Ashley Pittser

Department of Math, East Central University

MATH 1223: Intro to Statistics

Dr. Nicholas Jacob

January 15, 2022

US Mass Shootings Statistics

The data set I have picked provides information on mass shootings that have occurred in the US from 1966 to 2017. The information is provided from the data set found in Dr. Jacob’s GitHub, <<https://github.com/nurfnick/Data_Sets_For_Stats/blob/master/CuratedDataSets/massshootings.csv>>. The data in the set has 398 different cases of recorded mass shootings, and I believe it covers all of the shootings in the US, meaning the data is a population set. There are eight variables in the data set, four quantitative and four categorical. The four quantitative variables in the study are date, number of fatalities, number of people injured in the shooting, and the total number of victims. The categorical variables are the location if there were mental health issues with the shooter, the race of the shooter, and the sex of the shooter.

Table

Description automatically generated with medium confidence

As you can see from the clip, the set includes all varieties of shootings that have occurred in the US. In high school for an extracurricular club, I did a large report on the Las Vegas Shooting, and ever since this project I have been interested in the statistics behind shootings. I would like to see if there is any real data behind the idea that white male shooters are more common than other shooters. I would also like to see more about the mental health aspect of the data set. Many of the cases either have a clear indication of mental health problems or are unclear, but very few are a definite no to the mental health variable. I find this interesting and would like to learn more about it, and I believe this case study will help me do that.

Part 2

For Part 2, I will be looking at the frequency and relative frequency of the races of the data set. I will determine the frequency and relative frequencies of each type of race found in the sample by looking at these cases.

|  |  |  |
| --- | --- | --- |
| Race Types | Frequency | Relative Frequency |
| White American/European | 186 | 0.4673366834 |
| Black African/African American | 93 | 0.2336683417 |
| Asian | 30 | 0.0753768844 |
| Latino | 5 | 0.0125628141 |
| Native American | 6 | 0.0150753769 |
| Unknown | 47 | 0.1180904523 |
| Two or More Races | 3 | 0.0075376884 |
| Other Races Not Listed | 28 | 0.0703517588 |
| Total: | 398 | 1 |

Looking at this sample of shootings, it is easy to see that White American/European and Black African/African American shooters are more common than any other kind. White Americans/Europeans are the most common of shooters, with about 46% of shooters coming from this group. However, many shooters’ ethnicities were unknown or unlisted for this data set. For my two-way table, I am doing a comparison between the race and mental health issues variables. I hope this data will provide some more information on if there is a connection between mental health and shootings.

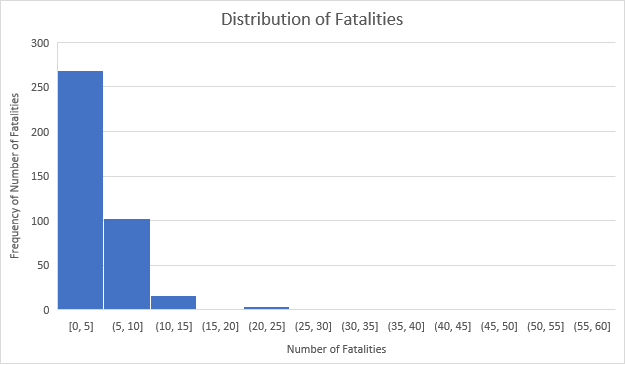
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Race | Mental Health Issues | No Mental Health Issues | Unknown Mental Health Issues | Total |
| White American/European | 80 | 54 | 52 | 186 |
| Black American/ African American | 32 | 27 | 34 | 93 |
| Asian | 12 | 8 | 10 | 30 |
| Latino | 1 | 4 | 2 | 7 |
| Native American | 3 | 1 | 2 | 6 |
| Unknown | 7 | 7 | 31 | 45 |
| Two or More Races | 0 | 1 | 2 | 3 |
| Other Races Not Listed | 5 | 8 | 15 | 28 |
| Total | 140 | 110 | 148 | 398 |

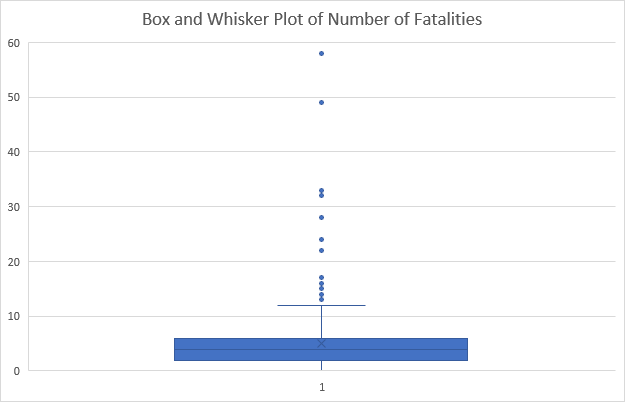
Looking at the data, I do not think you can state that there is some connection between mental health and shootings. Many of the shooters have unknown mental health statuses, with about 37% of the sample health status being unknown. Thirty-five percent of the set does have a mentally ill shooter and I would estimate that had there been fewer unknowns, there would’ve been a definite relationship between shootings and mental illness. Considering this problem, 43% of White American/European shooters were mentally ill and 34% of African Americans were mentally ill. I do not think a blanket statement can be made about different races having higher numbers of mentally ill, as this data only pertains to shootings, but I think mental health does play some part in shooting occurrences. With so many unknowns, sadly no definite conclusions can be drawn, besides the fact that the majority of shooters were white. I hypothesized that white shooters would be more common than others, and the data is supporting that idea.

Part 3

For part three, I will be looking at the number of fatalities of each shooting. I will find the mean, standard deviation, and the five-number summary. I will then graph the data with a histogram and a box plot.

|  |  |
| --- | --- |
| Mean | 5 |
| Standard Deviation | 5.714723 |
| Minimum | 0 |
| Quartile 1 | 2 |
| Median | 4 |
| Quartile 3 | 6 |
| Maximum | 58 |





Looking at the information and the tables, the first thing I notice is that the data is not very dispersed. The number of fatalities for shootings is quite low, with some exceptions, shown as outliers in the box and whiskers plot. Any number of fatalities over twelve was considered an outlier and out of 398 shootings, only twelve shootings had any more than 12 fatalities. The next thing I can see is that the data is certainly skewed to the right, with the majority of shootings not having over ten fatalities. Seeing this information, I can conclude that while there is a large number of shootings in the history of the US, very rarely do these shootings have large numbers of fatalities.

Part 4

For project part 4, I will create a hypothesis test for one categorical variable and one quantitative variable. The categorical variable I will test is the gender of the shooters and the quantitative variable I will test is the total number of injured for each shooting.

Categorical variable

For the gender variable, my null hypothesis is that 75% of shooters are men. My alternate hypothesis is that the number of male shooters is not 75%. While many shooters are left unknown in the data set, I still believe that the gender will be mostly male of the known shooters.

: p= .75

: p ≠ .75

Quantitative Variable

For the quantitative variable of the total number of injured, my null hypothesis is that the mean number of injured is twenty-five people. My alternative hypothesis is that the mean is not equal to 25 people.

: µ = 25

: µ ≠ 25

Part 5

For part five I will bootstrap and analyze the bootstrap distribution for my categorical and quantitative variables. Doing this will allow me to find the standard error and confidence intervals, allowing me to reject or accept my null hypothesis.

Categorical Variable

Using the count that 368 out of the 398 shootings had at least one male shooter, I took this number and created a data set using Stat Key. I created a distribution to find the proportion of male shooters in the bootstrapped data set.

Chart

Description automatically generated

Looking at this bootstrapped distribution, it has a normal distribution. The standard error is 0.013. The 95% confidence level has a lower number of .897 and an upper number of .949. Seeing this information, I can reject my null hypothesis, as more than 75% of shooters in the original and the bootstrapped data set are male.

Quantitative Variable

The graphs below are the bootstrapped distribution for my quantitative variable, the total number of injured. The two graphs on the side are histograms of the original and bootstrapped data, the original on top and the bootstrapped data on the bottom. The mean of the bootstrapped set is 11.026. The standard error for this variable is 1.451. The 95% confidence level had a lower number of 8.781 and an upper number of 14.048. Seeing the mean of the bootstrap and the 95% confidence interval, I reject my null hypothesis.

Chart

Description automatically generated

Part 6

For part 6, I will continue testing the categorical hypothesis. I will also compare my bootstrapped 95% confidence interval and the interval I get from this round of testing. Below is a table containing all the necessary information for the test.

|  |  |  |
| --- | --- | --- |
| Statistic | Formula | Result |
| N | Sample Size | 398 |
| X | Number of times the male gender is found in a sample | 368 |
| Proportion | Found in Hypothesis | .75 |
| P-Value | =NORM.S. DIST (Z, True) | .00001 |
| P Hat | X/N | .9246 |
| Standard Error | =SQRT(p\*(1-p)/n | .0217505 |
| Z statistic | = (Phat- p)/SE | 8.045 |
| Z\* | =NORM.S.INV (0.975) | 1.95996398 |
| Confidence Interval Lower | =phat-(2\*SE) | .881099 |
| Confidence Interval Upper | =phat+(2\*SE) | .968101 |

My null hypothesis for my categorical variable test was that 75% of the shooters in the data set were male. The 95% confidence interval for this test was between .881099 and .968101. Looking at this test’s confidence interval, I fail to reject my hypothesis as it falls between the lower and upper confidence intervals. In the bootstrapped example, the confidence interval was between .897 and .949. I did reject my hypothesis for the bootstrapped sample but did not for this one so there appears to be quite a difference between the two test’s results.

Part 7

For Part 7, I will be repeating part 6 for my quantitative hypothesis. I will also again compare my bootstrapped sample results to my results for this part. Part 6 and Part 7 will be extremely similar in setup and process.

|  |  |  |
| --- | --- | --- |
| Statistic | Formula | Result |
| MU | in hypothesis | 25 |
| X-Bar | sum of numbers/n | 6.251256281 |
| Sigma | Standard Deviation | 26.64567745 |
| N | Sample Size | 398 |
| Standard Error | Sigma/SQRT(n) | 1.335627 |
| T Statistic | X-Bar-Mu/SE | -14.0327 |
| Alpha | For 95% CI | 0.05 |
| T\* | T.INV (alpha, n-1) | -1.6488 |
| CI Low | X-Bar-T-Critical Value\*SE | 18.0864987 |
| CI High | X-Bar+T-Critical Value\*SE | 22.4810627 |

Looking at my tests results, my confidence interval has an upper number of about eighteen and a lower number of about twenty-two. I reject my hypothesis of twenty-five for the mean, as twenty-five does not fall in my confidence interval. With my bootstrapped sample, my confidence interval was 8.781 and 14.048, and I also rejected my hypothesis. Unlike my categorical variable, where my bootstrap sample and my test sample had different results for rejecting or not rejecting the hypothesis, my quantitative variable had a rejection for both the bootstrapped sample and the test sample.

Part 8

For my final part, I am going to look back at my part 2 two-way table and create two conditional probabilities, using the formula provided.

Conditional Probability Formula:

Text

Description automatically generated

Two-Way Table from Part 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Race | Mental Health Issues | No Mental Health Issues | Unknown Mental Health Issues | Total |
| White American/European | 80 | 54 | 52 | 186 |
| Black American/ African American | 32 | 27 | 34 | 93 |
| Asian | 12 | 8 | 10 | 30 |
| Latino | 1 | 4 | 2 | 7 |
| Native American | 3 | 1 | 2 | 6 |
| Unknown | 7 | 7 | 31 | 45 |
| Two or More Races | 0 | 1 | 2 | 3 |
| Other Races Not Listed | 5 | 8 | 15 | 28 |
| Total | 140 | 110 | 148 | 398 |

My first conditional probability will be the probability of no mental health issues as a Black American/African American shooter.

**P=27/93=.29**

I computed this using the formula. The numerator of the formula was found by finding the number of Black American/African Americans that had no mental health issues, which was twenty-seven. I found the denominator of the formula by finding the total number of Black American/African Americans, which was ninety-three. I then did the calculation of 27/93 and got .29, meaning 29%. This probability means that there is a 29% chance that Black African/African American shooters did not have mental health issues, meaning more Black African/African American shooters committed the shooting with mental health issues than no mental health issues.

My second conditional probability will be the probability of mental health issues as a White American/European.

**P=80/186=.43**

I again found this using the formula. The numerator was the number of White American/European Shooters who did have mental health issues and the denominator was the total number of White American/European shooters. This probability means that 43% of shootings done by White Americans/Europeans were done by people with mental issues. Again, I feel that the data is affected by the number of shooters who had unknown mental health statuses. I feel the probabilities would be different if more mental health statutes were known.