Significance: How Strong Is the Evidence?

Section 1.1

1.1.1 **C.**

1.1.2 **A.**

1.1.3 **C.**

1.1.4 **A.**

1.1.5 **B.**

1.1.6 **D.**

1.1.7

a. The true proportion of times the racket lands face up

b. Parameter

c. 0.50

d. 48 out of 100 does not constitute strong evidence that the spinning process is not fair, because if the spinning process was fair (50% chance of racket landing face up), getting 48 out of 100 spins landing face up is a typical result.

e. Plausible that the spinning process is fair

1.1.8

a. 24 out of 100 does constitute strong evidence that the spinning process is not fair, because if the spinning process was fair (50% chance of racket landing face up), getting 24 out of 100 spins landing face up is an atypical result.

b. Statistically significant evidence that spinning is not fair

1.1.9

a. LeBron's long-run proportion of making a field goal

b. Statistic

c. 0.50

d. Flip a coin 1,095 times and record the number of heads. Repeat this 1,000 times keeping track of the number of heads in each set of 1,095.

e. Approximately ½ of 1,095 (547 or 548) will be one of the most likely values.

1.1.10

a. Parameter

b. Statistic

c. 1,383

d. The number of heads (or proportion of heads) out of 1,383 flips

e. ½ of 1,383 or about 691 or 692

1.1.11

a. Parameter

b. Statistic

c. The correct matches are shown below:

Column A	Column B
Coin flip	Kevin shoots a field goal
Heads	Kevin makes his field goal
Tails	Kevin misses his field goal
Chance of heads	Long-run proportion of field goals Kevin makes
One repetition	One set of 100 field goal shots by Kevin

1.1.12

a. Parameter

b. Statistic

c. The correct matches are shown below:

Column A	Column B
Coin flip	Author plays a game of Minesweeper
Heads	Author wins a game
Tails	Author loses a game
Chance of heads	Long-run proportion of games that the author wins
One repetition	One set of 20 Minesweeper games played by author

1.1.13

a. 100 dots

b. Each dot represents the number of times out of 20 attempts the author wins a game of Minesweeper when the probability that the author wins is 0.50.

c. 10, because that is what will happen on average if the author plays 20 games and wins 50% of her games.

d. No, we are not convinced that the author's long-run proportion of winning at Minesweeper is above 0.50 because 12 is a fairly typical outcome for the number of wins out of 20 games when the long-run proportion of winning is 0.50. Stated another way, 0.50 is a plausible value for the long-run proportion of games that the author wins Minesweeper based on the author getting 12 wins in 20 games.

e. No, 0.50 is just a plausible (reasonable) explanation for the data. Other explanations are possible (e.g., the author's long-run proportion of wins could be 0.55).

f. Yes, it means that there were special circumstances when the author played these 20 games and so these 20 games may not be a good representation of the author's long-run proportion of wins in Minesweeper.

1.1.14

a. 100

b. Each dot represents the number of times out of 10 attempts the toast lands buttered side down when the probability that the toast lands buttered side down is 0.50.

c. 5, because that is what will happen on average if the toast is dropped 10 times and 50% of the drops it lands buttered side down.

d. No, we are not convinced that the long-run proportion of times the toast lands buttered side down is above 0.50 because 7 is a fairly typical outcome for the number of times landing buttered side down out of 10 drops of toast when the long-run proportion of times it lands buttered side down is 0.50. Stated another way, 0.50 is a plausible value for the long-run proportion of times that the toast lands buttered side down based on getting 7 times landing buttered side down in 10 drops.

e. No, 0.50 is just a plausible (reasonable) explanation for the data. Other explanations are possible (e.g., the long-run proportion of times the toast lands buttered side down could be 0.60).

1.1.15

a. Statistic

b. Parameter

c. Yes, it is possible to get 17 out of 20 first serves in if Mark was just as likely to make his first serve as to miss it.

d. Getting 17 out of 20 first serves in if Mark was just as likely to make the serve as to miss it is like flipping a coin 20 times and getting heads 17 times. This is fairly unlikely, so 17 out of 20 first serves in is not a very plausible outcome if Mark is just as likely to make his first serve as to miss it.

1.1.16

a. Observational unit is each cup, variable is whether the tea or milk was poured first.

b. The long-run proportion of times the woman correctly identifies a cup

c. 8, $\hat{p} = 1.0$

d. Yes, it's possible she could get 8 out of 8 correct if she was just randomly guessing with each cup.

e. Getting 8 out of 8 correct if she was randomly guessing is like flipping a coin 8 times and getting heads every time—a fairly unlikely result. Thus, 8 out of 8 seems unlikely.

1.1.17

a. Toss a coin 8 times to represent the 8 cups of tea. Heads represents a correct identification of what was poured first, tea or milk, and tails represents an incorrect identification of what was poured first. Count the number of heads in the 8 tosses, this represents the number of correct identifications of what was poured first out of the 8 cups. Repeat this process many times (1,000). You will end up with a distribution of the number of correct identifications out of 8 cups when the chance of a correct identification is 50%. If 8 correct out of 8 cups rarely occurs, then it is unlikely that the woman was just guessing as to what was poured first.

b. Using the applet shows that 8 out of 8 occurs rarely by chance (~4 times out of 1,000), confirming the fact that 8 out of 8 is quite unlikely to occur just by chance.

c. Yes, the simulation analysis gives strong evidence that the woman is not simply guessing. If she were guessing she'd rarely get 8 out of 8 correct.

d. Statistically significance evidence she is not guessing

1.1.18

a. The conclusion you've drawn is incorrect, because 5 out of 8 is a likely result if someone is just guessing. In particular, when you do a simulation with probability of success = 0.5, sample size (n) = 8, getting 5 heads happens quite frequently.

b. No, this does not prove that you cannot tell the difference. It's plausible (believable) you are not guessing, but we haven't proven it.

c. Applet inputs are: probability of success (π) = 0.5, sample size (n) = 16, number of samples = 1000. Applet output suggests that 14 out of 16 is a fairly unlikely result (~2 out of 1000 times). Thus, this result also provides strong evidence that the person actually has ability better than random guessing. The applet value for π stays the same because 0.50 still represents guessing, and $n = 16$ now because there are 16 cups of tea.

1.1.19

a. The long-run proportion of times that Zwerg chooses the correct object

b. Zwerg is just guessing or Zwerg is choosing the correct object because she understands the cue.

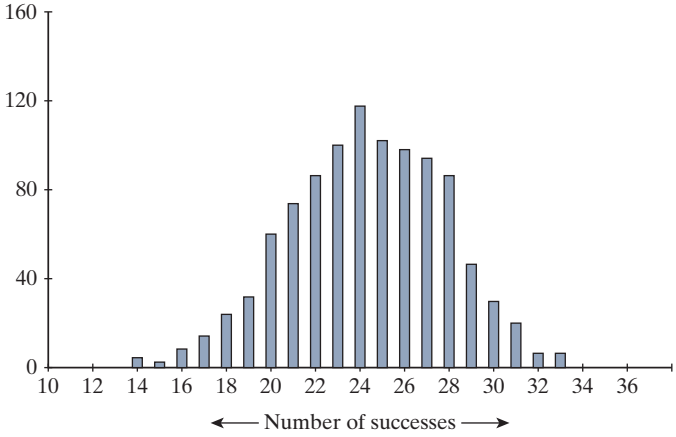
c. 37 out of 48 attempts seems fairly unlikely to happen by chance, because 24 out of 48 is what we would expect to happen in the long run.

d. 50%

1.1.20

a. 37 times out of 48 attempts

b. Applet input: probability of success is 0.5, sample size is 48, number of samples is 1,000



c. Yes, it appears as if the chance model is wrong, as it is highly unlikely to obtain a value as large as 37 when there is a 50% chance of picking the correct object.

d. We have strong evidence that Zwerg can correctly follow this type of direction more than 50% of the time.

e. The results are statistically significant because we have strong evidence that the chance model is incorrect.

1.1.21

a. Zwerg is just guessing or Zwerg is picking up on the experimenter cue to make a choice.

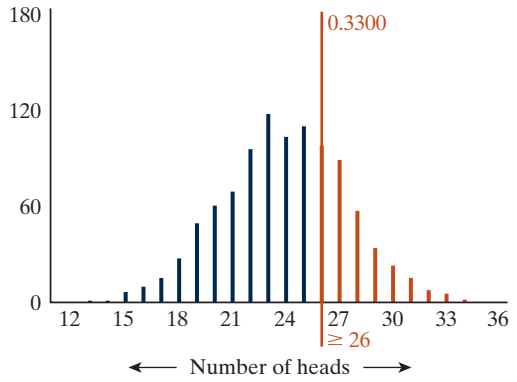
b. 26 out of 48 seems like the kind of thing that could happen just by chance because 24 out of 48 is what we would expect on average in the long run.

c. 50%

1.1.22

a. 26 times out of 48 attempts

b. Applet input: probability of success is 0.5, sample size is 48, number of samples is 1,000. This distribution is centered at 24.



c. We cannot conclude the chance model is wrong because a value as large or larger than 26 is fairly likely.

d. We do not have strong evidence that Zwerg can correctly follow this type of direction more than 50% of the time.

e. The chance model (Zwerg guessing) is a plausible explanation for the observed data (26 out of 48), because the observed outcome was likely to occur under the chance model.

f. Less convincing evidence that Zwerg can correctly follow this type of direction more than 50% of the time. We could have anticipated this because 26 out of 48 is closer to 24 out of 48 than is 37 out of 48.

g. This does not prove that Zwerg is just guessing. Guessing is just one plausible explanation for Zwerg's performance in this experiment. We cannot rule out guessing as an explanation for Zwerg getting 26 out of 48 correct.

1.1.23

a. The long-run proportion of times that Janine's short serve lands in bounds when serving left-handed

b. Janine has a 50-50 chance of landing in- bounds and so 23 out of 30 happened by chance; Janine's chance of landing her serve in bounds is greater than 50%.

c. 23 out of 30 seems somewhat unlikely to occur if she has a 50-50 chance of landing the serve in-bounds

d. 50%

1.1.24

a. 23 out of 30 attempts

b. Applet input: probability of success is 0.5, sample size is 30, number of samples is 1,000. Centered ~15

c. Yes, it appears as if the chance model is wrong, as it is highly unlikely to obtain a value as large as 23 when there is a 50% chance of getting the serve in-bounds.

d. We have strong evidence that Janine can land the majority of her serves in bounds.

e. The results are statistically significant because we have strong evidence that the chance model is incorrect.

1.1.25

a. Janine has a 50-50 chance of landing in- bounds and so 17 out of 30 happened by chance; Janine's chance of landing her serve in-bounds is greater than 50%.

b. 17 out of 30 seems like the kind of thing that could happen just by chance because 15 out of 30 is what we would expect on average in the long run.

c. 50%

1.1.26

a. 17 times out of 30 attempts

b. Applet input: probability of success is 0.5, sample size is 30, number of samples is 1,000. The distribution is centered at 15.

c. We cannot conclude the chance model is wrong because a value as large or larger than 17 is fairly likely.

d. We do not have strong evidence that Janine can land the majority of her serves in-bounds when serving right-handed.

e. The chance model (Janine lands 50% of her serves in-bounds when serving right-handed) is a plausible explanation for the observed data (17 out of 30).

f. This does not prove that Janine lands 50% of her right-handed short-serves in bounds. This is just one plausible explanation for Janine's performance. We cannot rule out a 0.50 long-run proportion of serves in bounds as an explanation for Janine landing 17 out of 30 serves in bounds.

1.1.27

a. 0.50

b. 20

c. 1,000 (or some large number)

d. 12 out 20 is a fairly likely value because it occurred frequently in the simulated data.

1.1.28

a. 0.50

b. 100

c. 1,000 (or some large number)

d. 60 out of 100 is somewhat unlikely because it occurred somewhat infrequently in the simulated data.

e. The sample size was different (20 serves vs. 100 serves).

1.1.29 **B.**

1.1.30

a. 12 coin flips

b. It is unlikely, because at least 11 correct (heads) happened very rarely on 12 coin flips.

c. Yes, we have very strong evidence that Milne does better than just guess because 11 correct out of 12 is very unlikely to happen by just guessing.

d. Even stronger evidence, because 12 correct is even more extreme than 11 correct.

e. Compared to 11, 12 is farther away in the tail and farther away from 6 (which is ½ of 12).

1.1.31

a. 20 times

b. Yes, because 16 heads out 20 coin flips is very rare.

c. Yes, because 16 heads in 20 tosses is very rare, so we have strong evidence that Mercury's choices are not random.

1.1.32

a. 20 times

b. No, because 11 heads in 20 coin flips is not unusual.

c. No, we do not have evidence that Panzee's choices are that different from what could have happened if Panzee was just randomly choosing containers.

1.1.33

a. We would flip 40 coins where one side (say heads) represents Donaghy's foul calls favoring the team that received heavier betting and the other side (say tails) represents Donaghy's foul calls not favoring the team that received heavier betting.

b. Because 28 out of 40 is way out in the upper tail of the chance distribution, it is very unlikely that Donaghy's foul calls would favor the team that received heavier betting 28 out of 40 times if the chance model is true.

1.1.34

a. We are assuming that Buzz's probability of choosing the correct button does not change and that previous trials don't influence future guesses.

b. The parameter is his actual probability of pushing the correct button.

Section 1.2

1.2.1 **D.**

1.2.2 **C.**

1.2.3 **A.**

1.2.4 **B.**

1.2.5 **A.**

1.2.6 **D.**

1.2.7 **C.**

1.2.8 **B.**

1.2.9 **B.**

1.2.10 **C.**

1.2.11 **B.**

1.2.12 **A.**

1.2.13

a. 0.25

b. 25 (because $0.25 \times 100 = 25$)

1.2.14

a. H_0 = Null hypothesis

b. H_a = Alternative hypothesis

c. $\hat{p}$ = sample proportion

d. π = long-run proportion (parameter)

e. n = sample size

1.2.15 $\hat{p}$ is the value of the observed statistic, while the p-value is the probability that the observed statistic or more extreme occurs if the null hypothesis is true; p-value is a measure of strength of the evidence.

1.2.16

a. The long-run proportion of times that Sarah chooses the correct photo, π

b. $7/8 = 0.875$.

c. Null: The long-run proportion of times Sarah chooses the correct photo is 0.50. Alternative: The long-run proportion of times Sarah chooses the correct photo is more than 0.50.

$$H_0: \pi = 0.50, H_a: \pi > 0.50$$

d. Because 4 of the 100 simulated outcomes gave a result of 7 or more, the p-value is 0.04.

e. We have strong evidence that Sarah is not simply guessing, because 7 out 8 rarely occurs by chance (if just guessing)

f. If Sarah doesn't understand how to solve problems and is just guessing at which picture to select, the probability she would get 7 or more correct out of 8 is 0.04.

g. A single dot represents the number of times Sarah would choose the correct picture (out of 8) if she were just guessing.

1.2.17

a. Null: The long-run proportion of times Hope will go to the correct object is 0.50, Alternative: The long-run proportion of times that Hope will go to the correct object is more than 0.50

b. $H_0: \pi = 0.50$, $H_a: \pi > 0.50$

c. 0.23 (23 dots are 0.60 or larger)

d. No, the approximate p-value is 0.23, which provides little to no evidence that Hope understands pointing.

e. 0.70

f. i.

1.2.18 Researcher A has stronger evidence against the null hypothesis because his p-value is smaller.

1.2.19

a. Roll a die 20 times, and keep track of how many times 'one' is rolled. Repeat this many times.

b. Using a set of five black cards and one red card, shuffle the cards and choose a card. Note the color of the card and return it to the deck. Shuffle and choose a card 20 times keeping track of how many times the red card is selected. Repeat this many times.

c. Roll 30 times, then repeat.

d. Shuffle and choose a card 30 times, then repeat.

e. Roll a die 20 times, and keep track of how many times a 'one, two, three, or four' is rolled. Repeat this many times.

f. Using a set of one black card and two red cards, shuffle the cards and choose a card. Shuffle and choose a card 20 times keeping track of how many times the red card is selected. Repeat this many times.

1.2.20

a. Observational units: 40 heterosexual couples who agreed on their response to which person was the first to say "I love you," Variable: Whether the man or woman said "I love you" first; this is a categorical variable

b. Null: The proportion of all couples where the male said "I love you" first is 0.50. Alternative: The proportion of all couples where the male said "I love you" first greater than 0.50

c. π is the proportion of all couples where the male said "I love you" first

d. $28/40 = 0.7$ is the sample proportion; we use the symbol $\hat{p}$ to denote this quantity.

e. Flip a coin 40 times and keep track of the number of heads. Repeat the 40 coin flips, 1,000 times. Calculate the proportion of sets of 40 coin flips where 28 or more heads were obtained. That proportion is the p-value.

f. Applet: $\pi = 0.50$, $n = 40$, number of samples = 1,000. To find the p-value, we find the proportion of times a value greater than or equal to 28 is observed. The p-value is approximately 0.008.

g. The p-value is the probability of observing a value of 28 or greater, assuming that for 50% of couples the man said "I love you" first.

h. The small p-value gives us strong evidence that for more than 50% of couples the man said "I love you" first.

1.2.21

a. Obs units = university students, Variable = male or female said "I love you" first

b. Null: The long-run proportion of university student relationships in which the male says "I love you" first is 0.50, Alternative: The long-run proportion of university student relationships in which the male says "I love you" first is more than 0.50,

c. 59/96 = 0.61 is the sample proportion, we use the symbol $\hat{p}$ to denote this quantity.

d. We could flip a coin 96 times and keep track of the number of heads. Then do many, many more sets of 96 coin flips, keeping track of the number of heads each time.

e. Probability of heads: 0.5, number of tosses: 96, number of repetitions: 1,000. To find the p-value, we find the proportion of times a value is greater than or equal to 0.61. The p-value is approximately 0.016.

f. The p-value (0.016) is the probability of 0.61 or larger assuming the null hypothesis is true.

g. We have strong evidence that the long-run proportion of university student relationships in which the male says "I love you" first is more than 0.50.

1.2.22

a. Obs units: each of the 40 monkeys. Variable: correct choice or not (categorical)

b. The long-run proportion of times that a monkey will make the correct choice, π

c. 30/40 = 0.75. Statistic. We use the symbol $\hat{p}$ to denote this quantity.

d. Null hypothesis: The long-run proportion of times that rhesus monkeys make the correct choice when observing the researcher jerk their head is 0.50 (just guessing). Alternative hypothesis: The long-run proportion of times that rhesus monkeys make the correct choice is more than 0.50.

$$H_0: \pi = 0.50, H_a: \pi > 0.50$$

e. Flip a coin 40 times and record the number of heads. Repeat this process 999 more times, yielding a set of 1,000 values of the number of heads received in 40 coin tosses. Compute the p-value as the proportion of times 30 or larger was obtained by chance in the 1,000 sets of 40 coin tosses. If the p-value is small (indicating 30 or larger rarely occurs by chance), then this is convincing evidence that rhesus monkeys can interpret human gestures better than by random chance.

f. The approximate p-value from the applet (using $\pi = 0.50$, $n = 40$, number of samples = 1,000) is 0.001 (probability of 30 or greater). This small p-value means that 30 out of 40 is strong evidence that the rhesus monkeys are not guessing, which may lead us to believe that rhesus monkeys may be able to understand a head jerk to indicate which box to choose.

1.2.23

a. Obs units: each of the 40 monkeys. Variable: correct choice or not (categorical)

b. The long-run proportion of times that a monkey will make the correct choice, π

c. 31/40 = 0.775. Statistic. We use the symbol $\hat{p}$ to denote this quantity.

d. Null hypothesis: The long-run proportion of times that rhesus monkeys make the correct choice when the researcher looks towards the correct box is 0.50 (just guessing). Alternative hypothesis: The long-run proportion of times that rhesus monkeys make the correct choice is more than 0.50.

$$H_0: \pi = 0.50, H_a: \pi > 0.50$$

e. Flip a coin 40 times and record the number of heads. Repeat this process 999 more times, yielding a set of 1,000 values of the number of heads received in 40 coin tosses. Compute the p-value as the proportion of times 31 or larger was obtained by chance in the 1,000 sets of 40 coin tosses. If the p-value is small (indicating 31 or larger rarely occurs by chance), then this is convincing evidence that rhesus monkeys can interpret human gestures better than by random chance.

f. The approximate p-value from the applet (using $\pi = 0.50$, $n = 40$, number of samples = 1,000) is 0.001 (probability of 31 or greater). This small p-value means that 31 out of 40 is strong evidence that the rhesus monkeys are not guessing, which may lead us to believe that rhesus monkeys can interpret gestures to indicate which box to choose.

1.2.24 The p-value is approximately 0.25. We don't have strong evidence that the author's long-run proportion of wins in Minesweeper is greater than 0.50. The null hypothesis (long-run proportion of wins is 0.50) is a plausible explanation for her winning 12 out of 20 games.

1.2.25 The p-value is approximately 0.134. We don't have strong evidence that the author's long-run proportion of wins in Spider Solitaire is greater than 0.50. The null hypothesis (long-run proportion of wins is 0.50) is a plausible explanation for him winning 24 out of 40 games.

1.2.26

a. The long-run proportion of times a spun penny lands heads

b. The p-value = 0.16. There is little to no evidence that a spun penny lands heads less than 50% of the time.

c. Null would be the same, Alternative would be > 0.50. To calculate the p-value, find the probability that 29 or larger (0.58 or larger) occurred.

1.2.27

a. Null: $\pi = 0.50$, Alternative: $\pi > 0.50$

b. The p-value = 0.31. There is little to no evidence that a coin that starts out heads will land heads more than 50% of the time.

c. No, the p-value does not prove the null hypothesized value (0.50) is correct, just that it is a plausible value for the parameter.

d. In the long run it will land heads 51% of the time, in any particular set of 100 flips it is likely the coin won't land heads exactly 51 out of 100 times.

1.2.28 A simulation analysis using a null hypothesis probability of 0.75 yields a p-value of 0.10, meaning that the set of 20 free throws by your friend (and making 12/20 of them) provides little to no evidence that your friend's long-run proportion of free throws made is worse than the NBA average.

1.2.29 A simulation analysis using a null hypothesis probability of 0.75 yields a p-value of 0.02, meaning that the set of 40 free throws by your friend (and making 24/20 of them) provides strong evidence that your friend's long-run proportion of free throws made is worse than the NBA average.

1.2.30

a. Null: Mercury will randomly choose between two containers versus Alternative: Mercury will choose the container with more bananas more often than the other container.

b. Null, $H_0: \pi = 0.50$ versus Alternative, $H_a: \pi > 0.50$

c. 0.02

d. Yes, because very rarely (p-value = 0.02) would Mercury pick the container with more bananas at least 16 times in 20 trials by just randomly choosing.

e. Answers may vary depending on what you consider "strong evidence." For a p-value to be at most 0.05, the sample proportion would have to be at least 0.75; so, 15 or more out of 20.

1.2.31

a. Null, H_0: $\pi = 0.50$ versus Alternative, H_a: $\pi > 0.50$

b. About 0.06

c. We have only moderate evidence because the p-value is fairly small, but not small enough to be considered strong evidence.

1.2.32

a. Null, H_0: $\pi = 0.25$ versus Alternative, H_a: $\pi > 0.25$

b. About 0.003

c. We have very strong evidence because the p-value is very small; it is unlikely that 42 correct matches out of 113 would happen by just guessing.

1.2.33

a. Null, H_0: $\pi = 0.25$ versus Alternative, H_a: $\pi > 0.25$

b. About 0.09

c. We have only moderate evidence because the p-value is somewhat small; it is somewhat unlikely that 35 correct matches out of 113 would happen by just guessing, but not unlikely enough to give us strong evidence.

1.2.34 When there are many outcomes, the probability of a single outcome, even if it is close to what we would expect, can be very small. We need to look at more outcomes to get a better understanding of the likeliness or unlikeliness of an outcome.

1.2.35

a. He claims he can flip a coin and make it come up heads every single time.

b. Five

Section 1.3

1.3.1 **C.**

1.3.2 $(0.45 - 0.30)/0.091 = 1.65$

1.3.3 **A**, because the standard deviation is smaller

1.3.4

D. Even though C is farther away from 0 ($z = 3$), because a positive standardized statistic puts the observed statistic in the right tail, this means the observed statistic was a number (much) larger than 0.25, which is not evidence for the alternative hypothesis that $\pi < 0.25$.

1.3.5

a. False **b.** True **c.** False **d.** False

1.3.6 **B.**

1.3.7 **C.**

1.3.8 **A.**

1.3.9 **D.**

1.3.10

a. 0.06

b. No, greater, 0.05

c. 1.72

d. No, less, 2

1.3.11

a. −3.47 (100 out of 400; 25%), −3.80 (20 out of 120; 16.7%), −4.17 (65 out of 300; 21.7%). Because of the random nature of obtaining the SDs of the null distributions, these standardized statistics can vary by as much as about plus or minus 0.2.

b. 65 out of 300 is the strongest evidence, 100 out of 400 is the least strong evidence.

1.3.12

a. Friend D because they played more games

b. Friend D because this is more evidence against the null hypothesis

c. Friend D because this is more evidence against the null hypothesis

d. Friend D because a smaller standard deviation leads to a larger standardized statistic

1.3.13 Friend G because the value of their statistic (30 out of 40) is larger than Friend F (15 out of 40) and thus will be farther in the tail of the distribution.

1.3.14 Simulation yields a standard deviation of the null distribution of 0.112, and a standardized statistic of approximately 0.89, which provides little or no evidence that the author's long-run proportion of wins in Minesweeper is higher than 0.50.

1.3.15 Simulation yields a standard deviation of the null distribution of 0.125, and a standardized statistic of approximately $(0.9375 - 0.50)/0.125 = 3.5$, which provides very strong evidence that the long-run proportion of Buzz pushing the correct button is higher than 0.50.

1.3.16 Simulation yields a standard deviation of the null distribution of approximately 0.094, and a standardized statistic of approximately 0.76, which provides little to no evidence that the long-run proportion of times Buzz pushes the correct button is higher than 0.50.

1.3.17

a. The long-run proportion of all couples that lean their heads to the right while kissing, π.

b. Null: $\pi = 0.50$, Alternative: $\pi > 0.50$

c. $80/124 = 0.645 = \hat{p}$

d. $(0.645 - 0.5)/0.045 = 3.22 = z$

e. The observed proportion of couples leaning their heads to the right while kissing is 3.22 standard deviations away from the null hypothesized parameter value of 0.5.

f. We have strong evidence that the proportion of couples that lean their heads to the right while kissing is more than 0.50.

1.3.18

a. $(0.645 - 0.60)/0.044 = 1.02$

b. The standardized statistic is smaller. This makes sense because the null hypothesis is now closer to the observed statistic (less extreme).

1.3.19

a. Null: The long-run proportion of all couples that have the male say "I love you" first is 0.50. Alternative: The long-run proportion is more than 0.50.

b. $z = (0.70 - 0.50)/0.079 = 2.53$

c. The observed proportion of couples where the males says "I love you" first is 2.53 standard deviations above the null hypothesized parameter value of 0.50.

d. We have strong evidence that the proportion of couples for which the male says "I love you" first is more than 0.50.

1.3.20

a. Null: The long-run proportion of times that rhesus monkeys choose the correct box is 0.50. Alternative. The long-run proportion of times that rhesus monkeys choose the correct box is greater than 0.50.

b. $z = (0.75 - 0.5)/0.079 = 3.16$

c. The observed proportion of rhesus monkeys that chose the box the experimenter gestured towards is 3.16 standard deviations away from the null hypothesized parameter value of 0.5.

d. We have strong evidence that the long-run proportion of times that rhesus monkeys choose the correct box is greater than 0.50.

1.3.21

a. The long-run proportion of times the lady correctly identifies which was poured first, π

b. Null: $\pi = 0.50$, Alternative: $\pi > 0.50$

c. $8/8 = 1 = \hat{p}$

d. 0.50, because that is the value of the parameter if the null hypothesis is true. The standard deviation will be positive because the standard deviation must be at least 0, and is only equal to zero if there is no variability in the values (there will be variability in the simulated statistics).

e. $z = (1 - 0.50)/0.177 = 2.82$

f. The observed proportion of times the lady correctly identified which was poured first is 2.82 standard deviations away from the null hypothesized parameter value of 0.50.

g. We have strong evidence that the long-run proportion of times that the lady makes the correct identification is greater than 0.50.

1.3.22

a. The long-run proportion of times that Zwerg makes the correct choice when the object is pointed at, π

b. Null: $\pi = 0.50$, Alternative: $\pi > 0.50$

c. $37/48 = 0.77 = \hat{p}$

d. 0.5, because that is the value of the parameter if the null hypothesis is true. The standard deviation will be positive because the standard deviation must be at least 0, and is only equal to zero if there is no variability in the values (there will be variability in the simulated statistics).

e. $(0.77 - 0.5)/0.073 = 3.70$

f. The observed proportion of times Zwerg made the correct choice is 3.70 standard deviations away from the null hypothesized parameter value of 0.5.

g. We have strong evidence that the long-run proportion of times that Zwerg makes the correct choice is greater than 0.50.

1.3.23

a. The long-run proportion of times that Zwerg makes the correct choice when using a marker, π

b. Null: $\pi = 0.50$, Alternative: $\pi > 0.50$

c. $26/48 = 0.54 = \hat{p}$

d. 0.5, because that is the value of the parameter if the null hypothesis is true. The standard deviation will be positive because the standard deviation must be at least 0, and is only equal to zero if there is no variability in the values (there will be variability in the simulated statistics).

e. $(0.54 - 0.5)/0.073 = 0.55$

f. The observed proportion of times Zwerg made the correct choice is 0.55 standard deviations away from the null hypothesized parameter value of 0.5.

g. We have little to no evidence that the long-run proportion of times that Zwerg makes the correct choice when using a marker is greater than 0.50.

1.3.24

a. The long-run proportion of times that 10-month olds choose the helper toy, π

b. Null: $\pi = 0.50$, Alternative: $\pi > 0.50$

c. $14/16 = 0.875 = \hat{p}$

d. $z = (0.875 - 0.50)/0.125 = 3$

e. The observed proportion of times the 10-month-old babies chose the helper toy is 3 standard deviations above the null hypothesized parameter value of 0.50.

f. We have strong evidence that the long-run proportion of times that 10-minth-old babies choose the helper toy is greater than 0.50.

1.3.25

a. Yes, the p-value will be small because the standardized statistic is large

b. A p-value of approximately 0.002. The p-value is the probability observing 14/16 or larger assuming the null hypothesis is true.

c. We have strong evidence that the long-run proportion of times that 10-month-old babies choose the helper toy is greater than 0.50.

d. Yes, because both the p-value and the standardized statistic are measuring the strength of evidence (how far out in the tail the observed value is), and so should lead to the same conclusion.

1.3.26

a. The long-run proportion of people that choose the number 3, π

b. Null: $\pi = 0.25$, Alternative: $\pi > 0.25$

c. $14/33 = 0.42 = \hat{p}$

d. The mean = 0.248 and SD = 0.076

e. $(0.42 - 0.248)/0.076 = 2.26$

f. The observed proportion of people that chose the number 3 is 2.26 standard deviations above the null hypothesized parameter value of 0.25.

g. We have strong evidence that the long-run proportion of people that will choose the number 3 is greater than 0.25.

1.3.27

a. Yes, because the standardized statistic is far from zero.

b. The p-value is approximately 0.02, and is the probability of observing 14/33 or larger assuming the null hypothesis is true.

c. We have strong evidence that the long-run proportion of people that will choose the number 3 is greater than 0.25.

d. Yes, because both the p-value and the standardized statistic are measuring the strength of evidence (how far out in the tail the observed value is), and so should lead to the same conclusion.

1.3.28

a. The long-run proportion of people that choose a big number, π

b. Null: $\pi = 0.50$, Alternative: $\pi > 0.50$

c. $19/33 = 0.58 = \hat{p}$

d. The mean = 0.50 and the SD = 0.084

e. $(0.58 - 0.5)/0.084 = 0.952$

f. We have little to no evidence that the long-run proportion of people that will choose a ”big number” is greater than 0.50.

1.3.29

a. No, because the standardized statistic is not far from zero.

b. The p-value is approximately 0.153, and is the probability of observing 19/33 or larger assuming the null hypothesis is true

c. We have little to no evidence that the long-run proportion of people that will choose a big number is greater than 0.50.

d. Yes, because both the p-value and the standardized statistic are measuring the strength of evidence (how far out in the tail the observed value is), and so should lead to the same conclusion.

1.3.30

a. 0.50

b. Null, H_0: $\pi = 0.50$ versus Alternative, H_a: $\pi > 0.50$

c. 0.144

d. 2.89

e. Because the standardized statistic is much larger than 2, we have convincing evidence that Milne's answers are better than random guesses.

f. Stronger evidence, because the standardized statistic is 3.46 which is farther from 0 than is 2.89.

1.3.31

a. Null, H_0: GY just guesses, $\pi = 0.50$ versus Alternative, H_a: GY is correct more often than just guessing, $\pi > 0.50$, where π represents the chance that GY guesses correctly on any trial.

b. Sample proportion; SD of the sample proportion is about 0.034.

c. 6.03

d. Very strong evidence that GY is more likely than random chance to correctly identify whether or not the square was present because the standardized statistic is so large.

e. p-value is <0.001 that is very small; yes, leads to same conclusion as the standardized statistic.

1.3.32

a. Null, H_0: GY just chooses between the two wagers randomly, $\pi = 0.50$ versus Alternative, H_a: GY chooses the higher wager more often than the lower wager, $\pi > 0.50$, where π represents the chance that GY chooses the higher wager.

b. Sample proportion; SD of the sample proportion is about 0.043.

c. –0.58

d. No evidence that GY is more likely to take the high wager than the low wager because the standardized statistic is close to zero.

e. p-value is about 0.75 that is not small at all; yes, leads to the same conclusion as the standardized statistic.

f. Because the 67/141 is 0.475 which less than 0.5, and in the opposite direction than that conjectured by the alternative hypothesis.

1.3.33

a. Null, H_0: $\pi = 0.50$ versus Alternative, H_a: $\pi > 0.50$

b. 0.003; there is about a 0.3% chance that Paul got all 8 predictions correct by just guessing.

c. Based on the very small p-value, we have very strong evidence that Paul is doing better than just guessing.

d. Based on the small p-value, it is surprising that an animal would correctly choose the winning team in 8 of 8 competitions. This is a fairly small sample size, however, and it would be interesting to see whether Paul could sustain this accuracy through a much larger number of competitions.

1.3.34

a. Null, H_0: $\pi = 0.50$ versus Alternative, H_a: $\pi > 0.50$

b. $z = (0.8667 - 0.50)/0.129 = 2.84$; because the standardized statistic is greater than 2, we have strong evidence that dogs are more likely to approach the positive donor.

c. p-value = 0.003; because this is less than 0.05, we again have strong evidence that dogs are more likely to approach the positive donor.

1.3.35

a. Null, H_0: $\pi = 0.50$ versus Alternative, H_a: $\pi > 0.50$

b. $z = (0.6667 - 0.50)/0.129 = 1.29$; because the standardized statistic is less than 2, we do not have strong evidence that dogs are more likely to approach the positive donor.

c. p-value = 0.150; because this is greater than 0.05, we again do not have strong evidence that dogs are more likely to approach the positive donor.

Section 1.4

1.4.1 **D.**

1.4.2 **A.**

1.4.3 **D.**

1.4.4 **B.**

1.4.5 **C.**

1.4.6 **A.**

1.4.7

a. Larger, because the sample proportion is closer to the hypothesized value (0.50) of the long-run proportion.

b. Smaller, because less likely to get extreme values of the statistic from a larger sample.

c. Larger, because now including at least as extreme values of the sample proportion in both tails of the null distribution.

1.4.8

a. Smaller, as in, closer to 0, because the sample proportion is closer to the hypothesized value (0.5) of the long-run proportion.

b. Larger, as in, farther from 0; because less likely to get extreme values of the statistic from a larger sample.

c. Same, because the sample proportion is still the same distance away from 0 in the null distribution.

1.4.9

a. Null, H_0: Students randomly choose between the two cookies, $\pi = 0.50$ versus Alternative, H_a: Students have a genuine preference for Chips Ahoy over Chipsters, $\pi > 0.50$, where π represents the long-run proportion of students that choose Chips Ahoy over Chipsters.

b. 0.03

c. Null, H_0: $\pi = 0.50$ versus Alternative, H_a: $\pi \neq 0.50$

d. 0.08

1.4.10

a. 0.05

b. 0.11 (0.30 or less and 0.70 or more)

1.4.11

a. Smaller

b. Smaller

c. Larger

1.4.12

a. True

b. False

1.4.13

a. Stronger. The statistic (18/20 = 0.90) is much farther away from the null hypothesized value (0.50) than before (12/20 = 0.60).

b. Stronger. The statistic is the same (0.60) but the sample size is much larger (100 vs. 20).

c. No, 12/30 = 0.40 is less than the null hypothesis value of 0.50: thus this is not evidence that the long-run proportion of wins in Minesweeper is more than 0.50.

1.4.14

a. The long-run proportion of male births is greater than 0.50.

b. The long-run proportion of male births

c. Null: The long-run proportion of male births is 0.50. Null: $\pi = 0.50$.

d. Alternative: One-sided. He wanted to demonstrate that male births outnumbered female births.

1.4.15 Probably weak because they are quite close (0.516 and 0.50)

1.4.16 Strong; the sample size is extremely large ($n = 938{,}223$).

1.4.17 Twice

1.4.18

a. Tiny

b. Huge

c. One-sided

1.4.19

a. Change: sample size. Stay the same: distance, one- or two-sided.

b. Stronger

1.4.20

a. Change: distance. Stay the same: sample size, one- or two-sided.

b. Weaker

1.4.21

a. Double it. The alternative would now be two-sided.

b. Weaker

1.4.22

a. Sample size changes; distance and one-sided are the same.

b. Stronger

1.4.23

a. Distance changes; sample size and one-sided are the same.

b. Weaker

1.4.24

a. The long run proportion of times that Krieger chooses the correct object, π.

b. 50%

c. Null: The long-run proportion of times that Krieger chooses the correct object is 0.50; Alternative: The long-run proportion of times that Krieger chooses the correct object is more than 0.50.

1.4.25

a. 6 out of 10 for a proportion of 0.60

b. Mean = 0.50, SD = 0.16

c. The p-value (The probability of obtaining 0.60 or larger when the true chance Krieger chooses the correct object is 0.05) is approximately 0.38.

d. We do not have strong evidence that Krieger will choose the correct object more than 50% of the time. It is plausible that Krieger will choose the correct object 50% of the time.

e. Stronger

1.4.26

a. Values of the long-run proportion less than 0.50

b. Increase, a two-sided p-value will be approximately twice as big as the corresponding one-sided p-value.

c. The two-sided p-value will approximately double to 0.75.

d. Stronger

1.4.27

a. Decrease, larger sample size.

b. The p-value decreased to 0.244, yes it did behave as predicted.

c. Stronger

1.4.28

a. The long-run proportion of times Krieger makes the correct choice when the experimenter leans towards the object, π.

b. 50%

c. Null: The long run proportion of times Krieger makes the correct choice when the experimenter leans towards the object is 0.50; Alternative: The long-run proportion of times Krieger makes the correct choice when the experimenter leans towards the object is more than 0.50.

d. Decrease, 9 out of 10 is farther out in the tail of the null distribution than 6 out of 10.

1.4.29

a. 9 out 10 (0.90)

b. Mean = 0.50, SD = 0.16

c. 0.01

d. Approximately 0.80

1.4.30

a. $z = (0.90 - 0.50)/0.16 = 2.5$

b. Using both the p-value and the standardized statistic, we have strong evidence that the long-run proportion of times that Krieger makes the correct choice is more than 0.50.

c. Yes, the p-value got smaller (evidence got stronger).

1.4.31

a.

	Analysis method		
Exercises	**Sample size n**	**Null value π_0**	**Value of $\hat{p}$**
A: 1.4.8 – 1.4.12	938,223	0.50	0.516
B: 1.4.25	82	0.50	1.00

b. Analysis Method A provides strong evidence, but Method B is overwhelming.

1.4.32

a. One-sided

b. Moderate. 6 out of 7 is not that strong.

c. p-value = 0.0547 + 0.0078 = 0.0625. We have moderate evidence that individuals living in the country have healthier lungs than those of individuals living in cities.

1.4.33

a. One-sided

b. Six out of 8 seems fairly likely to occur by chance; thus it is plausible that bees are just as likely to sting a target that has already been stung as they are to sting a target that is pristine.

c. p-value = 0.1094 + 0.0313 + 0.0039 = 0.1446. We have little to no evidence that bees are more likely to sting a target that has already been stung compared to a pristine target.

1.4.34

a. In a race for U.S. president, is the taller candidate more likely to win? Alternatively, is $\pi > 0.5$.

b. Null: The long-run proportion of races where the taller candidate wins in U.S. presidential elections is 0.5. Alternative: The long-run proportion of races where the taller candidate wins in U.S. presidential elections is larger than 0.5. Using symbols: H_0: $\pi = 0.5$, H_a: $\pi > 0.5$; where π is the long-run proportion of races where the taller candidate won.

c. Probability of heads: 0.5, number of tosses: 26, approximate p-value = 0.0047.

d. We have very strong evidence against the null and in support of the taller candidate winning the race more often than would be predicted by random chance.

e. It is somewhat arbitrary to only look at 20th and 21st century elections.

1.4.35

a. A: p-value = 0.3036, B: p-value = 0.0047, C: p-value = 0.9616, D: p-value = 0.1400

b. Looking at the set of p-values suggests there is little evidence that taller candidates are more likely to win. In particular, looking at all presidential elections since 1796 yields a p-value of 0.1400. Looking at an arbitrary subset of presidential elections (Exercise 1.4.34) suggested a potentially significant result, but looking at more data suggested otherwise.

1.4.36

D = Sample size is $n = 25$, alternative hypothesis is right-sided: $\pi > ½$.

A = Sample size is $n = 225$, alternative hypothesis is right-sided: $\pi > ½$.

C = Sample size is $n = 25$, alternative hypothesis is left-sided: $\pi < ½$.

B = Sample size is $n = 225$, alternative hypothesis is left-sided: $\pi < ½$.

E = Sample size is $n = 25$, alternative hypothesis is two-sided: $\pi \neq ½$.

F = Sample size is $n = 225$, alternative hypothesis is two-sided: $\pi \neq ½$.

1.4.37

a. Increasing: B,C

b. Decreasing: A,D

c. Up-down: E,F

1.4.38

a. 100%

b. 50%

c. 0%

1.4.39

a. Smaller, decreasing, A,D

b. Larger, increasing, B,C

c. Up-down, E,F

d.

Alternative hypothesis	Strongest evidence	Shape of curve
$\pi > 0.50$	$\hat{p} = 1$	Decreasing
$\pi < 0.50$	$\hat{p} = 0$	Increasing
$\pi \neq 0.50$	$\hat{p} = 0$ or $\hat{p} = 1$	Up-down

1.4.40

a. Lower

b. Larger

c. Always lies above

d. Less steep

1.4.41

	Increasing	Decreasing	Up- Down
Steeper	B	A	F
Flatter	C	D	E

1.4.42

a. That would be considered cheating. The hypotheses are always statements about the parameter(s) of interest and are stated before we begin to explore the data. We use the data to test our hypotheses, not to form them.

b. Two-sided tests always yield larger p-values than one-sided tests, making it harder to find strong evidence in support of your research conjecture.

Section 1.5

1.5.1 **C.**

1.5.2 No, the maximum standard deviation is always obtained at 0.5, regardless of sample size.

1.5.3 **D.**

1.5.4 **C.**

1.5.5 **D.**

1.5.6 **A.**

1.5.7 **C.**

1.5.8 The predicted value of the standard deviation of a null distribution of sample proportions. The prediction is accurate when the sample size is large.

1.5.9 The value of the standardized statistic, z.

1.5.10

a. Decrease

b. 0.50

c. 0 or 1

1.5.11 The standardized statistic; a measurement of how many standard deviations from the mean the observed statistic is on the null distribution

1.5.12

a. The p-value from option 1 is more valid.

b. The validity conditions are not met for this test because the light was green only 4 times (which is less than 10). We can also see this is a problem in the applet because the normal overlay does not match up nicely with the skewed null distribution.

1.5.13

a. Null: The long-run proportion of times that a penny lands heads when spun is 0.20 Alternative: The long-run proportion of times that a penny lands heads when spun is greater than 0.20

b. The simulation based p-value of 0.077, because the validity conditions are not met. There are not at least 10 times where the penny landed heads and at least 10 times where the penny landed tails in the sample.

c. No, we do not have strong evidence that a penny will land heads more than 20% of the time in the long run, because the p-value is only 0.077 (but there is *moderate* evidence).

1.5.14

a. The symbol π represents the long-run proportion of times of the coin landing heads up.

b. Null: $\pi = 0.50$. Alternative: $\pi > 0.50$

c. If the result was heads 52% of the time out 1000, then 520 must have been heads and 480 tails. Both of these are greater than 10.

d. A standardized statistic of 1.26 means that our observed proportion of 0.52 is 1.26 standard deviations above 0.50 in the null distribution.

e. We have little to no evidence that it is more likely for a coin to land the same side up as it started than not.

1.5.15

a. 0.152; this is very close to the hypothesized parameter value of 0.15.

b. No, the validity conditions are not met. There are not at least 10 successes and 10 failures in the data (only 8 and 2).

1.5.16

a. 0.15; this is the hypothesized parameter value of 0.15

b. 0.019; $\sqrt{\pi(1-\pi)/n} = \sqrt{0.15(1-0.15)/361} = 0.019$

c. Yes

d. Because the validity conditions are met for this dataset (larger sample size)

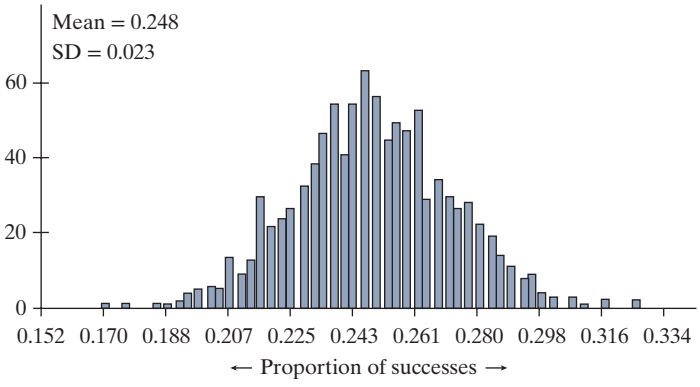
1.5.17

a. The long-run proportion of times that a person identifies the correct image

b. Null: The long-run proportion of times that a person identifies the correct image equals 0.25. Alternative: The long-run proportion of times that a person identifies the correct image greater than 0.25

c. Approximately 0.25 will identify the correct image if they have no psychic ability. This is the null hypothesis.

d. See graph below



e. The p-value is approximately 0.002 because we only got a result 0.322 or larger 2 out of 1000 times by chance. Thus, there is strong evidence that proportion of correct guesses is larger than 0.25.

f. Using the Theory-based inference applet (which uses the normal approximation for the null distribution) yields a standardized statistic of 3.02 and p-value of 0.0012, again showing strong evidence that the proportion of correct guesses is larger than 0.25. It is not surprising that the two approaches give similar results as the sample size is very much larger (in particular, there are 106 successful guesses and 223 unsuccessful guesses. Both values are much larger than 10).

1.5.18

a. 32/97 = 0.330 (Morris et al.) vs. 106/329 = 0.322 (Bern and Honorton). The proportions are about the same (Morris is just slightly more).

b. The p-value for Morris will be larger because the sample size is smaller

c. The simulation p-value is approximately 0.048; this is strong evidence that the Ganzfeld receivers choices are better than just chance

d. The p-value = 0.0346. It is similar to what we got from simulation, which is not surprising because the validity conditions are met (32 successes and 65 failures).

e. They are more similar in the previous study because the sample size is larger and, even though the validity conditions are met here, the p-values will continue to get closer and closer together as the sample size increases.

1.5.19

a. Null: The long-run proportion of times that the male says "I love you" first is 0.50.

Alternative: The long-run proportion of times that the male says "I love you" first is more than 0.50.

b. 0.0057

c. We have very strong evidence that the long-run proportion of times that the male says "I love you" first is more than 0.50

d. $z = 2.53$. That the proportion of males that say "I love you first" is 2.53 standard deviations above the mean of the null distribution.

e. $0.0057 \times 2 = 0.0114$

1.5.20

a. Null: The long-run proportion of rhesus monkeys that choose the correct box is 0.50

Alternative: The long-run proportion of rhesus monkeys that choose the correct box is more than 0.50

Null: $\pi = 0.50$

Alternative: $\pi > 0.50$

b. 0.0003

c. We have very strong evidence that the long-run proportion of Rhesus monkeys that choose the correct box is more than 0.50

d. $Z = 3.48$. That the observed proportion of rhesus monkeys that chose the correct box is 3.48 standard deviations above the mean of the null distribution.

e. $0.0003 \times 2 = 0.0006$

1.5.21

a. Null: The long-run proportion of times that a player starts with scissors is 0.333

Alternative: The long-run proportion of times that a player starts with scissors is different than 0.333

Null: $\pi = 0.333$

Alternative: $\pi \neq 0.333$

b. p-value = 0

c. We have very strong evidence that the long-run proportion of times that a player starts with scissors is different than 0.333.

1.5.22

a. Null: The long-run proportion of times that a player starts with rock is 0.333.

Alternative: The long-run proportion of times that a player starts with rock is different than 0.333.

Null: $\pi = 0.333$

Alternative: $\pi \neq 0.333$

b. p-value = 0

c. We have very strong evidence that the long-run proportion of times that a player starts with rock is different than 0.333.

1.5.23

a. Null: The long-run proportion of times that people assign the name Tim to the face on the left is 0.50.

Alternative: The long-run proportion of times that people assign the name Tim to the face on the left is more than 0.50.

b. 0.0721

c. We have moderate evidence that the long-run proportion of times that people assign the name Tim to the face on the left is more than 0.50.

d. $z = 1.46$. That the proportion of the sample that chose Tim is 1.46 standard deviations above the mean of the null distribution

1.5.24

a. Null: The long-run proportion of times that the most competent-looking candidate wins is 0.50.

Alternative: The long-run proportion of times that the most competent-looking candidate wins is more 0.50.

b. $n = 279, \hat{p} = 0.677$

c. 0

d. We have very strong evidence that the most competent-looking candidate wins more than 50% of the time.

1.5.25

a. Null: The long-run proportion of matches that the red uniform wins is 0.50.

Alternative: The long-run proportion of matches that the red uniform wins is not 0.50.

b. 0.0681

c. We have moderate evidence that the long-run proportion of matches the red uniform wins is not 0.50.

d. $z = 1.82$. That the proportion of the sample in which red won is 1.82 standard deviations above the mean of the null distribution

1.5.26

a. Null: The long-run proportion of times that the red uniform wins a boxing match is 0.50.

Alternative: The long-run proportion of times that the red uniform wins a boxing match is not 0.50.

b. 0.0896

c. We have moderate evidence that the proportion of times the red uniform wins a boxing match is not 0.50.

d. $z = 1.70$. That the proportion of the sample in which red won a boxing match is 1.70 standard deviations above the mean of the null distribution

1.5.27

a. The long-run proportion of times a six is rolled

b. Null: The long-run proportion of times a six is rolled is 0.167.

Alternative: The long-run proportion of times a six is rolled is more than 0.167.

c. 0.1497

d. We have little to no evidence that the long-run proportion of times a six is rolled is more than 0.167.

1.5.28

a. The long-run proportion of times a one is rolled.

b. Null: The long-run proportion of times a one is rolled is 0.167.

Alternative: The long-run proportion of times a one is rolled is less than 0.167.

c. 0.0840

d. We have moderate evidence that the long-run proportion of times a one is rolled is less than 0.167.

1.5.29

a. The long-run proportion of times that Mario wins

b. Null: The long-run proportion of times Mario wins is 0.50.

Alternative: The long-run proportion of times Mario wins is not 0.50.

c. p-value = 0.2733

d. We have little to no evidence that the long-run proportion of times Mario wins is not 0.50.

e. 100 games gives $z = 2$ ($p = 0.0455$). So, if out of 100 games Mario wins 60, this would be strong evidence that Mario's long-run proportion of times he wins is not 0.50.

1.5.30

a. Null, H_0: Students randomly choose a number between 1 and 10, $\pi = 0.50$ versus Alternative, H_a: Students have a genuine preference for numbers greater than 5, $\pi > 0.50$, where π represents the long-run proportion of students that a number greater than 5, when asked to choose a number between 1 and 10.

b. 0.61

c. Yes, more than 10 successes (62) and 10 failures (39)

d. 0.011

e. We have strong evidence (based on the small p-value = 0.011) that students have a genuine preference for numbers greater than 5.

1.5.31

a. Null, H_0: Cardiac arrests are just as likely on Mondays as on either Wednesdays or Fridays, $\pi = 0.33$ versus Alternative, H_a: Cardiac arrests are more likely on Mondays than on either Wednesdays or Fridays, $\pi > 0.33$, where π represents the long-run proportion of cardiac arrests among dialysis patients that happen on Mondays.

b. 0.454

c. Yes, more than 10 successes (93) and 10 failures (112)

d. < 0.001

e. 0.0001

f. Very strong evidence that of cardiac arrests among dialysis patients more than a third happen on Mondays (compared to Wednesdays or Fridays)

1.5.32

a. Null, H_0: Cardiac arrests are just as likely on Tuesdays as on either Thursdays or Saturdays, $\pi = 0.33$ versus Alternative, H_a: Cardiac

arrests are more likely on Tuesdays than on either Thursdays or Saturdays, $\pi > 0.33$, where π represents the long-run proportion of cardiac arrests among dialysis patients that happen on Tuesdays.

b. 0.342

c. Yes, more than 10 successes (65) and 10 failures (125)

d. 0.384

e. 0.3613

f. No evidence that of cardiac arrests among dialysis patients more than a third happen on Tuesdays (compared to Thursdays or Saturdays)

1.5.33

a. Null, H_0: The probability a spun coin will land on heads is 0.50, $\pi = 0.50$ versus Alternative, H_a: The probability a spun coin will land on heads is not 0.50, $\pi \neq 0.50$, where π represents the probability the spun coin will land on heads.

b. The sample statistic symbol is $\hat{p}$. Its value will vary.

c. Answers will vary.

d. Answers will vary.

e. Answers will vary.

1.5.34

a. 0.05525

b. 0.0289

c. The simulation-based p-value is more valid because the validity conditions for the normal approximation are not met here.

d. 0.0547, closer to the simulation-based p-value

1.5.35

a. Simulation-based p-value = 0.168, normal-approx.-based p-value = 0.103, exact binomial p-value = 0.1719; the simulation-based and exact binomial p-values are close to each other.

b. Simulation-based p-value = 0.378, normal-approx.-based p-value = 0.2635, exact binomial p-value = 0.377; the simulation-based and exact binomial p-values are close to each other.

End of Chapter 1 Exercises

1.CE.1 Null

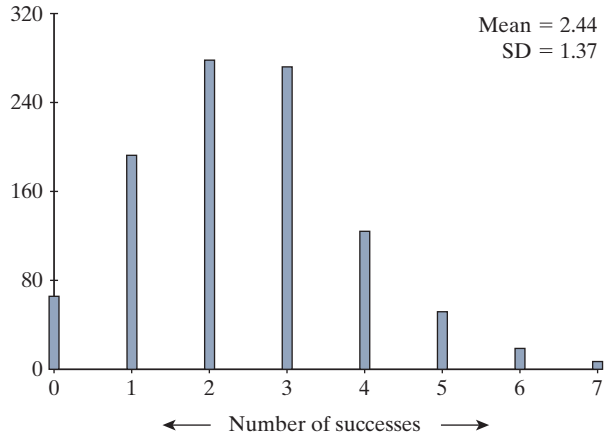
1.CE.2 Probability value in the null hypothesis

1.CE.3

a. Null: The probability the statistics professor wins is 0.02, Alternative: The probability the statistics professor wins is larger than 0.02.

b. Start with 5 playing cards—one red and four black. Shuffle the cards. Randomly choose one card, record if it is red or not, and then place the card back in the deck. Shuffle and randomly choose cards until 12 cards have been selected. Record the number of red cards selected out of the 12 selections. Repeat this entire process 999 more times to generate a distribution of counts of red cards. If 7 out of 12 red cards chosen rarely happened in the 1,000 repetitions, then this would be convincing evidence that 7 out of 12 was unlikely to have occurred by chance.

c. The observed data (7 wins out of 12 attempts) provide convincing evidence that the statistics professor's probability of winning in one week was larger than would be expected if the 5 competitor's were equally likely to win because 7 out of 12 rarely happens by chance, when everyone is equally likely to win; in particular the p-value is 0.005.



d. The p-value is the probability of observing 7 or more successes out of 12 attempts when each attempt has a 20% chance of being correct.

e. The theory-based approach is not appropriate here because the resulting simulated distribution of statistics is not normal. This is because the sample size is not large enough. In particular, there are only 7 successes and 5 failures, instead of at least 10 of each.

1.CE.4

Step 1: Ask a research question

Jamie and Adam wanted to investigate which side buttered toast prefers to land on when it falls through the air.

Step 2: Design a study and collect data

They set up a specially designed rig and dropped 48 pieces of toast from the roof of the Mythbusters' headquarters. They wish to test the following null and alternative hypotheses:

Null: There is no preference for which side the buttered toast lands on; both sided are equally likely (have a 50% chance of landing face down)

Alternative: One of the sides tends to land face down more than the other.

They recorded which sided landed down (buttered or not buttered side) for each of the 48 attempts.

Step 3: Explore the data

In 19 out of 48 attempts the buttered side landed down.

Step 4: Draw inferences

Statistic: 19/48 = 0.396

Simulation: Used applet to generate a distribution (using probability = 0.5, sample size = n = 48, number of samples = 1000), to generate a two-sided p-value of 0.19

Strength of evidence: We do not have strong evidence that one side of buttered toast tends to fall face down more often than the other.

Step 5: Formulate conclusions

We don't know if the results can be generalized to other situations (different bread? inside vs. outside? device used to drop bread?)

Step 6: Look back and ahead

While no evidence was found that one side falls to the ground more than the other, further studies are needed to ensure that the results apply to all bread, inside, and when a person drops it instead of a machine.

1.CE.5

a. Even though 34.6% (the percentage of players suspended for PED use who were from the United States) is less than the percentage of all baseball players born in the United States (57.3%), it's possible that this could have happened by chance (just like it's possible to flip a coin and get heads 8 times out of 8); the question is how likely this (34.6%) would happen just by chance. If it is quite unlikely then we say that the result is statistically significant, meaning that there is something about U.S. baseball players which make them less likely to be suspended for PED use.

b. The likelihood of having only 34.6% of 595 suspended baseball players be from the United States when the percentage of all baseball players who are from the United States is 57.3% is extremely unlikely; in other words, 34.6% is (statistically) significantly less than 57.3%.

c. $z = -11.19$. This tells us that the observed proportion (0.346) is 11.19 SDs less than the mean of the null distribution, confirming that we have extremely strong evidence that the null hypothesized value of the parameter is incorrect.

1.CE.6

a. Even though 61.8% (the percentage of players suspended for PED use who were from Latin America) is more than the percentage of all baseball players born in Latin America (34.6%), it's possible that this could have happened by chance (just like it's possible to flip a coin and get heads 8 times out of 8); the question is how likely is it that this (61.8%) would happen just by chance. If it is quite unlikely then we say that the result is statistically significant, meaning that there is something about Latin American baseball players which make them more likely to be suspended for PED use.

b. The likelihood of having 61.8% of 595 suspended baseball players be from Latin America when the percentage of all baseball players who are from the United States is 34.6% is extremely unlikely; in other words, 61.8% is (statistically) significantly more than 34.6%.

c. $z = 13.97$. This tells us that the observed percentage (61.8%) is more than 13.95 SDs less than the mean of the null distribution, confirming that we have extremely strong evidence that the null hypothesized value of the parameter is incorrect.

1.CE.7

a. Null: The long-run proportion of times that New Zealand students associate the name Tim with the face on the left is 0.50.

Alternative: The long-run proportion of times that New Zealand students associate the name Tim with the face on the left is more than 0.50.

Null: $\pi = 0.50$

Alternative: $\pi > 0.50$

b. The p-value is approximately 0, with a standardized statistic of 6.45. This is extremely strong evidence that the long-run proportion of times that New Zealand students associate the name Tim with the face on the left is more than 0.50.

c. Yes, because there are at least 10 successes (105 Tim on left) and at least 10 failures (30 Bob on left).

d. p-value = 0; $Z = 6.45$; This is extremely strong evidence that the long-run proportion of times that New Zealand students associate the name Tim with the face on the left is more than 0.50.

1.CE.8

a. Null: The long-run proportion of times that New Zealand students associate the name Bob with the face on the left is 0.50.

Alternative: The long-run proportion of times that New Zealand students associate the name Bob with the face on the left is less than 0.50.

Null: $\pi = 0.50$

Alternative: $\pi < 0.50$

b. The p-value is approximately 0, with a standardized statistic of –6.45. This is extremely strong evidence that the long-run proportion of times that New Zealand students associate the name Bob with the face on the left is less than 0.50.

c. Yes, because there are at least 10 successes (30 Bob on left) and at least 10 failures (105 Tim on left).

d. p-value = 0; $z = -6.45$; This is extremely strong evidence that the long-run proportion of times that New Zealand students associate the name Bob with the face on the left is less than 0.50.

1.CE.9 Not necessarily. A larger sample size yields a smaller p-value if the value of the statistic is the same; there is no guarantee the value of the statistic (proportion of heads) will be the same in Jose and Roberto's separate samples

1.CE.10 A is the only correct answer.

1.CE.11 Roll a die. If it comes up 1 or 2 then call it a success, otherwise failure. Repeat the process 50 times keeping track of the total number of successes out of 50. Then, repeat sets of 50 rolls keeping track of the number of successes within the 50 rolls.

1.CE.12 Flip two coins. If they both come up heads call it a "success," otherwise 'failure.' Repeat the process 25 times keeping track of the total number of successes out of 25. Then, repeat sets of 25 pairs of flips keeping track of the number of successes within the 25 paired flips.

1.CE.13

a. The long-run proportion of times that Rick makes a free throw underhanded

b. Null: The long-run proportion of times that Rick makes a free throw underhanded is 0.90.

Alternative: The long-run proportion of times that Rick makes a free throw underhanded is more than 0.90.

Null: $\pi = 0.90$

Alternative: $\pi > 0.90$

1.CE.14

a. The long-run proportion of times that Lorena makes a 10-foot putt

b. a) Null: The long-run proportion of times that Lorena makes a 10-foot putt is 0.60.

Alternative: The long-run proportion of times that Lorena makes a 10-foot putt is more than 0.60.

Null: $\pi = 0.60$

Alternative: $\pi > 0.60$

1.CE.15

a. The long-run proportion of times someone chooses an odd number

b. Null: The long-run proportion of times that someone chooses an odd number is 0.50.

Alternative: The long-run proportion of times that someone chooses an odd-number is more than 0.50.

c. 1029/1770 = 0.581

d. Yes, the validity conditions are met. There are 1029 successes and 741 failures, both well above the minimum of 10.

e. 6.85

f. p-value = 0

g. We have very strong evidence that the long-run proportion of times someone chooses an odd number is more than 0.50.

1.CE.16

a. The long-run proportion of times someone chooses 7

b. Null: The long-run proportion of times that someone chooses 7 is 0.10.

Alternative: The long-run proportion of times that someone chooses 7 is more than 0.10.

c. 503/1770 = 0.284

d. Yes, the validity conditions are met. There are 503 successes and 1267 failures, both well above the minimum of 10.

e. 25.83

f. p-value ≈ 0

g. We have very strong evidence that the long-run proportion of times someone chooses 7 is more than 0.10.

1.CE.17

a. We don't have strong evidence that the probability a spun tennis racket lands with the label up is different from 0.50.

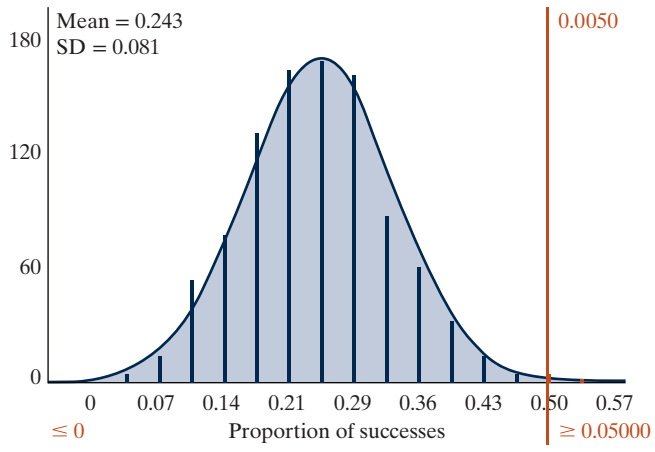
b. If the probability that a spun tennis racket lands with the label is actually 0.50, then it is quite likely to get 46 spins out of 100 with the label up; thus, a reasonable (plausible) explanation for the author's data (46 out 100) is that the tennis racket is fair (0.50 chance of label landing up).

1.CE.18 Null hypothesis probability = 0.50

Statistic = 0.46

p-value = 0.484

1.CE.19 Answers will vary.



1.CE.20 It would be more surprising for the sample proportion to be 0.70 in a sample of 60 compared to a sample of 30 students, if the population proportion is 0.60. This is because it is more unusual to see extreme results in a larger sample (p-value = 0.0719) rather than a small sample (p-value = 0.1763).

Chapter 1 Investigation

1. The observational units are the 28 students.

2. The variable recorded is which tire each student indicates (Right front, right rear, left front, left rear). This is a categorical, non- binary variable. We could also define the variable to be "right front" or "not right front" as that is the primary outcome of interest in our research question.

3. The parameter is the long-term proportion of students who will pick the right front tire.

4. Null: The probability that students choose the right front tire is 0.25 ($\pi = 0.25$). Alternative: The probability that students choose the right front tire is more than 0.25 ($\pi > 0.25$).

5. In this sample,14 out of 28 or 50% of the students selected the right front tire. This is more than we would expect if students choose randomly and equally among the four tires (25% of the time in the long run).

6. Yes, anything is possible, although some outcomes will be less likely or believable when the null hypothesis is true.

7. Our statistic is $\hat{p} = 0.50$.

8. Probability of success (π) = 0.25, sample size (n) = 28, number of samples = 1,000

9. The center is located at 0.25. Yes, it makes sense that this is the center because 0.25 is the specified null hypothesis proportion.

10.

a. The p-value from the simulation should be around 0.004. In other words, in a large number of samples (from a process with = 0.25), roughly .4% of those samples should have a sample proportion of 0.50 or larger.

b. The standard deviation of the simulated sample proportions should be around 0.082, so the standardized statistic is approximately $z = (0.50 - 0.25)/0.082 = 3.05$. (This number may vary a bit because the SD of the simulated null distribution will vary a bit.) Therefore, a sample proportion of 0.50 is more than 3 standard deviations above the hypothesized process probability of 0.25.

c. The One Proportion applet reports a theory-based p-value of 0.0011. The validity conditions are met because there are 14 "successes" (choose right front) and 14 "failures" (choose something else), both of which exceed 10.

11. Yes, all three methods in question 10 give strong evidence against the null hypothesis; the p-values are quite small and the standardized statistic is large (e.g., above 2).

12. We have strong evidence that students pick the right front tire more than 25% of the time because we would rarely have 50% of a sample of 28 students choose the right front tire if the long-run proportion of students choosing the right front tire is 0.25.

13. Hard to say. The question is: "Who are these students?" Will they perform similarly to other people in the same situation? It's hard to say that these students will necessarily act like people in general. We can probably say that we can infer these results to people similar to those that were in the study.

14. Answers will vary. Some things to consider include selecting a broader representation of students to participate in the study, examining exactly how the question is posed to students and whether that impacts their choices, considering whether there might be gender differences or a tendency for different responses among individuals who have recently had a flat tire, and whether this tendency is similar across cultures (including countries where motorists drive on other side of the road).

15. We would expect to find weaker evidence because the sample size is smaller and the statistic (0.50) has stayed the same.

16. The p-value should be around 0.04 which is larger than we got before and hence weaker evidence, as expected.

17. Null: The probability that students choose the right front tire is 0.25 ($\pi = 0.25$). Alternative: The probability that students choose the right-front tire is different than 0.25 ($\pi \neq 0.25$).

18. The p-value is 0.0023 for the two-sided test—about twice as large as before.

19. Yes, we have strong evidence the probability is different than one-fourth because the p-value of 0.0023 is still small enough to be considered strong evidence against the null hypothesis.

It is interesting to note here that even though we pass the technical conditions, the sample size of 28 is still pretty small and the theory-based method is not in close agreement with the simulation approach.

Chapter 1 Research Article

1. The researchers are examining the nature and development of attitudes toward similar and dissimilar others in human infancy

2. (two options, among others) (a) Dissimilar others are perceived as unkind trustworthy and unintelligent (Brewer 1979, etc.) (b) Humans may engage, support or ignore violence directed towards individuals who differ from themselves (Prentice and Miller, 1999)

3. 16

4. To figure out what kind of food the babies preferred (graham crackers or green beans) so that information could be used later in the study

5. Food preference: green beans or graham crackers, categorical with 2 outcomes (green beans or graham crackers)

6. To have babies establish which rabbit is similar to them, and which is dissimilar

7. No

8. The researchers needed to make sure that the babies understood what they were seeing (one puppy be nice to the rabbit, and one puppy be mean to the rabbit).

9. Puppy preference: Harmful or Helpful

10. 12/16 chose helper when viewing activities involving the rabbit similar to them, compared to 4/16 who chose the harmful puppy when viewing activities involving the rabbit similar to them.

11. 100% similar chose helper; 0% dissimilar chose helper

12. Fifty-three percent is fairly close to 50% (the null hypothesis) and the sample size (36) is not large.

13. To modify the experiment so that babies have a "neutral" option to provide strong comparisons between groups

14. Null hypothesis: The long-run proportion of times an infant chooses the harmer dog is 0.50 when the dog interacts with the dissimilar rabbit; Alternative hypothesis: The long-run proportion of times an infant chooses the harmer dog is not 0.50 when the dog interacts with the dissimilar puppet..

15. Answers will vary. Sixty-three percent of 14-month-olds in the study chose graham crackers over green beans when given a choice between the two.

16. Answers will vary. Null hypothesis: The long-run proportion of times that a 14-month-old chooses helper character instead of the neutral character is 0.50 when the dog interacts with the similar rabbit; Alternative hypothesis: The long-run proportion of times that a 14-month-old chooses the helper character instead of the neutral character is not 0.50 when the dog interacts with the similar rabbit. The p-value is 0.08, meaning that there is moderate evidence that the long run proportion of times that a 14-month-old chooses the helper characters instead of the neutral character is not 0.50.

17. If the infants in the study are special in some way (e.g., particularly developmentally advanced or not; different ethnicity, socio-economic status, etc.) then the results from this study may not generalize to all infants.

18. Answers will vary. (a) Select babies to represent different ethnicities/SES in order to improve the ability to generalize the results. (b) Give babies different foods to choose from initially to ensure that there is no impact of green beans/graham crackers in particular.

19. If babies generally liked the graham cracker rabbit/dog and didn't like the green bean rabbit/dog, then the researchers' conclusions about similarity/dissimilarity would be invalidated, because the differences between the similar and dissimilar conditions would be better explained by a different variable (green beans/graham crackers).

20. The researchers refer to prior research that links adult and child similarity preferences and group psychology, which, they argue, suggests that their results are more inborn (nature) rather than the result of accumulated experiences (nurture).