
ISyE 2028 – Basic Statistical Methods - Fall 2015

Bonus Project: “Big” Data Analytics

Final Report

Chris Noerjadi

ISYE 2028

Professor Yao Xie

Bonus Project

November 24, 2015

US Veterans Suicide Data

Abstract

A 2013 article in News21 stated that the annual suicide rate among veterans is more than double the rate of the civilians population. The article further stated that record from 48 states showed the annual suicide rate among veterans to be about 30 for every 100,000 of the population. In comparison, the civilians suicide rate was cited to be about 14 per 100,000 (Hargarten, Burnson, Campo & Cook, 2013). The obvious reasoning behind this disparity is that veterans are more prone to suicides because they have more stress burdens, particularly the stress of war. But is that the real reason? I want to explore the reasons behind this proportion and to research the validity of the data.

Methodology

- Data obtained from Individual state health and vital records departments, U.S. Census Bureau American Community Survey one-year estimates spanning 2005 to 2011, and Centers for Disease Control annual national mortality reports.
- 5-numerical summaries analysis using Excel software
- 95% Confidence Interval for Population Proportion with large size sample method.
- Hypothesis Testing using Inference on a Population Proportion method/

For my first objective, I constructed 5 numerical summaries of Veterans and Civilians suicide rate data. Some of the data were outside of the outliers and I omitted it from the calculation as necessary. As you can see from the data, veterans suicide rate numbers are much higher than the civilians counterpart.

Vet Suicides Rate - 5 point indicator	
Min	12.3
Max	50.6
Median	31.5
Average (mean)	32.5
Variance	79.0
Std Deviation (σ)	8.9

(Massachussets)
(Nevada)

Civ Suicides Rate - 5 point indicator	
Min	8.55
Max	23.93
Median	15.24
Average (mean)	15.58
Variance	12.79
Std Deviation (σ)	3.58

(New Jersey)
(Montana)

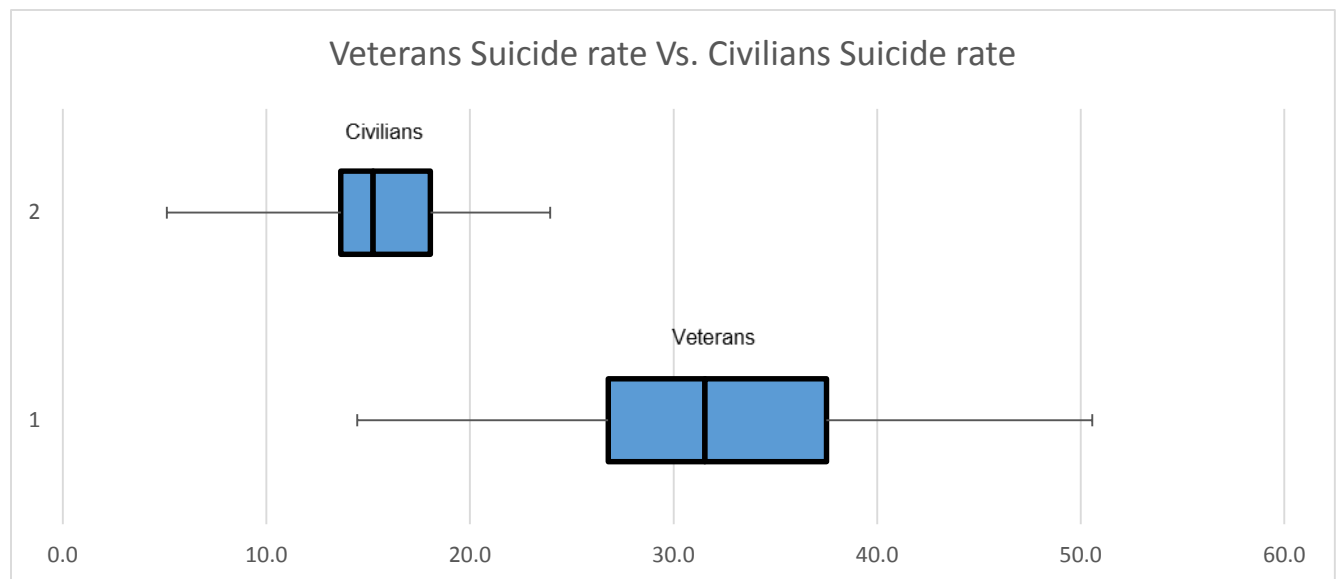
Min	12.3
Q1	26.8
Q2	31.5
Q3	37.5
Max	50.6

Min	8.55
Q1	13.66
Q2	15.24
Q3	18.06
Max	23.93

IQR (Q3 - Q1)	10.72
---------------	-------

IQR (Q3 - Q1)	4.40
---------------	------

*Rates are based per 100,000 people



For my second objective, I wanted to construct two-sided confidence interval with 95% confidence level for the data. Since this data was about suicide rate proportion, I chose confidence interval for a Population Proportion with large sample size method:

$X_1, X_2, \dots \sim \text{Bernoulli}(p)$

Using CLT, we can estimate a confidence interval for the population proportion.

Let \hat{P}

$$\hat{P} = \frac{\sum_{i=1}^n X_i}{n} \sim N\left(p, \frac{p(1-p)}{n}\right) \Rightarrow \Pr\left\{-Z_{\alpha/2} \leq \frac{\hat{P} - p}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}} \leq Z_{\alpha/2}\right\} = 1 - \alpha$$

$$\hat{P} - Z_{\alpha/2} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \leq p \leq \hat{P} + Z_{\alpha/2} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

$$\hat{P} = \frac{31.1}{100,000} = 3.11 * 10^{-4} \quad Z_{\alpha/2} = Z_{.05/2} = 1.96 \quad n = 100,000$$

$$3.11 * 10^{-4} - 1.96 \sqrt{\frac{3.11 * 10^{-4} (1 - 3.11 * 10^{-4})}{100,000}} \leq p \leq 0.1947 + 1.96 \sqrt{\frac{0.1947 (1 - 0.1947)}{230,015}}$$

$$2.02 * 10^{-4} \leq p \leq 4.20 * 10^{-4}$$

Hence, we are 95% confident that p falls somewhere between $2.02 * 10^{-4}$ and $4.20 * 10^{-4}$

Lastly, for my third objective, I wanted to construct a hypothesis testing for the data by using Inference on a Population Proportion method. I want to test the hypothesis that the suicide rate for veterans are more than double the rate of civilians suicide rate. The suicide rate for veterans was calculated to be *31.1 per 100,000 people* while civilians suicide rate was calculated to be *13.8 per 100,000 people*. Thus we can assume that

$$p = \frac{13.8}{100,000} * 2 = 2.76 * 10^{-4} \quad Z_{0.05} = 1.65$$

- a) Parameter of Interest: The parameter of interest is the suicide rate proportion p .
- b) Null hypothesis: $H_0: p = 2.76 * 10^{-4}$
- c) Alternative hypothesis: $H_1: p > 2.76 * 10^{-4}$

d) The test statistic is:
$$Z_0 = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

where $\hat{P} = 2.76 * 10^{-4}$ $n = 100,000$ $p_0 = \frac{\text{veteran+civilian suicide rate}}{100,000} = 4.49 * 10^{-4}$

- e) Rejection region: Reject H_0 if $Z_0 > Z_\alpha$ or if P – value $< \alpha$
- f) Computations: the test statistic is:

$$Z_0 = \frac{4.49 * 10^{-4} - 2.76 * 10^{-4}}{\sqrt{\frac{2.76 * 10^{-4} (1 - 2.76 * 10^{-4})}{100,000}}} = 3.29$$

$P\text{value} = 1 - \phi(Z_0) = 1 - \phi(3.29) = 1 - 1 = 0$

- g) Conclusions: Since p-value is less than ($\alpha = 0.05$), we reject the null hypothesis. Thus, the suicide rate proportion p is bigger than $2.76 * 10^{-4}$. This further confirms that veterans suicide rate is more than double the rate of civilians suicide rate

In conclusion, all of the statistical testing seemed to indicate the claim that veterans suicide rate is more than double the rate of civilians suicide rate, is indeed true. However, I was still curious on the reason behind the massive difference between veterans and civilians suicide rate. Upon further research, I learned that although the data is accurate, the suicide rate itself was not portrayed in the right context. For example, the average age of veteran suicides within the data set was nearly 60 years old, not representative of Iraq and Afghanistan veterans generation (Bare, 2015). Furthermore, the rate of veterans who did commit suicide, was slightly higher among veterans who never deployed to Afghanistan or Iraq, suggesting that the causes extend beyond the trauma of war (Zarembo, 2015). Thus, my initial reasoning behind this rate disparity is not necessarily true. Hence in conclusion, it seemed that the veterans suicide data is more complicated than it appears. Although I do not want to downplay the accuracy nor the seriousness of the issue, there are definitely flaws in this data. Perhaps, we can divide this current suicide data rate into different age groups and then compare the two rates. I think that would provide a better image of comparison between veterans and civilians suicide rate.

References

Bare, S. (2015, June 2). The Truth About 22 Veteran Suicides A Day. Retrieved November 12, 2015,

from <http://taskandpurpose.com/truth-22-veteran-suicides-day/>

Hargarten, J., Burnson, F., Campo, B., & Cook, C. (2013, August 24). Veteran Suicides Twice as High

as Civilian Rates. Retrieved October 22, 2015, from

<http://backhome.news21.com/article/suicide/>

Zarembo, A. (2015, January 14). Detailed study confirms high suicide rate among recent veterans.

Retrieved November 12, 2015, from [http://www.latimes.com/nation/la-na-veteran-suicide-](http://www.latimes.com/nation/la-na-veteran-suicide-20150115-story.html)

[20150115-story.html](http://www.latimes.com/nation/la-na-veteran-suicide-20150115-story.html)

state	vet suicides	all suicides	vet suicides %	vet rate	civ rate	Civ Suicides
Alabama	1,045	4,343	24.06%	37.1	15.1	3,298
Alaska	206	1,033	19.94%	41.2	28.3	827
Arizona	1,617	7,156	22.60%	42.6	19.2	5,539
Arkansas	634	2,941	21.56%	36.4	17.4	2,307
California	4,814	25,576	18.82%	33.9	11.8	20,762
Colorado	1,260	5,829	21.62%	44.3	20.0	4,569
Connecticut	318	2,206	14.42%	19.0	11.0	1,888
Delaware	119	707	16.83%	22.0	14.3	588
Georgia	1,319	7,225	18.26%	26.7	13.4	5,906
Hawaii	160	1,082	14.79%	20.2	15.2	922
Idaho	445	1,809	24.60%	49.5	20.0	1,364
Illinois	1,407	7,940	17.72%	25.5	10.6	6,533
Indiana	984	5,721	17.20%	29.2	15.8	4,737
Iowa	495	2,472	20.02%	29.3	13.8	1,977
Kansas	588	2,643	22.25%	37.2	15.9	2,055
Kentucky	505	4,313	11.71%	22.4	18.8	3,808
Louisiana	693	3,662	18.92%	31.2	14.2	2,969
Maine	296	1,286	23.02%	31.0	15.7	990
Maryland	636	3,549	17.92%	20.2	10.9	2,913
Massachusetts	361	3,680	9.81%	12.3	10.3	3,319
Michigan	1,605	8,264	19.42%	31.5	13.9	6,659
Minnesota	760	4,084	18.61%	27.7	13.3	3,324
Mississippi	526	2,644	19.89%	35.8	15.5	2,118
Missouri	1,260	5,702	22.10%	35.4	16.0	4,442
Montana	387	1,464	26.43%	55.9	23.9	1,077
Nevada	812	3,571	22.74%	50.6	23.3	2,759
New Hampshire	250	1,215	20.58%	29.5	15.4	965
New Jersey	597	4,276	13.96%	17.6	8.6	3,679
New Mexico	582	2,726	21.35%	47.4	23.5	2,144
New York	1,503	9,841	15.27%	21.2	8.6	8,338
North Carolina	1,570	7,873	19.94%	30.5	14.7	6,303
North Dakota	146	663	22.02%	37.8	16.7	517
Ohio	1,853	9,262	20.01%	28.6	13.6	7,409
Oklahoma	873	3,987	21.90%	38.5	18.6	3,114
Oregon	1,040	4,317	24.09%	43.5	18.3	3,277
Pennsylvania	2,164	10,587	20.44%	30.2	13.9	8,423
Rhode Island	128	741	17.27%	23.3	11.8	613
South Carolina	907	4,030	22.51%	32.7	15.0	3,123
South Dakota	175	864	20.25%	35.5	18.7	689
Tennessee	1,232	6,378	19.32%	35.2	17.5	5,146
Texas	3,405	18,339	18.57%	30.2	13.4	14,934
Utah	460	2,868	16.04%	43.8	20.3	2,408
Vermont	100	659	15.17%	26.9	18.2	559
Virginia	1,348	6,477	20.81%	25.9	14.4	5,129
Washington	1,570	6,218	25.25%	36.9	15.3	4,648
West Virginia	462	1,894	24.39%	39.6	16.2	1,432
Wisconsin	1,003	5,093	19.69%	33.0	15.2	4,090
Wyoming	173	805	21.49%	47.2	25.6	632
Total	44,793	230,015	19.47%	31.1	13.8	185,222