ISCAM 3: CHAPTER 1 EXERCISES

1. Go Take a Hike!
The book *Day Hikes in San Luis Obispo County* by Robert Stone gives information on 72 different hikes that one can take in the county. For each of the 72 hikes, Stone reports the distance of the hike (in miles), the anticipated hiking time (in minutes), the elevation gain (in feet), and the region of the county in which the hike can be found (North County, South County, Morro Bay, and so on, for a total of eight regions).
(a) What are the observational units here?
(b) How many variables are there?
(c) Classify each variable as quantitative or categorical.
(d) If you create a new variable that is the ratio of hiking time to hiking distance, would that be a quantitative or categorical variable?
(e) If you create a new variable called *length of hike* that is coded as "short" for a hike whose distance is less than 2 miles, "medium" for a hike whose distance is at least 2 but not more than 4 miles, and "long" for a hike whose distance is at least 4 miles, would this be a quantitative or categorical variable?
(f) For each of the following, specify whether or not it is a legitimate variable for the observational units you specified in (a):
  • Longest hike in the book
  • Average elevation gain of the hikes in the book
  • Whether or not the hike is in the Morro Bay region
  • Proportion of hikes with an elevation gain of more than 500 feet
  • Anticipated hiking time, reported in hours

2. Identifying Variables
(a) Consider the 50 states as observation units for a study. Suggest one quantitative and one categorical variable that you could measure about the states. (Be careful to express you answers clearly as variables.)
(b) Consider the nine members of the Supreme Court as the observational units for a study. Suggest one quantitative and one categorical variable that you could measure about the justices. (Be careful to express your answers clearly as variables.)

3. Bookstores
Consider all of the books in a bookstore as the observational units in a study.
(a) Identify three quantitative variables that could be recorded on these books.
(b) Identify three categorical variables that could be recorded on these books.

4. Rock n' Roll
We saw a claim on "trivia coaster" from the Rock n' Roll Hall of Fame that 30% of all songs with a color in the title are about the color blue.
(a) Clearly state in words the observational units and the variable that this claim refers to.
(b) Is the variable quantitative or categorical?
(c) Clearly state in words the population and parameter that this claim refers to.
(d) Clearly state, both in words and in symbols, the null and alternative hypotheses for testing the claim.

5. Predicting Elections from Faces?

Do voters make judgments about political candidates based on his/her facial appearance? Can you correctly predict the outcome of an election, more often than not, simply by choosing the candidate whose face is judged to be more competent-looking? Researchers investigated this question in a study published in *Science* (Todorov, Mandisodka, Goren, & Hall, 2005). Participants were shown pictures of two candidates and asked who has the more competent looking face. Researchers then predicted the winner to be the candidate whose face was judged to look more competent by most of the participants. For the 32 U.S. Senate races in 2004, this method predicted the winner correctly in 23 of them.

(a) In what proportion of the races did the "competent face" method predict the winner correctly?
(b) Describe (in words) the null model to be investigated with this study.
(c) Describe how you could (in principle) use a coin to produce a simulation analysis of whether these data provide strong evidence that the "competent face" method would correctly predict the election winner more than half the time. Include enough detail that someone else could implement the full analysis and draw a reasonable conclusion.
(d) Use the One Proportion Inference applet to conduct a simulation (using 1000 repetitions), addressing the question of whether the researchers' results provide strong evidence in support of the researchers' conjecture that the "competent face" method would correctly predict the election winner more than half the time. Submit a copy of the applet output (e.g., "print screen" key or a screen capture), and indicate where the observed research result falls in that distribution. Also report the approximate p-value from this simulation analysis.
(e) Write a paragraph, as if to the researchers, describing what your simulation analysis reveals about whether the data provide strong evidence in support of their conjecture.

These researchers also predicted the outcomes of 279 races for the U.S. House of Representatives in 2004. The "competent face" method correctly predicted the winner in 189 of those races.

(f) Use the applet to conduct a simulation analysis of these data. Again submit a copy of the "what if?" distribution, and indicate where the observed research result falls in that distribution. Also report the approximate p-value, and summarize your conclusion, again as if to the researchers.

6. Can Dolphins Communicate?

A famous study from the 1960s explored whether two dolphins (Doris and Buzz) could communicate abstract ideas. The dolphins were trained to push the right button if a headlight was shone steadily, but the left button if the headlight blinked on and off. Then the researcher placed a large wooden wall in the middle of the pool. Doris was on one side of the wall and could see the headlight, whereas Buzz was on the other side of the wall where he could not see the headlight. When the light was shone, Doris would swim near the wall and whistle. Buzz would then whistle back and press a button. If he pushed the correct button (corresponding to the light Doris was shown), they both got a fish. Dr. Bastian repeated this procedure again and again, and Buzz pushed the correct button 22 out of 25 times. Is this convincing evidence that Buzz and Doris could communicate?

(a) Calculate the statistic for this study.
(b) State the null and alternative hypotheses (in symbols and in words) to investigate this research question.
(c) Use the One Proportion Inference applet to estimate a p-value for this significance test. Include a copy of your computer results (e.g., screen capture).
(d) Write a one-sentence interpretation of this p-value.
(e) Summarize the conclusions you would draw about the research question based on this p-value.

7. Rock-Paper-Scissors

For informal sports events, players often play "rock-paper-scissors" to decide who serves first or who is the home team etc. Two players simultaneously show one of the three objects. The player showing the dominant object (e.g., rock beats scissors) wins. The optimal strategy is to alternate among the three objects. Play the game rock/paper/scissors against the computer using the website:
http://www.nytimes.com/interactive/science/rock-paper-scissors.html. Select the novice version of the computer to play against. Play for at least 30 rounds, but keep going for as long as you'd like. Keep track of which option you choose (rock or paper or scissors) for every round that you play (the computer will record this for you but that information will soon scroll off the screen, so make your own notes). *Try to recreate how you would play against a person and don't view your prior results when making your next selection.*
(a) Identify the observational units in this study.
(b) Identify the variable of interest.
An article published in *College Mathematics Journal* (Eyler, Shalla, Doumaux, & McDevitt, 2009) found that players tend to not prefer scissors, choosing it less than 1/3 of the time. We will investigate whether your data suggest that *you* tend to choose scissors less than one-third of the time.
(c) Calculate the statistic in this study and create a bar graph of your results. (In R and Minitab, you can do this with just the "summarized" data, you don't have to enter the individual outcomes.) Are your results in the direction conjectured by these researchers (choosing scissors less than 1/3 of the time)?
(d) Define the parameter of interest in this study.
(e) State appropriate null and alternative hypotheses about this parameter according to the theory suggested in the *CMJ* article.
(f) Explain how you could use an ordinary six-sided die to simulate a what-if distribution under this null hypothesis. Be sure to indicate what each possible outcome of the die (1, 2, 3, 4, 5, 6) would represent.
(g) Based on your sample results, are you convinced that you choose scissors less than one-third of the time in the long run? Clearly explain your reasoning.

8. Which Tire?

A statistics class at Cal Poly collected data on a well-known campus legend. Each student was asked to specify one of the four tires to answer in a situation where you have to make up which tire had recently been flat on your car. The prior conjecture is that a higher number than would be expected due to chance alone would pick the *right front* tire. In this class, 24 of 54 students in class chose the right front tire (a tire identified in advance as being one that people tend to pick out of the four). You will conduct a test of whether these data provide evidence that Cal Poly students tend to choose the right front tire more often than would be expected if the four tire choices were equally likely.
(a) Identify the observational units and variable in this study. Also classify the variable as categorical or quantitative. If the variable is categorical, also indicate whether it is binary.
(b) State the appropriate null and alternative hypothesis, in symbols and in words.
(c) Use technology to produce a bar graph of the student responses. Submit this graph, and comment on what it reveals.
(d) Use technology to determine the (exact binomial) p-value for the test of your hypotheses in (b).
(e) Write a sentence describing what this p-value is the probability of.
(f) Write a couple of sentences summarizing the conclusion that you would draw from this analysis and also explaining the reasoning process that underlies your conclusion.
(g) Suppose that another statistics class conducts this same study in their own class, which has exactly half as many students. Suppose further that this class obtains the same proportion of students choosing the right front tire. Determine the exact p-value in this case. Describe how the p-value and your conclusion would differ for this class of 27 students compared to the first class of 54 students, and comment on why this makes intuitive sense.

9. Which Tire? (cont.)
Reconsider the previous exercise, where you tested whether sample data suggest that more than 25% of a population would answer "right front" when asked to name a tire that had gone flat. Suppose that you read of a study in which 30% of a random sample answered "right front."
(a) What further information would you require to assess whether this sample result constitutes strong evidence that more than 25% of the population would answer "right front"? Also explain why this information is needed.
(b) Determine the p-value of the test for the following sample sizes, in each case supposing that the sample proportion answering "right front" is 0.3: $n = 10, 50, 250, 500$. (Feel free to use technology, but explain what you ask the technology to do in each case.)
(c) Determine the smallest sample size for which this sample result would be statistically significant at the 0.05 level.
(d) Repeat (c) for the 0.01 level.

10. Competitive Advantage of Uniform Color?
Does uniform color give athletes an advantage over their competitors? To investigate this question, Hill and Barton (*Nature*, 2005) examined the records in the 2004 Olympic Games for four combat sports: boxing, tae kwon do, Greco-Roman wrestling, and freestyle wrestling. Competitors in these sports were randomly assigned to wear either a red or a blue uniform. The researchers investigated whether competitors wearing one color won significantly more often than those wearing the other color. They analyzed results for a total of 457 matches. Of these, red won the match 248 times, while blue won 209 times.
(a) Identify the observational units and binary, categorical variable of interest. Indicate which outcome you will consider "success."
(b) Explain how and why randomness was used in this study.
(c) State the null and alternative hypotheses based on this research question. [*Hint*: Before you look at the data.]
(d) Calculate a binomial p-value for investigating this research conjecture.
(e) Compute a 95% binomial confidence interval for the parameter. Write a sentence interpreting what this interval says, including how you are defining the parameter.
(f) Is the confidence interval result consistent with your p-value result? Explain.
(g) Now determine a 90% confidence interval for the parameter. Comment on how it differs from the 95% interval. [*Hint*: Refer to both the midpoints of the intervals and their widths.]
(h) Summarize your results as if to an athletic director at a university. Include discussion about how you are willing to "generalize" these results beyond these 457 matches.

11. Competitive Advantage of Uniform Color (cont.)
For the Competitive Advantage of Uniform Color data (Exercise 10), assuming $\pi = 0.5$:
(a) State the appropriate null and alternative hypotheses, both in symbols and in words.
(b) Use the One Proportion Inference applet to simulate 1000 repetitions of these matches, under the null hypothesis. Submit a screen capture of the applet results.
The researchers found that the competitor wearing red defeated the competitor wearing blue in 248 matches, and the competitor wearing blue emerged as the winner in 209 matches.
(c) Use the applet simulation results to approximate the two-sided p-value from these data. Also report which values are being counted to determine this approximate p-value.
(d) Use technology to determine the exact binomial (two-sided) p-value. Submit the output with your answer.
(e) Summarize what your analysis reveals about how much evidence the data provide for concluding that uniform color does give one athlete an advantage over the other.

(f) Use technology to determine a 95% confidence interval for the parameter. Also write a sentence interpreting what this interval says.

(g) Now determine a 99% confidence interval for the parameter. Comment on how it differs from the 95% interval. [*Hint*: Refer to both the midpoints of the intervals and their widths.]

(h) Are these confidence intervals consistent with your earlier test (parts a-e)? Explain briefly.

12. Plausible Values?

In Investigation 1.3, you considered the sample of 361 patients with 71 deaths and whether 0.15 was a plausible value for the probability of a death during a heart transplantation at St. George's hospital at this time. Use the One Proportion Inference applet to test other values for π (e.g., 0.20, 0.21, 0.22, …). Create a list of which values (multiples of .01 are fine) that have a *two-sided* p-value of at least 0.05. That is, what are the *plausible* values of π based on the observed sample result? Include a screen capture of your applet results for the smallest value of π you consider plausible and the largest value of π you consider plausible.

13. A Cup of Tea

One of the first descriptions of a true randomized experiment was given by Ronald Fisher in 1935 (*The Design of Experiments*). He described a tea party in Cambridge where a woman claimed that she could tell whether a cup of tea with milk had the milk added to the tea or the milk poured into the cup first and then the tea added. Fisher proposed an experiment to determine whether she truly could tell the difference.

(a) Explain how randomization could be used in such an experiment.

(b) If she is given one cup of tea, what is the probability that she could give the right answer even if she is only guessing?

(c) Suppose she will be given 10 cups. Let X represent the number of cups for which she makes the correct identification of which ingredient was poured first. Can X be considered a binomial random variable? Explain.

(d) What are the parameters of the binomial distribution in (c)? Explain why the exact value for one of the parameters is unknown.

(e) Suppose that she is just guessing on each cup and also that she doesn't know how many cups there are of each type. Use the One Proportion Inference applet to determine how many identifications she needs to get correct out of ten so that the probability of doing so by guessing is less than 0.05.

(By the way – Fisher never said how many she did get correct!)

14. Hound of the Baskervilles

Phillips et al. (2001) report on a "Hound of the Baskervilles" effect, in which they conjecture that Chinese- and Japanese-Americans, perhaps frightened by the negative connotations of the number 3 in their culture, die more often on the 3rd day of a month than on the 2nd or 4th day of the month. They analyzed death records of Chinese- and Japanese-Americans from California for the years 1989–1998, focusing on what they termed "chronic," inpatient, coronary deaths. They found 404 deaths occurring on days 3, 4, or 5 of a month, with 157 of those deaths occurring on day 3.

(a) What proportion of these 404 deaths occurred on day 3? Is this higher than 1/3? Produce a bar graph to display these data.

(b) If in fact 1/3 of all deaths on day 3, 4, or 5 occur on day 3, what is the probability that a sample of 404 such deaths would reveal 157 or more on day 3?

(c) Is this probability small enough to suggest that Chinese- and Japanese-Americans do die on the 3rd day of the month more often than would be expected by chance? Explain.

Smith (2002) conducted a follow-up study in which he broadened the types of deaths to include all inpatient coronary deaths and all coronary deaths for the same years. He also studied deaths in the subsequent three years 1999-2001. Some of his findings are in the table:

Types of death	Years	# Deaths on day 3, 4, 5	Proportion of deaths on day 3
All coronary	1989–1998	1391	33.9%
Inpatient coronary	1989–1998	540	36.5%
Inpatient, chronic coronary	1999–2001	133	34.6%

(d) For each of these three categories of deaths, determine the p-value for a binomial test of whether the underlying death rate exceeds 1/3.

(e) Based on these p-values, do Smith's data seem to support the contention that Chinese- and Japanese-Americans die more often on the 3rd day of the month than on the 2nd or 4th day? Explain.

15. Pilot Study

On a recent trip one of the authors took four flights and encountered eight pilots (two per flight). Seven of these eight pilots were men, and one was a woman. Let the random variable X represent the number of women in a random sample of 8 commercial pilots.

(a) Would it be reasonable to model the distribution of X with a binomial distribution? Explain and clearly define the parameters of the distribution in words.

(b) Suppose that half of all commercial pilots are women. Determine the probability of encountering one or fewer women in a random sample of eight commercial pilots. (Show the details of your calculations (technology output) to support your answer.)

(c) Is the probability in (b) small enough to convince you that, based on these sample data, fewer than half of all commercial pilots are women, using the 0.05 level of significance?

(d) According to Women in Aviation International (WIA), women comprised 5% of all commercial pilots in 2001. Use this parameter value to calculate and graph the probability distribution of X. [*Hint*: List all possible values and their probabilities. You may use technology.]

(e) Assuming that the WIA parameter value is correct and using the results from (d), is *one* the most common value for the number of women in a random sample of eight commercial pilots? Would you say that it is a surprising value? Explain.

16. Pilot Study (cont.)

Reconsider the previous question, about encountering one woman in a sample of eight commercial pilots. Let X be a binomial random variable with parameters $n = 8$ and π = proportion of all commercial pilots that are female.

(a) Express $P(X = 1)$ as a function of π.

(b) Graph this function, for all possible values of π from 0 through 1.

(c) Use calculus to determine the value of π that maximizes this $P(X = 1)$ function. Verify from the graph that this value of π really does maximize the function.

(d) Explain why this value of π that maximizes this $P(X = 1)$ function makes intuitive sense.

(e) More generally, use calculus to determine the value of π that maximizes the function $P(X = k)$, where k can be integer from 0 through 8, inclusive. Does your answer make intuitive sense? Explain. (This question introduces you to a widely used statistical technique for estimating unknown parameters, known as maximum likelihood estimation.)

17. Rolling Die
Suppose that your instructor rolls a pair of dice five times.
(a) What is the probability that the sum of the dice would be either 7 or 11 all five times, assuming that the dice were fair? [*Hint*: Let the random variable X represent the number of rolls in which the sum is 7 or 11, and use the binomial distribution to calculate this probability. You will first need to determine π, the probability that the sum would be 7 or 11 on any one roll by listing the sample space of 36 equally likely outcomes. Then use that probability for the binomial parameter.]
Show the details of your calculations, or technology output, to support your answer.
(b) Would this result (five consecutive rolls of 7 or 11) constitute reasonably strong evidence that the dice were not fair? Explain your reasoning, based on your answer to (a).
(c) Determine the probability of obtaining a sum of 7 or 11 in *n* consecutive rolls of a pair of fair dice as a function of *n*, where *n* is a positive integer.
(d) Graph this probability as a function of *n*, and comment on the behavior of the function.
(e) Determine the smallest value of *n* for which this probability dips below 0.05. Then find the smallest such values of *n* for which the probability dips below 0.01, 0.001, and 0.00001.

18. Winning Survivor
Contestants compete for a million dollar prize on the hit reality game show *Survivor*. For the first few years, when the competition had been weaned down to just three remaining contestants, one of them had won "immunity" and got to decide which of the other two will advance to the final round. During that final round, each of the two finalists presents his/her case to the "jury" on why he/she deserves to win the million dollars. In the first eight series of *Survivor*, the person of the final two who had immunity won the overall game only twice. For this question you will consider whether these results provide convincing evidence that the person with immunity is the ultimate winner (of the final two) less often than we would expect by chance.
(a) Define the parameter of interest in words.
(b) State the appropriate null and alternative hypotheses (in symbols and in words).
(c) Let X represent the number of contestants with immunity who go on to become the ultimate winner (of the final two). Can X be considered a binomial random variable? Explain.
(d) Calculate the p-value for the hypotheses stated in (b). Show the details of your calculations (or technology output) to support your answer.
(e) Is this p-value statistically significant at the 5% level?
(f) Do the data provide convincing evidence that the contestant with immunity becomes the ultimate winner less often than we would expect by chance? Explain, including any cautions that you have about generalizing these results to the general process.
(g) Suppose that none of those with immunity had gone on to win the overall game. How surprising would this result be if those with immunity did not have a disadvantage in the final two? Support your answer with an appropriate probability calculation.

19. Multiple-Choice
One of the authors moved to California in 2001 and faced the traumatic situation of taking a multiple choice exam to qualify for a California driver's license. This exam consisted of 36 multiple-choice questions, with three options provided for each question. Candidates needed to answer at least 31 questions correctly in order to pass. Let the random variable X be the number of questions answered correctly on the exam.
(a) Suppose that the candidate guesses randomly among the three options on each question. What probability distribution does X have? (Specify the parameters of the distribution as well as its name.)

(b) If the candidate guesses randomly on each question, what would be the expected number of questions that he would answer correctly? What would the standard deviation of the number correct be?

(c) If the candidate guesses randomly on each question, what is the probability of passing the exam? (Show the details of your calculations, or technology output, to support your answer, here and throughout.)
[*Minitab note*: You may need to double click on the X-axis to change the max value in order to see the results.]

(d) If the candidate can eliminate one wrong option on each question and guess randomly between the other two, what is the probability of passing the exam? (Identify the probability distribution of X in this case also.)

(e) If the candidate studies to the point of having a 0.9 probability of answering each question, independently from question to question, what is the probability of passing the exam? (Identify the probability distribution of X in this case also.)

(f) Let π represent the probability of answering each question correctly, independently from question to question. If the candidate wants to have at least a 99% chance of passing the exam, what is the smallest value of π that will achieve this? [*Hint*: Use trial-and-error, and report π to two decimal places.]

The rules actually allow a candidate to take the exam three times, and passing (with at least 31 of 36 correct) at least once is sufficient to qualify for a license.

(g) Suppose now that the candidate studies to the point of having a 0.8 probability of answering each question, independently from question to question. What is the probability of passing the exam at least once in the three allotted attempts?

20. Online Solitaire

Enamored of the solitaire game on his new computer, author A sets out to estimate his probability of winning the game. Suppose he plans to play the game 100 times and record how often he wins the game.

(a) Let *W* represent the number of times he wins. Would *W* follow a binomial distribution? Discuss each of the binomial conditions.

(b) Define the parameter of interest in words.

(c) The author ended up winning 25 times and losing 192 times (217 trials). Use technology to determine a 95% confidence interval for the parameter in (b). Justify your choice of interval procedure.

(d) Author B played 444 games and won 74. Conjecture how a 95% confidence interval for Author B will compare to that of Author A (comment on width and midpoint).

(e) Calculate a 95% confidence interval for Author B and compare to that of Author A. Were your conjectures in (d) correct? If not, explain why they compare the way they do.

(f) Is 1/9 a plausible value for the probability of a win at this game for either author? Justify your choice.

21. Wearing Helmets

A study conducted by the non-profit Safe Kids campaign in 2004 found that 33% of kids riding a bicycle on a residential street wore a safety helmet. Suppose that you watch a sample of 30 kids riding a bike on a residential street in your community. Let the random variable *H* represent the number of those 30 kids who are wearing a helmet.

(a) Would *H* follow a binomial distribution? Explain.

(b) If the rate of helmet wearing is the same in your community as in the Safe Kids study, how many kids would you expect to be wearing a helmet?

(c) Explain how you might use a six-sided die to simulate the distribution of *H*.

(d) Suppose that 15 of the 30 kids in your sample are wearing a helmet. Is that enough to convince you, at the 10% significance level, that the helmet wearing rate in your community is higher than 33%? Address this question with all of the steps of a test of significance.

22. Rock-Paper-Scissors (cont.)
Recall the "Rock-Paper-Scissors" scenario from Exercise 7. A friend conjectures that men are more likely to show rock than paper or scissors. To test this, you ask 50 male friends to play the game with you and you note the first choice of each male.
(a) State the appropriate hypotheses to be tested, in symbols and in words. (Pay particular attention to whether this is a one-sided or a two-sided test.)
(b) Let the random variable X represent the number of males that show "rock" as their first choice. What distribution does X follow if the null hypothesis is true? Explain.
(c) Suppose 23 men choose rock first. Determine the p-value from these data, and indicate whether the null hypothesis would be rejected at any commonly used significance level.
(d) Use technology to determine a 95% confidence interval for the parameter. Interpret this interval, including a clear statement of what the parameter is.
(e) Is the confidence interval consistent with your test result? Explain.

23. Detecting Cancer
In a study whose results were published in the *British Medical Journal* (Willis et al., 2004), researchers tested whether dogs' sense of smell can be helpful in detecting cancer in humans. Each dog used in the study was presented with seven urine samples to smell, with only one of the urine samples coming from a patient suffering from bladder cancer. Taken as a group, the dogs were tested on 54 trials, and they correctly identified the urine from the cancer patient in 22 trials. Conduct the appropriate test to assess whether these data are statistically significant. Be sure to state the appropriate hypotheses, check whether the binomial model applies, and calculate the p-value. Also summarize your conclusions and explain how they follow from your analysis.

24. Sarah the Chimpanzee
Early research has found chimpanzees able to solve complex problems, like fitting sticks together to make a rake to gather food. In a 1978 study published in *Science*, Premack and Woodruff asked "To what extent does the chimpanzee comprehend the elements of a problem situation and potential solutions?" An adult chimpanzee (Sarah) was shown 30-second videotapes of a human actor struggling with one of several problems (for example, not able to reach bananas hanging from the ceiling, a record player not playing). Then Sarah was shown two photographs, one that depicted a solution to the problem (like stepping onto a box, plugging in the record player) and one that did not match that scenario. On seven of eight problems, the animal consistently chose the correct photograph.
(a) Identify the observational units and variable in this study.
(b) State the null and alternative hypotheses for testing whether Sarah is more likely to pick the correct photograph than if she was just randomly guessing between the two solutions each time.
(c) Is it reasonable to model this study as a binomial process? Explain.
(d) Determine the p-value for Sarah and provide a detailed interpretation of the p-value you find.
(e) Summarize the conclusions you would draw from this study. Do you think Sarah got lucky or do you think something other than random chance was at play? How strong is the evidence?

25. Armrest Battles
In a study reported in 1982 by Kai, Khairullah, and Coulmas, researchers observed 426 pairs of passengers in "mixed-sex" seating arrangements on airplanes to see whether either person used the joint armrest. Observations were made after a beverage or meal was served. Passengers who were asleep or obviously couples were not counted. Data were collected by noting whether the man, the woman, both, or neither was using the armrest. Of the 426 pairs, the man used the armrest 284 times, the woman 57 times,

both 37 times, and neither 48 times. We will consider only the 341 pairs for which one person or the other used the armrest.
(a) State the null and alternative hypotheses for testing whether men are more likely than women to use the shared armrest. (Be sure to define the parameter π in this context.)
(b) Use the binomial distribution to calculate the exact p-value of this test.
(c) Identify a potential alternative explanation (rather than just being male) to explain the higher armrest use by males in this study.
(d) Calculate a 95% confidence interval for the probability that the shared armrest is used by the male. Include a one-sentence summary of this interval.
(e) Is 0.5 among the values in your confidence interval? Explain why this is consistent with your analysis in part (b).

26. Halloween Treat Choices (cont.)
Reconsider Investigation 1.9, in which you analyzed data from a study of 284 children's Halloween treat preferences. Let the parameter π represent the underlying probability that a child would choose the candy over the toy.
(a) Determine and interpret a 90% binomial confidence interval for π.
(b) Determine a 95% binomial confidence interval for π.
(c) Determine a 99% binomial confidence interval for π.
(d) How do these intervals compare to each other? Explain why this makes sense.
(e) Do some or all of these intervals include the value 0.5? Explain the relevance of this to the original research question in this study.

27. Halloween Treat Choices (cont.)
Reconsider the previous exercise.
(a) How would you expect those confidence intervals to change if you defined the parameter π to be the underlying probability that a child would choose the *toy* over the candy? Explain.
(b) Recalculate one of the previous intervals (90%, 95%, or 99%) with this new definition of the parameter π. Does the resulting interval validate what you expected in (a)? If not, explain how this interval compares to the previous one (with the other definition of π).

28. Relieving Back Pain
A study published in the journal *Neurology* (May 22, 2001) examined whether the drug botulinum toxin A is helpful for recurring pain among patients who suffer from chronic low back pain. The 31 subjects who participated in the study were randomly assigned to one of two treatment groups: 16 received a placebo of normal saline and the other 15 received the drug itself. The subjects' pain levels were evaluated at the beginning of the study and again after eight weeks. The researchers found that 2 of the 16 subjects who received the saline experienced a substantial reduction in pain, compared to 9 of the 15 subjects who received the actual drug.
(a) Use these data to conduct a two-sided test of whether one-third of all patients would experience substantial pain reduction from a placebo. Report the hypotheses and p-value, along with a summary of your conclusion. Also provide specific details about how the p-value is calculated.
(b) Repeat (a) for testing whether one-tenth of all patients would experience substantial pain reduction from a placebo.
(c) Determine a 95% confidence interval for estimating the proportion of all patients with this condition who would experience substantial pain reduction from a placebo.

(d) Does this interval include the value 0.333? Does it include the value 0.1? Explain why your answers make sense in light of the p-values in (a) and (b).

(e) Without doing any calculations, describe how you would expect your answers to (a), (b), and (c) to change if the study had involved 160 patients receiving the placebo, 20 of whom experienced pain reduction. [*Hint*: Notice that the sample size is larger but the proportion of successes is unchanged.]

29. Hospital Mortality Rates (cont.)

Reconsider Investigation 1.3, in which you analyzed data from a sample of ten transplant operations performed by St. George's Hospital in London. Continue to consider testing whether the underlying mortality rate at the hospital exceeds the benchmark rate of 15%.

(a) Determine the smallest number of patient deaths (out of these ten operations) that would have been needed to produce a statistically significant result at the 5% level. Show the details of your calculations, or technology output, to support your answer.

(b) Repeat (a) for the 1% significance level.

(c) Repeat (a) and (b) for the larger sample of 361 patients whose records were analyzed.

30. Improved Batting Averages

Reconsider Investigation 1.6, in which you considered a baseball player who has historically had a 0.250 probability of success but now has a 0.333 probability. Continue to suppose that the player has a sample of 20 at-bats (trials) in which to convince his manager to reject the null hypothesis that $\pi = 0.250$ in favor of the alternative that $\pi > 0.250$. Now suppose that you want to choose the rejection region that minimizes the sum of the probabilities of Type I and Type II errors. Consider rejection regions of the form $\{X \geq k\}$, where X represents the number of hits (successes) in the sample of 20 at-bats (trials).

(a) For values of k from 0 to 20, calculate the probability of Type I error.

[*Hints*: It may be useful to keep in mind that $P(X \geq k) = 1 - P(X \leq k-1)$.]

(b) Produce a graph of these Type I error probabilities as a function of k.

(c) For values of k from 0 to 20, determine the probability of Type II error.

(d) Produce a graph of these Type II error probabilities as a function of k.

(e) Take the sum of the Type I and Type II error probabilities, and graph this sum as a function of k.

(f) What rejection region (value of k) minimizes the sum of these error probabilities?

31. Improved Batting Averages (cont.)

Reconsider the previous exercise. Now suppose that you consider a Type I error to be twice as serious as a Type II error.

(a) Calculate and graph $2 \times P(\text{Type I error}) + P(\text{Type II error})$, as a function of k.

(b) What rejection region (value of k) minimizes the sum of these error probabilities?

(c) Repeat, now supposing that you consider a Type I error to be four times as serious as a Type II error.

(d) Do you notice a pattern in how these rejection regions change, as you consider Type I error to be increasingly more important than a Type II error? Explain why this makes sense.

32. Improved Batting Averages (cont.)

Reconsider the previous exercises. Suppose that the player has a sample of n at-bats (trials) in which to convince his manager to reject the null hypothesis that $\pi = 0.250$ in favor of the alternative that $\pi > 0.250$, using the $\alpha = 0.05$ significance level.

(a) Intuitively, without doing any calculations but perhaps sketching some graphs, do you expect the power of this test to be an increasing function of n, a decreasing function of n, or neither? Explain.

(b) Determine how large the sample size n needs to be, in order for the power to be at least 0.75. Explain the details of your calculation, but feel free to use technology to check your answer.
(c) Now suppose that the player's success probability is denoted by π_a, not necessarily equal to 0.333. For a fixed sample size n and significance level α, do you expect the power of this test to be an increasing function of π_a, a decreasing function of π_a, or neither? Explain.
(d) With a sample size of $n = 50$ and the $\alpha = 0.05$ significance level, determine how large π_a has to be in order for the power to be at least 0.75. Explain the details of your calculation, but feel free to use technology to check your work.

33. Halloween Treat Choices (cont.)

Reconsider Investigation 1.9, in which you analyzed data from a study of 284 children's Halloween treat preferences. Let the parameter π represent the underlying probability that a child would choose the candy over the toy.
(a) State an appropriate null and alternative hypothesis involving this parameter (in symbols and in words), for testing whether there is strong evidence of a preference for either the toys or the candy.
(b) Explain what a Type I error would represent in this context.
(c) Explain what a Type II error would represent in this context.
(d) Suppose 60% of trick-or-treaters prefer the toy. Explain what "power" would represent in this context (loosely outline how you would determine the power of the test, you do not need to perform the calculations).
(e) Explain the effect of increasing the sample size on the power of this test.
(f) Now suppose that 55% of all children prefer candy over a toy as a Halloween treat. Determine the power of the test that you performed, using the 0.05 level of significance.
(g) How does the power in (f) compare to that in (d)? Explain why this makes sense.

34. Cola Discrimination?

A teacher doubted whether his students could distinguish between two different brands of cola soft drink (say, Coke and Pepsi). He presented each of his 48 students with three cups of cola. Two contained the same brand, and the third contained the other brand. Each student was asked to identify the cup containing cola that differed from the other two cups. Let π represent the probability that a student can correctly identify the "odd" brand. The hypotheses to be tested are $H_0: \pi = 1/3$ vs. $H_a: \pi > 1/3$.
(a) Describe (in words) what Type I error means in this situation.
(b) Describe (in words) what Type II error means in this situation.
(c) Describe (in words) what power means in this situation.
For the remaining questions, you may use either the Power Simulation applet for an approximate answer or R/Minitab for an exact answer. (Include screen captures of applet results or R/Minitab output with your answers.)
(d) Determine the rejection region for this test, using the $\alpha = 0.05$ significance level.
(e) Calculate the power of this test, using the $\alpha = 0.05$ significance level, when the success probability is actually $\pi = 0.5$. Also be sure to write this probability as P(X ___ k), where you indicate the appropriate probability distribution of X, fill in the blank with the appropriate inequality, and indicate the appropriate value of k.)
(f) Conjecture how the power would change if the success probability were larger. Explain why this makes sense intuitively. Then calculate the power when $\pi = 2/3$, and comment on whether this supports your answer.
(g) Conjecture how the power would change if the significance level were smaller. Explain why this makes sense intuitively. Then calculate the power using $\alpha = 0.01$ (for an alternative value of $\pi = 0.5$), and comment on whether this supports your answer.

(h) Conjecture how the power would change if the sample size were larger. Explain why this makes sense intuitively. Then calculate the power using $n = 96$ (with $\alpha = 0.05$ for an alternative value of $\pi = 0.5$), and comment on whether this supports your answer.

35. The Monty Hall Problem

In the famous "Monty Hall problem," a game show contestant is asked to choose one of three doors, behind only one of which is a prize. Then the host opens one of the other doors, one which he knows does not have the prize, and asks the contestant if she would then prefer to switch to the remaining door. Suppose that you want to play a series of games with the switching strategy to investigate whether the probability of winning with that strategy is higher than 1/3.

(a) State the null and alternative hypotheses, in symbols and in words.

(b) Suppose that you play this game for a sample of 25 trials. Determine the rejection region corresponding to the $\alpha = 0.10$ significance level.

It can be shown that the probability of winning the prize doubles from 1/3 to 2/3 by switching doors.

(c) Determine the probability of committing a Type II error in this study.

(d) Repeat (b) and (c) for the $\alpha = 0.05$ significance level.

(e) Repeat (b) and (c) for the $\alpha = 0.01$ significance level.

(f) Describe how the probability of Type II error changes as the significance level α decreases. Explain why this makes intuitive sense.

36. Heart Transplant Mortality (cont.)

Reconsider the situation of Investigation 1.11, in which you tested whether the true mortality rate of heart transplant procedures at the hospital exceeded 0.15. Recall that data on the previous 361 patients revealed 71 deaths.

(a) Do these data provide very strong evidence that the underlying mortality rate exceeded 0.15? Would you reject the null hypothesis at the $\alpha = 0.01$ significance level? Explain, based on the p-value of the test.

(b) Repeat (a), if there had been 70 deaths, rather than 71, in the sample of 361 patients.

(c) Repeat (a), if there had been 60 deaths, rather than 71, in the sample of 361 patients.

(d) In which two of these three cases do you make the same decision at the $\alpha = 0.01$ significance level?

(e) In which two of these three cases are the sample results most similar?

(f) Explain why your answers to (d) and (e) have convinced many statisticians to find the p-value more informative than only a statement of whether a null hypothesis is rejected at one particular significance level.

37. Heart Transplant Mortality (cont.)

Reconsider the previous exercise. Continue to consider testing whether the data suggest that the underlying mortality rate of heart transplant procedures at this hospital exceeded 0.15, based on a sample of 361 patients.

(a) Determine the rejection region, using the $\alpha = 0.05$ significance level.

(b) Determine the probability of Type II error, still using the $\alpha = 0.05$ significance level, if the underlying mortality rate actually equals 0.20.

(c) Repeat (a) and (b) using the $\alpha = 0.01$ significance level. Describe how the rejection region and Type II error probability change between the two values of α, and explain why these changes make intuitive sense.

38. Compounding Errors

Suppose that you conduct 10 independent tests of significance, using the $\alpha = 0.05$ significance level for each test. Also suppose that, unknown to you, the null hypothesis is actually true for every test. Let the random variable R be the number of tests for which you (mistakenly) reject the null hypothesis.
(a) Does rejecting the null hypothesis constitute a Type I or a Type II error in this situation?
(b) Explain why the random variable R can be considered as having a binomial distribution, and specify its values of n and π.
(c) Determine the expected value of R, the number of tests for which you (mistakenly) reject the null hypothesis.
(d) Determine the probability that you mistakenly reject the null hypothesis for at least one of these tests (i.e., that $P(R > 0)$).
(e) Repeat (c) and (d) if you conduct 20 independent tests of significance, using the $\alpha = 0.05$ significance level for each test, where the null hypothesis is actually true for every test.
(f) Repeat (e), supposing that you use the $\alpha = 0.10$ significance level for each test.
This exercise illustrates a concern with conducting a large number of tests of significance.

39. Compounding Errors (cont.)

Reconsider the previous exercise. Let n represent the number of independent tests to be performed, and let α represent the significance level. Continue to assume that, unknown to you, the null hypothesis is actually true for every test. Continue to let the random variable R be the number of tests for which you (mistakenly) reject the null hypothesis.
(a) Determine the expected value of R, as a function of n and π.
(b) Is this expected value an increasing or a decreasing function of n? Explain why your answer makes intuitive sense.
(c) Is this expected value an increasing or a decreasing function of α? Explain why your answer makes intuitive sense.
(d) Determine the probability that you mistakenly reject the null hypothesis for at least one of these tests (i.e., that $P(R > 0)$), as a function of n and α.
(e) Is this probability an increasing or a decreasing function of n? Explain why your answer makes intuitive sense.
(f) Is this probability an increasing or a decreasing function of α? Explain why your answer makes intuitive sense.

40. Rock n' Roll (cont.)

Reconsider Exercise 4. Suppose that we obtain a list of all songs with a color in the title, and suppose that we take a random sample of 75 songs from that list and determine the sample proportion, $\hat{p}$, of songs that are about the color blue. Suppose for now that the claim (30% of all songs with a color in the title are about the color blue) is true.
(a) Verify that the CLT for a sample proportion applies here.
(b) What does the CLT say about the sampling distribution of $\hat{p}$? (Mention shape, center, and spread, and also draw a well-labeled sketch of the sampling distribution.)
(c) Use the CLT to approximate the probability that between 24% and 36% of the songs in the sample will be "blue." Include a one-sentence summary of what the calculated probability signifies.
(d) Use the binomial distribution to calculate the probability in (c) exactly. [*Hint:* First use the sample size of 75 to convert the 24% and 36% into counts.]
(e) How close did the normal approximation in (c) come to the exact binomial probability in (d)?
(f) Use the continuity correction to improve on the normal approximation in (c).

(g) Repeat (c)–(f) for finding the probability that between 20% and 40% of the songs in the sample will be "blue."

41. Rock n' Roll (cont.)

Reconsider again the claim that 30% of all songs with a color in the title are about the color blue. Again suppose that we obtain a list of all songs with a color in the title, and suppose we take a random sample of 75 songs from that list and determine the sample proportion, $\hat{p}$, of songs that are about the color blue.

(a) Using the CLT, determine the values of $\hat{p}$ that would lead to rejecting the null hypothesis at 30% of all songs with a color in the title are about the color blue at the 5% level of significance.

(b) Suppose that actually 40% of the population of all songs with a color in the title are about the color blue. Sketch a graph of the resulting sampling distribution, using 40% as the value for π, and using the values you determined in (a), determine the probability that we would fail to reject the null hypothesis of 30% at the 5% level of significance. Is this a Type I or a Type II error?

42. Competitive Advantage of Uniform Color? (cont.)

For the Competitive Advantage of Uniform Color data (Exercise 10), assuming $\pi = 0.5$:

(a) Is the sample size in this study large enough for the Central Limit Theorem (CLT) for a sample proportion to apply?

(b) Assuming the CLT applies, specify the shape, mean, and standard deviation predicted by the CLT for the distribution of sample proportions. Include a well-labeled "sketch" (by hand or computer rendered) of this distribution. [*Hint*: By well-labeled, we mean a horizontal axis label and indication of scaling on the horizontal axis showing the mean and one standard deviation on each side of the mean.]

(c) Compare the predicted distribution with simulation results from the One Proportion Inference applet (comment on each of shape, mean, and SD). Be sure to include your output.

(d) Use technology to estimate the probability of obtaining a sample proportion of at least 0.543 using the normal approximation. (Be very clear how you have done so and include your output.) How does this compare to the p-value you found using the binomial distribution?

43. Which Tire? (cont.)

Reconsider Exercise 8. Here are results from a different class:

Left front	Left rear	Right front	Right rear
11	5	19	2

(a) What proportion of students picked the right front tire? Is this in the direction of the research conjecture?

(b) Would it be valid to apply the one-proportion *z*-test with these data? Explain.

(c) Regardless of your answer to (b), calculate the one-proportion *z*-test using technology. Include your output and report the test statistic and p-value (both with and without a continuity correction). [*Hint*: To do the continuity correction in Minitab or Normal Probability Calculator applet, go back to the normal distribution, using the mean and SD specified by the CLT.]

(d) Write a one-sentence interpretation of the test statistic in this context.

(e) What *test decision* (reject or fail to reject the null hypothesis) would you make based on this p-value?

(f) Regardless of your answer to (b), calculate and interpret a 95% *z*-confidence interval for the parameter of interest.

(g) Explain what is meant by the phrase "95% confidence" in this context.

(h) Reconsider the class of 54 students, in which 24 picked the right front tire. *Without carrying out any new inference procedure calculations*, explain how the test statistic, p-value, and confidence interval for the 54 student class to compare to the results you found here. Justify your answers.

44. Magical Numbers
Think of a number satisfying these conditions:
- The number has two digits.
- Both digits are odd.
- The two digits are different from each other.

(a) Report the number that you think of.
(b) How many numbers satisfy these conditions?

A magician on television claimed that the number 37 is somehow special and so people tend to pick it more often than would be expected by chance.

(c) If the numbers satisfying these conditions are all equally likely, what is the probability that a person would pick the number 37?
(d) State the null and alternative hypotheses for testing the magician's claim.
(e) Describe what Type I and Type II errors signify in this situation.
(f) With a sample of 25 people, how many would have to pick 37 in order for the result to be statistically significant at the 0.05 level?
(g) Suppose that the probability of picking 37 is actually 0.10, twice as high as equal likeliness would suggest. Determine the probability that a sample of 25 people would produce a result in the rejection region of (f). Write a sentence or two interpreting what this probability reveals.
(h) What term describes the probability that you calculated in (g) – *significance, confidence, power*, or *influence*?

45. Magical Numbers (cont.)
Previous studies have investigated the question of whether people tend to think of an odd number when they are asked to think of a single-digit number (0 through 9). Combining results from several studies, Kubovy and Psotka (1976) used a sample size of 1770 people, of whom 741 thought of an even number and 1029 thought of an odd number.

(a) Use the normal approximation (Wald procedure) to determine a 95% confidence interval for the relevant parameter. Show how to calculate this interval by hand.
(b) Interpret what this confidence interval in (a) says. [*Hint*: You are 95% confident of what?]
(c) Determine a 99% confidence interval and a 99.9% confidence interval for the relevant parameter based on the sample data. (Feel free to use technology.)
(d) Comment on how the three confidence intervals compare to each other. [*Hint*: Be sure to comment on both width and midpoint.]
(e) Based on these confidence intervals, would you say that the sample data provide strong evidence that people have a tendency to think of an odd number rather than an even number? Explain how you are deciding.
(f) Suppose that you want to estimate the relevant parameter to within ± 0.02 with 99% confidence. Determine how many people you would have to ask. [*Hint*: Use the result of the previous studies in determining the sample size.]
(g) Re-answer (f) without using the result of the previous studies, so using 0.5 as an estimate for the observed proportion. Comment on how this changes the necessary sample size.

The researchers found that 503 of the 1770 people thought of the number 7.

(h) Determine a 95% confidence interval for the long-run probability that a person thinks of the number 7. Does this confidence interval suggest that people have a tendency to think of the number 7 more often than by random chance? Explain.

46. Competitive Advantage from Uniform Color? (cont.)

Recall the study of 457 matches in four combat sports (boxing, tae kwon do, Greco-Roman wrestling, freestyle wrestling) at the 2004 Olympic Games. Competitors in these sports were randomly assigned to wear either a red or a blue uniform. The researchers found that the competitor wearing red defeated the competitor wearing blue in 248 matches, and the competitor wearing blue emerged as the winner in 209 matches.

(a) Identify the observational units and variable in this study.
(b) Verify the conditions for using the Wald (z-) procedure to determine a 95% confidence interval for the probability that the competitor wearing red wins a match.
(c) Calculate this 95% confidence interval.
(d) Interpret what this interval reveals: We are 95% confident that …
(e) Interpret what the 95% confidence *level* means in this context.
(f) Repeat (c) for a 99% confidence interval.
(g) Describe how these two intervals compare, in terms of both their midpoints and widths.
(h) Do these intervals suggest that one uniform color or the other provides a competitive advantage? Explain.
(i) Suppose that the sample size had been four times larger, and the sample proportion had been identical to the actual study. Determine a 95% confidence interval in this case, and comment on how it compares to the interval in (c). [*Hint*: Be as specific as possible, and be sure to comment on both midpoint and width.]
(j) Determine how large a sample size would be necessary to estimate the actual probability to within 5 percentage points with 90% confidence.

47. Baseball Big Bang?

A reader wrote in to the "Ask Marilyn" column in *Parade* magazine to say that his grandfather told him that in 3/4 of all baseball games, the winning team scores more runs in one inning than the losing team scores in the entire game. (This phenomenon is known as a "big bang.") Marilyn responded that this probability seemed to be too high to be believable. Let π denote the actual probability that a Major League Baseball game results in a "big bang."

(a) Restate the grandfather's assertion as a null hypothesis, in symbols and in words.
(b) Report Marilyn's response as an alternative hypothesis, in symbols and in words.

To investigate this claim, one of the authors examined the 45 Major League baseball games played on September 17 – 19, 2010, and found that 21 of these 45 games contained a big bang.

(c) Calculate the sample proportion of games that had a big bang, and denote it with the appropriate symbol.
(d) If the grandfather's claim is true, how many standard deviations below the mean is the observed sample proportion? Also denote this with the appropriate symbol.
(e) Use the normal distribution to determine the approximate p-value, first without using the continuity correction and then with using the continuity correction. Also produce (and submit) an appropriately labeled and shaded graph for each of these normal calculations.
(f) Would you conclude that the sample data provide strong evidence to support Marilyn's contention that the proportion cited by the grandfather is too high to be the actual value? Explain your reasoning, as if writing to the grandfather, who has never taken a statistics course.

(g) Marilyn went on to assert that she believes the actual probability of a big bang to be 0.5. Conduct a two-sided test of this hypothesis. Report the hypotheses, test statistic, and p-value. Again perform the calculations with and without using the continuity correction. Also calculate the exact p-value from the binomial distribution. Produce (and submit) appropriately labeled shaded graphs for all of these calculations. Comment on whether the continuity correction is helpful here. State the test decision at the $\alpha = .10$ significance level, and summarize your conclusion.

48. Political Jokes?
The University of Pennsylvania's National Annenberg Election Survey of 2004 studied the humor of late-night comedians. They performed a content analysis of the jokes made by Jon Stewart during the "headlines" segment of *The Daily Show* from July 15 – September 16, 2004. They found that 83 of the 252 jokes were of a political nature.
(a) Identify the observational units and the variable in this study.
(b) Produce a graphical and numerical summary, along with a one-sentence description, of this sample.
(c) Consider these 252 jokes to be a random sample from the joke-producing process for this show. Check technical conditions, and use these sample data to produce 90%, 95%, and 99% confidence intervals for the underlying probability that a joke is political in nature.
(d) How do the widths of these confidence intervals compare? Explain why that makes sense.
(e) How do the midpoints of these confidence intervals compare? Explain why that makes sense.
(f) Based on these confidence intervals, what can you say about the p-value for testing whether half of *The Daily Show* jokes are political in nature? Explain.
(g) Give a value of π that would be rejected at the 5% level but not at the 1% level. Explain.

49. Halloween Treat Choices (cont.)
Inspired by the Halloween treat study described in Investigation 1.9, a pair of students (Avery & Botts, 2004) conducted their own study of whether trick-or-treaters would choose fruit-flavored candy (high in sugar) or chocolate candy (high in both sugar and fat). The students gave justification for why they felt the sample they obtained was representative of all trick-or-treaters in their community. Prior to conducting the study, the students wanted to test whether half of children would choose the chocolate candy, while some classmates conjectured that two-thirds would choose the chocolate candy, and others proposed that three-fourths would choose the chocolate candy.
(a) State the null and alternative hypotheses, in words and in symbols, for testing the students' suspicion that half of children would choose chocolate candy.
The students found that of 127 of 191 children in the study chose the chocolate candy.
(b) Calculate the proportion of children in the study who chose the chocolate candy. Is this a parameter or a statistic? What symbol do we use to denote it?
(c) Verify that the technical conditions for the one proportion *z*-test are satisfied in this study.
(d) Calculate the z-test statistic and p-value, and include a well-labeled sketch of the sampling distribution denoting the area represented by the p-value. Would you reject the null hypothesis at the 0.10 level? What about the 0.05 and 0.01 levels? What is the smallest level of significance for which you would reject the null hypothesis here?
(e) Conduct significance tests for the other two hypotheses suggested above. Report the hypotheses, technical condition checks, test statistics, and p-values. Summarize your conclusions.
(f) After verifying the appropriate technical conditions, determine and interpret a 90% confidence interval for the population proportion who would choose the chocolate candy. Is this interval consistent with the p-values in (d)? Explain.
(g) Without knowing more about the study, to what population would you be willing to generalize your findings?

50. Halloween Treat Choices (cont.)
Reconsider the previous question.
(a) If two-thirds of the trick-or-treaters in this community prefer the chocolate candy, what is the power of
   the test at the 5% level of significance? (Include relevant sketches.)
(b) Repeat (a) if three-fifths of the trick-or-treaters in this community prefer the chocolate candy.
Now consider how details of your analysis would change if the parameter of interest was the proportion
of children who would choose the fruit candy, rather than the proportion who would choose the chocolate
candy.
(c) How would this affect (if at all) your confidence interval calculation and interpretation? Explain.
(d) How would this affect (if at all) your p-value calculation and test conclusion for testing the value .5
   against the two-sided alternative? Explain.

51. *Cohen v. Brown University*
In 1991, a suit was filed against Brown University after Brown terminated funding for its women's
gymnastics and volleyball teams and its men's water polo and golf teams. The suit charged that Brown
was violating Title IX of the Education Amendments of 1972, the federal law that prohibits sex
discrimination by all educational institutions receiving federal funds. This requires men and women to
have equivalent opportunities for participation. A main component of the plaintiff's case was that while
51% of the undergraduate student body was women, only 38% of the 897 students engaged in
intercollegiate athletics were women. If there is no gender discrepancy, then Title IX assumes that the
proportion of women athletes should be similar to the proportion of women in the overall student body.
However, we know the sample result can deviate from this population proportion "just by chance."
Suppose we were randomly selecting students to be athletes, the question is whether this random process
could lead to such a disparity in these proportions. Although we know the proportion of women in the
population of all university students to be 0.51, we don't know the probability of an athlete being female.
Let π refer to the probability of a Brown university athlete being female.
(a) What are the observational units and variable of interest in this study? Also specify the population or
   process of interest. [Note: Although we have not actually taken a random sample here, we will
   assume the sample of 897 students is representative of this overall process.]
(b) State a null and an alternative hypothesis about the value of π (in symbols and in words).
(c) We can assess the strength of evidence against the null hypothesis by treating the 897 current athletes
   as a random sample from the process of athlete determination at Brown. Carry out a one-proportion z-
   test to assess whether the observed sample proportion ($\hat{p} = 0.38$) is surprising under the null
   hypothesis. Check and comment on the technical conditions of this procedure. Would it be beneficial
   to include a continuity correction? Explain.
(d) Based on this p-value, what conclusion do you draw about whether this discrepancy could have arisen
   by chance? In other words, does your analysis suggest that the proportion of women involved in
   intercollegiate athletics is significantly lower than the proportion of women students at the university?
   Explain.

52. Random Babies
Recall the scenario from Investigation B: Suppose that on one night at a certain hospital, four mothers
give birth to baby boys. As a very sick joke, the hospital staff decides to return babies to their mothers
completely at random. In this situation, the theoretical probability of obtaining zero matches is 9/24, or
0.375. For now let π denote this value, so $\pi = 0.375$. Suppose that you conduct 1000 repetitions of
simulating this process of returning babies to their mothers at random, so each repetition has probability
0.375 of resulting in zero matches. Let $\hat{p}$ denote the proportion of your 1000 repetitions that result in
zero matches.

(a) Describe what the Central Limit Theorem says about the sampling distribution of $\hat{p}$ if you were to repeatedly conduct simulations of 1000 repetitions each. (Comment on shape, center, and spread, and draw a well-labeled sketch of the sampling distribution.)

(b) Between what two values will $\hat{p}$ fall, with probability 0.05? [*Hint*: Go 1.96 standard deviations on either side of π.]

(c) If you want the simulation to approximate π to within ± 0.01 with 99% confidence, how many repetitions are necessary?

(d) Now let π represent the theoretical probability that all four mothers get the correct baby. Then π = 1/24, but suppose that you want the simulation to approximate π to within ± 0.005 with 99% confidence. Determine how many repetitions of the simulation would be required.

The previous exercise shows that you can use the Central Limit Theorem to determine how many repetitions of a simulation are necessary to estimate the underlying theoretical probability with a desired level of accuracy and precision.

53. *Swain v. Alabama*

In *Swain v. Alabama* (1965) it was alleged that there was discrimination against blacks in grand jury selection. Swain, a black man, was convicted in Talladega County, Alabama, of raping a white woman and was sentenced to death. At that time in Alabama, only men over the age of 21 were eligible for jury duty. Census data suggested that about 26% of those eligible for grand jury service were black, yet a "random sample" of 1050 individuals called to appear for possible grand jury duty yielded only 177 blacks.

(a) Produce numerical and graphical summaries of the sample results.

(b) Define the parameter of interest and indicate the symbol used to represent it.

(c) State the null and alternative hypotheses.

(d) Describe what a Type I and a Type II error would represent in this situation. What are the consequences of each type of error?

(e) Is the normal model valid here? If so, use the CLT to calculate the test statistic and p-value (include a sketch). If not, calculate the p-value based on the binomial distribution.

(f) Do you reject or fail to reject the null hypothesis based on this p-value? The Supreme Court ruled that this disparity was small. Do you agree with this decision?

54. Detecting Cancer (cont.)

Recall from the Exercise 23 the study on whether dogs' sense of smell can be helpful in detecting cancer in humans. Each dog used in the study was presented with seven urine samples to smell, with only one of the urine samples coming from a patient suffering from bladder cancer. Taken as a group, the dogs were tested on 54 trials, and they correctly identified the urine from the cancer patient in 22 trials. Even though the technical conditions for the one proportion *z*-test are not quite satisfied here, we can still calculate a *z*-statistic, which will tell us how many standard deviations the observed results falls from what the null hypothesis would predict. But we should not proceed to calculate a p-value from the normal distribution, because the normal approximation to the binomial distribution is not valid here.

(a) State the null and alternative hypotheses for testing whether the dogs are able to correctly identify the urine from the cancer patient more often than random chance would predict.

(b) Show that the technical conditions for the one proportion *z*-test are not quite satisfied here.

(c) Determine how large the sample size would need to be in order for calculation of the p-value of the one proportion *z*-test to be appropriate here.

(d) Check technical conditions, and then use the Wald or adjusted Wald procedure to produce and 95% confidence interval for the population proportion of correct identifications. As always, include a one-sentence interpretation of the interval.

(e) Comment on what these intervals reveal about the dogs' effectiveness at detecting urine cancer.

55. Pilot Study (cont.)
Recall from Exercises 15 and 16 that we encountered one female in a sample of eight pilots on a recent trip.
(a) Are the technical conditions for the Wald confidence interval procedure satisfied here?
(b) Use the adjusted Wald procedure to estimate the population proportion of all commercial pilots who are women, based on these data.
(c) Using this sample as a pilot study (pardon the pun), determine how large a sample would be needed to estimate the population proportion to within $\pm$ 0.075 with 90% confidence. [*Hint*: Remember to report n, not $n + 4$.]

56. Simulating Wald Intervals
One of the criticisms against the Wald method is that its performance can be quite haphazard, meaning that it doesn't always improve as n increases with a fixed value of π, and it doesn't always improve as π gets closer to 0.5 for a fixed value of n. You will investigate this claim using the Simulating Confidence Intervals applet. Start with the population proportion $\pi = 0.5$ and the sample size $n = 14$.
(a) Is the sample size condition for the validity of the Wald method satisfied here?
(b) Simulate the drawing of 200 random samples, and the construction of 95% confidence intervals for π, in this situation. Then keep clicking on "sample" until you have produced at least 5000 samples and their resulting intervals. What percentage of the nominal 95% confidence intervals succeed in capturing the actual value of π?
(c) Is this percentage reasonably close to 95%? How well would you say the Wald procedure performs in this situation?
(d) Repeat (a)–(c) with the population proportion $\pi = 0.5$ and the sample size $n = 40$.
(e) In which of these situations ($n = 14$ or $n = 40$) should the Wald method have performed better? Why? Is that how it turned out?
(f) Repeat this analysis by comparing the performance of the Wald procedure when $n = 100$ and $\pi = 0.106$ to its performance when $n = 100$ and $\pi = 0.107$. Summarize your findings.
(g) Apply the adjusted Wald procedure in all four of these situations, and comment on its performance relative to that of the Wald procedure. Does its performance appear to improve as n increases?

57. Hometown SUVs
Suppose that you want to estimate the proportion of vehicles on the road in your hometown that are sport utility vehicles (SUVs). You decide to stand at the intersection closest to your home between 7 and 8 A.M. every morning for a week, keeping track of how many vehicles go by and how many of them are SUVs.
(a) Identify the population, sample, parameter, and statistic (all in words) in this study.
(b) Give some reasons why your sampling method is probably not unbiased.

58. Hometown SUVs (cont.)
Reconsider the previous question. Now suppose that change your sampling plan and go to a local car dealership. You ask the manager to let you inspect a random sample of vehicles sold in the past year.
(a) Identify the sampling frame in this study.
(b) Is this sampling method likely to be unbiased for estimating the proportion of SUVs among all vehicles on the road in your town? Explain.

59. Representative Samples?
Suppose that you are asked to use the students in your current statistics class as a sample from the population of all students at your school. This is not literally a *random* sample, but whether or not this sample is *representative* of the population could depend on the variable of interest.
For each of the following variables, discuss whether you think the sample would be representative of the population. Briefly explain your answer.
(a) Sex
(b) Time spent sleeping last night
(c) Knowledge of statistics
(d) Political party affiliation
(e) Number of movies seen in past year
(f) Age

60. Average Word Lengths
Suppose that you want to estimate the average length of a word (measured by number of letters) in this book.
(a) Is this number a parameter or a statistic? Explain. Also indicate what symbol would be commonly used to denote it.
Suppose that your sampling plan is to open the book haphazardly, set your finger on that page haphazardly, record the number of letters in the word that your finger lands on, and then repeat this process.
(b) Does this constitute a simple random sample? Explain.
(c) Is this sampling plan likely to be unbiased? If so, explain why. If not, indicate whether the sample mean is likely to over- or under-estimate the population mean, and explain your answer.
Now suppose that you decide to select a page at random, examine all of the words on that page, and calculate the average length of the words on that page.
(d) Is this number a parameter or a statistic? Explain. Also indicate what symbol would be commonly used to denote it.
(e) Answer (c) for this sampling plan.
(f) Explain how you could use a multi-stage sampling plan in this situation. Provide enough detail that someone else could implement your plan without further guidance from you.

61. Baseball Careers
Think about the average number of years in the career of a professional baseball player among all players who have played in the major leagues.
(a) Is this a parameter or a statistic? Explain, and indicate what symbol we would use to denote it.
(b) Suppose we estimate this average with the average number of years in the career of a player in the Baseball Hall of Fame. Is this likely to be an underestimate or an overestimate? Explain.
(c) If we estimate this average with the average years that current players have been playing, are we likely to obtain an overestimate or an underestimate? Explain which and why.

62. Sampling Senators
Suppose we want a sample of five members of the current U.S. Senate and decide to ask you to name five senators. We then determine how many years these senators have served in the Senate and calculate their average service time.
(a) Identify the population, sample, parameter, and statistic in this study.

(b) Do you think this sampling method will be biased? If so, do you think it will tend to overestimate or underestimate the parameter of interest? Explain.

63. Sampling Senators (cont.)
Suppose that we take the current members of the U.S. Senate as a sample from the population of all adult Americans. For each of the following, indicate whether you would expect the sample statistic to be an underestimate, and overestimate, or a fairly accurate estimate of the population parameter. Explain your answers.
(a) Proportion of women
(b) Average years of education
(c) Average annual income
(d) Standard deviation of ages
(e) Proportion that live below the poverty line

64. Emotional Support?
In the mid-1980s sociologist Shere Hite undertook a study of women's attitudes toward relationships, love, and sex by distributing 100,000 questionnaires through women's groups. Of the 4500 women who returned the questionnaires, 96% said that they give more emotional support than they receive from their husbands or boyfriends. Around the same time, an ABC News/*Washington Post* poll surveyed a national random sample of 767 women, finding that 44% claimed to give more emotional support than they receive. Consider the population of interest for both surveys to be all American women.
(a) Identify (in words) the parameter of interest for both polls.
(b) With each poll, determine a 90%, 95%, and 99% confidence interval for the parameter. Calculate one of these confidence intervals by hand, but feel free to use technology (applet, R, Minitab) for the others.
(c) Which poll has the smaller margin-of-error? Explain why this poll has the smaller margin-of-error.
(d) Which poll's results do you think are more representative of the truth about the population of all American women? Explain.
(e) Which polling method do you think is more likely to be biased in a particular direction? Explain your answer, and also indicate whether you think that poll's statistic is an overestimate or underestimate of the population parameter.
(f) Determine the sample size that would be needed to estimate the population parameter to within $\pm$ 0.025 with 95% confidence. Use both 0.5 and the statistic from the ABC News/*Washington Post* poll to perform this calculation, and comment on how your answers differ.
(g) Based only on your confidence intervals from the ABC News/*Washington Post* poll, does 0.5 appear to be a plausible value for the proportion of all American women who would answer "yes" to this question? Explain.
(h) Conduct a (normal-based) significance test of whether 0.5 is a plausible value for the proportion of all American women who would answer "yes" to this question, based on the data from the ABC News/*Washington Post* poll. Report the hypotheses, test statistic, and p-value. State your test decision at the α = 0.10, 0.05, and 0.01 significance levels, and summarize your conclusion.

65. The *Literary Digest*
Recall from Investigation 1.13 that the *Literary Digest* conducted a large sample survey of 2.4 million people and found 57% indicating they would vote for Alf Landon in the upcoming 1936 election.
(a) Use technology to produce a 99.9% confidence interval for the proportion of all voters who were planning to vote for Alf Landon based on this sample data.

(b) Explain why this interval is so narrow.
(c) Explain why the results of this interval provided such an erroneous estimate (the actual value of π turned out to be 0.37 as Landon received only 37% of the vote in the nation election).

66. Renaming Brides

How often do brides keep their own last name, or adopt their husband's last name, or use a combination of both names? Researchers investigated these questions by examining a random sample of 600 heterosexual wedding announcements appearing in *The New York Times* between the years 2001 and 2005 (Kopelman et. al., 2009). They found that 18% of the 600 brides kept their own last name.

(a) Check the sample size condition(s) for whether a one-proportion z-interval can be used to estimate the population proportion of brides between 2001 and 2005 who kept their own last name.
(b) Calculate a 95% confidence (one-proportion z-) interval for the population proportion of brides between 2001 and 2005 who kept their own last name.
(c) Interpret the confidence interval that you determined in (b).
(d) To what population would you feel comfortable generalizing this confidence interval: all brides in the world between 2001 and 2005, or all brides in the United States, or …? Justify your response.
(e) Determine the sample size needed to estimate the population proportion of brides between 2001 and 2005 who kept their own last name to within $\pm$ 0.03 with 95% confidence. Show the details of your calculation.

The Kopelman et. al. researchers reported that in addition to the 18% of the 600 brides who kept their own last names, another 10% of the 600 brides kept their maiden name in a modified form, for example by using her maiden name as her middle name or by hyphenating her maiden name with her husband's name. Moreover, the researchers reported that 45% of the 600 brides adopted their husband's last name, and no information about names could be determined for the remaining 27% of the 600 brides.
Now eliminate from the sample the 27% of the 600 brides for whom name information could not be determined, and combine those who kept their own last name and those who kept their maiden name in modified form into one group.

(f) What proportion of the new sample (those brides for whom name information could be determined) kept their last name in full or modified form?
(g) Use your answer to (f) to produce a 95% confidence interval for the relevant parameter. Also interpret this interval.

67. Volunteerism

In September of 2003 the Bureau of Labor Statistics conducted a study of volunteerism in the United States as part of its ongoing Current Population Survey. A random sample of about 60,000 people (age 16 and over) was asked whether they had participated in any volunteer activities through or for an organization in the past year. It turned out that 28.8% said that they had participated in volunteer activities.

(a) Produce a graphical and numerical summary, with a one-sentence description, about this sample.
(b) Use these data to conduct a one proportion z-test of the hypothesis that 30% of adult Americans participated in volunteer activities that year, against the alternative that this percentage was not 30%. State the hypotheses in symbols and words, check the technical conditions, calculate the test statistic and p-value, and indicate whether the result is statistically significant at the 0.01 level. Include a well-labeled sketch of the sampling distribution and indicate the area represented by the p-value. Also summarize your conclusion about the population and explain how it follows from your test.
(c) Follow up on this test by producing a 99% confidence interval for the population proportion who participated in volunteer activities that year. Include a one-sentence interpretation of this interval.

(d) Is the value 0.3 contained within the confidence interval? Is this consistent with your test decision? Explain.

(e) Explain the difference in the following statements. Also indicate which is the appropriate conclusion to draw from the data in this particular study, and explain your choice.

- The sample data provide strong evidence that the proportion of adult Americans who participated in volunteer activities was very different from 30%.
- The sample data provide very strong evidence that the proportion of adult Americans who participated in volunteer activities was different from 30%.

68. A Penny for Your Thoughts (cont.)

In June of 2004, the Harris organization asked a random sample of 2136 adult Americans: "Would you favor or oppose abolishing the penny so that the nickel would be the lowest denomination coin?" It turned out that 59% of those adults answered "oppose."

(a) Identify the observational units and variable in this study.

(b) Identify the statistic and parameter in this study.

(c) Check whether the technical conditions for the Wald confidence interval procedure are satisfied in this study.

(d) Determine and interpret a 95% confidence interval for the population proportion who oppose abolishing the penny. (Use the Wald interval if appropriate, the adjusted Wald procedure if not.)

(e) Does this interval suggest that more than half of all American adults oppose abolishing the penny? Explain.

(f) If the study had involved a random sample of 1136 adult Americans, and all else had turned out the same, would the width of this confidence interval change? If so, how? Explain briefly.

(g) If the study had involved a random sample of 1136 adult Americans, and all else had turned out the same, would the midpoint of this confidence interval change? If so, how? Explain briefly.

(h) If another random sample of 2136 adult Americans were asked this question, is the probability 0.95 that the sample proportion who answer "oppose" would fall within the interval from (d)? Explain briefly.

69. Backpack Weights

A growing problem in American schools involves students who develop back problems, possibly as a result of carrying too much weight in their backpacks. It has been proposed that students should carry no more than 10% of their body weight in their backpack. To investigate how much students carry in their backpacks, student researchers at Cal Poly sampled 100 students. They asked these students to report their body weight, and they weighed how much was carried in their backpack. These data, along with the ratio of backpack weight to body weight, are in the file backpack.txt.

(a) Use technology to create a new variable: whether or not the student carries at least 10% of his/her body weight in his/her backpack. Is this a quantitative or a categorical variable?

(b) Examine graphical and numerical summaries of this variable, and comment on its distribution in this sample.

(c) Conduct a significance test of whether the sample data suggest that the less than half of all Cal Poly students carry at least 10% of their body weight in their backpack. Report hypotheses, comment on technical conditions, and calculate the test statistic and p-value. Include a well-labeled sketch of the sampling distribution for the test statistic and indicate the area represented by the p-value. Also summarize your conclusion and explain how it follows from your test.

(d) Construct and interpret a 90% confidence interval for the population proportion who carry at least 10% of their body weight in their backpack. [*Hint*: Be sure to check whether the technical conditions for the Wald procedure are satisfied; use the adjusted Wald procedure if not.]

(e) Construct and interpret a 95% confidence interval for the population proportion who carry at least 15% of their body weight in their backpack.

70. Left-handed Vegetarians

In a class of 36 statistics students, 6 reported that they were left-handed and 3 that they were vegetarians.
(a) Is this sample a random one from the population of all students at the college? Would you expect it to be representative of the population of all students at the college with regard to these variables?
For the remainder of this exercise, treat the sample as if it were a random one from the population.
(b) Conduct a binomial test of whether the population proportion of left-handers differs from 10%. First construct a bar graph to display the sample data, then state the hypotheses (in symbols and in words), and finally report the smallest significance level α at which you would reject this hypothesis. Also summarize your conclusion.
(c) Repeat (b) for testing whether the population proportion of vegetarians differs from 10%.
(d) With which variable do the sample data provide stronger evidence that the corresponding population parameter is not .10? Explain your answer.

71. Left-handed Vegetarians (cont.)

Reconsider the class results from the previous exercise.
(a) For each of the two variables (left-handedness and vegetarianism), determine a 95% binomial confidence interval for the population parameter.
(b) Determine the width of each interval.
(c) Because the sample size is the same for both variables, one might expect the two intervals to have the same widths. Do they? If not, which is wider? Explain why this makes sense.

72. Time Travelers

Reconsider the previous two exercises. The 36 students were also asked: "If time travel were possible, would you prefer to travel to the past or the future?" The past was preferred by 20 students and the future by 16.
(a) Does this sample provide much evidence to doubt that the percentage break-down in the population would be 50/50? Produce a bar graph, conduct a binomial test, and explain your conclusion.
(b) Determine and interpret a 95% binomial confidence interval for the population parameter. Also clearly identify in words what this parameter represents.
(c) If you define "success" to be "future" and another student defines success to be "past," how would your p-values for testing whether the population proportion differs from one-half compare? Explain.
(d) How would the competing definitions in (c) affect the confidence interval for the population proportion? Explain.

73. Left-Handed Vegetarians (cont.)

Reconsider the previous exercises which reported the results of a class survey on three questions:
- Of the 36 students, 3 were vegetarian.
- Of the 36 students, 6 were left-handed.
- Of the 36 students, 20 preferred to travel to the past and 16 to the future.

(a) For each of these three questions, estimate the relevant population proportion (again treat these as if they were random samples from the population) with a 95% Wald confidence interval, a 95% adjusted Wald confidence interval, and a 95% binomial confidence interval.

(b) Comment on how similar the three types of intervals are. For which question are they most similar? For which question are they most different?

74. 2015 Wimbledon

Suppose that you want to gather data to test whether half of men's professional tennis matches last for the maximum possible number of sets. Before collecting the data, you do not have a conjecture for whether this proportion would be greater than or less than one-half if it does not equal one-half.

(a) State the appropriate null and alternative hypotheses, in symbols and in words.

In the 2015 Wimbledon Championship 16 of the 64 first round men's singles tennis matches lasted for 5 sets (the maximum number possible).

(b) Were these data gathered from a random sample of men's professional tennis matches in 2015? Were these data gathered from a random sample of matches in the 2015 Wimbledon? Explain.

(c) Do you have reason to suspect that this data collection method might be biased with regard to how long professional tennis matches last? Explain.

(d) Despite any concerns about the sampling method, use these data to conduct a test of the hypotheses in (a). Report the p-value, and describe how you calculated it. Do these data provide convincing evidence (at the α = 0.05 level) against the null hypothesis?

75. 2015 Wimbledon (cont.)

Reconsider the previous exercise. Of the 64 women's singles tennis matches in the first round of the 2015 Wimbledon Championship, 20 matches lasted for 3 sets (the maximum number possible). Suppose that you use these data to test whether half of all professional women's singles matches last for 3 sets.

(a) Without doing any calculations, explain whether the p-value of this test will be larger or smaller than the p-value based on the men's data, where 16 of 64 matches lasted for the maximum number of sets.

(b) Conduct the test based on the women's data. Report the p-value, and describe how you calculated it. Do these data provide convincing evidence (at the α = 0.05 level) against the null hypothesis?

(c) What is the smallest significance level at which the null hypothesis would be rejected? Explain. [*Hint*: Do not confine yourself only to the conventional significance levels.]

(d) Without calculating it, indicate whether you expect a 95% confidence interval for the underlying parameter to include the value 0.5. Explain.

76. Prescribing Placebos

Many medical studies compare a new treatment against a placebo, because many patients respond well even to a placebo. This has led some to question whether there might be circumstances in which it would be appropriate for a physician to administer a placebo to a patient. Nitzan and Lichtenberg (2004) surveyed 89 physicians and head nurses in Israel, chosen from across all medical and surgery units at two hospitals. They found that 53 admitted administering placebos to patients.

(a) Identify the population of interest and the sample selected.

(b) Identify the parameter and statistic (both in words and with a symbol). Also calculate the value of the statistic.

(c) How large must the population (total number of physicians and head nurses in Israel) be in order for the binomial model to be a reasonable one here? Explain.

(d) Use technology to determine a 95% binomial confidence interval for π, the proportion of all physicians and head nurses in Israel who prescribe placebos to patients.

(e) Now determine a 90% binomial confidence interval for π, and comment on how it compares to the 95% interval.

(f) Would you use these intervals to estimate the proportion of American physicians who prescribe placebos? Explain.

77. Prescribing Placebos (cont.)

Reconsider the previous exercise about prescribing placebos. The researchers stated that they had expected no more than 10% of clinicians to use placebos.

(a) State the null and alternative hypotheses, in symbols and in words, for testing whether the sample data provide evidence that more than 10% of clinicians have used placebos (i.e., to invalidate what the researchers expected).

(b) Conduct a binomial test of the hypotheses that you stated in (a). Report the p-value and summarize your conclusion.

(c) Based on the confidence intervals that you determined in the previous exercise, what can you say about the p-value for conducting a two-sided test of whether the population proportion differs significantly from 0.5? Explain. (Do not conduct the test; base your answer on the confidence intervals.)

78. Standard Deviation Properties

The standard deviation of a sample proportion is given by $\sqrt{\pi(1-\pi)/n}$.

(a) Explain what each of the three symbols ($\hat{p}$, π, and n) in this equation represent.

(b) Take the derivative of this function with respect to π and find the value of π that maximizes SD($\hat{p}$) for a fixed value of n. [*Hint*: You may also want to graph the function to verify you have found a maximum.]

(c) Now consider changing the sample size for a fixed value of π. Does the standard deviation decrease more by adding 500 subjects to a sample size of 500 or to sample size of 2500? Explain.

(d) Take the (first and) second derivative of SD($\hat{p}$) as a function of n to determine whether this function of n is concave up or concave down.

(e) Explain what your analysis in (b) and (c) reveal about the "diminishing returns" of increasing sample size. [*Hint*: you may want to create and examine at a graph of SD($\hat{p}$) vs. n for values of n from say 0 to 3000.]

79. Margin-of-Error Properties

The margin-of-error of a confidence interval for a population proportion π using the Wald procedure is $z*\sqrt{\pi(1-\pi)/n}$. Some books recommend a shortcut formula that approximates this margin-of-error for a 95% CI for π quite simply by $1/\sqrt{n}$.

(a) Explain why this is a reasonable approximation. [*Hint*: What simplification can you make in the given formula when you are using a 95% confidence level? What about different values of π ?]

(b) Show that this approximation is conservative, in that it slightly overestimates the actual margin-of-error.

(c) Reconsider Practice Problem 1.10 and re-answer part (b) – solving for the necessary sample size using this approximation.

(d) Suggest two different ways (that the researcher has direct control over) to reduce the margin-of-error in a study.

80. Margin-of-Error Properties (cont.)

Consider the margin-of-error of a confidence interval for a population proportion π using the Wald procedure: $z *\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$. Suppose that you want the margin-of-error to be no larger than some pre-specified error bound, M. That is, we want to satisfy the inequality $z *\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} \le M$.

(a) Rearrange the terms in the inequality to solve for the sample size n necessary to achieve such a margin-of-error, as a function of z^*, M, and $\hat{p}$.

(b) Is the expression in (a) an increasing function of z^*, a decreasing function of z^*, or neither? Consequently, is it an increasing function of the confidence level, a decreasing function of the confidence level, or neither? Explain why this makes intuitive sense.

(c) Is the expression in (a) an increasing function of M, a decreasing function of M, or neither? Explain why this makes intuitive sense.

(d) Is the expression in (a) an increasing function of $\hat{p}$, a decreasing function of $\hat{p}$, or neither? Use calculus to determine the value of $\hat{p}$ that maximizes this function.

(e) Explain why your answer to (d) suggests that the conservative approach to sample size determination is to use 0.5 as a guess for $\hat{p}$.

81. Margin-of-Error and Sample Size

Consider again the margin-of-error of a confidence interval for a population proportion π using the Wald procedure: $z *\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$.

(a) Determine whether this is an increasing or decreasing function of n, or neither.

(b) Take the second derivative of this function of n to determine whether this function of n is concave up or concave down.

(c) Sketch the behavior of this margin-of-error as a function of n for 95% confidence with $\hat{p} = 0.5$.

(d) Does the margin-of-error decrease more by adding 500 subjects to a sample size of 500 or to a sample size of 2500? Explain.

(e) Explain what your analysis reveals about the "diminishing returns" of increasing sample size.

82. Reducing the Margin-of-Error

Suggest two different ways for reducing the *margin-of-error* in a confidence interval for a population proportion π.

83. Sample Size Determination

Suppose you want to estimate the proportion of students at a very large university who are nearsighted. The prevalence of nearsightedness in the general U.S. population is 45%. Using this as a preliminary guess of π, how many students would need to be included in a random sample if you wanted the margin-of-error of a 95% confidence interval to be at most 1%?

84. The Wilson Adjustment

There are two probability rules concerning a constant times a random variable: $E(cX) = \underline{c}E(X)$ and $SD(cX) = c \cdot SD(X)$. Similar rules for the sum of a random variable X and a constant c are: $E(X + c) = E(X) + c$ and $SD(X + c) = SD(X)$.

(a) Explain why each of these last two rules makes intuitive sense.

(b) Let the random variable X have a binomial distribution with parameters n and π, let the random variable $\hat{p} = \dfrac{X}{n}$ represent the sample proportion, and let the random variable $\tilde{p}$ represent the Wilson adjustment: $\tilde{p} = \dfrac{X+2}{n+4}$. Rewrite $\tilde{p}$ in the form $cX + k$, where c and k are constants.

(c) Use your expression in (b) and both sets of rules to determine the expected value of $\tilde{p}$. Is this expected value equal to π?

(d) Use your expression in (b) and both sets of rules to determine the standard deviation of $\tilde{p}$.

(e) How does the standard deviation of $\tilde{p}$ compare to the standard deviation of $\hat{p}$?

85. Technical Conditions

The condition that you have studied for the reasonableness of the normal approximation to the binomial distribution is: $n\pi \geq 10$ and $n(1 - \pi) \geq 10$. An alternative that has been proposed is:
$$n\pi(1-\pi) \geq 5.$$

(a) Suppose that $\pi = 0.5$. According to the conventional condition, how large would the sample size n have to be in order for the normal approximation to be reasonable?

(b) Still in the $\pi = 0.5$ case, how large does this alternative condition say that the sample size n must be in order for the normal approximation to be reasonable? How does this compare to your answer from (a)?

(c) Repeat (a) and (b) for the $\pi = 0.25$ case.

(d) Repeat (a) and (b) for the $\pi = 0.9$ case.

(e) For the $n\pi(1 - \pi) \geq 5$ condition, express the required sample size n as a function of the success probability π. Graph this function, and comment on its behavior. [*Hint*: Feel free to use technology, and you might use values of π from 0 to 1 in multiples of 0.01.]

(f) Repeat (e) for the $n\pi \geq 10$ and $n(1 - \pi) \geq 10$ condition. [*Hint*: Remember that n must satisfy both conditions.]

(g) Write a few sentences comparing these conditions. Which seems to require a larger sample size? Does this vary depending on the value of π?

86. Sample Proportions

Let the random variables X and Y have binomial distributions with sample sizes n_1 and n_2, respectively, and let $\hat{p}_1$ and $\hat{p}_2$ represent the corresponding sample proportions of success. Recall that the pooled estimate of the underlying probability is $\hat{p} = \dfrac{X+Y}{n_1 + n_2}$.

(a) Show that $\hat{p}$ can be re-written as $\dfrac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$.

(b) Rewrite $\hat{p}$ as a weighted average of $\hat{p}_1$ and $\hat{p}_2$. In other words, express $\hat{p}$ as $\alpha\hat{p}_1 + (1-\alpha)\hat{p}_2$ for an appropriate value of α.

(c) Describe the two circumstances under which $\hat{p}$ would equal the (ordinary) average of $\hat{p}_1$ and $\hat{p}_2$.

87. Inference Subtleties

The following questions address some finer distinctions about the inference procedures you learned in this chapter.

(a) If you reject a hypothesis at the 0.05 level, what can you say about whether it would be rejected at the 0.07 level? How about at the0 .03 level? Explain.

(b) If a 95% confidence interval for a population parameter contains a particular value, what can you say about whether a 93% confidence interval would contain that value? How about a 97% confidence interval? Explain.

(c) If a 95% confidence interval for a population parameter contains a particular value, is it always true that you will fail to reject this hypothesized value in favor of a one-sided alternative? Explain.

88. Current Smokers

The 2010 National Health Interview Survey (NHIS) took a representative sample of 27,157 American adults using a multi-stage cluster design. One of their findings was that 19.3% of the individuals (age 18 and over) sampled identified themselves as current smokers (defined as ever smoking 100 cigarettes in their lifetime and smoking now every day or some day). The report listed the standard error of this statistic as 0.3%.

(a) If the sampling design had been a simple random sample, what would the standard error have been?

(b) Is the reported standard error from the multistage cluster design larger or smaller than the standard error from a simple random sample of the same size? Explain why makes sense.

(c) Identify in words the parameter of interest in this study.

(d) Use the reported standard error to produce a 99% confidence interval for this parameter.

(e) Would a test of whether the data provide evidence that less than 20% of American adults were smokers in 2010 produce a significant result at the $\alpha = 0.01$ level? Explain. [*Hint*: Remember to use the reported standard error.]

89. Current Smokers (cont.)

Reconsider the smoking data from the 2010 NHIS from the previous question. The report also mentioned that 21.2% of the male adults interviewed and 17.5% of the female adults interviewed identified themselves as current smokers.

(a) Do you expect the standard errors of these statistics to be less than, greater than, or equal to the 0.003 standard error of the previous question? Explain.

The report listed the standard errors to be 0.005 for the males and 0.004 for the females.

(b) Do you need to reconsider your answer to (a) in light of this information?

(c) Answer questions (c)–(e) of the previous exercise separately for males and for females.

90. High Anxiety

The Anxiety Disorders Association of America claims on its website (http://www.adaa.org/about-adaa/press-room/facts-statistics, August 2015) that 18% of adults suffer from some form of anxiety disorder.

(a) Describe how you might design a study to assess this claim.

(b) State the hypotheses that you would test, both in symbols and in words.

(c) If you want to estimate the population proportion who suffer from an anxiety disorder to within ± 0.05 with 95% confidence, how large a sample would be necessary? (Use the association's claim as a reasonable initial estimate of this proportion.)

(d) How would your answer to (c) change (if it all) if you wanted to estimate the proportion of women who suffer from an anxiety disorder? In other words, would the sample size need to be larger or smaller or the same for this population?

(e) How would your answer to (c) change (if it all) if you wanted to estimate the population proportion who suffer from an anxiety disorder in your state?

(f) Re-answer (c) using the more conservative estimate that the proportion is 0.5 rather than the
association's claim of 0.18. Is the required sample size larger now? Does the increase in sample size
appear to be substantial?

91. Leaving Office?

The news website MSNBC.com reports that one week during the Clinton-Lewinsky scandal, readers of
the site were invited to vote in an unscientific poll that asked whether President Clinton should leave
office. The site received over 200,000 votes, of which 73% said "yes."
(a) Describe the population and parameter of interest here.
(b) Use these sample data to determine a 99% confidence interval for the proportion of adult Americans
who felt during the week in question that Clinton should leave office. Also report the margin-of-error.
(c) Do you think that this interval does a reasonable job of estimating the population proportion who felt
that Clinton should leave office? Explain.
During the same week, an NBC News–Wall Street Journal poll contacted a random sample of 2005
people, with 34% answering that Clinton should leave office.
(d) Re-answer questions (b)–(c) based on these sample data.
(e) Which of these two intervals (the one based on the MSNBC.com poll or the one based on the NBC
News- Wall Street Journal poll) is narrower? Explain why that makes sense.
(f) Do these two intervals overlap? Are they similar at all?
(g) Which interval do you think provides a more reasonable estimate of the proportion of adult Americans
who felt that Clinton should leave office? Explain.

Additional Practice with Probability Distributions

92. Binomial or Not?

A binomial random variable counts the number of successes with a fixed probability of success and a
fixed number of independent trials. Explain why the following are *not* binomial random variables.
(a) Let X = the number of computer solitaire games you play before your first win.
(b) Let Y = the time you wait in line at a convenience store each time you visit.
(c) Let W = the number of long words in a sample of ten words from the Gettysburg Address.

93. Binomial Properties

Suppose we have a binomial random variable with n trials and probability of success π.
(a) In the One Proportion Inference applet, check the box on the far right to Hide Coins. Enter 25 as the
Number of tosses and check the Exact Binomial button. Also check the Summary Statistics box.
Include a screen capture of the binomial distribution.
(b) Check the Show Sliders box below the graph. Slowly move the first white square to the right (larger
values of π). Include a screen capture with π above 0.90. How would you describe the shape of this
distribution? How have the mean and standard deviation changed? Why? (*Hint*: Use the formulas
for E(X) and SD(X) to justify your answer.)
(c) For a fixed value of n, for what value of π is E(X) maximized? [*Hint*: Remember the restriction on
possible values of π.]
(d) Now move the slider close to zero and include a screen capture with π below 0.10. How would you
describe the shape of this distribution? How have the mean and standard deviation changed? Why?
(*Hint*: Use the formulas for E(X) and SD(X) to justify your answer.)

(e) Now keeping π below 0.10, move the Sample Size slider to the right (larger value of n). Include a screen capture with $n > 100$. How would you describe the shape of this distribution? How have the mean and standard deviation changed? (*Hint*: Use the formulas for E(X) and SD(X) to justify your answer.)

(f) For a fixed value of n (say 100), find the value of π that maximizes the standard deviation, SD(X). [*Hint*: Use trial and error in the applet or use calculus to take the derivative with respect to π of the standard deviation formula. Set this derivative equal to zero and solve for π.)

(g) For a fixed value of n, for what two values of π is SD(X) minimized? [*Hint*: You should not need to use calculus.]

94. Favored Candidate

Consider the population of roughly 200 million adult Americans. Suppose that 105 million favor candidate A in an upcoming election and 95 million favor candidate B. Suppose that a sample of 1000 is to be chosen at random. Let X be the number in the sample who favor candidate A.

(a) Identify the (exact) probability distribution of X (its name and its parameter values).

(b) Identify a probability distribution that could be used to approximate X (also its name and parameter values).

(c) Do you expect the approximation to be accurate here? Explain.

95. Freshmen Voting Patterns

A student project group (Norquest, Bayer, & McConville, 2004) wanted to determine whether there was a preference for the primary presidential candidates (George Bush and John Kerry) among first-year students at their university. They believed that these younger students would be more liberal and more inclined to vote for the democratic candidate (Kerry). They took a sample of 30 students from the 705 first-years at their school by assigning the first-year residence halls a number between 1 and 5 (inclusive) and rolling a die to select a residence hall. Then from the selected residence hall they took every seventh room to be part of the sample. The survey was distributed by going from room to room with the surveys, giving each resident of the room a copy of the survey to fill out, and immediately collecting the surveys after they were completed. If one of the residents was not in the room at the time of the first visit, the group repeatedly returned to the room until the person was contacted. The surveys were anonymous and the group members did not look at the completed surveys until they were randomly scrambled. The survey contained three questions, the first question was whether the student planned to vote for Kerry or Bush, and the other two questions concerned other social issues. For half of the surveys Kerry was listed first and for half of the surveys Bush was listed first.

(a) Explain why this sampling method is *not* a simple random sample but can still be considered a random sample and likely to be representative of the population.

(b) Identify some precautions taken by these students to avoid *nonsampling* errors.

In this study the sample results were that 22 of the 30 first-years planned to vote for John Kerry, whereas 8 planned to vote for George Bush.

(c) Suppose 352 first-year students at this school (essentially 50% of the population) plan to vote for Kerry. Use the hypergeometric distribution to calculate the probability of observing 22 or more students in a random sample of 30 students favoring Kerry. Would the sample result obtained by these students be surprising?

(d) The student project group suspected that about 2/3 of the first-year students at this school are planning to vote for Kerry. If 470 first-year students at this school plan to vote for Kerry (2/3 of this population), would the sample result obtained by these students be surprising? Explain.

(e) Would you consider the binomial approximation to the hypergeometric distribution appear valid for this study? Explain.

(f) Regardless of your answer to (e), use the binomial distribution to calculate the probability that at least 22 of 30 students would say they planned to vote for Kerry assuming the students' conjecture that 2/3 of the population of first-year students were planning to vote for Kerry.

(g) Does the normal approximation to the binomial appear to be valid for this study? Explain.

(h) Regardless of your answers to (g), carry out a one-sample *z*-test to test the students' conjecture. Compare the hypergeometric, binomial, and normal probabilities.

96. Binomial vs. Hypergeometric

Suppose that in a population of 10 items, 3 are defective and 7 are not. Suppose that two items are chosen at random for inspection. Let X be the number of defective items inspected. The random variable X has a hypergeometric distribution, with parameters $N = 10$, $M = 3$, and $n = 2$. (Note that because we are counting the number of defectives in the sample, "success" means "defective" here.)

(a) Explain why X does not have a binomial distribution. [*Hint*: Which of the binomial conditions is not satisfied here?]

(b) Determine the probabilities that $X = 0$, that $X = 1$, and that $X = 2$.

(c) Repeat (b) assuming that the population size is $N = 100$, with 30 defective and 70 not.

(d) Repeat (b) assuming that the population size is $N = 1000$, with 300 defective and 700 not.

(e) Repeat (b) assuming that the population size is $N = 10,000$, with 3000 defective and 7000 not.

If the sampling had been with replacement, then X would have followed a binomial distribution with $n = 2$ and $\pi = M/N = 0.3$.

(f) Determine the probabilities that this binomial random variable equals 0, 1, and 2.

(g) Does the binomial distribution do a good job of approximating the hypergeometric distribution for any of these values of the population size N? Explain why this makes sense.
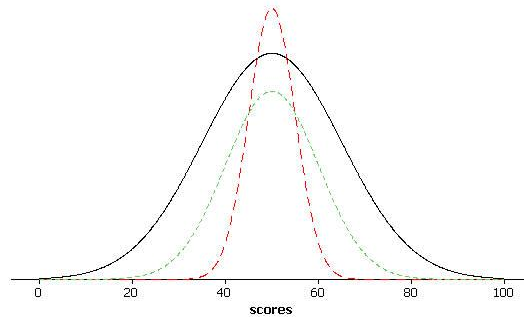
97. Normal Curves

Consider the following four normal curves modeling exam scores for four classes:
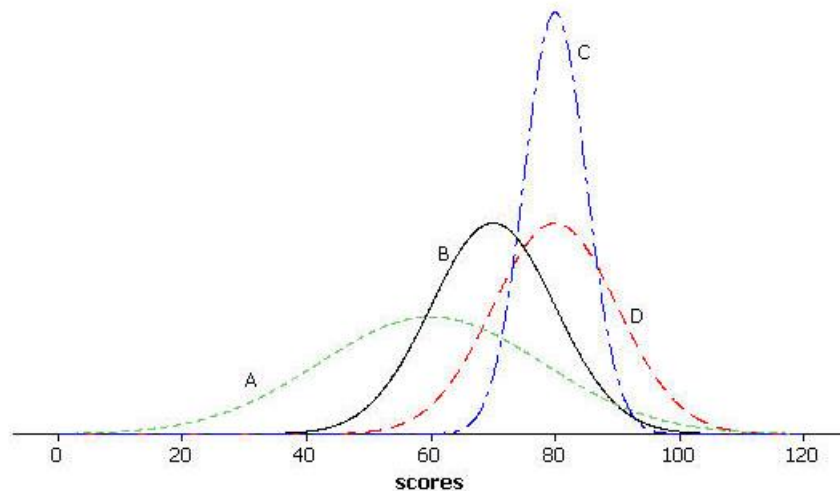
Approximate (as well as you can from the graphs) the mean and standard deviation of the exam score model for each class.

98. Bell-shaped Curves

Consider the following three bell-shaped curves:

Explain why these cannot all be normal curves.

99. Normal Probability Density Functions

Consider the standard normal probability density function, and let Z denote a standard normal random variable. [*Hint*: Remember that "standard" indicates that $\mu = 0$ and $\sigma = 1$, so the function

is $f(z) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$.]

(a) Evaluate this function at $z = 0$.
(b) Does your answer to (a) represent the probability that a standard normal random variable equals zero, $P(Z = 0)$? [*Hint*: You might want to draw a sketch, and remember what probabilities correspond to with a density curve.]
(c) Now consider a normal model for a random variable X with $\mu = 0$ and $\sigma = 5$. Write out the probability density function in this case, and then evaluate it at $x = 0$. Does this represent the probability that X = 0 with this model? Explain.
(d) Now consider a normal model with $\mu = 0$ and $\sigma = 0.1$. Evaluate the probability density function at $x = 0$. Does this represent the probability that X = 0 with this model? Explain what's different about this answer and the previous ones.
(e) Explain how it is possible that evaluating a legitimate pdf at a particular value of x can yield a value greater than one, yet probabilities calculated from legitimate pdfs are never greater than one.

100. Normal Probability Density Curves

Consider the standard normal probability density function.
(a) Use calculus to determine where the maximum value occurs with the standard normal probability density function. [*Hint*: Take the derivative with respect to x, set the derivative equal to zero, and solve for x. Then check the second derivative to make sure that you have found the value of x where the *maximum* occurs.]
(b) Use calculus to determine where the inflection points occur with the standard normal probability density function.
(c) Repeat (a) and (b) for the normal pdf with parameters μ and σ. Show that the maximum occurs at $x = \mu$ and the inflection points occur at $x = \mu - \sigma$ and $x = \mu + \sigma$.

The following two exercises use Minitab. There is not a simple power function in R for one proportion. Minitab uses the normal approximation.

101. *Swain v. Alabama* (cont.)

Reconsider the *Swain v. Alabama* case (Exercise 53) where census data suggested that about 26% of those eligible for grand jury service were black. Suppose, due to some discriminating efforts, the rate at which blacks were selected for grand jury duty service was 0.25.

(a) Calculate the power of the test (with a sample size of $n = 1050$) against this alternative value ($\pi_a = 0.25$), using the $\alpha = 0.05$ level of significance. Include an interpretation of what the power calculation signifies.

(b) To verify your calculations using Minitab: Choose **Stat > Power and Sample Size > 1 Proportion**. Specify the sample size, the Comparison proportion (π_a), and the hypothesized proportion (π_0). Leave the Power values box empty. Under the "Options" button, specify the form of the alternative hypothesis, and keep 0.05 as the level of significance. Click the OK button twice. How does your calculation compare to Minitab's?

(c) Repeat (b) using 0.20 as the alternative value of the parameter. How does the power change for this new alternative value of the true underlying rate at which blacks are selected for grand jury duty? Provide an intuitive explanation of why this makes sense.

102. *Swain v. Alabama* (cont.)

Return to the context the previous exercise, where you used Minitab to perform the power calculation for two different alternative values of π, the underlying probability of a black person being selected for jury duty in Alabama in 1965.

(a) Choose **Stat > Power and Sample Size >1 Proportion** as before. In the Comparison proportion box, enter: .005:.995/.005. Minitab will create a graph of power as a function of the Comparison probabilities. Describe the behavior of this graph and explain why it makes intuitive sense. (You might want to "zoom in" on the more interesting part of the graph by double-clicking on the horizontal axis and then changing the minimum and maximum values for the scale.)

(b) Repeat (a) with a sample size of 550. Describe the behavior of the "power curve," explain why this behavior makes sense.

(c) Repeat (a) with a sample size of 150. Describe the behavior of the "power curve," and explain why this behavior makes sense.

NEW EXERCISES

103. Clinton e-mails

While Secretary of State, Hilary Clinton used a private e-mail server rather than a "secure" government server. This led to controversy of whether she had allowed sensitive government information to be insecure. In 2015, she was asked to give access to her email account to the State Department. She submitted 30,000 emails. A random sample of 40 e-mails was selected and two were deemed to contain "Top Secret/Sensitive Compartmented Information," one of the government's highest levels of classification.

(a) Describe how such a random sample could be selected.

(b) Calculate and interpret a 95% confidence interval for the probability of a Clinton email containing top secret information.

104. Binomial nightmares

Suppose you are having a nightmare involving a 100 multiple-choice exam for which you have not studied… in a foreign language... in a Syriac alphabet. Each question has four possible answers but only one correct response. Let Y represent the number of questions you answer correctly.

(a) Would it be reasonable to model Y with a binomial distribution? Comment on whether you think of the necessary conditions is met, including any assumptions you have to make. Also clarify the parameter values (n and π) that you would use.
(b) Calculate the mean and standard deviation of the binomial distribution.
(c) Assuming the binomial model is valid, calculate the exact probability of answering more than half the questions correctly. (Include your output.)
(d) Assuming the binomial model is valid, calculate the probability of answering at least 35 questions correctly.
(e) Use the normal distribution to approximate the probability of answering more than half the questions correctly.
(f) Is the normal approximation valid in this case? Justify your answer.
(g) Explain and illustrate how you could use a continuity correction to improve your estimate of the p-value in (d).

105. Common Core Opinions

The Common Core is a set of guidelines that describe what children should learn and be able to do in math and reading from kindergarten through 12th grade. These national academic standards were adopted by 40 states, but there has been widespread confusion about what they entail, especially for parents. Fairleigh Dickinson University's PublicMind conducted a telephone survey from Dec. 8 through Dec. 15, 2014 using a randomly selected sample of 964 adults nationwide. PublicMind interviews are conducted by Opinion America of Cedar Knolls, NJ, with professionally trained interviewers using a CATI (Computer Assisted Telephone Interviewing) system. When questions are asked the order of the option choices are determined randomly. Random selection is achieved by computerized random-digit dialing. Landline households are supplemented with a separate, randomly selected sample of cell-phone respondents interviewed in the same time frame. One question was whether the respondent believed the common core includes teaching about evolution. Forty-three percent said yes and 17% said no (the rest didn't know or refused to answer).
(a) Identify the population of interest. Is this a large population?
(b) Focusing on the percentage that said yes, describe the parameter of interest in context.
(c) Calculate and interpret a 95% one-sample z-confidence interval for the parameter in (b). Include your work.
(d) Report the margin-of-error from your interval in (c) and provide a one-sentence interpretation of what this margin-of-error measures as if to a college freshman.
(e) Of the participants with at least some college-level education, 46% said yes. Do you expect the margin-of-error for this result to be larger or smaller or roughly the same as in (c)? Explain your reasoning.
(f) Identify a method used by this survey in an attempt to prevent "sampling error."
(g) Identify a potential "non-sampling error" and how this survey methodology tried to prevent it in this study.
(h) Identify another potential "non-sampling error" that could still exist.

106. Paul the Octopus

People use animals to predict outcomes of sporting events and elections mostly by having the animals pick between two containers of food, each representing one of the two teams or the two candidates involved. Probably the most famous of these is Paul the Octopus from Germany who was correct in all eight of his predictions for the 2010 World Cup. Do his results provide strong evidence that Paul is doing something other than randomly guessing and perhaps has some psychic power?

(a) Can we use a "coin tossing" simulation to represent the scenario where Paul is simply guessing? Explain (and discuss any limitations you see in using this model to represent Paul's behavior).

(b) If we let π represent Paul's probability of picking the winning team, what is the value of π if Paul is randomly guessing?

(c) If Paul has some psychic power, what does this imply about the value of π ?

(d) Use the One Proportion Inference applet to find a p-value. Include a copy of your output showing both the input values and the results.

(e) *Interpret* this p-value in the context of the study (what is it the probability of?)

(f) Based on your p-value, write out a conclusion in the context of the research question.

(g) As was stated earlier, many animals are used to predict outcomes of sporting events with most of them not doing any better than random chance and therefore not getting much publicity. Paul made the news because he was so accurate. Do you think it is surprising that some animal would be very successful like Paul was? Explain.

(h) Do these data provide convincing evidence that Paul's probability of picking the winning team is larger than 0.70? Justify your answer.

107. Hazelwood vs US

One of the first times the U.S. Supreme Court considered statistical significance in an employment discrimination case was in *Hazelwood School District vs. United States* (1977). The U.S. government sued the City of Hazelwood, a suburb of St. Louis, MO, on the grounds that it discriminated against African Americans in its hiring of school teachers. The evidence introduced noted that of the 405 teachers hired in 1972 and 1973 (the years following the passage of the Civil Rights Act), only 15 had been African-American. By comparison, according to 1970 census figures, of the almost 20,000 elementary and secondary teachers employed in the St. Louis area, 15.4% were African American. We want to decide whether the data on these 405 teachers is convincing evidence that the *Hazelwood hiring process* had a probability of a new hire being African American that was less than 0.154.

(a) Define the observational units of interest.

(b) Define the variable of interest.

(c) Define the parameter of interest (in context).

(d) State appropriate null and alternative hypotheses.

(e) Calculate an exact p-value for your hypotheses in (d). Be sure to include appropriate output.

(f) What *conclusion* can you draw from the p-value in (e)? Be sure to discuss whether you can conclude that being African American caused a teacher to have a lower probability of being hired in Hazelwood.

108. Hazelwood vs US cont.

Continuing with the context from the previous problem: The St. Louis City School District had recently followed a policy attempting to maintain a 50% African-American teaching staff. If you exclude the St. Louis City School District, then proportion of eligible teachers in the St. Louis area that were African-American was 0.057. This means the city school district seemed to be pulling up the overall proportion of African Americans hired, so let's compare the other school districts to Hazelwood, without this "outlier".

(a) Calculate a new p-value for Hazelwood. Include your output.

(b) Is the p-value larger or smaller than in problem 1? Explain why.

The *Hazelwood* case set forth the "standard deviations analysis":

> Because a fluctuation of more than two or three standard deviations would undercut the hypothesis that decisions were made randomly with respect to race.."

(c) Using the binomial distribution, how many standard deviations is 15 from the expected number of new hires that are black if $\pi = 0.154$? (Show your calculation details.)

(d) Repeat (c) assuming $\pi = 0.057$.

(e) Discuss briefly how this case illustrates "the importance of the choice of the relevant labor market area" in the cases of discrimination.

109. Cuing Dogs

Can dogs understand visual cues? Researchers Udell, Giglio, and Wynne ("Domestic dogs (*Canis familiaris*) use human gestures but not nonhuman tokens to find hidden food, *Journal of Comparative Psychology,* 122(1), 2008) positioned dogs about 2.5 meters from an experimenter. Two cups were placed, one on each side of the experimenter.
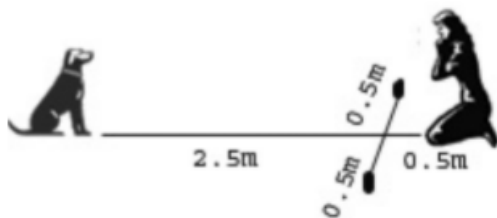
Familiar Communicative

Dog	Group	Point	Bow look dog	Head turn	Glance with eyes	Average
Harley	ABC	10**	9**	5†	3	6.75

The experimenter would perform some sort of gesture (pointing, bowing, looking) toward one of the cups or there would be some other nonhuman gesture (e.g., a mechanical arm pointing, a doll pointing, or a stuffed animal looking) toward one of the cups. The researchers would then see whether the dog would go to the cup that was indicated. One dog tested, a four-year-old mixed breed need Harley, was given 10 trials with the human experimenter. In nine of the 10 "bow look dog" trials, Harley chose the correct cup.

(a) Define the parameter of interest and state the appropriate null and alternative hypotheses (in symbols).

(b) Find the exact binomial p-value, a simulation-based p-value, and a theory-based (one proportion z-test) p-value. (Remember to include all relevant output.)

(c) Which p-values are most similar to each other? Which p-value would you report?

(d) A "continuity correction" would use the normal distribution to find p-value $= P(X \geq 8.5)$ rather than $P(X \geq 9)$, to account for the discrete probability at $X = 9$. What is the continuity-corrected p-value for this study? Is it larger or smaller than the one proportion *z*-test p-value? Is it more "valid"? (How are you deciding?) (Remember to include all relevant output.)

110. Why Men Cheat

The August 19, 2013 issue of *Sports Illustrated* included a brief article titled "Why Men Cheat," in which the author L. Jon Wertheim argued that Major League Baseball (MLB) players from the United States are less likely than those from non-US (often poorer) countries to take performance enhancing drugs (PEDs). Around 57.3% of all professional baseball players between 2005 and 2013 were from the United States. Of the 595 players suspected for PED use in these years, 206 were from the United States.

(a) Let π represent the probability that a player suspended for PED use is from the United States and use the sample proportion $\hat{p} = 206/595$ to carry out a *one-sample z-test* to test the following hypotheses

H_0: $\pi = 0.573$

H_a: $\pi < 0.573$

(b) Include a one-sentence interpretation of the *z*-test statistic from your output in this context.

(c) Include a one-sentence interpretation of the p-value from your output in this context.

(d) Are the validity conditions for the normal approximation in (a) valid for these data? Justify your answer.

(e) Produce <u>and</u> interpret a *one-sample z-confidence interval,* with 95% confidence.

111. Look Alikes

Researchers have tested the theory that "dogs look like their owners," typically asking participants to match photos of dogs with their owners. Japanese psychologist Sadahiko Nakajima attempted to explore which features had the strongest connection. In his study, participants were shown two sets of 20 photos of dog-owner pairs, one set showed actual dog-owner pairings, the other set showed random pairings. To eliminate extraneous factors, the photos were very basic color headshots cropped at the shoulders and shown against a plain white background (see below). In one study (Anthrozoos, 26(4), 2013), participants where shown only pictures of the owner eyes and the dog eyes (bottom right photo). The participants' task was to: "Choose the set of dog-owner pairs that physically resemble each other, Set A or Set B." In this study the participants were 54 Japanese undergraduates.
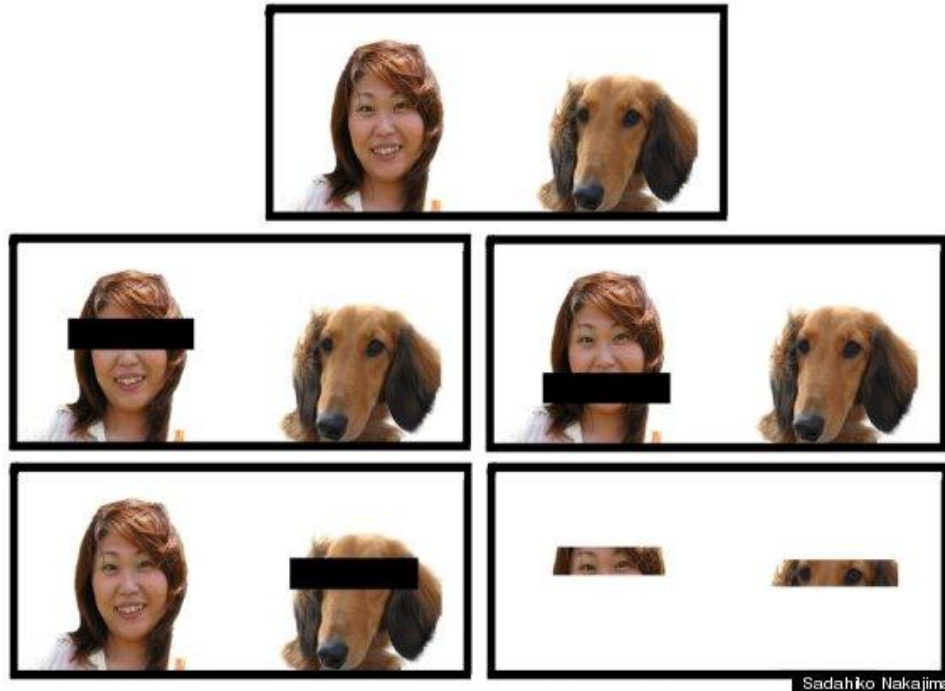
(a) Identify the observational units and variable in this study.
(b) State the appropriate null and alternative hypotheses for considering whether there is convincing evidence that people are more likely to match a dog and owner photo than we would expect by random chance alone. You should do so using symbols, defining the symbols you use.
(c) In this study 40 of the 54 Japanese undergraduates identified the correct set of photos. Use technology to carry out a "binomial test" for your hypotheses in (b). Be sure to include appropriate evidence of input used, and your technology output including a p-value.
(d) Provide a detailed interpretation of the p-value from (c) in context. [*Hint*: Apply the lessons you learned about how to interpret probability, e.g., what is being repeated, what is happening, what are you assuming?]
(e) What conclusion can you draw from the p-value?
(f) Using the binomial model (i.e. theoretical results), how many standard deviations is the observed result from the expected value under the null hypothesis?
(g) Nakajima was so surprised by these results that he ran this study again using 55 new students. Forty-two chose the correct list. Combine the results from these two studies together and find the p-value for the hypotheses in (b). Include your technology output. Report the new p-value. Is this new p-value larger or smaller than in (c)? Give an intuitive explanation for why the p-value is larger or smaller for the new results.

112. Marine Cancer Detection

Recall the 2011 article published by the medical journal *Gut—An International Journal of Gastroenterology and Hepatology* (Sonoda et al.) which reported the results of a study conducted in Japan in which a dog was tested to see whether she could detect colorectal cancer. The dog used was an eight-year-old black Labrador named Marine. (As her name might suggest, she was originally trained for water rescues.) The study was designed so that the dog first smelled a bag that had been breathed into by a patient with colorectal cancer. This was the standard that the dog would use to judge the other bags. Marine then smelled the breath in five different bags from five different patients, only one of which contained breath from a different colorectal cancer patient; the others contained breath from noncancer patients. The dog was then trained to sit next to the bag which she thought contained breath from a cancer patient (i.e., had the cancer scent). If she sat down next to the correct bag, she was rewarded with a tennis ball. Marine completed 33 attempts of this experimental procedure, with a different set of five patients each time: four noncancer patients and one cancer patient. And, each time, whether she correctly identified the bag with the breath of the cancer patient was recorded. In this study, Marine was correct in 30 of the 33 attempts.

(a) Why is it important to use new patients each time? [*Hint*: How does this help us meet the binomial model?]

(b) Identify the observational units and variable of interest. Indicate which outcome you will consider "success."

(c) Define the parameter of interest (in context).

(d) State appropriate null and alternative hypotheses for deciding whether Marine's probability of correct identification *differs* from 3/4.

(e) Use technology to carry out a binomial test for the hypotheses in (d).

(f) Using a two-sided hypothesis test and trial and error, find the smallest value of π for which you would first fail to reject the null hypothesis. Feel free to use the One Proportion Inference applet.

(g) Using trial and error, find the largest value of π for which you would fail to reject the null hypothesis. Again feel free to use the One Proportion Inference applet.

113. Monday Heart Attacks

Patients with kidney disease often have a procedure called dialysis done to clean their blood if their kidneys can't do this properly. This procedure is often done three days per week, with Monday, Wednesday, and Friday often being those days. In terms of the dialysis treatments, these days are all the same except for the gap of two off days before the Monday treatment. Does this gap make a difference? Cardiac arrest and sudden death for patients undergoing dialysis for kidney disease are possibilities. In recent years, it has been noted that these types of patients have cardiac arrests on Mondays more often than what would be expected. A study published in *Kidney International* (Karnik et al., 2001) looked at 205 dialysis patients that had cardiac arrests on Monday, Wednesday, or Friday, the same days they had

dialysis. They found that 93 of these happened on a Monday. Do we have convincing evidence of a larger **or smaller** probability of cardiac arrests occurring on Mondays compared to the other two days?
(a) Define the parameter of interest in words.
(b) State the null and alternative hypotheses for these research question.
(c) Calculate the sample proportion (to four decimal places).
(d) Does the normal approximation to the binomial distribution appear to be valid for these data? Explain your reasoning.
(e) Carry out a one-sample *z*-test using an applet or R or JMP, including your output. Report the test statistic (*z*-score) and p-value. *Provide a one-sentence <u>interpretation</u> of each in context.*
(f) Use technology to compute a 95% confidence interval (one proportion z-interval) for the parameter. Include your output and provide a one-sentence interpretation of the interval in context.
(g) Does 1/3 appear to be a reasonable value of π? Explain how your results in (e) and (f) lead to the same conclusion.

114. Home Field Advantage
For the first 88 games in the 2019 English Premier League soccer games, the home team won 54 of the matches (See Practice Problem A.B).
(a) Use technology to create a bar graph of these results. Summarize the results in context.
(b) Suppose we treat these 88 games as independent observations of a random process. Can we use a "coin tossing" simulation to represent the situation where there I no home field advantage? Explain (and discuss any limitations you see in using this model to represent the underlying process). Also complete a "simulation model" table.
(c) If we let π represent the (long-run) probability of the home team winning, what are the null and alternative hypotheses for considering whether these data provide convincing evidence of a home-field advantage?
(d) Use the One Proportion Inference applet to find a p-value using simulation. Include a copy of your computer results showing both the input values and the results.
(e) *Interpret* this p-value in the context of the study (what is it the probability of?)
(f) Based on your p-value, write out a conclusion in the context of the research question.

115. Castaneda v. Partida
One of the first legal cases that relied heavily on statistical analysis was *Castaneda v. Partida*. Many employment discrimination cases have since relied heavily on this landmark case. One of the main considerations is what comparison to make. The Castaneda case involved a claim of underrepresentation of Mexican-Americans in grand jury selection. Over an 11-year period, 870 people had been summoned to serve on grand jury in Hidalgo County, but only 338 had been Mexican-American. The Court held that because about 79.1% of the population of Hidalgo County were Mexican-American, then about 79.1% of those summoned to serve as grand jurors should be Mexican-American.
(a) Carry out a "binomial test" for the judge (Define the parameter of interest in context, state the null and alternative hypotheses, report the exact binomial p-value (including appropriate output), report your test decision (reject or fail to reject the null hypothesis) and summarize/justify your conclusions.)
(b) Based on the binomial distribution, what is the expected number of Mexican-Americans in the jury pool?
(c) Based on the binomial distribution, what is the theoretical standard deviation of the number of Mexican-Americans?
(d) How many standard deviations is the observed number of Mexican-Americans from the expected number?

(e) Rather than calculating the binomial probability, many court cases have adopted the general rule: "[I]f the difference between the expected value and the observed number is greater than two or three standard deviations, then the hypothesis that the ... drawing was random would be suspect to a social scientist." Do you reach the same conclusion?

(f) However, both of these methods rely on the assumptions of the binomial model. Do you consider the assumptions met in this case? What about in employment discrimination cases? Explain your reasoning. (Consider reviewing this reference.)