

# Data 606 Chapter 1 Homework

Leticia Salazar

2021-09-04

## CHAPTER 1 - INTRODUCTION TO DATA

Smoking habits of UK residents. (1.10, p.20) A survey was conducted to study the smoking habits of UK residents. Below is the 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 matrix represents the results of the residents smoking habits*

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

*There were 1691 participants included 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.

- *index: numerical*
- *sex: categorical*
- *age: numerical [discrete]*
- *marital: categorical*
- *gross income: categorical [ordinal]*
- *smoke: categorical*
- *amt weekends: numerical [ordinal]*
- *amt weekdays: numerical [ordinal]*

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 rate 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 the study?

*The interest sample in the study are 160 children between the ages of 5 to 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 the study cannot be generalized to the population because all we know is that 160 children were taken in for the study. There is no additional details that states diversity / randomness within this group. There is just too much information missing in order for it to be generalized.*

---

Reading the Paper. (1.26, 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 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, we can conclude that smoking causes dementia later in life? Explain your reasoning.

*We cannot conclude that smoking causes dementia later in life because we don't know if there's a direct relationship between smoking and dementia / Alzheimer's. More information would need to be provided, possibly a further study with a larger population to see if this can be concluded.*

*(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 articles says, " The study shows that sleeps 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 is not justified because the relationship being "concluded" is not 100% true. To best describe the conclusion, those students who had disruptive behavior or bullying can have sleeping disorders. I do think that another conclusion to be drawn can be that those students that are being bullied can have sleeping disorders or behavioral issues.*

---

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- representative proportions of 18-30, 31-40 and 41-55 year olds from the population. Next, randomly assign half the subects 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 a randomized study.*

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

*The treatment group is the subject who have to exercise twice a week. The control group is those who won't be exercising.*

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

*Yes, this study does make use of blocking by using the ages 18-30, 31-40 and 41-55.*

(d) Does this study make use of blinding?

*There is no use of blinding in this study, not for the people conducting the study or for the subjects.*

(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?

*Based on this study being randomized the conclusions can be generalized to the population. There are a few things that will have to stay random in order for it to work as well as a large enough population.*

(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?

*I wouldn't have any reservations about this study proposal to get funding. This study would be beneficial to many people especially if conducted well.*