Mass Shootings from 1982 – 2019

[redacted]

Class: SU20-INTRO TO PROBABILITY AND STATS_01
Instructor: Nicholas Jacob

I will examine the data from the mass shootings done from 1982 to 2019. The data can be found here https://www2.stetson.edu/~jrasp/data.htm. The original data set was not arranged by year, but I wanted to make it easier to read as the mass shootings in chronologic order. There was some additional data that was repetitive and not useful for my analysis, so I cleaned that data out.

The categorical variables are the cases, location (state/city), date, summary of the shooting, location (where it happened), prior signs of mental health, mental health details, were the weapons obtained legally, weapon type, race, gender, and type of shooting. The quantitative variables are the number of fatalities, age of the shooter, the number of injured, and total victims. Below is an image of four cases of my data set.

case	location	date	summary	fatalities	injured	total_victims	location	age_of_shootei	prior_signs_m	mental_health_details	weapons_obtai	weapon_type	race	gender	type
Welding shop shooting	Miami, Florida	8/20/1982	Junior high school tea	8	3	11	Other	51	Yes	His second wife left him becau	Yes	One shotgun	white	Male	Mass
Dallas nightclub shooting	Dallas, Texas	6/29/1984	Abdelkrim Belachheb,	6	1	7	Other	39	Yes	During his last meal with his w	No	One semiautomatic ha	white	Male	Mass
San Ysidro McDonald's massacre	San Ysidro, California	7/18/1984	James Oliver Huberty,	22	19	41	Other	41	Yes	The day before the shooting, h	Yes	One semiautomatic ha	white	Male	Mass
United States Postal Service shooting	Edmond, Oklahoma	8/20/1986	Postal worker Patrick	15	6	21	Workplace	44	Unclear	He was worried he had inherite	Yes	Three semiautomatic	white	Male	Mass

What I am most interested in learning from these mass shootings that took place during this timeframe is how many of these shooters had prior signs of mental illness. If the data shows that most of these shooters had some kind of mental illness beforehand, it would suggest that we need to pay closer attention to those who have a mental illness, especially if they show any signs of hostility towards classmates or fellow workers. If having mental problems is one of the main signs that could cause a mass shooter, then we could develop better programs that would help these individuals to have the tools they need to not become the next mass shooter.

I chose to make two tables that would show the frequency and relative frequency of those mass shooters who had prior signs of mental illness and who bought their guns legally. Those who had a mental illness prior to the shootings are significantly higher than those who did not. Since my table includes forty-one shooters whose prior mental condition is unknown, it makes the relative frequency of those who had prior mental illness appear to be lower. If we removed the unclear, the relative frequency would about 80%, which indicates that mental illness is a key factor in those who would commit a mass shooting.

As you look at the second table, about 71% of the weapons they used were obtained legally. If you removed the unclear accounts, about 84% of the guns were bought legally.

Our third table, which is a two-way table, show a correlation between having a mental illness prior to the shooting and having guns legally. This correlation would suggest that if we want to help reduce the number of mass shootings, then we need to have stricter gun rules for those with mental illnesses. However, we need to keep in mind that having a mental illness doesn't mean a person will become a mass shooter or harm anyone. So, creating restrictions would need to be done fairly. Perhaps further study might show that certain kinds of mental illnesses might be correlated with mass shootings.

We will never be able to stop all mass shootings, but the data I have composed indicates that we could decrease the number of shootings by having more substantial restrictions on owning guns if people have mental illnesses. If the parents have guns legally with no mental illness but their child is mentally ill, then they should secure their weapons, so the child has no access to them.

Prior signs of mental illness

Response	**Frequency**	**Relative Frequency**
Yes	59	0.504273504
No	17	0.145299145
Unclear	41	0.35042735
Total	117	1

Weapons obtained legally

Response	**Frequency**	**Relative Frequency**
Yes	83	0.709401709
No	16	0.136752137
Unclear	18	0.153846154
Total	117	1

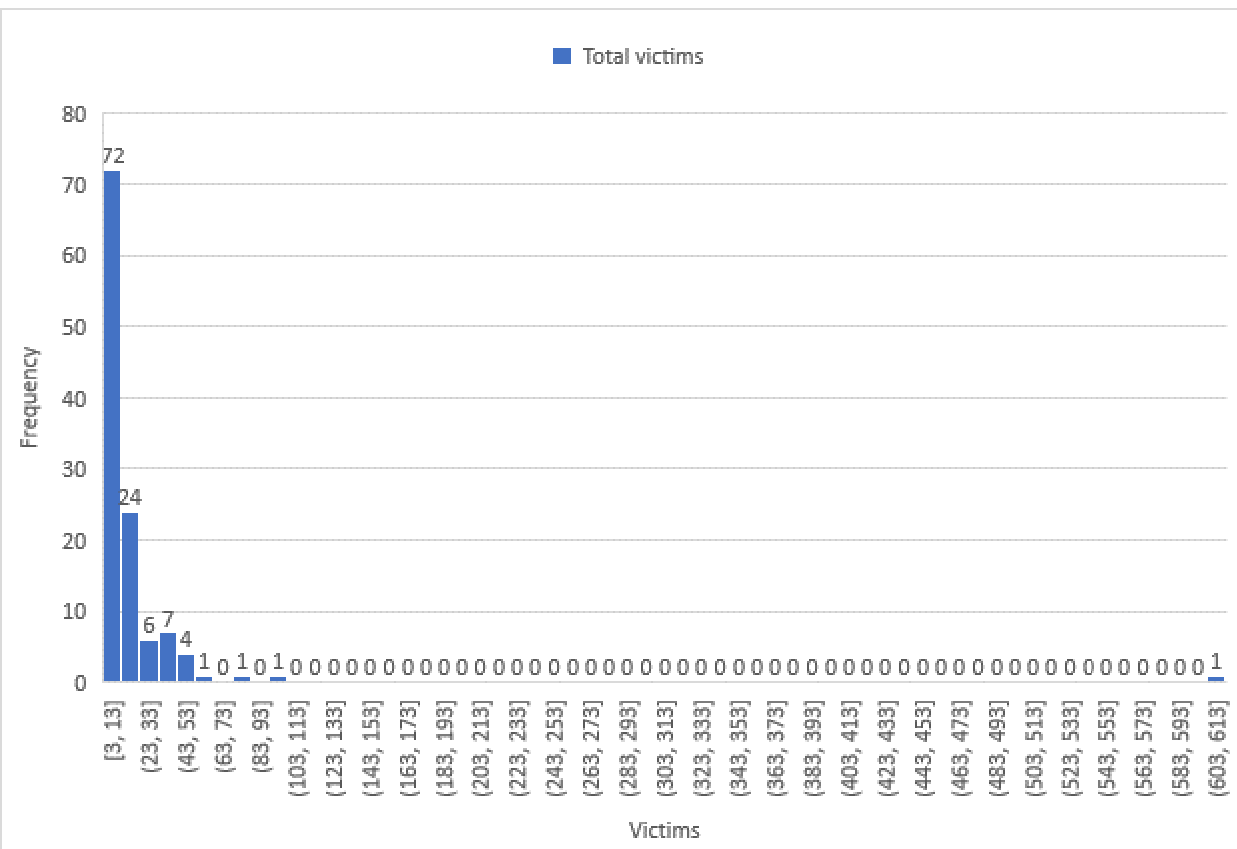
Two-way table

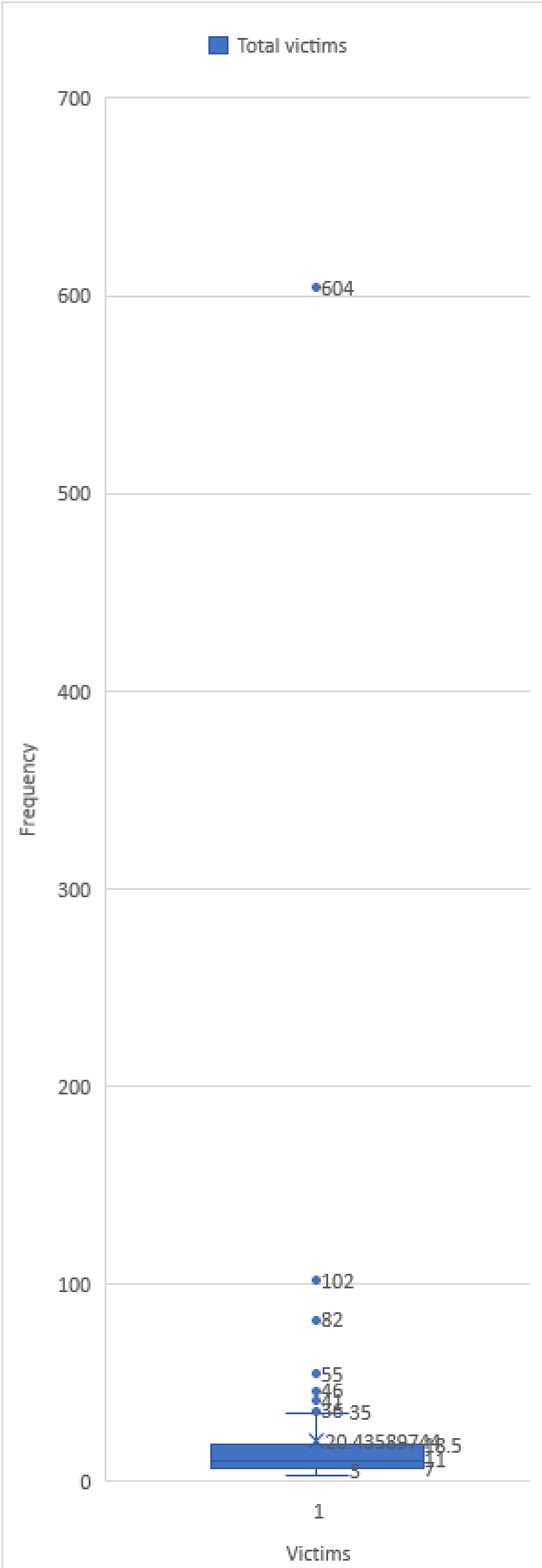
	Weapons obtained legally	**Weapons obtained illegally**	**Weapons obtained unclear**
Prior signs of mental illness	46	8	5
No prior signs of mental illness	12	4	0
Unclear prior signs of mental illness	25	3	14

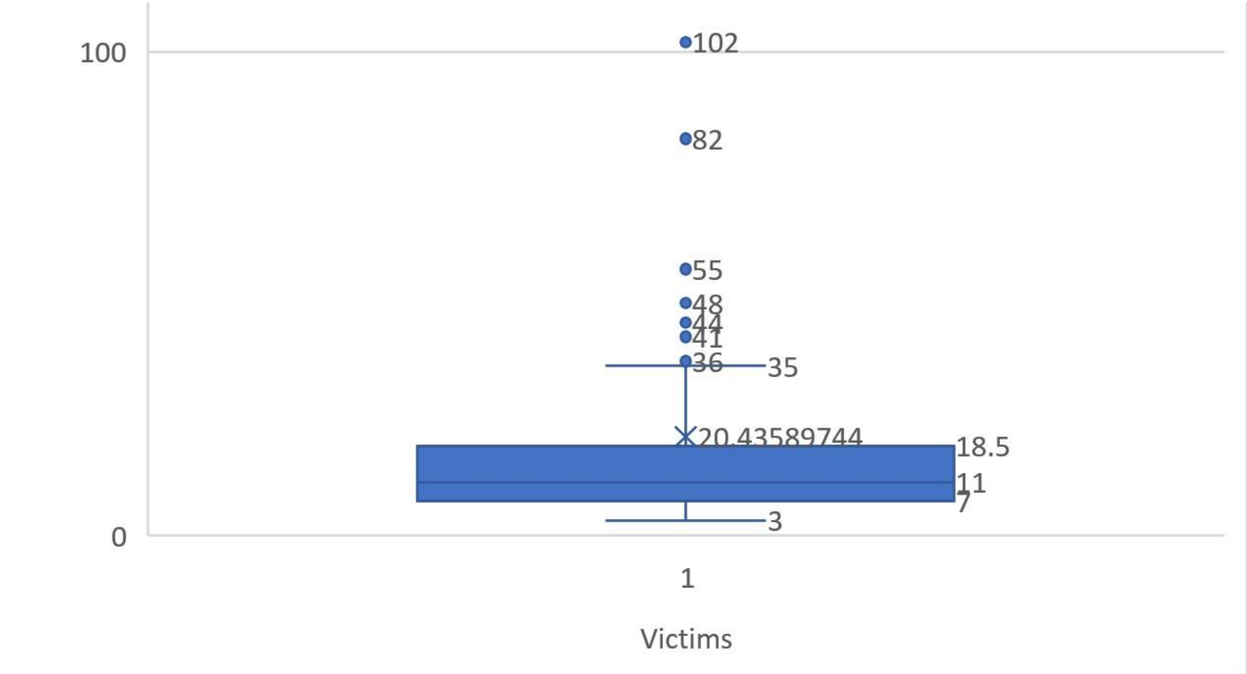
Next, I will examine the statistical summary of the total number of victims in these mass shootings, and I will provide a histogram and a box plot. I also included a close up of the main data of the box plot to make it easier to see. The charts will show that the distribution is skewed to the right as the mean is greater than the median. Also, you will notice in the box plot that there are several outliers ranging from 36 to 604. If I removed the extreme outlier (604), the mean would change to 15.40517241, which shows how out outliers can inflate the average. As you can see from the histogram, the majority of victims injured or killed during a mass shooting are between three and thirteen. We can conclude that on average that a mass shooting will have between three to thirteen victims that are injured or killed.

Total victims - Summary

Mean	20.43589744
Median	11
Standard Dev.	56.27184141
Variance	3190.248011
Q1	7
Q2	11
Q3	18
IQR	11
Min	3
Max	(604.00)




Total victims
700
600
500
400
300
200
100
0
Frequency
604
102
82
55
46
41
36 35
20.43589744
18.5
11
3
7
1
Victims



Hypothesis test for a quantitative variable

I suspect that if we examine all the mass shootings in my chart that we will discover that there are more injuries than deaths.

Ho: = The mean of the injuries will be > than the mean of the deaths in the mass shootings in my report.

Ha: ≠ The mean of the injuries will be < than the mean of the deaths in the mass shootings in my report.

This will be easy to determine if my hypothesis is supported because all I have to do is compare the mean of the deaths in the chart to the mean of the injuries. If my hypothesis is supported, the numbers will show it, and they do. The mean of deaths is 8.10, and the mean of injuries 12.33. Therefore, my hypothesis is supported. I will add that I based my hypothesis on the fact that when a mass shooting happens, people tend to run as fast as they can, which puts them at a higher risk of injuring themselves. Also, as the shooter shoots into a crowd, he or she is not going to have kill shot every time, which means that there will be more people injured from the stray bullets as well.

Simple formula: (i=injures d=deaths)

Ho: $\mu i = \mu d$
Ha: $\mu i < \mu d$

Hypothesis test for a categorical variable

I suspect that the majority of mass shooters have some form of mental illness and showed signs of mental illness before they committed the mass shooting.

Ho: = The frequency of mass shooters who had prior signs of mental illness is equal to the mass shooters who had no prior signs of mental illness.

Ha: ≠ The frequency of mass shooters who had prior signs of mental illness is < than the mass shooters who had no prior signs of mental illness.

The main reason I suspect the majority of mass shooters would have some form of mental illness is because what person in his or her right mind would decide to kill as many people as he or she can. We can easily determine if the hypothesis is supported by looking at the frequency table below. It clearly supports my hypothesis because 59 cases had prior signs of mental illness, and 17 did not.

Prior signs of mental illness

Response	**Frequency**	**Relative Frequency**
Yes	59	0.504273504
No	17	0.145299145
Unclear	41	0.35042735
Total	117	1

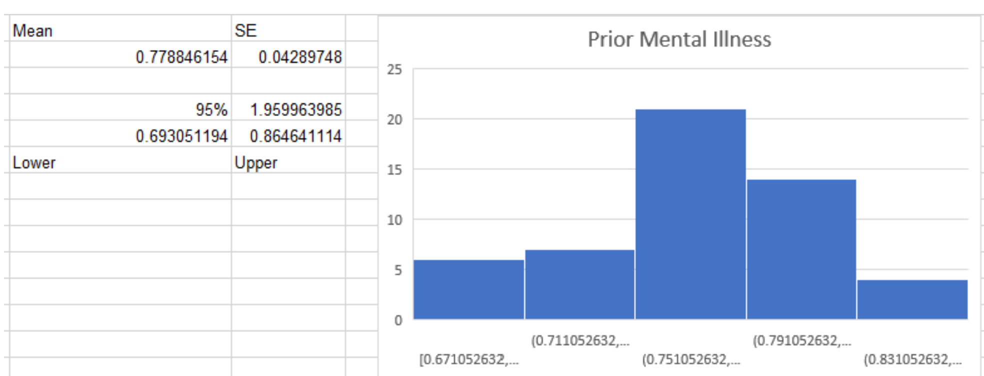
Simple formula:

Ho: p Mental Illness=1/2

Ha: p Mental Illness >1/2

Bootstrap test for my categorical variable

To test my hypothesis further, I used bootstrapping with 52 generated samples to determine if my hypothesis should be accepted or rejected. I discovered that my bootstrap mean was .778846154, and the standard error is .042. Computing the 95% confidence level with a z score of 1.96, I found the range to be .6930511954 to .864641114, which means that I can be 95% confident that the majority of mass shooters had some previous signs of mental illness. So, I failed to reject my hypothesis.

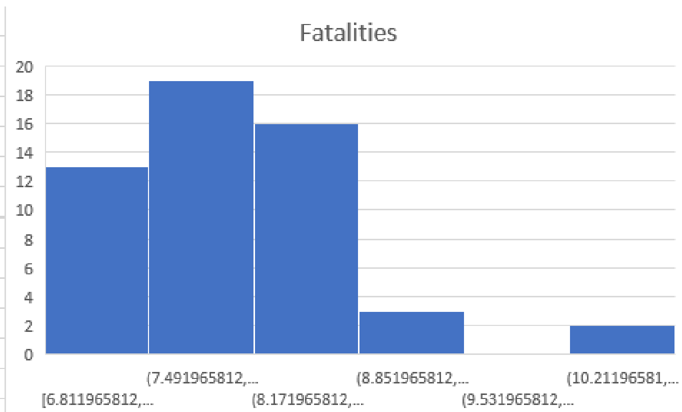


Bootstrap test for my quantitative variable

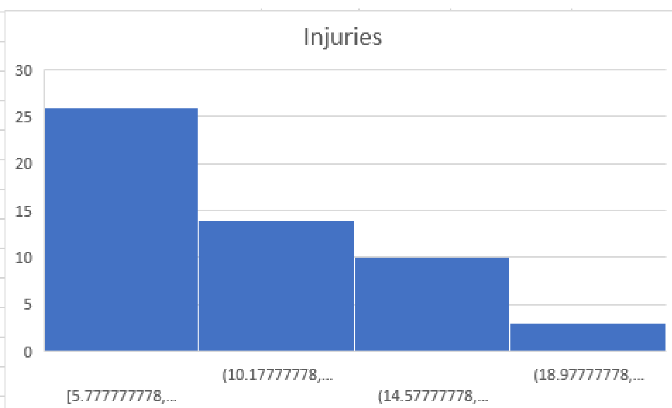
To test my hypothesis further, I used bootstrapping with 52 generated samples for the injured and the fatalities to determine if my hypothesis should be accepted or rejected. I discovered that my bootstrap mean for the fatalities is 8.010804709, and my bootstrap mean for injuries is 11.43202709. The standard error for fatalities is .728691332 and 4.73563476.

If my hypothesis were being tested by the bootstrap mean, I would have failed to reject my null hypothesis because the number of injuries is greater than the number of deaths. However, based on the 95% confidence level, I must reject the hypothesis because, based on the compared ranges for fatalities and injuries, it is possible for the deaths to be greater than the injuries because there is a much greater range with the injury numbers. As you can see in the charts below, the fatalities range between 6.553422046 and 9.468187372. The injuries range from 1.960757572 to 20.90329661.

Mean	SE
8.010804709	0.728691332
95%	1.959963985
6.553422046	9.468187372
Lower	Upper



Mean	SE
11.43202709	4.73563476
95%	1.959963985
1.960757572	20.90329661
Lower	Upper



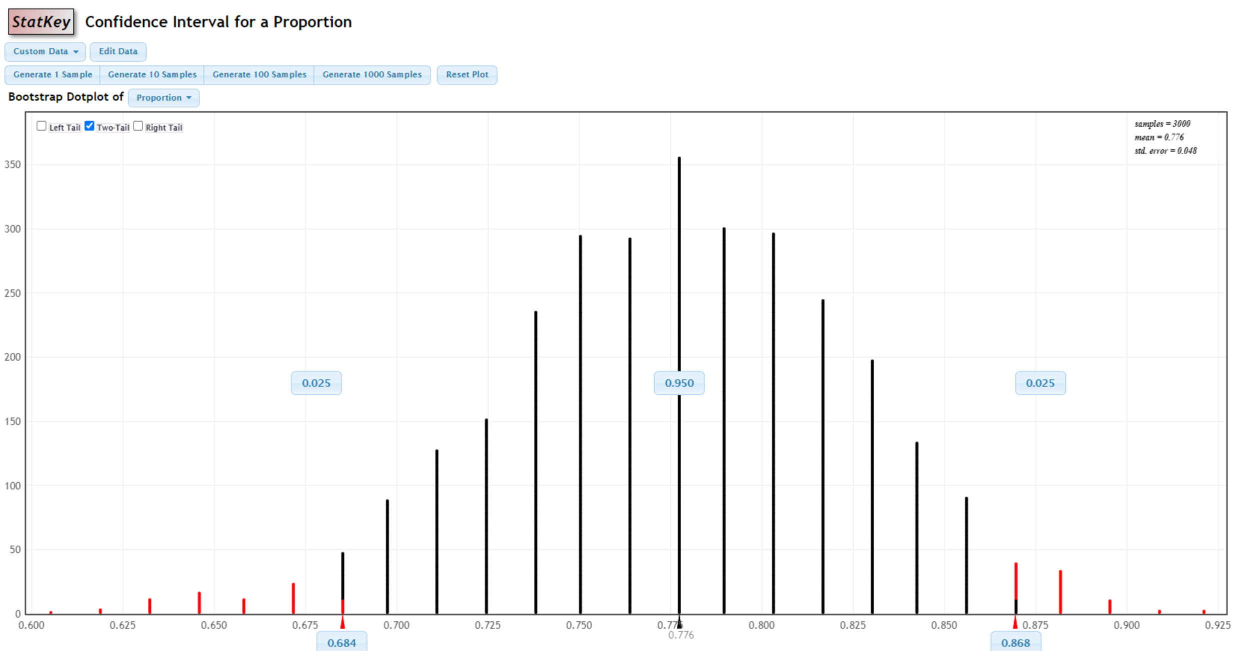
From Module 6, I will use the formulas we learned to test my hypothesis again. Let's begin with our categorical data the records if the shooter had prior signs of mental illness.

Simple formula:

Ho: p Mental Illness=1/2

Ha: p Mental Illness >1/2

There are 76 shooters, with 59 of them having prior signs of mental illness. I suspect that more than half of the shooters had prior signs of mental illness. I will use an Alpha of .05. My sample size is 76, the proportion is ½, and my statistic is .776315787. The SE is .057353933. Z = 4.817730411 and Z* 1.6448553627. Using StartKey, my 95% CI ranges from .684 to .868. This is almost identical to my initial bootstrap test above, which once again offers support that my hypothesis is true. Using these numbers, I have failed to reject the null hypothesis.



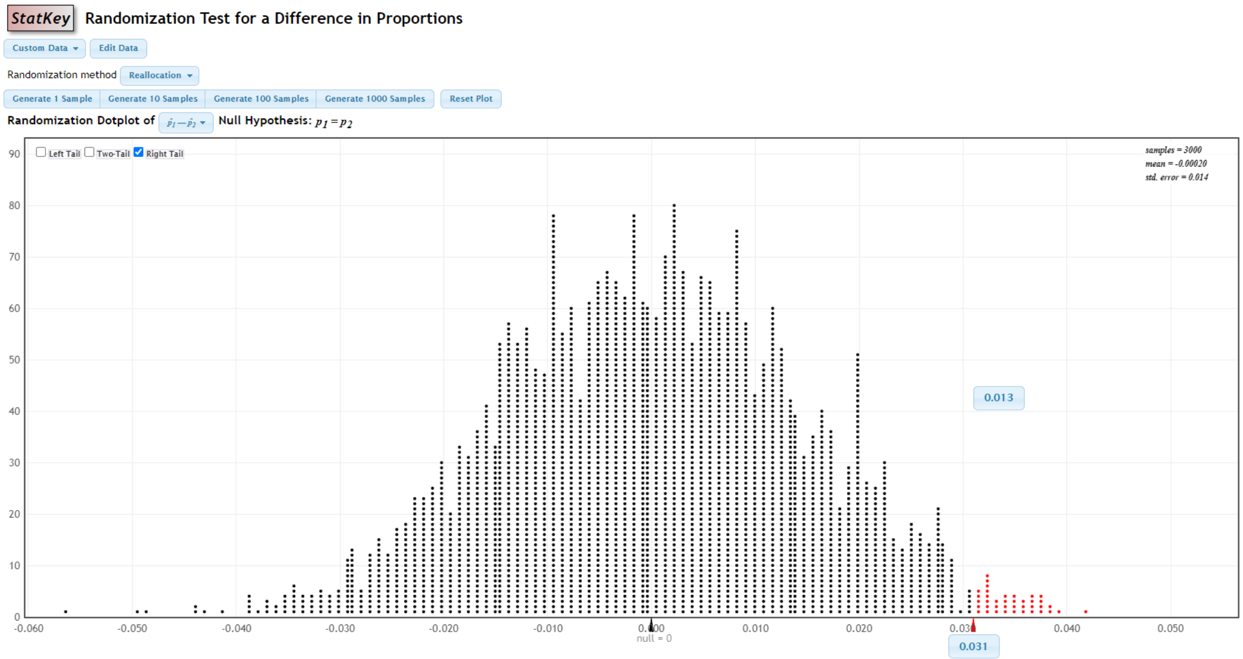
Samplesize	proportion	statistic	
76	0.5	0.776315789	0.05
n	P	p hat	alpha
Simple formula: (i=injures d=deaths)			
Ho: µi = µd			
Ha: µi < µd			
z*	z	SE	
1.644853627	4.817730411	0.057353933	
	(phat-p)/SE	sqrt(p*(p*(1-p)/n	

Next, let's examine the quantitative data using the new formula from Module 6. I am very interested in how this will compare to the previous module.

Simple formula: (i=injures d=deaths)

Ho: µi = µd
Ha: µi < µd

We have a sample size of 2391. Out of the 2391, there were 948 fatalities and 1443 injuries. The SE = 0.014460896. P1-P2 = -0.20702635. Z = -14.31628741. P = 8.6564E-47. I have included a snapshot of Excel numbers. However, when I put this problem into StartKey, you can see from the chart below that is only a .013 chance that the deaths and injuries will be different from one another. However, from the Excel nonpooled numbers, the 95% Cl difference can range from -0.234755148 to -0.17929755. Though this formula gave me different numbers from the previous example above, I would still have to reject the null hypothesis. Also, my instructor told me to include the following Cl numbers which range from -0.027728799 and 0.027728799.



Fatalities	Injuries		
2391	2391	4782	
948	1443	2391	
0.396486826	0.603513174	0.5	
p1	p2	Pooled Proportion	
Pooled SE	p1-p2		
0.014460896	-0.20702635		
z	p	tail	two times
-14.31628741	8.6564E-47	1	1.73E-46
		0	
Not pooled SE			
0.014147606			
z*			
1.959963985			
95% CI			
-0.234755148	-0.17929755		

In this section, I am only retesting the qualitative part of my hypothesis with a test. Using the formula we learned this week, I came up with the following numbers.

Module 7	
Average	Standard Dev.
4.230769231	45.9914506
SE	
0.940561764	
Tstat	
4.498130153	
Postive	
0.034017483	
CL	
4.198773687	4.262764774
Lower	Upper

If you compare these numbers to my previous test, they are quite different. With such a high confidence level, I will accept my null hypothesis. I amazed at how different each statistical test is. My question would be, which is the most reliable method?

For the last part of our project I will provide two conditional probabilities from a modified version of my two-way table that was done earlier in this project.

Two-way table			
Column1	**Weapons obtained legally**	**Weapons obtaintaid illegally**	**total**
Prior signs of mental illness	39	0	39
No prior signs of mental illness	11	4	15
Total	50	4	54

How many shooters had a prior sign of mental illness? This would be determined by dividing 39/54, which would = .72. In other words, 72% had prior signs of mental illness.

How many shooters had no prior signs of mental illness and obtained weapons illegally? This would be determined by dividing 4/54 which would = .074. In other words, 7.4% had no prior signs of mental illness who had obtained weapons illegally.

Please refer to the following formula that is to be included in this final project.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

I have enjoyed examining the statistics of my project. I am thankful for the opportunity to have had a glimpse it to what goes into looking at projects like mine.

You can view my edited Excel file from this link
https://www.dropbox.com/s/3xezqdqoth3v17q/Mother%20Jones%20-%20Mass%20Shootings%20Database%5EJ%201982%20-%202019%20revised.xlsx?dl=0