

**Chapter 3 – Surveys and Sampling****SECTION EXERCISES****SECTION 3.1****1.**

- a. False. Sampling error cannot be avoided, even when the sample is unbiased. Sampling error is always present when a sample statistic is used to estimate a population parameter.
- b. True.
- c. True.
- d. False. Randomization will match the characteristics in a way that is unbiased. We can't possibly think of all the characteristics that might be important or match our sample to the population on all of them.

**2.**

- a. False. The fraction isn't important. Although increasing the size of the sample reduces sampling error, it does not guarantee that the sample is representative of the population. If a sample is selected in a bad way (e.g., convenience), even a very large sample will be biased.
- b. False. Given modern methods, it is best to randomize.
- c. False. A sample selected properly is also representative. For example, a simple random sample is representative of the population.
- d. True.

**SECTION 3.2****3.**

- a. Population—Professional food preparers in the U.S.
- b. Sampling Frame—*Chef's Collaborative* membership listing.
- c. Parameter—Proportion of professional food preparers in the U.S. who believe that food safety has improved.
- d. Sampling method—Simple random sample (SRS).

**4.**

- a. Population—All frequent flyer customers of the airline.
- b. Sampling Frame—Airline database of frequent flier customers.
- c. Parameter—Proportion who plan to use one of the airline's new hubs in the next 6 months.
- d. Sampling method—Simple random sample (SRS).

**SECTION 3.3****5.**

- a. No. It would be nearly impossible to get exactly 500 males and 500 females by selecting people from India at random.
- b. Stratified sampling. This allows GfK Roper to divide the population of India into two groups based on gender (called strata) and then randomly select 500 individuals from each group. The resulting sample will consist of an equal number of males and females.

**6.**

- a. No. It would be nearly impossible to get exactly 50 students from each class by selecting them at random from the student body.
- b. Stratified sampling. This allows the Business students to divide the student body into four groups based on class (freshman, sophomore, junior and senior) and randomly select 50 students from each group (called strata). The resulting sample will consist of an equal number of students from each class.

7. The consumer advocacy group used systematic sampling. To ensure the sample is random, the consumer advocacy group should have started the systematic sampling by selecting the first member from the *Chef's Collaborative* list at random.
8. The airline used stratified sampling. The population of frequent fliers was divided into groups (strata) defined by tier level (silver, blue and red).
9.
  - a. Population—Human resources directors of Fortune 500 companies.
  - b. Parameter—Proportion who don't feel surveys intruded on their workday.
  - c. Sampling Frame—List of HR directors at Fortune 500 companies.
  - d. Sample—23% of HR directors don't feel that surveys intruded on their workday.
  - e. Method—Questionnaire mailed to all (nonrandom).
  - f. Bias—Nonresponse bias. Those that respond are obviously receptive to participating in surveys. Since who responds is related to the question itself, there is likely a difference between those who respond and those who don't respond to the survey with regard to the question.
10.
  - a. Population—unspecified.
  - b. Parameter—Proportion of the population who think businesses should pay for their employees' health insurance.
  - c. Sampling Frame—none since the population is unspecified.
  - d. Sample—Individuals who visited the website and responded.
  - e. Method—Voluntary response (no sampling method employed).
  - f. Bias—Voluntary response sample. Those who visit the website and respond may be predisposed to a particular answer. This is a potential source of bias.

#### SECTION 3.4

11.
  - a. The population of interest is professional food preparers in the U.S.
  - b. Members of the *Chef's Collaborative* who attended the recent symposium on "Food Safety in the 21<sup>st</sup> Century" that was held in Las Vegas.
  - c. The sampling frame is not necessarily representative of the entire group of food preparers. Those who attended the symposium may have different opinions from those who didn't. It is likely that those who attended the symposium have a special interest in the topic. Moreover, rather than selecting members to call in a random fashion, he started from the top of the list. The list was generated as members enrolled for the symposium; therefore those at the top of the list enrolled early (and perhaps were more enthusiastic about the issue). Consequently, there may be differences in opinion between members near the top of the list and those near the bottom. Finally, the script is biased (leading respondents to answer in a particular way) and may lead to an inflated estimate of the true population proportion who think food safety has improved.
12.
  - a. The population of interest is all frequent flier customers of the airline.
  - b. Customers who have recently registered for the "Win the trip to Miami" contest on the Internet.
  - c. The sampling frame may not be representative of all frequent fliers since those who are interested in a trip to Miami may be more likely to fly there. Also, only those with Internet access would have been able to register for the contest. Since customers without Internet access may have different flying preferences than those with access, there is potential bias due to undercoverage. The survey question is biased (encouraging customers to respond in a particular way) and may lead to an inflated estimate of the true proportion of airline customers who would consider traveling through the Miami hub.

- 13.
- a. *Question 1* seems appropriately worded, although using the phrase “state-of-the-art” may give the impression that a higher price could be charged for the service. By noting that the monthly cost for the service is less than what it would cost to have a daily cup of cappuccino, *Question 2* predisposes the respondent to agree that \$50 is a reasonable price. *Question 2* does not seem to be appropriately worded.
  - b. *Question 1* is more neutrally worded than *Question 2*. In addition, *Question 1* addresses the issue more directly: the willingness of customers to pay \$50 per month for the new service.
- 14.
- a. Both questions introduce bias by leading respondents to answer in a particular way. *Question 3* describes dial-up Internet connections as being “slow” in comparison with the high speed Internet service being offered and implies that speed affects the enjoyment of Web services. *Question 4* implies that high speed Internet service is important for children’s education so it leads respondents to say that they would subscribe to the service.
  - b. *Question 3* is more appropriately worded than *Question 4*. However, there is room for improvement. For example, it could be reworded as: “Would high speed access improve your experience using the Internet?”
- 15.
- a. True.
  - b. True
  - c. False. Measurement error refers to inaccurate responses. Sampling error refers to sample-to-sample variability and is always present when using a sample statistic to estimate a population parameter.
  - d. True
- 16.
- a. False. This is a voluntary response sample and will likely be biased. Typically respondents to these types of invitations have strong opinions about the issue and won’t be representative of the population.
  - b. False. Constructing a survey that has too many questions increases the time it takes for respondents to participate. A survey that is too long may reduce the response rate and introduce non-response bias.
  - c. False. A large sample does not ensure a valid survey. If the large sample is not representative of the population, the survey results will be biased.
  - d. True.
- 17.
- a. This is a multistage design, with a cluster sample in the first stage and a simple random sample in the second stage. Churches are treated as clusters of the population, and three churches (clusters) are selected. Then a random sample of members is selected from each church.
  - b. Each church (cluster) is assumed to be representative of the population. This may not be the case if they are different (e.g., in terms of prominent ethnicity of its members such as Italian, Irish, etc.). If each church is not representative of the population, then bias will be introduced at the cluster stage.
- 18.
- a. This is a multistage design involving cluster sampling and census. In the first stage, one day is selected at random. In the second stage, five boats (boats are treated as clusters) are selected within that day. In the final stage a census of the number and types of fish is taken for each boat.

- b. If the one day that is selected is not representative of all fishing days, that will introduce bias. If the five boats selected within that day are not representative in terms of the numbers and types of fish typically caught by all fishing boats, then that will also introduce bias.

## CHAPTER EXERCISES

### 19. Software licenses.

- a. This sample was a voluntary response, not a random sample.
- b. There is no confidence in the estimate sampled. Voluntary response samples are almost always biased, and so conclusions drawn from them are almost always wrong.

### 20. Drugs in baseball.

- a. This sample was a cluster sample, with the teams selected being the cluster. A cluster is recognized by a random sample selection of teams within all of baseball. A sample is then taken from the selected teams.
- b. It is a reasonable solution to the problem of randomly sampling players. You can sample an entire team at once relatively easily. You could select your random sample prior to showing up unannounced but it would be less efficient to search for the players randomly selected.

### 21. Gallup.

- a. The population of interest is all adults in the United States aged 18 and older.
- b. The sampling frame, a list of individuals from which the sample will be drawn, consists of U.S. adults with landline telephones, which are the only numbers available for a study like this.
- c. An increasing number within the population (e.g., many college students and others with mobile only service) don't have landline phones, which could create a bias.

### 22. Defining the survey.

- a. They are using a multistage design in which the countries selected are clusters, and then a random sample is drawn within each cluster. They don't specify how the random samples are taken.
- b. The difference in population size has no effect on the precision of estimates from these surveys. Only the sample size matters and the sample chosen should be representative of the entire population.

### 23. Alternative medicine.

- a. Population—All Consumer Union subscribers.
- b. Parameter—Proportion of Consumer Union subscribers who have used and benefited from alternative medicine.
- c. Sampling Frame—All Consumer Union subscribers.
- d. Sample—Subscribers who responded.
- e. Sampling method—Questionnaire to all subscribers.
- f. Bias—Nonresponse. Those who respond could have strong feelings about the topic and affect the results.

**24. Global warming.**

- a. Population—U.S. adults.
- b. Parameter—Proportion of sample who believe that global warming has already begun and the proportion of sample who think global warming will never happen.
- c. Sampling Frame—U.S. adults aged 18 and over.
- d. Sample—1012 randomly selected adults.
- e. Sampling method—Random selection method not specified.
- f. Bias—Probably not biased. A large sample that was randomly selected was interviewed so it follows that the conclusions could be generalized.

**25. At the bar.**

- a. Population—Adult bar patrons.
- b. Parameter—Proportion of sample who thought drinking and driving was a serious problem.
- c. Sampling Frame—All chosen bar patrons.
- d. Sample—Every 10<sup>th</sup> person leaving the bar.
- e. Sampling method—Systematic sampling (every 10<sup>th</sup> person).
- f. Bias—Probably biased toward thinking drinking and driving is not a serious problem. The sample consisted of bar patrons leaving the bar. A large percentage of them had something to drink, most likely leading to a biased viewpoint. In addition, bar patrons don't reflect what all adults think about drinking and driving.

**26. Election poll.**

- a. Population—City voters.
- b. Parameter—Not clearly specified; likely, the proportion of voters who think certain issues are important or favor certain issues.
- c. Sampling Frame—All city resident voters.
- d. Sample—Every city resident voter in one block from each district.
- e. Sampling method—Multistage design for a combination approach. A single block chosen from each district (not clear what method used in selection) represents a cluster. The sample represents all of the residents that could be found and willing to participate. This is a convenience sample.
- f. Bias—Parameter(s) of interest not clearly specified. Convenience sampling within block clusters is not random and could produce biased results.

**27. Toxic waste.**

- a. Population—Soil located near former waste dumps.
- b. Parameter—Concentrations of toxic chemicals.
- c. Sampling Frame—Any accessible soil surrounding a former waste dump.
- d. Sample—Soil samples taken from 16 locations near a former waste dump.

- e. Sampling method–Not specified how the sample locations were chosen.
- f. Bias–Not specified how soil sample locations were chosen and therefore cannot assume they were chosen randomly, perhaps accessibility or some other factors. Unless this is known, it is possible that bias can affect the results if soil taken is more or less polluted than a random selection would produce.

**28. Housing discrimination.**

- a. Population–Landlords in a particular area.
- b. Parameter–Proportion of landlords illegally denying fair access to rental apartments.
- c. Sampling Frame–All advertised apartments.
- d. Sample–Apartments actually visited and inquired about.
- e. Sampling method–Not specified how the apartments visited were chosen.
- f. Bias–Likely to be a fair study as long as the apartments visited were randomly chosen and not all in one section of town.

**29. Quality control.**

- a. Population–Snack food packages.
- b. Parameter–Proportion of snack food packages passing inspection, weight of bags.
- c. Sampling Frame–All snack food packages produced in a day.
- d. Sample–Packages in 10 randomly selected cases, 1 bag from each case for inspection.
- e. Sampling method–Multistage sampling due to a combination of methods. The selection of the 10 cases is a cluster and the sampling selection of an individual bag from each case is probably a random sample, although this is not specified.
- f. Bias–Should be unbiased as long as the individual bag chosen is random. There could be differences in the first bag of a case versus the last bag.

**30. Contaminated milk.**

- a. Population–Dairy farms.
- b. Parameter–Not clearly specified although perhaps the proportion of dairy farms passing inspection.
- c. Sampling Frame–All dairy farms although not specifically stated this way.
- d. Sample–Not specified. Probably a random selection of farms and then a random selection of milk samples.
- e. Sampling method–Multistage sampling due to a combination of methods. The selection of dairy farms is a cluster and the sampling selection of an individual sample(s) from each dairy is probably a random sample, although this is not specified.
- f. Bias–Should be unbiased as long as the farms and the milk samples are randomly selected.

- 31. Instant poll.** The station's faulty prediction is most likely the result of bias. Only people watching the local TV station news have the opportunity to respond. The responders who volunteered to participate may have different viewpoints than those of other voters, who either chose not to respond or didn't have the opportunity to participate (didn't see the news program).
- 32. Paper poll.** The newspaper's faulty prediction is most likely the result of sampling error. The description of the stratified sampling method does suggest that the sample is representative of the voting population. However, it is unclear whether the percentages by party, age, etc. were accurate when compared to the entire voting population. Random selection of individuals within each strata means that the sample statistics will vary from the population parameter. In addition, no measure of a sampling error percentage was given for the result.
- 33. Cable company market research.**
- a. Sampling strategy is volunteer response. Bias is introduced because only those individuals who see the ad and feel strongly about the issue will respond. The opinions may not be representative of the rest of the public.
  - b. Sampling strategy is a cluster of one town selected to be sampled. Bias is introduced because one town may not be representative of all towns.
  - c. Sampling strategy is an attempted census, accessing all customers. Bias is introduced because of nonresponse to the mailing survey.
  - d. Sampling strategy is stratified by town, selecting 20 customers at random from each town to be surveyed, including follow up. This strategy should be unbiased and representative of the public opinion about the cable issue.
- 34. Cable company market research 2.**
- a. Sampling strategy is volunteer response. Bias is introduced because only those individuals who see the ad and feel strongly about the issue will respond. The opinions may not be representative of the rest of the public.
  - b. Sampling strategy is still volunteer response. Bias is introduced because only those who are strongly motivated to express their opinions will attend the meetings.
  - c. Multistage sampling, with a cluster sample within each town, consisting of those who live on a randomly selected street. Bias is introduced if there is a large percentage of residents on the selected street who do not participate or if the selected street is unrepresentative of the town as a whole.
  - d. Sampling strategy is systematic sampling. This strategy should be unbiased and fairly representative of the public opinion about the cable issue.
- 35. Amusement park riders.**
- a. This is a systematic sample (every 10<sup>th</sup> person in line).
  - b. It is likely to be representative of all of those waiting in line to go on the roller coaster. It would be useful to compare those who have waited and are now at the front with those who are in the back of the line. Otherwise, survey every 10<sup>th</sup> person about to board the roller coaster for a more consistent response.
  - c. The sampling frame consists of persons willing to wait in line for the roller coaster on a particular day within a given time frame.

- 36. Playground.** The managers will get responses only from those who bring children to the park. It is very possible that parents and others who are dissatisfied with the playground's size and condition do not come to the playground.
- 37. Another ride.** Biases exist because it could be that only those who think it is worth waiting for the roller coaster ride are likely to still be in line. Those who don't like roller coasters or don't want to stay in lines are not part of the sampling frame. Therefore, the poll won't get a fair picture of whether park patrons overall would favor more roller coasters.
- 38. Playground bias.** The first sentence points out problems and issues that the respondent may not have noticed, and might lead them to feel they should disagree. The last phrase mentions higher fees which could make people reject the proposed improvements to the playground.
- 39. (Possibly) Biased questions.**
- a. This statement is biased because it leads the responder toward yes because of the word "pollute". The word "pollute" conjures up a negative image leading the responder to agree that companies should pay for this behavior. Another way to phrase it would be "Should companies be responsible for costs of environmental cleanup?"
  - b. This statement is biased because it leads the responder to no because of the words "enforce" and "strict" that conjure up images that could lead a responder to having negative reaction. Another way to phrase it would be "Should companies have dress codes?"
- 40. More possibly biased questions.**
- a. This statement seems unbiased, stating the question without extra leading phrases.
  - b. This statement is biased because it leads the responder to agreeing with space exploration because it has been a "great tradition" of the past. The responder is unlikely to disagree with an activity that is in line with a "great tradition" of the past. A better way to phrase the question would be "Do you favor continued funding for the space program?"
- 41. Phone surveys.**
- a. It would be difficult to achieve a random sample in this case because not everyone in the sampling frame has an equal chance of being chosen. People with unlisted phone numbers, without phones, and those at work or away from the home at the designated calling time cannot be contacted.
  - b. Another strategy would be to generate random numbers and call at random times or select random numbers from the phonebook and call at random times (this doesn't solve the unlisted phone number issue).
  - c. In the original plan, families that have one person at home are more likely to be included in the study. Using the second plan, more people are potentially included although people without phones or those not home when called are still not included.
  - d. This change does improve the chance of selected households being included in the study.
  - e. The random digit dialing does address all existing phone numbers, including unlisted numbers. However, there is still the issue of residents not being home at the time of the call. In addition, people without phones are still left out of the study.
- 42. Cell phone survey.** Cell phones are not used equally in all demographic groups. Retired persons may not have cell phones. Individuals who cannot afford a cell phone or those who choose not to have one except for emergency purposes would also be left out of the study. As cell phones replace land lines and as they become more affordable and commonplace, this strategy would be a viable option.
- 43. Change.**
- a. Answers will vary



- b. The parameter being estimated is the true mean amount of change that you carry daily just before lunch.
- c. Population is now the amount of change carried by your friends. The average parameter estimates the mean of these amounts.
- d. The 10 measurements in c) are more likely to be representative of your class (peer group with similar needs) but unlikely for larger groups outside of your circle of friends.

**44. Fuel economy.**

- a. The mean gas mileage for your last 6 fill-ups (sample statistic).
- b. The mean gas mileage for your vehicle (population statistic).
- c. The results may not represent typical driving habits for an average driver of this car: speed, highway or city driving, aggressive driving, etc.
- d. The EPA would be trying to estimate the mean gas mileage for all cars of this make and model.

**45. Accounting.**

- a. Assign numbers 001 to 120 (3 digits required because the maximum number is 120) representing each order in a day. Use random numbers to select 10 transactions to check for accuracy.
- b. Separate the transactions and sample each type (wholesale and retail) proportionately. This would be a stratified random sample.

**46. Happy workers?**

- a. If all types of employees are sampled every month in relative proportions to their job types, the study should be without bias. It is not specified how the sample will be chosen. Because the company is doing the study on itself, there could be some inherent bias introduced due to self-interest.
- b. The random sample approach assigns numbers to all employees, 001 to 439 (3 digits are required because the maximum number is 439). Then a random number table or random number generator software would be used to select the sample.
- c. The random number approach does not recognize the separate labor groups where there are few managers and mostly laborers. A random sample would favor the largest group, the laborers.
- d. A better solution would be to stratify by job type (proportionately to the numbers within each type).
- e. Answers will vary. Assign numbers 01 to 14 to each person (2 digits are required because the maximum number is 14). Use a random number table or random number generator software to select 2 managers.

**47. Quality control.**

- a. Randomly select 3 cases and then randomly select one jar from each case.
- b. Assign numbers 01 to 20 to cases 07N61 to 07N80 respectively. Then generate three random numbers between 01 and 20 and select the appropriate case. Then assign random numbers 01 to 12 to each of the 12 jars within each case. For each case selected, generate a random number between 01 and 12 and select the corresponding jar within each case.
- c. The method described involves two separate sampling methods and, therefore, it is multistage sampling.

**48. Fish quality.** There are some sources of biases for the results. The scientists asked fishermen to bring any fish they caught to the field station for inspection. They did not require it and, therefore, the fishermen that participated volunteered to do so and possibly do not represent the catch of all fishermen. Fishermen with discolored fish might be more likely to bring them in for inspection. In addition, it would have to be assumed that the fish caught and brought in for inspection are representative of all fish downstream of the chemical plant. It could be misleading to suggest that 18% of fish in the river have discolored scales.

**49. Sampling methods.**

- a. Yellow pages may not include all doctor listings. If regular line listings are used, the list may include all doctors. If ads are used, not all doctors would be included and the ones with ad would not be typical of all doctors.
- b. This sampling method is not appropriate. The cluster sample chosen (the randomly selected page) will only contain a handful of businesses and maybe only one or two business types.

**50. More sampling methods.**

- a. It does not specify the method for calling local businesses and whether all or a random sample were called. Some people will say that they are willing to sign a petition when asked over the phone but later may not be willing to sign it. The mention of signing a petition may bias business owners to respond positively.
- b. If the food court is the largest and perhaps only food court in the airport, then the results would be fairly representative. If travelers don't like the food available, they probably aren't eating there and made other choices.

### **Brief Case – Market Survey Research**

While answers will vary, the questionnaire should begin with a description of the new product idea. If possible, an image of a prototype for the new product could be provided. Questions should address issues such as consumer behavior including habits and usage of such products, attitudes toward the new product and opinions about its various features, acceptance of the new product and likelihood of future purchase.

*Some sample questions follow for a new smartphone:*

*Rate your level of agreement with the following statements on a 5 point Likert scale (1 = Completely disagree; 2 = Somewhat disagree; 3 = Neither disagree nor agree; 4 = Somewhat agree; 5 = Completely agree).*

I currently own a smartphone.

I use my mobile phone to send text messages every day.

I use my mobile phone to check e-mail messages every day.

I need a smartphone to run advanced applications.

The new smartphone is superior to the mobile phone (or smartphone) I am using currently.

*Below is a list of features available on the new smartphone design. Rate the importance of each on a 5 point scale (1 = Not important; 2 = Somewhat important; 3 = Important; 4 = Very Important; 5 = Extremely Important; 9 = Don't Know).*

- 3G support
- Full touch screen .... etc.

*Will you purchase the new smartphone when it becomes available on the market?*

1 = No

2 = Maybe

3 = Probably

4 = Definitely

For this high technology product, conducting an online survey in conjunction with a cell phone service provider may be a reasonable approach. Current customers could be contacted via e-mail. Responses will help in decisions regarding the design of the new product (by determining customer expectations and preferences for various features) and launch (by predicting its success).

### **Brief Case – The GfK Roper Reports Worldwide Survey**

*What is the population of interest?*

The population of interest is consumers worldwide. This can be further defined according to market or industry segment. For example, from the questions posed in this case we may define the population as worldwide retail consumers of food and personal care products.

*Why might it be difficult to select an SRS from this population?*

Obtaining a suitable sampling frame for the population of interest could be very difficult considering the global scale. A suitable sampling frame is required to select an SRS. A more reasonable approach might be to use a multistage sampling scheme that involves cluster sampling and SRS within each cluster.

*What are some potential sources of bias?*

Potential sources of bias may result from an incomplete sampling frame or undercoverage of portions of the population (i.e., countries excluded from GfK Roper Consulting studies). As with any study, attention should be paid to potential bias from poorly worded questions and/or bias due to nonresponse. In this study special care should be used in wording questions so that they are properly interpreted in light of cultural differences among countries.

*Why might it be difficult to ensure a representative number of men and women and all age groups in some countries?*

Countries vary in terms of the factors that may affect the ability to access or obtain valid responses from representative numbers of men, women and individuals in all age groups. Depending on the sampling scheme used, these factors may result in the over- or under- representation of certain segments of the population. For example, in some countries there may be significant differences between males and females in education level, literacy rate, working outside the home, or owning property (<http://www.nytimes.com/2010/07/01/world/01iht-poll.html>). In some cultures interviewing women may be prohibited (i.e., Islamic). With regard to age, technology use is generally more prevalent among younger individuals. The differences in some countries between younger and older segments of the population with regard to the use of cell phones and/or computers may be quite significant.

*What might be a reasonable sampling frame?*

Since GfK Roper Consulting monitors consumer attitudes and trends for several major market sectors, it most likely uses several means of contacting consumers including telephone (landline and mobile), surveys (online and mail) and in-home face-to-face interviews. Therefore reasonable sampling frames would be lists of telephone numbers or mailing lists of addresses.

