The Mortgage Lending Refusal Rates for Caucasians and Minorities

Cathy Evans, Brandon Long, Barak Schmookler, Jillian Varner

North Carolina State University

Instructor: Dr. Sujit Ghosh

ST380

Section 001

December 4th, 2009
Table of Contents

Data........................................................................................................................................3
Executive Summary..................................................................................................................4
Description of Data..................................................................................................................4
Statistical Analysis..................................................................................................................5
Major Findings..........................................................................................................................12
Discussion................................................................................................................................12
Appendix.................................................................................................................................14
Data

Refusals in Mortgage Lending

**Number of cases:**
20

**Variable Names:**
1. Name of bank
2. MIN = refusal rate for minority applicants
3. WHITE = refusal rate for white applicants
4. HIMIN = refusal rate for high income minority applicants
5. HIWHITE = refusal rate for high income white applicants

<table>
<thead>
<tr>
<th>Bank</th>
<th>MIN</th>
<th>WHITE</th>
<th>HIMIN</th>
<th>HIWHITE</th>
</tr>
</thead>
<tbody>
<tr>
<td>HARRIS TRUST</td>
<td>20.9</td>
<td>3.7</td>
<td>21.4</td>
<td>2.2</td>
</tr>
<tr>
<td>NCNB TEXAS</td>
<td>23.23</td>
<td>5.5</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>CRESTAR</td>
<td>23.1</td>
<td>6.7</td>
<td>11.3</td>
<td>3.6</td>
</tr>
<tr>
<td>MERCANTILE</td>
<td>30.4</td>
<td>9</td>
<td>17.3</td>
<td>5.5</td>
</tr>
<tr>
<td>FIRST NB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEXAS COMMERCE</td>
<td>62.2</td>
<td>20.6</td>
<td>33.3</td>
<td>10.3</td>
</tr>
<tr>
<td>COMERICA</td>
<td>39.5</td>
<td>13.4</td>
<td>33.6</td>
<td>9.4</td>
</tr>
<tr>
<td>FIRST OF AMERICA</td>
<td>38.4</td>
<td>13.2</td>
<td>29.5</td>
<td>7.3</td>
</tr>
<tr>
<td>BOATMAN'S NAT'L</td>
<td>26.2</td>
<td>9.3</td>
<td>21.7</td>
<td>7.4</td>
</tr>
<tr>
<td>1ST COMM'L</td>
<td>55.9</td>
<td>21</td>
<td>39.1</td>
<td>15.8</td>
</tr>
<tr>
<td>PROVIDENT NAT'L</td>
<td>49.7</td>
<td>20.1</td>
<td>36.6</td>
<td>15.3</td>
</tr>
<tr>
<td>WORTHEN</td>
<td>44.6</td>
<td>19.1</td>
<td>28.6</td>
<td>10.1</td>
</tr>
<tr>
<td>HIBERNIA NAT'L</td>
<td>36.4</td>
<td>16</td>
<td>32.9</td>
<td>9.2</td>
</tr>
<tr>
<td>SOVRON</td>
<td>32</td>
<td>16</td>
<td>21</td>
<td>13</td>
</tr>
<tr>
<td>BELL FEDERAL</td>
<td>10.6</td>
<td>5.6</td>
<td>5.8</td>
<td>4.2</td>
</tr>
<tr>
<td>SEC PAC AZ</td>
<td>34.3</td>
<td>18.4</td>
<td>24.2</td>
<td>14.1</td>
</tr>
<tr>
<td>CORE STATES</td>
<td>42.3</td>
<td>23.3</td>
<td>38.3</td>
<td>15</td>
</tr>
<tr>
<td>CITIBANK AZ</td>
<td>26.5</td>
<td>15.6</td>
<td>27.3</td>
<td>16.1</td>
</tr>
<tr>
<td>MFERS HANOVER</td>
<td>51.5</td>
<td>32.4</td>
<td>41.3</td>
<td>25.1</td>
</tr>
<tr>
<td>CHEMICAL</td>
<td>47.2</td>
<td>29.7</td>
<td>41.1</td>
<td>26.8</td>
</tr>
</tbody>
</table>
Executive Summary

The Association of Community Organization for Reform Now (known as ACORN) collected data from different banks to determine if there was biased over racial difference on mortgage rejection and acceptance rates for fixed income groups. The data collected only presents a set number of twenty banks and their data in terms of percentage of mortgage loans refused from high income Caucasians, high income minorities, low income Caucasians, and low income minorities. The actual number of people taken into account from each individual bank is not listed. The data shows a wide range of percentage difference in mortgage refusals of racial groups for both high income groups and in low income groups.

The purpose of this project is to answer if banks would be more likely to reject minorities than Caucasians. To answer such questions data was analyzed based on distribution of data, the likelihood function, box plots, and summary statistics. Also an error of data collected was determined.

The results show a higher rejection rate for minorities at all income levels. Also, the results show a very positive linear relationship among banks and their refusal rates of Caucasians versus minorities. The same is true for the refusal rates of high income Caucasians and high income minorities.

Description of Data

ACORN stands for the Association of Community Organization for Reform Now. A meeting was held by a committee under ACORN to discuss data involving the refusal rates in mortgage lending found from 20 national banks in major cities. The data included rejection rates of Caucasian applicants, minority applicants, high-income Caucasian applicants, and high-income minority applicants as the major variables to compare with their refusal rate in mortgage lending.

The data was collected in the beginning of the 1990s and released in October of 1991. At this time the Community Reinvestment Act, which was instituted in 1977 to reduce redlining, had been in practice for almost fifteen years, expected to have by that point taken effect. Redlining is the name given to discriminatory crediting practices against lower income housing areas, including banking, access to healthcare, access to jobs, insurance and even increasing food prices in supermarkets. Redlining was typically a practice to keep lower income families and
minorities, who tend to statistically have lower household incomes, out of prestigious neighborhoods.

There was also a countrywide slump in the housing market between 1986 and 1991, the year in which our data was collected by ACORN. Because we were literally just exiting the slump, our data should be relatively independent from extreme economic conditions since the study occurred at the upturn of the slump.

**Statistical Analysis**

In statistically analyzing our data, we first ran summary statistics and found the following:

<table>
<thead>
<tr>
<th>Statistic</th>
<th>white</th>
<th>minority</th>
<th>high income white</th>
<th>high income minority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>10.60</td>
<td>3.70</td>
<td>5.80</td>
<td>2.20</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>26.43</td>
<td>9.225</td>
<td>21.30</td>
<td>7.375</td>
</tr>
<tr>
<td>Median</td>
<td>37.40</td>
<td>15.800</td>
<td>29.05</td>
<td>9.750</td>
</tr>
<tr>
<td>Mean</td>
<td>36.88</td>
<td>15.625</td>
<td>27.52</td>
<td>11.300</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>45.25</td>
<td>20.225</td>
<td>36.95</td>
<td>15.075</td>
</tr>
<tr>
<td>Max.</td>
<td>62.20</td>
<td>32.400</td>
<td>41.30</td>
<td>26.800</td>
</tr>
</tbody>
</table>

These statistics indicate that rejection rates were noticeably higher for minorities in comparison to Caucasians, and also for high income minorities and high income Caucasians.

A box plot, shown in figure 2 (All figures are located in the Appendix), confirms this information and in conjunction with the numerical standard deviations (13.05090 for minorities, 7.795807 for white, 10.92203 for high income minorities and 6.516376 for high income white, we can conclude that the standard deviation is generally higher for the data concerning minorities than the data concerning Caucasians when economic standing is taken into consideration, and that there is a greater variation in the percentage of mortgage loan rejections for minorities than for Caucasians. This may be from widely differing financial situations, or from widely differing implicit discrimination.

A further illustration of this can clearly be seen in the scatter plots shown in figure 3, which plot white and minority versus bank number, high income white and high income minority versus bank number, minority and high income minority versus bank number and white and high income white versus bank number respectively. As the graph moves from left to right, each vertical column shows a separate bank.
For the given data, we will conduct the following hypothesis tests:

1) \( H_0: \) The mortgage rejection rate for low to middle income minorities is the same as that for low to middle income Caucasians.
   \( H_a: \) The mortgage rejection rate for low to middle income minorities is higher than that for low to middle income Caucasians.

2) \( H_0: \) The mortgage rejection rate for high income minorities is the same as that for high income Caucasians.
   \( H_a: \) The mortgage rejection rate for high income minorities is higher than that for high income Caucasians.

3) \( H_0: \) The overall mortgage rejection rate for minorities is the same as that for Caucasians.
   \( H_a: \) The overall mortgage rejection rate for minorities is higher than that for Caucasians.

Figures 3, shown below, display density estimates of mortgage rejection rates for each of the four groups. These graphs show that the data can be modeled with normal distributions. (The bandwidth for the density estimates is only 20, which indicates that normal distributions are good estimates.) Since the spread is similar for the four distributions, we can create the following model:

\[
M_1, \ldots, M_{20} \sim \text{i.i.d. } N(\mu_M, \sigma) \\
W_1, \ldots, W_{20} \sim \text{i.i.d. } N(\mu_W, \sigma) \\
HM_1, \ldots, HM_{20} \sim \text{i.i.d. } N(\mu_{HM}, \sigma) \\
HW_1, \ldots, HW_{20} \sim \text{i.i.d. } N(\mu_{HW}, \sigma)
\]

with the \( M_i \)'s, \( W_i \)'s, \( HM_i \)'s, and \( HW_i \)'s equal to the rejection rates for minorities, Caucasians, high income minorities, and high income Caucasians respectively. We can also create the equivalent model:

\[
M_1, \ldots, M_{20} \sim \text{i.i.d. } N(\mu_M, \sigma) \\
W_1, \ldots, W_{20} \sim \text{i.i.d. } N(\mu_M + \delta_W, \sigma) \\
HM_1, \ldots, HM_{20} \sim \text{i.i.d. } N(\mu_{HM}, \sigma) \\
HW_1, \ldots, HW_{20} \sim \text{i.i.d. } N(\mu_{HM} + \delta_{HW}, \sigma)
\]

The table below describes the above parameters.
Let us rewrite the second model to gain insight. There are a total of 80 data points. Let \((Y_1, \ldots, Y_{40})\) be the rejection rates for low to middle income minorities and Caucasians. Let \((Z_1, \ldots, Z_{40})\) be the rejection rates for high income minorities and Caucasians. For each \(i\) and \(j\) from 1 to 40 define the following variables:

\[
X_{W,i} = \begin{cases} 
1 & \text{if } i\text{'th rate is white} \\
0 & \text{otherwise}
\end{cases}
\]

\[
X_{HW,j} = \begin{cases} 
1 & \text{if } j\text{'th rate is high income white} \\
0 & \text{otherwise}
\end{cases}
\]

We can now rewrite the model as follows:

\[
Y_i = \mu_M + \delta_W X_{W,i} + e_i
\]

\[
Z_j = \mu_{HM} + \delta_{HW} X_{HW,j} + E_j
\]

with \(e_1, \ldots, e_{40} \sim \text{i.i.d. } N(0, \sigma)\) and \(E_1, \ldots, E_{40} \sim \text{i.i.d. } N(0, \sigma)\)

We should note that even if the data could not be approximated with normal distributions, the equations for \(Y_i\) and \(Z_j\) would still be correct. Also, as long as the central limit theorem applies, the regression that we will perform using \(R\) will be independent of the original distributions.

With the model described, we can use \(R\) to fit the model to the data. Considering the \(Y_i\)'s first, we get the following output from the code appearing in the Appendix:
The model estimates $\mu_M$ as 36.881 with a SD of 2.404, and it estimates $\delta_W$ as -21.256 with a SD of 3.399. Significant R-squared and P values both indicate a significant relationship exists between race and mortgage loan rejection rates.

For the $Z_i$’s we get the following output from the code appearing in the Appendix:

Residuals:

<table>
<thead>
<tr>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>-26.281</td>
<td>-7.200</td>
<td>0.175</td>
<td>5.518</td>
<td>25.319</td>
</tr>
</tbody>
</table>

Coefficients:

| Estimate Std. Error t value Pr(>|t|) |
|-------------------------------------|
| (Intercept) 36.881  2.404  15.344 < 2e-16 |
| h2$\text{TYPEWHITE}$ -21.256  3.399  -6.253 2.56e-07 |

Residual standard error: 10.75 on 38 degrees of freedom
Multiple R-squared: 0.5072, Adjusted R-squared: 0.4942
F-statistic: 39.1 on 1 and 38 DF, p-value: 2.559e-07

The model estimates $\mu_M$ as 27.515 with a SD of 2.011, and it estimates $\delta_{HW}$ as -16.215 with a SD of 2.844. Significant R-squared and P values both indicate a significant relationship exists between race and mortgage loan rejection rates of high income households.
From the central limit theorem (since the original data was normally distributed), we can approximate the likelihood functions for $\mu_M$, $\delta_W$, $\mu_{HM}$, and $\delta_{HW}$ as normal distributions. Graphs of these likelihood functions can be seen in figure 4. We can say with 95% confidence that the variables are within two standard deviations of their mean estimates.

$\mu_M$ is around $36.881 \pm 4.808 = (32.073, 41.689)$

$\delta_W$ is around $-21.256 \pm 6.897 = (-28.054, -14.458)$

$\mu_{HM}$ is around $27.515 \pm 4.022 = (23.493, 31.537)$

$\delta_{HW}$ is around $-16.215 \pm 5.688 = (-21.903, -10.527)$

One can clearly see from figure 1 that there are plenty of variables that have some linear relation between the two of them. We can attempt to fit these to linear models, as in figure 6. Using a the R commands in the appendix, we get the following values for linear regressions versus the min variable:

### white:

|                | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------|----------|------------|---------|----------|
| (Intercept)    | 16.020   | 4.136      | 3.873   | 0.00111  |
| white          | 1.335    | 0.238      | 5.609   | 2.54e-05 |

Residual standard error: 8.089 on 18 degrees of freedom
Multiple R-squared: 0.6361, Adjusted R-squared: 0.6159

### HiWhite

|                | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------|----------|------------|---------|----------|
| (Intercept)    | 23.583   | 4.951      | 4.763   | 0.000155 |
| hiwhite        | 1.177    | 0.382      | 3.081   | 0.006438 |

Residual standard error: 10.85 on 18 degrees of freedom
Multiple R-squared: 0.3453, Adjusted R-squared: 0.3089
and versus himin

Coefficients:

        Estimate Std. Error   t value  Pr(>|t|)
(Intercept)  9.4908    3.2561    2.915  0.00924
white       1.1535    0.1874    6.155 8.21e-06

---

Residual standard error: 6.368 on 18 degrees of freedom
Multiple R-squared: 0.6779,  Adjusted R-squared: 0.66

hiwhite

Coefficients:

        Estimate Std. Error   t value  Pr(>|t|)
(Intercept) 14.7402    3.7807    3.899  0.00105
hiwhite     1.1305    0.2917    3.876  0.00111

---

Residual standard error: 8.284 on 18 degrees of freedom
Multiple R-squared: 0.4549,  Adjusted R-squared: 0.4247

But one can note that the R-squared value in all these cases is less than seventy percent. We can make a much better regression by taking our values as the variables min-white and himin-hiwhite, or the discrepancies for minorities versus the discrepancies for the high-income minorities. See figure 5. The linear regression for this was:

Coefficients:

        Estimate Std. Error   t value Pr(>|t|)
(Intercept)   9.5667    3.0864    3.100  0.006184
hidiscrimination  0.7209    0.1711    4.213  0.000522

---

Residual standard error: 6.047 on 18 degrees of freedom
Multiple R-squared: 0.4965,  Adjusted R-squared: 0.4686
So this does not explain much more data. A better analysis may be to fit min and all of the other variables, or himin and all of the other variables. These give:

min
Coefficients:

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| (Intercept) | 10.2501 | 3.8247 | 2.680 | 0.0164 |
| white | 1.7687 | 0.6222 | 2.843 | 0.0118 |
| hiwhite | -1.3425 | 0.5722 | -2.346 | 0.0322 |
| himin | 0.5148 | 0.2419 | 2.128 | 0.0492 |

---

Residual standard error: 6.128 on 16 degrees of freedom
Multiple R-squared: 0.8143, Adjusted R-squared: 0.7795

himin
Coefficients:

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| (Intercept) | 2.58240 | 4.15041 | 0.622 | 0.5426 |
| white | 0.63179 | 0.67830 | 0.931 | 0.3655 |
| min | 0.42854 | 0.20136 | 2.128 | 0.0492 |
| hiwhite | -0.06586 | 0.60502 | -0.109 | 0.9147 |

---

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.591 on 16 degrees of freedom
Multiple R-squared: 0.7793, Adjusted R-squared: 0.738
F-statistic: 18.84 on 3 and 16 DF, p-value: 1.682e-05

Notice that this fits more of the data, but the statistical significance of each variable's fit goes down. This seems to indicate that one will not get a very good fit for more than one individual variable. In addition, it is interesting to note that hiwhite is negatively correlated with min and
himin, which indicates that a bank more likely to loan to a high-income white person is less likely to loan to a minority. Moreover, it is good to note that if the slope is over one, then the correlation indicates that the bank is more likely to reject the minorities or high income minorities as their rejection rates of the other group increases.

**Major Findings**

We can conclude that there is strong evidence that the mortgage rejection rate for low to middle income minorities is higher than that for low to middle income Caucasians. Also, there is strong evidence that the mortgage rejection rate for high income minorities is higher than that for high income Caucasians. Since we overturned the null hypotheses for the first two tests, we can overturn the null hypothesis for the third test. We can conclude that there is strong evidence that the overall mortgage rejection rate for minorities is higher than that for Caucasians.

In looking at when the data was taken there are many possible economic reasons as to why the data yields its results. First, the data was collected at the end of a housing slump, which means banks may have been strict in whom they gave mortgage loans. Banks' knowledge of person’s income, credit report, history of payment plans, and the down payment on the house all contribute to whether a bank will accept or reject a mortgage loan. In the early 1990s Caucasians may have been more likely to have a higher salary, better credit reports, and more liquid assets than a minority, which is a possible reason as to why banks refused more loans to minorities, along with widely differing amounts of implicit racism.

With regard to the regression analysis, when comparing two variables, one predicts the other about fifty percent of time, but if more than two variables are fit, then the statistical significance decreases for each individual predictor.

While no conclusions can be made about whether banks make distinctions on a case-by-case basis, the data did seem to indicate that there is a distinction between one bank and the next, which can give us the tools to analyze discrimination by banks.

**Discussion**

The number of banks used as data in this project is perhaps the biggest source of error. With only 20 banks taken into account across the United States, the result of the data analysis is very limited. To make the hypothesis even more reliable, a greater amount of banks' mortgage loan refusal rates of Caucasians and minorities must be used in order to prove the reliability of
the hypothesis in the project. If more banks are used, then a random sampling process can be used to reduce the bias of the data.

Additionally, the mortgage loan rejection rate is a composite statistic, based on previous analysis done by ACORN. The previous analysis was not described by the DASL site, which leaves some room for error, as each data point may be hiding statistical anomalies such as a skewed distribution or unconventional analysis. On top of this, we run into the problems of possible systematic error, such as not taking into account local economic conditions or housing markets, the number of Caucasians and minorities applying for mortgage loans, and the range and number of banks surveyed in the study. Furthermore, the rejection rate is ill defined, that is, we have little knowledge of how the statistic bears out in real life. These both hinder our analysis.

Lastly, the banks used for the study are located in different places across America, thus the demographics of customers are different for each bank. For example, one bank might have ten percent minority customers and ninety percent Caucasian customers, while another could have seventy percent minority and thirty percent Caucasian. Another factor to consider is the economic standing of the citizens. Some cities classically have lower personal incomes. These areas would obviously have a higher percentage of mortgage loan rejections than a bank in an upper middle class area would have.
Appendix

mydata=read.table("banks.txt",header=T)
> head(mydata)
   MIN  WHITE  HIMIN  HIWITE
1  20.90  3.7  21.4  2.2
2  23.23  5.5  21.4  2.2
3  23.10  6.7  11.3  3.6
4  30.40  9.0  17.3  5.5
5  42.70 13.9  38.0  7.6
6  62.20 20.6  33.3 10.3

> min=mydata$MIN
> white=mydata$WHITE
> himin=mydata$HIMIN
> hiwhite=mydata$HIWHITE
> banks=data.frame(min,white,himin,hiwhite)
> summary(banks)

        min       white       himin       hiwhite
Min.  :10.600  Min.  : 3.700  Min.  : 5.800  Min.  : 2.200
1st Qu.:26.430  1st Qu. : 9.225  1st Qu.:21.300  1st Qu.: 7.375
Median :37.400  Median :15.800  Median :29.050  Median : 9.750
Mean  :36.880  Mean  :15.625  Mean  :27.520  Mean  :11.300
3rd Qu.:45.250  3rd Qu.:20.225  3rd Qu.:36.950  3rd Qu.:15.075
Max.  :62.200  Max.  :32.400  Max.  :41.300  Max.  :26.800
> plot(banks)

Figure 1: A plot of all variables
Mortgage Refusal Rates

Boxplot(banks)

Figure 2: A boxplot of minority, white, high income minority and high income white rejection rates.

#standard deviations of the data
> sd(min)
[1] 13.05090
> sd(white)
[1] 7.795807
> sd(himin)
[1] 10.92203
> sd(hiwhite)
[1] 6.516376
#scatterplot with letters

par(mfrow=c(2,2))
plot(rate[type=="min"],col=2,pch="m",ylim=c(0,70),xlim=c(0,20),main="Minority vs. White")
points(rate[type=="white"],col=3,pch="w",ylim=c(0,70),xlim=c(0,20))
plot(rate[type=="himin"],col=4,pch="m",ylim=c(0,70),xlim=c(0,20),main="High Minority vs. High White")
points(rate[type=="hiwhite"],col=5,pch="w",ylim=c(0,70),xlim=c(0,20))
plot(rate[type=="himin"],col=6,pch="h",ylim=c(0,70),xlim=c(0,20),main="Minority vs. High Minority")
points(rate[type=="min"],col=7,pch="l",ylim=c(0,70),xlim=c(0,20))
plot(rate[type=="hiwhite"],col=8,pch="h",ylim=c(0,70),xlim=c(0,20),main="White vs. High White")
points(rate[type=="white"],col=9,pch="l",ylim=c(0,70),xlim=c(0,20))

Figure 3: Relationships of rejection rates of individual banks
# Density Estimates

```R
> par (mfrow=c(2,2))
> plot (density ( min, bw=20)
+ , yaxt="n", ylab="", xlab="% Rejection", main="Minority")
> plot (density ( white, bw=20)
+ , yaxt="n", ylab="", xlab="% Rejection", main="White")
> plot (density ( himin, bw=20)
+ , yaxt="n", ylab="", xlab="% Rejection", main="HIMIN")
> plot (density ( hiwhite, bw=20)
+ , yaxt="n", ylab="", xlab="% Rejection", main="HIWHITE")
```

![Density Estimates](image)

**Figure 4: Density Estimates of Minority, White, High Income Minority and High Income White rejection rates**
Figure 5: Linear Regressions for various data relationships.
#Define vars first
par(mfrow=c(1,1))
Discrimination=min-white
hidiscrimination=himin-hiwite
par(mfrow=c(1,1))
plot(Discrimination~hidiscrimination)
abline(9.5667,.7209,col="red")

Figure 6: Linear Regression illustrating the discrimination in mortgage loan rejections.
# regressions
fit.mw=lm(min~white)
fit.mhw=lm(min~hiwhite)
fit.hmw=lm(himin~white)
fit.hmhw=lm(himin~hiwhite)
fit.red=lm(Discrimination~hidiscrimination)
fit.all=lm(min~white+hiwhite+himin)
fit.hiall=lm(himin~white+min+hiwhite)