

# STAT 3400 - Homework #1

Alex Ojemann

Due January 25, 2023

## Problem 1.4.10

- a. This is an experiment rather than an observational study because we are separating the population into a treatment group and a control group and seeing if the treatment group has different results.
- b. Each adolescent participating in the study is a case and there are 2260 of them.
- c. The response variable is whether each participant got COVID. It's a nominal categorical variable because the possible responses are yes and no which don't have an order.
- d. The explanatory variables in this study is whether each patient received the treatment or the placebo. It's also a nominal categorical variable because the possible values are treatment and placebo which don't have an order.

## Problem 2.5.12

- a. This is an observational study rather than an experiment because no treatment is explicitly applied.
- b. This study can't be used to conclude a causal relationship between increased stress and muscle cramps because observational studies don't reflect causal relationships because we can't isolate the response variable.
- c. Potential confounding variables include physical burden and dehydration, which could both cause increased stress and muscle cramps.

## Problem 4.8.4

- a. Views on raising taxes and political affiliation appear to be associated because a substantially higher portion of Democrats think it's better to raise taxes on the rich compared to Republicans and Independents and a substantially lower portion of Democrats think it's better to raise taxes on the poor compared to Republicans and Independents, while a substantially higher portion of Republicans and Independents aren't sure compared to Democrats.
- b. Democrats could be poorer than Republicans and Independents, which would lead more of them to want higher taxes for the rich instead of the poor. Republicans and Independents could believe that the government should have less power over people, leading them to respond not sure rather than to raise taxes for either the rich or poor at a much higher rate than Democrats.

## Problem 4.8.6

- a. The bottom graph is better to understand the shipping choices of people of different ages because it uses the age groups as the different bars which allows their shipping choices to be compared more easily.
- b. The bottom graph is better to understand the the age distribution across different types of shipping choices because it uses the shipping choices as the different bars which allows their age groups to be compared more easily.
- c. USPS would be their biggest competitor because they have the largest segment in the 55+ bar in the bottom chart.
- d. FedEx should market more to customers ages 55+ to balance out their age distribution because 55+ has the smallest section in FedEx's bar in the top chart.

## Problem 5.10.2

1. Positive linear association
2. No association
3. Positive nonlinear association
4. Negative nonlinear association (very close to linear but appears slightly concave down)

## Problem 5.10.16

- a. This distribution is skewed right because there are a meaningful number of points above 6 million which is well to the right of where the normal distribution has a significant number of points. Because the distribution is skewed, the median and IQR are better summary statistics to use than the mean and standard deviation.
- b. This distribution is symmetric because the first, second, and third quartiles are each 300,000 apart, there are very few houses above 300,000 above the third quartile, and there are no houses below 300,000 below the first quartile (0). Because the distribution is symmetric, the mean and standard deviation are better summary statistics to use than the median and IQR.
- c. This distribution is skewed right because almost all the students are at 0 and a few students are far to the right of that. Because the distribution is skewed, the median and IQR are better summary statistics to use than the mean and standard deviation.
- d. This distribution is skewed right because almost all the employees earn a regular working salary and the high-level executives earn much, much more than that. Because the distribution is skewed, the median and IQR are better summary statistics to use than the mean and standard deviation.
- e. This distribution is symmetric because the first, second, and third quartiles are each one week apart and we aren't explicitly told that a meaningful amount of data exists several weeks above or below the median of 39. Because the distribution is symmetric, the mean and standard deviation are better summary statistics to use than the median and IQR.