Problems to be handed in:

1) Submit electronically exercises 1 to 5, and 36 to 38 from Unit A3 of CyberStats.

A3 Ex. 1. What are the response and explanatory variables?

Solution:\(2\): Often, you can think of the explanatory variable as the variable that defines the groups that you want to compare. So, here the explanatory variable is “regular use of Vitamin C.” The two groups formed from this explanatory variable would be “those that do use Vitamin C regularly” and “those that don’t.” Often, you can think of the response variable as the outcome measure with respect to which you want to compare two or more groups. So, here, the (binary) response variable is “whether or not you get a cold.”

A3 Ex. 2. Briefly describe how the researcher would do an observational study to examine this relationship.

Solution:\(2\): In an observational study, you just observe people’s natural behavior. So, here, you’d just identify a group of people who take Vitamin C regularly, and identify a group of people who do not take Vitamin C regularly. Then, you’d compare the percentage of people who get a cold in each group.

A3 Ex. 3. What confounding factors do you think would be present in this observational study? Briefly explain.

Solution:\(2\): A confounding factor is a variable that is related to both the explanatory variable and the response variable. Other (“extraneous”) variables that probably explain whether or not you get a cold are eating, drinking, exercise, and sleeping habits. In order for these extraneous variables to be considered confounding factors, they must also be related to the explanatory variable, that is, whether or not you take Vitamin C regularly. If people who take vitamin C regularly are more health conscious – that is, if they eat, drink, exercise and sleep better – than people who don’t take vitamin C regularly, then eating, drinking, exercise, and sleeping habits are all confounding factors. The term confounding comes from the idea that confounding factors “confuse” or “confound” the relationship between the explanatory variable and the response variable. We can’t be sure if differences in the response variable (whether or not you get a cold) are due to the explanatory variable (whether or not you take Vitamin C) or the confounding factors (whether or not you eat, drink, exercise, and sleep well). The relationship is “confounded.”

A3 Ex. 4. Suppose that 100 people are available for an experiment. Describe a randomized experiment for this problem.
**Solution**: Rather than just observing people’s natural behavior, in a randomized experiment, a researcher randomly assigns different people to the different levels of the explanatory variable. So, here, we’d randomly assign 50 of the people to take Vitamin C, and randomly assign 50 people to not take Vitamin C (or to take a placebo pill). Then, we’d compare the percentage of people in each group who get a cold.

**A3 Ex. 5.** Why would it be better to do an experiment rather than an observational study to examine this relationship?

**Solution**: In general, it is better to do an experiment over an observational study, since in an experiment the groups should be, on average, fairly similar with respect to everything except the treatment of interest. If the groups are similar, except for the treatment, then if there is a different in the groups, we can say the treatment caused the difference. We say there is a cause-and-effect relationship.

**A3 Ex. 36.** Is the research described in the article an observational study or an experiment?

**Solution**: Because the researchers just observed naturally occurring groups of walkers and non-walkers, the study is an observational study.

**A3 Ex. 37.** What are the response and explanatory variables for this study?

**Solution**: The primary response variable is “death.” The explanatory variable is “whether or not you walk two or more miles per day.”

**A3 Ex. 38.** Explain why the headline and the first sentence of the article are inappropriate descriptions of the results. What are some alternative explanations for the results?

**Solution**: The headline and first sentence suggest that walking two or more miles per day causes the reduction in the risk of death. Other factors may explain the reduction. The group of men who walk two or more miles per day may exercise even more. Also, they may eat and drink healthier than the non-walkers.

2) Submit electronically exercises 2, 3, 4, and 5 from Unit A2 of *CyberStats*.

**A2 Ex. 2.** A major airline has flights operating out of 50 cities worldwide. It is planning to conduct a survey of 5,000 customers next summer to determine how it can improve its check-in facilities and services. What kind of sampling plan would you advise it to use (stratified, simple random sampling, etc.)? Explain your reasoning.

**Solution**: Assuming the airline wants information about the check-in facilities and services for all 50 cities, a stratified sample should be used. The airline could take a sample of 100 customers from each city.

**A2 Ex. 3.** In the above situation, describe the steps it should use to select the sample of 5,000 customers according to the plan you recommended.
Solution(2): Ideally, the airline would obtain a list of all passengers checking in at each city over the summer, and select a simple random sample of 100 of them. Those customers would then be sent a survey.

However, this may not be practical, because the customers flying at the beginning of the summer should be surveyed shortly after they use the services, and the entire list wouldn’t be available yet at that time. An alternative would be to randomly choose two passengers each day from the list of those checking in at that city that day, and hand them the survey when they check in.

A2 Ex. 4. A health insurance company sent a survey to a simple random sample of 2,000 of its customers. Of the 500 who responded, over 50 percent complained that they have to wait too long to get an appointment with their primary care physician. The insurance company sent a letter to all of the primary care doctors who accept their insurance and told them that a “significant majority” of their customers felt they had to wait too long for an appointment. Explain the problem.

Solution(2): Many of the individuals contacted do not respond. This problem is worse than it may originally sound, because those who are willing to respond to questions are generally those who feel most strongly about the issues. The problem is particularly acute when surveys are conducted by mail (a letter was sent). The term “significant majority” may not be defined.

A2 Ex. 5. The student government at a university wanted to find whether students would support an additional fee for athletics. They were not able to obtain a list of all students, but managed to get a list of students who registered to park their cars on campus. They used that list to select a simple random sample of students. Explain the problem.

Solution(2): The sample is drawn from a list that isn’t representative of the population. Only students who parked their cars on campus were in the sampling frame.

3) Submit electronically exercises 13 to 16 from Unit A4 of CyberStats.

A4 Ex. 13. What is the population of interest to the faculty senate?

Solution(2): The population of interest is the population of all students at the university.

A4 Ex. 14. What is the sampling frame?

Solution(2): The sampling frame is the collection of all students enrolled in statistics classes at the time of the survey.

A4 Ex. 15. What is the sample?

Solution(2): The sample is the 500 students to whom the survey was mailed.

A4 Ex. 16. Is the sample representative of the population of interest? Explain.

Solution(2): The extent to which the sample represents the population of interest depends on what types of students are enrolled in statistics classes that term. Depending on the university requirements and the term in question, there may be more freshman or seniors than in the general student body, and certain majors are likely to be over- or under-represented.
4) According to Cynthia Crossen (Tainted truth: The manipulation of fact in America. New York: Simon and Schuster, 1994, p. 106): “It is a poller’s business to press for an opinion whether people have one or not. ‘Don’t knows’ are worthless to pollers, whose product is opinion, not ignorance. That’s why so many polls do not even offer a ‘don’t know’ alternative.”

a) Explain how this feature might lead to bias in a survey?

Solution(2): People who prefer to answer that they don’t know, but are forced to choose an answer, will use something other than an informed opinion to do so. For example, if a question were to ask “Do you support George W. Bush’s new tax bill in Congress?” and people who were unfamiliar with the bill were forced to answer, their answers would probably be more of a reflection of how they feel about taxes or about the current Congress.

b) By considering each of the components of a study (Lecture 1-5), describe how opinion polls with this feature can be poor studies.

Solution(2): The lack of a “don’t know” as a possible response would be uncovered by Component 2, “the exact ... questions asked.” The other components are all relevant to an opinion poll as well, although a poll with out an “don’t know” is no more suspect to them than the usual poll.

c) Suggest some ways to change opinion polls to remove this difficulty without reducing their effectiveness to pollsters.

Solution(1): A natural way is to add answers to the poll that measure the strength of the belief. In the tax bill example, this could be by including alternatives such as “strongly believe”, “weakly believe”, etc. In addition, a separate question could be asked which identifies how much the person knows about the bill: “Are you very familiar with George W. Bush’s new tax bill in Congress?” This would balance the strength of belief with amount of knowledge. Of course, the simplest way is to just include the “don’t know” category.

5) A recent advertisement for a piece of personal fitness equipment, “Captain Crunch,” claimed, “In just two months, research shows that people who used Captain Crunch lost three inches on their waistline.”

Explain how each of the components of a survey should have been chosen if the research results are fair and unbiased.

a) Component 1:

Solution(2): The participants should have been randomly selected from among typical potential Captain Crunch users.

b) Component 2:

Solution(2): The participants should have been measured by the researchers rather than reporting the waistline size loss themselves. The measurements should have been taken in the same way for each participant.
c) Component 3:

**Solution** (2): The Captain Crunch should have been used in a natural setting, at home, rather than in a lab where someone else would be making sure participants used it regularly. The setting should have been the same for all participants.


d) Component 4:

**Solution** (2): The advertisement would be misleading if there were major differences between the group tested and the typical user, like if this group started out grossly overweight or if they were on a lean diet at the same time.


e) Component 5:

**Solution** (1): The claimed difference is 3 inches which for a typical person appears to be a lot. However, we don’t know the size of the participants before they started the program. It would be important to measure the variation in waistline over the same period by people not on the program. It is possible that there is great variation in this, especially for very large people. However, the effect does appear to be large.