar{x} = -1$ and $s_x = 0.816$. $t = -3.873$, df = 9, and the *P*-value is between 0.001 and 0.0025 (0.0019). *C*: Because the *P*-value of $0.0019 < \alpha = 0.05$, we reject H_0. We have convincing evidence that the mean difference (no coffee − coffee) in word recall < 0.

T9.13 *S*: H_0: $\mu = \$158$ versus H_a: $\mu \neq \$158$, where μ is the true mean amount spent on food by households in this city. *P*: One-sample t test for μ. Random: Random sample. 10%: The sample size (50) < 10% of households in this small city. Normal/Large



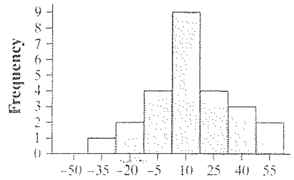
D: $\bar{x} = 11.5164$ and $s_x = 0.0950$. $t = 0.77$, df = 19, and the P-value is between 0.40 and 0.50 (0.4494). C: Because our P-value of $0.4494 > \alpha = 0.05$, we fail to reject H_0. We do not have convincing evidence that the true mean hardness of these tablets is different from 11.5.

9.79 D: With df = 19, (11.472,11.561). C: We are 95% confident that the interval from 11.472 to 11.561 captures the true mean hardness measurement for this type of pill. The confidence interval gives 11.5 as a plausible value for the true mean hardness μ, but it gives other plausible values as well.

9.81 S: H_0:$\mu = 200$ versus H_a:$\mu \neq 200$, where μ is the true mean response time of European servers. P: One-sample t interval to help us perform a two-sided test for μ. Random: The servers were selected randomly. 10%: The sample size (14) < 10% of all servers in Europe. Normal/Large Sample: The sample size is small, but a graph of the data reveals no strong skewness or outliers. D: (158.22, 189.64). C: Because our 95% confidence interval does not contain 200 milliseconds, we reject H_0 at the $\alpha = 0.05$ significance level. We have convincing evidence that the mean response time of European servers is different from 200 milliseconds.

9.83 (a) Yes. Because the P-value of $0.06 > \alpha = 0.05$, we fail to reject H_0: $\mu = 10$ at the 5% level of significance. Thus, the 95% confidence interval will include 10. (b) No. Because the P-value of $0.06 < \alpha = 0.10$, we reject H_0:$\mu = 10$ at the 10% level of significance. Thus, the 90% confidence interval would not include 10 as a plausible value.

9.85 (a) If all the subjects used the right thread first and they were tired when they used the left thread, then we wouldn't know if the difference in times was because of tiredness or because of the direction of the thread. (b) S: H_0: $\mu_d = 0$ versus H_a: $\mu_d > 0$, where μ_d is the true mean difference (left − right) in the time (in seconds) it takes to turn the knob with the left-hand thread and the right-hand thread. P: Paired t test for μ_d. Random: The order of treatments was determined at random. Normal/Large Sample: There is no strong skewness or outliers.



D: $\bar{x} = 13.32$ and $s_x = 22.94$. $t = 2.903$, df = 24, and the P-value is between 0.0025 and 0.005 (0.0039). C: Because the P-value of $0.0039 < \alpha = 0.05$, we reject H_0. We have convincing evidence that the true mean difference (left − right) in time it takes to turn the knob >0.

9.87 (a) H_0:$\mu_d = 0$ versus H_a:$\mu_d > 0$, where μ_d is the true mean difference in tomato yield (A − B). (b) df = 9. (c) *Interpretation*: Assuming that the average yield for both varieties is the same, there is a 0.1138 probability of getting a mean difference as large or larger than the one observed in this experiment. *Conclusion*: Because the P-value of $0.1138 > \alpha = 0.05$, we fail to reject H_0. We do not have convincing evidence that the true mean difference in tomato yield (A − B) > 0. (d) I: Finding convincing evidence that Variety A tomato plants have a greater mean yield, when in reality there is no difference. II: Not finding convincing evidence that Variety A tomato plants have a higher mean yield, when in reality Variety A does have a greater mean yield. They might have made a Type II error.

9.89 Increase the significance level α or increase the sample size n.

9.91 When the sample size is very large, rejecting the null hypothesis is very likely, even if the actual parameter is only slightly different from the hypothesized value.

9.93 (a) No, in a sample of size $n = 500$, we expect to see about (500)(0.01) = 5 people who do better than random guessing, with a significance level of 0.01. (b) The researcher should repeat the procedure on these four to see if they again perform well.

9.95 b

9.97 d

9.99 c

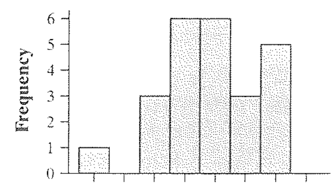
9.101 a

9.103 (a) Not included. The margin of error does not account for undercoverage. (b) Not included. The margin of error does not account for nonresponse. (c) Included. The margin of error is calculated to account for sampling variability.

Answers to Chapter 9 Review Exercises

R9.1 (a) H_0: $\mu = 64.2$; H_a: $\mu \neq 64.2$, where μ = the true mean height of this year's female graduates from the local high school. (b) H_0: $p = 0.75$; H_a: $p < 0.75$, where p = the true proportion of all students at Mr. Starnes's school who completed their math homework last night.

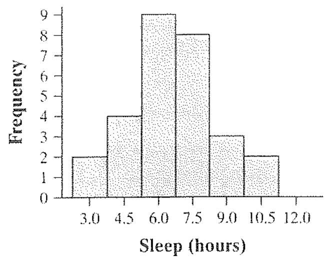
R9.2 Random: Random sample. 10%: The sample size (24) < 10% of the population of adults. Normal/Large Sample: The histogram below shows that the distribution is roughly symmetric with no outliers.



R9.3 (a) $H_0 : \mu = 300$ versus H_a: $\mu < 300$, where μ = the true mean breaking strength of these chairs. (b) I: Finding convincing evidence that the mean breaking strength <300 pounds, when in reality it is 300 pounds or higher. *Consequence*: falsely accusing the company of lying. II: Not finding convincing evidence that the mean breaking strength <300 pounds, when in reality it <300 pounds. *Consequence*: allowing the company to continue to sell chairs that don't work as well as advertised. (c) Because a Type II error is more serious, increase the probability of a Type I error by using $\alpha = 0.10$. (d) If the true mean breaking strength is 294 pounds, there is a 0.71 probability that we will find convincing

$2(0.0267) = 0.0534$. Because the P-value of $0.0534 > \alpha = 0.05$, we fail to reject H_0. There is not convincing evidence that the true mean amount of the active ingredient in Aspro tablets from this batch of production differs from 320 mg.

page 583: 1. S: $H_0: \mu = 8$ versus $H_a: \mu < 8$, where μ is the true mean amount of sleep that students at the professor's school get each night. P: One-sample *t* test for μ. Random: Random sample. 10%: The sample size (28) < 10% of the population of students. Normal/Large Sample: The histogram below indicates that there is not much skewness and no outliers.



D: $\bar{x} = 6.643$ and $s_x = 1.981$. $t = -3.625$ and the P-value is between 0.0005 and 0.001. *Using technology*: P-value = 0.0006. C: Because our P-value of $0.0006 < \alpha = 0.05$, we reject H_0. There is convincing evidence that students at this university get less than 8 hours of sleep, on average.

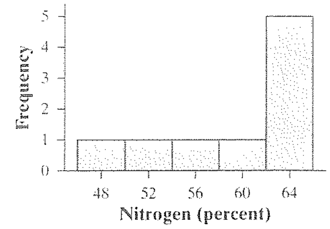
page 586: 1. S: $H_0: \mu = 128$ versus $H_a: \mu \neq 128$, where μ is the true mean systolic blood pressure for the company's middle-aged male employees. P: One-sample *t* test for μ. Random: Random sample. 10%: The sample size (72) < 10% of the population of middle-aged male employees. Normal/Large Sample: $n = 72 \geq 30$. D: $t = 1.10$ and P-value = 0.275. C: Because our P-value of $0.275 > \alpha = 0.05$, we fail to reject H_0. There is not convincing evidence that the mean systolic blood pressure for this company's middle-aged male employees differs from the national average of 128. 2. We are 95% confident that the interval from 126.43 to 133.43 captures the true mean systolic blood pressure for the company's middle-aged male employees. The value of 128 is in this interval and therefore is a plausible mean systolic blood pressure for the males 35 to 44 years of age.

page 589: S: $H_0: \mu_d = 0$ versus $H_a: \mu_d > 0$, where μ_d is the true mean difference (air – nitrogen) in pressure lost. P: Paired *t* test for μ_d. Random: Treatments were assigned at random to each pair of tires. Normal/Large Sample: $n = 31 \geq 30$. D: $\bar{x} = 1.252$ and $s_x = 1.202$. $t = 5.80$ and P-value ≈ 0. C: Because the P-value of approximately $0 < \alpha = 0.05$, we reject H_0. We have convincing evidence that the true mean difference in pressure (air – nitrogen) > 0. In other words, we have convincing evidence that tires lose less pressure when filled with nitrogen than when filled with air, on average.

Answers to Odd-Numbered Section 9.3 Exercises

9.65 Random: Random sample. 10%: The sample size (45) < 10% of the population size of 1000. Normal/Large Sample: $n = 45 \geq 30$.

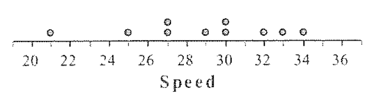
9.67 The Random condition may not be met, because we don't know if this is a random sample of the atmosphere in the Cretaceous era. Also, the Normal/Large Sample condition is not met. The sample size < 30 and the histogram below shows that the data are strongly skewed to the left.



9.69 (a) $t = \dfrac{125.7 - 115}{29.8/\sqrt{45}} = 2.409$. (b) For this test, df = 44. Using Table B and df = 40, we have 0.01 < P-value < 0.02. *Using technology*: `tcdf(lower:2.409, upper:1000, df:44)` = 0.0101.

9.71 (a) Using Table B and df = 19, we have 0.025 < P-value < 0.05. *Using technology*: P-value = 0.043. 5%: Because the P-value of $0.043 < \alpha = 0.05$, we reject H_0. There is convincing evidence that $\mu < 5$. 1%: Because the P-value of $0.043 > \alpha = 0.01$, we fail to reject H_0. There is not convincing evidence that $\mu < 5$. (b) *Using technology*: P-value = 0.086. 5%: Because the P-value of $0.086 > \alpha = 0.05$, we fail to reject H_0. There is not convincing evidence that $\mu \neq 5$. 1%: same as part (a).

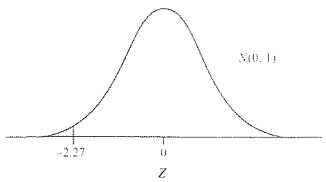
9.73 (a) S: H_0: $\mu = 25$ versus H_a: $\mu > 25$, where μ is the true mean speed of all drivers in a construction zone. P: One-sample *t* test for μ. Random: Random sample. 10%: The sample size (10) < 10% of all drivers. Normal/Large Sample: There is no strong skewness or outliers in the sample.



D: $\bar{x} = 28.8$ and $s_x = 3.94$. $t = 3.05$, df = 9, and the P-value is between 0.005 and 0.01 (0.0069). C: Because the P-value of $0.0069 < \alpha = 0.05$, we reject H_0. We have convincing evidence that the true mean speed of all drivers in the construction zone > 25 mph. (b) Because we rejected H_0, it is possible we made a Type I error—finding convincing evidence that the true mean speed > 25 mph when it really isn't.

9.75 (a) S: $H_0: \mu = 1200$ versus $H_a: \mu < 1200$, where μ is the true mean daily calcium intake of women 18 to 24 years of age. P: One-sample *t* test for μ. Random: Random sample. 10%: The sample size (36) < 10% of all women aged 18 to 24. Normal/Large Sample: $n = 36 \geq 30$. D: $t = -6.73$ and P-value = 0.000. C: Because the P-value of approximately $0 < \alpha = 0.05$, we reject H_0. There is convincing evidence that women aged 18 to 24 are getting less than 1200 mg of calcium daily, on average. (b) Assuming that women aged 18 to 24 get 1200 mg of calcium per day, on average, there is about a 0 probability that we would observe a sample mean ≤ 856.2 mg by chance alone.

9.77 S: $H_0: \mu = 11.5$ versus $H_a: \mu \neq 11.5$, where μ is the true mean hardness of the tablets. P: One-sample *t* test for μ. Random: The tablets were selected randomly. 10%: The sample size (20) < 10% of all tablets in the batch. Normal/Large Sample: There is no strong skewness or outliers in the sample.



9.37 (a) *P*-value = 0.0143. 5%: Because the *P*-value of 0.0143 $< \alpha = 0.05$, we reject H_0. There is convincing evidence that $p > 0.5$. 1%: Because the *P*-value of 0.0143 $> \alpha = 0.01$, we fail to reject H_0. There is not convincing evidence that $p > 0.5$. (b) *P*-value = 0.0286. Because this *P*-value is still less than $\alpha = 0.05$ and greater than $\alpha = 0.01$, we would again reject H_0 at the 5% significance level and fail to reject H_0 at the 1% significance level.
9.39 S: $H_0: p = 0.37$ versus $H_a: p > 0.37$, where p = true proportion of all students who are satisfied with the parking situation after the change. P: One-sample z test for p. Random: Random sample. 10%: The sample size (200) < 10% of the population of size 2500. Large Counts: $200(0.37) = 74 \geq 10$ and $200(0.63) = 126 \geq 10$. D: $z = 1.32$, *P*-value = 0.0934. C: Because the *P*-value of 0.0934 $> \alpha = 0.05$, we fail to reject H_0. We do not have convincing evidence that the true proportion of all students who are satisfied with the parking situation after the change > 0.37.
9.41 (a) S: $H_0: p = 0.50$ versus $H_a: p > 0.50$, where p is the true proportion of boys among first-born children. P: One-sample z test for p. Random: Random sample. 10%: The sample size (25,468) < 10% of all first-borns. Large Counts: $25{,}468(0.50) = 12{,}734 \geq 10$ and $25{,}468(0.50) = 12{,}734 \geq 10$. D: $z = 5.49$, *P*-value≈0. C: Because the *P*-value of approximately $0 < \alpha = 0.05$, we reject H_0. There is convincing evidence that first-born children are more likely to be boys. (b) First-born children, because that is the group that we sampled from.
9.43 Here are the corrections: H_a: $p > 0.75$; p = the true proportion of middle school students who engage in bullying behavior; 10%: the sample size (558) < 10% of the population of middle school students; $np_0 = 558(0.75) = 418.5 \geq 10$ and $n(1 - p_0) = 558(0.25) = 139.5 \geq 10$; $z = \dfrac{0.7975 - 0.75}{\sqrt{\dfrac{(0.75)(0.25)}{558}}} = 2.59$; *P*-value = 0.0048. Because the *P*-value of 0.0048 $< \alpha = 0.05$, we reject H_0. We have convincing evidence that more than three-quarters of middle school students engage in bullying behavior.
9.45 S: $H_0: p = 0.60$ versus $H_a: p \neq 0.60$, where p is the true proportion of teens who pass their driving test on the first attempt. P: One-sample z test for p. Random: Random sample. 10%: The sample size (125) < 10% of all teens. Large Counts: $125(0.60) = 75 \geq 10$ and $125(0.40) = 50 \geq 10$. D: $z = 2.01$, *P*-value = 0.0444. C: Because our *P*-value of 0.0444 $< \alpha = 0.05$, we reject H_0. There is convincing evidence that the true proportion of teens who pass the driving test on their first attempt is different from 0.60.
9.47 (a) D: (0.607,0.769). C: We are 95% confident that the interval from 0.607 to 0.769 captures the true proportion of teens who pass the driving test on the first attempt. (b) Because 0.60 is not in the interval, we have convincing evidence that the true proportion of teens who pass the driving test on their first attempt is different from 0.60.
9.49 No. Because the value 0.16 is included in the interval, we do not have convincing evidence that the true proportion of U.S. adults who would say they use Twitter differs from 0.16.
9.51 (a) p = the true proportion of U.S. teens aged 13 to 17 who think that young people should wait to have sex until marriage. (b) Random: Random sample. 10%: The sample size (439) < 10% of the population of all U.S. teens. Large Counts: $439(0.5) = 219.5 \geq 10$ and $439(0.5) = 219.5 \geq 10$. (c) Assuming that the true proportion of U.S. teens aged 13 to 17 who think that young people should wait to have sex until marriage is 0.50, there is a 0.011 probability of getting a sample proportion that is at least as different from 0.5 as the proportion in the sample. (d) Yes. Because the *P*-value of 0.011 $< \alpha = 0.05$, we reject H_0. There is convincing evidence that the true proportion of U.S. teens aged 13 to 17 who think that young people should wait to have sex until marriage differs from 0.5.
9.53 (a) I: Finding convincing evidence that more than 37% of students were satisfied with the new parking arrangement, when in reality only 37% were satisfied. *Consequence:* The principal believes that students are satisfied and takes no further action. II: Failing to find convincing evidence that more than 37% are satisfied with the new parking arrangement, when in reality more than 37% are satisfied. *Consequence:* The principal takes further action on parking when none is needed. (b) If the true proportion of students that are satisfied with the new arrangement is really 0.45, there is a 0.75 probability that the survey provides convincing evidence that the true proportion > 0.37. (c) Increase the sample size or significance level.
9.55 $P(\text{Type I}) = \alpha = 0.05$ and $P(\text{Type II}) = 0.22$.
9.57 (a) If the true proportion of Alzheimer's patients who would experience nausea is really 0.08, there is a 0.29 probability that the results of the study would provide convincing evidence that the true proportion < 0.10. (b) Increase the number of measurements taken (n) to get more information. (c) Decrease. If α is smaller, it becomes harder to reject the null hypothesis. This makes it harder to correctly reject H_0. (d) Increase. Because 0.07 is further from the null hypothesis value of 0.10, it will be easier to detect a difference between the null value and actual value.
9.59 c
9.61 b
9.63 (a)$X - Y$ has a Normal distribution with mean $\mu_{X-Y} = -0.2$ and standard deviation $\sigma_{X-Y} = \sqrt{(0.1)^2 + (0.05)^2} = 0.112$. To fit in a case, X − Y must take on a negative number. (b) We want to find $P(X - Y < 0)$ using the $N(-0.2, 0.112)$ distribution. $z = \dfrac{0 - (-0.2)}{0.112} = 1.79$ and $P(Z < 1.79) = 0.9633$. *Using technology*: 0.9629. There is a 0.9629 probability that a randomly selected CD will fit in a randomly selected case. (c) $P(\text{all fit}) = (0.9629)^{100} = 0.0228$. There is a 0.0228 probability that all 100 CDs will fit in their cases.

Section 9.3

Answers to Check Your Understanding

page 579: 1. $H_0: \mu = 320$ versus $H_a: \mu \neq 320$, where μ = the true mean amount of active ingredient (in milligrams) in Aspro tablets from this batch of production. 2. Random: Random sample. 10%: The sample of size 36 < 10% of the population of all tablets in this batch. Normal/Large Sample: $n = 36 \geq 30$. 3. $t = \dfrac{319 - 320}{3/\sqrt{36}} = -2$ 4. For this test, df = 35. Using Table B and df = 30, the tail area is between 0.025 and 0.05. Thus, the *P*-value for the two-sided test is between 0.05 and 0.10. *Using technology*: `2tcdf(lower:-1000,upper:-2,df:35)` =

9.13 $\alpha = 0.10$: Because the P-value of $0.2184 > \alpha = 0.10$, we fail to reject H_0. We do not have convincing evidence that the proportion of left-handed students at Simon's college is different from the national proportion. $\alpha = 0.05$: Because the P-value of $0.2184 > \alpha = 0.05$, we fail to reject H_0. We do not have convincing evidence that the proportion of left-handed students at Simon's college is different from the national proportion.
9.15 $\alpha = 0.05$: Because the P-value of $0.0101 < \alpha = 0.05$, we reject H_0. We have convincing evidence that the true mean score on the SSHA for all students at least 30 years of age at the teacher's college >115. $\alpha = 0.01$: Because the P-value of $0.0101 > \alpha = 0.01$, we fail to reject H_0. We do not have convincing evidence that the true mean score on the SSHA for all students at least 30 years of age at the teacher's college >115.
9.17 Either H_0 is true or H_0 is false—it isn't true some of the time and not true at other times.
9.19 The P-value should be compared with a significance level (such as $\alpha = 0.05$), not the hypothesized value of p. Also, the data never "prove" that a hypothesis is true, no matter how large or small the P-value.
9.21 (a) $H_0{:}\mu = 6.7$; $H_a{:}\mu < 6.7$, where μ represents the mean response time for all accidents involving life-threatening injuries in the city. (b) I: Finding convincing evidence that the mean response time has decreased when it really hasn't. A consequence is that the city may not investigate other ways to reduce the mean response time and more people could die. II: Not finding convincing evidence that the mean response time has decreased when it really has. A consequence is that the city spends time and money investigating other methods to reduce the mean response time when they aren't necessary. (c) Type I, because people may end up dying as a result.
9.23 (a) $H_0{:}\mu = \$85{,}000$; $H_a{:}\mu > \$85{,}000$, where μ = the mean income of all residents near the restaurant. (b) I: Finding convincing evidence that the mean income of all residents near the restaurant exceeds \$85,000 when in reality 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. II: Not finding convincing evidence that the mean income of all residents near the restaurant exceeds \$85,000 when in reality 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 and you lose potential income.
9.25 d
9.27 c
9.29 (a) P(woman) = 0.4168, so (24,611)(0.4168) = 10,258 degrees were awarded to women. (b) No. P(woman) = 0.4168, which is not equal to P(woman | bachelors) = 0.43.
(c) P(at least 1 of the 2 degrees earned by a woman)
= 1 − P(neither degree is earned by a woman) =

$$1 - \left(\frac{14{,}353}{24{,}611}\right)\left(\frac{14{,}352}{24{,}610}\right) = 0.6599$$

Section 9.2

Answers to Check Your Understanding

page 560: S: $H_0{:}p = 0.20$ versus $H_a{:}p > 0.20$, where p is the true proportion of all teens at the school who would say they have electronically sent or posted sexually suggestive images of themselves. P: One-sample z test for p. Random: Random sample. 10%: The sample size (250) < 10% of the 2800 students. Large Counts: $250(0.2) = 50 \geq 10$ and $250(0.8) = 200 \geq 10$. D:

$$z = \frac{0.252 - 0.20}{\sqrt{\frac{0.20(0.80)}{250}}} = 2.06$$

and $P(Z \geq 2.06) = 0.0197$. C: Because the P-value of $0.0197 < \alpha = 0.05$, we reject H_0. We have convincing evidence that more than 20% of the teens in her school would say they have electronically sent or posted sexually suggestive images of themselves.
page 563: S: $H_0{:}p = 0.75$ versus $H_a{:}p \neq 0.75$, where p is the true proportion of all restaurant employees at this chain who would say that work stress has a negative impact on their personal lives. P: One-sample z test for p. Random: Random sample. 10%: The sample size (100) < 10% of all employees. Large Counts: $100(0.75) = 75 \geq 10$ and $100(0.25) = 25 \geq 10$.

$$D{:}\ z = \frac{0.68 - 0.75}{\sqrt{\frac{0.75(0.25)}{100}}} = -1.62$$

and $2P(Z \leq -1.62) = 0.1052$. C: Because the P-value of $0.1052 > \alpha = 0.05$, we fail to reject H_0. We do not have convincing evidence that the true proportion of all restaurant employees at this large restaurant chain who would say that work stress has a negative impact on their personal lives is different from 0.75.
page 564: The confidence interval given in the output includes 0.75, which means that 0.75 is a plausible value for the population proportion that we are seeking. So both the significance test (which didn't rule out 0.75 as the proportion) and the confidence interval give the same conclusion. The confidence interval, however, gives a range of plausible values for the population proportion instead of only making a decision about a single value.
page 569: 1. A Type II error. If a Type I error occurred, they would reject a good shipment of potatoes and have to wait to get a new delivery. However, if a Type II error occurred, they would accept a bad batch and make potato chips with blemishes. This might upset consumers and decrease sales. To minimize the probability of a Type II error, choose a large significance level such as $\alpha = 0.10$
2. (a) Increase. Increasing α to 0.10 makes it easier to reject the null hypothesis, which increases power. (b) Decrease. Decreasing the sample size means we don't have as much information to use when making the decision, which makes it less likely to correctly reject H_0. (c) Decrease. It is harder to detect a difference of 0.02 (0.10 − 0.08) than a difference of 0.03 (0.11 − 0.08).

Answers to Odd-Numbered Section 9.2 Exercises

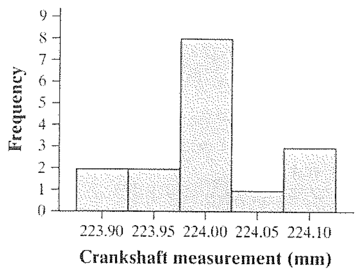
9.31 Random: Random sample. 10%: The sample size (60) < 10% of all students. Large Counts: $60(0.80) = 48 \geq 10$ and $60(0.20) = 12 \geq 10$.
9.33 $np_0 = 10(0.5) = 5$ and $n(1 - p_0) = 10(0.5) = 5$ are both < 10.
9.35 (a)

$$z = \frac{0.683 - 0.80}{\sqrt{\frac{0.80(0.20)}{60}}} = -2.27$$

(b) $P(Z \leq -2.27) = 0.0116$.
Using technology: `normalcdf (lower:−1000, upper: −2.27, μ:0, σ:1) = 0.0116`. The graph is given below.

entered. (b) It is likely that more than 171 respondents have run red lights because some people may lie and say they haven't run a red light. The margin of error does not account for these sources of bias; it accounts only for sampling variability.

R8.7 (a) S: μ = the true mean measurement of the critical dimension for the engine crankshafts produced in one day. P: One-sample t interval for μ. Random: The data come from an SRS. 10%: the sample size (16) is less than 10% of all crankshafts produced in one day. Normal/Large Sample: the histogram shows no strong skewness or outliers.



D: Using df = 15, (223.969, 224.035). C: We are 95% confident that the interval from 223.969 to 224.035 mm captures the true mean measurement of the critical dimension for engine crankshafts produced on this day. (b) Because 224 is a plausible value in this interval, we don't have convincing evidence that the process mean has drifted.

R8.8 Solving $1.96\left(\frac{3000}{\sqrt{n}}\right) \le 1000$ gives $n \ge 35$.

R8.9 (a) The margin of error must get larger to increase the capture rate of the intervals. (b) If we quadruple the sample size, the margin of error will decrease by a factor of 2.

R8.10 (a) When we use the sample standard deviation s_x to estimate the population standard deviation σ. (b) The t distributions are wider than the standard Normal distribution and they have a slightly different shape with more area in the tails. (c) As the degrees of freedom increase, the spread and shape of the t distributions become more like the standard Normal distribution.

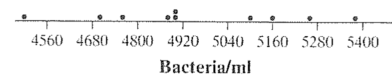
Answers to Chapter 8 AP® Statistics Practice Test

T8.1 a
T8.2 d
T8.3 c
T8.4 d
T8.5 b
T8.6 a
T8.7 c
T8.8 d
T8.9 e
T8.10 d

T8.11 (a) S: p = the true proportion of all visitors to Yellowstone who would say they favor the restrictions. P: One-sample z interval for p. Random: the visitors were selected randomly. 10%: the sample size (150) is less than 10% of all visitors to Yellowstone National Park. Large Counts: $n\hat{p} = 89 \ge 10$ and $n(1-\hat{p}) = 61 \ge 10$. D: (0.490, 0.696). C: We are 99% confident that the interval from 0.490 to 0.696 captures the true proportion of all visitors who would say that they favor the restrictions. (b) Because there are values smaller than 0.50 in the confidence interval, the U.S. Forest Service cannot conclude that more than half of visitors to Yellowstone National Park favor the proposal.

T8.12 (a) Because the sample size is large ($n = 48 \ge 30$), the Normal/Large Sample condition is met. (b) Maurice's interval uses a z critical value instead of a t critical value. Also, Maurice used the wrong value in the square root—it should be $n = 48$. *Correct:* Using df $= 40, 6.208 \pm 2.021\left(\frac{2.576}{\sqrt{48}}\right) = (5.457, 6.959)$. *Using technology*: (5.46, 6.956) with df = 47.

T8.13 S: μ = the true mean number of bacteria per milliliter of raw milk received at the factory. P: One-sample t interval for μ. Random: The data come from a random sample. 10%: the sample size (10) is less than 10% of all 1-ml specimens that arrive at the factory. Normal/Large Sample: the dotplot shows that there is no strong skewness or outliers.



D: Using df = 9, (4794.37, 5105.63). C: We are 90% confident that the interval from 4794.37 to 5105.63 bacterial/ml captures the true mean number of bacteria in the milk received at this factory.

Chapter 9

Section 9.1

Answers to Check Your Understanding

page 541: 1. (a) p = proportion of all students at Jannie's high school who get less than 8 hours of sleep at night. (b) $H_0: p = 0.85$ and $H_a: p \neq 0.85$. 2. (a) μ = true mean amount of time that it takes to complete the census form. (b) $H_0: \mu = 10$ and $H_a: \mu > 10$.
page 549: 1. Finding convincing evidence that the new batteries last longer than 30 hours on average, when in reality their true mean lifetime is 30 hours. 2. Not finding convincing evidence that the new batteries last longer than 30 hours on average, when in reality their true mean lifetime > 30 hours. 3. Answers will vary. A consequence of a Type I error would be that the company spends the extra money to produce these new batteries when they aren't any better than the older, cheaper type. A consequence of a Type II error would be that the company would not produce the new batteries, even though they were better.

Answers to Odd-Numbered Section 9.1 Exercises

9.1 $H_0: \mu = 115$; $H_a: \mu > 115$, where μ is the true mean score on the SSHA for all students at least 30 years of age at the teacher's college.

9.3 $H_0: p = 0.12$; $H_a: p \neq 0.12$, where p is the true proportion of lefties at his large community college.

9.5 $H_0: \sigma = 3$; $H_a: \sigma > 3$, where σ is the true standard deviation of the temperature in the cabin.

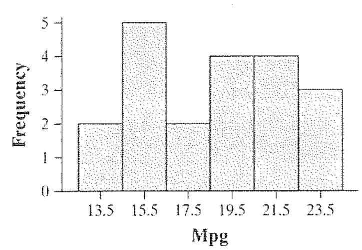
9.7 The null hypothesis is always that there is "no difference" or "no change" and the alternative hypothesis is what we suspect is true. Correct: $H_0: p = 0.37$; $H_a: p > 0.37$.

9.9 Hypotheses are always about population parameters. Correct: $H_0: \mu = 1000$ grams; $H_a: \mu < 1000$ grams.

9.11 (a) The attitudes of older students do not differ from other students, on average. (b) Assuming the mean score on the SSHA for students at least 30 years of age at this school is really 115, there is a 0.0101 probability of getting a sample mean of at least 125.7 just by chance in an SRS of 45 older students.

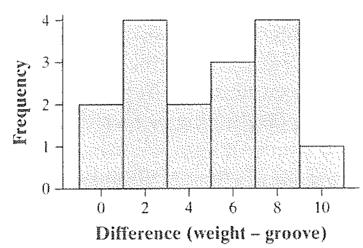
95% confident that the interval from 2.001 to 2.699 captures the true mean size of the muscle gap for the population of American and European young men. (b) The large sample size ($n = 200 \geq 30$) allows us to use a t interval for μ.

8.69 S: μ = the true mean fuel efficiency for this vehicle. P: One-sample t interval for μ. Random: the records were selected at random. 10%: it is reasonable to assume that 20 is less than 10% of all records for this vehicle. Normal/Large Sample: the histogram does not show any strong skewness or outliers.



D: Using df = 19, (17.022, 19.938). C: We are 95% confident that the interval from 17.022 to 19.938 captures the true mean fuel efficiency for this vehicle.

8.71 (a) S: μ = the true mean difference in the estimates from these two methods in the population of tires. P: One-sample t interval for μ. Random: A random sample of tires was selected. 10%: the sample size (16) is less than 10% of all tires. Normal/Large Sample: The histogram of differences shows no strong skewness or outliers.



D: Using df = 15, (2.837, 6.275). C: We are 95% confident that the interval from 2.837 to 6.275 thousands of miles captures the true mean difference in the estimates from these two methods in the population of tires. (b) Because 0 is not included in the confidence interval, there is convincing evidence of a difference in the two methods of estimating tire wear.

8.73 Solving $2.576\dfrac{7.5}{\sqrt{n}} \leq 1$ gives $n \geq 374$.

8.75 b

8.77 b

8.79 (a) Because the sum of the probabilities must be 1, $P(X = 7) = 0.57$. (b) $\mu_X = 5.44$. If we were to randomly select many young people, the average number of days they watched television in the past 7 days would be about 5.44. (c) Because the sample size is large ($n = 100 \geq 30$), we expect the mean number of days $\bar{x}$ for 100 randomly selected young people (aged 19 to 25) to be approximately Normally distributed with mean $\mu_{\bar{x}} = \mu = 5.44$. Because the sample size (100) is less than 10% of all young people aged 19 to 25, the standard deviation is $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{2.14}{\sqrt{100}} = 0.214$. We want to find $P(\bar{x} \leq 4.96)$. $z = \dfrac{4.96 - 5.44}{0.214} = -2.24$ and $P(Z \leq -2.24) = 0.0125$. *Using technology:* `normalcdf(lower: -1000, upper:4.96, μ:5.44, σ:0.214)` = 0.0124. There is a 0.0124 probability of getting a sample mean of 4.96 or smaller. Because this probability is small, a sample mean of 4.96 or smaller would be surprising.

Answers to Chapter 8 Review Exercises

R8.1 (a) $\dfrac{1 - 0.94}{2} = 0.03$, and the closest area is 0.0301, corresponding to a critical value of $z^* = 1.88$. *Using technology:* `invNorm(area:0.03, μ:0, σ:1)` = −1.881, so $z^* = 1.881$. (b) Using Table B and 50 degrees of freedom, $t^* = 2.678$. *Using technology:* `invT(area:0.005, df:57)` = −2.665, so $t^* = 2.665$.

R8.2 (a) $\bar{x} = \dfrac{430 + 470}{2} = 450$ minutes. Margin of error = 470 − 450 = 20 minutes. Because $n = 30$, df = 29 and $t^* = 2.045$. Because $20 = 2.045\dfrac{s_x}{\sqrt{30}}$, standard error $= \dfrac{s_x}{\sqrt{30}} = 9.780$ minutes and $s_x = 53.57$ minutes. (b) The confidence interval provided gives an interval estimate for the *mean* lifetime of batteries produced by this company, not individual lifetimes. (c) No. A confidence interval provides a statement about an unknown population mean, not another sample mean. (d) If we were to take many samples of 30 batteries and compute 95% confidence intervals for the mean lifetime, about 95% of these intervals will capture the true mean lifetime of the batteries.

R8.3 (a) p = the proportion of all adults aged 18 and older who would say that football is their favorite sport to watch on television. It may not equal 0.37 because the proportion who choose football will vary from sample to sample. (b) Random: The sample was random. 10%: The sample size (1000) is less than 10% of all adults. Large Counts: $n\hat{p} = 370 \geq 10$ and $n(1 - \hat{p}) = 630 \geq 10$. (c) $0.37 \pm 1.96\sqrt{\dfrac{0.37(0.63)}{1000}} = (0.3401, 0.3999)$. (d) We are 95% confident that the interval from 0.3401 to 0.3999 captures the true proportion of all adults who would say that football is their favorite sport to watch on television.

R8.4 (a) μ = mean IQ score for the 1000 students in the school. (b) Random: the data are from an SRS. 10%: the sample size (60) is less than 10% of the 1000 students at the school. Normal/Large Sample: $n = 60 \geq 30$. (c) Using df = 50, $114.98 \pm 1.676\left(\dfrac{14.8}{\sqrt{60}}\right) = (111.778, 118.182)$. *Using technology:* (111.79, 118.17) with df = 59. (d) We are 90% confident that the interval from 111.79 to 118.17 captures the true mean IQ score for the 1000 students in the school.

R8.5 Solving $2.576\sqrt{\dfrac{0.5(0.5)}{n}} \leq 0.01$ gives $n \geq 16{,}590$.

R8.6 (a) S: p = the true proportion of all drivers who have run at least one red light in the last 10 intersections they have entered. P: One-sample z interval for p. Random: the drivers were selected at random. 10%: The sample size (880) is less than 10% of all drivers. Large Counts: $n\hat{p} = 171 \geq 10$ and $n(1 - \hat{p}) = 709 \geq 10$. D: (0.168, 0.220). C: We are 95% confident that the interval from 0.168 to 0.220 captures the true proportion of all drivers who have run at least one red light in the last 10 intersections they have than 10% of the population of all students taking the SAT twice. Large Counts: $n\hat{p} = 427 \geq 10$ and $n(1-\hat{p}) = 2733 \geq 10$. *D*: (0.119, 0.151). *C*: We are 99% confident that the interval from 0.119 to 0.151 captures the true proportion of students retaking the SAT who receive coaching.

8.41 (a) We do not know the sample sizes for the men and for the women. (b) The margin of error for women alone would be greater than 0.03 because the sample size for women alone is smaller than 1019.

8.43 (a) Solving $1.645\sqrt{\frac{0.75(0.25)}{n}} \leq 0.04$ gives $n \geq 318$. (b) Solving $1.645\sqrt{\frac{0.5(0.5)}{n}} \leq 0.04$ gives $n \geq 423$. In this case, the sample size needed is 105 people larger.

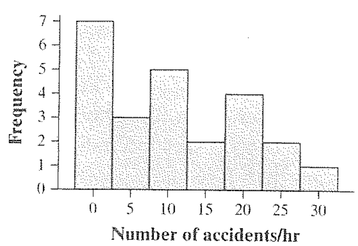
8.45 Solving $1.96\sqrt{\frac{0.5(0.5)}{n}} \leq 0.03$ gives $n \geq 1068$.

8.47 (a) Solving $0.03 = z^*\sqrt{\frac{0.64(0.36)}{1028}}$ gives $z^* = 2.00$. The confidence level is likely 95%, because 2.00 is very close to 1.96. (b) Teens are hard to reach and often unwilling to participate in surveys, so nonresponse is a major "practical difficulty" for this type of poll.

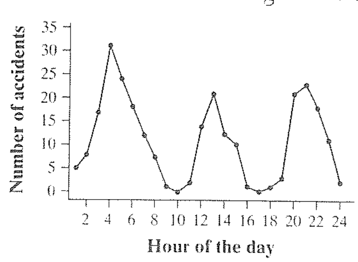
8.49 a

8.51 c

8.53 (a) A histogram of the number of accidents per hour is given below.



(b) A graph of the number of accidents is given below.



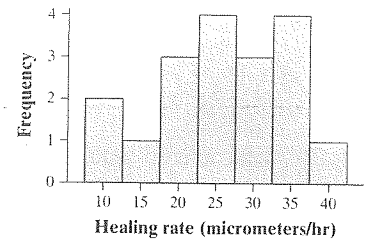
(c) The histogram in part (a) shows that the number of accidents has a distribution that is skewed to the right. (d) The graph in part (b) shows that there is a cyclical nature to the number of accidents.

Section 8.3

Answers to Check Your Understanding

page 514: 1. df = 21, $t^* = 2.189$. *Using technology*: `invT(area:0.02, df = 21)` $= -2.189$, so $t^* = 2.189$. 2. df = 70, $t^* = 2.660$ (using df = 60). *Using technology*: `invT(area: 0.005, df = 70)` $= -2.648$, so $t^* = 2.648$.

page 522: *S*: We are trying to estimate μ = the true mean healing rate at a 95% confidence level. *P*: One-sample *t* interval for μ. Random: The description says that the newts were randomly chosen. 10%: The sample size (18) is less than 10% of the population of newts. Normal/Large Sample: The histogram below shows no strong skewness or outliers, so this condition is met.



D: $25.67 \pm 2.110\left(\frac{8.32}{\sqrt{18}}\right) = 25.67 \pm 4.14 = (21.53, 29.81)$. *C*: We are 95% confident that the interval from 21.53 to 29.81 micrometers per hour captures the true mean healing time for newts.

page 524: Using $\sigma = 154$ and $z^* = 1.645$ for 90% confidence, $30 \geq 1.645\frac{154}{\sqrt{n}}$. Thus, $n \geq \left(\frac{1.645(154)}{30}\right)^2 = 71.3$, so take a sample of 72 students.

Answers to Odd-Numbered Section 8.3 Exercises

8.55 (a) $t^* = 2.262$. (b) $t^* = 2.861$. (c) $t^* = 1.671$ (*using technology*: $t^* = 1.665$).

8.57 Because the sample size is small ($n = 20 < 30$) and there are outliers in the data.

8.59 (a) No, because we are trying to estimate a population proportion, not a population mean. (b) No, because the 15 team members are not a random sample from the population. (c) No, because the sample size is small ($n = 25 < 30$) and there are outliers in the sample.

8.61 $SE_{\bar{x}} = \frac{9.3}{\sqrt{27}} = 1.7898$. If we take many samples of size 27, the sample mean blood pressure will typically vary by about 1.7898 from the population mean blood pressure.

8.63 (a) Because $19.03 = \frac{s_x}{\sqrt{23}}$, $s_x = 91.26$ cm. (b) They are using a critical value of $t^* = 1$. With df = 22, the area between $t = -1$ and $t = 1$ is approximately `tcdf(lower: −1, upper: 1, df: 22)` $= 0.67$. So, the confidence level is 67%.

8.65 (a) *S*: μ = the true mean percent change in BMC for breast-feeding mothers. *P*: One-sample *t* interval for μ. Random: the mothers were randomly selected. 10%: 47 is less than 10% of all breast-feeding mothers. Normal/Large Sample: $n = 47 \geq 30$. *D*: Using df = 40, (−4.575, −2.605). *Using technology*: (−4.569, −2.605) with df = 46. *C*: We are 99% confident that the interval from −4.569 to −2.605 captures the true mean percent change in BMC for breast-feeding mothers. (b) Because all of the plausible values in the interval are negative (indicating bone loss), the data give convincing evidence that breast-feeding mothers lose bone mineral, on average.

8.67 (a) *S*: μ = the true mean size of the muscle gap for the population of American and European young men. *P*: One-sample *t* interval for μ. Random: the young men were randomly selected. 10%: 200 is less than 10% of young men in America and Europe. Normal/Large Sample: $n = 200 \geq 30$. *D*: Using df = 100, (1.999, 2.701). *Using technology*: (2.001, 2.699) with df = 199. *C*: We are

8.9 (a) We are 95% confident that the interval from 0.63 to 0.69 captures the true proportion of those who favor an amendment to the Constitution that would permit organized prayer in public schools. (b) Point estimate $= \hat{p} = \frac{0.63 + 0.69}{2} = 0.66$ and margin of error $= 0.69 - 0.66 = 0.03$. (c) Because the value $2/3 = 0.667$ (and values less than 2/3) are in the interval of plausible values, there is not convincing evidence that more than two-thirds of U.S. adults favor such an amendment.

8.11 Because only 84% of the intervals actually contained the true parameter, these were probably 80% or 90% confidence intervals.

8.13 Answers will vary. One practical difficulty is response bias: people might answer "yes" because they think they should, even if they don't really support the amendment.

8.15 *Interval*: We are 95% confident that the interval from 10.9 to 26.5 captures the true difference (girls − boys) in the mean number of pairs of shoes owned by girls and boys. *Level*: If this sampling process were repeated many times, approximately 95% of the resulting confidence intervals would capture the true difference (girls − boys) in the mean number of pairs of shoes owned by girls and boys.

8.17 Yes. Because the interval does not include 0 as a plausible value, there is convincing evidence of a difference in the mean number of shoes for boys and girls.

8.19 (a) Incorrect. The interval provides plausible values for the *mean* BMI of all women, not plausible values for individual BMI measurements. (b) Incorrect. We shouldn't use the results of one sample to predict the results for future samples. (c) Correct. A confidence interval provides an interval of plausible values for a parameter. (d) Incorrect. The population mean doesn't change and will either be a value between 26.2 and 27.4 100% of the time or 0% of the time. (e) Incorrect. We are 95% confident that the population mean is between 26.2 and 27.4, but that does not absolutely rule out any other possibility.

8.21 b

8.23 e

8.25 (a) Observational study, because there was no treatment imposed on the pregnant women or the children. (b) No. We cannot make any conclusions about cause and effect because this was not an experiment.

Section 8.2

Answers to Check Your Understanding

page 496: 1. Random: not met because this was a convenience sample. 10%: met because the sample of 100 is less than 10% of the population at a large high school. Large Counts: met because 17 successes and 83 failures are both at least 10. 2. Random: met because the inspector chose an SRS of bags. 10%: met because the sample of 25 is less than 10% of the thousands of bags filled in an hour. Large Counts: not met because there were only 3 successes, which is less than 10.

page 499: 1. p = the true proportion of all U.S. college students who are classified as frequent binge drinkers. 2. Random: met because the statement says that the students were chosen randomly. 10%: met because the sample of 10,904 is less than 10% of all U.S. college students. Large Counts: met because 2486 successes and 8418 failures are both at least 10. 3. $\frac{1 - 0.99}{2} = 0.005$ and the closest area in Table A is 0.0051 (or 0.0049), corresponding to a critical value of $z^* = 2.57$ (or 2.58). *Using technology*: `invNorm(area:0.005, μ:0, σ:1)` $= -2.576$, so $z^* = 2.576$. $0.228 \pm 2.576\sqrt{\frac{0.228(1 - 0.228)}{10904}} = 0.228 \pm 0.010 = (0.218, 0.238)$. 4. We are 99% confident that the interval from 0.218 and 0.238 captures the true proportion of all U.S. college students who are classified as frequent binge drinkers.

page 503: 1. Solving $1.96\sqrt{\frac{0.80(0.20)}{n}} \leq 0.03$ for n gives $n \geq 682.95$. We should select a sample of at least 683 customers. 2. The required sample size will be larger because the critical value is larger for 99% confidence (2.576) versus 95% confidence (1.96). The company would need to select at least 1180 customers.

Answers to Odd-Numbered Section 8.2 Exercises

8.27 Random: met because Latoya selected an SRS of students. 10%: not met because the sample size (50) is more than 10% of the population of seniors in the dormitory (175). Large Counts: met because $n\hat{p} = 14 \geq 10$ and $n(1 - \hat{p}) = 36 \geq 10$.

8.29 Random: may not be met because we do not know if the people who were contacted were a random sample. 10%: met because the sample size (2673) is less than 10% of the population of adult heterosexuals. Large Counts: not met because $n\hat{p} = 2673(0.002) \approx 5$ is not at least 10.

8.31 $\frac{1 - 0.98}{2} = 0.01$, and the closest area is 0.0099, corresponding to a critical value of $z^* = 2.33$. *Using technology*: `invNorm(area:0.01, μ:0, σ:1)` $= -2.326$, so $z^* = 2.326$.

8.33 (a) Population: seniors at Tonya's high school. Parameter: true proportion of all seniors who plan to attend the prom. (b) Random: the sample is a simple random sample. 10%: The sample size (50) is less than 10% of the population size (750). Large Counts: $n\hat{p} = 36 \geq 10$ and $n(1 - \hat{p}) = 14 \geq 10$.

(c) $0.72 \pm 1.645\sqrt{\frac{0.72(0.28)}{50}} = 0.72 \pm 0.10 = (0.62, 0.82)$.

(d) We are 90% confident that the interval from 0.62 to 0.82 captures the true proportion of all seniors at Tonya's high school who plan to attend the prom.

8.35 (a) *S*: p = the true proportion of all full-time U.S. college students who are binge drinkers. *P*: One-sample z interval for p. Random: the students were selected randomly. 10%: the sample size (5914) is less than 10% of the population of all college students. Large Counts: $n\hat{p} = 2312 \geq 10$ and $n(1 - \hat{p}) = 3602 \geq 10$. *D*: (0.375, 0.407). *C*: We are 99% confident that the interval from 0.375 to 0.407 captures the true proportion of full-time U.S. college students who are binge drinkers. (b) Because the value 0.45 does not appear in our 99% confidence interval, it isn't plausible that 45% of full-time U.S. college students are binge drinkers.

8.37 Answers will vary. Response bias is one possibility.

8.39 *S*: p = the true proportion of all students retaking the SAT who receive coaching. *P*: One-sample z interval for p. Random: the students were selected randomly. 10%: the sample size (3160) is less