[redacted]

Math 1223

Dr.Jacob

01/17/2022

USA protests 2020

The data set that I chose to review is protests that happened in the USA during the year 2020. The data displays the protest, where it happened, what kind of protest it was, along with the date of the event. This data was made available to me via GitHub. Here is the link for the data. https://github.com/nurfnick/Data_Sets_For_Stats/blob/master/CuratedDataSets/USA_2020_Dec12.xlsx. 2020 was a very traumatic year for many civilians involved in standing up for what they believe in. Variables I have found interest in and that I will be using in my analysis for the categorical variables are the type of protest as well as what movement they were supporting, and possibly why. The quantitative variables that are going to be used will be the number of participants, and how many fatalities occurred as a result of the protest. I am interested in this data because during the year 2020, there were various protest that were happening for multiple reasons, but all I ever happened to hear about was the Black Lives Matter movement. Therefore, I am anxious to make the discovery of just how many protests occurred in the year of 2020 and were not related to the Black Lives Matter movement. I will say I am researching this data in hopes of discovering if the location has any correlation with what was being protested and why or why not.

EVENT_DATE	YEAR	TIME_PRECISION	EVENT_TYPE	SUB_EVENT_TYPE	ACTOR1	ASSOC_ACTOR_1
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	Cancel the Rents Movement
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	

01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	Cancel the Rents Movement
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	Labour Group (United States)
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	Health Workers (United States)
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	Women (United States)
01-May-2020	2020	1	Protests	Protest with intervention	Protesters (United States)	Cancel the Rents Movement
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	Christian Group (United States)
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	Health Workers (United States)
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	Labour Group (United States)
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	Labour Group (United States)
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	Cancel the Rents Movement
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	Labour Group (United States)
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	Labour Group (United States)
01-May-2020	2020	1	Protests	Protest with intervention	Protesters (United States)	Health Workers (United States)
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	Health Workers (United States); NNU: National Nurses United
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	
01-May-2020	2020	1	Protests	Peaceful protest	Protesters (United States)	

Project Part 2

For part two of the project, I am looking at the frequency and relative frequency of the protests. The table below shows the data I collected. The frequency and relative frequency of the categorical variables regarding types of protests and what movement was being protested for. I have provided the findings below. In Table 2.1 we can see which type of protest had the highest percentage, as well as we can see which movement had the highest frequency. Turning our attention to table 2.2 we can see exactly what type of protest corresponded with a specific movement.

Table 2.1

Frequency	frequency	Relative frequency	Relative frequency
peaceful protest	17048	0.9208664182	92%
protest with intervention	657	0.0354885756	4%
arrests	14	0.0007562254	0%
attack	52	0.0028088370	0%
Violent demonstration	742	0.0400799438	4%
total:	18513	1.0000000000	100%
cancel the rents movement	61	0.007194245	1%
labour Group	578	0.068168416	7%
Black lives matter	7644	0.901521406	90%
Police Forces	26	0.003066399	0%
health workers	170	0.020049534	2%
total:	8479	1	100%

TABLE 2.2

	Cancel the Rents Movement	Labour Group (United States)	BLM: Black Lives Matter	Police Forces	Health Workers (united states)
Peaceful protest	54	0	0	0	0
Protest with intervention	6	3	356	0	2
Arrests	0	0	0	0	0
Attack	0	0	0	0	0
Violent Demonstration	1	1	454	0	0
total:	61	4	810	0	2

Project Part 3

For part three of this project, I will be looking at the quantitative variable regarding latitude of the protests that occurred during the year of 2020. Throughout research, I was able to make the conclusion that the middle of the United States is located at a latitude that is around 38. I will be using all of the data given to me to later discover if more protests happened in the North compared to the South by using the latitude of the protests.

Table 3.1

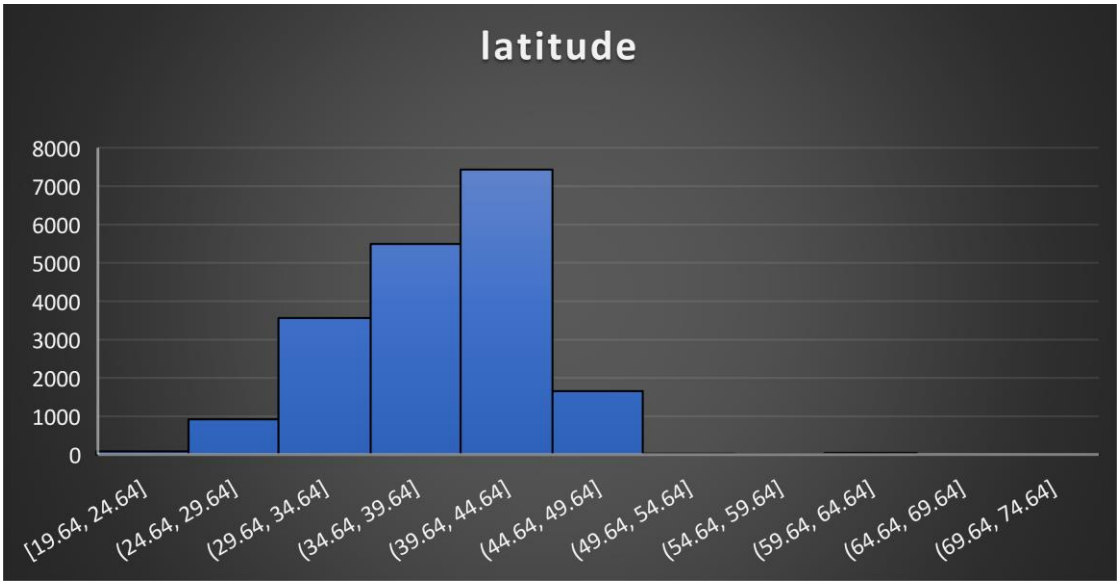
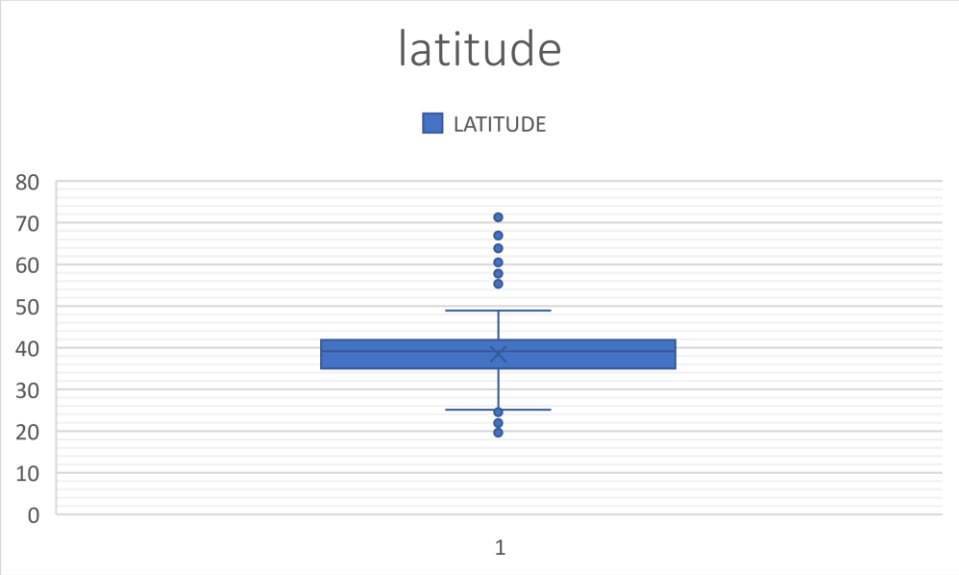


Table 3.2

Summary	Result
Min:	19.64
Q1:	34.993
Q2:	39.183
Q3:	41.85
Max:	71.291
Mean:	38.4706
Range:	51.65
Std. Dev:	5.1615

Table 3.3



Project Part 4

For this next part of the project, I will be attempting to design a hypothesis test for both, a quantitative and categorical variable. The hypothesis over the two can be found below. I will be questioning and analyzing on what latitudinal mark some of the protest occurred on, as well as what types of protests will occur the most within the span of the protests during the year of 2020.

Quantitative Hypothesis:

My null hypothesis for the quantitative variable is that the latitudinal degree of where the protests occurred during the month of May will be equal to 38. The alternative hypothesis is that the data will show results that it is greater than 38. Constructing these tests will hopefully tell me whether or not more protests occurred in the North or the South.

H_0: μ=38

H_a: μ > 38

Categorical Hypothesis:

My categorical variable is the types of protests. Looking at all the protests, I am predicting that "peaceful protests" have the highest percentage of the data. I am also predicting that the cases will have 80 % counting for peaceful protests. My alternative hypothesis is that peaceful protest will count for less than 80%.

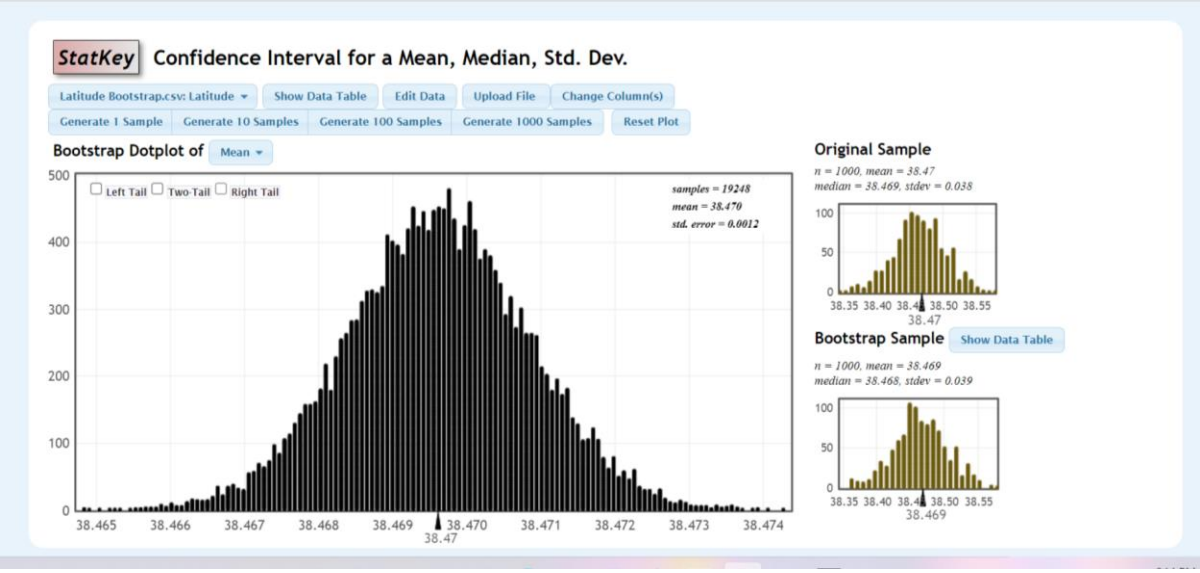
H_0: p=0.80

H_a: p < 0.80

Project Part 5

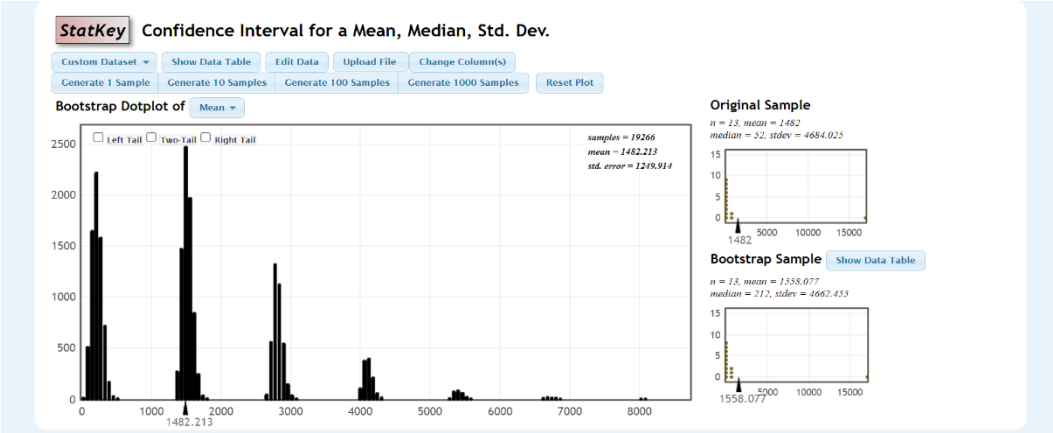
Regarding the fifth part of this project, I will be testing the data using bootstrapping, to see if anything correlates with my hypotheses. Both, the quantitative and the categorical hypotheses will be tested with hopes of the discovery as to which hypothesis is correct, the null or the alternative.

Quantitative Variable: Table 5.1

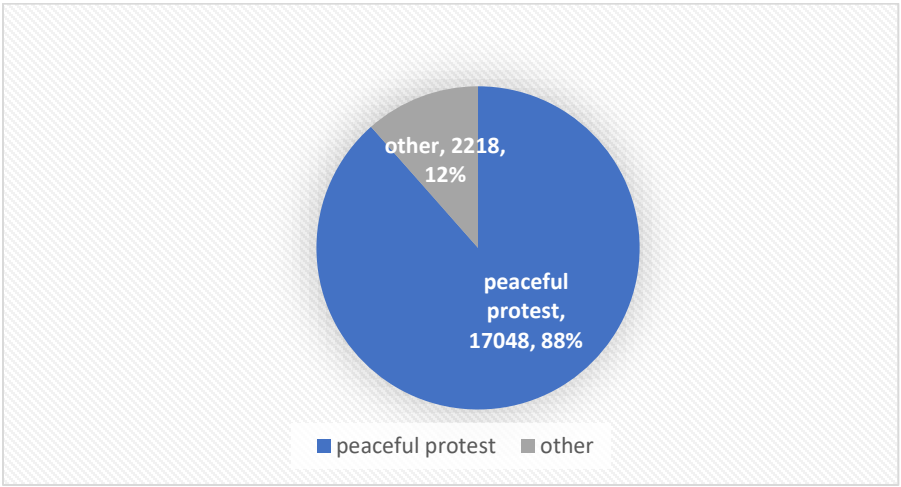


My quantitative variable was tested for the latitudinal degree of where the protests occurred. With a null hypothesis stating that the average latitude would be equal to 38, and an alternative hypothesis of it being greater than 38. With the average computing to 38.4695 we can reject the null hypothesis and accept the alternative hypothesis. We would accept the alternative hypothesis because the alternative hypothesis states that the average would be greater than 38, which it is. To further explain, by having a number of 38.4695 that's not 38.00 which means it is greater than 38. Using a 95% confidence interval level, one would be able to conclude that the findings were 38.46 (lower) and 38.54 (higher).

Categorical variable :Table 5.2



My categorical variable was testing the percentage of types of protests, the peaceful protest in particular. Judging off the bootstrapping distribution we can conclude that we must reject our null hypothesis due the fact that the null hypothesis stated that the percentage of peaceful protest would count 80% of the cases. The total number of cases looked at was the 19,266. I first, analyzed the data by categorizing all the protests, I initially used a bar-graph just to show myself what protests had what total number. Then to finalize it, I plugged those numbers into a pie chart to show the percentages of what makes up the protests. To summarize my findings, of the protests, peaceful protests make up 88.48 % of the types of protests. It is for this reason that we must reject the null hypothesis because the null hypothesis stated that, peaceful protests counted for 80% of the selected cases. However, our alternative hypothesis would be incorrect as well, in the idea that it stated that peaceful protests would count for less than 80% which it does not, therefore we cannot accept the alternative hypothesis either. As far as 95% confidence interval, one can conclude the findings to be between 0.8803(low) and 0.8893 (high). In my opinion when looking at percentages, it is between 88.03%(low) and 88.93% (high).



Project Part 6

For this part of the project, I will be repeating the hypothesis test on the categorical value while implementing the appropriate formulas. These formulas will be computing data for the necessary variable, which is "what percentage of protests accounted for peaceful protests?"

Statistic	***Formula***	***Result***
N	Sample size:	19266
X	number of times (PP) appears:	17048
Proportion	(In Hypothesis):	0.8
P-value	NORMS.S. DIST(Z,true):	0.000446
P-Hat	X/N:	0.88487491
Standard Error	SQRT(p*(1-p)/N:	1270.798
Z-statistic	(p-hat -p)/SE:	3.32321
Z*	NORM.S.INV:	1.96
CI Low	p-hat-(2*SE)	0.8803
CI High	p-hat+(2*SE)	0.8893

To analyze this data, we see that our null hypothesis as well as our alternative hypothesis was incorrect. Due to the idea that it stated that the data computed would be less than 80%, which it is not. Looking at the data above, we see that from the bootstrapping test, our 95% confidence interval is reading at 0.8803 or 88.03% (low) and 0.8893 or 88.93% (high). In addition to testing our hypothesis while using bootstrapping we must reject both, as a result of both hypotheses were proven to be wrong.

Project part 7

For this part of the project, I will be re-analyzing my quantitative variable by testing them with the formulas below. These formulas will help in the production of drawing a conclusion for the quantitative variable data.

statistic	***formula***	***result***
MU	(IN HYPOTHESIS)	38
X-Bar	(Sum of numbers)	38.4706
Sigma	(Standard Deviation)	5.1615
N	(Sample Size)	19248
Standard Error	Sigma/SQRT (n)	0.0012
T- statistic	X-Bar-MU/SE	12.6493
Alpha	For 95% Confidence Interval	0.05
T*	T.INV(alpha, n-1)	1.644933
CI Low	X-bar-T-critical Value * SE	38.46
CI High	X-bar+ T- critical Value * SE	38.54

To analyze the data for our quantitative hypothesis, we see that the confidence interval is displaying 38.46 (low) and 38.54(high). As a result of these data readings, we can conclude that the idea to accept the alternative hypothesis is correct. Even while observing the bootstrapping test results along with all the other data computed, the alternative hypothesis states that the latitudinal reading is greater than 38, which it is. The null hypothesis states that it would be equal to 38, because of this, it is correct to reject the null hypothesis and accept the alternative hypothesis.

Project Part 8

For this final part of the project, I will be utilizing the two-way table from part two to calculate for a conditional probability. I am going to be determining the probability of peaceful protests, given that they were supporting the "cancel the rents movement"

P (A | B) = P(A∩B) / P(B)

	Cancel the Rents Movement	Labour Group (United States)	BLM: Black Lives Matter	Police Forces	Health Workers (united states)
Peaceful protest	54	0	0	0	0
Protest with intervention	6	3	356	0	2
Arrests	0	0	0	0	0
Attack	0	0	0	0	0
Violent Demonstration	1	1	454	0	0
total:	61	4	810	0	2

Observing the data, we see that there are 54 peaceful protests supporting the cancel the rents movement as opposed to the total of 61 protests that were supporting the movement.

P= 54/61~88%

Observing the data to determine the probability of protests with intervention given that they are supporting the Black Lives Matter movement. Resulting from the data, we can conclude that 356 protests with intervention were in support of the Black Lives Matter movement compared to the total number of 810 protests that were in support of the Black Lives Matter movement.

P=356/810~43%

To conclude our findings, one can agree with the statement of: "of the 61 protests that occurred for the "Cancel the rents" movement, 54 of them were peaceful protests". That tells us that approximately 88% of the cancel the rents movement protests were peaceful protests.

In addition to the first conditional probability, to conclude the second conditional probability findings one could agree with the statement that: "of the total number 810, 43% of that, which is 356, which counts for the Black Lives Matter movement." To explain that in a simpler way, there were a total of 810 protests that occurred in support of the Black Lives Matter movement. Out of those 810 protests, 356 of them were classified as protests with intervention. Which is approximately 43% of the Black Lives Matter movement.

Another conclusion to make would be the idea that when it comes to the cancel the rents movement, at random there is 88% chance that a protest will be peaceful. Looking at Black Lives Matter on the other hand, at random there is a 43% chance that a protest will involve intervention.