

Chapter 3 Solutions

Solutions for Section 3.1

- 3.1 (a) i) Observational. ii) Descriptive.
(b) i) Observational. ii) Comparative. iii) Response: Baby's birthweight. Explanatory: Mother's weight gain during pregnancy.
(c) i) Experimental. ii) Comparative. iii) Response: Occurrence of heart disease. Explanatory: Amount of fiber in diet.
(d) i) Observational. ii) Comparative. iii) Response: Annual rate of return. Explanatory: Group of mutual fund.
- 3.2 (a) i) Observational. ii) Descriptive.
(b) i) Experimental. ii) Comparative. iii) Response: Proportion of contributors. Explanatory: Fundraising method.
(c) i) Observational. ii) Comparative. iii) Response: Presence of respiratory problems. Explanatory: Whether or not the subjects cook on a gas stove.
(d) i) Observational. ii) Descriptive
- 3.3 (a) The lower death rate may be due to heightened awareness of the staff to heart patients, a change in patient mix over time, or changes in medical treatments, and not the new screening procedure.
(b) Proofreading accuracy may decrease over amount of time spent proofreading. Also, a high error rate in the beginning may motivate readers to be more vigilant in the later part of the document. For each person, one should randomize the order of the low or high error rate sections of text to avoid this confounding.
- 3.4 (a) Nonresponse bias is present. The extra practice is voluntary, and participants are more likely to be serious runners than nonparticipants. Therefore, the practice improvement could be confounded with commitment or ability level.
(b) Because of the cost of living in a retirement community, elderly in a retirement community are generally more affluent than those in the general community. They have more money for medical care, and may maintain a higher level of health, confounding the effect of retirement community living.
- 3.5 Response: Mortality rates among heart attack survivors. Explanatory: Exercise. Confounding factors: Diet, weight, and age could all be correlated with willingness to join the exercise program, and could affect mortality. The control group is inappropriate because:
i) Physician disapproval of exercise program could be because of other health concerns, and
ii) Lack of interest in exercise program could reflect a lack of motivation to control diet and take medication reliably.
- 3.6 Response: Risk of heart disease. Explanatory: Coffee consumption. Confounding factor: Smoking.

Solutions for Section 3.2

3.7

- (a) i) Survey. ii) Descriptive.
- (b) i) Retrospective Study. ii) Comparative. iii) Response: Term of pregnancy. ~~Explanatory~~ Mother's smoking habits.
- (c) i) Survey. ii) Descriptive.
- (d) i) Survey. ii) Descriptive.

3.8 (a) Self-selection bias: People who didn't participate value their time more, and **may be** wealthier and able to eat different and more expensive foods.
 Bias in interview: The high interviewer turnover and the lack of training **could lead to** biases in how the questions were asked.
 Bias in sampling: If the design was disturbed, then the sample is suspect of **not being a** representative random sample of the target population.

(b) See above description of each type of bias.

3.9 The timing of the poll is during morning rush hours, when most listeners are **people with** regular working hour jobs, creating a convenience sample of employed, more affluent **adults**. This only samples this particular segment of the population and induces bias in **the survey** results. Also, in call-in polls, people who feel strongly for or against a particular **issue are** more likely to voice their opinion. The results reflect proportions of people **with strong** opinions which may differ from the overall proportions for or against an issue.

3.10 Bias will be a problem. Elderly people alive at the end of 6 years are generally **healthier** and in less need of nursing home care than those who died. This would **underestimate the** use of nursing home services.

3.11 People with more education are more likely to have access to, or even recognize **the need to** consult, a medical specialist. This gives a selection bias of people with AD that is **related** to education level. People with more education are overrepresented as AD patients **in the** specialist's studies.

Solutions for Section 3.3

- 3.12 (a) Systematic sampling.
 (b) Stratified sampling, by oldest child's year in school.
 (c) Simple random sampling.
 (d) Multistage cluster sampling. The high schools are stratified by elementary/high school, then a sample taken of each one, forming a cluster. Further clusters of the **grade levels** within each school are formed, and finally a SRS is taken within this **sampling frame**.

3.13 The first three are conceptually simple sampling plans. Stratified sampling in (b) **may give** more precise estimates than (a) or (c) if preferences on dealing with growth are **related** to the age of the oldest child in school. Systematic sampling in (a) and SRS in (c) are equally efficient, but (a) may be easier to implement. Plan (d) is more complicated, **first** stratifying by high school/elementary school, then taking a cluster sample within each **type** of school, and within each school selected, taking a cluster sample of the grade levels to form the sampling frame. The stratification of high school/elementary school can **improve** the precision of the estimate if a child's age is related to parent's opinions on school growth. Plan (d) may be easier to implement than plan (b).

- 3.14 (a) Systematic sampling.
 (b) Simple random sampling.
 (c) Stratified sampling by gender.
- 3.15 SRS: choose each customer randomly. Systematic: Choose one out of the first 10,000 customers on the lists, and then go through the lists selecting each 10,000th subsequent customer. Stratified: Within each "local" area, choose the appropriate proportion of customers based on the number of people based in the area. Cluster: Choose a sample of "local" areas, then choose a sample of customers within each area.
- Stratified sampling will give the most precise estimates but is the most difficult to implement. SRS and systematic sampling give less precise estimates than stratified sampling for the same sample size, but systematic sampling is easier to implement. Cluster sampling is the most convenient to implement, but gives the least precise estimates for the same sample size relative to the other sampling methods, and it may be very biased if it omits the major city.
- 3.16 SRS: Each member surveyed is chosen randomly from the pooled member list. Systematic: A starting point is chosen from the pooled member list, then each k th member after that is surveyed. Stratified: Within each local chapter, a SRS is chosen. Multistage Cluster: A sample of states is chosen. Within each state, a sample of local chapters is chosen, and within each chapter, a SRS of members is selected.
- Stratified sampling will give the most precise estimates but is the most difficult to implement because every chapter will need to be contacted for the survey. Systematic sampling could give a more precise estimate than SRS for the same sample size, if the overall membership list is grouped by chapter, thereby insuring that a lot of different chapters are sampled. Cluster sampling is the most convenient to implement, but gives the least precise estimates relative to its sample size of all the sampling methods, and it may be very biased if it omits the largest chapters.
- 3.17 A 2-stage stratified sample: Create 4 strata of 100 male managers, 50 female managers, 1900 male employees, and 1450 female employees. Since the sample size relative to the population size is $700/3500 = 1/5$, from each strata sample $1/5$ th of the people. This sample will reflect the proportion of males and females in management.
- 3.18 (a) Stratified sampling using job category strata, because computer usage is highly dependent on job classification.
 (b) Systematic sampling, because it is logistically easy to interview every 10th patient. Also, the sample can be collected on an ongoing basis. In contrast, a SRS requires the total population size to be identified before a sample can be drawn.
 (c) SRS or Systematic sampling.
 (d) Multi-stage cluster sampling, with college as a cluster, because there are too many colleges to use each college as a stratum.

Solutions for Section 3.4

- 3.19 Response: Fabric strength (loss in fabric weight). Treatments: Chemical agent, wash times. Blocks: Martindale wear tester batches and bolts of fabric.