Shana Green

DATA 606 - Homework 1

Due Date: 8/30/2020

Smoking habits of UK residents. (1.10, p. 20) A survey was conducted to study the smoking habits of UK residents. Below is a data matrix displaying a portion of the data collected in this survey. Note that "£" stands for British Pounds Sterling, "cig" stands for cigarettes, and "N/A" refers to a missing component of the data.

	sex	age	marital	grossIncome	smoke	amtWeekends	amtWeekdays
1	Female	42	Single	Under £2,600	Yes	12 cig/day	12 cig/day
2	Male	44	Single	£10,400 to £15,600	No	N/A	N/A
3	Male	53	Married	Above £36,400	Yes	6 cig/day	6 cig/day
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1691	Male	40	Single	£2,600 to £5,200	Yes	8 cig/day	8 cig/day

(a) What does each row of the data matrix represent?

Each row of the data represents the number of UK residents.

(b) How many participants were included in the survey?

According to the data matrix, there were 1691 participants in the survey.

(c) Indicate whether each variable in the study is numerical or categorical. If numerical, identify as continuous or discrete. If categorical, indicate if the variable is ordinal.

sex (categorical)

age(numerical, discrete)

marital (categorical)

grossIncome (categorical, ordinal)

smoke (categorical)

amtWeekends (numerical, discrete)

amtWeekdays (numerical, discrete)

Cheaters, scope of inference. (1.14, p. 29) Exercise 1.5 introduces a study where researchers studying the relationship between honesty, age, and self-control conducted an experiment on 160 children between the ages of 5 and 15[1]. The researchers asked each child to toss a fair coin in private and to record the outcome (white or black) on a paper sheet, and said they would only reward children who report white. Half the students were explicitly told not to cheat and the others were not given any explicit instructions. Differences were observed in the cheating rates in the instruction and no instruction groups, as well as some differences across children's characteristics within each group.

(a) Identify the population of interest and the sample in this study.

The population of interest are children between the ages of 5 and 15.The sample is 160 children between the ages of 5 and 15.

(b) Comment on whether or not the results of the study can be generalized to the population, and if the findings of the study can be used to establish causal relationships.

The results of this study cannot be generalized to the population! There is no information provided in the statement that mentioned the sample was taken randomly. We cannot generalize to a target population in the results of this study. Since the statement above indicates an experimental test, we can use this result to correlate a causal relationship.

---

[1] Alessandro Bucciol and Marco Piovesan. "Luck or cheating? A field experiment on honesty with children". In: Journal of Economic Psychology 32.1 (2011), pp. 73-78. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1307694

Reading the paper. (1.28, p. 31) Below are excerpts from two articles published in the NY Times:

(a) An article titled Risks: Smokers Found More Prone to Dementia states the following:

"Researchers analyzed data from 23,123 health plan members who participated in a voluntary exam and health behavior survey from 1978 to 1985, when they were 50-60 years old. 23 years later, about 25% of the group had dementia, including 1,136 with Alzheimer's disease and 416 with vascular dementia. After adjusting for other factors, the researchers concluded that pack-a-day smokers were 37% more likely than nonsmokers to develop dementia, and the risks went up with increased smoking; 44% for one to two packs a day; and twice the risk for more than two packs."

Based on this study, can we conclude that smoking causes dementia later in life? Explain your reasoning.

This study is considered observational, so we cannot conclude smoking causes dementia later in life. This study, however, does show correlation between increased smoking and dementia.

(b) Another article titled The School Bully Is Sleepy states the following:

"The University of Michigan study, collected survey data from parents on each child's sleep habits and asked both parents and teachers to assess behavioral concerns. About a third of the students studied were identified by parents or teachers as having problems with disruptive behavior or bullying. The researchers found that children who had behavioral issues and those who were identified as bullies were twice as likely to have shown symptoms of sleep disorders."

A friend of yours who read the article says, "The study shows that sleep disorders lead to bullying in school children." Is this statement justified? If not, how best can you describe the conclusion that can be drawn from this study?

This example is very similar to part (a); this study is also observational. I would tell my friend that I cannot conclude that sleep disorders lead to bullying school children. I can tell my friend there is correlation between sleep disorders and bullying.

Exercise and mental health. (1.34, p. 35) A researcher is interested in the effects of exercise on mental health and he proposes the following study: Use stratified random sampling to ensure representative proportions of 18-30, 31-40 and 41-55 year olds from the population. Next, randomly assign half the subjects from each age group to exercise twice a week, and instruct the rest not to exercise. Conduct a mental health exam at the beginning and at the end of the study, and compare the results.

(a) What type of study is this?

This is an experimental study.

(b) What are the treatment and control groups in this study?

The group who exercises twice a week is the treatment group. The control group is the group that does not exercise.

(c) Does this study make use of blocking? If so, what is the blocking variable?

This study does use blocking and the blocking variable are the age groups. The blocked age groups are 18-30, 31-40, and 41-55.

(d) Does this study make use of blinding?

Blinding is not used in this study. The participants are well aware if they are in the exercise or non-exercise group.

(e) Comment on whether or not the results of the study can be used to establish a causal relationship between exercise and mental health, and indicate whether or not the conclusions can be generalized to the population at large.

This study can be used to establish a causal relationshp between exercise and mental health. The conclusion can be generalized to the population at large because random sampling was conducted.

(f) Suppose you are given the task of determining if this proposed study should get funding. Would you have any reservations about the study proposal?

Suppose active individuals were assigned to the non-exercise group and vice versa. The deisgn of the group assignment could have a negative effect on the results of the study.