DA6213

Exercise #2

Name: Emily Bates

Having talked about privacy and big data, let’s see how US adult consumers feel about it. This exercise utilizes the 2015 National Consumer Survey. The data is available in Excel, SPSS and SAS formats. You should be able to get at it from one of these forms and move it to whatever platform you want. Here are the instructions for this exercise.

1. Create a bar graph that displays the data for the variable commodity – this variable represents how consumers feel about their willingness to provide personal data for something of value. What does the bar graph tell you in general?

Generally the bar plot is showing that most people fall into the neither willing or unwilling to provide personal data for something of value group or they fall into the group that is a little bit willing.

A graph with blue bars

Description automatically generated

1. Create a bar graph that displays the data for the variable like\_know – this variable represents the extent to which consumers would like to know how their data is being used. What does the bar graph tell you in general?

Generally, this plot shows us that people tend to want to know how their data is being used, as they mostly fall into agree a lot, and agree a little.

A graph of a bar graph

Description automatically generated with medium confidence

1. Perform the appropriate statistical test to see whether or not men and women differ in terms of how they feel about exchanging their personal data for something of value. Interpret what you found – be sure to comment on the magnitude of any differences you find.

Because we have an independent variable with two levels (gender) and an ordinal dependent variable, I will use the Wilcoxon-Mann Whitney test. The p-value is 1.91e-6, indicating we reject he null hypothesis. There is a statistically significant difference between men and women in terms of how they feel about exchanging personal data for something of value. The estimated difference in location (median shift) between men and women (using the Hodges-Lehmann estimator) is -2.129606e-5. When considering the commodity variable scale of 1-5, practically speaking this doesn’t seem like a substantial difference even though it's statistically significant.

1. Perform the appropriate statistical test to see whether or not people who drink coca cola differ in terms of how they feel about knowing how their data is being used. Interpret what you found – be sure to comment on the magnitude of any differences you find.

Because we have an independent variable with two levels (classic\_coke) and an ordinal dependent variable, I will use the Wilcoxon-Mann Whitney test. The p-value is 0.001151, indicating we reject the null hypothesis. There is a statistically significant difference between people who drink coke and those who don’t drink coke in terms of the extent to which they would like to know how their data is being used. The estimated difference in location (median shift) between men and women (using the Hodges-Lehmann estimator) is 2.307603e-5. Similar to the previous question, it seems that practically this isn’t a huge difference given the scale of 1-5. However, the difference is still statistically significant.