Homework 2-0

Please answer the following questions based on your reading in Chapter 2 of the “Doing Data Science” text. Brief, but complete answers are best (28 points total).

1. (13pts) Consider the following scenario:

A survey was conducted of UMass students. The survey consisted of one question: “On a scale of 1 to 5, rate your opinion of how effective the energy drink ‘Red Calf’ was on your academic performance this semester”. A total of 50 students responded to the survey.

Based on the above scenario, answer the following questions:

1. (5pts) Explain how the terms “population” and “sample” apply to this study.

The population in this case would be all the UMass students and the sample of the study would be the 50 students that responded to the survey.

1. (3pts) List three sources of bias in the study.

One source of bias would be that the 50 students that responded most likely responded since they either like the drink or hate the drink and it wasn’t chosen at random from the population. Another source of bias could be that only UMass students are involved and the drink could be popular in one school but not another. Also, depends on who conducted the survey say if was a professor, student, or the Red Calf itself.

1. (5pts) Explain how you would address one of the sources of bias you identified.

Instead of having a number scale they could have a few questions about it. Another change is to give the survey to other colleges as well and conduct the survey through a computer.

1. (5pts) In the text, the question is asked: “Can N=all”? In your own words, explain what this question means and express your opinion on this topic: agree or disagree, and say why.

This question means that the population of N is every person even if the experiment wasn’t conducted on every person. This also questions whether we can truly get accurate data, which I believe is not possible. This is because most of the time when a survey is conducted the data collected is never fully accurate.

1. (5pts) In your own words, explain what a model of data is. Describe a factor that would contribute to make a model perform poorly as a predictor of outcomes?

A model of data is a diagram that allows us to understand trends and relationships in the data. A poor predictor would be if the axes are incorrectly scaled causing a bad data visualization.

1. (5pts) In your own words, briefly explain the term “Exploratory Data Analysis”.

This is a form or approach to analyzing data that allows us to view and summarize main characteristics. This gives us insights further than modeling or hypothesis testing.