Introduction

In the wake of the covid pandemic, I curiously watched my parents who are Americans as proud as can be suddenly begin to malign protesters following the George Floyd protests. As someone who believes that there is nothing more american than protesting, my first amendment loving parents suddenly were angrily and worriedly discussing the need for anti protest laws at the dinner table. I believe that the media in the United States plays a large part in stoking fear and resentment in people like my parents to bolster their bottom line, and so I set about to see if I could prove it.

Hypothesis

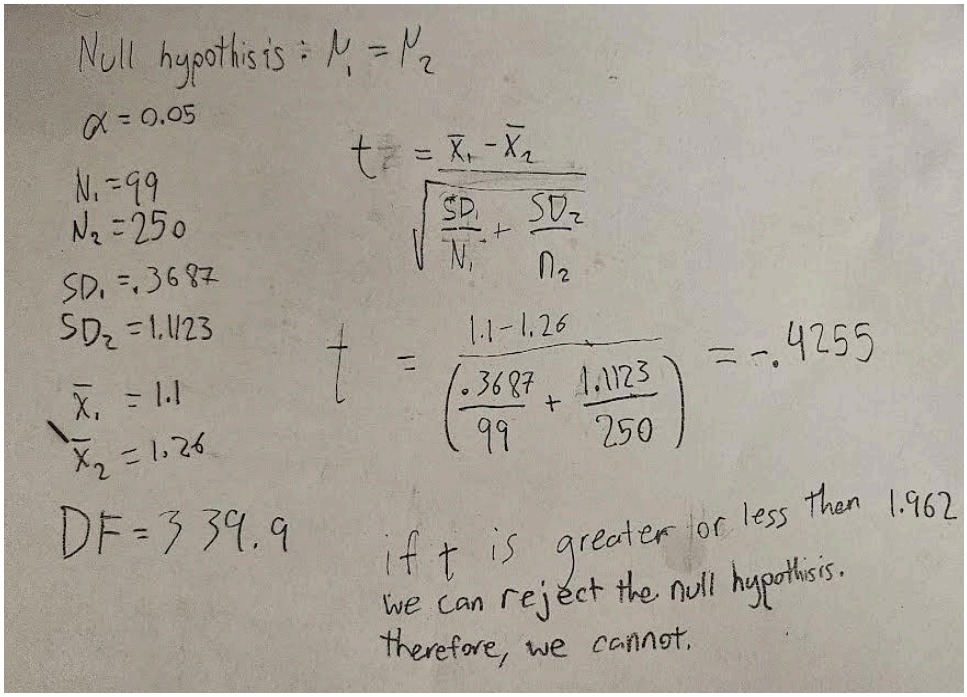
I hypothesize that comparing the sentiment of people's tweets before the George Floyd protests and after will reveal that people will tweet more negatively in regards to protests. I will sentimentally analyze people's tweets including words such as "protests", "protesting", and "protest" before the protests, and in the two weeks.

Experimental design

I will do a two sample T test to compare the mean sentiment before and after. If the mean sentiment after is below the mean sentiment before by a wide enough margin, then I can say with 95% confidence that people tweeted more negatively in regards to protests after the protests occurred. Given that the vast majority of people have had no first hand experience with the protests, this would be evidence of negative media bias towards the protests.

Modifications

I modified my code to calculate the standard deviation, collect the data in the proper way to be analyzed and export the data to a convenient text document.



With N1 being tweets before the protests and N2 being tweets after the protests we fail to reject the null hypothesis which is that there is no statistical difference between peoples sentiment before or after the protests. Furthermore the sentimental analysis suggests an opposite relationship with sentiment and protest then I hypothesized, though to no statistical significance

Closing remarks

The results of this experiment are surprising, but they are a welcome surprise. In order to find statistical significance in the result I would require an even larger sample size. I believe that these results can be generalized out of the dataset to some degree, but there are a couple issues that could be rectified: The twitter user base is younger on average than other social medias such as facebook. Young people are generally more progressive than their older counterparts and thus would be more likely to support the George Floyd protests. If I wanted to capture my parents age group, a facebook dataset would be better and it would be interesting to see the results. Limiting the tweets which are processed to only the USA would also raise the integrity of the results. Also in the 2 weeks prior to the protests there were 54 instances of people talking about protests in the dataset, but in the 2 weeks following there were 250 which is neat, but not relevant to my conclusions.

With the results of the test suggesting an opposite relationship then I expected it is also important to consider that my own bias from the family I was locked down with could limit my perspective on how I believed the general public reacted to the George Floyd protests. While I would have felt ‘smart’ if I discovered a statistically significant negative change to peoples discussion around protests, I am overall glad that my preconception was wrong.