TT

TEXTBOOK TESTBANKS

Solutions Manual for Introduction to Statistical Investigations 2nd Tintle

SOLUTIONS MANUAL SAMPLE PREVIEW

PUBLISHER

John Wiley & Sons

COPYRIGHT

2021

RESOURCE TYPE

Solutions Manual

ISBN

9781119683452

Generalization: How Broadly Do the Results Apply?

Section 2.1

2.1.1 **B.**

2.1.2 **A.**

2.1.3 **A.**

2.1.4 **C.**

2.1.5 **C.**

2.1.6 **C.**

2.1.7 **D.**

2.1.8 **D.**

2.1.9 **A.**

2.1.10 **A.**

2.1.11 **B.**

2.1.12

A. False

B. True

C. True

2.1.13 **A.**

2.1.14 **B.**

2.1.15

a. F.

b. C.

c. A.

2.1.16

a. SD $= \sqrt{0.41(0.59)/30} = 0.090$

b. The SD should be about 0.085 which is a bit smaller than that from part (a). It is different because the population is less than 10 times the sample size. It should be at least 20 times the sample size.

c. The SD should be about 0.090 which is what we got in part (a). The larger population size helped.

2.1.17 The graph of the most recent sample represents whether or not a word was short, a categorical variable. However, in the graph of the proportions, the horizontal axis represents the proportion of short words in a sample, a quantitative variable.

2.1.18

a. Mean $= 0.25$, and SD $= 0.068$

b. Mean $= 0.25$, and SD $= 0.022$

2.1.19

a. Mean $= 0.20$, SD $= \sqrt{0.20(0.80)/1{,}414} = 0.0106$

b. $(271/1{,}414 - 0.200)/0.0106 = -0.7873$. So the sample proportion is 0.7873 standard deviations below the mean.

c. It is not very unlikely to get a sample proportion of 271/1,414 because it is within 2 standard deviations of the mean.

2.1.20 Using a phone call as the method of asking this question is probably a biased method. Those answering a person on a phone call were much more unlikely to say that they exercise less than once per week. Having an interaction with a person probably makes some people not give the socially undesirable answer.

2.1.21

a. Likely representative, because the distribution of blood type is probably not different among those that eat at the cafeteria compared to that of the U.S. population.

b. Likely not representative because those in the cafeteria may eat most of their meals in the cafeteria and not regularly eat fast food

c. Perhaps representative because the proportion with brown hair is probably not too different among those that eat at the cafeteria compared to that of all the students at the school.

d. This may not be representative. Those in the cafeteria (as well as those at the school) could differ quite a bit racially from the U.S. population and thus would differ in the proportion that has brown hair.

2.1.22 Although some of these could be argued the other way, all of the samples would likely not be representative.

a. If you have food that finches like and other birds do not, you would overestimate the proportion of finches.

b. You could have food that finches do not like, and you would rarely see more than one eating at a time.

c. The proportion of male birds could be species-dependent and depending on the type of food you have could affect the type of species and hence affect the proportion of male birds that come to your feeder.

d. This proportion could then be different than the proportion of males in your area as well as those that typically visit feeders.

2.1.23

a. The population is all the likely voters in the city.

b. Because it is a random sample, I would think the proportion that favor the incumbent in the sample is similar to that for the population.

2.1.24

a. The proportion of all city voters that plan to vote for the incumbent

b. The proportion of those in the sample that plan to vote for the incumbent (0.65).

2.1.25

a. The variable is who they plan to vote for or whether or not they plan to vote for the incumbent.

b. Categorical

c. Proportion

d. Bar graph

2.1.26

a. Null: 50% of the all the city voters plan to vote for the incumbent. Alternative: A majority of all the city voters plan to vote for the incumbent.

b. The proportion of all city voters that plan to vote for the incumbent, 0.65.

2.1.27

a. p-value ≈ 0

b. We have strong evidence that a majority of all the city voters plan to vote for the incumbent.

c. We can infer these results to all likely city voters because they came from a random sample.

d. Using theory-based methods is appropriate and we also obtain a p-value of approximately 0.

2.1.28

a. The population is all adults in the United States

b. It is perhaps greater than the population. People were allowed to self-select themselves to be part of the sample. This method will often overestimate the population proportion because people that really care about the issue will be the ones to respond.

2.1.29

a. The proportion of U.S. adults that are unhappy with the verdict.

b. The proportion of the sample (0.82) that are unhappy with the verdict.

2.1.30

a. Whether or not someone is unhappy with the verdict

b. Categorical

c. Proportion

d. Bar graph

2.1.31

a. Null: The proportion of U.S. adults that are unhappy with the verdict is 0.75. Alternative: The proportion of U.S. adults that are unhappy with the verdict is greater than 0.75.

b. 0.82

2.1.32

a. p-value ≈ 0

b. We have very strong evidence that the proportion of all U.S. adults that are opposed to the verdict is greater than 0.75.

c. Not comfortable generalizing to all U.S. adults. Instead, comfortable generalizing to a population like the one that participated in the survey—watchers of the TV news program who were motivated enough to participate.

d. Theory-based is appropriate because at least 10 successes and 10 failures in the data, $z = 3.83$ and p-value $= 0.0001$.

2.1.33

a. The population is all the sharks at the zoo.

b. The proportion should be similar to the population because it came from a random sample.

2.1.34

a. The parameter is the proportion of sharks in the zoo that have the disease.

b. The statistic is the proportion of sharks in the sample (0.20) that have the disease.

2.1.35

a. The variable measured is whether or not they have the disease.

b. Categorical

c. Proportion

d. Bar graph

2.1.36

a. Null: The proportion of all sharks at the zoo that have the disease is 0.25. Alternative: The proportion of all sharks at the zoo that have the disease is less than 0.25.

b. $3/15 = 0.20$

2.1.37

a. p-value $= 0.46$

b. We do not have any evidence that the proportion of sharks in the zoo that have the disease is less than 0.25.

c. The sharks at the zoo

d. A theory-based approach using the normal distribution is not reasonable to use because there were only three sharks with the disease. We need at least 10.

2.1.38

a. All the customers of the store

b. The 100 people asked to fill out the survey

c. The proportion of all customers that visit the store because of the sale on coats

d. The proportion of the sample that said they visited the store because of the sale on coats (0.40)

2.1.39

It may not be representative because it was not a random sample.

2.1.40

a. The new proportion of $1{,}523/2{,}216 = 0.687$ is closer to the proportion that actually voted.

b. SD $= \sqrt{0.592(0.408)/2{,}216} = 0.0104$

c. $(0.687 - 0.592)/0.0104 = 9.13$. Because the sample proportion is 9.13 SD above 0.592, it is significantly larger. Results like this are very unlikely to happen by chance.

2.1.41 The question is awkwardly phrased. "Does it seem possible or impossible ... it never happened?" The latter part ("impossible it never happened") involves a double negative.

2.1.42 This question is more clearly phrased; they got rid of the double negative.

2.1.43 Yes, assistance to the poor likely elicits a more favorable response toward programs than the term welfare.

2.1.44 Probably the question that did not give two options will yield a higher percentage of affirmative responses.

2.1.45 A sample of 1,500 might not have many or any members of a rare subpopulation, but that does not matter because members of a rare subpopulation are such a small part of the entire population and hence do not really affect an overall population proportion.

2.1.46 False. Increasing the sample size will not affect bias but will only affect sampling variability.

2.1.47 In the Doris and Buzz example, Dr. Bastian randomly determine if the light would flash or if it would be shining steadily.

2.1.48 Randomness occurs in the chance model by flipping a coin to determine if Buzz would push the correct button if he was just guessing (a constant probability). In reality, Buzz may not have a constant probability of choosing the correct button. He may be learning along the way, he may get tired, or his stomach may get full of fish.

Section 2.2

2.2.1 **B.**

2.2.2 **B.**

2.2.3 **A.**

2.2.4 **D.**

2.2.5 **B.**

2.2.6 **C.**

2.2.7 **B.**

2.2.8 **B.**

2.2.9 **D.**

2.2.10 **C.**

2.2.11 **A.**

2.2.12 **A, D.**

2.2.13 **E.**

2.2.14 Graph (a) is a distribution of sample means from samples of size 30; we know this because it is the distribution with the smaller SD.

2.2.15

a. False

b. False

c. False

2.2.16 The horizontal axis of the graph of the most recent sample is the length of the words, but the horizontal axis of the graph of the statistic is the mean length of the words in a sample of 10 words. Both of these are quantitative.

2.2.17

a. Because the distribution is skewed to the left, the mean will be to the left of the median; hence, 65.86°F is the mean and 67.50°F is the median.

b. Mean: Larger; Median: Larger; Standard deviation: Smaller

2.2.18 Because your score of 84 is below the median of 87, more students had exam scores higher than yours than below.

2.2.19

a. The students in the class

b. The variable is the number of states visited and it is quantitative.

c. The graph is centered about 8, most of the data is between 2 and 16 (and skewed to the right a bit) with outliers 25, 30, and 43.

d. 7.5

e. The mean would be larger because the distribution is skewed to the right.

f. The mean would be smaller, the median would stay the same, and the standard deviation would be smaller.

2.2.20

a. The distribution is skewed to the right.

b. Because the distribution is skewed to the right, we should expect the mean to be higher than the median.

c. The median is \$35, and the mean is \$45.68. The mean is higher, as expected.

d. If a \$150 haircut is changed to \$300 we should expect the median to stay the same (because \$150 or \$300 is just another larger value), but the mean should increase (because the total of all haircut costs will increase). When the change is made, we can see the median does stay the same and the mean increases to \$48.68.

2.2.21

a. **E.**

b. **C.**

c. **B.**

2.2.22

a. $\text{SD} = \frac{2.119}{\sqrt{30}} = 0.387$

b. The SD is about 0.366. It is smaller than what is predicted because the population size is too small. The population size should be at least 20 times the sample size and it is less than 10 times.

c. The SD is about 0.386 or 0.387, much closer to what was predicted in (a).

2.2.23

a. Mean = 10, and SD = 0.8

b. Mean = 10, and SD = 0.4

2.2.24

a. Decreases

b. Increases

c. 100

2.2.25 The sample size must be four times as large.

2.2.26

a. Mean = 100 points, SD = 3.35 points

b. 80

c. 320

2.2.27

a. **i.** Mean = 8 hours, SD = 0.47 hours, and bell-shaped;
ii. Yes

b. **i.** Mean = 8 hours, SD = 0.47 hours, and slight skewed to the right;
ii. yes

c. **i.** Mean = 8 hours, SD = 0.47 hours, and bell-shaped;
ii. Yes

d. Regardless of the shape of the population distribution, the mean and SD were about 8 and 0.47, respectively. When the population had a symmetric shape, the sampling distribution had a bell shape. But when the population had a skewed distribution, the sampling distribution was also skewed.

2.2.28

a. Students at her school

b. I would guess that the average number of hours students in a statistics class watch TV per day is similar to that of students in the entire school.

2.2.29

a. The parameter is the average number of hours students at the school watch TV per day.

b. The statistic is the average number of hours students in the sample watch TV per day (1.2 hours).

2.2.30

a. The variable is the number of hours of TV is watched per day.

b. Quantitative

c. Mean or median

d. Dotplot

2.2.31

a. If a student watched more than 10 minutes of TV yesterday

b. Categorical

c. Proportion

d. Bar graph

2.2.32

a. Null: 50% of the students at the school watched at least 10 minutes of TV yesterday. Alternative: More than 50% of the students at the school watched at least 10 minutes of TV yesterday.

b. The proportion (or percentage) of all students that watched at least 10 minutes of TV yesterday

c. $21/30 = 0.70$

2.2.33

a. 0.0214

b. We have strong evidence that students watched more than 10 minutes of TV yesterday.

c. Concerned because not a random sample of students at the school, although that is the population of interest; generalize to this population with caution.

d. A theory-based approach is not reasonable because there are not at least 10 students in the sample who watched less than 10 minutes of TV yesterday.

2.2.34

Student	Hours per day	Watched TV yesterday
Alejandra	2	no
Ben	4	yes
Cassie	0.5	no
...	...	...

2.2.35 The study was not done using a random sample. To take a random sample of 30 students at the school, we first need to obtain a list of all the students at the school and randomly choose from that list. One way to do this is to assign every student a number and then have a random number generator give you 30 random numbers. The students' names that match the numbers chosen will be your random sample.

2.2.36

a. How long people spend reading or learning about local politics

b. Quantitative

c. Mean or median

d. Dotplot

2.2.37

Voter	Time spent reading/ learning about local politics	Voting for incumbent?
#1	0	Yes
#2	0	Yes
#3	60	No
...	...	...

2.2.38 The study was done using a random sample. It could have been done by obtaining a list of all the voters in the city and then assigning every voter a number. Then have a random number generator give 267 random numbers. The voter's names that match the numbers chosen will be the random sample.

2.2.39

a. The variable is the time spent reading or watching news coverage about the trial in the last 3 days.

b. Quantitative

c. Mean or median

d. Dotplot

2.2.40

Respondent	Time spent reading/ watching about the trial	Not happy with verdict?
#1	240	Not happy
#2	90	Not happy
#3	30	Happy
...	...	...

2.2.41 The study was not done using a random sample. Because national polls like this don't have available lists of all U.S. adults from which to sample, typically random-digit dialing is done. Phone numbers are randomly generated. This could have been done for this poll.

2.2.42

a. The shark's blood oxygen content

b. Quantitative

c. Mean or median

d. Dotplot

2.2.43

Shark	Has disease?	Blood oxygen content
#1	Yes	1.2%
#2	No	5.6%
#3	No	6.2%
...	...	...

2.2.44 The study was done using a random sample. It could have been done by obtaining a list of all the sharks at the zoo and assigning each a number. Then have a random number generator give 15 random numbers. The sharks that match the numbers chosen will be the random sample.

2.2.45

a. All full-time students at the school

b. The 150 students in the sample

c. The average daily study time for all full-time students at the school

d. The average daily study time for the students in the sample (2.23 hours)

2.2.46 Because it is a random sample it should be representative of the population.

2.2.47

a. If one coffee bar opens earlier than the other, the lines may be different at the bars when they first open (if one opens before first class starts it may have a longer line than one that opens after first class starts).

b. A more representative sample of students might be observed at different times of the day and on different days of the week.

2.2.48

a. Maybe overstate, as students may tend to think they are getting more sleep than they are actually getting

b. Maybe overstate, as students may tend to think they are volunteering more than they are actually volunteering

c. Maybe overstate, as students may tend to think they are attending church more often than they actually are attending

d. Maybe overstate, as students may tend to think they are studying more than they actually are studying

e. Maybe overstate, as students may tend to think they are wearing a seat belt more often than they actually wear a seat belt

Section 2.3

2.3.1 B.

2.3.2 C.

2.3.3 D.

2.3.4 B.

2.3.5 C.

2.3.6 C.

2.3.7

a. The number of hours Cal Poly students watch TV per day and it is quantitative

b. μ = the average number of hours Cal Poly students watch TV per day

c. H_0: $\mu = 2.84$ hrs. H_a: $\mu \neq 2.84$ hrs

2.3.8

a. Statistic

b. $\bar{x} = 3.01$

c. You would have to fabricate a large dataset to represent the population of times for all Cal Poly students with the variability similar to that of the sample data and a mean of 2.84. From that data you would take a sample of 100 and find its mean. Repeat this at least 1,000 times to develop a null distribution. To determine the p-value, determine the proportion of simulated statistics that are at least as far away from 2.84 as that of 3.01.

d.

Simulation		Real study
One repetition	=	A random sample of 100 students
Null model	=	Population mean hours of TV watching is 2.84 hours
Statistic	=	The average TV watching time in the sample

2.3.9

a. If the mean daily TV watching time for Cal Poly students is 2.84 hours, the probability we would get a sample mean as extreme as 3.01 from a random sample of 100 students is 0.16.

b. Because this is a one-sided test, Dr. Elliot's p-value should be about half that of Dr. Sameer's.

2.3.10

a. $s = 1.97$ is a statistic.

b. $(3.01 - 2.84)/(1.97/\sqrt{100}) = 0.86$

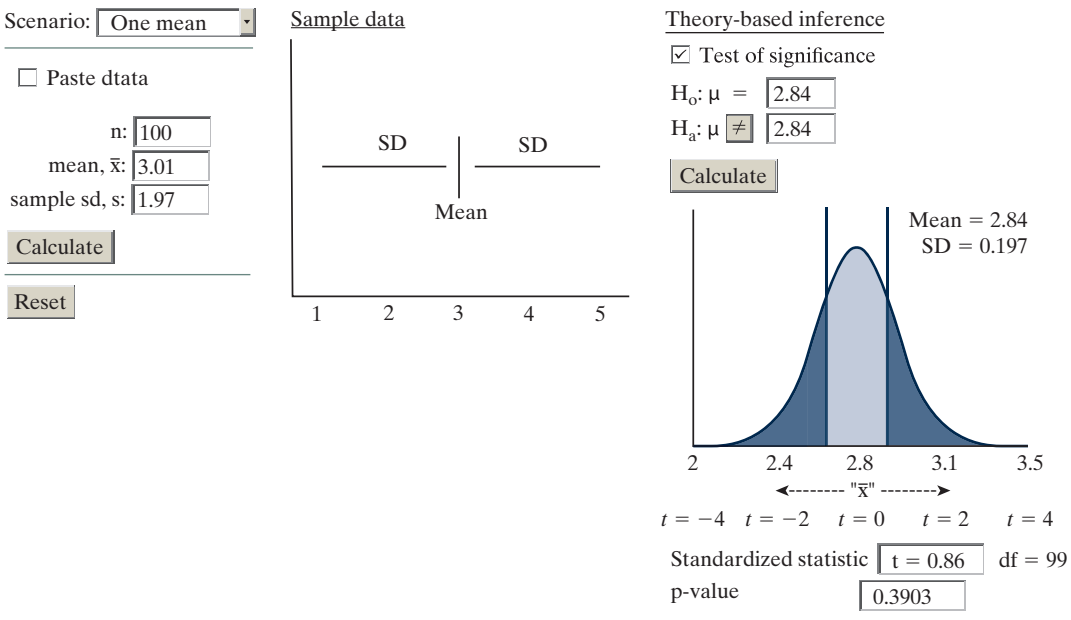
c. Because the standardized statistic is 0.86 (and that is less than 2) we do not have strong evidence that the average time Cal Poly students watch TV is different than 2.84 hours.

2.3.11

a. Yes, because the sample size is large

b. The standardized statistic is 0.86 and the p-value is 0.3903.

c. See Solution 2.3.11c. Because the p-value is greater than 0.05 we do not have strong evidence that the average time Cal Poly students watch TV is different than 2.84 hours.



Solution 2.3.11c

2.3.12 The standardized statistic should be more than 2 because a *t*-distribution has more area (probability) in the tail, we would have to move the standardized statistic out farther to reduce the probability to what you would find beyond 2 in the tail of a normal distribution.

2.3.13

a. The SPF value for sunscreen used by students at a certain school

b. μ = mean SPF of sunscreens used by all students at this school.

c. H_0: $\mu = 30$ versus H_a: $\mu > 30$

d. $n = 48, \bar{x} = 35.29, s = 17.19$

e. No, it just came from students in her class

f. They probably don't differ much from the students as a whole on this issue.

g. It may not be representative for students at a Midwestern college where it is very cloudy.

2.3.14

a. You would have to fabricate a large dataset to represent the population of SPF numbers of sunscreens for all students at the school with the variability similar to that of the sample data and a mean of 30. From that data you would take a sample of 48 and find its mean. Repeat this at least 1,000 times to develop a null distribution. To determine the p-value, determine the proportion of simulated statistics that are at or more than 35.29.

b.

Simulation		Real study
One repetition	=	A sample of 48 students
Null model	=	Population mean is 30
Statistic	=	Average SPF number in the sample

2.3.15

a. Yes, because the sample size is larger than 20

b. The standardized statistic: $t = 2.13$ and the p-value = 0.0191 (applet output is shown in Solution 2.3.15b).

c. If the mean SPF for all students at the school is 30, the probability we would get a sample mean as large or larger than 35.29 from a random sample of 48 is 0.0191.

d. We have strong evidence that the average SPF used by students at the school is more than 30.

2.3.16

a. The diameter of needles and it is quantitative

b. H_0: $\mu = 1.65$ mm H_a: $\mu \neq 1.65$ mm

c. $n = 35, \bar{x} = 1.64, s = 0.07$

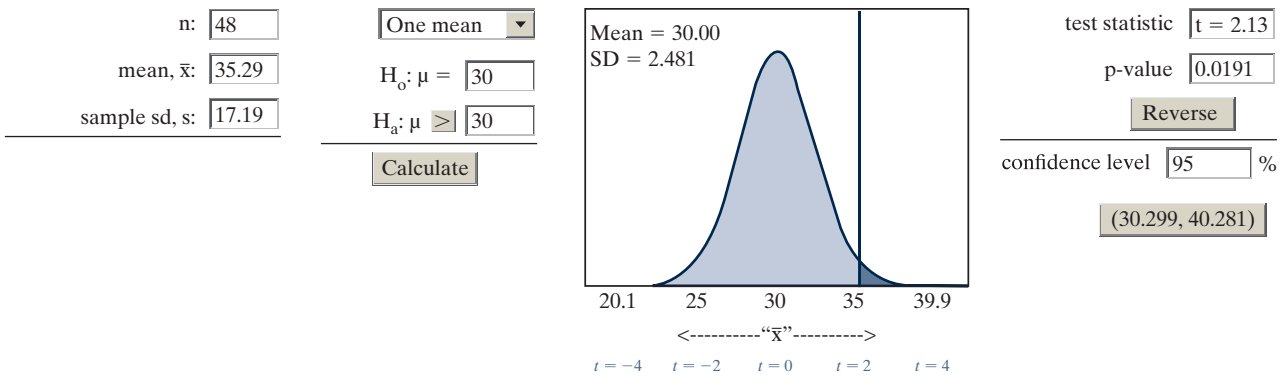
d. See Solution 2.3.16d. The mean should be about 1.65 and the standard deviation of the distribution should be about $\frac{0.07}{\sqrt{35}} \approx 0.01$.

2.3.17

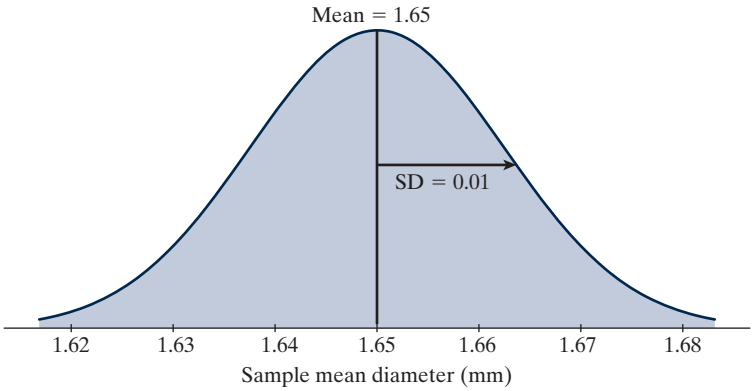
a. You would have to fabricate a large dataset to represent the population of needles with the variability similar to that of the sample data and a mean of 1.65. From that data you would take a sample of 35 and find its mean. Repeat this at least 1000 times to develop a null distribution. To determine the p-value, determine the proportion of simulated statistics that are at least as extreme or more extreme than 1.64.

b.

Simulation		Real study
One repetition	=	A random sample of 35 needles
Null model	=	Population mean is 1.65mm
Statistic	=	Average diameter of needles in the sample



Solution 2.3.15b



Solution 2.3.16d

2.3.18

a. The standardized statistic is –0.85 and the p-value is 0.4039.

b. Because the p-value is much greater than 0.05, we do not have strong evidence that the average diameter of the population of needles is different than 1.65mm.

2.3.19

a. The distribution is fairly symmetric.

b. The mean and the median will be about the same because the distribution of temperatures is fairly symmetric.

c. The mean is 98.105 and the median is 98.100. Yes, they are very close.

d. The actual standard deviation is 0.699.

2.3.20

a. Null: The average body temperature for males is 98.6°F ($\mu = 98.6$°F), Alternative: The average body temperature for males is not 98.6°F ($\mu \neq 98.6$°F).

b. The standardized statistic is -5.71 and the p-value is 0.

c. Because the p-value is less than 0.05 we have strong evidence that the average male body temperature is not 98.6°F.

d. Any generalization should be done with caution, but we can probably generalize it to healthy male adults similar to those that were in the study.

2.3.21

a. Null hypothesis: The average body temperature for females is 98.6 °F ($\mu = 98.6$°F). Alternative: hypothesis: the average body temperature for females is not 98.6°F ($\mu \neq 98.6$°F).

b. Hard to tell, there is a lot of variability in the data

c. The standardized statistic is -2.24 and the p-value = 0.0289.

d. Because the p-value is less than 0.05 we have strong evidence that the average body temperature for females is different than 98.6°F.

2.3.22

a. Null hypothesis: The average commute time in the California city is 27.5 minutes. Alternative hypothesis: The average commute time in the California city is different than 27.5 minutes.

b. The standardized statistic is –2.23 and the p-value is 0.0308.

c. Because the p-value is less than 0.05 we have strong evidence that the average commute time in the California city is different than 27.5 minutes.

d. Perhaps we can generalize just to the people like those that she and her parents know. They could easily not be representative of the city residents as a whole.

2.3.23

a. Rejecting a true null hypothesis is finding an innocent person guilty.

b. Not rejecting a false null hypothesis is finding a guilty person not guilty.

c. Our judicial system is supposed to be set up to make it difficult to find an innocent person guilty, the one described in part (a).

2.3.24

a. We have strong evidence the patient has the disease, when in fact they are healthy. The consequence is potentially frightening a patient into thinking they are diseased when, in fact, they are not.

b. We do not have evidence the patient has the disease, when in fact they are diseased. The consequence is giving a patient a false/untrue sense of security.

c. Although both mistakes are problematic, telling someone they are healthy when they are not could have serious, negative, long-term personal and community health impacts (e.g., they could give others the disease).

2.3.25

a. We have strong evidence the subject is not telling the truth.

b. It is plausible the subject is telling the truth (we cannot rule out the fact that they are telling the truth).

c. We have strong evidence the subject is not telling the truth, when in fact they are telling the truth.

d. We do not have strong evidence the subject is not telling the truth (it's plausible they are telling the truth), when in fact they are lying.

2.3.26

a. We have strong evidence the incoming message is not legitimate.

b. It is plausible the incoming message is legitimate.

c. We have strong evidence the incoming message is not legitimate, but it is legitimate.

d. It is plausible the incoming message is legitimate, when it is actually not legitimate.

2.3.27

a. We have strong evidence the new treatment is better, when it actually is not.

b. We do not have evidence the new treatment is better, when it actually is.

2.3.28

a. We find strong evidence that Buzz is not guessing, but he is guessing.

b. We do not have evidence that Buzz is not guessing (guessing is plausible), when Buzz is actually not guessing.

2.3.29 A Type I error is possible here, which means that we conclude there is strong evidence Buzz is not guessing, when in fact Buzz is guessing.

2.3.30

a. The true, long-run, average needle diameter (μ).

b. Null: The long-run average needle diameter is 1.65 mm ($\mu = 1.65$ mm). Alternative: The long-run average needle diameter is different than 1.65 mm ($\mu \neq 1.65$ mm).

c. We find evidence that the long-run average needle diameter is different than 1.65 mm, when it is actually 1.65 mm. The consequence is that production may stop when it should not have.

d. We do not have evidence that the long-run average needle diameter is different from 1.65 mm, when it is actually different than 1.65 mm. The consequence is producing needles with the wrong diameter and selling them in the marketplace, which ultimately may be bad for business.

2.3.31 Type I error would be the producer's risk because they risk shutting down manufacturing when it should not have and Type II error would be the consumer's risk because they are risking using needles which (purportedly) have an average diameter of 1.65 mm when the true average may not be 1.65 mm.

2.3.32 If a dataset is symmetric, the mean and median will be about the same and (*mean* – *median*) will be about 0. As the dataset gets more skewed to the high numbers the mean will get larger than the median and (*mean* – *median*) will get larger. Similarly as the dataset

gets more skewed to the low numbers the mean will get smaller than the median and (*mean* – *median*) will get smaller. When we divide by the SD we standardize this difference and thus this makes this statistic a good measure of skewedness.

2.3.33 Deviations from the average always sum to be zero. So if you know $n - 1$ deviations, you can figure out the other one. This means that only $n - 1$ of the numbers of a dataset are free. The last one is determined by the rest. The more degrees of freedom for a t-distribution, the more it looks like a normal distribution.

2.3.34 A p-value of 0.04999 or 0.050001 are very similar and came from very similar results. It is not like something magical happens when a p-value drops below 0.05. p-values around 0.05 are really all about the same, just as p-values close to any other number are all about the same. Therefore, we should not have dramatically different conclusions for p-values of 0.04999 and 0.050001.

Section 2.4

2.4.1 **B.**

2.4.2 **A.**

2.4.3 **B.**

2.4.4 **B.**

2.4.5 **C.**

2.4.6 **B.**

2.4.7 **C.**

2.4.8 **D.**

2.4.9 **B.**

2.4.10 **A.**

2.4.11 **B.**

2.4.12 **A.**

2.4.13 **A.**

2.4.14 **B.**

2.4.15

a. Get larger

b. Stay the same

c. Get larger

d. Get larger

2.4.16

a. False

b. True

c. False

d. True

2.4.17

a. True

b. False

c. True

d. False

2.4.18

a. The mean and median would increase by 5 points; the standard deviation would stay the same because the entire distribution would move up but the variability stays the same.

b. The mean would increase because 5 would be added to the total of the scores. The median would stay the same because the high score remains the high score. The standard deviation would increase because there would be more variability with the larger number.

c. The mean would increase because 5 would be added to the total of the scores. We cannot really say for certain how or whether the median or standard deviation would change. It depends on the original distribution of scores. For example, if adding the 5 points to the lowest score could still keep it the lowest score or it could change it to the highest score.

2.4.19 No, the distribution of means is not symmetric but is skewed right. Therefore, a *t*-distribution will not model this well.

2.4.20 Randomly choose a coin out of your collection, note its year, and put it back. Do this same thing 99 more times so you have a sample of 100 years. Find the mean of those 100 years. Repeat this process many, many times to develop a bootstrap sampling distribution.

2.4.21

a. (3, 3, 3), (3, 3, 6), (3, 6, 6), (3, 3, 9), (3, 9, 9), (6, 6, 6), (6, 6, 9), (6, 9, 9), (9, 9, 9), (3, 6, 9)

b. There are seven possible means: 3, 4, 5, 6, 7, 8, 9.

c. There are only three possible medians: 3, 6, 9.

2.4.22

a. H_0: The mean body temperature for males is 98.6°F ($\mu = 98.6$°F). H_a: The mean body temperature for males is not 98.6°F ($\mu \neq 98.6$°F).

b. $\bar{x} = 98.105$°F, $s = 0.699$

c. SD ≈ 0.086

d. $(98.105 - 98.6)/0.086 = -5.76$. Because the sample mean 98.105°F is more than 5 SDs below the hypothesized mean of 98.6°F, we have very strong evidence that the mean body temperature for males is different (less) than 98.6°F.

e. As the p-value < 0.0001, we come to the same conclusion.

f. We can probably generalize to all healthy adult males in the United States between the ages of 18 and 40.

2.4.23 $98.105 \pm 2(0.086) = 97.933$°F to 98.277°F

2.4.24

a. The standardized statistic is $t = -5.71$. It is very similar to what was found using a bootstrap sampling distribution.

b. The p-value is < 0.001. It is very similar to what was found using a bootstrap sampling distribution.

2.4.25

a. H_0: The mean body temperature for females is 98.6°F ($\mu = 98.6$ °F). H_a: The mean body temperature for females is not 98.6°F ($\mu \neq 98.6$°F).

b. $\bar{x} = 98.394$°F, $s = 0.743$

c. SD ≈ 0.092

d. $(98.394 - 98.6)/0.092 = -2.24$. Because the sample mean 98.394°F is more than 2 SDs below the hypothesized mean of 98.6°F, we have strong evidence that the mean body temperature for females is different (less) than 98.6°F.

e. As the p-value ≈ 0.022, yes, we come to the same conclusion.

2.4.26

a. The standardized statistic is $t = -2.24$. It is exactly what was found using a bootstrap sampling distribution.

b. The p-value is 0.0289. It is very similar to what was found using a bootstrap sampling distribution.

2.4.27

a. H_0: The mean increase in the laugh rating is 0 ($\mu = 0$). H_a: The mean increase in the laugh rating greater than 0 ($\mu > 0$).

b. $\bar{x} = 0.295$, $s = 0.427$

c. SD ≈ 0.067

d. As the p-value < 0.0001, we have strong evidence that the mean increase in rating is greater than 0 (or jokes are funnier with a laugh track).

2.4.28

a. H_0: The median increase in the laugh rating is 0. H_a: The median increase in the laugh rating greater than 0.

b. Median = 0.330, $s = 0.427$

c. SD ≈ 0

d. As the p-value < 0.0001, we have strong evidence that the median increase in rating is greater than 0 (or jokes are funnier with a laugh track).

2.4.29

a. H_0: Students' actual scores are the same as predicted on average or the mean difference is 0 ($\mu = 0$). H_a: The students' actual scores are lower than predicted on average or the mean difference is less than 0 ($\mu < 0$).

b. $\bar{x} = -2.481$, $s = 8.635$

c. p-value ≈ 0.07

d. We do not have strong evidence that the mean difference in scores is less than 0, or we do not have strong evidence that students' actual scores tend to be less than the predicted in the long run.

2.4.30

a. H_0: The median difference in scores is 0. H_a: The median difference in scores is less than 0.

b. Median = –1.000, $s = 8$

c. p-value ≈ 0.28

d. We do not have strong evidence that the median difference score is less than 0 in the long run (or we do not have strong evidence that students tend to predict lower scores than their actual score).

2.4.31

a. H_0: People tend to pick their own face, on average ($\mu = 0$). H_a: People tend to pick a face that is different than theirs, on average ($\mu \neq 0$).

b. $\bar{x} = 6.296$, $s = 12.449$

c. SD ≈ 2.346

d. (6.296 – 0)/2.346 = 2.68. Because the standardized statistic is more than 2, we have strong evidence that the mean score is different (greater) than 0, or people, on average, pick faces that are better looking than their own in the long run.

e. As the p-value ≈ 0.01, we come to the same conclusion.

2.4.32

a. H_0: People tend to pick their own face, on average ($\mu = 0$). H_a: People tend to pick a face that is different than theirs, on average ($\mu \neq 0$).

b. $\bar{x} = 12.000$, $s = 19.712$

c. SD ≈ 4.93

d. (12 – 0)/4.93 = 2.43. Because the standardized statistic is more than 2, we have strong evidence that the mean score is different (greater) than 0, or people, on average, pick faces that are better looking than their own in the long run when looking at mirror images of their faces.

2.4.33

a. 3.8 cups per week

b. SD ≈ 1.4, p-value ≈ 0.02

c. 5.133 cups per week, SD ≈ 2.5, p-value ≈ 0.50

d. The p-value increased. This happened because the sample mean moved closer to the value hypothesized under the null. The SD of the bootstrap sampling distribution also increased. This would also cause the p-value to increase.

e. For the original data, the median is 2, the SD of the sampling distribution is ≈ 1.3, and the p-value ≈ 0.004. For the data where the 20 changes to a 40, the median is still 2, the SD of the sampling distribution is still about 1.3, and the p-value is still about 0.004.

f. The observed median did not change because 40 is just a high number just like the 20 was. The SD of the null did not change much because again, having the 20 or 40 (or even multiples of these) in your bootstrap sample does not affect the median of that sample. Therefore, a very similar sampling distribution will be obtained. Because the observed median and the SD of the sampling distribution did not change much, the p-value will not change much.

End of Chapter 2 Exercises

2.CE.1

a. No, all games during a certain period were selected, instead of randomly choosing some.

b. Number of runs is likely more representative of the population (all games) as attendance fluctuates dramatically during the year due to weather, opponent, and timing of the games during the season.

2.CE.2

a. No, you did not make a list of all students and sample from the list.

b. Answers will vary. One possibility is blood type—unlikely that blood type is associated with whether or not you are likely to walk in front of the library.

c. Answers will vary. One possibility is GPA—students walking near the library may be more likely to have higher GPAs.

2.CE.3

a. No, the instructor did not list all games in the 2010 season and randomly choose some.

b. Null hypothesis: 75% of all major league baseball games have a "big bang" ($\pi = 0.75$). Alternative: hypothesis: Less than 75% of all games have a "big bang" ($\pi < 0.75$).

c. $\hat{p} = \frac{21}{45} = 0.467$

d. Using applet with 0.75 = probability of success, $n = 45$, and number of samples = 1000, calculate probability of 0.467 or less. The p-value is approximately 0.

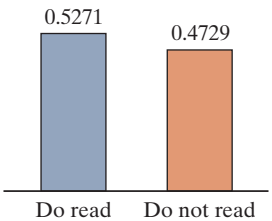
e. The data provide strong evidence that the true proportion of all major league baseball games with a big bang is less than 0.75, because it's extremely unlikely that in a sample of 45 games only 21 would have a big bang if the true proportion of all games with a big bang was 0.75.

2.CE.4

a. 263/499 = 0.527

b. statistic, because it is based on the sample. The appropriate symbol is $\hat{p}$.

c. Bar chart



d. p-value is 0.2444 (probability of heads = 0.5, number of tosses = 499, number of reps = 1,000, as extreme as 263, two-sided). We do not have evidence that the proportion of all Israelis that read while using the toilet differs from one half.

e. Although this is not a random sample and we should be cautious generalizing, generalizing to all Israelis seems reasonable if we believe that sampling individuals in public gathering areas are like those that aren't in public gathering areas. It is a judgment call as to whether generalizing to the United States or another country is reasonable because the behavior may be quite different in different cultures.

2.CE.5

a. The observational units are the 600 brides in the sample. The variable is whether or not they kept their own name.

b. The population is the U.S. brides in 2001 to 2005. The sample is the 600 brides with wedding announcements in the *NY Times*.

c. The proportion of all U.S. brides that keep their own names

d. Yes, concerned because people with wedding articles in *NY Times* are likely to be different than "typical" U.S. adults. Probably reasonable to generalize to other brides who could/would have their wedding announcement in the *NY Times*.

2.CE.6 With a p-value on 0.0396, we have strong evidence that the proportion of brides (who could/would have their wedding announcement in the *NY Times* from 2001 to 2005) that kept their name is different (or greater than) 0.15.

2.CE.7

a. Nonbiased

b. Biased; another population would be "students who visit the library"

c. Biased; another population would be "students who visit the student center"

d. Biased; another population would be "students who go to basketball games"

e. Biased; another population would be students who live on campus

f. Biased; another population would be students who have a car on campus

2.CE.8

a. Library

b. Basketball game, cars

c. Library, student center

d. Dorms

2.CE.9

a. The sample size is large, but it would be good to know that the data was not strongly skewed to feel better about running a theory-based test on the data.

b. Null: The average adult body temperature is 98.6°F. Alternative: The average adult body temperature is not 98.6°F. p-value is 0.0000, we have strong evidence that average body temperature is different than 98.6°F.

c. t-statistic of 6.32 means our sample mean is more than 6 standard deviations from 98.6.

d. Yes, the t-statistic is very different than 0 (in particular > 3) which means there is very strong evidence against the null hypothesis.

2.CE.10

a. Skewed to the left; most students have reasonably good GPAs, except a few who are struggling.

b. **i.** 3.29—as expected;

ii. 0.172—as expected;

iii. Slightly skewed to the left

c. **i.** 3.29—as expected;

ii. 0.082—as expected;

iii. Fairly bell-shaped

2.CE.11

a. 86.7, as expected; close to the population mean 86.736

b. 2.106, as expected; close to $\frac{9.608}{\sqrt{20}} = 2.148$

c. The distribution is almost symmetric; it is very slightly skewed to the left.

d. Mean ≈ 86.7, SD ≈ 1.36, nearly bell-shaped

2.CE.12

a. Smaller

b. Larger

c. Smaller

d. Larger

e. Larger

f. Smaller

2.CE.13

a. Smaller

b. Larger

c. Larger

d Smaller

e. Larger

f. Smaller

2.CE.14

a. Smaller

b. Larger

c. Smaller

d. Stays the same

e. Larger

f. Larger

g. Smaller

Chapter 2 Investigation

1. It was stated that it was a random sample, so therefore it is an unbiased sampling method.

2. Telling the respondents that a very large proportion of people fake phone calls might increase the proportion of the respondents that would admit to faking a phone call.

3. Yes, because it says it was a random sample of all cell phone users

4. It was stated that it was a random sample, so therefore it is an unbiased sampling method. However, our population now would be U.S. college students instead of all adult Americans.

5. Telling the respondents many college students use apps to help them make fake phone calls might increase the proportion of the respondents that would admit to faking a phone call.

6. No, because it was not a random sample of all cell phone users. We can generalize to the college student population that was being sampled.

7. It was not a random sample, so therefore we cannot assume that it is an unbiased sampling method.

8. Telling the respondents that a very small proportion of people fake phone calls might decrease the proportion of the respondents that would admit to faking a phone call.

9. No, hard to generalize because not a random sample

10. Have more than 1 in 10 cell phone users faked cell phone calls within the last 30 days?

11. Each of the cell phone users in the sample (1,858 in the Pew sample)

12. Whether or not the person has faked a cell phone call in the past 30 days

13. The population proportion of all American cell phone users who have faked a cell phone call within the past 30 days

14. Null: The population proportion of all American cell phone users who have faked a cell phone call within the past 30 days is 0.10.

Alternative: The population proportion of all American cell phone users who have faked a cell phone call within the past 30 days is more than 0.10.

15. Yes, every sample may yield somewhat different results because they are randomly taken.

16. Statistic, because it is based on the sample

17. 13% of the people in the study report faking a cell phone call within the past 30 days.

18. Yes, because any result for the sample statistic is possible

19. Probability of success = 0.10, sample size = 1,858, number of samples = 1,000

20. 0.10 because that's the value in the population we are simulating samples from. It makes sense that, on average, our sample proportions are equal to the population proportion.

21. Yes, we don't get exactly 0.10 every time. What we get varies.

22. Sample proportions between 0.08 and 0.12 are typical values of the sample statistic *if* the population proportion is 0.10.

23. Because the sample actually yielded 0.13 this suggests that the population proportion is not 0.10. This is convincing evidence that the population proportion of people who faked a cell phone call in the past 30 days is more than 0.10.

24. The approximate p-value is 0. This is the proportion of times we obtained 0.13 or larger when assuming the population proportion was 0.10.

25. This conclusion does not hold for people in general, but it does hold for all American cell phone users because we took a random sample of all American cell phone users. Because we took a random sample of the population of all American cell phone users, we know that the sample is representative of that population.

26. A random sample of American cell phone users gave us strong evidence that more than 10% of all American cell phone users have faked a cell phone call within the past 30 days. Further research might investigate particular demographic groups who are more/less likely to fake cell phone calls and to pursue popular reasons why people are faking cell phone calls.

Chapter 2 Research Article

1.

a. Will infants prefer a helping toy vs. a hindering toy given a choice between the two?

b. Will infants look at interactions between the climbing toy and the hindering toy longer than between the climbing toy and the helping toy?

c. Will infants choose the pushing up toy more than the pushing down toy?

d. Will infants choose the helper toy more than the neutral toy?

e. Will infants choose the neutral toy more than the hinderer toy?

f. Will infants look at interactions between the climbing toy and the helping toy longer than between the climbing toy and the neutral toy?

g. Will infants look at interactions between the climbing toy and the neutral toy longer than between the climbing toy and the hinderer toy?

2. Researchers are interested in learning whether preverbal infants (6-month-olds and 10-month-olds) assess individuals based on their interactions with others.

3. Ten-month-old babies and 6-month-old babies. All from the greater New Haven, CT, area. Little other information is available.

4. Ethnicbackground,socioeconomicstatus,socializationexperiences, and so on may be helpful.

5. The infants were "recruited" (see Methods), meaning that they probably used lists of new babies from the hospital or newspaper and contacted people asking them to participate. Detailed information on the recruiting strategy, however, is not provided in the article.

6. Experiment #1: Choose helper toy or hinderer toy? Categorical, 2 outcomes (helper, hinderer). Looking time at hindering toy (quantitative), looking time at helping toy (quantitative).

Experiment #2: Choose pusher-up toy or pusher-down toy? Categorical, 2 outcomes (pusher-up, pusher-down).

Experiment #3: Choose neutral or hinderer toy? Categorical, 2 outcomes (neutral, hinderer). Choose neutral or helper toy? Categorical, 2 outcomes (neutral, helper). Looking time neutral toy (when also shown helper; quantitative), looking time helper toy (quantitative). Looking time neutral toy (when also shown hinderer; quantitative), looking time hinderer toy (quantitative).

7. 10-month-olds: 14/16, 6-month-olds: 12/12 prefer helper vs. hinderer

8. 10-month-olds: average length of look at helper toys: 3.82 s, average length of look at hinderer toys: 4.96 s; 6-month-olds: average length of look at helper toys: 6.7 s, average length of look at hinderer toys: 5.7 s

9. 10-month-olds: 6/12, 6-month-olds: 4/12 prefer pusher up vs. pusher down

10. 10-month-olds: 7/8, 6-month-olds: 7/8 prefer helper vs. neutral

11. 10-month-olds: 7/8, 6-month-olds: 7/8 prefer neutral vs. hinderer

12. The p-value is 0.002. We have strong evidence that (in the long run) 10-month-old infants prefer the helper toy over the hinderer toy.

13. The p-value is 0.0002. We have strong evidence that (in the long run) 6-month-old infants prefer the helper toy over the hinderer toy.

14. The validity conditions are not met. In particular, there were only two 10-month-old infants who chose the helper (not 10 or more), and there were no 6-month-old infants who chose the helper (not 10 or more).

15. Quantitative

16. We have strong evidence that the average difference in looking times of the 10-month-old infants between the helper and hinderer was different than zero.

17. With a p-value of 0.44, we do not have strong evidence that the average difference in looking times is different than zero for the 6-month-olds

18. No, they make it sound like they've proven the null hypothesis is true (average difference in looking times is zero).

19. Infants like those in the study (similar background, etc.); however, because we don't know much about the infants in the study, it's difficult to be more specific

20. If the characteristics of the helper/hinderer (color/shape) are not counter-balanced, then you don't know if the preference of the infants is for the color/shape or for the helper/hinderer. It is ideal to counter-balance as many possibly important characteristics of the experiments as possible to rule them out as possibly explaining the significant preference of the infants.

21. The first sentence of the final paragraph of the paper summarizes two key findings well: "Our findings indicate that humans engage in social evaluation far earlier in development than previously thought, and support the view that the capacity to evaluate individuals on the basis of their social interactions is universal and unlearned."

The first aspect of this conclusion seems reasonable given the prior research cited in this article and our belief that the study was conducted in a fairly sound manner. However, the second conclusion may be a bit of a stretch based on this experiment alone and would require evidence from other studies to support the statement more fully; though, this final point is certainly open to debate.

22. The demographic profile of the infants used (ethnicity; socioeconomic); the supposition that the choice of toys will translate into choices of partners, friends, and so on, among others

23. Two ideas (there are many more) are to:

a. Try the study on infants with different demographic profiles to argue that it is universal behavior, and not only observed among a particular ethnic or socioeconomic strata.

b. Use live subjects (young children) instead of inanimate objects and see whether the preferences persist.