Data606-HW1

George Cruz

9/5/2020

Introduction to Data

1.a. The rows are named a *case* or *observational unit*. 1.b. Apparently, 1691 participants. 1.c. Variables: *sex*. categorical, nominal. *age*. numerical, discrete. *marital*. categorial, nominal. *grossIncome*. categorical,ordinal. *smoke*. categorical, nominal. *amtWeekends*. numerical, discrete. (I think it can be averaged and summarized) *amtWeekdays*. numerical, discrete.

Cheaters, scope of inference

a. Population is all children between 5 and 15 years old. The sample is 160 children between those ages.
b. The researchers conducted an experiment in which they have no way of verifying the answers received. Their goal was to study the relationship between honesty, age and self-control but they gave themselves little control on how to measure that relationship. Granted, you could assume the number of white answers to average 50% and consider any deviation from it to be cheating. This does not really guarantee accuracy. I don't believe a causal relationship can be established. In terms of generalization, we don't know the composition of the sample and if it really represents the population.

Readind the paper

a. No. Correlation does not imply causality. To establish causality, we would need to perform an experiment with a random sample and not only with voluntary supplied data.

(b)Again, a correlation does not imply causality. At most, we can establish that there is correlation between sleep disorders and bullying not that one causes the other. There could be one or more variables that cause these two issues and cannot be determined by the data collected.

Exercise and mental health

a. Blocked randomized controlled experiment
b. Treatment group: the participants who were instructed to exercise. Control group: participants who were instructed not to exercise.
c. Yes, age is the blocking variable.
d. No.
e. The experiment design appears to be correct. The way the experiment is designed will only allow us to determine a causal relationship between exercise and mental health moving forward, as we do not know the exercise habits of the participants before the experiment.
f. Yes, I believe the sample should take into consideration exercise habits before the experiment and randomly distribute the individuals into treatment and control group to properly measure the effects of exercise(or lack thereof) on mental health not only on age groups but also based on prior exercising habits.