Bryan Persaud

title: “Chapter 1 - Introduction to Data” author: "" output: pdf_document: extra_dependencies: [“geometry”, “multicol”, “multirow”] —

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 represents the smoking habits of a UK resident.

(b) How many participants were included in the survey? # There were 1691 participants in this 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. # There are some variables that are numerical and some that are categorical. The variables that are numerical are age, amtWeekends, and amtWeekdays. Out of these varialbles, age is continuous and amtWeekend and amtWeekday are discrete. The variables that are categorial are sex, marital, grossIncome, and smoke. Out of these variables, grossIncome would be ordinal. The other variables would be nominal.

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. 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 is the 160 children between the ages of 5 and 15. The sample in this study were half the children who were explicitly told not to cheat and the others who were not given any explict instructions.

(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 the study cannot be genralized to the population. This is because the study does not represent all of the children in their specific age group. Also the children were told if they answered white they would be rewarded, so this affects the children's decision to answer honestly. The findings of the study can be used to establish a casual relationship since in this experiment the results were based on half of the children told not to cheat.

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. # We cannot conclude that smoking causes dementia later in life based on this study because the data was analyzed from a survey and not from a experiment using control groups. It is tough to decide that smoking causes dementia based on the results of this survey. Also it wasn't just dementia that was found, there were some who participated that ended up with Alzherimer's disease. There was also no information provided on the other factors used to come to the conclusion that smoking causes demetia.

(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? # The statement that sleep disorders lead to bullying is not justified based on the study showed. The conclusion that can be drawn is that children who had behavioral issues and those who were identified as bullies are more likely to have symptoms of sleep disorders.

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 rep- resentative 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? # The type of study is an experiment that uses a control group and a treatment group where the tretment group is assigned to exercise.

(b) What are the treatment and control groups in this study? # The treatment group in this study is half of the subjects from each age groups 18-30, 31-40, and 41-55 and to have them exercise twice a week. The control group is the other half of the age groups 18-30, 31-40, and 41-55 and to have them not exercise.

(c) Does this study make use of blocking? If so, what is the blocking variable? # This study does make use of blocking. The blocking variable is age. This study uses blocking by using age groups.

(d) Does this study make use of blinding? # No, this study does not make use of blinding. Everyone in the study is instructed as whether they need to exercise or not and a mental health exam is conducted at the beginning and at the end of the study.

(e) Comment on whether or not the results of the study can be used to establish a causal rela- tionship between exercise and mental health, and indicate whether or not the conclusions can be generalized to the population at large. # The results of this study can be used to establish a casual relationship between exercise and mental health. With a big enough sample size, the conclusionn can be generalized to the population at large.

(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? # The reservations about the study proposal I would have is the effect on not only the mental health but on the physical health as well. The subjects in the control group are told not to exercse as a way to measure mental health, but this can lead to a decrease in their physical and mental health. This experiment might lead some to develop health issues.