DATA 606 Data Project Proposal

Kenan Sooklall

Data Preparation

```
# load data
```

Research question

You should phrase your research question in a way that matches up with the scope of inference your dataset allows for.

Researchers have shown how covid-19 has affect scent, my project will expand that analysis to taste. https://www.washingtonpost.com/business/2020/12/01/covid-scented-candle-reviews/

Cases

What are the cases, and how many are there?

There are many different kinds of protein powder sold on Amazon.com with varying reviews. Some people think they taste horrible others not so much. My goal is to see if there was a significant change in bad/tasteless reviews in 2020 due to covid-19. According to many reports, side affects of covid-19 causes one to loose their sense of taste and smell. My wife was one of these cases.

I want to see if the number of negative reviews for protein powder went up significantly in 2020 due to covid-19 or not.

Data collection

Describe the method of data collection.

The data will be collected from Amazon.com specifically reviews from 2020 and as far back as they go. The goal would be as much a possible and as recent as possible.

Type of study

What type of study is this (observational/experiment)?

This is an observational study since I won't be manipulate anything.

Data Source

If you collected the data, state self-collected. If not, provide a citation/link.

The data will be collected from Amazon.com https://www.amazon.com/product-reviews/B01NAEHLFO/ref=cm_cr_arp_d_show_all?pageNumber=1&reviewerType=all_reviews

Dependent Variable

What is the response variable? Is it quantitative or qualitative? The response variable is the change in taste reviews in 2020 due to covid, it will be a quantitative, a count.

Independent Variable

You should have two independent variables, one quantitative and one qualitative. The data will be qualitative, ie taste good/bad and quantitative in the proportions

Relevant summary statistics

Provide summary statistics for each the variables. Also include appropriate visualizations related to your research question (e.g. scatter plot, boxplots, etc). This step requires the use of R, hence a code chunk is provided below. Insert more code chunks as needed.

$H_o = \mu_{protien-badtaste-2020} = \mu_{protien-badtaste-2019before}$

$H_a = \mu_{protien-badtaste-2020} \neq \mu_{protien-badtaste-2019before}$

I can't provide any visualization at the moment since I haven't scrapped the data yet