ISCAM 2: CHAPTER 2 EXERCISES

1. Gender and Blood Donations

The 2004 General Social Survey (GSS), a large-scale national survey of representative adult Americans conducted every two years, asked respondents whether or not they had donated blood at least once in the past year. Of the 1271 male respondents, 239 had given blood in the previous 12 months. Of the 1427 female respondents, 201 had giving blood. Of interest is whether there is a statistically significant relationship between gender and this "formal helping behavior." [Notes: In Investigation 2.1, the researchers literally took two different random samples from two different groups of teens. In this study, the GSS took one random sample of 2698 adult Americans and classified each respondent according to two variables: gender and whether or not donated blood. We can use the same analysis tools (that investigated the difference in two populations) to investigate the relationship between these two binary variables.]

(a) Identify and classify the two variables in this study.
(b) Create a two-way table of counts to summarize these data using the gender variable as the column variable.
(c) Calculate conditional proportions to measure the difference in the likelihood of blood donations between the men and women in this sample. Also sketch out (by hand) a segmented bar graph.
(d) Carry out a test of significance to determine whether the difference in the sample proportion giving blood is statistically significant. [Define the parameter of interest, specify the null and alternative hypotheses, verify the validity of the normal model, use technology to calculate the test statistic and p-value, and state your conclusions in context.] Also produce a 95% confidence interval for the parameter. Be sure to specify the population(s) you are willing to generalize your results to.
(e) Suppose you had defined the parameter by subtracting in the other direction (e.g., female – male instead of male – female). How would that change
- The observed statistic?
- The test statistic?
- The p-value?
- The confidence interval?

(f) Use technology to test your conjectures in (e) (with the original samples sizes).
(g) Suppose the sample sizes for men and women had been 600 and 700 but the sample proportions had been the same. Would the p-value you found be larger or smaller or the same?
(h) Use technology to test your conjecture in (g).

2. The Governator

In 2003, the state of California conducted their first recall election. Over one hundred candidates vied to replace Gray Davis as Governor, with actor Arnold Schwarzenegger winning 49% of the votes cast. In the days before the election, newspapers had published accusations of inappropriate treatment of women by Schwarzenegger, so some suspected that his support from women would be weaker than his support from men. Before the polls close, pundits try to predict the election results by using "exit polls," asking people as they leave the voting booth who they voted for. In one such poll, CNN interviewed 2023 men and 2191 women. Forty-nine percent of the men and 43% of the women said they voted for Arnold.

(a) Specify the populations.
(b) Specify the parameter(s) of interest.
(c) Produce and comment on a segmented bar graph to compare the proportion voting for Arnold for men and women. Is there preliminary evidence in these samples that men were more likely than women to say they voted for Arnold?
(d) Set up null and alternative hypotheses to assess where there is convincing evidence that mean were more likely to say they voted for Arnold than women.

(e) Carry out a simulation analysis, as in Investigation 2.1, to approximate the p-value for this test. Be sure it's clear how you conducted the simulation, the "what if" distribution you found (including mean and standard deviation), and how you found the p-value.
(f) Are the technical conditions for the normal approximation to be valid met for this study?
(g) Use the normal approximation to calculate the test statistic and approximate the p-value. Remember to include your output! Compare your answer to the simulated p-value.
(h) Provide a detailed interpretation of this p-value *in the context of this study*. Be very specific about the source of "random chance" and what you mean by "results like this."
(i) What conclusion will you draw based on this p-value?
(j) Calculate and interpret a 90% confidence interval for the difference in the population proportion of men and women who would claim to have voted for Arnold. Be sure to specify the "direction" of the difference.

Extra Credit: The simulation you conducted in part (e) is not an exact match for how these data were collected. Explain why not <u>and</u> describe how you would modify the simulation to more exactly match the data collection process for this study. [*Vague Hint*: Different sample sizes.]

3. Native Californians?

As a transplant to California, author A wondered whether California residents were more or less likely to have been born in California (i.e., native Californians) back in 1950 or more recently, say in 2000. To investigate this question, he took a random sample of 500 CA residents from the 1950 Census and an independent random sample of 500 CA residents from the 2000 Census. The results are shown in the table below:

	1950	**2000**
Born in CA	219	258
Not born in CA	281	242
Total	500	500

(a) For each year, calculate the proportion of California residents who were born in California. Use appropriate symbols to represent them. Also calculate the difference between these proportions.
(b) Produce a segmented bar graph to display the conditional proportions who were born in California in these two years. Comment on what the graph, along with your calculations from (a), reveal.
(c) State the appropriate null and alternative hypotheses, in words and in symbols, to address the research question of whether California residents were more or less likely to have been born in California back in 1950 or in 2000.
(d) Use technology to conduct a simulation analysis to approximate a p-value for this significance test. Be sure to report the appropriate parameter values (n and π) for the binomial distribution that you simulate from. Also submit a well-labeled histogram of your simulation results. Finally explain how you calculate the approximate p-value and report its value.
(e) Check the conditions for whether the normal approximation is appropriate for this significance test.
(f) Calculate the z-test statistic and p-value based on the normal distribution.
(g) Summarize your conclusion from these analyses, with regard to the research question. Also explain the reasoning process that leads to your conclusion.

4. Home Court Disadvantage?

Many sports enthusiasts believe there is an advantage to playing on the home court or home field. They argue that advantages such as the familiarity of the environment and the loud support of their fans can help the home team win the game. But is fan support always a benefit? And does the size of the crowd make a difference? The 2008-9 Oklahoma City Thunder, a National Basketball Association team in its second year after moving from Seattle, found that their win-loss record was actually worse for home games with a sell-out crowd (3 wins and 15 losses) than without have a sell-out crowd (12 wins and 11 losses). (These data were noted in the April 20, 2009 issue of *Sports Illustrated* in the Go Figure column.)

(a) Identify the observational units and variables. Which variable would you consider the explanatory variable?
(b) Calculate the difference in the proportion of wins in each group.
(c) State the null and alternative hypotheses to investigate whether there was a sell-out effect (advantage or disadvantage) on the underlying probability of winning a home game for the Oklahoma City Thunder that year.
(d) Explain whether you would use Fisher's Exact Test or a two-sample *z*-test to obtain a p-value to compare the two sample proportions.
(e) Use the Analyzing Two-way Tables applet to find the exact p-value from Fisher's Exact Test. (Note: Enter the data in the table, run the simulation with 1000 repetitions, then check the options for a two-sided p-value and for displaying the exact p-value, which uses the "smaller p-value" approach.) Include your output and provide a one-sentence interpretation of this p-value.
(f) Based on this p-value, would you reject the null hypothesis at the 5% level of significance?
(g) Calculate and interpret a 95% confidence interval for the relative risk of losing a game with a sold out crowd versus a non-sell out for this long-run "process."
(h) Based on this analysis, would you be willing to conclude that the sell-out home crowds *caused* the Thunder to perform more poorly (be less likely to win)?
(i) Suggest another variable that could explain why the winning percentage of the Thunder was significantly worse in home games with a sell-out crowd than in home games with a smaller crowd.
(j) In fact, 22 of the 41 home games for the Thunder were against teams that won more than half of their games that season. Of these 22 games, 13 were sell-outs. Of the 19 games against opponents that won less than half of their games that season, only 5 of those games were sell-outs. Does this support the statement that the Thunder tended to have more sell-outs against "good" teams as opposed to "weaker" teams? Support your answer by calculating relevant proportions.
(k) When the Thunder played a "good" team, they lost 18 of 22 games. When they played a "weaker" team, they won 11 of 19 games. Does this support the statement that the Thunder tended to lose more often when they played better teams? Again support your answer by calculating relevant proportions. [*Hint*: Be sure to calculate these proportions in a consistent manner.]

5. Email or Paper & Pencil?

Are people more likely to lie with e-mail than with pencil and paper? A study reported at a meeting of the Academy of Management involved 48 graduate students studying business at a particular university who participated in a bargaining game (Naquin, Kurtzberg, & Belkin, 2008). The response variable of interest was whether the person misrepresented (lied about) the size of the pot when negotiating with another player. Some of the participants were randomly assigned to use e-mail for their communication, whereas others used paper and pencil. It turned out that 24 of 26 who used e-mail were guilty of lying about the pot size, compared to 14 of 22 who used paper and pencil.

(a) Identify the explanatory variable and the response variable in this study. Be sure to state the variables clearly as *variables* (as opposed to numerical summaries or conclusions or research questions).
(b) For each explanatory variable group, determine the conditional proportion who lied. Do these proportions differ in the direction conjectured by the researchers?
(c) Use the **Analyzing Two-way Tables** applet to conduct a simulation analysis with 1000 repetitions. Report the approximate p-value from this simulation. Include a copy of your output, including one showing the two-way table that you entered.
(d) Use the hypergeometric distribution (Fisher's Exact Test) to calculate the exact p-value. Show the details of your calculation. (You can use the applet to check your answer, but you should demonstrate that you can get determine the p-value without the applet using the hypergeometric distribution as

well. You may use technology.)

(e) Provide a complete, detailed interpretation (in one or two sentences) of what this (approximate) p-value measures in this context (i.e., what is it the probability of?). Be sure to define what you mean by "by chance" and "this extreme."

(f) Check the technical conditions for the two-sample *z*-test, and comment on whether they are satisfied.

(g) Regardless of your answer to (f) calculate the corresponding p-value using the two-sample *z*-test. How do the p-values compare?

(h) Suppose the study had involved 52 e-mail users and 44 paper and pencil users, but that the conditional proportions who lied were the same. Determine the exact p-value in this case (show your work), and comment on whether/how it changes from the real data. Explain why this makes sense. (Your explanation can appeal to information from the two-sample *z*-test, but you should perform the hypergeometric calculations.)

6. Praising Intelligence or Effort

Psychologists investigated whether praising a child's intelligence, rather than praising his/her effort, tends to negative consequences such as undermining their motivation (Mueller & Dweck, 1998). Children participating in the study were given a set of problems to solve. After the first set of problems, half of the children were randomly assigned to be praised for their intelligence, whereas the other half was praised for their effort. The children were then given another set of problems to solve and later told how many they got right. They were then asked to write a report about the problems for other children to read, including information about how many they got right. Some of the children misrepresented (i.e., lied about) how many they got right, as shown in the following table:

	Praised for intelligence	Praised for effort	Total
Misrepresented their score (lied)	11	4	15
Did not misrepresent (did not lie)	18	26	44
Total	29	30	59

(a) Identify the explanatory and response variables in this study.

(b) For each group, determine the proportion who lied, and identify them with appropriate symbols.

(c) Describe how you could use index cards to conduct a simulation analysis for determining whether the difference between these proportions is statistically significant. Include the following information in your description:
   - i) how many cards you would use
   - ii) how many would be marked how
   - iii) how many you would deal out
   - iv) which kinds of cards you would count
   - v) what you would compare the results to, after you conducted a large number of repetitions

(d) Use the Analyzing Two-way Tables applet to conduct a simulation with 1000 repetitions. Submit a screen capture of the resulting histogram, and report the empirical p-value from the applet.

(e) Use R or Minitab to perform Fisher's Exact Test. Along with reporting the p-value, provide an appropriate graph, and also express the p-value as P(X ____ *k*), where you:
   - i) insert the appropriate inequality in the _____ space,
   - ii) report the appropriate value of *k*,
   - iii) indicate what kind of probability distribution X has (name),
   - iv) provide the numerical values (inputs) associated with that probability distribution.

(f) Provide a complete, detailed interpretation (in one or two sentences) of what this p-value means in this context (i.e., what is it the probability of, assuming what?)

(g) Based on this p-value, is the observed difference between the groups statistically significant at the α = 0.05 level? Explain how you know.

(h) Summarize and justify your conclusion about whether the data provide evidence that praising a child's intelligence leads to more negative consequences than praising his/her effort.

7. Effectiveness of AZT

In 1993, one of the first studies aimed at preventing maternal transmission of AIDS to infants gave the drug AZT to pregnant, HIV-infected women (Connor et al., 1994). Roughly half of the women were randomly assigned to receive the drug AZT, and the others received a placebo (a "fake" treatment, same appearance as the drug but with no active ingredients). The HIV-infection status was then determined for 363 babies, 180 from the AZT group and 183 from the placebo group. Of the 180 babies whose mothers had received AZT, 13 were HIV-infected, compared to 40 of the 183 babies in the placebo group.

(a) Identify the observational units in this study.
(b) Explain why you think use of a placebo would be important in a study like this.
(c) If you were going to carry out a simulation to analyze these data, how many player cards would you have? How many would be marked "red" and how many "blue"? How many would you deal out? What "statistic" would you record for each shuffle? How would you determine the p-value?
(d) Suppose instead of measuring "whether or not the baby is HIV-infected," the researchers gave a numerical score to the health of each baby. You want to know whether the difference in the mean scores given to these two groups is statistically significant. Explain briefly how you would modify your simulation.

8. Conserving Hotel Towels?

Many hotels have begun a conservation program that encourages guests to re-use towels rather than have them washed on a daily basis. A recent study examined whether one method of encouragement might work better than another. Different signs explaining the conservation program were placed in the bathrooms of the hotel rooms, with random assignment determining which rooms received which sign. One sign mentioned the importance of environmental protection, whereas another sign claimed that 75% of the hotel's guests choose to participate in the program. The researchers suspected that the latter sign, by appealing to a social norm, would produce a higher proportion of hotel guests who agree to re-use their towels. Researchers used the hotel staff (a mid-sized, mid-priced hotel in the Southwest that was part of a well-known national hotel chain) to record whether guests staying for multiple nights agreed to reuse their towel after the first night.

(a) Identify the observational units, explanatory variable, and response variable in this study.
(b) State the null and alternative hypotheses in symbols, and be sure to define the parameter in the context of this study.

The following table displays the observed data in this study:

	Social norm	Environmental protection	Total
Guest opted to re-use towel	98	74	172
Guest did not opt to re-use towel	124	137	261
Total	222	211	433

(c) Calculate the conditional proportions of re-use in each group. Also calculate the difference between them and the ratio of these proportions.
(d) Interpret what this ratio reveals in this context.
(e) Use a two-sample z-test to test the hypotheses that you stated in (a). Report the test statistic and p-value.
(f) Report your test decision at the $\alpha = 0.10$, 0.05, and 0.01 significance levels. Also summarize what these test decisions reveal about the strength of evidence for the researchers' conjecture.
(g) Produce and interpret a 90% confidence interval for the difference in probabilities of re-using towels between these two signs.
(h) Produce and interpret a 90% confidence interval for the ratio of probabilities of re-using towels (relative risk) between these two signs.

9. The Winner?

Enamored of the solitaire game on his new computer, author A sets out to estimate his probability of winning the game and wins 25 games while losing 192 games. Anxious to outperform author A, author B plays 444 games of solitaire and wins 74.

(a) Do these data arise from sampling from two random processes or random sampling from two populations?
(b) Set up the null and alternative hypotheses for comparing the success probabilities of these authors.
(c) Calculate and interpret a 95% confidence interval for the difference in success probabilities between these two authors *using the Wilson adjustment* (see end of Investigation 2.1). Make sure it's clear how the interval was calculated and include your output.
(d) Based on your confidence interval, we will reject or fail to reject the null hypothesis in (b) at the 5% level of significance?
(e) In particular, are you convinced author B is a better solitaire player than author A? Explain.

10. Pulling All-Nighters

A study published in the January 2008 issue of the journal *Behavioral Sleep Medicine* involved a survey of 120 students at St. Lawrence University, a small liberal arts college in upstate New York. Researchers found that students who claimed to have never pulled an all-nighter have average GPAs of 3.1, compared to 2.9 for those students who do claim to have pulled all-nighters.

(a) Identify the explanatory and response variables in this study. Classify each as categorical or quantitative.
(b) Is this an observational study or a randomized experiment? Explain how you know.
(c) Suppose that the difference between these two averages is shown to be statistically significant. Can you legitimately conclude that pulling all-nighters causes a student's GPA to decrease? If so, explain why. If not, identify a potential confounding variable, and explain how it provides an alternative explanation for why the all-nighter group has a significantly lower average GPA.

11. Self-Conscious Exercising

In a study reported in the journal *Health Psychology* (Ginis, Jung, & Gauvin, 2003), researchers investigated whether the presence or absence of mirrors during an exercise session would affect women's attitudes toward the session. The subjects were 58 sedentary women, who answered questions about their self-image and mood before an exercise session. They then rode a stationary exercise bike for a 20-minute session. A week later the women returned for another 20-minute session, for which they were randomly assigned to exercise in front of either a mirrored or curtained wall. After each ride, they answered similar questions about their self-image and mood. It turned out that women tended to feel worse after riding in the mirrored room than in the curtained room.

(a) Identify the explanatory and response variables in this study. Classify each as categorical or quantitative.
(b) Is this an observational study or an experiment? If an experiment, is it a randomized comparative experiment? If observational, would you consider it a case-control, cohort, or cross-classified study? Explain how you are deciding.
(c) Explain why the random assignment is important in this study, as opposed to allowing women to choose for themselves whether to exercise in front of the mirror or the curtain.
(d) Does the design of the study allow for drawing a cause-and-effect conclusion between the mirrors and the women's worsened moods? Explain.
(e) Does the design of the study allow for generalizing conclusions to all adults? Explain.
(f) The first table in the research article describes characteristics of the sample, including the variables of age, body mass index, smoking status, and student status. Some of the characteristics reported

include:

(g)

Variable	Mirror			Curtain		
	Proportion	Mean	Std. Dev.	Proportion	Mean	Std. Dev.
Smoking	0.071			0.067		
Student	0.786			0.734		
Age		20.86	1.65		20.60	1.57
Body mass index		23.35	3.76		24.23	6.19

When the researchers compared the groups on these variables (e.g., proportion that smoked between mirror group and curtain group), they did not find any statistically significant differences between the treatment groups. Do you think they were happy about this or disappointed about this? Explain.

12. Effectiveness of AZT (cont.)

Recall the study in Exercise 7 about the effectiveness of AZT.

(a) Calculate *and interpret* the relative risk of HIV comparing the placebo group to the AZT group.

(b) Suggest why the researchers might prefer to look at the relative risk in this study rather than the difference in conditional proportions.

(c) Calculate and interpret a 95% confidence interval for the relative risk of HIV transmission between the population of placebo takers and the population of AZT takers. (Make sure you show your work.)

(d) Briefly explain, in your own words, why we are working with the *log* of the relative risk in these calculations.

(e) State the null and alternative hypotheses (in symbols *and* in words) for testing whether the risk of HIV-transmission is higher among placebo users than AZT users).

(f) Based on your confidence interval, do these data provide convincing evidence that placebo takers are more likely to transmit HIV to their babies than AZT takers? Explain how you are deciding.

(g) What population are you willing to generalize these results to?

(h) Does the design of this study allow you to conclude that use of AZT *causes* a lower risk of HIV-transmission? Justify your answer.

13. Pet Birds and Lung Cancer?

Researchers investigated whether owning a pet bird might be associated with having lung cancer. They studied a sample of 239 lung cancer patients and a sample of 429 people who did not have lung cancer, chosen to have similar characteristics to those with lung cancer. They asked all subjects whether they owned a pet bird in adulthood.

(a) Identify the explanatory and response variables in this study.

(b) What kind of observational study is this: case-control, cohort, or cross-classification? Explain briefly.

The researchers found that 98 of the lung cancer patients owned a pet bird, and 101 of those without lung cancer owned a pet bird.

(c) Organize these data into a 2×2 table, with the explanatory variable in columns.

(d) Calculate the odds ratio of having lung cancer, comparing those who owned a pet bird to those who did not.

(e) Use the normal approximation to test whether these data provide strong evidence that the probability of lung cancer differs between those who owned a pet bird and those who did not.
Report the hypotheses, test statistic and p-value, along with the test decision at the .05 significance level. Also verify that the technical conditions are satisfied, and summarize your conclusion from this test.

(f) Produce a 95% confidence interval for the odds ratio of having lung cancer between the two

populations. Also interpret this interval.
(g) Summarize your conclusion from this study and your analysis. Be sure to address the issues of causation and generalizability as well as statistical significance.

14. Nicotine Lozenge?

Helping smokers to quit continues to be a very important and challenging public health goal. In a recent study of the effectiveness of a nicotine lozenge, smokers who wanted to quit were recruited to participate thorough advertisements near four sites in the United Kingdom and 11 sites in the United States. Those smokers who met the screening qualifications were randomly assigned to one of two groups: one group received nicotine lozenges and the other group received placebo lozenges. The subjects were compared on various background variables at the beginning of the study, and at the end of the study they were compared on whether or not they successfully abstained from smoking. Of the 459 subjects in the nicotine group, 42.9% were male. Of the 458 subjects in the placebo group, 40.2% were male. Conducting a two-sided, two-sample *z*-test to compare these proportions gives a test statistic of $z = 0.84$ and a p-value of 0.399.

(a) What test decision would you make at the 0.05 significance level? Explain why the researchers would be pleased that the data resulted in this test decision.

At the end of the 52-week study, 17.9% of the nicotine group had successfully abstained from smoking, compared to 9.6% of the placebo group.

(b) Calculate and interpret an odds ratio from these data.
(c) Use the normal approximation to test whether these data provide strong evidence that the nicotine lozenge is more effective than the placebo lozenge, using the .01 significance level. Report the hypotheses, test statistic and p-value. Also verify that the technical conditions are satisfied, and summarize your conclusion from this test.
(d) Based on this study and your test result, is it legitimate to draw a cause-and-effect conclusion between the nicotine lozenge and the increased rate of abstaining from smoking? Explain.
(e) Explain what a Type II error would mean in the context of this study.
(f) Produce a 95% confidence interval for the odds ratio of successfully abstaining from smoking between the two groups. Also interpret this interval.
(g) Now consider only the subjects that received nicotine lozenges. Produce a 95% confidence interval to estimate the population proportion who would successfully abstain from smoking for 52 weeks when using the nicotine lozenge. Would you conclude that the nicotine lozenge is very successful for smokers who want to quit?

15. Odds Ratio vs. Relative Risk

Consider a generic two-way table classifying whether a person smokes and whether he/she has a disease:

	Smoker	**Non-smokers**
Disease	A	B
No disease	C	D

(a) Derive the expression for the difference in conditional proportions with the disease for smokers vs. non-smokers in terms of A, B, C, and D.
(b) Derive the expression for the relative risk of disease for the smokers compared to the non-smokers.
(c) Show the mathematical relationship between these two quantities. (Work the formulas algebraically so RR is on one side of the equation and the difference (plus some other values that can be calculated using A, B, C, and D) are on the other.)
(d) Another statistic we will consider is the *odds ratio*: (A/C) / (B/D). Show the mathematical relationship between the relative risk and the odds ratio in terms of A, B, C, and D.
(e) Using your expression in (d), how will the odds ratio and relative risk compare when both proportions of success are small?

16. Minority Coaches Discrimination?
Journalist Sandy Tolan reported in the book *Hank and Me* that in 1999 there were fewer minorities coaching at third base than at first base. Tolan argued that because third base is the more challenging of the two positions and typically leads to more managing responsibilities, this discrepancy constitutes evidence of discrimination against minority coaches. Of the 60 base coaches in Major League Baseball that year (30 at first base and 30 at third), 21 were minorities. Of these 21, only 6 coached at third base. Consider these data to representative of the overall hiring process in Major League Baseball.
(a) Fill in the following two-way table to represent these data:

	Minority coach	Non-minority coach	Total
First base coach			
Third base coach			
Total			

(b) What proportion of the minorities coached at third base in 1999? What proportion of the non-minorities coached at third base?
(c) What proportion of coaches coached at third base in 1999? Explain why it is not appropriate to calculate the relative risk for these data.
(d) Calculate the odds ratio for coaching at third base in these two groups. Do these data appear to support the journalist's argument? Explain.
(e) Calculate and interpret a 90% confidence interval for the underlying odds ratio for Major League Baseball's hiring process.
(f) Based on your interval and the study design, would it be appropriate to conclude that the coaches' minority status is causing the lower probability of coaching at third base? Explain.
(g) Suggest a potential confounding variable in this study and explain how it is confounding in this context (that is, how it provides an alternative explanation for the observed relationship).

17. Diet Weight Gains
In a study published in the October 2003 issue of the journal *Pediatrics*, researchers reported on a study that examined whether kids who diet may actually gain more weight than kids who do not diet. The study was based on questionnaire data provided by more than 16,000 U.S. boys and girls aged 9 to 14 from 1996 to 1998.
(a) Is this an observational study or an experiment? Explain.
(b) Identify the explanatory variable and the response variable in this study. Classify each as categorical or quantitative.
(c) Suppose the study found that kids who diet gain significantly more weight on average than kids who do not diet. Would it be reasonable to conclude that dieting *causes* weight gain in kids? Explain.

18. Sampling Senators
Suppose we want a sample of 5 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.

19. Challenger
The following two-way table displays data related to NASA shuttle launches prior to the fatal *Challenger* launch on January 28, 1986. The variables are

	Temp below 71°F	Temp above 71°F
O-ring damage	6	1
No O-ring damage	8	8
Total	14	9

(a) Among shuttle launches that took place when the temperature was below 71°F, what proportion had O-ring damage?
(b) Among shuttle launches that took place when the temperature was above 71°F, what proportion had O-ring damage?
(c) Construct a segmented bar graph based on your calculations in (a) and (b).
(d) Is it true that most of the launches at lower temperatures had O-ring damage? Explain.
(e) Is it true that most of the launches with O-ring damage took place at lower temperatures? Explain (including any additional calculations you need to support your answer).

20. Challenger (cont.)
Reconsider the space shuttle data in the previous exercise.
(a) Calculate the relative risk of O-ring damage at the lower temperatures as compared to the higher temperatures. (Show your work.)
(b) Calculate the odds ratio of O-ring damage for the low-temperature group compared to the high-temperature group. (Show your work.)
(c) Write a paragraph summarizing your finding from these calculations to the NASA officials.

21. Challenger cont.
Reconsider the space shuttle O-ring data.
(a) Apply Fisher's Exact Test to these data. Interpret what the p-value represents, and summarize your findings.
(b) Is this an observational study or a controlled experiment? Explain.
(c) How does your answer to (b) affect the conclusion that you can draw from Fisher's Exact Test? Explain.

22. Have a Nice Trip
For the 24 subjects in the tripping experiment (Investigation 2.4):
(a) One could suggest tossing a coin 24 times and if the coin lands heads, sending the person to group A, but if the coin lands tails, sending the person to group B. Discuss a minor weakness of this randomization method. Do you think this weakness could lead to confounding? Explain.
(b) Explain how you could use a coin to carry out the randomization to achieve an equal number of subjects in the two groups. Discuss a weakness of this randomization method. Do you think this weakness could lead to confounding? Explain.

23. U.S. Open
The following table pertains to the results of the 128 first-round matches in the 2004 U.S. Open tennis tournament:

	Men's singles	Women's singles
2 sets	0	40
3 sets	22	24

4 sets	23	0
5 sets	15	0
Other "retired"	4	0

(a) Identify the observational units.
(b) Identify the two variables presented in the table.
(c) What proportion of the men's matches (excluding the "retired" matches) lasted for at least 4 sets?
(d) What proportion of the men's matches (excluding the "retired" matches) lasted for at most 4 sets?
(e) What proportion of the matches that were completed in 3 sets were played by women?
(f) Who had a higher proportion of matches that lasted for the maximum possible number of sets – men or women? Support your answer with appropriate calculations.
(g) Construct and comment on a segmented bar graph to display the distributions of match lengths between the two genders.

24. Top 100 Films

In 1998 the American Film Institute created a list of the top 100 American films ever made (www.afi.com/Docs/tvevents/pdf/movies100.pdf). Two friends decide to watch one of these films and want to find a film that neither has seen. The following table indicates whether or not each friend has seen the film.

	Beth yes	Beth no	Total
Allan yes	42	8	
Allan no	15	35	
Total			100

(a) Create a segmented bar graph to display these data, and comment on what it reveals.
(b) Are the variables "whether or not Allan has seen the film" and "whether or not Beth has seen the film" independent? Justify your answer. Interpret the implication of your answer in this context – what does this tell you about Allan and Beth's taste in movies?
(c) Beth tries the same thing with her friend Dave:

	Beth yes	Beth no	Total
Dave yes	20	40	
Dave no	37	3	
Total			

Are the variables "whether or not Beth has seen the film" and "whether or not Dave has seen the film" independent? How does the direction of the dependence compare to that of Allan and Beth?
(d) Now suppose Ellen has seen 80% of the films Dave has seen. Complete the two-way table for Ellen and Dave so that the variables "whether or not Ellen has seen the movie" and "whether or not Dave has seen the movie" are independent.

25. Emergency Visits

The Community Tracking Study household survey, "a nationally representative telephone survey of the civilian, non-institutionalized population" (1999 and 2001) produced estimates of how satisfied different patients were with their emergency room visit. One question asked them to rate how well the doctor listened. This was then related to the type of insurance that the person carried, and the results are reproduced below.

	Rated as good or excellent listener	Not rated as good or excellent listener
Private insurance	23402	19935
Medicare	9936	6090
Medicaid	9209	9208

Self-pay or no charge (not insured)	6847	11172

(a) Is this an observational study or an experiment?
(b) Which group seems to be more satisfied with how the doctor listened?
(c) Create numerical and graphical summaries to describe how the perception of the doctor's listening varied across the insurance groups. Write a paragraph summarizing your results.

26. Popcorn Production
In May 2000, eight people who had worked at the same microwave-popcorn production plant reported to the Missouri Department of Health with fixed obstructive lung disease. These workers had become ill between 1993 and 2000 while employed at the plant. On the basis of these cases, in November 2000 researchers began conducting medical examinations and environmental surveys of workers employed at the plant to assess their occupational exposures to certain compounds (Kreiss, et al., 2002). As part of the study, current employees at the plant underwent spirometric testing. This measures FVC – forced viral capacity – the volume of air that can be maximally, forcefully exhaled. On this test, 31 employees had abnormal results, including 21 with airway obstruction. A total of 116 employees were tested.
(a) Calculate the baseline risk of airway obstruction in this plant.
Researchers were curious as to whether the baseline risk was high among microwave-popcorn workers. The plant itself was broken into several areas, some of which were separate from the popcorn production area. Air and dust samples in each job area were measured to determine the exposure to diacetyl, and marker of organic chemical exposure. Employees were classified into two groups: a "low-exposure group" and a "high-exposure group," determined by how long an employee spent at different jobs within the plant and the average exposure in that job area. Of the 21 participants with airway obstruction, 6 were from the low-exposure group and 15 were from the high-exposure group. Both exposure groups contained 58 total employees.
(b) Construct the two-way table, segmented bar graph, and conditional proportions for comparing airway obstructions in the low and high exposure groups.
(c) Compute the relative risk and odds ratio of airway obstruction for the high-exposure group compared to the low-exposure group. Include an interpretation of these numbers.
(d) Do the high-exposure employees appear more likely to have airway obstructions than the low-exposure group? Give reasons to support your answer.

27. More Popcorn Problems
Reconsider the previous exercise.
(a) Explain why it was appropriate to calculate the difference in proportions and relative risk in the popcorn study by not for the Wynder and Graham smoking study of Investigation 2.10.
(b) Instead of conditioning on the level of exposure (in the popcorn study), calculate the conditional proportions of being in the high-exposure group between those with airway obstruction and those without. Do these values give evidence of an association between these variables? Explain, being clear about how you interpret these values.

28. Popcorn Production (cont.)
Reconsider the two previous exercises. Twenty employees working on the plain-popcorn packaging line, the bag-printing areas, the warehouse, the offices, and the outdoor areas were considered an internal reference group to the employees who worked in the other work areas. Trained interviewers administered a standardized questionnaire to employees of the popcorn plant (written informed consent was obtained from all participating employees). One of the questions on the survey concerned "extertional shortness of breath" (shortness of breath when hurrying on level ground or walking up a slight hill), which was

reported by 31 participants – 1 from the reference group of 20 employees and 30 from 92 of the production employees.

(a) Identify the explanatory and response variables. Specify whether they are quantitative or categorical.
(b) Construct the two-way table and segmented bar graph comparing the exertional shortness of breath of the reference group to the production employees.
(c) Calculate the baseline rate of exertional shortness of breath, the relative risk, and the odds ratio to compare the exertional shortness of breath of the reference group and the production employees. Include interpretations of these numbers.
(d) Do the production employees appear more likely to have exertional shortness of breath than the reference group? Give reasons to support your answer.
(e) Not all of the workers in the plant completed the questionnaire. Suggest some cautions you might have in interpreting the results because of this.

29. Popcorn Production (cont.)

Recall the popcorn workers from Exercise 26. Another issue is how these workers compared to the general population. Information about the general population was obtained from the National Health and Nutrition Examination Survey (NHANES III). The NHANES study found 8% of current or former smokers with airway obstruction, 2.3% of nonsmokers with airway obstruction, and 5.5% as the baseline rate of airway obstruction in the general population. At the popcorn plant, 8 of 64, 13 of 52, and 21 of 116 workers with airway obstruction in the smokers, nonsmoker, and overall groups, respectively, were observed. Of interest is whether the rate of airway obstruction is higher among the population of popcorn workers than in the general population.

(a) Consider the population of all microwave popcorn workers who are current or former smokers. Let π represent the proportion of this population with airway obstruction. Carry out a test of significance to determine whether there is evidence that the airway obstruction rate in this population is higher than the national rate for former and current smokers. (State your hypotheses (in symbols and words, calculate the test statistic, report the (approximate) p-value, and state your conclusion in context.)
(b) Repeat (a) for the population of microwave popcorn workers who are nonsmokers and the corresponding national rate.
(c) Repeat (a) for the population of microwave popcorn workers and the corresponding national rate.
(d) Which group (or groups) has the strongest evidence that the rate of airway obstruction among popcorn factory workers is higher than the national rate? Explain.

30. Familiar Faces?

In a study of whether other species are comforted by the sight of a familiar face, researchers placed 40 sheep on their own in a darkened barn and showed them various faces. Some were shown faces of sheep familiar to them, whereas others were shown faces of goats. The stress levels of the sheep were then measured by its number of bleats, heart rate, movement within the barn, and cortisol and adrenaline levels in its blood.

(a) Identify the observational units in this study.
(b) Is this an observational study or an experiment? Explain.
(c) Explain what is wrong with identifying each of the following as a variable in this study (say more than simply "this is not a variable"):
   - The number of sheep whose stress level increased after seeing a goat face
   - The proportion of sheep who were shown a familiar sheep's face
   - Whether sheep are comforted by seeing a familiar face
(d) Clearly identify the explanatory and response variables in this study, and classify them as categorical or quantitative.

(e) If the focus of the study is on the "familiar" aspect of the face, how might the design of the study be improved? [*Hint*: Is there a better alternative than photos of goat faces to present sheep with?] Explain why your design is an improvement.

31. Ginko for Memory?

Solomon et al. (2002) report on a study evaluating whether ginkgo, an over-the-counter agent marked as enhancing memory, improves memory in elderly adults. Participants randomly received either 40 mg of ginkgo three times a day or a placebo treatment. Many different variables were measured such as standardized neuropsychological tests of memory, attention, and concentration, as well as a "caregiver clinical global impression of change" completed by a companion.

(a) Identify the observational units, response variable, and explanatory variable. Speculate as to whether the variables were quantitative or categorical.
(b) Is this an observational study or an experiment? Explain.

The following two-way table displays the results of the companion response on the Caregiver Global Impression of Change Rating Scale.

Response	Ginkgo	Placebo
1 – Very much improved	0	0
2 – Much improved	2	1
3 – Minimally improved	31	35
4 – No change	77	70
5 – Worse	0	0

(c) Produce a segmented bar graph comparing the two groups. Describe what this plot reveals.
(d) Calculate the relative risk of some improvement in memory (response = 1, 2, or 3) for the Ginkgo group compared to the placebo group. Be sure to show your work and include a 1-2 sentence summary of your results.
(e) Would it be reasonable to conclude that the use of Ginkgo improved the memory ratings of these subjects from this study? Explain.

32. The Titanic

The following tables display the status of the passengers of the Titanic, their gender and whether they were a child or an adult, as determined by Dawson (1995). (The historical sources didn't completely agree with each other.)

	Child	
	Male	**Female**
Survived	29	28
Died	35	17

	Adult	
	Male	**Female**
Survived	338	316
Died	1329	109

(a) What was the odds ratio of survival for a child compared to an adult?
(b) What was the odds ratio of survival for a female compared to a male?
(c) Was the association between survival status and gender stronger for the children or for the adults? Justify your conclusion.
(d) Is it reasonable to conclude from your calculations in (a) and (b) that females were about ten times more likely to survive than males? Explain.
(e) Fill in the cells of the following (hypothetical) table so that the gender and survival status variables are independent.

	Male	Female	Total
Survived			
Died			
Total			

(f) Calculate the odds ratio for the table you created in (e). Does its value make sense? Explain.

33. Donner Party
In 1846, the Donner and Reed families left Springfield, IL, for California. In July, they reached Fort Bridger, WY. The leaders decided to attempt a new and untested route to the Sacramento Valley, where they were delayed by difficult river and desert crossings. The group became stranded in the eastern Sierra Nevada mountains when the region was hit by heavy snows in late October. By the time the last survivor was rescued on April 21, 1847, 40 of the 87 members had died from famine and exposure to extreme cold. Among the adults (15 years and older), 5 of the 15 females died, and 20 of the 30 males died.
(a) Create a two-way table and segmented bar graph to summarize the gender differences in survival for the Donner Party.
(b) Were females more likely to survive than males? Justify your answer.
(c) Would it be reasonable to use these results as proof that females were more apt to survive than males based solely on their gender? Explain.

34. Sleepless Drivers
Connor et al. (2002) reported on a study that investigated whether sleeplessness is related to car crashes. The researchers identified all drivers or passengers of eligible light vehicles who were admitted to a hospital or died as a result of a car crash on public roads in Auckland, New Zealand between April 1998 and July 1999. They identified a sample of drivers who had been involved in a crash resulting in injury, and another sample of drivers who had not been involved in a crash. The researchers interviewed all participants in the study with questions about their sleep; one question involved measuring their sleepiness through the Stanford sleepiness score. The following table summarizes their findings:

	Most alert (score 1)	Moderately alert (scores 2-3)	Least alert (scores 4-7)	Total
Crash drivers	175	272	63	510
Non-crash drivers	322	256	8	586

(a) Is this an observational study or a controlled experiment? Explain.
(b) If this is an observational study, is it a case-control, cohort, or cross-sectional study? If it is an experiment, is it a randomized experiment? Explain.
(c) Produce a segmented bar graph displaying the distribution of crash/not for each of the three sleepiness categories, and comment on what it reveals.
(d) Produce a segmented bar graph displaying the distribution of sleepiness between crash victims and non-crashers, and comment on what it reveals. Would you draw a similar conclusion from this graph as the previous one?
(e) What proportion of the drivers in this study was involved in car crashes? Does this study allow you to estimate the proportion of New Zealand drivers involved in car crashes? Explain.

35. Sleepless Drivers (cont.)
The New Zealand researchers also recorded whether a driver had obtained less than 5 hours of sleep in the previous 24 hours. They found that 65 of 529 drivers in the case group had less than 5 hours, compared to 30 of 584 drivers in the control group.
(a) Calculate a 90% confidence interval for the population odds ratio, τ. Remember to provide an interpretation of the sample odds ratio and of this confidence interval.
(b) Does this interval provide convincing evidence that the odds of a crash are higher for the population with less sleep? Explain.

36. Woman's College Basketball
A February 2004 Harris poll asked a nationwide sample of 2204 adults whether or not they followed college basketball. When asked about women's basketball, 7% said they did follow. When the same

question was asked in December 1998, 8% of 1005 respondents said they did follow women's college basketball.

(a) Is there evidence of a change in the women's basketball following over this 6-year period? (Examine both the difference in proportions and the odds ratio.)
(b) Construct and interpret a 95% confidence interval for the population odds ratio, τ.

37. Internet Use

The Pew Internet and American Life Project examines how Americans use the internet. In 2002 the organization took samples of people from across the country and asked questions about their use of the internet. Consider the following information from this study:

	Northeast	South	Midwest	West
Sample size	3973	4332	4929	5137
# of internet users	2417	2372	2831	3259

(a) Turn this information into a genuine two-way table of counts, with region of the country as the explanatory variable and whether or not the person uses the internet as the response.
(b) Would you conclude that people in the Midwest use the internet more than people in the Northeast, because there were more internet users in the Midwest than in the Northeast in this study? Explain.
(c) For each of the four regions, calculate the proportion of interviewees who are internet users.
(d) Construct a segmented bar graph to display these data, and comment on what it and your calculations in (b) reveal.

38. Odds ratios versus Relative Risk

Consider a generic two-way table classifying whether a person smokes and whether he/she has a disease:

	Smoker	Non-smoker
Disease	A	B
No Disease	C	D

(a) Derive the expression of the odds ratio of disease for smokers compared to non-smokers in terms of A, B, C, and D.
(b) Derive the expression for the odds ratio of being a smoker for the disease group compared to the no disease group. How does this expression compare to that from (a)?
(c) Derive the expression for the relative risk of disease for the smokers compared to the non-smokers.
(d) Derive the expression for the relative risk of being a smoker for the disease group compared to the no disease group. Is this expression the same as (c)?

39. Odds ratio versus Relative Risk (cont.)

The odds ratio and relative risk are closely related. Again consider the following generic two-way table classifying whether a person smokes and whether he/she has a disease:

	Smoker	Non-smoker
Disease	A	B
No Disease	C	D

(a) Let $\hat{p}_1 = \dfrac{A}{A+C}$ be the proportion of success in the smoker group, and let $\hat{p}_2 = \dfrac{B}{B+D}$ be the proportion of success in the non-smoker group. Derive the mathematical relationship for which the odds ratio is equal to the relative risk times a multiplicative factor involving these success proportions.
(b) Using your expression in (a), how will the odds ratio and relative risk compare when both proportions of success are small?

(c) For what types of values for the proportions of success will the odds ratio and relative risk be quite different from each other?
(d) *Prove* that the relative risk and odds ratio will be equal if the proportions of success are equal.
(e) More than that, prove that the relative risk and odds ratio will both be equal to one if the proportions of success are equal.

40. Red Dye #2

The following three-way table presents the results of a study to assess whether red dye #2 causes cancer in laboratory rats. The three variables are: dosage (high or low), presence of cancerous tumor (yes or no), and whether the rat survived the 131-week study (died or sacrificed):

	Died		**Sacrificed**	
	Low	**High**	**Low**	**High**
Cancerous tumor	4	7	0	7
No cancerous tumor	26	16	14	14
Total	30	23	14	21

(a) Is this an observational study or an experiment? Explain.
(b) Identify the explanatory variable. How do you suspect that the researchers assigned this variable to the rats?
(c) Identify the primary response variable.
(d) Compare the proportion of high-dosage rats who had a cancerous tumor to the same proportion for low-dosage rats. Also create a segmented bar graph to display these proportions. Does the difference appear to support the claim that red dye #2 causes cancer? Explain.
(e) Calculate the relative risk and odds ratio of having a cancerous tumor between the two dosage groups. Show your work and write a sentence interpreting each of these calculations in this context.
(f) Among the rats that died during the experiment, what proportion had a cancerous tumor? Among the rats that were sacrificed at the end of the experiment, what proportion had a cancerous tumor? Are these proportions similar?
(g) Consider the alternative explanation: "If the rats had been allowed to live longer, perhaps more of them would have developed cancer. In particular, maybe more of the low dosage group would contract cancer eventually." Explain how the calculations in (f) help to refute this argument.

41. Fisher vs. Newton

Suppose that Dr. Fisher teaches an 8am statistics class and Dr. Newton teaches an 11am calculus class. During lunch one day they begin comparing notes on their students' attitudes in class, and they quickly agree that the calculus students are much more engaged and enthusiastic than the statistics students. Consider the students' subject (statistics or calculus) as the explanatory variable and a measure of the students' engagement and enthusiasm as the response variable.
(a) Identify two confounding variables, and explain how they prevent you from concluding that the students' subject is the cause of the difference in their engagement and enthusiasm levels. (Be sure to state them clearly as variables.)
(b) Suppose that next semester Dr. Fisher will teach his statistics class at 11am and Dr. Newton will teach his calculus class at 8am. Then they will compare all of the data gathered on their students' levels of engagement and enthusiasm from both semesters. Does this plan eliminate all potentially confounding variables? Explain.

42. Second-hand Smoke and Children

Consider the question of whether exposure to second-hand smoke is harmful to the health of children.
(a) Describe a prospective (cohort) observational study that could address this question.
(b) Describe a retrospective (case-control) observational study that could address this question.
(c) Describe a cross-classified observational study that could address this question.
(d) Describe how you could (in principle) design an experiment to address this question.
(e) Would it be ethical to conduct an experiment to address this question? Explain.

43. Breast-fed Babies

A study of 9,357 German children (von Kries et al., 1999) that mothers who breast-feed their babies feel warmer and more reception toward the babies than mothers who bottle-feed.
(a) Identify the observational units in these studies.
(b) Identify the explanatory and the response variable in these studies.
(c) Do you suspect that these are observational studies or controlled experiments? Explain.
(d) Suppose that you read a magazine article that cites this research and concludes that breast-feeding has positive effects on mothers' attitudes toward their children. Would you agree with this conclusion? Explain.

44. ADHD Children

A recent study (Milberger et al., 1997) study found that children whose mothers smoked during pregnancy were more likely to be diagnosed with attention deficit hyperactivity disorder than children of nonsmoking mothers.
(a) Is this an observational study or an experiment? Explain briefly.
(b) Identify the explanatory variable and the response variable in this study.
(c) Is it valid to conclude from this study that a mother's smoking causes this disorder in her children? Explain briefly.
(d) Explain why it would not be reasonable to conduct a randomized, comparative experiment to investigate whether a mother's smoking causes this disorder in her children.

45. St. John's Wort for Depression?

An April 9, 2002 news release from the National Institutes of Health describes a study that investigated whether the herb St. John's wort is effective for treating depression. An excerpt from this press release follows:

> An extract of the herb St. John's wort was no more effective for treating major depression of moderate severity than placebo, according to research published in the April 10 issue of the *Journal of the American Medical Association*. The randomized, double-blind trial compared the use of a standardized extract of St. John's wort (*Hypericum perforatum*) to a placebo for treating major depression of moderate severity. The multi-site trial, involving 340 participants, also compared the FDA-approved antidepressant drug sertraline to placebo as a way to measure how sensitive the trial was to detecting antidepressant effects.

(a) Is this an experiment or an observational study? Explain.
(b) Identify the explanatory and response variables in this study.
(c) Explain why randomization was important in the context of this study.
(d) Explain what it means for this study to be "double-blind" and also why this is an important component of the study design.

46. School Uniforms

In the mid-1990s some public schools instituted a policy of requiring students to wear school uniforms, a policy long associated with private schools, in the hope of decreasing behavioral problems and increasing academic performance. President Clinton gave national attention to this issue by commending the policy in his 1996 State of the Union Address.

(a) Suppose that a particular school district implements a mandatory school uniform policy and then observes that it has fewer behavioral problems and increased class attendance in the following year. Would it be reasonable to conclude that the uniform policy caused these changes? If your answer is yes, explain why. If your answer is no, explain why and identify a potential confounding variable.

(b) Suppose that a researcher takes a random sample of schools that do have a uniform policy and a random sample of schools that do not have a uniform policy. The researcher then follows the students for several years and records various measures of their academic performance. If the students from schools with a uniform policy tend to perform better than students from schools without such a policy, would it be reasonable to conclude that the uniform policy is effective? Explain, as in (a).

47. Nightlights and Near-Sightedness (cont.)

Zadnik et al. (2000) conducted a study of children's lighting conditions and myopia (near-sightedness). They studied 1220 children who were seen by various optometrists in schools across the U.S. The researchers found that of 417 children who had slept with no light on, 20% became myopic (near-sighted); of 758 children who slept with a night light on, 17% became myopic; and of 45 children who slept in a fully lit room, 22% became myopic.

(a) Create a segmented bar graph to compare the conditional distributions of myopia across the three lighting categories.

(b) Does this graph reveal much of an association between lighting condition and myopia? How would your findings from this study compare to those of Investigation 2.2?

(c) These researchers pointed out that their subjects came from schools across the country, whereas the subjects in the Quinn et al. study had come to one specialty clinic. Does this raise the issue of cause-and-effect or the issue of generalizability-of-results? Explain.

48. Randomization?

In a clinical trial, data collection usually begins at "baseline," when the subjects are first recruited into the study but before they are randomly assigned to treatment or control groups. Suppose that you are called in as a consultant on two clinical trials for treatments that aim to reduce the risk of heart attacks. You ask to see summaries of baseline data on subjects' weight and are presented with the following:

		Number of subjects	Mean weight	SD of weights
Trial 1	**Treatment**	101	154 lb	25 lb
	Control	99	175 lb	26 lb
Trial 2	**Treatment**	98	163 lb	27 lb
	Control	102	166 lb	25 lb

(a) In which trial do you have reason to suspect that randomization was not performed correctly? Explain.

(b) How will this concern affect the conclusions that you can draw from this study? Explain.

49. Randomizing Volleyball Players

Suppose that the names of 6 volleyball players are: Alex, Bob, Cheryl, Diego, Elizabeth, and Flo.

(a) Explain how you could use index cards to simulate the process of randomly assigning three of them to be on the same team. Include sufficient detail that someone else could implement your random assignment plan without any further knowledge or instructions.

(b) Now explain how you could use a fair, six-sided die to simulate this process. Again include sufficient detail that someone else could implement your plan. Be sure to provide instruction about what to do if you roll the die twice and obtain the same number each time.
(c) Now explain how you could use the random number feature on a calculator or computer, to simulate this random assignment.

50. Randomizing Volleyball Players (cont.)
Reconsider the previous question about assigning six volleyball players to two teams of three players each.
(a) How many different such assignments are possible?
(b) Determine the (exact, not simulated) probability that random assignment would put all three women on one team and all three men on the other. Is this probability small enough that you would be very surprised to see random assignment produce teams where the genders are completely separated?

51. Randomizing
Suppose that the subjects in an experiment are to be randomly divided into two groups.
(a) Suppose that there are 8 subjects, of whom 4 are men and 4 are women. Determine the probability that 2 subjects of each gender are assigned to each group.
(b) Now consider the general case that there are $4N$ subjects, of whom $2N$ are men and $2N$ are women. Derive an expression for the probability that N subjects of each gender are assigned to each group, as a function of N.
(c) Produce and submit a graph of your function from (b), for values of N ranging from 1 to 10. Does the function appear to be increasing or decreasing? Explain why this makes sense.

52. Creating Committees
Suppose that an n-person committee consists of 2 women and (n–2) men, and suppose that two representatives are to be chosen at random.
(a) Before conducting the analysis, do you expect the probability of both women being selected to increase, decrease, or remain the same as the committee size n increases? Explain.
(b) Use the multiplication rule to determine the number of possible pairs that could be chosen (i.e., the number of outcomes in the sample space), as a function of n.
(c) Determine the probability that both women are chosen, as a function of n.
(d) Evaluate this probability for the following values of n: 6, 8 (which you already analyzed), 10, 20, 30, and 50. [*Hint*: Feel free to use technology.]
(e) Graph this function. Comment on its behavior, especially on whether it is increasing or decreasing and how rapidly.
(f) What is the smallest value of n for which this probability first dips below 0.05? Below 0.01? Below 0.001?
(g) Let the random variable X represent the number of women chosen. Determine the probability distribution of X. [*Hint*: Each of the three probabilities will be a function of n. Verify that the three probabilities sum to one.]
(h) Determine the expected value of X as a function of n. Graph this function. Comment on its behavior, and explain why this makes sense.

53. Hypergeometric Probabilities

Suppose that a group of 20 subjects contains 10 men and 10 women. Suppose that these subjects are to be randomly assigned to two treatment groups.

(a) Create a graph of the hypergeometric probability distribution of X, the number of women randomly assigned to group A.
(b) Use the hypergeometric distribution to determine the probability that 5 of each gender end up in each group. Is this outcome more likely than not?
(c) Use the hypergeometric distribution to determine the probability that all 10 of each gender end up in different groups.
(d) If the experiment does end up with all men in one group and all women in the other, would you have reason to doubt that randomization was applied correctly? Explain.

Now suppose that a group of 50 men and 50 women are to be randomly assigned to two treatment groups.

(e) Without doing any calculations, produce a rough sketch of the probability distribution of the number of women randomly assigned to group A.
(f) Would the probability of an exactly even gender split be larger, smaller, or the same as in (b)? Explain, without providing any calculations.
(g) Would the probability of finding all men in one group and all women in the other be larger, smaller, or the same as in (c)? Explain, without providing any calculations.

54. Testing for AIDS

The ELISA test for AIDS was used in the screening of blood donations in the early 1990's. As with most medical diagnostic tests, the ELISA test was not infallible. If a person actually carries the AIDS virus, experts estimated that this test gave a positive result 97.7% of the time. If a person did not carry the AIDS virus, ELISA gave a negative result 92.6% of the time. Estimates at the time were that 0.5% of the American public carried the AIDS virus. You will determine the (conditional) probability that a person actually carried the AIDS virus given that he/she tested positive on the ELISA text.

(a) First, without doing any calculations, take a guess for the value of the probability that a person who tests positive carries the virus:

Imagine a hypothetical population of 1,000,000 people for whom these percentages hold exactly. (The population size is chosen to be so large in order to make the calculations all work out to be integers.) For questions (b)-(e) below, record your answers in the appropriate cells of the table.

	Positive test	**Negative test**	**Total**
Carries AIDS virus	(c)	(c)	(b)
Does not carry AIDS	(d)	(d)	(b)
Total	(e)	(e)	1,000,000

(b) Assuming that 0.5% of the population of 1,000,000 people carries AIDS, how many such carriers are there in the population? How many non-carriers are there?
(c) Consider for now only the carriers. If 97.7% of them test positive, how many people test positive? How many carriers does that leave who test negative?
(d) Now consider only the non-carriers. If 92.6% of them test negative, how many test negative? How many non-carriers does that leave who test positive?
(e) Determine the total number of positive test results and the total number of negative test results.
(f) Among those who test positive, what proportion actually carry the disease? How does this compare to your prediction above?
(f) Explain why this probability turns out to be small compared to the 97.7% and 92.6% figures cited above. (Be sure to refer to calculations in the table.)

55. Sample Sizes and Fisher's Exact Test

Suppose that you randomly assign 8 subjects into two treatment groups (comparing a new drug to a standard drug) of 4 each. You then record the results (improvement in symptoms or lack thereof) for each subject.

(a) Determine the p-value of Fisher's Exact Test in for the most extreme possible outcome: all 4 subjects receiving the new drug experience improve and all 4 subjects with the standard drug do not.

(b) Determine all possible outcomes of this experiment for which the p-value of Fisher's Exact Test would be less than 0.05. [*Hint*: Consider the number of "successes" to be anything between 0 and 8, inclusive.]

(c) Based on your answers to (a) and (b), does a sample size of 8 subjects provide much opportunity for determining whether the new drug is more effective than the standard? Explain.

56. Sample Sizes and Fisher's Exact Test (cont.)

Reconsider the previous problem. Now suppose that there are *n* subjects in each group.

(a) Express the p-value of Fisher's Exact Test for the most extreme outcome (that all *n* subjects receiving the new drug improve and all *n* with the standard drug do not), as a function of *n*.

(b) Use technology to calculate this p-value for values of *n* from 1 through 20, inclusive.

(c) Produce a graph of this p-value as a function of *n*. (Remember that *n* must be an integer, so do not "connect the dots.")

(d) Explain why it makes sense that this p-value is a decreasing function of *n*.

57. Red Dye # 2 (cont.)

Reconsider the red dye #2 experiment described in Exercise 40, in which 88 rats were randomly assigned to receive either a high or a low dosage of red dye #2. It turned out that 4 of the 44 rats in the low dosage group contracted a cancerous tumor, compared to 14 of the 44 rats in the high dosage group.

(a) Construct a 2× 2 table to display the experiment data. (Remember to put the explanatory variable on the columns and the response on the rows.)

(b) Use hypergeometric probabilities to determine the probability of getting (exactly) four cancerous rats in the "low dosage" group, if in fact there were actually no effect of dosage. Include the details of your calculation.

(c) Repeat (b) for the probability of three, two, one, and zero cancerous rats in the "low dosage" group.

(d) Use these calculations to determine the p-value for Fisher's Exact Test.

(e) Do the experimental results indicate that red dye #2 causes cancer in laboratory rats? Explain the reasoning behind your conclusion, referring both to these calculations and to the design of the study.

58. Gender Discrimination?

Recall the bank manager gender discrimination study, Rosen and Jerdee (1974) reported on a study of 48 male bank supervisors (attending a management institute at the University of North Carolina) who were each sent the same personnel file and asked to judge whether the person should be promoted to a branch manager position. The files were identical except that half of them were randomly assigned to be the file of a female and the other half indicated that the file was that of a male. The researchers suspected that a higher proportion of "males" would be recommended for promotion. Of the 24 "male" files, 21 were recommended for promotion. Of the 24 "female" files, 14 were recommended for promotion.

(a) Is this an experiment or observational study?

(b) In this study, the subjects believed that they were all participating in an identical exercise dealing with personnel problems in the banking industry. Explain why it is important for the subjects to be "blind" to the fact that there were other versions and that the researchers were focusing on the gender of the applicant.

(c) Calculate the p-value from Fisher's Exact Test for this study.
(d) Write a paragraph describing the results of your analysis to the researchers. In addition to stating your conclusion, explain the reasoning process that leads to this conclusion. What does this p-value tell you?
(e) Does the design of this study allow you to draw a cause-and-effect conclusion regarding gender and likelihood of promotion? Justify your answer.

59. Gender Discrimination (cont.)
Recall the bank manager gender discrimination study (from the previous exercise). The results that you already examined were based on the subjects being told that the nature of the manager's job was "routine." The researchers also examined how the subjects would respond to the applicant's gender when the nature of the managerial position was described as "complex." They expected that the greater tendency to promote the male would be even stronger for the more complex job. In this case, 5 of 25 women were recommended for promotion and 11 of 20 men were recommended for promotion.
(a) Do these data provide convincing evidence that men were more likely to be recommended for promotion than women for a complex job? Justify your answer with appropriate calculations (show your work and/or output) and explanations. Clearly explain in your own words what the calculated probability represents.
(b) How do these results compare to those of the "routine" job? Does the tendency to promote men seem even stronger for the complex job? Explain.

60. Thoughts of Death
Psychologist recently studied whether voters' view can be influenced by appeals to fear. One aspect of their study involved asking subjects to think either about their own death or about a neutral topic. The researchers then presented subjects with descriptions of three hypothetical candidates, each with a different leadership style. One candidate was described as charismatic, another as task-oriented, and the third as relationship-oriented. Of the 100 students asked to think about the neutral topic, only 4 chose the charismatic candidate, compared to 30 of the 100 students who had been asked to think about their own death.
(a) Is this an observational study or an experiment? Explain.
(b) Identify the explanatory and response variable in this study. Classify each as categorical or quantitative.
(c) Create the two-way table for displaying these data.
(d) Determine the p-value of Fisher's Exact Test to assess the degree to which these data support the researchers' suspicion that thoughts of death increase the appeal of charismatic candidates. Show your work and/or output and summarize your conclusions.

61. Letrozole vs. Tamoxifen
The November 6, 2003 issue of the *New England Journal of Medicine* reported on a study of the effectiveness of letrozole in postmenopausal women with breast cancer who had completed five years of tamoxifen therapy. Over 5000 women were enrolled in the study. They were randomly assigned to receive either letrozole or a placebo. The primary end result studied was disease-free survival.
(a) Is this an experiment or an observational study? Explain.
(b) Identify the observational units (subjects) in this study.
(c) Identify the explanatory variable and the response variable in this study. Are they quantitative or categorical?
The article reported that 92.8% of the 2575 women who received letrozole achieved disease-free survival,

compared to 86.8% of the 2582 women in the placebo group.

(d) Construct a two-way table to display the experimental results (you will have to round to the nearest integer). Remember to use the explanatory variable for the columns and the response variable for the rows.
(e) Produce a segmented bar graph and describe what is revealed by this graph.
(f) Apply Fisher's Exact Test to these data. [*Hint*: Because the numbers involved are so large, use technology to do the calculation. Be sure to include a copy of your output.]
(g) Explain what the p-value of Fisher's Exact Test means in this context.
(h) What conclusion would you draw about whether letrozole increases the likelihood of disease-free survival? Also explain the reasoning process by which your conclusion follows.
(i) Considering the design of the study, is a cause-and-effect conclusion warranted here? Explain.
(j) If the sample sizes had been smaller, and the sample proportions had turned out the same, how would the p-value of the test have changed, if at all? Explain.

62. Wearing Caps

An instructor suspected that male students are more likely to wear a baseball cap in class than female students. While proctoring an exam, he noted that 5 of 18 male students were wearing a cap, whereas none of the 6 female students were. Suppose for now that you were to randomly distribute the five caps to be worn among these 24 students, and let the random variable X be the number of males that would be given caps.

(a) Explain why X has a hypergeometric distribution, and state the values of its parameters.
(b) Determine and interpret the expected value of X; that is, the expected number of caps distributed to men.
(c) List the possible values of X and their probabilities. Also provide a graph of this probability distribution. [*Hint*: Feel free to use technology.]
(d) Do the observed data provide strong evidence that male students tend to wear caps in class more than female students do? Report the p-value of Fisher's Exact Test, and explain the reasoning behind your answer.
(e) Explain and show two other ways that the p-value for Fisher's Exact Test could have been calculated.

63. Lost Bills and Tickets

Students in a statistics class were randomly assigned to receive one of the following two versions of a similar question:

- Suppose that you have decided to see a play for which the admission charge is $20 per ticket. As you prepare to enter the theater, you discover that you have lost a $20 bill. Would you still pay $20 for a ticket to see the play?
- Suppose that you have decided to see a play and paid the admission price of $20 per ticket. As you prepare to enter the theater, you discover that you have lost the ticket. The seat was not marked, and the ticket cannot be recovered. Would you pay $20 for another ticket?

Before collecting the students' responses, the instructor conjectured that those who lost the ticket would be less likely to buy another ticket than those who lost the $20 bill. The students' responses are summarized in the following table:

	Lost $20 bill	**Lost ticket**	**Total**
Yes, would pay to see play	22	11	33
No, would not pay to see play	5	17	22
Total	27	28	55

(a) Identify the observational units and variables in this study.
(b) Is this an experiment or an observational study? Explain.

(c) Produce and describe a segmented bar graph to display these data.
(d) Calculate and interpret the relative risk and odds ratio of paying to see the play between these two groups from these data.
(e) How likely is it for random variation to produce experimental data as favorable to the instructor's conjecture as these, if in fact the question wording had no effect on people's answers? Explain.
(f) Does the design of the study allow you to draw a cause-and-effect conclusion between the wording of the question and people's responses? Explain.
(g) To what population would you be willing to generalize these results? Explain.

64. Exploring the Hypergeometric Distribution

Recall that in Investigation 2.4 you randomly assigned 24 subjects (8 women and 4 men) to two treatment groups, A and B, of 12 subjects each. In that Investigation, you used an applet to repeat the random assignment a large number of times. The point was to see how well random assignment balances out various factors (such as gender and height) between the two treatment groups. Now you can use hypergeometric probabilities to analyze what would happen in the long run.

(a) How many different ways are there to make the random assignment of 124 people into two groups of 12 each? [*Hint*: This is the same as the number of ways of choosing 12 people from a group of 24 people.]
(b) Let the random variable X represent the number of women in group A. What are the possible values for X?
(c) Calculate the hypergeometric probability of obtaining 4 women and 8 men in group A, P(X = 4).
(d) Let Y represent the number of men in group A. What is the hypergeometric probability of obtaining 8 men and 4 women in group A, P(Y = 8)?
(e) Returning to the number of women in group A, calculate the hypergeometric probability for each possible value of X – the number of women in group A – specified in (b). [*Hint*: You may want to use technology to carry out the calculations.]
(f) Make a graph of this probability distribution and calculate the expected number of women in group A, E(X).
(g) Based on your probability distribution, is it more likely that the genders will be equally divided between the two groups or that they will not? Explain.
(h) Determine the probability that randomization puts all of the women in one group. How surprising is this even? [*Hint*: Either group A or group B could be the group with all women.]
(i) How many men would you need to see in group A before you would start to become skeptical that the outcome arose by chance alone? [*Hint*: For what value of y is $P(Y \geq y) \leq 0.05$?]

65. 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.

66. Friendly Observers

A study published in the *Journal of Personality and Social Psychology* (Butler & Baumeister, 1998), investigated a conjecture that having an observer with a vested interest would decrease subjects' performance on a skill-based task. Subjects were given time to practice playing a video game that required them to navigate an obstacle course as quickly as possible. They were then told to play the game one final time with an observer present. Subjects were randomly assigned to one of two groups. One group (A) was told that the participant and observer would each win $3 if the participant beat a certain threshold time, and the other group (B) was told only that the participant would win the prize if the threshold were beaten. The threshold was chosen to be a time that they beat in 30% of their practice turns. The following results are very similar to those found in the experiment: 3 of the 12 subjects in group A beat the threshold, and 8 of 12 subjects in group B achieved success.
(a) Use a simulation to approximate the p-value in this situation.
(b) Use Fisher's Exact Test to calculate the exact p-value.
You should have determined that the exact p-value is 0.0498. Now you will determine how many repetitions of a simulation are necessary to estimate the exact p-value with a desired level of accuracy and precision.
(c) Suppose we are trying to estimate the true p-value 0.0498, and we use simulation to obtain an empirical p-value (using the proportion of random assignments in the simulation that produced 3 or fewer successes in group A). Of course, if we performed another repetition of the simulation we would likely obtain a different empirical p-value. We can think of these empirical p-values as sample proportions and model their distribution. According to the Central Limit Theorem, in 100 repetitions of the simulation, what is the standard deviation of these empirical p-values? (Comment on shape, center, and spread, and draw a well-labeled sketch of the sampling distribution.)
(d) Would it be surprising for the empirical p-value based on 100 repetitions to be 0.07 or larger? Explain. [*Hint*: How many standard deviations away from0 .0498 is 0.07?]
(e) Would it be surprising for the empirical p-value based on 100 repetitions to be .14? Explain.
(f) In 1000 repetitions of the simulation, what is the standard deviation of the empirical p-values?
(g) Would it be surprising for the empirical p-value based on 1000 repetitions to be 0.07? Explain.
(h) How many repetitions of the simulation should you perform in order to approximate the true p-value to within ± 0.005 with 99% confidence?

67. Comparing p-values.

In the Minority Baseball Coaches study (Exercise 16), the p-value was 0.015, and in the Friendly Observer's study (Exercise 66), the p-value was 0.0498.
(a) Which study provides stronger evidence that the difference between the samples was larger than we would expect by chance? Explain.
(b) Which study provides stronger evidence that the difference in the response was caused by the explanatory variable? Explain.

68. Gender Candy Choices

Recall from the Chapter 1 exercises the study of whether children prefer fruit-flavored candy (high in sugar) or chocolate candy (high in sugar and fat) for Halloween treats. In addition to recording which candy each child chose, the students conducting the study recorded the gender of the child and whether or not the child was with an adult chaperone. The students were interested in whether the proportion choosing chocolate differed significantly between girls and boys or between chaperoned and non-chaperoned children.

(a) State the null and alternative hypotheses, in symbols and in words, for testing whether girls and boys prefer the chocolate candy to the fruit-flavored equally often.
(b) State the null and alternative hypotheses, in symbols and in words, for testing whether children accompanied by a chaperone are more likely to choose the fruit-flavored candy.

It turned out that 63 of 104 boys and 64 of 87 girls chose the chocolate candy.

(c) Calculate conditional proportions and construct a segmented bar graph for comparing the candy choices of boys and girls. Comment on what these numerical and graphical summaries reveal.
(d) Conduct the appropriate test, first checking the technical conditions and then calculating 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. Is the difference significant at the 0.05 level? How about the 0.02 level?
(e) Calculate and interpret a 90% confidence interval for estimating the difference in population proportions who prefer chocolate candy.
(f) Explain what is meant by "90% confidence" in (e).

Of the 82 children who were with an adult, 56 chose the chocolate candy. Of the 109 children who were not with an adult, 71 chose the chocolate candy.

(g) Analyze these data with numerical and graphical summaries, with a significance test (including stating the hypotheses and checking the technical conditions), and with a confidence interval. Interpret your results and summarize your findings (including discussion of the numerical and graphical summaries).

69. Flu Vaccines

In January of 2004, the Centers for Disease Control analyzed preliminary data on the effectiveness of a flu vaccine given to workers at Children's Hospital in Denver, Colorado. The hospital sent an anonymous survey to approximately 3100 hospital workers, and 1866 responded. From these 1866 responses, 1818 were included in the study, after some were eliminated for not responding to all of the questions. Of the 1818 hospital workers in the study, 1009 had opted to receive the vaccine before November 1, and an additional 425 had opted to receive the vaccine on or after November 1, leaving 402 who opted not to receive the vaccine. The 425 who received the vaccine on or after November 1 were also excluded, because it was not clear that the vaccine would have had time to prove effective by December 17, when the study results were compiled for analysis. Of the 1009 who had been vaccinated before November 1, 149 experienced flu-like symptoms by December 17, compared to 69 of the 402 who had not received the vaccine.

(a) Is this an observational study or a controlled experiment? Explain.
(b) Identify the explanatory and response variables, and classify them as categorical or quantitative.
(c) Organize the data in a two-way table, calculate conditional proportions, and construct a segmented bar graph. Comment on what this descriptive analysis reveals about this sample.
(d) Conduct an appropriate significance test of whether the rate of experiencing flu-like symptoms is significantly lower in the sample that received the vaccine than in the sample that did not. Identify the procedure that you use, check the technical conditions, state the hypotheses in words and symbols, sketch the randomization distribution, calculate the test statistic and p-value, and summarize your conclusions. (Show the details of your calculations, or supply relevant computer output.)
(e) Is the difference in observed proportions statistically significant at the 0.10 level?
(f) If the difference in observed proportions is statistically significant, would you be able to draw a causal connection between the vaccine and a lower rate of flu-like symptoms? Explain.

70. Flu Vaccines (cont.)
Reconsider the previous question about the study on the effectiveness of a flu vaccine.
(a) Construct a 90% confidence interval for the difference in rates of experiencing flu-like symptoms between the two populations.
(b) Does this interval include the value 0? Explain why this is not surprising, in light of the p-value from the previous question.
(c) Construct a 90% confidence interval for the odds ratio of experiencing flu-like symptoms between the two populations.
(d) Does this interval include the value 1? Explain why this is not surprising, in light of the p-value from the previous question and the confidence interval in (b).

71. Smoking Abstinence
A study published in the *Journal of Clinical Pharmacological Therapy* (2001) involved 450 smokers who had previously attempted to quit smoking with the drug bupropion SR. The subjects were randomly assigned to receive either the drug bupropion SR (again) or a placebo. The subjects' goal was to abstain from smoking through weeks 4 through 7 of the study. It turned out that 61 of 226 subjects in the treatment group met this goal, compared to 11 of 224 in the placebo group.
(a) Is this an observational study or a controlled experiment? Explain.
(b) Identify the explanatory and response variables, and classify them as categorical or quantitative.
(c) Organize the data in a two-way table, calculate conditional proportions, and construct a segmented bar graph. Comment on what this descriptive analysis reveals.
(d) Conduct an appropriate test of whether the success rate for abstaining from smoking is significantly higher in the treatment group than in the placebo group. Check the technical conditions, state the hypotheses in symbol(s) and words, sketch the sampling distribution, calculate the test statistic and p-value, and summarize your conclusions. (Show the details of your calculations, or supply relevant computer output.)
(e) Is the difference in observed proportions of success statistically significant at the .01 level?
(f) Construct and interpret a 90% confidence interval for the treatment effect.
(g) Construct and interpret a 99% confidence interval for the treatment effect.
(h) Are you justified in drawing a causal connection between the drug and the higher rate of smoking abstinence in this study? Explain.

72. Smoking Abstinence (cont.)
Reconsider the study described in the previous question. A second response variable considered was successful abstinence from smoking through six months. Of the 226 subjects in the treatment group, 27 met this stricter measure of success, compared to 5 of 224 in the placebo group.
(a) Test whether these data provide significant evidence that the treatment is more effective than the placebo. (Include all steps of a test of significance.)
(b) Construct and interpret a 95% confidence interval for the treatment effect. (Use whichever is more appropriate between the Wald or adjusted Wald techniques. Or use both and compare their results.)

73. 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) Is this an experiment or an observational study? Explain.
(b) Explain the importance of using the "placebo treatment" of saline in this study.
(c) Create the two-way table for summarizing these data, putting the explanatory variable in the columns and the response variable in the rows.
(d) Calculate the conditional proportions of pain reduction in the two groups. Display the results in a segmented bar graph. Comment on what this preliminary analysis reveals.
(e) If there were no association between the treatment and pack pain relief, about how many of the 11 "successes" would you expect to see in each group? Did the researchers observe more successes in the saline group than expected (if the drug had no effect) or fewer successes than expected? Is this in the direction conjectured by the researchers?
(f) Calculate the p-value from Fisher's Exact Test for this study.
(g) Check the technical conditions to determine whether it would be appropriate to conduct a two-sample z-test on these data.
(h) Regardless of your answer to (g), conduct the two-sample z-test. Report the hypotheses, test statistic, and p-value.
(i) Does this approximate p-value from the z-test come close to the exact p-value from Fisher's Exact Test? Does this support your conclusion in (g) about whether the two-sample z-test is appropriate for these data? Explain.
(j) Does the design of this study allow you to draw a cause-and-effect conclusion regarding botulinum and pain reduction? Explain.

74. Newspaper Credibility

A nationwide sample of 1002 adults, 18 years or older, were interviewed via telephone (under the direction of Princeton Survey Research Associates) during the period May 6–16, 2002. One of the questions asked was:

> Please rate how much you think you can BELIVE each organization on a scale of 4 to 1. On this four-point scale, "4" means you can believe all or most of what the organization says. "1" means you believe almost nothing of what they say. How would you rate the believability of (USA *Today*, NPR, MSNBC, etc.) on this scale of 4 to 1?

In this exercise, we will focus on people's responses to the credibility issue for "The daily newspaper you are most familiar with." About 6–7% of respondents said they were not able to rate their daily newspaper. Of the 932 respondents who were able to rate the daily newspaper they were most familiar with, 591 rated the paper as "largely believable" (a 3 or 4 on the scale). When the same question was asked 4 years earlier (May 7–13, 1998), 922 said they could rate their daily paper and of those, 618 rated the paper as "largely believable."

(a) Report the hypotheses, in symbols and in words, for this test.
(b) Report the one-sided p-value that technology produces for this comparison.
(c) Using the $\alpha = 0.05$ significance level as the standard, would you reject the null hypothesis and conclude that newspaper credibility declined between these two years?
(d) Suppose that one person in the 2002 sample changed his/her answer from "largely believable" to "not largely believable." Repeat (b) and (c) for these data.
(e) Did this one person's change affect the p-value much? Did it affect the test decision at the $\alpha = 0.05$ level?
(f) Explain how this exercise reinforces the idea that reporting a p-value is more informative than simply reporting a "reject or not" decision at a given significance level.

75. U.S. Volunteerism
In the September 2003 study of volunteerism in the U.S. conducted by the Bureau of Labor Statistics, 25.1% of men and 32.2% of women surveyed said that they had done volunteer work for or through an organization in the previous year.
(a) Construct and comment on an appropriate graphical display for comparing the sample results between the two groups.
(b) Suppose that the sample sizes had been 30,000 in each gender (half of the roughly 60,000 sampled monthly in the Current Population Survey) and that the samples had been drawn independently. Conduct a test of whether the difference in sample proportions is statistically significant. (Like always, state the hypotheses in words and symbols, check the technical conditions, sketch the sampling distribution, calculate the test statistic and p-value, and summarize your conclusions. Show the details of your calculations, or supply relevant computer output.)
(c) Construct and interpret a 99% confidence interval for the difference in population proportions (subtracting the men's rate from the women's rate) who did volunteer work in that year.
(d) If you had subtracted in the other order (men's rate minus women's rate), how would the interval have changed?

76. U.S. Volunteerism (cont.)
Reconsider the previous question and the study about volunteerism.
(a) Suppose that the sample sizes had been the same for men and for women. Determine the smallest sample size so that the difference between the observed sample proportions of 0.251 and 0.322 would be significant at the 0.05 level (with a two-sided test).
(b) For the sample sizes that you found in (a), would the same difference in sample proportions (0.071) have been significant if the sample proportions had been 0.451 and 0.522? Report the test statistic and p-value in this case.
(c) Repeat (b) with the same difference in sample proportions (0.071), but assuming that the sample proportions had been 0.051 and 0.122.
(d) Summarize your findings from this analysis.

77. U.S. Volunteerism (cont.)
Reconsider the volunteerism study described in the previous exercise. Consider a significance test of H_0: $\pi_f - \pi_m = 0.06$ vs. H_a: $\pi_f - \pi_m \neq 0.06$.
(a) Explain in words what these hypotheses say.
(b) Conduct this significance test. Report the test statistic and p-value. Would you reject the null hypothesis at the $\alpha = 0.05$ level? (*Hint*: Subtract 0.06 rather than 0 in the numerator of the test statistic, and do not use the pooled estimate in the denominator.)
(c) Repeat for testing the hypotheses H_0: $\pi_f - \pi_m = 0.07$ vs. H_a: $\pi_f - \pi_m \neq 0.07$.
(d) Repeat for testing the hypotheses H_0: $\pi_f - \pi_m = 0.05$ vs. H_a: $\pi_f - \pi_m \neq 0.05$.

78. U.S. Volunteerism (cont.)
Reconsider the volunteerism study described in previous exercises. Suppose that the sample sizes had been 30,000 men and 30,000 women. Suppose that the sample proportions of volunteerism had been 0.29 and 0.30.
(a) Is this difference in sample proportions statistically significant at the $\alpha = 0.01$ level? Justify your answer with an appropriate test of significance, reporting the test statistic and p-value.
(b) Determine a 95% confidence interval for the difference in proportions volunteering between the two populations.

(c) Would you describe this difference as *statistically* significant but not *practically* significant? Explain your answer, including an explanation of what these phrases mean.

79. Low Birth Weights

The 2002 birth weight data reported in the National Vital Statistics Reports revealed that 6.9% of 2,298,156 births to non-Hispanic white women were of low birth weight, compared to 13.4% of 578,335 births to non-Hispanic black women.

(a) Construct a 99% confidence interval for the odds ratio of low birth weight births for black women compared to white women.
(b) Explain why the confidence interval turns out to be so narrow.
(c) Does the interval include the value 1, or is it entirely above 1, or is it entirely below 1? Explain what this reveals.
(d) Does the interval include the value 2? Explain what this reveals.

80. Graduate Admissions Discrimination

The University of California at Berkeley was charged in a 1973 lawsuit with having discriminated against women in their graduate admissions process for the fall quarter of 1973. The data revealed that 533 of 1198 male applicants had been accepted, whereas 113 of 449 female applicants had been accepted.

(a) Identify the explanatory and response variables for this study.
(b) Conduct a two-proportion *z*-test of whether the observed difference in sample proportions is statistically significant. Use the 0.001 significance level. (Feel free to use technology.) Report the hypotheses, check technical conditions, sketch the sampling distribution, calculate test statistic and p-value, and state your conclusion.
(c) Based on your answers to (b), is it reasonable to conclude that being male caused an applicant to have a higher probability of being accepted? Explain.
(d) The table below classifies the same applicants according to which program they applied to. For each program, determine the proportion of male applicants who were accepted and also the proportion of female applicants who were accepted. Is this consistent to what you found in (b)?

	Men		**Women**	
	Accepted	**Denied**	**Accepted**	**Denied**
Program A	511	314	89	19
Program F	22	351	24	317
Total	533	665	113	336

You should have discovered an apparent paradox in these data: Men have a significantly higher proportion who were accepted than women overall, yet within each program the acceptance rates are quite similar between men and women. In fact, women have a higher acceptance rate than men in most programs, and the difference in acceptance rates is quite small for the programs that had a higher acceptance rate for men.

(e) Write a paragraph that uses the data provided to explain why the apparent "paradox" occurs, as if to a judge or jury struggling to understand what these data reveal. [*Hint*: Re-examine carefully the table in (d).]
(f) Explain how the idea of a confounding variable applies to these data.

81. A Nurse Accused

Statistical evidence played an important role in the murder trial involving Kristen Gilbert, a nurse who was accused of murdering hospital patients by giving them fatal doses of heart stimulant. (See for example Cobb & Gehlbach, 2004.) The following table summarizes eighteen months of data, examining whether Gilbert was working during an eight-hour shift and whether a death occurred on the shift:

	Gilbert on shift	Gilbert not on shift	Total
Death occurred on shift	40	34	74
No death occurred on shift	217	1350	1567
Total	257	1384	1641

(a) Identify the observational/experimental units.
(b) Identify the explanatory and response variables. Classify each variable as categorical or quantitative.
(c) Is this an observational study or an experiment? Explain.
(d) Provide numerical and graphical summaries of these results. Describe what these summaries reveal.
(e) Is a two-sample *z*-procedure appropriate for these data? Explain.
(f) Obtain a hypothetical p-value for these data. Interpret this p-value in the context of this study. What conclusion do you feel the jurors should have come to, based on this p-value? Explain.
(g) Suggest a potential confounding variable that the defense attorney could use to explain why there was a significantly higher proportion of deaths on Gilbert's shifts (and explain the confounding).

82. Civil Action

The novel and movie *A Civil Action* is based on a case of suspected contamination of municipal wells servicing Woburn, Massachusetts. Shortly after the contaminated wells were discovered, the town was found to have an elevated rate of childhood leukemia. According to one source, there were 16 birth defects out of 414 births when the contaminated wells were being used, and 3 birth defects out of 228 births when the contaminated wells were not being used. Carry out all steps of a test of significance to determine whether there is evidence that the rate of birth defects was higher while the polluted wells were in use. [*Hint*: Be very clear and careful in how you define the population parameter(s).] Also clearly state what conclusion you would come to, based on these data.

83. Gay Marriage

A poll conducted March 6–8, 2004, by *The Wall Street Journal/NBC News* asked 1,018 respondents their opinions about gay marriage. When asked to state whether they would favor or oppose "a constitutional amendment making it illegal for gay couples to marry," 43% responded in favor and 52% opposed (5% were unsure). When asked whether they would favor or oppose "a constitutional amendment that defined marriage as a union between a man and a woman and made same-sex marriages unconstitutional," 54% favored the amendment, 42% opposed (1% said it depends, and 3% were not sure).
(a) Produce, include, and comment on a graphical summary of these sample results.
(b) Suppose these were independent random samples of 1,018 respondents each, is there a statistically significant difference in the percentage favoring the amendment between these two phrasings of the question? (Include a statement of hypotheses in symbols and in words, a discussion of the technical conditions, a sketch of the sampling distribution, the test statistic and p-value, and your conclusion in context.)
(c) Calculate and interpret a 95% confidence interval for the difference in the population proportion favoring the amendment with these two wordings.
(d) Write a paragraph summarizing the results of your analysis being sure to comment on the scope of conclusions you can draw (Can you generalize these results to a larger population? Can you draw a cause and effect conclusion? Justify your decisions.).
(e) Explain why the analyses carried out in (b) and (c) are actually not valid.

84. Penny Thoughts

Students in a statistics class were asked whether they would vote to retain or abolish the penny. Of 10 men, 6 voted to retain the penny. Of 15 women, 14 voted to retain the penny. Treat these as independent samples of male statistics students and female statistics students at this university.

(a) Calculate an approximate 95% confidence interval (without the Wilson adjustment) for the difference in the proportion of male students at this university who would vote to retain the penny and the proportion of female students at this university who would vote to retain the penny.
(b) Are the technical conditions for this procedure satisfied? Explain.
(c) Repeat (a) using the Wilson adjustment. How do the interval compare?
(d) Why should you be cautious about interpreting this interval for the population of all students at this university?

85. Comparing Success Rates

Suppose that a researcher comes to you and says that she has categorical data for which she wants to compare the success rates between two groups. She tells you that in group A, 60% of the subjects succeeded in the task, and in group B 40% of the subjects succeeded in the task.

(a) What information would you request from her in order to determine if this difference is statistically significant? Explain briefly.
(b) What further information would you request from her in order to determine if this difference provides evidence of a cause-and-effect connection between group membership and success? Explain briefly.

86. Teaching Morals

Lee et al. (2014) examined whether some classic moral stories actually influence whether or not kids lie. They examined the stories of "Pinocchio," "The Boy Who Cried Wolf," and "George Washington and the Cherry Tree," as stories commonly used by teachers and parents to promote honesty, though in different ways (negative consequences of lying for the first two vs. positive consequences of truth telling in the third). Two hundred and sixty-eight Canadian children aged 3–7 years were recruited for the study (children begin to tell lies around 2–3 years of age). Children participated in a "temptation-resistance task" that has been used widely to study whether children choose to lie to hide a transgression (essentially peeking at the answer when left alone in the room for one minute) and then were read one of the three stories or a control story – "The Tortoise and the Hare"). (The reader did not know whether or not they had peeked.) After the story, the child was asked whether or not he or she had peeked.

Suppose the results for the children who peeked turned out like this:

	Tortoise and Hare (control)	**George Washington**	**Pinocchio**	**Boy Who Cried Wolf**	**Total**
Confessed	20	22	13	16	71
Did not (lied)	44	22	31	30	127
Total	64	44	44	46	198

(a) Were the children who heard "George Washington" significantly more likely to confess than the control group?
(b) Were the children hearing "The Boy Who Cried Wolf" significantly more likely to confess than the control group?
(c) What are the implications of comparing your results to (a) and (b)? What follow-up study would this suggest?
(d) The age distribution of children in each of the four conditions was similar. Why is that an important consideration?

The following three exercises review material from Chapter 1.

87. 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.

88. Freshmen Voting Patterns (cont.)

Recall the student project discussed in Investigation 1.15.

(a) When the sample size *n* is larger, we saw in Chapter 1, that the binomial distribution can be approximated by the normal distribution. Would you consider this approximation valid for this study? Explain.
(b) Regardless of your answer to (a), 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.
(c) Does the normal approximation to the binomial appear to be valid for this study? Explain.
(d) Regardless of your answers to (c), carry out a one-sample *z*-test to test the students' conjecture . Compare the hypergeometric, binomial, and normal probabilities.

89. Lady Tasting Tea

In Exercise 13 of Chapter 1, we introduced you to Fisher's famous "Lady Tasting Tea" experiment, where Muriel Bristol claimed she could tell whether the milk was added to the tea or the milk was first added to the cup of tea. Fisher (1935) describes a potential experiment to test this claim: The lady would be given 8 cups, 4 with the milk added first and 4 with the tea added first. The order of the cups would be randomized and she would be asked to determine which was which. According to Salsburg (2001), a colleague, Hugh Smith, was at the part, and claims the lady identified all 8 correctly. So the two-way table is:

	Tea first	Milk first	Total
Lady claims tea first	4	0	4
Lady claims milk first	0	4	4
Total	4	4	8

(a) Is this study an experiment or an observational study? Explain your reasoning.
(b) If we let X represent the number of tea-first cups that she correctly identifies, explain why this is *not* a binomial random variable.
(c) How many different possible ways are there to arrange the cups at random?
(d) Use Fisher's Exact Test to calculate the exact p-value. Show your work and/or output.
(e) What would you conclude about the lady's ability to distinguish which "infusion" was used?

Additional Problems with Simpson's Paradox

Exercise 80 provides an example of *Simpson's Paradox*: Although men were accepted at a higher rate than women overall, when the data were disaggregated by a third variable, program, the direction of the association actually reversed. The "paradox" arose because program was related to both of the original two variables, program and gender. The following exercises also pertain to Simpson's Paradox.

90. Kidney Surgeries

A recent study (reported in Julious & Mullee, 1994) compared two types of procedures for removing kidney stones: open surgery and percutaneous nephrolithotomy (PN), a "keyhole" surgery that removes the stone through the skin, designed to have much less disturbance than an open operation. In this study,

- For stones less than 2 cm, 81 of 87 cases of open surgery were successful compared to 234 of 270 cases of PN.
- For stone at least 2 cm, 192 of 263 cases of open surgery were successful compared to 55 of 80 cases of PN.

Is there evidence of Simpson's paradox (see definition above) with this data? Include calculations to support your answer. If so, explain to someone not in a statistic class what is giving rise to the paradox in this study.

91. Paradoxical Movies

Moore (2005) gathered data on a sample of 100 movies taken from Leonard Maltin's *Movie and Video Guide (1996)*. He classified the movies according to three variables:

- Length: whether it is short (< 90 minutes) or long
- Age: whether it is old (made in 1965 or earlier) or new
- Quality: whether Maltin rated it as good (rating of 3 or higher, out of a possible 5) or bad

The following tables summarize the data:

Short movies

	Old	**New**
Good	6	0
Bad	29	7
Total	35	7

Long movies

	Old	**New**
Good	9	16
Bad	6	27
Total	15	43

(a) From these tables, create the two-way table that summarizes the results for age of movie and quality of movie (combine the tables across the length categories and use the age variable in the columns).
(b) From this new table, calculate the conditional proportion of old movies that are good and the conditional proportion of new movies that are good. Which is higher?
(c) Verify that Simpson's paradox occurs here: Show that for both long movies and short movies, old movies have a higher proportion of good ones than new movies do, yet for the two lengths combined, this relationship is reversed. (Show the calculations that verify this.)
(d) Explain in a few sentences why the paradox occurs here, arguing from the data given and relating your explanation to the context.

92. Simpson's Paradox

Reconsider the Simpson's paradox phenomenon.

(a) Construct a hypothetical example to show that Simpson's paradox could arise in comparing survival percentages between two hospitals by filling in the following tables so that hospital X has a higher survival rate than hospital Y for both patient conditions (critical and stable), yet hospital Y has a higher survival rate overall. [*Hint*: Make sure that the overall counts are the sum of the critical and stable patient counts for all cells of the tables.]

Overall	Hosp. X	Hosp. Y
Survive	80	90
Not	20	10
Total	100	100

Critical condition	Hosp. X	Hosp. Y
Survive		
Not		
Total	80	10

Stable condition	Hosp. X	Hosp. Y
Survive		
Not		
Total	20	90

(b) Write a few sentences explaining why the apparent paradox happens in this setting.
(c) Which hospital would you rather go to? Explain.

93. Simpson's Paradox (cont.)
Construct your own hypothetical data to illustrate Simpson's paradox in the following context. Show that it is possible for one softball player (Alex) to have a higher proportion of hits than another (Bob) in July and in August, and yet Alex can have a lower proportion of hits for the two months combined. [*Hints*: You might want to give each player the same number of at-bats (maybe 200) for the two months combined, and you may want to use Excel to help you automatically update the calculations as you try different numbers. Try to make the differences in the players' proportions of hits (number of hits divided by the number of at-bats) as large as you can (don't worry about these proportions being realistic)]. Also provide a brief explanation that explains the apparent paradox, as if to a baseball fan with limited knowledge of statistics, for the example that you construct.

94. Simpson's Paradox (cont.)
The following data report the number of flights that were "on time" and "not on time" for Continental Airlines and American Airlines in November 2002 for all flights to Houston, Chicago, and Los Angeles.

	Houston		Chicago		Los Angeles	
	On-time	**Late**	**On-time**	**Late**	**On-time**	**Late**
Continental	7318	1017	466	135	544	145
American	598	70	8330	1755	2707	566

(a) Calculate the on-time arrival for Continental for each city. Do the same for American. Which airline has a higher proportion of on-time arrivals in these cities?
(b) Calculate the overall on-time arrival rate for each airline. Is Simpson's paradox present with these data? Explain.
(c) Suggest an explanation for the cause of the paradox – what is the lurking variable here?
(d) Based on these data, if you were to fly to one of these three cities, which airline would you rather fly on? Explain.