

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	2
1.1	Data Preparation	2
2	Two-Sample Tests	4
2.1	Age vs. Whiteness: Parametric Analysis	4
2.2	Age vs. Whiteness: Nonparametric Analysis	7
2.3	More Two-Sample Analysis	8
3	One-Way Analysis of Variance Methods	10
3.1	Age vs. Race: Parametric Analysis	10
3.2	Age vs. Race: Nonparametric Analysis	11
3.3	More One-Way Analysis	13
4	Conclusion	16
4.1	Further Discussion	17

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.

As shall be seen in the following, 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, which is a prerequisite for all of the basic techniques taught in beginning statistics courses. 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.

Before we begin, two very important hesitations must be addressed. The first is that our conclusions only apply to fatal encounters in the year 2015. With only one year in our sample, we cannot assess the influence of time-based effects on our parameters of interest. Indeed, as police brutality began growing more prominent in the public imagination starting in 2014, it is likely that some police departments have responded to scrutiny of their peacekeeping practices by overhauling them.

Secondly, our data is not a typical simple random sample (SRS) because it technically includes our entire population of interest: 2015 victims of deadly force. As such, we have no need to estimate means and deviations because we can calculate them directly. That said, our conclusions are still meaningful for comprehending and comparing these data. If a higher power wound back the clock to 2015 and set the conditions back to how they were, we would expect slightly different outcomes due to the contribution of randomness to our parameters.

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
<i>age</i>	quantitative	Age of deceased	<i>The Guardian</i>
<i>pop</i>	quantitative	Population of tract on which the killing occurred. Tracts are subdivisions of counties containing a few thousand residents each.	U.S. Census
<i>share_white</i>	quantitative	Share of tract population that is non-Hispanic white	U.S. Census
<i>share_black</i>	quantitative	Share of tract population that is black (alone, not in combination)	U.S. Census
<i>share_hispanic</i>	quantitative	Share of tract population that is Hispanic/Latino (any race)	U.S. Census
<i>p_income</i>	quantitative	Tract-level median personal income (USD)	U.S. Census
<i>h_income</i>	quantitative	Tract-level median household income (USD)	U.S. Census
<i>county_income</i>	quantitative	County-level median household income (USD)	U.S. Census
<i>comp_income</i>	quantitative	Ratio of <i>h_income</i> to <i>county_income</i>	Calculated by <i>FiveThirtyEight</i>
<i>pov</i>	quantitative	Tract-level poverty rate (official)	U.S. Census
<i>urate</i>	quantitative	Tract-level unemployment rate	Calculated by <i>FiveThirtyEight</i>
<i>college</i>	quantitative	Share of tract population ≥ 25 years old with BA or higher	Calculated by <i>FiveThirtyEight</i>
<i>gender</i>	categorical	Gender of deceased	<i>The Guardian</i>
<i>raceethnicity</i>	categorical	Race/ethnicity of deceased	<i>The Guardian</i>
<i>cause</i>	categorical	Cause of death	<i>The Guardian</i>
<i>armed</i>	categorical	How/whether deceased was armed	<i>The Guardian</i>
<i>county_bucket</i>	categorical	Tract-level household income, quintile within county	Calculated by <i>FiveThirtyEight</i>
<i>nat_bucket</i>	categorical	Tract-level household income, quintile nationally	Calculated by <i>FiveThirtyEight</i>
<i>white</i>	binary	Coded 1 when <i>raceethnicity</i> = “White”, 0 otherwise	Computed by the author
<i>male</i>	binary	Coded 1 when <i>gender</i> = “Male”, 0 when <i>gender</i> = “Female”	Computed by the author
<i>armedyes</i>	binary	Coded 0 when <i>armed</i> = “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.

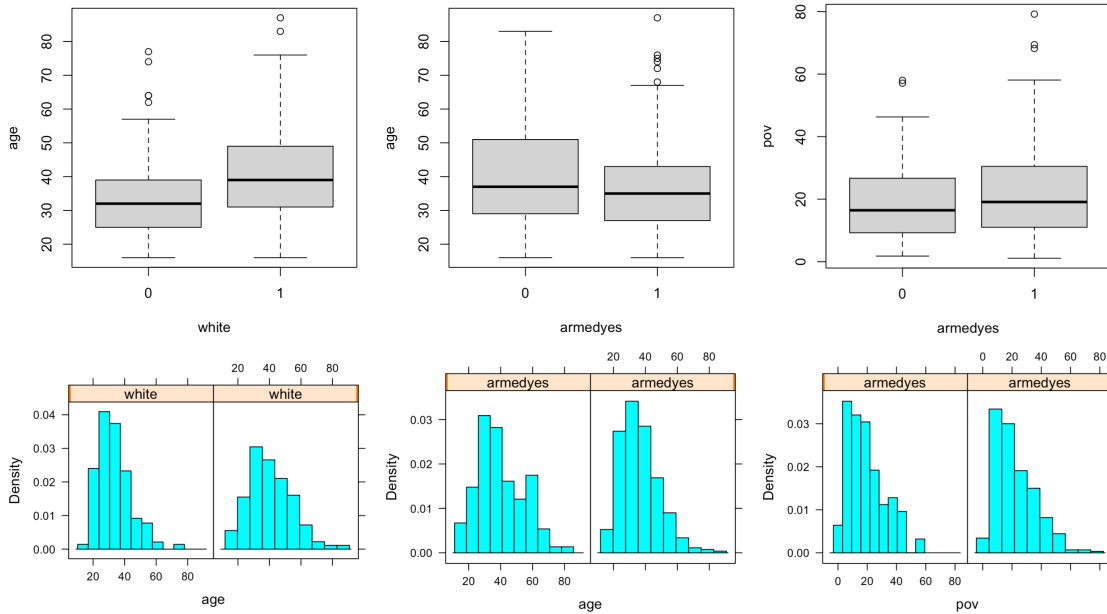


Figure 1: Box plots and histograms of three quantitative-binary variable pairs

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 average age of these two populations, our first instinct might be to conduct a t -test. Let μ_W be the mean age of white victims and μ_N be the mean of nonwhite victims. Our null hypothesis is that these two means are equal, or equivalently, their difference $\mu_W - \mu_N = 0$. Our estimate of the difference is simply the difference in the mean ages of our two samples, which turns out to be 7.259 years.

But is this difference significant? 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 reject the null hypothesis and declare with 95 percent confidence

that white victims are between 4.973 and 9.544 years older than nonwhite victims on average.

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.

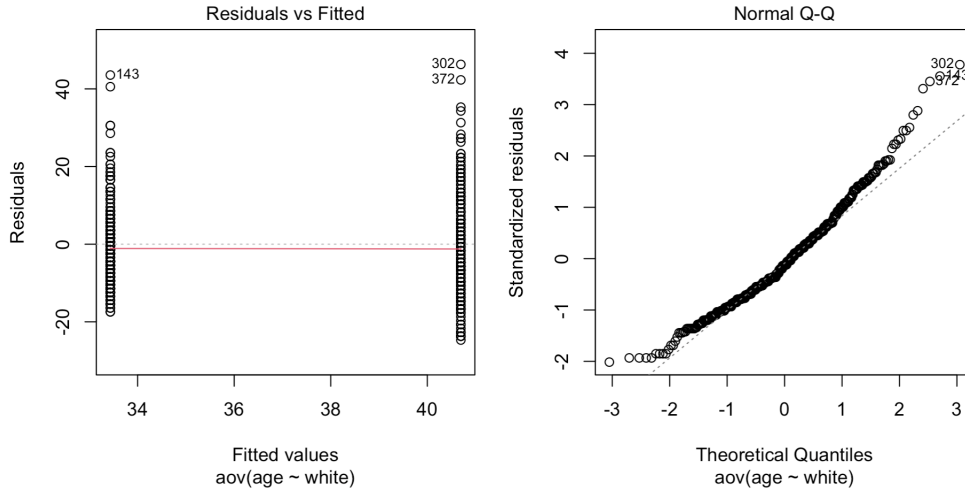


Figure 2: Diagnostic plots for the two-sample t -test of *age* vs. *white*

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 these two samples would be invalid.

¹To 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.

Perhaps a transformation on *age* will satisfy the conditions for the *t*-test. In Figure 2 we see upper outliers in the boxplots and a right skew in the histograms. The square root of age might have a more symmetric distribution. Sure enough, our plots of \sqrt{age} vs. *white* and the new *t*-test residuals (shown in Figure 3) depict a distribution that looks much closer to the normal curve. However, Levene's test of the transformed variance returns a *p*-value of 0.00844, and the Anderson-Darling test returns a *p*-value of 0.0396. We still reject their null hypotheses of equally-varying and normal residuals.

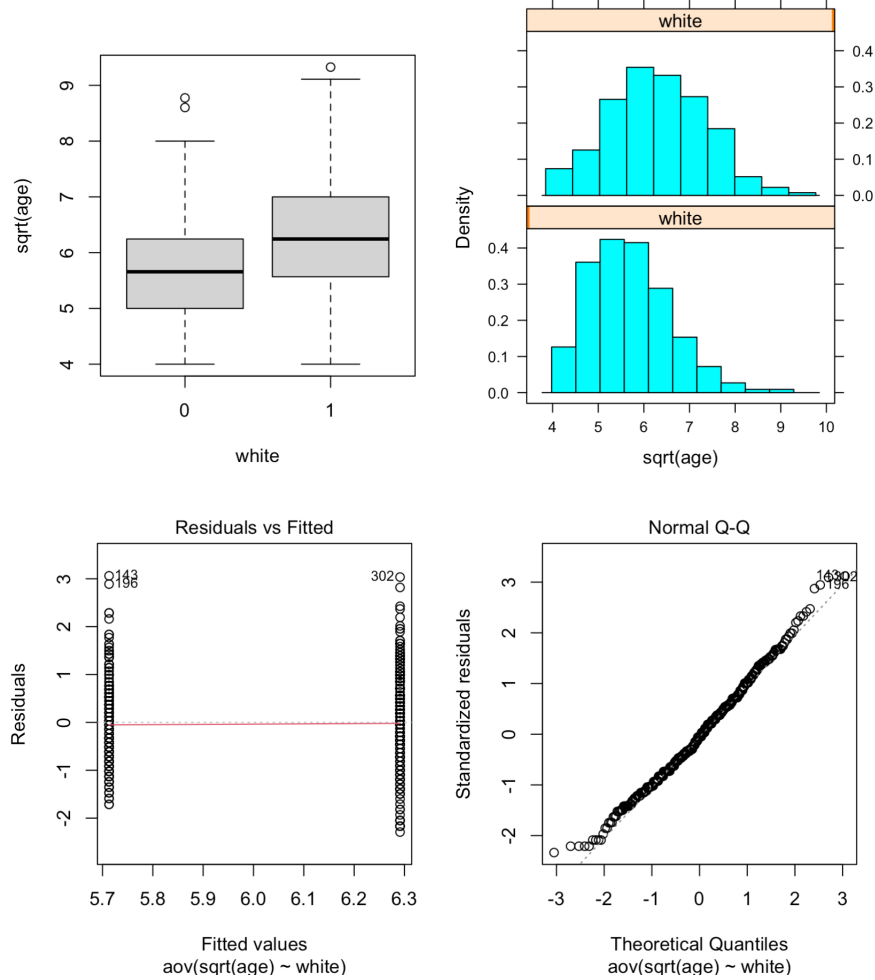


Figure 3: Plots of \sqrt{age} vs. *white*

For the sake of comparison with our nonparametric procedures, we convert the estimates from the *t*-test on \sqrt{age} against *white* back into regular years and report them. Though their distribution is still invalid for the *t*-test, it is an improvement on the untransformed model and its conclusions are slightly more representative of the truth. Our new *p*-value is 1.381×10^{-9} . Our new estimate is the difference between the squared estimates of mean \sqrt{age} , which comes out to 6.941. We form the new

confidence interval by taking the squared mean \sqrt{age} of white victims and subtracting the squared difference between the mean \sqrt{age} of white victims and the lower bound of our confidence interval, then doing the same with the upper bound. This new interval is (4.810, 9.004).

In light of the problematic distribution of age against whiteness, we turn to non-parametric methods for a more robust analysis of the two variables.

2.2 Age vs. Whiteness: Nonparametric Analysis

We will first determine whether the distributions underlying our two populations are dispersed equally. Setting $\gamma^2 = \frac{\text{var}(\sqrt{age}_{white})}{\text{var}(\sqrt{age}_{nonwhite})}$, 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 these 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

$$\begin{aligned} H_0 : \gamma^2 &= 1 \\ H_3 : \gamma^2 &\neq 1 \end{aligned} \quad \text{Reject } H_0 \text{ if } |Q| \geq z_{\alpha/2}.$$

Using the custom R function `jackknife()`, we find that $Q = 2.420$ and obtain a p -value of 0.00777, which is quite similar to that of Levene's test. 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.

Now we move on to determining whether the median ages are different for our two populations. In light of our rejection of the null hypothesis with the Miller jackknife, we should conduct the Fligner-Policello test, which does not assume equality of variance. We need not transform and untransform *age* for this test: it is based on the ranks of observations relative to each other, and since the square root is an increasing function, it does not change the rankings of observations at all. Therefore, the test will not conclude anything differently before and after transformation. Additionally, since our sample has many observations and because it reduces computing time greatly, we will use the large-sample approximation (LSA) of the Fligner-Policello test.

Fligner-Policello test

$$H_0 : \theta_{white} = \theta_{nonwhite}$$

$$H_3 : \theta_{white} \neq \theta_{nonwhite}$$

$$\text{Reject } H_0 \text{ if } |\hat{U}| \geq z_{\alpha/2}.$$

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 the null hypothesis 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 median $\theta_{white} - \theta_{nonwhite}$, 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 see that $\hat{\Delta} = 7.000$. In other words, we estimate that white victims are roughly 7 years younger than nonwhite victims on average. The output also finds a confidence interval of (5.000, 9.000), but that is meaningless if the variances of our populations are unequal.

Table 2 summarizes the results obtained from our tests of age broken down by whiteness. It appears that the true difference in means (and our sample size) is large enough to make the conclusions of these tests very similar.

parameter	procedure	p-value	estimate	95 pct. CI
Location	t -test	$1.381 * 10^{-9}$	$\hat{\mu}_W - \hat{\mu}_N = 6.941$	(4.810, 9.004)
	Fligner-Policello	$2.111 * 10^{-10}$	$\hat{\Delta} = 7.000$	not meaningful
Dispersion	Levene's test	0.00844	—	—
	Miller jackknife	0.00777	$\bar{\gamma}^2 = 1.396$	(1.065, 1.829)

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

2.3 More Two-Sample Analysis

For the remainder of the paper, we will continue applying a square root to *age* and all other quantitative variables, as it improves conditions in every case.

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. This time, the Anderson-Darling test still rejects the normality condition, but Levene's test returns a p -value of 0.0850, hovering right on the brink of the common significance level of $\alpha = 0.05$. The Miller jackknife returns a similar p -value of 0.0584. Therefore, in addition to the Fligner-Policello test, we proceed with the Wilcoxon rank sum test using `wilcox.test()` and obtain a meaningful 95 percent confidence interval for the difference in population medians $\theta_A - \theta_U$. 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.

parameter	procedure	p-value	estimate	95 pct. CI
Location	t -test	0.0328	$\hat{\mu}_U - \hat{\mu}_A = -3.274$	$(-6.403, -0.266)$
	rank sum	0.0400	$\hat{\Delta} = -3.000$	$(-6.000, -0.0000587)$
	Fligner-Policello	0.0454	—	—
Dispersion	Levene's test	0.0850	—	—
	Miller jackknife	0.0584	$\bar{\gamma}^2 = 0.709$	$(0.574, 1.010)$

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

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. In this case, we see that the rank sum test is slightly less confident than the t -test, while the Miller jackknife is more confident than Levene's test. Full results are shown in Table 4.

parameter	procedure	p-value	estimate	95 pct. CI
Location	t -test	0.1483	$\hat{\mu}_A - \hat{\mu}_U = 2.009$	$(-0.749, 4.568)$
	rank sum	0.1657	$\hat{\Delta} = 1.800$	$(-0.800, 4.400)$
Dispersion	Levene's test	0.6036	—	—
	Miller jackknife	0.3010	$\bar{\gamma}^2 = 1.078$	$(0.814, 1.427)$

Table 4: Results for *pov* vs. *armedyes*

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 Kruskal-Wallis test for unequal treatment effects, and the Steel, Dwass, Critchlow-Fligner multiple comparisons procedure.

3.1 Age vs. Race: Parametric Analysis

“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.

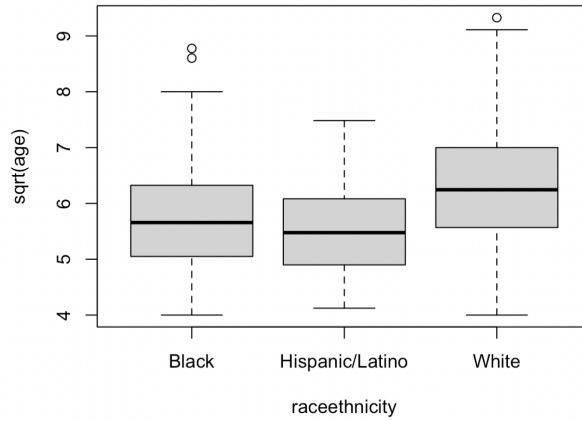


Figure 4: Box plots of $\sqrt{\text{age}}$ vs. three levels of *raceethnicity*

We see in Figure 4 that the median $\sqrt{\text{age}}$ 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 4.273×10^{-9} . If this test were valid, we would conclude quite confidently that at least one of the treatment effects τ_W, τ_B, τ_L is not equal to the rest.

We would then proceed with a multiple comparisons procedure, such as Tukey’s HSD. Using the base R function `TukeyHSD()`, we obtain intervals for the three configurations of differences in means, adjusted for differences in variance, with 95 percent overall confidence. As shown in Table 6, Tukey’s HSD claims that, on average, white victims are older than both Black and Latino victims.

These distribution-reliant claims ring hollow when we check their conditions for inference. Shown in Figure 5, the residuals appear to have different-sized variances for the three groups and lift up considerably from the tails of the expected normal distribution. The picture gets more complicated when we conduct statistical tests: though Levene’s test returns a p -value of .0243, the Anderson-Darling test yields a p -value of 0.0564, which rejects the null hypothesis of normality only at a weak significance level of 0.1. We may consider the ANOVA model somewhat more meaningful in this case, but our distribution-free methods will give us much more robust conclusions.

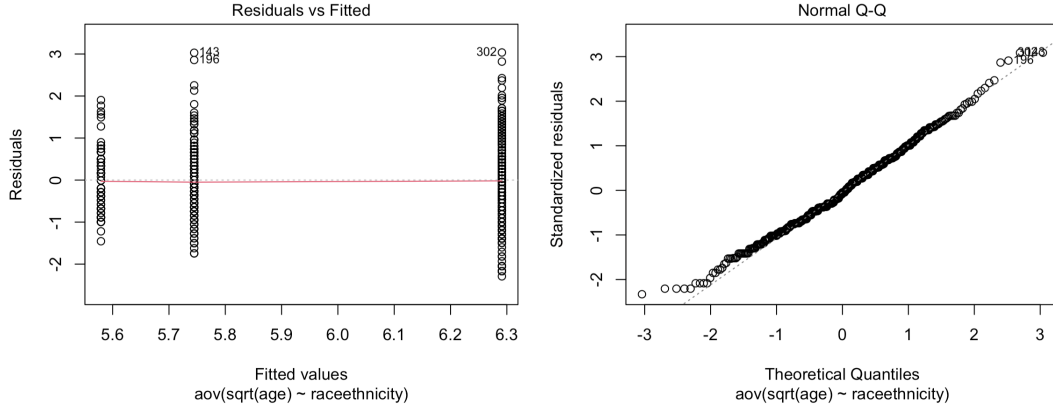


Figure 5: Diagnostic plots for the ANOVA model of *age* vs. *raceethnicity*

3.2 Age vs. Race: Nonparametric Analysis

Turning to our nonparametric toolbox, we apply the Kruskal-Wallis test, which ranks the combined samples to measure whether the observations in at least one of them are generally ranked higher or lower than those of the other two samples. This test assumes the populations of interest have equal variance, so its results are questionable considering the rejection of the null hypothesis of Levene’s test. That said, it is much more resilient to unequal variance than the ANOVA model, so its conclusions are more robust. We will once again perform this rank-based test on the untransformed *age* variable and use the LSA to reduce computing time.

Kruskal-Wallis test

$$H_0 : \tau_W = \tau_B = \tau_L$$

$$H_1 : \text{Not all } \tau_k \text{ equal, } k \in \{W, B, L\} \quad \text{Reject } H_0 \text{ if } H' \geq \chi_{2,\alpha}^2.$$

Using the R function `pKW()` from the package `NSM3`, we find that $H' = 37.515$ and obtain a p -value of $7.140 * 10^{-9}$. We reject the null hypothesis and conclude that at least one of the three groups has a different median age than at least one of the others. Note that the Kruskal-Wallis p -value is only marginally higher than that of the ANOVA model, and both firmly reject the null hypothesis.

Our conclusion from the Kruskal-Wallis test motivates us to identify significant pairwise differences with the Steel-Dwass-Critchlow-Fligner procedure, which applies the rank sum test to the three configurations of combined pairs of populations while maintaining overall confidence of $1 - \alpha$.

Steel, Dwass, Critchlow-Fligner procedure

For each pair (τ_u, τ_v) , decide $\tau_u \neq \tau_v$ if $|W_{uv}^*| \geq w_\alpha^*$.

Using the R function `pSDCFlig()` from the package `NSM3`, we obtain the three statistics W_{uv}^* and use the LSA. Based on its results (shown in Table 6), we decide with 95 percent overall confidence that white victims have a different median age than Black and Latino victims. We can now use a procedure devised by Spjøtvoll to find adjusted Hodges-Lehmann estimators for each difference in group medians. Using custom R code, we obtain the estimates $\hat{\theta}_{WB} = 6.849$ and $\hat{\theta}_{WL} = 8.309$; the estimate $\hat{\theta}_{BL} = 1.460$ is not significant. Finally, from the ordered differences of each population, we find three intervals for the true median with 95 percent overall confidence.

The results of this subsection are summarized in Table 6. We see that our nonparametric methods yield p -values that are about twice the size of the parametric p -values, except for the test of θ_{BL} , which is slightly smaller than its distribution-based counterpart. That said, every conclusion is quite strong in this case, and the outcomes of the two types of tests are no different.

parameter	procedure	p-value	estimate	95 pct. CI
General differences	one-way ANOVA	$4.273 * 10^{-9}$	—	—
	Kruskal-Wallis	$7.140 * 10^{-9}$	—	—
θ_{WB}	Tukey's HSD	0.0000019	$\hat{\mu}_{WB} = 6.577$	(3.595, 9.430)
	SDCFlig	0.0000038	$\hat{\theta}_{WB} = 6.849$	(4.000, 10.000)
θ_{WL}	Tukey's HSD	0.0000015	$\hat{\mu}_{WL} = 8.452$	(4.687, 12.001)
	SDCFlig	0.0000022	$\hat{\theta}_{WL} = 8.309$	(5.000, 12.000)
θ_{BL}	Tukey's HSD	0.5137	$\hat{\mu}_{BL} = 1.874$	(-2.132, 5.521)
	SDCFlig	0.4601	$\hat{\theta}_{BL} = 1.460$	(-2.000, 5.000)

Table 6: Results for *age* vs. *raceethnicity*

3.3 More One-Way Analysis

We proceed with three additional analyses of treatment effects with the same methods.

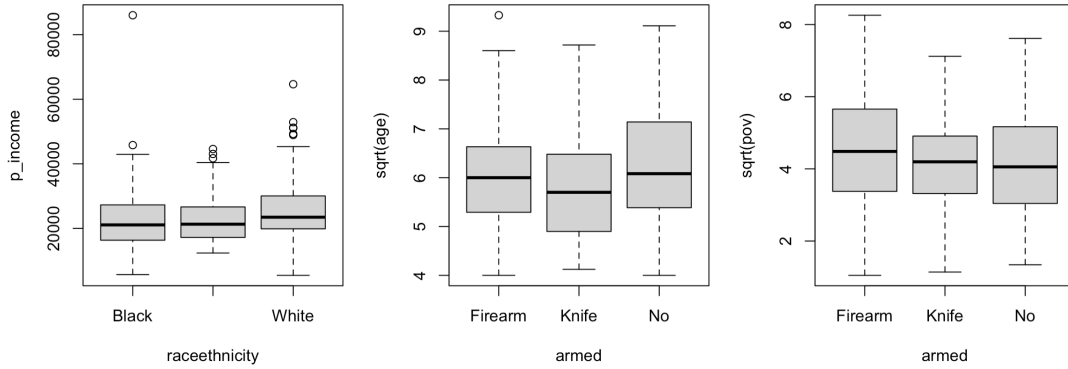


Figure 6: Box plots of $\sqrt{p_income}$ vs. *raceethnicity*, and of \sqrt{age} and \sqrt{pov} vs. *armed*

Results for the analysis of *p_income* against *raceethnicity* are summarized in Table 7. In this case, Levene's test yields a large *p*-value of 0.417, while Anderson-Darling strongly rejects normality of residuals. The square root transformation is found to worsen variance, so the untransformed model is used this time. Interestingly, the conclusions of our nonparametric tests are considerably stronger than those of our (invalid) parametric tests. The SDCFlig procedure even identifies a significant difference between median age for white vs. Latino victims, which Tukey's HSD does not.

parameter	procedure	p-value	estimate	95 pct. CI
General differences	one-way ANOVA	0.02071	–	–
	Kruskal-Wallis	0.0021	–	–
θ_{WB}	Tukey’s HSD	0.0428	$\hat{\mu}_{WB} = 2369.67$	(60.56, 4678.77)
	SDCFlig	0.0065	$\hat{\theta}_{WB} = 2657.40$	(608.00, 4623.00)
θ_{WL}	Tukey’s HSD	0.1046	$\hat{\mu}_{WL} = 2581.70$	(–398.68, 5562.09)
	SDCFlig	0.0301	$\hat{\theta}_{WL} = 2520.99$	(186.00, 4816.00)
θ_{BL}	Tukey’s HSD	0.5137	$\hat{\mu}_{BL} = 212.04$	(–3002.64, 3426.71)
	SDCFlig	0.9852	$\hat{\theta}_{BL} = 136.41$	(–2747.00, 2529.00)

Table 7: Results for *age* vs. *raceethnicity*

Our final two analyses involve the categorical variable *armed*. As before, we trim the data to only include the three levels of *armed* with the most observations. In the resulting dataset, there are 100 unarmed victims, 219 firearm-wielding victims, and 66 knife-wielding victims.

Table 8 displays the results of our analysis of *age* against *armed*. Levene’s test again does not reject its null hypothesis for these groups, while the Anderson-Darling test does. The outcomes of the parametric and nonparametric tests in this case are quite similar. One notable difference is that Tukey’s HSD returns a much higher *p*-value for θ_{FK} than for θ_{UF} , while the Kruskal-Wallis *p*-values for those two differences are virtually the same. Perhaps the upper-tailed outliers of *age* for the firearm- and knife-wielding groups (seen in the second plot in Figure 6) make the parametric methods consider those two distributions closer to each other than to the unarmed group.

parameter	procedure	p-value	estimate	95 pct. CI
General differences	one-way ANOVA	0.0914	–	–
	Kruskal-Wallis	0.0796	–	–
θ_{UF}	Tukey’s HSD	0.1620	$\hat{\mu}_{UF} = 2.852$	(–0.820, 6.524)
	SDCFlig	0.3590	$\hat{\theta}_{UF} = 2.000$	(–2.000, 6.000)
θ_{UK}	Tukey’s HSD	0.0905	$\hat{\mu}_{UK} = 4.313$	(–0.512, 9.139)
	SDCFlig	0.0803	$\hat{\theta}_{UK} = 4.000$	(0.000, 9.000)
θ_{FK}	Tukey’s HSD	0.7004	$\hat{\mu}_{FK} = 1.461$	(–2.811, 5.734)
	SDCFlig	0.3239	$\hat{\theta}_{FK} = 2.000$	(–2.000, 6.000)

Table 8: Results for *age* vs. *armed*

Lastly, Table 9 summarizes results for our analysis of *pov* against *armed*. The *p*-value of Levene’s test is an indecisive 0.05415, but the Anderson-Darling test strongly rejects normality, as usual. Across the board, our distribution-free methods are less

confident in rejection than our distribution-based methods. Both ANOVA and the Kruscal-Wallis test reject the null only at levels .1 and above.

parameter	procedure	p-value	estimate	95 pct. CI
General differences	one-way ANOVA	0.0672	—	—
	Kruscal-Wallis	0.081	—	—
θ_{UF}	Tukey's HSD	0.1874	$\hat{\mu}_{UF} = -2.682$	$(-6.283, 0.919)$
	SDCFlig	0.2043	$\hat{\theta}_{UF} = 2.470$	$(-6.000, 0.900)$
θ_{UK}	Tukey's HSD	0.8364	$\hat{\mu}_{UK} = 1.145$	$(-3.587, 5.877)$
	SDCFlig	0.9754	$\hat{\theta}_{UK} = 0.530$	$(-3.400, 4.600)$
θ_{FK}	Tukey's HSD	0.0815	$\hat{\mu}_{FK} = 3.827$	$(-0.363, 8.012)$
	SDCFlig	0.1456	$\hat{\theta}_{FK} = 3.000$	$(-0.700, 7.200)$

Table 9: Results for *pov* vs. *armed*

4 Conclusion

The nonparametric statistical techniques employed in this paper allow us to make several interesting conclusions about the recipients of deadly force from police in America in 2015. The reader is reminded not to extrapolate the following conclusions beyond the year 2015, and that our data includes the entire population of interest and thus its parameters are technically known. Our estimates of the parameters are not so useful in themselves, but our comparisons of the different populations do reveal whether or not differences between them are significant.

We estimate that white victims are roughly 7 years older than nonwhite victims on average, and that the distribution of age is tighter around the median for nonwhite victims compared to white victims. We expand this conclusion with analysis of variance methods, determining with 95 percent overall confidence that white victims are typically 4-10 years older than Black victims and 5-12 years older than Latino victims, and that there is no meaningful difference in median age between Black and Latino victims. All our conclusions for this variable pair are very strong, so each nonparametric test came to the same conclusion as its parametric counterpart. However, the nonparametric conclusions are more useful because they are much more robust.

We also estimate with 95 percent confidence that armed victims are between 0 and 6 years younger than unarmed victims on average. We find with a significance level of .1 that unarmed victims are between 0 and 9 years older than knife-wielding victims on average. On the other hand, we find no differences in the distributions of tract-level poverty rate based on whether or not the victim was armed, nor based on whether the victim specifically wielded a firearm, a knife, or no weapon. Nonparametric methods generally yield more conservative p -values in these analyses, but the Miller jackknife procedure comes closer to rejecting the null hypothesis of equal dispersion.

Finally, we find with 95 percent confidence that, on average, white victims were killed in tracts with average personal incomes between \$608 and \$4623 higher than in the tracts of Black victims, and between \$186 and \$4816 higher than in the tracts of Latino victims. Distribution-free techniques prove considerably more confident than their counterparts for this pair of variables.

Throughout our analyses, the nonparametric procedures proved to be resilient to both outliers and atypical distribution variances compared to their parametric cousins. Our data provided good examples of how transformations are unnecessary for rank-based nonparametric tests, giving them an advantage in this setting.

4.1 Further Discussion

Our conclusions point to other worthwhile investigations using these data. One interesting thread to pull would be how closely the victim population represents the general population. For instance, to look for significant racial disparities, one could group victims geographically and compare the racial proportions to the general population of each area using variables like *share_white*, *share_black*, and *share_hispanic*. There is certainly no way to design an ethical experiment about police use of deadly force, so future analysis would have to remain in the realm of relations without making inferences about causations.

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>.