Data analysis of Wrong Conviction

Name: Xinning Chu

Date: May 15, 2018

Instructor: Professor Qing Pan

George Washington University

Contents

Part 1. Introduction

According to the study of some scholars, more than 2000 people in the United states were wrongly convicted of crimes that they did not commit since 1989. These people are called ‘exonerees’, and there are two basic ways for them to seek compensation. Initially, 33 states have state statutes which permit exonerees to request compensation from a state court or state administrative body. However, not every state has such laws, while exonerees in all 51 states are able to file and prevail another kind of compensention----civil rights compensention.

A data set which contains information of the first 1900 persons wrongly convicted in a state court were created by Professor Gutman of the GW law school. There are 35 variables in it. Six of them are demographics, and the rest are tags from the National Registry. Most of these tags are binary variables. Among them six tags are the main variables of interest and chosen as response variables. Then the relationships between each response variable and explanary variables are analyzed.

In part 2, some values are calculated to explore essential features of variables, such as number of incarcerated exonerees. Not every exoneree is actually incarcerated, so knowing the percentage of incarcerated exonerees can help us have a better understanding of the background. Also, numbers of total incarcerated by race or gender are given in this part.

In part 3, three basic approaches are used to analyze the relationship between response variables and every factor. Firstly, contigency table can provide a basic picture of the interrelation between two variables and help find interactions between them. Thus, it's used to analyze the relationship between the likelihood of seeking or receiving two kinds of compensation and some explanary variables. However, in some cases, the response variable or explanary variable may be continous. Anova table and Spearman test are used as well. Based on these p-values, whether a variable is statistically significant or not can be judged.

Part 2. Simple Inquiries

Before doing this part, simple data cleaning is done. 8 cases in the data set are deleted because of the same name and information.

2.1 Numbers of Exonerees

Table 2.1.1 Number of incarcerated or not exonerees

	Incarcerated Exonerees	0-time Exonerees
Number	1710	182

Table 2.1.2 Number of incarcerated exonerees serving different time

Serving Time	1 year or less	More than 1 year
Number of incarcerated exonerees	181	1529

According to Table 1.1 and Table 1.2, the majority of exonerees were actually incarcerated, and most of them served more than 1 year. These numbers indicate that wrong conviction really had significantly negative effects on these people.

Table 2.1.3 Number and percentage of all exonerees by race and by gender

Race	Female	Male	Total	Percentage
Asian	0	13	13	6.871036%
Black	46	868	914	48.308668%
Caucasian	109	614	723	48.2135307%
Hispanic	22	199	221	11.6807611%
Native American	1	11	12	0.63424295%
Other	1	8	9	0.4756871%
Total	179	1713	1892	100%
Percentage	9.46088%	90.53911%	100%	

Table 2.1.4 Number and percentage of incarcerated exonerees by race and gender

Race	Female	Male	Total	Percentage

Asian	0	11	11	0.64327%
Black	36	823	859	50.233918%
Caucasian	84	554	638	37.30994%
Hispanic	8	176	184	10.76023%
Native American	1	9	10	0.58479%
Other	1	7	8	0.46783%
Total	130	1580	1710	100%
Percentage	7.60233%	92.39766%	100%	

From Table 2.1.3 and Table 2.1.4, Black has the largest number of both total exonerees and incarcerated exonerees, followed by Caucasian. The numbers of Asian, Native American, other races are very tiny, which can even be ignored. In addition, it is quite clear that in the whole population males are far more than females.

Table 2.1.5 Number and percentage of exonerees by race and by gender in states with a compensation statute

Race	Female	Male	Total	Percentage
Asian	0	11	11	0.66026%
Black	44	786	830	49.81992%
Caucasian	90	514	604	36.25450%
Hispanic	21	185	206	12.36494%
Native American	1	6	7	0.42016%
Other	1	7	8	0.48019%
Total	157	1509	1666	100%
Percentage	9.42376%	90.57623%	100%	

Table 2.1.6 Number and percentage of exonerees by race and by gender in states without a compensation statute

Race	Female	Male	Total	Percentage
Asian	0	2	2	0.96153%
Black	2	81	83	39.90384%
Caucasian	15	89	104	50%
Hispanic	0	14	14	6.73076%
Native American	0	4	4	1.92307%
Other	0	1	1	0.48076%
Total	17	191	208	100%
Percentage	8.17307%	91.82692%	100%	

As is shown in Table 2.1.5 and Table 2.1.6, the Black also accounts for the majority of exonerees in states with a compensation statue, while in states without such kind of statue Caucasian has the largest number. Percentages of exonerees by race and gender in states with a statue are slightly different with them in states without a statue, but they’re very close.

Table 2.1.7 Number and percentage of incarcerated exonerees by race and by gender in states with a compensation statute

Race	Female	Male	Total	Percentage
Asian	0	9	9	0.59920%
Black	34	745	776	51.59786%
Caucasian	69	465	534	35.55259%
Hispanic	8	162	170	11.31824%
Native American	1	5	6	0.39946%
Other	1	6	7	0.46604%
Total	113	1389	1502	100%
Percentage	7.52330%	92.47669%	100%	

Table 2.1.8 The average number of years lost by race and gender

Race	Female	Male	Average Year Lost of Total
Asian	0	7.2	7.2
Black	4.3	10.9	10.5
Caucasian	4.7	8.3	7.8
Hispanic	3.5	7.1	6.8
Native American	17.6	9.6	10.2
Other	2.6	7.1	6.6
Average Year Lost	4.5	9.5	9.0

According to Table 1.8, the average years lost of Black people are 10.4 years, which is the biggest. It can be seen that Native American also served a lot time in prison, though there are not many Native American incarcerated exonerees. Years lost of these two races are remarkably large, so that may have some potential reasons waiting to be identified.

2.2 Percentages of Each Race and Gender

Table 2.2.1 Percentages of Males and Females

	Male	Female
%Seeking State	46.7%	30.8%
%Seeking Civil Rights	45.8%	32.3%
%Seeking at least one	67.7%	50.8%
% Receiving State	34.0%	21.5%
%Receiving Civil Rights	26.0%	16.9%
%Receiving at least one	47.9%	31.5%

Table 2.2.2 Percentages of Races

	Asian	Black	Caucasian	Hispanic	Native	Other
%Seeking State	36.4%	53.2%	30.8%	45.1%	33.3%	25.0%
%Seeking Civil Rights	36.4%	48.4%	32.3%	46.7%	33.3%	12.5%
%Seeking at least one	63.6%	71.0%	50.8%	64.7%	83.3%	25.0%
% Receiving State	27.3%	41.2%	21.5%	37.0%	16.7%	12.5%
%Receiving Civil Rights	36.4%	27.1%	16.9%	28.8%	0.0%	12.5%
%Receiving at least one	63.6%	53.6%	31.5%	48.4%	33.3%	25.0%

It can be noticed in Table 2.2.1 and Table 2.2.2 that the likelihood of seeking a state claim of males is bigger than this of females, which has the same situation when it comes to seeking civil rights. Likewise, male is more likely to receive a state claim or civil rights.

2.3 Percentages of each tag

Table 2.3.1 Percentages of exonerated through and not through CIUs

	CIU=1	CIU=0
%Seeking State	51.9%	44.8%
%Seeking Civil Rights	37.2%	45.5%
%Seeking at least one	57.1%	67.3%
% Receiving State	41.7%	32.2%
%Receiving Civil Rights	18.6%	25.9%
%Receiving at least one	46.8%	46.7%

Table 2.3.2 Percentages of exonerated with or without guilty pleas

	Guilty Plea=1	Guilty Plea=0
%Seeking State	32.8%	48.0%
%Seeking Civil Rights	31.7%	47.4%
%Seeking at least one	51.0%	69.5%
% Receiving State	23.8%	34.9%
%Receiving Civil Rights	23.4%	25.6%
%Receiving at least one	39.7%	48.7%

Table 2.3.3 Percentages of people assisted or not assisted by IO

	IO=1	IO=0
%Seeking State	61.2%	40.9%
%Seeking Civil Rights	62.8%	39.5%
%Seeking at least one	85.7%	60.8%
% Receiving State	51.8%	27.6%
%Receiving Civil Rights	33.3%	22.9%

%Receiving at least one	65.9%	41.1%

Table 2.3.4 Percentages of DNA or not DNA exonerees

	DNA=1	DNA=0
%Seeking State	72.1%	38.7%
%Seeking Civil Rights	56.1%	41.9%
%Seeking at least one	88.7%	60.8%
% Receiving State	66.9%	24.5%
%Receiving Civil Rights	39.5%	21.7%
%Receiving at least one	80.8%	38.1%

Table 2.3.5 Percentages of death penalty or not death penalty cases

	Death Penalty=1	Death Penalty=0
%Seeking State	42.2%	45.7%
%Seeking Civil Rights	56.0%	43.9%
%Seeking at least one	68.1%	66.2%
% Receiving State	29.3%	33.3%
%Receiving Civil Rights	29.3%	25.0%
%Receiving at least one	46.6%	46.7%

From all the tables above, we can see percentages of exonerees with some tags seeking or receiving each type of compensation. Being exonerated through a CIU makes it more likely to seek and to win state compensation or a civil rights award, which is as same as being a DNA exonerees. Evidently, people without guity plea are also more likely to file or prevail both types of compensation. What's more interesting, the likelihood of exonerees assisted by IO receiving a claim is less than not assisted by IO, which can indicate that help from IO may not be very useful.

Table 2.3.6 Number of each crime

	Murder	Sexual	Drugs	Child	Robbery	Other	Total
Number	776	286	125	212	94	217	1710

Table 2.3.7 Percentages of each crime

	Murder	Sexual Assault	Drugs	Child Sexual Abuse	Robbery	Other
%Seeking State	59.8%	59.8%	16.8%	42.0%	46.8%	41.9%
%Seeking Civil Rights	39.5%	39.5%	16.8%	33.5%	27.7%	34.6%
%Seeking at least one	74.8%	74.8%	28.8%	61.8%	56.4%	60.4%
% Receiving State	51.7%	51.7%	9.6%	28.3%	31.9%	24.0%
%Receiving Civil Rights	23.1%	23.1%	9.6%	19.3%	12.8%	21.7%
%Receiving at least one	63.3%	63.3%	19.2%	42.5%	36.2%	37.8%

Table 2.3.6 shows percentages of each crime. We can conclude that the nature of the crime allegedly committed makes it more likely to seek and to win state compensation or a civil rights award, because exonerees with conviction of murder or sexual assault are more likely to seek or win compensation.

2.4 Numbers and Percentages of each tag in M-R

Table 2.4.1 Number and Percentages of FC

	Seeking Civil Rights	Receiving Civil rights
All exonerees (236)	156 (66.10%)	100 (42.37%)
Incarcerated exonerees (228)	155 (67.98%)	100 (43.85%)

Table 2.4.2 Number and Percentages of MWID

	Seeking Civil Rights	Receiving Civil rights
All exonerees (586)	263 (44.88%)	141 (24.06%)
Incarcerated exonerees(573)	262 (45.72%)	140 (24.43%)

Table 2.4.3 Number and Percentages of FMFE

	Seeking Civil Rights	Receiving Civil rights
All exonerees (471)	195 (41.40%)	117 (24.84%)
Incarcerated exonerees (426)	194 (45.54%)	117 (27.46%)

Table 2.4.4 Number and Percentages of PFA

	Seeking Civil Rights	Receiving Civil rights
All exonerees (1049)	553 (52.72%)	332 (31.65%)
Incarcerated exonerees (987)	527 (53.39%)	309 (31.31%)

Table 2.4.5 Number and Percentages of OM

	Seeking Civil Rights	Receiving Civil rights
All exonerees (878)	541 (61.62%)	311 (35.42%)
Incarcerated exonerees (827)	516 (62.39%)	298 (36.03%)

Table 2.4.6 Number and Percentages of ILD

	Seeking Civil Rights	Receiving Civil rights
All exonerees (459)	184 (40.09%)	77 (16.78%)
Incarcerated exonerees (436)	180 (41.28%)	75 (17.20%)

Part 3. More Complex Inquiries

3.1 The Likelihood of Seeking a Claim

In this section, we analyze the relationships between some factors and response variables. The likelihood of filing a state claim can be seen as a response variable, and its relationships with 15 explanatory variables are tested. Contingency tables are used in most analysis, while we use Spearman test when analyzing the relationship between number of years lost with the response variable, since the number of years lost is continuous. Before doing this, data cleaning is done, which deletes cases that states don't have a statue and premature cases. Moreover, we test the relationships between the likelihood of filing a civil right claim with 18 variables. The methods are similar. This time the premature cases are also removed. Results are shown in the following table.

Table 3.1.1 p-values (1)

p-value of…	Filing a state claim	Filing civil rights
Race	<0.0001	0.0092
Date of exoneration	<0.0001	0.12
Number of years lost	<0.0001	<0.0001
Gender	<0.0001	0.0112
Guilty plea	<0.0001	<0.0001
IO	<0.0001	<0.0001
Crime	<0.0001	<0.0001
DNA	<0.0001	<0.0001
Death penalty	0.8039	0.0033
Geographic area	Not being tested	<0.0001
Red/Blue State	Not being tested	<0.0001
CIU	Not being tested	<0.0001
M-R		
FC	0.0027	<0.0001
MWID	<0.0001	0.7322
FMFE	0.0715	0.3747
PFA	0.5375	<0.0001
OM	0.8248	<0.0001
ILD	0.4808	0.1095

According to the table above, for the variable 'the likelihood of filing a state claim,' most explanatory variables are statistically significant since the p-values are quite small, except death penalty, FMFE, PFA, OM and ILD. In other words, many variables have a significant relationship with the likelihood of filing a state claim, such as race, gender, IO and so on. Obviously, 4 tags among M-R are insignificant statistically, thus these tags are possibly useless.

The results of filing civil rights are quite different. Race, date of exoneration, MWID, FMFE and ILD are insignificant since the p-values are too large, while other variables are statistically significant. Generally speaking, M-R tags have little effects on the likelihood of filing claims.

Table 3.1.2 Correlation Coefficients (1)

StateClaimMade, N = 1438	Years Lost
Spearman Correlation Coefficients	0.42308
CivilRightsMade, N = 1677	Years Lost
Spearman Correlation Coefficients	0.33788

Some results of Spearman test are shown in the above table. It can be seen that years lost has a positive correlation with both two response variables, because the Spearman correlation coefficients are more than zero.

3.2 The Likelihood of Prevailing a Claim

This section is similar with the part 3.1, especially the methods we use to analyze the data, but the two response variables that we are interested in are the likelihood of prevailing a state claim and the likelihood of prevailing a civil rights claim. The relationships with the first response variable with 15 explanatory variables are tested. Before doing this, data cleaning is done, which deletes premature cases and pending cases, as well as people who don't make a state claim. Exonerees in states where don't have a claim statue are also deleted.

In addition, we test the relationships between the likelihood of filing a civil rights claim with 12 variables. Likewise, this time the premature and pending cases are removed.

In these two steps, contingency table is used when the explanatory variable is categorical, but Spearman test is used when the explanatory variable is continuous. Results are shown in the following tables.

Table 3.2.1 p-values (2)

p-value of…	Prevailing a state claim	Prevailing civil rights
Race	0.0025	0.4871
Date of exoneration	0.0005	0.0007
Number of years lost	<0.0001	0.0003

Gender	0.592	0.1342
Guilty plea	0.59	0.0009
IO	<0.0001	0.0035
Crime	<0.0001	0.5603
DNA	<0.0001	0.0003
Death penalty	0.4935	0.241
Geographic area	Not being tested	<0.0001
Red/Blue State	Not being tested	<0.0001
CIU	Not being tested	0.1782
M-R		
FC	0.522	0.0003
MWID	0.0029	0.1047
FMFE	0.1033	0.7202
PFA	0.2587	0.0033
OM	0.5281	0.0025
ILD	0.0091	0.0001

According to the table 3.2.1, for the variable 'the likelihood of prevailing a state claim', 7 explanatory variables are statistically significant since the p-values are smaller than 0.05. That's to say, these variables, like race and date of exoneration, have a significant relationship with the likelihood of filing a state claim. In particular, the p-values of IO, crime, and DNA are less than 0.0001. Thus, these three variables are highly correlated with the likelihood of prevailing a state claim. Obviously, 4 tags among M-R are insignificant statistically, which is as same as the last part. What's totally different that is the gender and guilty plea are insignificant, but their p-values are smaller than 0.0001 when we test the relationship with the likelihood of seeking a state compensation.

The results of winning a civil rights claim point out that race, gender, CIU, crime, MWID and FMFE are not statistically significant since the p-values are too large, while other variables are statistically significant.

It should be noticed that race is not correlated with the likelihood of seeking or prevailing a civil rights claim. However, people in different states really have a huge difference about the probability of filing or prevailing a civil rights claim, as geographic area and red/blue states are significant in two analyses.

Table 3.2.2 Correlation Coefficients(2)

StateClaimAward, N = 696	Years Lost
Spearman Correlation Coefficients	0.22457
CivilRightsAward, N = 617	Years Lost
Spearman Correlation Coefficients	0.14388

Some results of Spearman test are shown in the above table. It can be seen that years lost has a positive correlation with both two response variables, because the Spearman correlation coefficients are more than zero. Now we can safely draw a conclusion that the probability of seeking or winning any claim will be bigger when the years lost increases. The possible reason is that the exoneree should be likely to be compensated if he or she served more time in prison.

3.3 The amount received in civil rights claim per year lost

This section is also similar with the last two parts. Now we are interested in the amount of received in prevailing civil rights claims per lost year. The relationships with response variable with 18 explanatory variables are tested. Before doing this, it's necessary to do some data cleaning. Thus we delete a lot of cases, including people don't seek a civil rights claim and the amount is not sure (pending). In order to make it convenient to calculate, cases in which the values of years lost are zero are removed from the data set.

ANOVA table is used when the explanatory variable is categorical, but Spearman test is used when the explanatory variable is continuous. In this section contingency table is not able to be used because the response variable is not categorical. Results are shown in the following table.

Table 3.3.1 p-values (3)

p-value of…	Amount received in civil rights claims per year lost
Race	0.091
Date of exoneration	<0.0001
Number of years lost	0.0197
Gender	0.3346
Guilty plea	0.048
IO	0.3514
Crime	0.1394
DNA	0.0025
Death penalty	0.4962
Geographic area	<0.0001
Red/Blue State	0.0052
CIU	0.8473
M-R	
FC	0.0003
MWID	0.4324
FMFE	0.6109
PFA	0.149
OM	0.0494
ILD	0.3523

According to the table 3.3.1, for the variable ’amount received in prevailing civil rights claims per year lost’, 8 explanatory variables are statistically significant since the p-values are smaller than 0.05, such as numbers of years lost, geographic area and red/blue state. Hence, people in different geographic areas may receive different amount of money in prevailing civil rights claims per year lost. Also, the political position of a state can affect the amount per year lost.

Obviously, 4 tags among M-R are insignificant statistically, which is as same as the last two part. So these M-R tags can be ignored in the further research.

Race and gender are both insignificant, which means that the amount of money exoneree can receive in prevailing civil rights claims per year lost doesn’t depend on race or gender.

Figure 3.3.1 Distribution for Race

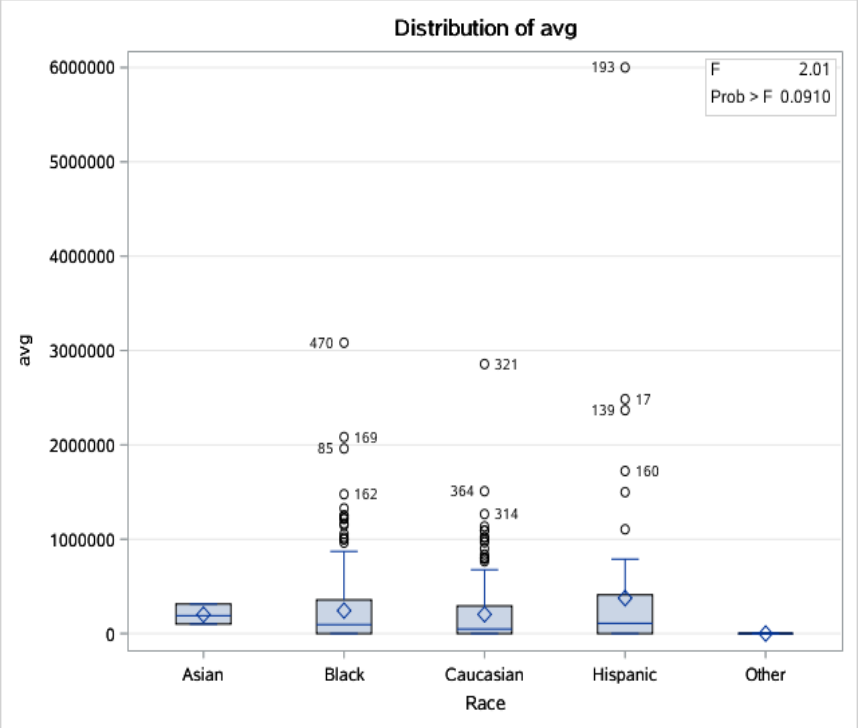


Figure 3.3.2 Distribution for Gender

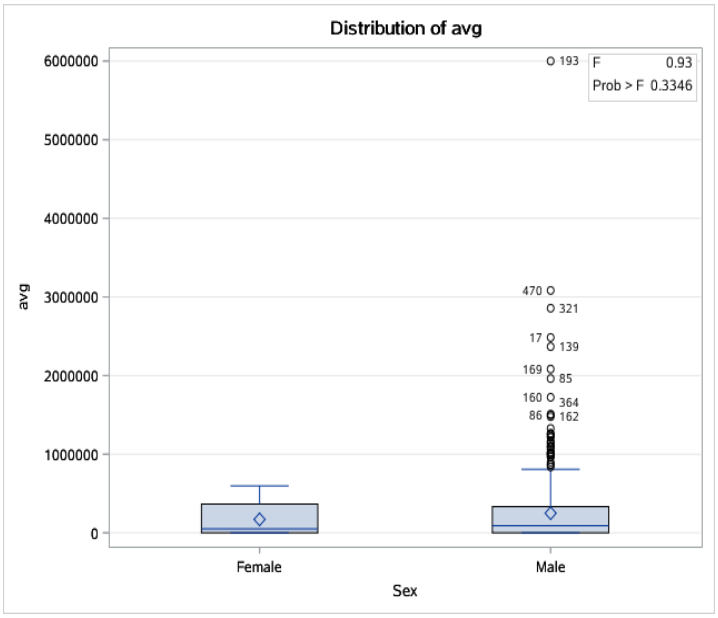


Figure 3.3.3 Distribution for Guilty Plea

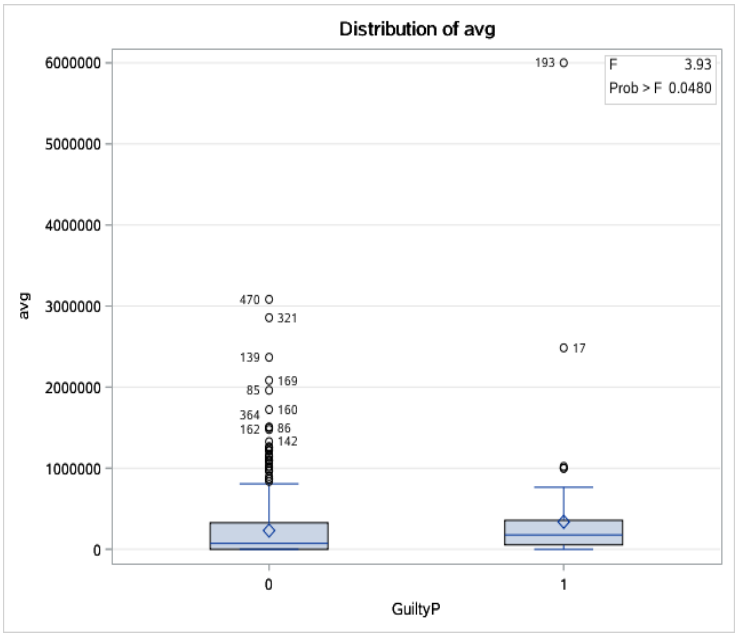


Figure 3.3.4 Distribution for IO

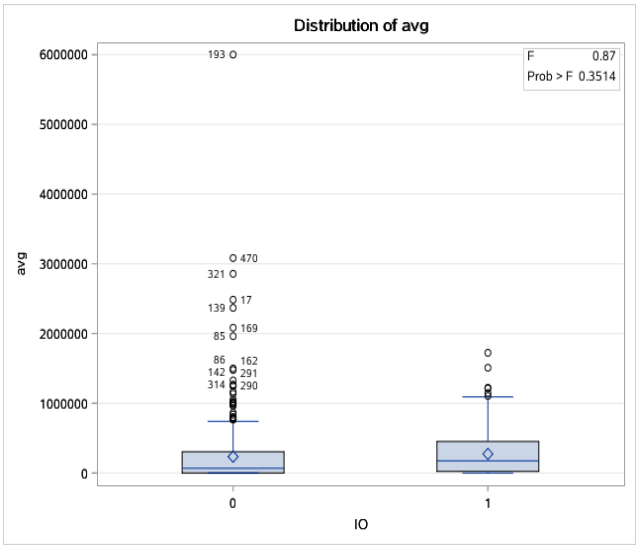


Figure 3.3.5 Distribution for Crime

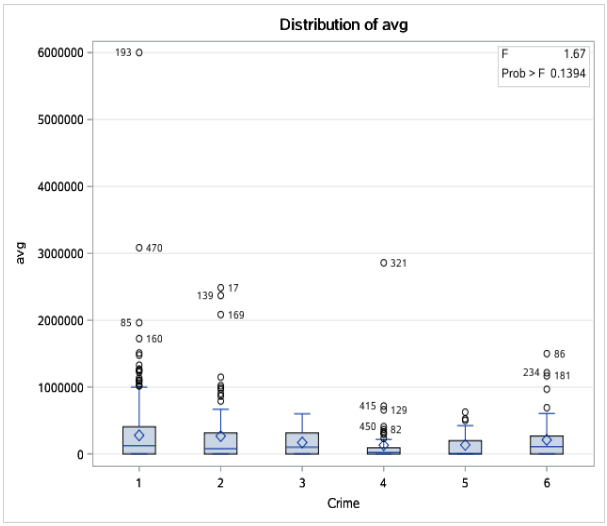


Figure 3.3.6 Distribution for DNA

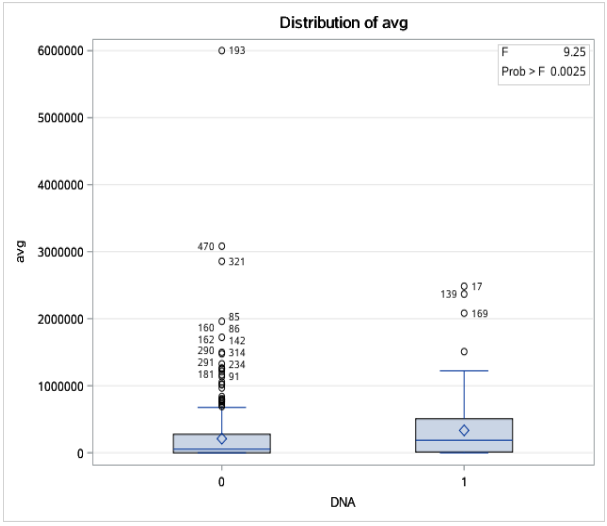


Figure 3.3.7 Distribution for Death

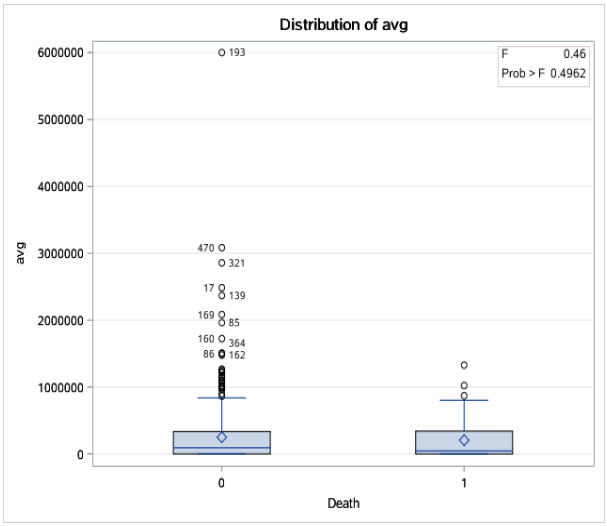


Figure 3.3.8 Distribution for FC

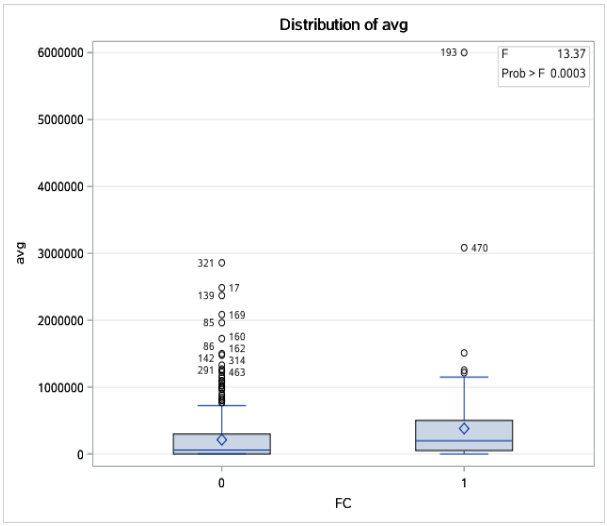


Figure 3.3.9 Distribution for MWID

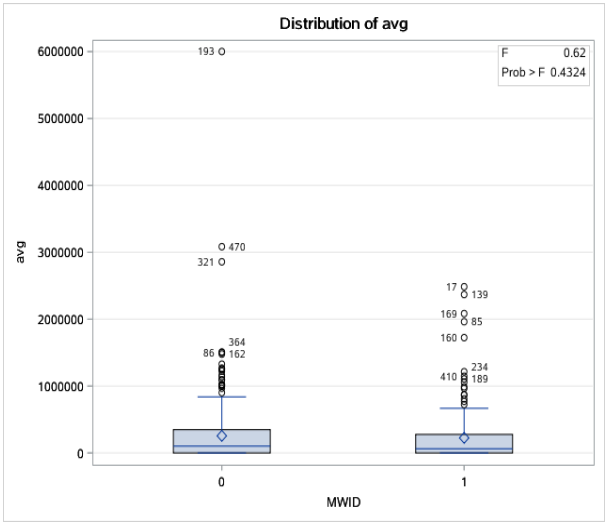


Figure 3.3.10 Distribution for FMFE

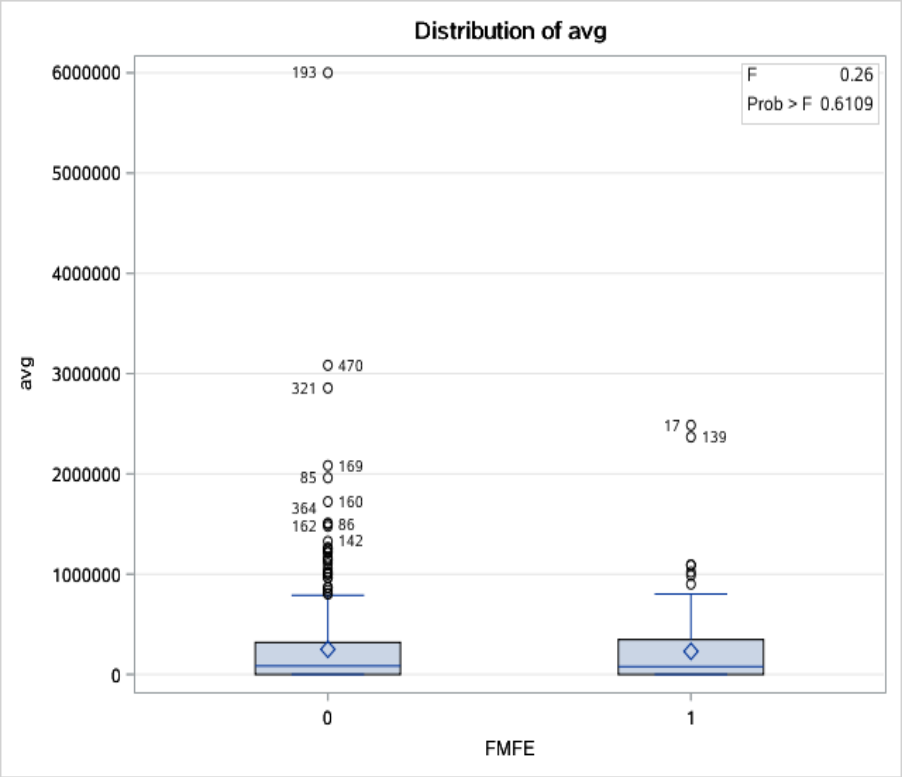


Figure 3.3.11 Distribution for Geographic Area

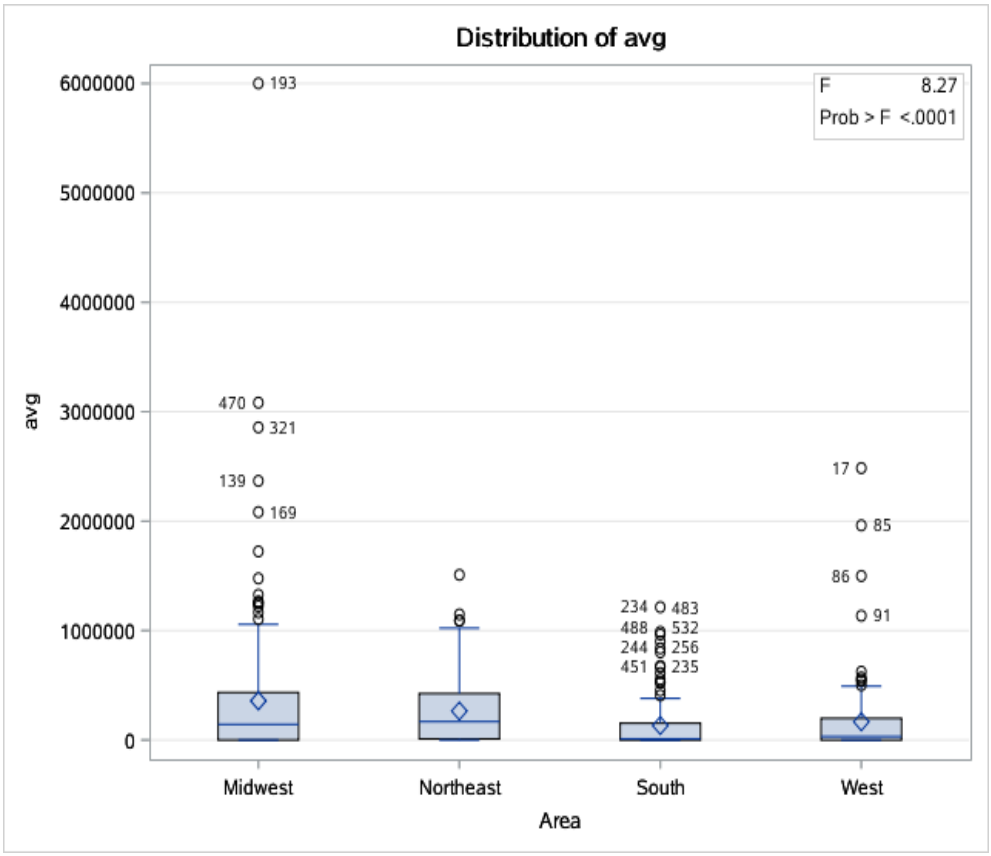
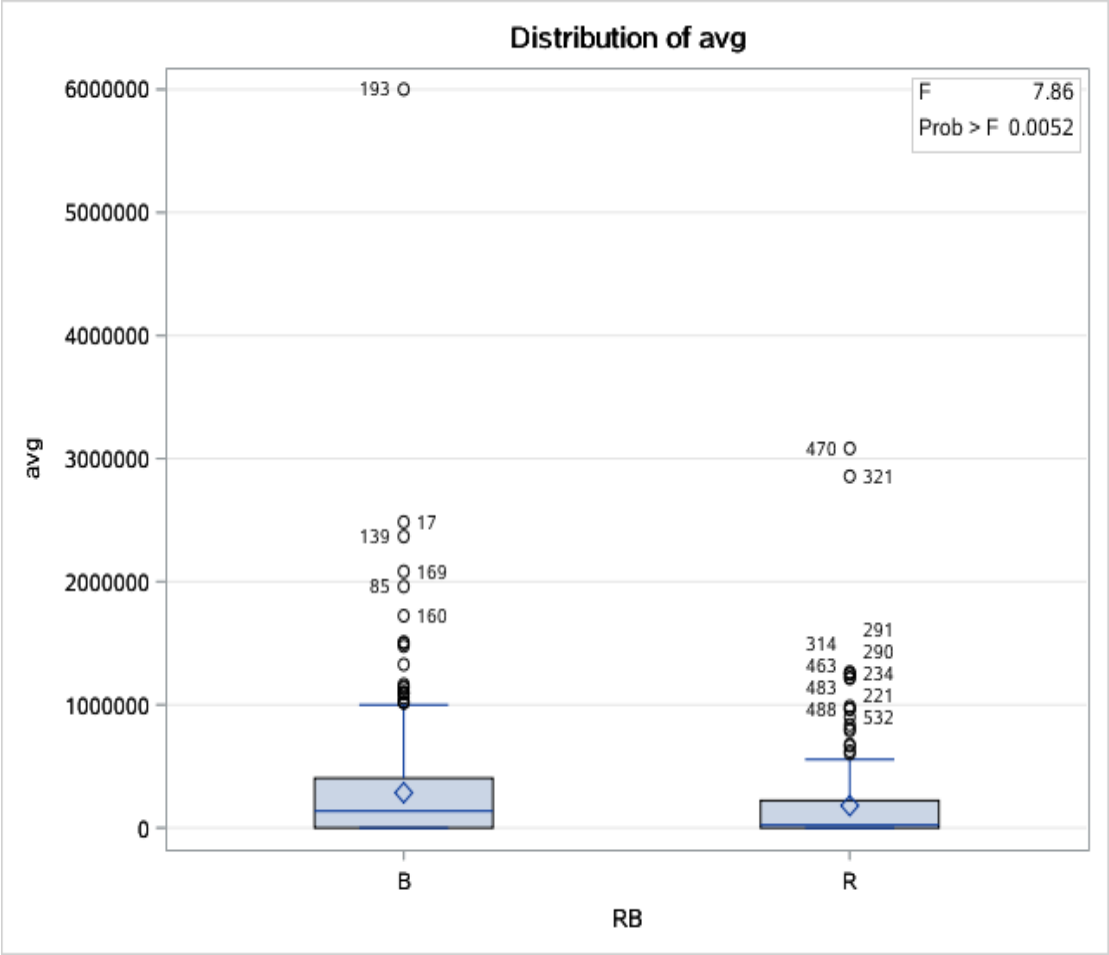


Figure 3.3.12 Distribution for Red/Blue State



The 12 figures above are created by SAS programming, which show distributions of average amount for some variables. According to Figure 3.3.11, average civil rights awards are larger in Midwest and Northeast than other areas. In addition, Figure 3.3.12 shows that average civil rights awards in Clinton states are larger than Trump states, which is possibly caused by the fact that blue states are wealthier and more developed. So they have enough budget to provide exonerees with a large amount of money.