

# Analyzing the Length of Sentencing between 2 Crimes in Texas

Group 2: Elizabeth Bekele, Gabriel Rodriguez, Nugyen Pham, Davin Dillon

## Introduction

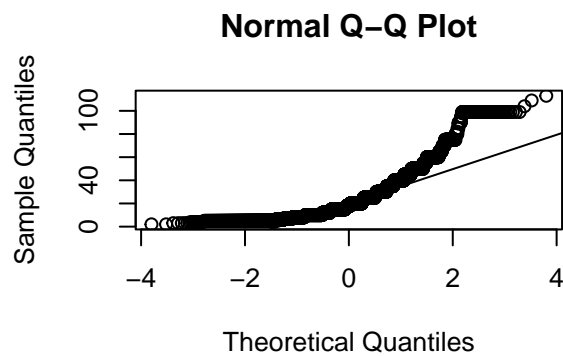
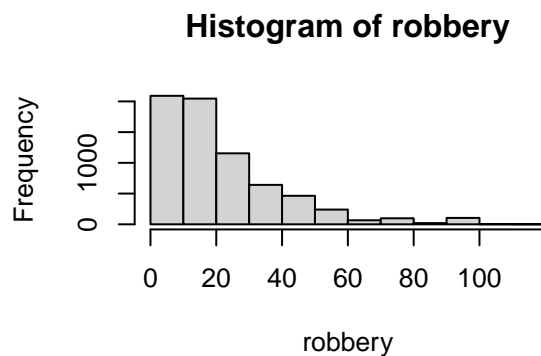
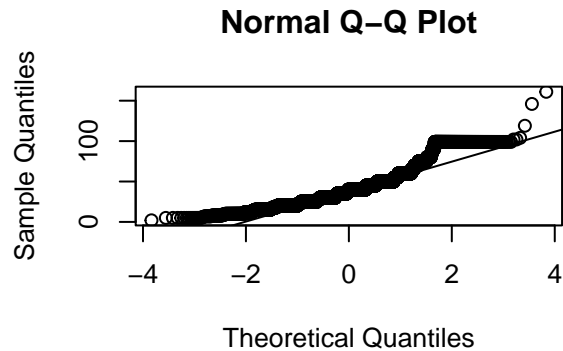
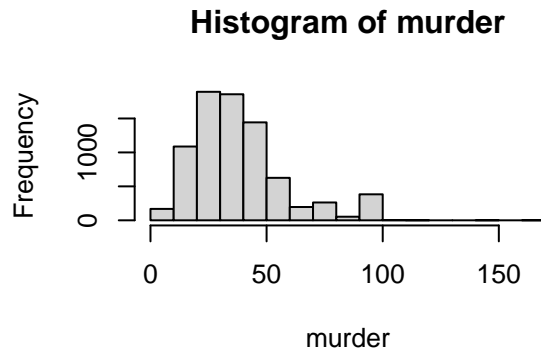
The data utilized in our analysis consists of attributes that describe current inmates that are in jail in the state of Texas. Our data is provided by the Texas Department of Criminal Justice (TDCJ) and is updated every 6 months. TDCJ oversees inmates in a variety of correctional facilities managed by the TDCJ. Our data contains information on just over 125,000 inmates. Some of the columns in the file include gender, race, offense, date of the offense, and many more. We will tailor the data set to contain only the information we need to answer our problem of interest.

Through the use of statistical methods such as interval estimation, resampling, and hypothesis testing, we want to investigate whether the Texas justice system is fair in the sentencing length of inmates who commit similar crimes regardless of their age, gender, and race. We will explore the average years sentenced for both murder and aggravated robbery. We chose these two crimes because they were the most populated in our data set. We will look into if the average years sentenced is equal or not among different subsets such as, gender, race, and age.

The justice system has the preconceived notion that you are innocent before proven guilty. Thus, if two individuals were to commit the same crime they ought to serve the same sentencing. However, time and time again we hear stories about inmates who are wrongfully committed. The Texas Innocence Project is a nonprofit organization that works towards correcting past injustices and preventing future wrongful convictions. Currently, they have exonerated 29 people. This should be alarming because there are people who are losing decades of their life due to implicit bias, lack of evidence, and other factors. Bias is an attitude towards an individual or group of people based on one's preconceived opinion without evidence. It often affects how we judge or treat a group of people. Implicit bias is a type of bias that occurs automatically without being aware of it. Both the judge and jury are given the responsibility of not holding strong ties in favor for or against the defendant prior to hearing all of the facts presented in a court of law. However, we cannot take into account an individual's implicit bias. With our project we hope to shed light to any injustices stemming from implicit bias when the sentencing of a defendant is determined. If injustice can be detected within this data set based in Texas, it's possible that it's occurring in other parts of the nation as well.

## Explore data

In this data, we are focusing on two groups: Murder and Aggravated Robbery. We want to infer upon the true mean years sentenced for each crime using interval estimation.



Based on the histogram of sentencing years for murder, we can see that it's skewed to the right and not normal. To confirm this, we conducted a QQ-Plot of the data and it aligns with our suspicions. Since the data is not normal we cannot use a method that relies on normality of the data. We instead will conduct a bootstrap t confidence interval.

The distribution for sentencing years of inmates who committed aggravated robbery is skewed to the right and not normal. The QQ-Plot confirms this because majority of the data does not fall upon the line. We will use bootstrap t to construct our confidence interval.

## Tier 1: Average Years Sentenced Based on Crime

### Murder

```
##      97.5%      2.5%
## 40.31073 41.21873
```

Our interval for murder comes out to be [40.31, 41.22]. Thus, we are 95% confident that the true mean years sentenced for an individual who has committed murder in Texas lies between 40.31 years and 41.22 years.

### Aggravated Robbery

```
##      97.5%      2.5%
## 23.29969 24.18444
```

Our interval for aggravated robbery is [23.30, 24.18]. We are 95% confident that the true average years sentenced for aggravated robbery in Texas is between 23.30 years and 24.18 years.

## Difference in Sentencing Years

Now, we want to infer upon the difference in true means between murder and aggravated robbery in Texas. To do this, we will construct a two sample bootstrap t confidence interval.

```
##      97.5%      2.5%  
## 16.39992 17.65747
```

We are 95% confident that the years sentenced for murder in Texas are, on average, between 16.40 years and 17.66 years higher than individuals who are sentenced for aggravated robbery. Based on this interval, we can also conclude that there is a significant difference in the average years sentenced for murder and aggravated robbery because the interval does not contain zero.

## Tier 2: Gender Statistics

Essentially we want to explore if the average years sentenced for a crime differs based on gender. Since the sample distribution of inmates who committed murder is not normal a subset from the non-normal distribution will also not be normal. The same is true for inmates who committed aggravated robbery. With this being said, to construct a range in which the true average years sentenced based on both crime and gender lies we will be using interval estimation. Our choice of interval estimation will be the bootstrap t confidence interval because it is better suited for skewed data.

### Average Murder Sentences

First, we will construct a bootstrap t confidence interval for the true average years sentenced in males who committed murder in Texas.

```
##      97.5%      2.5%  
## 40.45705 41.41681
```

Based on the interval, we are 95% confident that the true average sentencing years for men who committed murder in Texas is within 40.457 years and 41.417 years. Now, we will look at the confidence interval for females who committed murder in Texas.

```
##      97.5%      2.5%  
## 37.18063 40.26162
```

We are 95% confident that the true average sentencing for women who committed murder in Texas is between 37.181 and 40.262 years. Looking at the intervals alone, we can see that women are often given a lighter sentence in comparison to men who committed murder in Texas. However, to determine if this difference is truly significant we will construct a two sample bootstrap t confidence interval.

We are 95% confident that the years sentenced for males who committed murder are, on average, 0.745 to 3.858 years higher than women who committed murder in Texas. There is a significant difference between male and female inmates who committed murder because the interval does not contain zero.

## Average Aggravated Robbery Sentences

We are now going to construct a bootstrap t confidence interval for the true average years sentenced for men who committed aggravated robbery in Texas.

```
##      97.5%      2.5%  
## 23.57447 24.49007
```

We are 95% confident that the true average years sentenced for men who committed aggravated robbery in Texas is between 23.574 and 24.490 years. Now we will look at the confidence interval for the sentencing years of women who committed aggravated robbery.

```
##      97.5%      2.5%  
## 15.86751 18.75698
```

Our 95% confidence interval is (15.868, 18.757). We are 95% confident that the true average sentencing years for women in Texas who committed aggravated robbery lies between 15.868 years and 18.757 years. The confidence interval for women is shifted more to the left in comparison to men who were sentenced for aggravated robbery. This indicates that on average women are sentenced less years for aggravated robbery compared to men. However, to determine if it is a significant difference or just due to chance we will construct a two sample bootstrap t confidence interval for the difference in means.

```
##      97.5%      2.5%  
## 5.370379 8.332116
```

Our 95% confidence interval for the difference in means is (5.370, 8.332). We are 95% confident that the sentencing years for men who commit aggravated robbery are, on average, 5.370 to 8.332 years higher than women who commit aggravated robbery in Texas. There is a significant difference between male and female inmates who were sentenced for aggravated robbery because the interval does not contain zero.

## Tier 3 - Race Statistics

We will now infer statistics from our inmate data based on race. There are a total of 7,975 inmates who have been sentenced for murder. For the purposes of our statistical analysis we have grouped inmates into two race categories: "white" and "minority". There are a total of 1,877 white inmates and a total of 6,068 minority inmates. First, we want to calculate a confidence interval for the true mean sentenced years for both groups, and secondly we want to determine whether the true mean years sentenced are equal for both groups.

### Mean Sentenced Years

In order to help us measure the uncertainty of the true mean sentenced years for inmates convicted for murder, we have chosen to calculate a confidence interval for said true mean. More specifically, and due to the lack of normality in our data, we will calculate our confidence interval using the bootstrap T method.

#### Murder

```
## [1] "Murder - Population Mean Sentenced Years"
```

```
## [1] "White Inmates:"
```

```
##      97.5%      2.5%  
## 42.25285 44.11873
```

```
## [1] "Minority Inmates:"
```

```
##      97.5%      2.5%  
## 39.51002 40.52981
```

Based on our analysis for murder convicts, we are 95% confident that the true average years sentenced for white inmates convicted for murder in Texas is between 42.25 and 44.12 years. For minorities who committed murder we are 95% confident that the true average years sentenced lies between 36.51 and 40.53 years.

### Aggravated Robbery

```
## [1] "Aggravated Robbery - Population Mean Sentenced Years"
```

```
## [1] "White Inmates:"
```

```
##      97.5%      2.5%  
## 23.24536 25.42612
```

```
## [1] "Minority Inmates:"
```

```
##      97.5%      2.5%  
## 23.16006 24.11453
```

Based on our analysis for aggravated robbery convicts, we are 95% confident that the true average years sentenced for white inmates convicted for aggravated robbery in Texas is between 23.25 and 25.43 years. For minorities who committed aggravated robbery we are 95% confident that the true average years sentenced lies between 23.16 and 24.11 years.

### Significance of Difference of Means:

We now want to test if there is any significant difference between the true mean sentenced years for white inmates and minority inmates. No significant differences between the true mean sentenced years would indicate that all inmates are, on average, sentenced to the same number of years, for similar offenses, regardless of race. On the other hand, a significant difference would indicate that there is some race bias in true mean sentenced years for similar offenses. We will be testing for significance using a two-sample bootstrap t confidence interval. A two-sample bootstrap t confidence interval was chosen as our data meets the following assumptions: our two sample populations are not normally distributed, the sample population sizes are dissimilar, and we will also test the population variances for equality.

Check variances to satisfy model assumption

```
## [1] 0.278
```

```
## [1] 0.5
```

At a 5% significance level we fail to reject the null for both murder and aggravated robbery sample populations, and conclude that the true population variances are equal, and thus we can continue with our two-sample bootstrap t confidence interval.

```
## [1] "Murder - Difference Between True Mean Sentenced Years"
```

```
## [1] "White Inmates vs. Minority Inmates:"
```

```
##      97.5%      2.5%
## 2.101843 4.255000
```

```
## Observed difference: 3.170953
```

We are 95% confident that the years sentenced for white inmates who committed murder in Texas are, on average, between 2.10 and 4.26 years longer than that of minority inmates who committed murder. This allows us to conclude that our observed difference of 3.17 years is statistically significant because our confidence interval does not contain 0.

```
## [1] "Aggravated Robbery - Difference Between True Mean Sentenced Years"
```

```
## [1] "White Inmates vs. Minority Inmates:"
```

```
##      97.5%      2.5%
## -0.5192875 1.8932127
```

```
## Observed difference: 0.6553207
```

We are 95% confident that the true average difference in years sentenced for aggravated robbery among white and minority inmates in Texas is within -0.52 and 1.89 years. The observed difference of 0.66 years is not statistically significant because our confidence interval contains 0.

## Exploring Age Groups

The age that we are concerned with is not the inmates' current age, but rather the age they were given the time of their sentencing. We decided to group the ages at the time of sentencing into 5 bins ranging from 12 to 86 years old. The five categories are listed below.

```
##      group Age Categories
## 1      A      12-26
## 2      B      27-41
## 3      C      42-56
## 4      D      57-71
## 5      E      72-86
```

For a basic overview we looked at the observed average years sentenced per age group for murder in Texas.

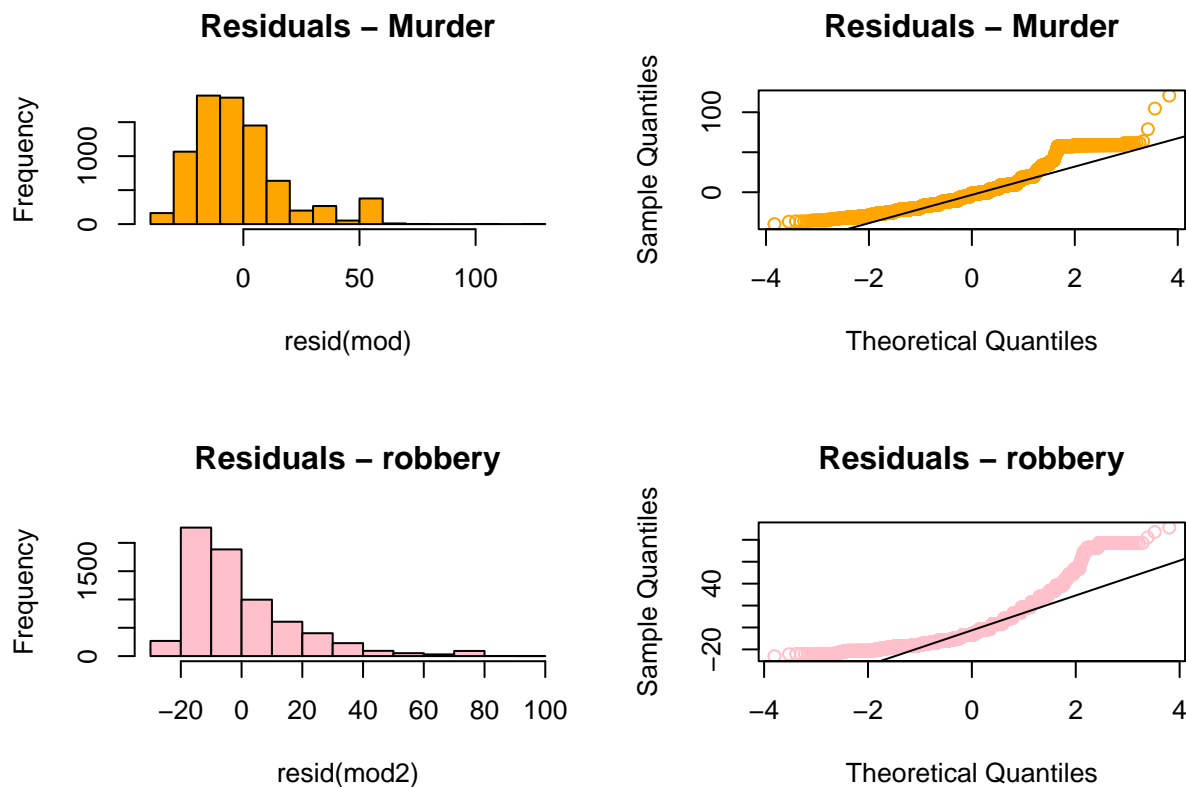
```
##      group Sentenced Age Average.Yrs.Sentenced
## 1      A      12-26      40.21280
## 2      B      27-41      41.38475
## 3      C      42-56      41.82964
## 4      D      57-71      37.74860
## 5      E      72-86      21.07692
```

Based on the data, inmates who were sentenced between the age of 42 and 56 received the longest average sentencing of 41.83 years. While inmates who were sentenced at the age of 72-86 received the shortest average sentencing of 21.08 years. The difference is well over 20 years, however, we cannot make conclusions based on these point estimates alone.

## Tier 4: Does Age play a factor in the amount of Years Sentenced?

### Murder

To determine if age plays a role in the severity of an inmates sentencing we will be conducting a permutation ANOVA test. Below you can see that the model residuals for both murder and aggravated robbery are not normal.



Permutation does not require normality of the residuals, however, requires the population variances to be common. To check this assumption we will apply the `equa4vartest()` twice and the `equa2vartest()` once to account for all 5 age groups.

```
## Murder (age groups) - Homogeneity in Variances Test:
```

```
## Groups A-D P-value: 0.94
```

```
## Groups A-C & E P-value: 0.055
```

```
## Groups D & E P-value: 0.067
```

In all 3 tests, we are able to conclude the population variances for inmates who committed murder are equal among the different age groups because the P-values are greater than alpha at the 5% significance level. We can now proceed to conduct the permutation ANOVA test. Our hypotheses are as follows:

$H_0$  : the average sentenced years from all age groups among murder convicts are equal

$H_A$  : the average sentenced years from all age groups among murder convicts are not equal

Since we are performing a permutation ANOVA test we will be permuting, or reordering, the indices of the Sentenced.Years column whilst keeping the age groups fixed.

```
## [1] 3e-04
```

After performing the permutation ANOVA test we reach a P-value of 3e-04 which is approximately 0. We reject the null hypothesis at a 5% significance level and conclude that age does have a significant effect on the true average years sentenced for an inmate who committed murder.

To inspect which age group(s) have the most effect on the average years they are sentenced for murder we will conduct a Pairwise Permutation Test.

The four lowest adjusted P-values include:

- 12-26 vs 72-86 at an adjusted P-value of 0.002494;
- 27-41 vs 72-86 at an adjusted P-value of 0.001851;
- 42-56 vs 72-86 at an adjusted P-value of 0.001851;
- 57-71 vs 72-86 at an adjusted P-value of 0.010520.

The common age group in these four comparisons is 72-86. Inmates who were convicted between the ages of 72 and 86 for murder are more likely to be treated differently in how long they are required to serve their sentencing. This age group is one of the reasons that the true average sentencing for murder is different among all age groups.

Now, we will look at how age plays a role in inmates who were convicted for aggravated robbery. Below, are the observed average sentences per age group for inmates who committed Aggravated Robbery.

##	group	Sentenced Age	Average.Yrs.Sentenced
## 1	A	12-26	21.86174
## 2	B	27-41	26.31495
## 3	C	42-56	29.09188
## 4	D	57-71	22.63750
## 5	E	72-86	37.50000

Based on these point estimates, inmates who committed aggravated robbery and were sentenced at the age of 72-86 years old were given the longest average sentencing of 37.5 years. The age group that received the shortest average sentencing at 29.09 years were inmates who committed aggravated robbery between the ages of 12 and 26.

Before conducting our method of testing, we will check the equality in variances for each age group.

```
## Aggravated Robbery (age groups) - Homogeneity in Variances Test:
```

```
## Groups A-D P-value: 0.208
```

```
## Groups A-C & E P-value: 0.002
```



## Groups D & E P-value: 0.107

After conducting our test, we can conclude that the 5 age groups do not share equality in variances. The age groups A, B, C, and E resulted in a P-value below 0.05 leading us to reject our null hypothesis at a 5% significance level.

For this reason we will conduct a two-sample bootstrap t test. In the first test, we will compare age groups A and B. In the second test, we will compare age groups C and D. We will omit age group E (72-86) because there are only 2 inmates who committed aggravated robbery and fall in this age group.

## Aggravated Robbery

Are young people given a shorter sentence for aggravated robbery compared to an older age group? First we will look to see if the difference in the average sentencing for inmates sentenced between the ages of 12-26 is less than inmates sentenced between the ages of 27-41 for aggravated robbery. Our hypotheses are as follows:

$$\theta_d = \mu_{AR:12-26} - \mu_{AR:27-41}$$

$$H_0 : \theta_d \geq 0$$

$$H_A : \theta_d < 0$$

## [1] 9.999e-05

Our test results in a P-value of approximately 0. At a 5% significance level we reject the null hypothesis and conclude that the average years sentenced for inmates who committed aggravated robbery between the ages of 12 and 26 is less than inmates sentenced between the age of 27 and 41.

Now we will observe if the difference in the average sentencing for inmates sentenced between the age of 42-56 is less than inmates sentenced between the ages of 57-71 for aggravated robbery. Our hypotheses are as follows:

$$\theta_d = \mu_{AR:42-56} - \mu_{AR:57-71}$$

$$H_0 : \theta_d \geq 0$$

$$H_A : \theta_d < 0$$

## [1] 0.9789021

After conducting a 2-sample bootstrap t test we obtain a P-value of 0.9789. At a 5% significance level we fail to reject the null hypothesis and conclude that the average years sentenced for inmates who committed aggravated robbery between the ages of 42 and 56 is greater than or equal to inmates sentenced between the age of 57 and 71.

## Conclusion

We divided our analysis into four tiers. In the first tier we investigated the average sentencing length for murder, aggravated robbery, and the difference among the two. We concluded that since the distribution of both murder and aggravated robbery inmates are not normal we constructed a bootstrap t confidence interval. We observed that inmates who committed murder were given longer sentences than inmates who committed aggravated robbery.

In the second tier we wanted to know how gender affected the length of sentencing for inmates who committed the same crime. A two-sample bootstrap t confidence interval was formed. We were able to see that on average men served longer sentences than women for both murder and aggravated robbery.

Since there were significant findings between the sentencing length and gender we wanted to know if the same were true for race and age. This led us to the third tier where we explored the average sentencing length among two groups, white and minority inmates. A two-sample bootstrap t confidence interval was created. We concluded that white inmates are, on average, serving longer sentences than minority inmates for murder, however, the same was not true for aggravated robbery.

Lastly, in the fourth tier we explored how five different age groups affected the length of sentencing among inmates who committed the same crime. Since the residuals were not normal a permutation ANOVA test was performed for murder inmates. We observed that age is a significant factor in the length of sentencing. For aggravated robbery a bootstrap t hypothesis test was performed because the population variances were not equal. We were not able to conclude that younger convicted felons who committed aggravated robbery in Texas were given a shorter sentence compared to older convicted felons.

If further analysis were to continue in the future, we suggest looking into the files of the inmates. Our data does not include the severity of the crime or if the inmate has committed previous crimes. Moreover, finding ways to prevent outliers and skewness in the race and age distributions would assist in providing more insight to the questioning of Texas' legal system.

## Sources

“Inside TDCJ.” Texas Department of Criminal Justice, Texas Department of Criminal Justice, [https://www.tdcj.texas.gov/kss\\_inside.html](https://www.tdcj.texas.gov/kss_inside.html).

“Releases & Exonerations Archives.” Innocence Project of Texas, [https://innocencetexas.org/case\\_type/releases-exonerations/](https://innocencetexas.org/case_type/releases-exonerations/).

“Implicit Bias.” National Institutes of Health, U.S. Department of Health and Human Services, <https://diversity.nih.gov/sociocultural-factors/implicit-bias>.

Chihara, Laura, and Tim Hesterberg. “Introduction to Hypothesis Testing: Permutation Tests, Introduction to Confidence Intervals: The Bootstrap.” *Mathematical Statistics with Resampling and R*, 2nd ed., John Wiley & Sons, Inc, Hoboken, NJ, 2019, pp. 50–119.