Analyzing Fatal Encounters With Police in 2015 Using Nonparametric Methods

Evan Wagner

12/14/2021

Abstract: We are interested in analyzing data on fatal encounters with police in America in 2015 but cannot assume that our observed values follow a normal distribution. We apply a variety of distribution-free statistical tests to make conclusions about 2015 police killings, comparing each to its normality-assuming counterpart.

***

Contents

1 Introduction

In recent years, a spate of high-profile violent encounters with American police officers has focused public attention on such incidents. Media outlets have responded by compiling detailed data on fatal incidents across the U.S. and making it available for statistical analysis. For one, *FiveThirtyEight* compiled data on all such incidents from the year 2015, pulling details of each incident from data gathered by *The Guardian* and adding location-related data from the American Community Survey (ACS) conducted in 2015 by the U.S. Census.[1] We will attempt to derive statistically significant conclusions from this dataset.

The nature of these data prevents us from applying many techniques typically used for statistical analysis. In short, we cannot assume that the populations underlying the parameters we wish to measure are normally distributed.

Luckily, several savvy statisticians have developed statistical techniques whose conclusions are valid without assuming its parameters follow a distribution, normal or otherwise. We will apply their distribution-free methods to our data on police killings to try and answer several questions of interest.

1.1 Data Preparation

In order to apply our techniques, changes were made to the raw data. Rows which contained at least one missing value of a quantitative variable were removed, as were entries labeled *"Unknown"* for some categorical variable; this reduced the number of observations from 467 to 438. Additionally, three new binary variables were created by forming two groups from the levels of categorical variables. The full list of variables is displayed in Table 1.

Variable	**Type**	**Description**	**Original Source**
age	quantitative (discrete)	Age of deceased in years	*The Guardian*
pop	quantitative	Population of tract on which the killing occurred	U.S. Census
share_white	quantitative	Share of tract population that is non-Hispanic white	U.S. Census
share_black	quantitative	Share of tract population that is black (alone, not in combination)	U.S. Census
share_hispanic	quantitative	Share of tract population that is Hispanic/Latino (any race)	U.S. Census
p_income	quantitative	Tract-level median personal income	U.S. Census
h_income	quantitative	Tract-level median household income	U.S. Census
county_income	quantitative	County-level median household income	U.S. Census
comp_income	quantitative	Ratio of *h_income* to *county_income*	Calculated by *FiveThirtyEight*
pov	quantitative	Tract-level poverty rate (official)	U.S. Census
urate	quantitative	Tract-level unemployment rate	Calculated by *FiveThirtyEight*
college	quantitative	Share of tract population $\geq$ 25 years old with BA or higher	Calculated by *FiveThirtyEight*
gender	categorical	Gender of deceased	*The Guardian*
raceethnicity	categorical	Race/ethnicity of deceased	*The Guardian*
cause	categorical	Cause of death	*The Guardian*
armed	categorical	How/whether deceased was armed	*The Guardian*
county_bucket	categorical	Tract-level household income, quintile within county	Calculated by *FiveThirtyEight*
nat_bucket	categorical	Tract-level household income, quintile nationally	Calculated by *FiveThirtyEight*
white	binary	Coded 1 when *raceethnicity* = "$White$", 0 otherwise	Computed by the author
male	binary	Coded 1 when *gender* = "$Male$", 0 when *gender* = "$Female$"	Computed by the author
armedyes	binary	Coded 0 when *armed* = "No", 1 otherwise	Computed by the author

Table 1: Variables in the dataset

2 Two-Sample Tests

If we split a quantitative variable into two groups based on a binary variable, we can test whether the two samples have equal medians and/or variance. Box plots were generated for each of the possible variable choices to identify potential differences. The author selected three pairings of variables whose boxplots seemed to show significantly different medians, shown in Figure 1 along with histograms to give the reader a sense of the distribution's shape. We see that each distribution is somewhat right-skewed, and the white age distribution seems skinnier than the non-white distribution in the first histogram.

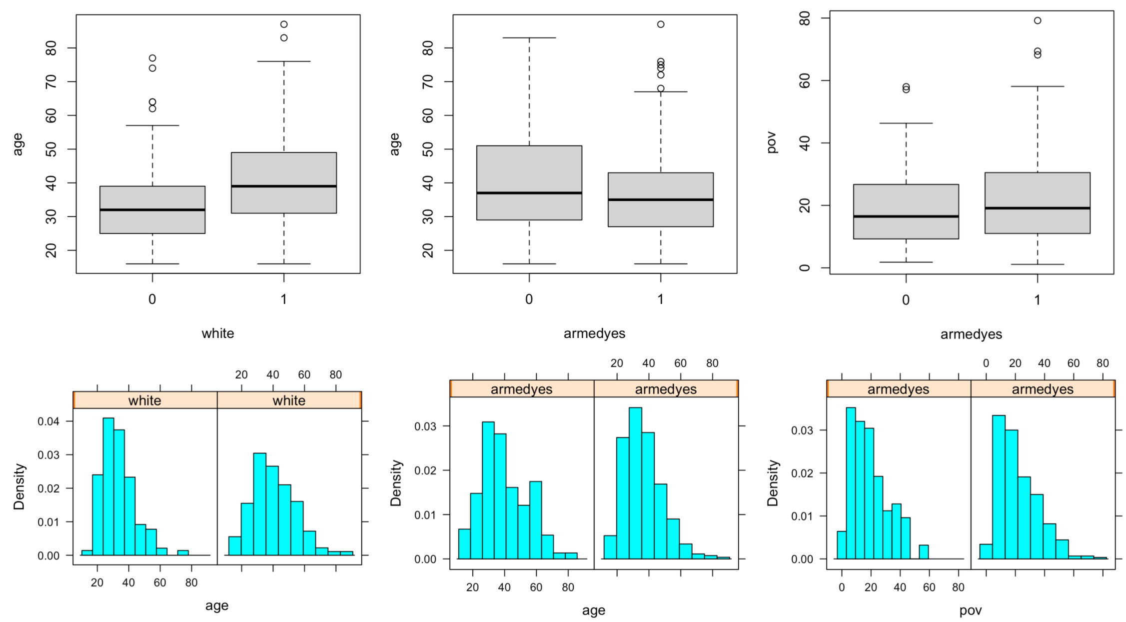

Figure 1: Box plots and histograms of three quantitative-binary variable pairs

2.1 Age vs. Whiteness: Parametric Analysis

Our data includes 229 white victims and 209 nonwhite victims. In order to find a significant difference between the mean ages of these two populations, our first instinct might be to conduct a t-test. Using the base R function `t.test()`, we obtain a p-value of $1.038 * 10^{-9}$ and a 95 percent confidence interval of $(4.973, 9.544)$. If this test were valid, we would say with 95 percent confidence that white victims are between 4.973 and 9.544 years older than nonwhite victims on average.

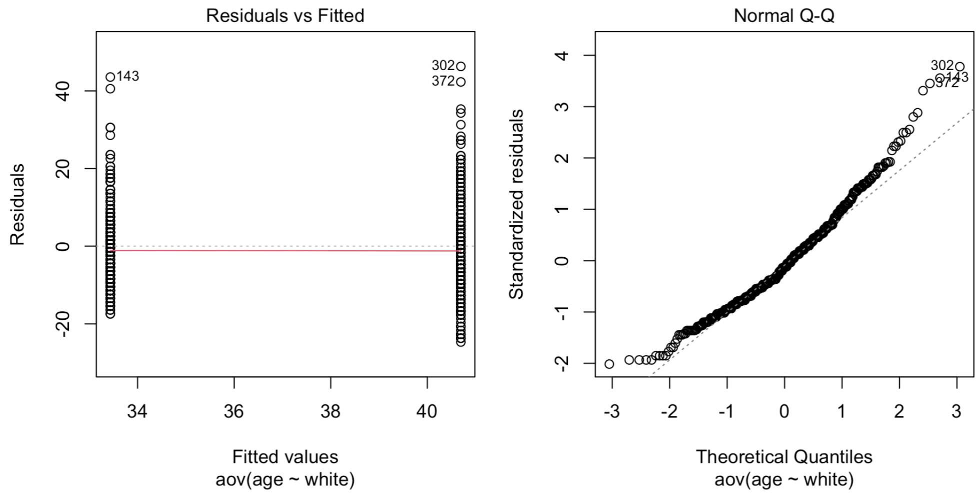

Figure 2: Diagnostic plots for the two-sample t-test of *age* vs. *white*

Unfortunately, a quick look at some diagnostic plots shows us that the conditions for inference based on the t-test are not met. Figure 2 shows a plot of the residuals vs. fitted values from our model,[1] as well as a normal probability plot of the residuals. The residuals for the nonwhite population seem to be less widely distributed than for the white population, calling the equal variance assumption into question. We can further test for unequal variance with Levene's test. Using the R function `levene.test()` from the package `lawstat`, we obtain a p-value of .000413 and conclude that the variances of the two populations are indeed unequal.

Furthermore, we have evidence that our populations are not distributed normally. The normal probability plot shows residuals that depart strongly from the expected normal curve at both tails. We can also conduct the Anderson-Darling test for normality. Using the R function `ad.test()` from the package `nortest`, we obtain a p-value of $1.013 * 10^{-7}$ and conclude that our residuals do not follow the normal distribution. Clearly, any inference we make from the standard t-test on the two populations would be invalid. Our familiar distribution-based methods have failed us, so we must turn to distribution-free methods.

[1]To create the plots in Figures 2 and **??**, the author used one-way ANOVA tests, which in the two-sample case are equivalent to t-tests.

2.2 Age vs. Whiteness: Nonparametric Analysis

We will first determine whether the distributions underlying our two populations are dispersed equally. Setting $\gamma^2 = \frac{var(age_{white})}{var(age_{nonwhite})}$, we seek to determine whether the ratio of the two population variances γ^2 is significantly far from 1 in either direction.

Our result from Levene's test suggests that the variances are unequal. However, we see from Figure 2 that the data contains several outliers, to which Levene's test is very sensitive. In comparison, the Miller jackknife procedure is much more resilient to outliers because its test statistic is adjusted based on the contribution of each observation to the overall mean. We can make a more robust conclusion about the variances of these two populations by applying the jackknife procedure.

Miller jackknife test

$H_0 : \gamma^2 = 1$
$H_3 : \gamma^2 \neq 1$ Reject H_0 if $|Q| \geq z_{\alpha/2}$

Using the custom R function `jackKnife()`, we find that $Q = 2.752$ and obtain a p-value of .00296. Thus we reject H_0 and conclude that the two distributions have different dispersions. Note that the Miller p-value is considerably larger than the Levene p-value, emphasizing that the jackknife procedure is less certain about variance because it adjusts for outliers.

We also obtain an estimate for the ratio between the variances of the populations, $\bar{\gamma}^2 = 1.581$, and the associated 95 percent confidence interval $(1.141, 2.191)$. In other words, we are 95 percent confident that the variance in age for white victims is between 1.141 and 2.191 times larger than that of nonwhite victims.

Having concluded unequal variance for the two distributions of *age*, conditions are not met to apply the Wilcoxon rank sum test. Instead, we will use the Fligner-Policello test for unequal medians, which does not assume equal variance. Since our sample has many observations and because it reduces computing time greatly, we will use the large sample approximation of this test.

Fligner-Policello test

$H_0 : \theta_{white} = \theta_{nonwhite}$
$H_3 : \theta_{white} \neq \theta_{nonwhite}$ Reject H_0 if $|\hat{U}| \geq z_{\alpha/2}$

Using the R function `pFligPoli()` from the package `NSM3`, we find that $\hat{U} = 6.353$ and obtain a p-value of $2.112 * 10^{-10}$. Thus we reject H_0 and conclude that the median age of white victims is different than that of non-white victims.

For an estimate of the true difference in location $\theta_{white} - \theta_{nonwhite}$, we can use the Hodges-Lehmann $\hat{\Delta}$, which is the median of the differences between each measurement of $age_{nonwhite}$ and age_{white}. Using the base R function `wilcox.test()`, we find that $\hat{\Delta} = 7.000$. In other words, we estimate that white victims are 7 years older than nonwhite victims on average. Because our two variances are unequal, we are unable to obtain a meaningful confidence interval for the estimated age difference between the two populations. Table 2 summarizes the results obtained in this subsection.

parameter	**parametric p-value**	**nonparametric p-value**	**estimate**	**95 percent conf. interval**
Location	$1.038 * 10^{-9}$ (t-test)	$2.112 * 10^{-10}$ (Fligner-Policello)	$\hat{\Delta} = 7.000$	unobtainable
Dispersion	.000413 (Levene's test)	0.00296 (Miller jackknife)	$\bar{\gamma}^2 = 1.581$	(1.141, 2.191)

Table 2: Results for *age* vs. *white*

2.3 More Two-Sample Nonparametric Analysis

Our second analysis, of age broken down by whether or not the victim was armed, goes quite similarly to our first. There are 338 armed and 100 unarmed victims in the dataset. Plots and tests indicate that the values of *age* vary neither equally nor normally around their means, rendering distribution-based methods inappropriate. The Miller jackknife returns a p-value of 0.0197, which is enough to reject the null hypothesis of equal population variances at alpha level .05 but not .01. Therefore, in addition to the Fligner-Policello test, we proceed with the Wilcoxon rank sum test using the base R function `wilcox.test()` and obtain a 95 percent confidence interval for $\theta_{armed} - \theta_{unarmed}$. Though both nonparametric tests for unequal means are less certain than the t-test, they still reject the null at alpha levels .05 and higher. Full results are shown in Table 3.

As for *pov* vs. *armedyes*, normal residuals are again soundly rejected, but neither variance test could conclude that the populations have different scales. We proceed only with the Wilcoxon rank sum test, which fails to find a difference in medians. Therefore, we do not have sufficient evidence to conclude that the populations are distributed differently in any way. Full results are shown in Table 4.

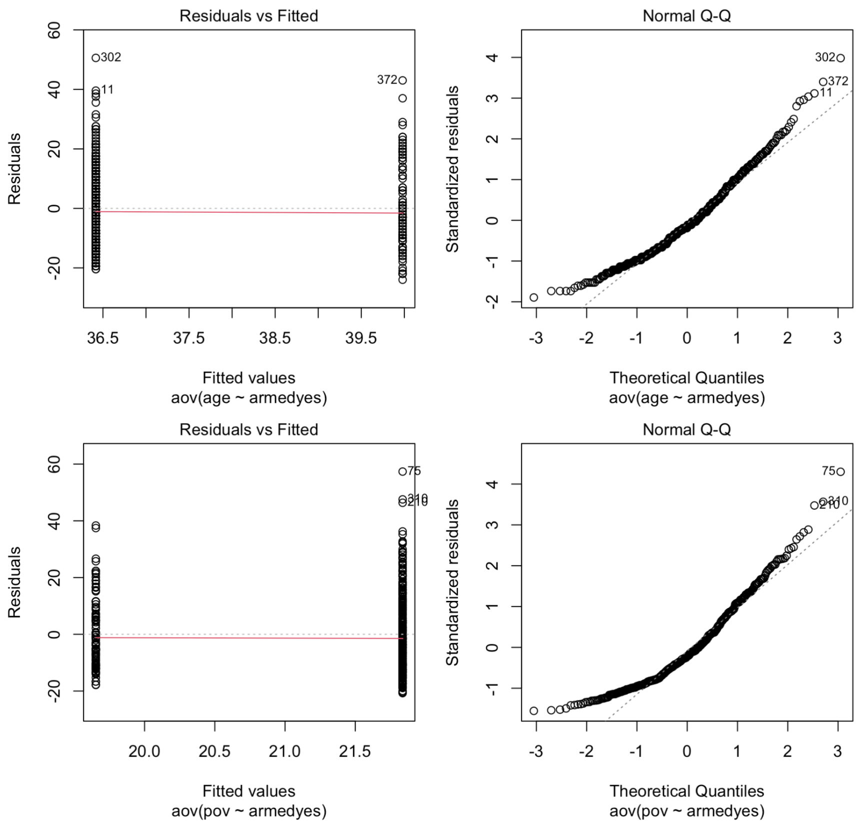

Figure 3: Diagnostic plots for the two-sample t-tests of *age* and *pov* vs. *armedyes*

parameter	**parametric p-value**	**nonparametric p-value**	**estimate**	**95 percent conf. interval**
Location	0.0263 (t-test)	0.0400 (rank sum) 0.0454 (Fligner-Policello)	$\hat{\Delta} = -3.000$	(-6.000, -.0000587)
Dispersion	0.0357 (Levene's test)	0.0197 (Miller jackknife)	$\bar{\gamma}^2 = 0.709$	(0.511, 0.984)

Table 3: Results for *age* vs. *armedyes*

parameter	**parametric p-value**	**nonparametric p-value**	**estimate**	**95 percent conf. interval**
Location	0.1331 (t-test)	0.1657 (rank sum)	$\hat{\Delta} = 1.800$	$(-0.800, 4.400)$
Dispersion	0.3443 (Levene's test)	0.1733 (Miller jackknife)	$\bar{\gamma}^2 = 1.190$	(0.828, 1.711)

Table 4: Results for *age* vs. *white*

3 One-Way Analysis of Variance Methods

In the previous section, we found a significant race-based difference in the ages of deadly force victims between two racial categories, white and nonwhite. In the parametric setting, we could extend our analysis to more than two categories using a one-way analysis of variance (ANOVA) model, then make comparisons between each difference with Tukey's honest significant differences (HSD). This section explores the analogous distribution-free tests: the Kruscal-Wallis test for unequal treatment effects, and the Steel, Dwass, Critchlow-Fligner multiple comparisons procedure.

3.1 Age vs. Race

"White"	"Black"	"Hispanic/Latino"	"Asian/Pacific Islander"	"Native American"
229	131	64	10	4

Table 5: Observation counts for the five levels of *raceethnicity*

Table 5 shows the count of observations for the five levels of *raceethnicity*. The last two groups have only 10 and 4 observations, respectively, making them too much smaller than the other three for valid comparisons. Thus we drop observations coded "Asian/Pacific Islander" and "Native American" and compare the remaining three.

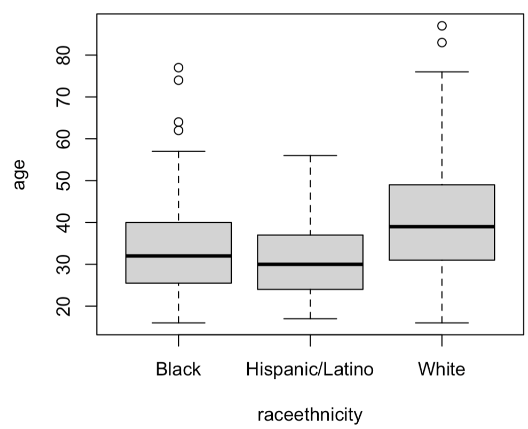

Figure 4: Boxplots of *age* vs. three levels of *raceethnicity*

We see in Figure 4 that the median age of white victims looks significantly higher than for both Black and Latino victims. Using the base R function `aov()`, we fit an ANOVA model to the data and obtain a p-value of $3.715 * 10^{-9}$. If this test were valid, we would conclude quite confidently that at least one of the treatment effects $\tau_{white}, \tau_{black}, \tau_{latino}$ is not equal to the rest.

4 Regression Methods

5 Conclusion

References

[1] B. Casselman, "Where Police Have Killed Americans In 2015." *FiveThirtyEight*, June 3, 2015. Retrieved December 14, 2021 from https://fivethirtyeight.com/features/where-police-have-killed-americans-in-2015/. Dataset: FiveThirtyEight Police Killings, Version 105. Retrieved December 14, 2021 from https://www.kaggle.com/fivethirtyeight/fivethirtyeight-police-killings-dataset/version/105.