

SAMPLING

Note. You can put your answers to Questions I and II directly on these pages if wish. Put your answers to the remaining problems on a separate page. ***Your written answers should be brief, for both your sake and mine.***

- I. Researchers want to learn about the political views of students at the State University campus with 15,000 students. Since it is not feasible to interview all students, they draw a *simple random sample* of 1,000. Then 838 of these students are successfully interviewed. (The remainder could not be located, refused to be interviewed, their responses were lost, etc.) Of the students interviewed, 407 reported that they approved of President Bush's performance as President and 431 reported that they disapproved.
- (a) The probability that a given student appears in the drawn sample is _____
- (b) The (drawn) *sampling fraction* is _____
- (c) The *completion* or (*response*) rate is _____
- (d) The *margin of* (sampling) *error* is approximately (check Table 3.4 in the Weisberg book or use the approximate formula given in handout)

- (e) Can we conclude with a high degree (e.g., 95% or better) of confidence that more State U. students disapprove of the President than approve? Why or why not?
- II. You have been given a spreadsheet with the raw data from the Spring and Fall 2009 Student Political Attitudes Surveys. Consider the 67 respondents to be the *population*. Note that the cases are (arbitrarily) numbered 1 through 67. Using this data, determine the value of the *population parameter* PERCENT OF STUDENTS WHO WATCHED MOST OR ALL OF TELEVISED PRESIDENTIAL DEBATES (Question 9). What is the value of this population parameter?

Population parameter = _____%

Now select a simple random sample of size $n = 10$ from this the population of $N = 67$. I suggest that you do this by going to the POLI 300 course web page; click on the link to *Statistical Applets*, and select Simple Random Sample. Set the *population hopper* to $N = 67$ by deleting the default figure of 100, typing in 67, and then clicking on the Reset button. Then set the *sample size* to $n = 10$ (already the default size) and click on the Sample button. Now calculate

the *sample statistic* PERCENT OF STUDENTS WHO WATCHED MOST OR ALL OF TELEVISED PRESIDENTIAL DEBATES for the sample of 10 students whose ID numbers appear in the *sample bin*. Repeat this operation for a total of ten samples of size $n = 10$. For each subsequent sample, click on **Reset** button and then on the **Sample** button. What further values for the sample statistic do you come up with, i.e., what percent watched most or all debates in each of the 10 samples?

Sample statistic #1 _____	Sample statistic #6 _____
Sample statistic #2 _____	Sample statistic #7 _____
Sample statistic #3 _____	Sample statistic #8 _____
Sample statistic #4 _____	Sample statistic #9 _____
Sample statistic #5 _____	Sample statistic #10 _____

(If for some reason you cannot use this sampling procedure, use the Table of Random Numbers that appears on the back page of Handout #2 on Random Sampling. This will be considerably more tedious, so you may want to take fewer than ten separate samples.)

Comment on the result of your sampling? Did you have to decide whether each of your 10 samples were *with replacement* or *without replacement*? Can you tell whether the Simple Random Sample applet samples with or without replacement? Using the formula given in the handout on Random Sampling, what is the approximate margin of error associated with each of your samples? Does this margin of error appear reasonable in light of the actual sample statistics you produced? Given that you have taken ten samples of size $n = 10$, they can be *pooled* together into a single larger sample of $n = 100$. What is the value of the sample statistic in this pooled sample and what is its margin or error? (Notice, incidentally, that your pooled sample of $n = 100$ is actually larger than the population of $N = 67$; some cases necessarily appear in the pooled sample several times; at the same time, there is no assurance that every case appears in the pooled sample even once, and such pooled samples therefore are still subject to sampling error.)

Note. Determine the margin of error (ME) for your samples of size $n = 10$ and $n = 100$ using the formula $ME = 100\%/\sqrt{n}$. The actual margin of error for your samples is smaller than the margin given by this formula. Can you explain why?

- III. **Note.** The following problems are taken or adapted from David S. Moore, *Statistics: Concepts and Controversies*, a text book previously used in this course. You need not provide long and detailed answers — just note the basic point(s) in a few words or couple of sentences.
1. You are on the staff of a member of Congress who is considering a bill that would provide government-sponsored insurance for nursing home care. You report that 1128 letters have been received on the issue, of which 871 oppose the legislation. “I’m surprised that most voters in my district oppose the bill. I thought it would be quite popular,” says the congresswoman. Are you convinced that a majority of the voters oppose the bill?
 2. Highway planners decided to make a main street in West Lafayette, Indiana, a one-way street. The *Lafayette Journal and Courier* took a one-day poll by inviting readers to call a telephone number to record their comments. The next day, the paper reported:

Journal and Courier readers overwhelmingly prefer two-way traffic flow in West Lafayette's Village area to one-way streets. By nearly a 7-1 margin, callers to the newspaper's Express Yourself opinion line on Wednesday complained about the one-way streets that have been in place since May. Of the 98 comments received, all but 14 said no to one-way.

 - (a) What population do you think the newspaper wants information about?
 - (b) Is the proportion of this population who favor one-way streets almost certainly larger or smaller than the proportion 14/98 in the sample? Why or why not?
 3. “Should the United Nations continue to have its headquarters in the United States?” A television program asked its viewers to call in with their opinions on that question. There were 186,000 callers, 67% of whom said “No.” A nationwide random sample of 500 adults found that 72% answered “Yes” to the same question. Explain to someone who knows no statistics why the opinions of only 500 randomly chosen respondents are a better guide to what all Americans think than the opinions of 186,000 callers.
 4. A university has 1000 male and 500 female faculty members. A survey of faculty opinion selects 100 of the 1000 men at random and then separately selects 50 of the 500 women at random. The 150 faculty members chosen make up the sample.
 - (a) Explain why this sampling method gives each member of the faculty an equal chance to be chosen.
 - (b) Nonetheless, this is not an SRS. Why not?
 5. A flour company want to know what fraction of Minneapolis households bake some or all of their own bread. The company selects a sample of 500 residential addresses in Minneapolis and sends interviewers to these addresses. The interviewers work during regular working hours on

- weekdays and interview only during those hours. What is the population parameter of interest? Would this procedure produce a sample statistic giving a good estimate of this parameter?
6. The Miami Police Department wants to know how African-American residents of Miami feel about police service. A sociologist prepares several questions about the police. The police department chooses an SRS of 300 mailing addresses in predominantly black neighborhoods and sends a uniformed black police officer to each address to ask the questions of an adult living there. What are the population and the sample? Why are the results likely to be biased even though the sample is an SRS?
7. For each **boldface number** in (a) through (d) indicate whether it is a parameter or a statistic. (For each boldface number x , write $x =$ population parameter or $x =$ sample statistic.)
- (a) The Bureau of Labor Statistics announces that last month it interviewed all members of the labor force in a sample of 50,000 households; **4.5%** of the people interviewed were unemployed.
- (b) A carload lot of ball bearings has an average diameter of **2.503** centimeters (cm). This is within the specifications for acceptance of the lot by the purchaser. The inspector happens to inspect 100 bearings from the lot with an average diameter of **2.515** cm. This is outside the specified limits, so the lot is mistakenly rejected.
- (c) A telemarketer in Los Angeles uses a random digit dialing device to dial residential phone numbers in that city at random. Of the first 100 numbers dialed, **43** are unlisted numbers. This is not surprising, because **52%** of all Los Angeles residential phones are unlisted.
- (d) Voter registration records show that **68%** of all voters in Indianapolis are registered as Republicans. To test a random digit dialing device, you use the device to call 150 randomly chosen residential telephones in Indianapolis. Of the registered voters contacted, **73%** are registered Republicans.
8. A poll of 1190 adults finds that 702 prefer balancing the budget over cutting taxes. The announced margin of error for this result is plus or minus 4 percentage points. While the news report does not explicitly give the confidence level, you can assume that the confidence level is 95%.
- (a) What is the value of the sample statistic for the percent of respondents who prefer balancing the budget?
- (b) Explain in words what the corresponding population parameter is in this setting.
- (c) Make a confidence statement about the value of the population parameter.

9. An agency of the federal government plans to take an SRS of residents in each state to estimate the percent of each state's population who are owners of real estate. The population of the states ranges from about 550,000 people in Wyoming to about 35 million in California.
- (a) Will sampling error differ substantially from state to state if an SRS of size 2000 is taken in each state? Explain your answer.
 - (b) Will sampling error differ substantially from state to state if an SRS of $1/10$ of 1 % (0.001) of the state's population is taken in each state? Explain your answer.
10. During the 1968 Presidential campaign, the segregationist candidate George Wallace regularly accused pollsters of underestimating his support. He would ask crowds at his rallies, "Have any of y'all ever been asked about this here election by Mr. Harris or Mr. Gallup?" The crowd shouted back "No" and "Never." The U.S. voting age population was then about 150 million, typical polls use random samples of about 1500, and about 20 polls were conducted during the course of the 1968 campaign. What was the approximate probability that a given person would be interviewed in one particular such poll? What was the probability that a given person would be interviewed at least once in the twenty or so polls conducted during the Fall 1968 campaign? Can you think of any reason (unrelated to their sample sizes) why "Mr. Harris and Mr. Gallup" (and other pollsters) may indeed have underestimated Wallace's voting support prior to the election.