Statistical Reasoning Name:

5.3 – Sample Surveys in the Real World

**HOW SAMPLE SURVEYS GO WRONG.**

**Sampling Frame:** the list of individuals from which we draw our sample

**Sampling Errors:** errors caused by the act of taking a sample – they cause sample results to be different from the results of a census

**▫ Random Sampling Error:** the deviation between the sample statistic and the population parameter caused by chance in selected in a random sample – the margin of error in a confidence interval includes random sampling error only

**▫ Undercoverage:** occurs when some groups in the population are left out of the process of choosing the sample

**Nonsampling Errors:** errors not related to the act of selecting a sample of the population – nonsampling errors can also be present in a census

**▫ Processing Errors:** mistakes in mechanical tasks such as doing arithmetic or entering responses into a computer

**▫ Response Errors:** occurs when someone gives an incorrect response/lies

 **▫ Nonresponse:** failure to obtain data from an individual selected for a sample - people can’t be reached or refuse to cooperate

**\*Undercoverage, nonresponse, and other practical difficulties can cause a large bias that aren’t covered by the margin of error\***

Another source of bias is the exact **wording of questions** – this happens when the wording of a question creates a bias or causes our sample to respond in a certain way

**SAMPLING METHODS.**

Previously, we have discussed simple random samples (SRS), convenience samples, and voluntary response samples. Convenience samples and voluntary response samples are often biased while simple random samples are typically fair and unbiased.

In the real world, obtaining a simple random sample is not always realistic. There are other types of samples that are often used which can also produce fair and unbiased samples.

**Stratified Random Samples:** the sampling frame is broken into groups by common characteristics called **strata**; a simple random sample from each strata is taken and combined to create the stratified sample

**Systematic samples:** samples members are selected from the sampling frame according to a random starting point but with a fixed periodic interval (selecting every 10th student on a list starting with student #3)

**Probability Samples:** samples chosen by chance – we must know which samples are possible and what chance/probability each possible sample has – all good samples are probability samples (stratified and systematic samples are examples)

Practice and Examples.

1) About 20% of the engineering students at a large university are women. The school plans to poll a sample of 200 engineering students about the quality of student life.

a) If an SRS of 200 is selected, about how many women do you expect to find in the sample?

b) If the polls wants to be able to report separately the opinions of male and female students, what type of sampling design would you suggest? Why?

2) Each of the following is a source of error in a sampling survey. Label each as a sampling error or a nonsampling error and explain your answers.

a) The telephone directory is used as a sampling frame.

b) The person cannot be contacted in five calls.

c) Interviewers choose people walking by on the sidewalk to interview.

3) Suppose you want to know the average amount of money spent by fans attending opening day for the Cleveland Indians baseball season. You get permission from the team’s management to conduct a survey at the stadium but they will not allow you to bother the fans in the club seating or box seat areas (the most expensive seating). Using a computer, you randomly select 500 seats from the rest of the stadium. During the game, you ask the fans in those seats how much they spent that day.

a) Provide a reason why this survey might yield a biased result.

b) Explain whether the reason you provided in (a) is a sampling error or nonsampling error.

4) A poll of 586 adults who used the Internet in the past week were asked whether “the Internet has made your life much better, somewhat better, somewhat worse, much worse, or has it not affected your life either way.” In all, 152 of the 586 said “much better.”

a) What is the population for this sample survey?

b) Use the quick method to find a margin of error. Then give a complete confidence statement for a conclusion about the population.

5) Determine which type of sampling method was used for each of the following.

a) Student organization looking to get signatures for a petition camp out in front of Class of 1950 Lecture Hall.

b) Select three students from a class to receive ice cream by putting all the students' names in a hat and picking out three names randomly.

c) Select three female students and three male students to receive ice cream by putting all the men's names in one hat and all the women's names in a different hat and picking out three names from each hat.

d) In Fall 1995, the BBC in Britain requested viewers to call the network and indicate their favorite poem.

e) Divide the class into four groups (freshman, sophomore, junior and senior) and take a random sample of two students from each group.

f) Priceline.com randomly e-mails a Customer Satisfaction Survey for certain transactions done on its site in which customers choose to either respond or not.

g) A researcher has a population of 100 third grade children from a local school district from which a sample of 25 children is to be selected. Each child’s name is put on a list, and each child is assigned a number from 1 to 100. Then the numbers 1 to 100 are written on separate pieces of paper and shuffled. Finally, the researcher picks 25 slips of paper and the numbers on the paper determine the 25 participants.

h) A sociologist conducts an opinion survey in a major city. Part of the research plan calls for describing and comparing the opinions of four different ethnic groups: African Americans, Asian Americans, European Americans, and Native Americans. For a total sample of 300, the researcher selects 75 participants from each of the four predetermined subgroups.

i) Every fifth person boarding a plane is searched thoroughly.

6) Which of the following are sources of sampling error and which are sources of nonsampling error? Explain your answer.

a) The subject lies about past drug use.

b) The typing error is made in recording the data.

c) Data are gathered by asking people to mail in a coupon printed in a newspaper.

d) A hot dog vendor is collecting data on where is the best location to sell hot dogs. The city licenses 4 locations. Jack starts at the site in the park that is the most packed.

7) A police officer is interested in the drug and alcohol use of all high school students. He does to his local high school to question some of the students.

a) Suppose the officer initially decides to question students by asking them face to face. If a student does not want to tell the officer the truth and simply declares himself drug free, what error is introduced into the study?

Nonresponse Undercoverage Response Error

b) .Suppose the officers grabs a random sample but only does so of the freshman students. What type of bias does this study suffer from?

Nonresponse Undercoverage Response Error

c )Suppose the officer distributes an anonymous survey to 100 random high school students but only receives 12 of the surveys back. What types of bias does this study suffer from?

Nonresponse Undercoverage Response Error

8) A survey of users of the Internet found that males outnumbered females by nearly 2 to 1. This was a surprise, because earlier surveys had put the ratio of men to women closer to 9 to 1. Later in the article, we find this information: “Detailed surveys were sent to more than 13,000 organizations on the Internet; 1,468 usable responses were received. According to Mr. Quarterman, the margin of error is 2.8 percent, with a confidence of 95 percent.

a) What was the response rate for this survey? (The response rate is the percent of the planned sample that responded.)

b) Use the quick method to estimate the margin of error of this survey. Is your result close to the 2.8% claimed?

c) Use your answers from (a) and (b) to construct the 95% confidence interval.

9) Does increasing the sample size increase or decrease the margin of error?

10) Does increasing the sample size narrow or widen the confidence interval?

12) A nationwide poll was taken of 1432 teenagers (ages 13-18). 630 of them said they have a TV in their room.

a) What is the sample proportion ($\hat{p}$)?

b) What is the margin of error?

c) What is the 95% confidence interval?

d) Interpret your 95% confidence interval.

13. A poll of 1070 teens aged 13-17 finds that 742 have received personal messages online from people they don’t know. The announced margin of error for this result is plus or minus three percentage points. A parents group is satisfied with 95% confidence of the survey but wants a smaller margin of error than ± 3%. How can we get a smaller margin of error, still with 95% confidence?

14) Suppose a 95% confidence interval is stated as $(0.315, 0.365)$. *(Think about working backwards and where the proportion is in relation to the two numbers in the interval and how that is related to the margin of error)*

a) What is the sample proportion?

b) What is the margin of error?

15) You are interested in the proportion of Stat 113 students that will end up with an A this semester. You survey 20 students enrolled in Stat 113. Identify the

a) population:

b) sample:

c) variable *(what is being studied)*:

16) The mean income of all subscribers to a particular magazine is **$26000**. We draw a random sample of 100 subscribers and find that their mean income is **$27300**. Identify the

a) population:

b) sample:

c) parameter:

d) statistic:

17) The average GPA for all female volleyball players in a particular college is **2.8**, and their mean height is **182cm**. Identify the parameter or statistic if there is any.

18) To reduce bias, you should: use random sampling or use a larger sample *(choose 1)*

19) To recue the variability, you should: use random sampling or use a larger sample *(choose 1)*

20) Out of 54 randomly selected patients of a local hospital who were surveyed, 49 reported that they were satisfied with the care they received. Construct and interpret a 95% confidence interval for the percentage of all patients satisfied with their care at the hospital.

21) Determine if each of the following is a population parameter or a sample statistic.

a) A survey of 1000 U.S. adults found that 40% think that the Internet is the best way to get news and information.

b) At a college, 90% of the members of the Board of Trustees approved the contract of the new president.

c) A survey of 733 small business owners found that 17% have a current job opening.

d) The 2182 students who accepted admission offers to Northwestern University in 2009 have an average SAT score of 1442. (Source: Northwestern University)