Race, Experiences, and Criminal Sentencing

Julie Norman, Abraham Alhomadi, and Keaton Manwaring

12/6/2020

In this paper, we will undertake an econometric examination using the Crime1 dataset in the Wooldrige Package. The dataset contain a total of 2725 observations on 16 distinct variables.It was collected in 1986, in the State of California. Looking at the data, we will aspire to answer the following research question that is the focus of this paper: “How does an individual’s income affect average sentencing outcome when they commit a crime? This question has been the subject of policy research for decades and, using the Crime1 dataset, we would like to contribute to the conversation on this issue of economics and criminal justice reform. Our belief as citizens is that we have a collective responsibility to promote better policies and the well-being of our nation through the efficient and just allocation of resources and ensuring that our criminal justice system is in fact”just" in its treatment of citizens who are convicted of a crime. In addition to our focus research question, we will investigate other equally critical questions, some of which include,“Does a person’s employment status increase or decrease average sentencing?”, " Do blacks get different average sentence lengths than nonblacks?“, and”how does an individual’s prior criminal history, if any, contribute to their average sentencing length?“.

Now, to carry out our investigation, we will begin by focusing our annalysis on the the 134 individuals in our dataset who have been sentenced. We’re filtering out the individuals who have not served prison time as our question is focused on what factors impact sentence length, not whether they’ve been found guilty at all. We’ll regress our dependent variable “avgsen”, which is the average sentence length of a convicted criminal, on variant explanatory variables: inc86 (legal income), nfarr86 (felony arrests), pcnv (proportion of prior convictions), durat (recent unemp duration), narr86(times arrested), Black, and Hispanic. Since our data treats black and hispanic as mutually exclusive, we’re going to combine our race variable into a single dummy that illustrates whether an individual is black, hispanic, or neither. Our multiregression model would start with the following model and build on from there to test for interaction between variables and to reflect for possible increase or decrease in sentencing length in the short- and long-term. This is our original regression model:

The additional questions will help put our focus question in perspective, and allow us to identify alternative explanations besides an individual’s income to why average sentencing length would differ from one person to another, given variables like employment status, criminal history, and race.

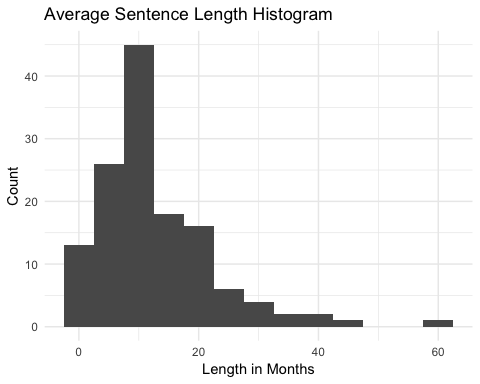
Lastly, for so long, policy makers have been trying to answer the above questions in order to improve economic and criminal justice policy making. They endeavor to identify our criminal justice system flaws and seek to address them, including issues like sentencing discrimination on the basis of income and race, and to improve the efficient and fair allocation of economic resources to all their citizens. This is an interesting economic inquiry, for it touches the heart of one of the challenges that has engulfed our nation. A stable and functioning economic system and a fair criminal justice system are important to maintaining a healthy life style, public safety to guard against social division because of discriminatory economic and criminal justice motives, and to promote a flourishing, economically strong country in the 21st century. For this reason, we are seeking to investigate whether average sentencing in, say, California or other states for that matter, are rendered based on effective, just, and fair laws, or are there discriminatory motives that impact our justice system rendering of sentencing. If discrimination does exist, then as citizens, we have a solemn obligation to put forth facts backed up by objective data and partake in making our country, as our founding fathers envisioned, a “more perfect union.”

Our Crime1 dataset is included in the wooldridge package. We did not collect the data on our own nor do we know precisely how the data was collected. However, we know that the data it contains was collected in the State of California and the arrests data are for the year 1986. The observations reflect men who were born in 1960 or 1961, and the subject men in the dataset all had criminal history and a minimum of one arrest before 1986. So, our regression model, as written above, will provide a multidimensional analysis to explain whether our independent variables listed do or do not have a causal relationship with our dependent variable, Avgsen. We wanted to include gender as a dummy variable in our regression model, but the dataset only contains data on men. Therefore, instead of looking at the relationship between gender and average sentencing, we will dwell on race and average sentencing.

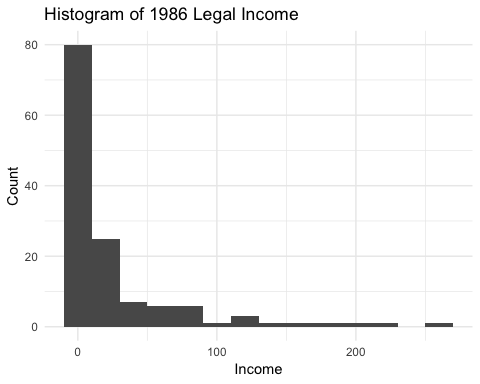
## Data Exploration

In order to better understand our data we’ll look at each variable to check for any violations of the first Gauss Markov assumption: linear parameters. While there’s logical expanations for relationships between our explanatory variable and our response variables, we also need to ensure that our data bares out that relationship.

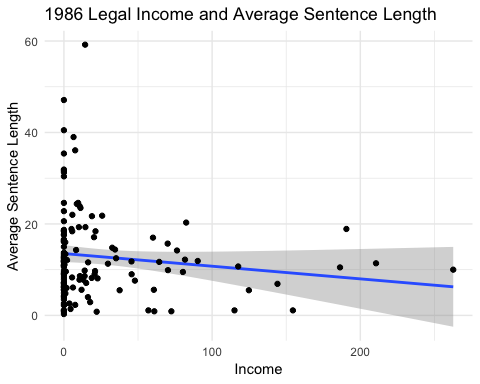
### Average Sentence Length

 The histogram shows that that the distribution of sentence length is bell shape with a slight right skew indicating that a log might be an appropriate adjustment for our model. However, since there are zero values in the variable a log transformation isn’t appropriate.

### Income from 1986

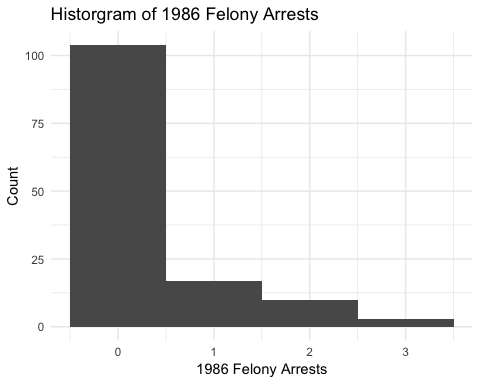


The histogram shows a strong right hand skew indicating that in order to meet the linear parameters assumption, we will need to take the log of the variable. However, since there are so many 0 values we won’t be transforming the data Next we need to visualize the relationship between the income and average sentence length.

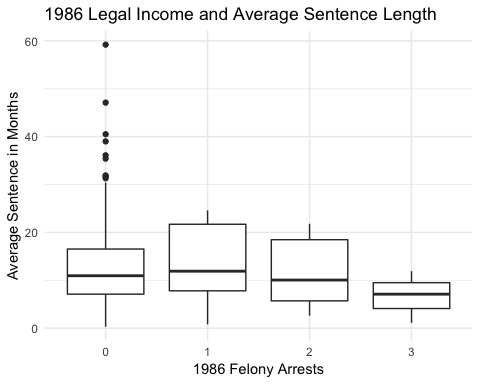


The model indicates a week negative association between the income variable and the sentence variable. Due to this, we shouldn’t expect a strong impact of this variable in our model.

### Felony arresting

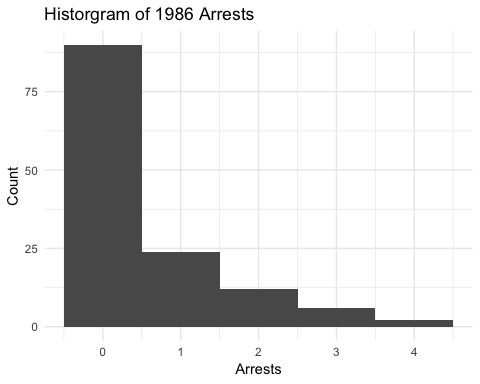


The histogram indicates a strong right hand skew in felony arrests. However, the ordinal nature of the data makes it unideal for a multiple regression model.

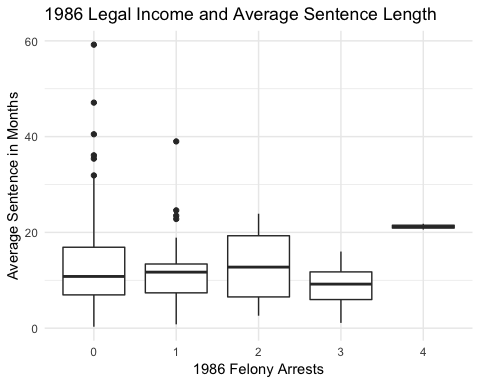


The side by side boxplots illustrates that amoung the average sentence length for each number of felony arrests has essentially no association on average. Due to this and the ordinal nature of the variable we should consider removing this variable from our model.

### Number of Arrests in 1986

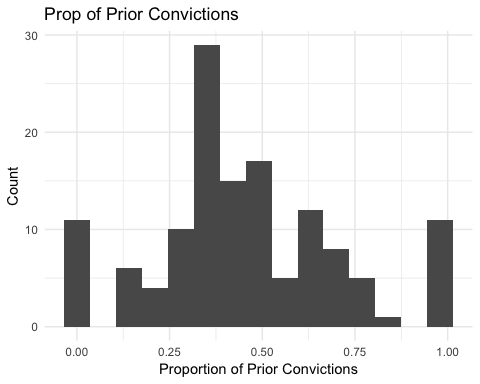


The histogram indicates a strong right hand skew in arrests. However, the ordinal nature of the data makes it again unideal for a multiple regression model.

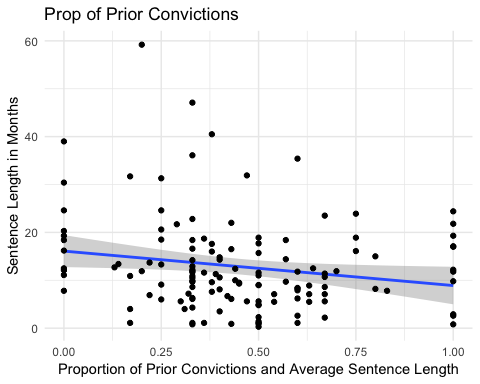


Again, he side by side boxplots indicates that amoung the average sentence length for each number of arrests there is essentially no association. While the average sentence for those with 4 felony arrests is higher than the rest, this is merely one data point and as it does not imply an overall trend in the data it’s unlikely to be influencial in our model. Due to this and the ordinal nature of the variable we should consider removing this variable from our model.

### Proportion of prior convictions

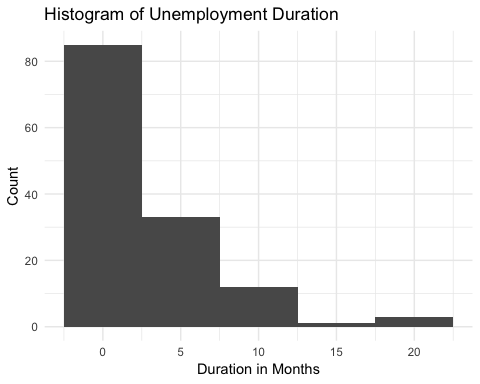


The histogram indicates that the distribution of th proportion of prior convictions is approximately normal, indicating that this variable is a good candidate for use in a multiple regression model.

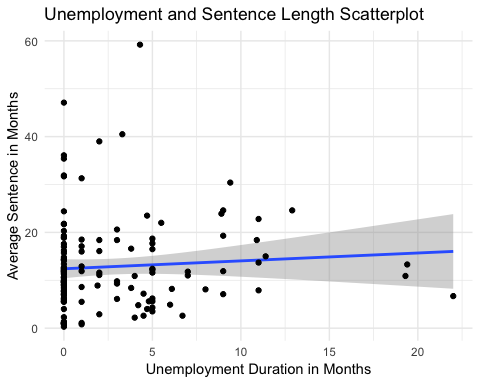


This scatterplot indicates a negative association between prior convictions and sentence length. This indicates that as convictions become more common for an individual, their sentence length decreases. This trend is particularly interesting as it does fit a typically assumption of repeat offenders having harsher punishments. While we can only speculate as to the exact nature of this relationship, its possible that other variables such as the types of crimes committed and/or race have a larger impact on average sentence then we wouldn’ve prevously thought. That makes this variable an ideal candidate for a multiple regression since we can see how this variable combined with other indicators appropriate predict average sentence length.

### Duration of Unemployment

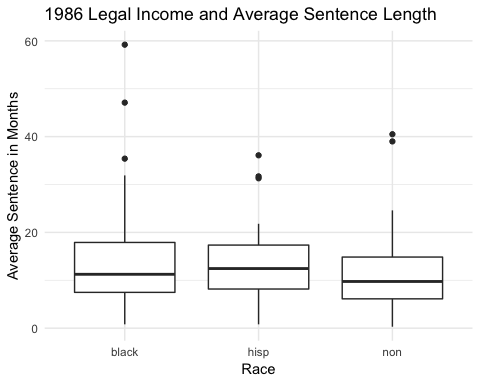


The histogram indicates a strong left hand skew in unemployment duration. This indicates that a logarithmic model might be more appropriate for this variable, however since we have zero values in our data a log isn’t appropriate. The measurement unit for duration of unemployment is not given in our data set but since all the other time variables are measured in months we’re assuming that here as well.



This scatterplot illustrates a weak positive association between unemployment length and average sentence length. Due too the weakness of the association

### Race Comparisons



The boxplots of average sentence length show minmimal variation accross different racial categories. This indicates that even if there is a statistical significance between the regressions, it is likely of little practical significance. However, since race is a major point in our research question we’ll keep this variable in our regression.

## Model Specification

Since we’re considering eliminating two of the variables, it’s important we include them in our models until we’ve determined the best form for the other variables in our model.

During our exploration of the variables, we found that due to skewness, a few of our variables could benefit from logs. However, since there are zero values in all of our variables this isn’t an option. However, just to be safe we’ll compare a variatey of quadratic functions on the variables that weren’t as skewed to see if it improves our model.

##   
## ==========================================================================================  
## Dependent variable:   
## -----------------------------------------------------------  
## avgsen I(avgsen + avgsen2) (avgsen) I(avgsen + avgsen2)  
## (1) (2) (3) (4)   
## ------------------------------------------------------------------------------------------  
## inc86 -0.019 -1.018 -0.020 -1.048   
## (0.019) (0.917) (0.020) (0.920)   
##   
## narr86 1.317 2.325 1.405 6.179   
## (1.661) (78.376) (1.668) (78.597)   
##   
## nfarr86 -2.979 -77.192 -3.081 -81.641   
## (2.158) (101.853) (2.168) (102.163)   
##   
## pcnv -6.520\*\* -306.316\*   
## (3.290) (155.278)   
##   
## I(pcnv + pcnv2) -2.592 -131.234\*   
## (1.626) (76.594)   
##   
## durat 0.143 0.741 0.144 0.754   
## (0.209) (9.871) (0.210) (9.911)   
##   
## racehisp -1.561 -89.443 -1.593 -91.831   
## (2.240) (105.721) (2.255) (106.237)   
##   
## racenon -3.119 -132.065 -3.161 -133.946   
## (1.944) (91.767) (1.955) (92.095)   
##   
## Constant 17.777\*\*\* 531.228\*\*\* 16.721\*\*\* 488.744\*\*\*   
## (2.232) (105.355) (2.051) (96.650)   
##   
## ------------------------------------------------------------------------------------------  
## Observations 134 134 134 134   
## R2 0.085 0.072 0.075 0.065   
## Adjusted R2 0.034 0.020 0.024 0.013   
## Residual Std. Error (df = 126) 9.511 448.855 9.562 450.516   
## F Statistic (df = 7; 126) 1.673 1.388 1.464 1.246   
## ==========================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

The strongly indicate that our non quadratic model is the best model to work on.

Next we’ll compare models with and without narr86 and nfarr86 to see if their impact justifies keeping them in the model.

##   
## ===========================================================  
## Dependent variable:   
## ---------------------------------------  
## avgsen (avgsen)   
## (1) (2)   
## -----------------------------------------------------------  
## inc86 -0.019 -0.017   
## (0.019) (0.019)   
##   
## narr86 1.317   
## (1.661)   
##   
## nfarr86 -2.979   
## (2.158)   
##   
## pcnv -6.520\*\* -6.655\*\*   
## (3.290) (3.290)   
##   
## durat 0.143 0.111   
## (0.209) (0.208)   
##   
## racehisp -1.561 -1.088   
## (2.240) (2.198)   
##   
## racenon -3.119 -3.000   
## (1.944) (1.939)   
##   
## Constant 17.777\*\*\* 17.420\*\*\*   
## (2.232) (2.146)   
##   
## -----------------------------------------------------------  
## Observations 134 134   
## R2 0.085 0.068   
## Adjusted R2 0.034 0.032   
## Residual Std. Error 9.511 (df = 126) 9.524 (df = 128)   
## F Statistic 1.673 (df = 7; 126) 1.866 (df = 5; 128)  
## ===========================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

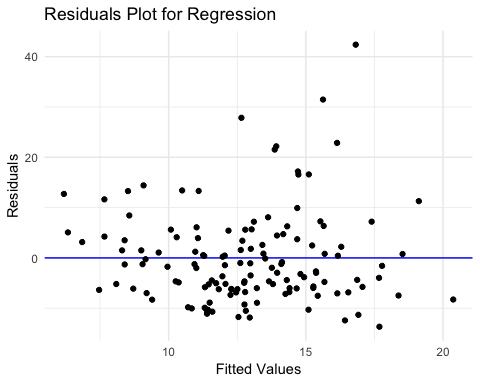
Including narr86 and nfarr86 not only increases our adjusted by 3 percentage points but they are also both statistically significant at the 90% confidence level so we’ll keep them in our model. As there is no indication in our scatterplots for a quadratic form, which we’ll confirm later, for now we can consider the following regression our current final model

## Model Assumptions

We used an Ordinary Least Square (OLS) estimating method to compute the coefficients of our parameters. Our model complies with all five of OLS multiple regression assumptions. First, our regression model is linear in parameters, as illustrated by our model above. Additionally, the sampling of the observations was undertaken randomly so we’ve met that condition as well. The third condition is that there’s no multicollinearity. Since the independent variables is 93.6621494 which is greater than zero and we have no variable repetition so we just need to check that we have no variables that are perfect multiples of each other

## $`Regress on inc86`  
## GVIF Df GVIF^(1/(2\*Df))  
## nfarr86 3.512049 1 1.874046  
## narr86 3.552611 1 1.884837  
## pcnv 1.013684 1 1.006819  
## durat 1.038258 1 1.018949  
## race 1.104268 2 1.025106  
##   
## $`Regress on nfarr86`  
## GVIF Df GVIF^(1/(2\*Df))  
## inc86 1.172797 1 1.082957  
## narr86 1.057301 1 1.028251  
## pcnv 1.014088 1 1.007019  
## durat 1.123485 1 1.059946  
## race 1.101730 2 1.024516  
##   
## $`Regress on narr86`  
## GVIF Df GVIF^(1/(2\*Df))  
## inc86 1.166666 1 1.080123  
## nfarr86 1.039765 1 1.019689  
## pcnv 1.013268 1 1.006612  
## durat 1.129069 1 1.062577  
## race 1.081621 2 1.019809  
##   
## $`Regress on pcnv`  
## GVIF Df GVIF^(1/(2\*Df))  
## inc86 1.170134 1 1.081728  
## nfarr86 3.505471 1 1.872290  
## narr86 3.561708 1 1.887249  
## durat 1.130924 1 1.063449  
## race 1.147090 2 1.034902  
##   
## $`Regress on durat`  
## GVIF Df GVIF^(1/(2\*Df))  
## inc86 1.073780 1 1.036234  
## nfarr86 3.479486 1 1.865338  
## narr86 3.555754 1 1.885671  
## pcnv 1.013236 1 1.006596  
## race 1.126372 2 1.030197  
##   
## $`Regress on Black`  
## inc86 nfarr86 narr86 durat hispan   
## 1.122249 3.505470 3.552835 1.115451 1.080318   
##   
## $`Regress on Hispanic`  
## inc86 nfarr86 narr86 durat black   
## 1.165819 3.375976 3.349365 1.111732 1.052138

The highest variance inflation factor of all the parameter regressions is 3.555754 while most are between one and two. Values between 1 and 5 are considered moderately correlated while values greater than 5 near perfect collinearity. Since our strongest correlation is a moderate correlation we’ve shown there’s no multicollinearity and we’ve met this assumption. The next assumption is the zero conditional mean assumption ie . We’ve already added logs to the model where appropriate and since the visualizations of our variables don’t indicate a need for quadratics in our model, we can double check our model specificatioon by looking at the model residuals.



Since the residuals are more clustered below. The final assumption is homoskedasticity, in otherwords that . Our visualizations don’t indicate any clear signs of homoskedasticity but to verify this we’ll conduct a breusch-pagan test and a white test. The hypothesis test is as follows: $ H\_{0}: Var(u|inc, narr86, nfarr86, pcnv, durat, race) = Var(u|X) = ^2$ and the alternative where these values are not equal.

First we will use the BP test using the F statistic, then White test using F statistic, and both tests using LM ratio. Then, we will discuss the difference in the results they render. Keep in mind that a large F statistic or a large Lagrange multiplier statistic, (LM) indicate a rejection of the null hypothesis.

##   
## studentized Breusch-Pagan test  
##   
## data: reg1  
## BP = 6.2918, df = 7, p-value = 0.5061

Using the BP test, we see that our p-value is relatively large (p-value = 0.197) meaning that our F-statistic is relatively small. Consequently, we fail to reject the null hypothesis. This means our unobserved distribution is heteroskedastic. Since we failed the bp test we’ll also do a white test, it is a less vigorious test for heteroskedasticity.

##   
## Call:  
## lm(formula = u\_hat^2 ~ y\_hat + I(y\_hat^2))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -173.11 -73.98 -44.20 -3.10 1657.27   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 68.4518 266.9796 0.256 0.798  
## y\_hat -10.4234 42.2433 -0.247 0.805  
## I(y\_hat^2) 0.8695 1.6329 0.532 0.595  
##   
## Residual standard error: 199 on 131 degrees of freedom  
## Multiple R-squared: 0.02986, Adjusted R-squared: 0.01505   
## F-statistic: 2.016 on 2 and 131 DF, p-value: 0.1373

We have an F statistic of 2.016 and a p value of 0.1373 >.05. Therefore we fail to reject the null hypothesis indicating we have an issue with heteroskedasticity. Due to the results of the bp and white test we will use robust standard errors to correct for the heteroskedasticity.

Below, we compare our previous regression with a new regression that includes robust standard errors. .

## [1] 6.183464 20.370425

##   
## ===========================================================  
## Dependent variable:   
## ----------------------------  
## avgsen (avgsen)   
## (1) (2)   
## -----------------------------------------------------------  
## inc86 -0.019 0.014   
## (0.019) (0.009)   
##   
## narr86 1.317 4.818\*\*\*   
## (1.661) (1.745)   
##   
## nfarr86 -2.979 -5.612\*\*\*   
## (2.158) (2.005)   
##   
## pcnv -6.520\*\* 1.548   
## (3.290) (2.581)   
##   
## durat 0.143 0.134   
## (0.209) (0.224)   
##   
## racehisp -1.561 -2.776   
## (2.240) (2.337)   
##   
## racenon -3.119 -3.508\*   
## (1.944) (1.789)   
##   
## Constant 17.777\*\*\* 12.017\*\*\*   
## (2.232) (2.536)   
##   
## -----------------------------------------------------------  
## Observations 134 134   
## R2 0.085 0.108   
## Adjusted R2 0.034 0.059   
## Residual Std. Error (df = 126) 9.511 0.067   
## F Statistic (df = 7; 126) 1.673 2.188\*\*   
## ===========================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Not only is there an increase in the adjusted by .025 points, we also have three additional statistically significant variables: narr86, nfarr86, racenon. Due to this output and the results of the robustness tests, we’ll utalise the equation with the robust standard errors. In that equation there are three numeric equations that are not statistically significant. In order to determine whether or not they should be in our model we need to determine whether or not they are jointly significant with the following hypothesis test with an F distribution.

## Linear hypothesis test  
##   
## Hypothesis:  
## inc86 = 0  
## pcnv = 0  
## durat = 0  
##   
## Model 1: restricted model  
## Model 2: avgsen ~ (inc86) + narr86 + nfarr86 + pcnv + durat + race  
##   
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 129 11975   
## 2 126 11398 3 577.31 2.1274 0.1

The results of the hyothesis test has a low F value of 2.1274 with ap value of 0.1 which indicates that these values are not jointly significant so we’ll remove those variables from our model, we can see the comparison with the removed variables here

stargazer(reg1\_wls, reg4\_wls, type = "text")

##   
## ===============================================================  
## Dependent variable:   
## -------------------------------------------  
## (avgsen)   
## (1) (2)   
## ---------------------------------------------------------------  
## inc86 0.014   
## (0.009)   
##   
## narr86 4.818\*\*\* 4.838\*\*\*   
## (1.745) (1.725)   
##   
## nfarr86 -5.612\*\*\* -6.085\*\*\*   
## (2.005) (1.963)   
##   
## pcnv 1.548   
## (2.581)   
##   
## durat 0.134   
## (0.224)   
##   
## racehisp -2.776 -2.780   
## (2.337) (2.307)   
##   
## racenon -3.508\* -2.903\*   
## (1.789) (1.700)   
##   
## Constant 12.017\*\*\* 13.969\*\*\*   
## (2.536) (1.618)   
##   
## ---------------------------------------------------------------  
## Observations 134 134   
## R2 0.108 0.088   
## Adjusted R2 0.059 0.060   
## Residual Std. Error 0.067 (df = 126) 0.067 (df = 129)   
## F Statistic 2.188\*\* (df = 7; 126) 3.123\*\* (df = 4; 129)  
## ===============================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

By removing the variables we’ve increased the adjusted and our F statistic indicating that this is the appropriate approach. So our final model is

## Model Interpretation and Implication

For every additional arrest an individual has, are model suggests an increase in average sentence length by 4.8 months. This result is statistically significant at the 99% confidence level but more importantly it has practical significantin that it supports the understanding that those under frequent suspicion tend to experience more prison time even if those arrests don’t lead to jailtime. The next variable has an unexpected association. Every additional felony arrest, our model suggests a 6.1 month decrease in average sentence length. This result is also statistically significant at the 99% confidence level. This result could stem from felonies having harsher punishments which encourages individuals to fight those charges harder in order to have a shorter sentence. Our model also includes dummy variables for race where individuals are considered hispanic, black, or neither. Relative to black individuals, being hispanic decreases average sentence length by 2.8 months and being in neither category decreases average sentence length 2.9 relative to black individuals. Suffice to say that black indivuals tend to have the longest sentences, but its particularly interesting that hispanic and non black/non hispanic individuals are only seperated by about three days difference in average sentencing. This indicates that that race plays a role in how individuals experience the criminal justice system, particularly for those of african heritage. However, the adjusted meaning that our variables can explain only 6% of the variation in average sentence length. So we know that despite our model being statistically significant at the 95% confidence lever, it is far from illustrating the entire story of an individual’s average sentence length.

## Conlcusion