# MLA

## *Primary Findings*

*Mark Russeff, Renato Albolea, Shuai Ma*

# 1. Abstract

TBD

# 2. Background

The following abstract appeared in Alicia H. Munnell, Geoffrey M.B. Tootell, Lynn E. Browne, and James McEneaney (1996), “Mortgage Lending in Boston: Interpreting HMDA Data,” American Economic Review 86, 25-53.

> *The Home Mortgage Disclosure Act was enacted to monitor minority and low-income access to the mortgage market. The data collected for this purpose show that minorities are more than twice as likely to be denied a mortgage as whites. Yet variables correlated with both race and creditworthiness were omitted from these data, making any conclusion about race’s role in mortgage lending impossible. The Federal Reserve Board of Boston collected* additional variables important to the mortgage lending decision….

As discussed in Munnell et al (1996), the HMDA data indicate whether an applicant’s mortgage application was approved and provide several demographic characteristics. In 1990, following the request of the Federal Reserve Board of Boston, lending institutions in the Boston area provided additional information relevant to mortgage lending decisions. In light of the relatively small number of mortgage loan applications made by minorities, these extra variables were collected for all applications by blacks and Hispanics and for a random sample of those by whites.

> *All applicants are non-Hispanic white, non Hispanic black, or Hispanic. In 1990 about 94% of Boston residents were white, Black, or Hispanic. (http://www.bostonplans.org/getattachment/83972a7a-c454-4aac-b3eb-02e1fddd71e3/)*

# 3. Research Question

- **Controlling for relevant characteristics, is race/ethnicity associated with the outcome of a mortgage loan application?**

# 4. Explortory Data Analysis

# 4.1 Data Wrangling

**(A) First View of Data** (The code and result are hid here since this is not key point of this paper, and makes this paper clean. The final data summary will be showed.)

1. First Thought:

- The following column names should be updated:
  - GDLIN - Credit_History (credit history meets the guideline)
  - OBRAT - Other_Obligations (other obligations as a percent of total income)
  - LOANPRC - Loan_Percentage (loan amount/purchase price)
- The following variables should be changed to factors:
  - MARRIED
  - GDLIN
  - BLACK
  - HISPAN
  - MALE
  - APPROVE

**(B) Second View of Data** (Code and results are hid.)

2. Second Thought (criteria for sub-setting data ):

- Remove the three “.” from “Married”
- Remove the two “666” from “Credit_History”
- Remove the fifteen “.” from “Male”
- Remove the value which higher than 1 from “Loan_Percentage”

**(C) Final Data Summary**

```
summary(base)
```

```
##  Married  Credit_History Other_Obligations Black    Hispan   Male    
##  0: 662   0: 167         Min.   : 0.00     0:1745   0:1833   0: 361  
##  1:1275   1:1770         1st Qu.:28.00     1: 192   1: 104   1:1576  
##                          Median :33.00                               
##                          Mean   :32.37                               
##                          3rd Qu.:37.00                               
##                          Max.   :95.00                               
##  Approve  Loan_Percentage 
##  0: 235   Min.   :  2.105 
##  1:1702   1st Qu.: 69.697 
##           Median : 80.000 
##           Mean   : 76.075 
##           3rd Qu.: 89.820 
##           Max.   :100.000 
```

# 4.2 Race and Gender

(Code is hid from this paper, but available based on request)

**Findings from race and gender:**

1. The proportion of Black and Hispanic applications account for 10% and 5% of all applications separately.
2. The proportion of Male and Female applications account for 19% and 81% of all applications separately.
3. The smallest category in the application is “Hispanic Female” which was 1%.
4. The largest category in the application is “Non-Black/Hispanic Male” which was 70%.

Total Number by Race and Gender

| | Gender | Black | Hispan | Not Hispan/Black | Total |
|---|---|---|---|---|---|
| 1 | Female | 50 | 20 | 291 | 361 |
| 2 | Male | 142 | 84 | 1350 | 1576 |
| 3 | Total | 192 | 104 | 1641 | 1937 |

Percentage by Race and Gender

| | Gender | Black | Hispan | Not Hispan/Black | Total |
|---|---|---|---|---|---|
| 1 | Female | 2.6% | 1.0% | 15.0% | 18.6% |
| 2 | Male | 7.3% | 4.3% | 69.7% | 81.4% |
| 3 | Total | 9.9% | 5.4% | 84.7% | 100.0% |

# 4.3 Race and Approval

(Code is hid from this paper, but available based on request)

**Findings from race and decision:**

1. Among all applicants the approved application from Black counts for 6.6%. The approval rate for applications from Black was **66.7%**.
2. Among all applicants the approved application from Hispanic counts for 4.2%. The approval rate for applications from Hispanic was **77.9%**.

3. Among all applicants the approved application from Non-Black/Hispanic counts for 77.1%.The approval rate for applications from Non-Blackwas **91.0%**.
4. The average approval rate was **87.9%**.

Total Number by Race and Decision

| | Approve | Black | Hispan | Not Hispan/Black | Total |
|---|---|---|---|---|---|
| 1 | Approved | 128 | 81 | 1493 | 1702 |
| 2 | Denied | 64 | 23 | 148 | 235 |
| 3 | Total | 192 | 104 | 1641 | 1937 |

Percentage by Race and Decision

| | Approve | Black | Hispan | Not Hispan/Black | Total |
|---|---|---|---|---|---|
| 1 | Approved | 6.6% | 4.2% | 77.1% | 87.9% |
| 2 | Denied | 3.3% | 1.2% | 7.6% | 12.1% |
| 3 | Total | 9.9% | 5.4% | 84.7% | 100.0% |

# 4.4 Race and Credit History

(Code is hid from this paper, but available based on request)

**Findings from race and credit history:**

1. Among all applicants that the Black applicant with good credit history counts for **7.2%**. 72.3% Black applicants had good credit history.
2. Among all applicants that the Hispanic applicants with good credit history counts for **4.6%**. 86.5% Hispanic applicants had good credit history.
3. Among all applicants that the Non-Black/Hispanic applicants with good credit history counts for **79.6%**. 93.9% Non-Black/Hispanic applicants had good credit history.
4. The average approval rate was **91.4%**.

Total Number by Race and Credit History

| | Credit_History | Black | Hispan | Not Hispan/Black | Total |
|---|---|---|---|---|---|
| 1 | Good | 139 | 90 | 1541 | 1770 |
| 2 | Not Good | 53 | 14 | 100 | 167 |
| 3 | Total | 192 | 104 | 1641 | 1937 |

Percentage by Race and Credit History

| | Credit_History | Black | Hispan | Not Hispan/Black | Total |
|---|---|---|---|---|---|
| 1 | Good | 7.2% | 4.6% | 79.6% | 91.4% |
| 2 | Not Good | 2.7% | 0.7% | 5.2% | 8.6% |
| 3 | Total | 9.9% | 5.4% | 84.7% | 100.0% |

# 4.5 Race, Credit History, Decision

**Findings from race, credit history, and decision:**

1. Among all applicants who were with “Good” credit history, the approval rate was **“Not Hispan/Black” > “Black” > “Hispan”**.
2. Among all applicants who were with “Not Good” credit history, the approval rate was **“Not Hispan/Black” > “Hispan” > “Black”**.

**(A)Approveal/Denial Rate with Good Credit History by Races**

Approval Rate by Race and Credit History - Black with Good Credit

| | Category | Decision | Counts | Rate |
|---|---|---|---|---|
| 1 | Black - Good | Approved | 122 | 0.878 |
| 2 | Black - Good | Denied | 17 | 0.122 |

Approval Rate by Race and Credit History - Hispan with Good Credit

| | Category | Decision | Counts | Rate |
|---|---|---|---|---|
| 1 | Hispan - Good | Approved | 78 | 0.867 |
| 2 | Hispan - Good | Denied | 12 | 0.133 |

Approval Rate by Race and Credit History - Hispan with Good Credit

| | Category | Decision | Counts | Rate |
|---|---|---|---|---|
| 1 | Not Hispan/Black - Good | Approved | 1465 | 0.951 |
| 2 | Not Hispan/Black - Good | Denied | 76 | 0.049 |

**(B)Approveal/Denial Rate with Not Good Credit History by Races**

Approval Rate by Race and Credit History - Black with Not Good Credit

| | Category | Decision | Counts | Rate |
|---|---|---|---|---|
| 1 | Black - Not Good | Approved | 6 | 0.113 |
| 2 | Black - Not Good | Denied | 47 | 0.887 |

Approval Rate by Race and Credit History - Black with Not Good Credit

| | Category | Decision | Counts | Rate |
|---|---|---|---|---|
| 1 | Hispan - Not Good | Approved | 3 | 0.214 |
| 2 | Hispan - Not Good | Denied | 11 | 0.786 |

Approval Rate by Race and Credit History - Black with Not Good Credit

| | Category | Decision | Counts | Rate |
|---|---|---|---|---|
| 1 | Not Hispan/Black - Not Good | Approved | 28 | 0.280 |
| 2 | Not Hispan/Black - Not Good | Denied | 72 | 0.720 |

# 4.6 Distribution of loan rate

- loan rate = loan amount / purchase price

**Findings from distribution of loan rate by races and genders:**

1. Fig 1 and Fig 2 below indicate that, regardless the race and gender, most of applicants applies for the loan amount around 75% to 90% of the purchase price.

Fig 1. Distribution of Loan Rate by Races

Loan Rate = Loan Amount / Purchase Price

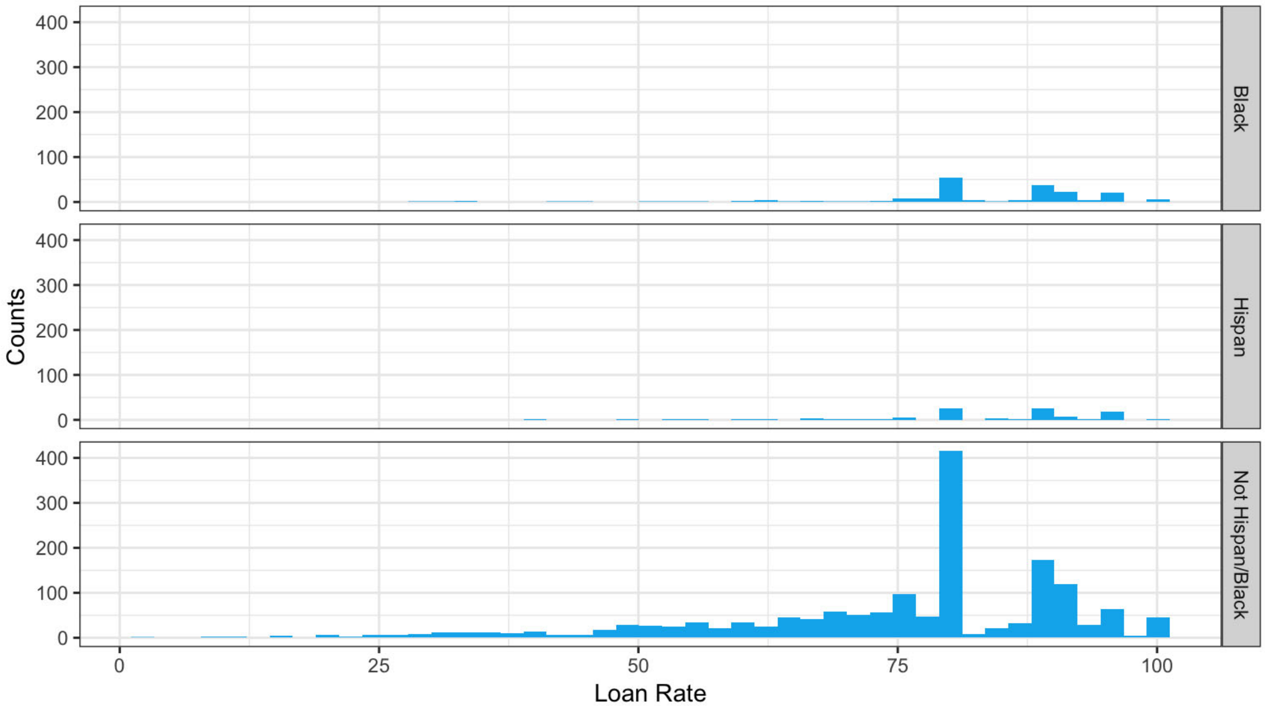
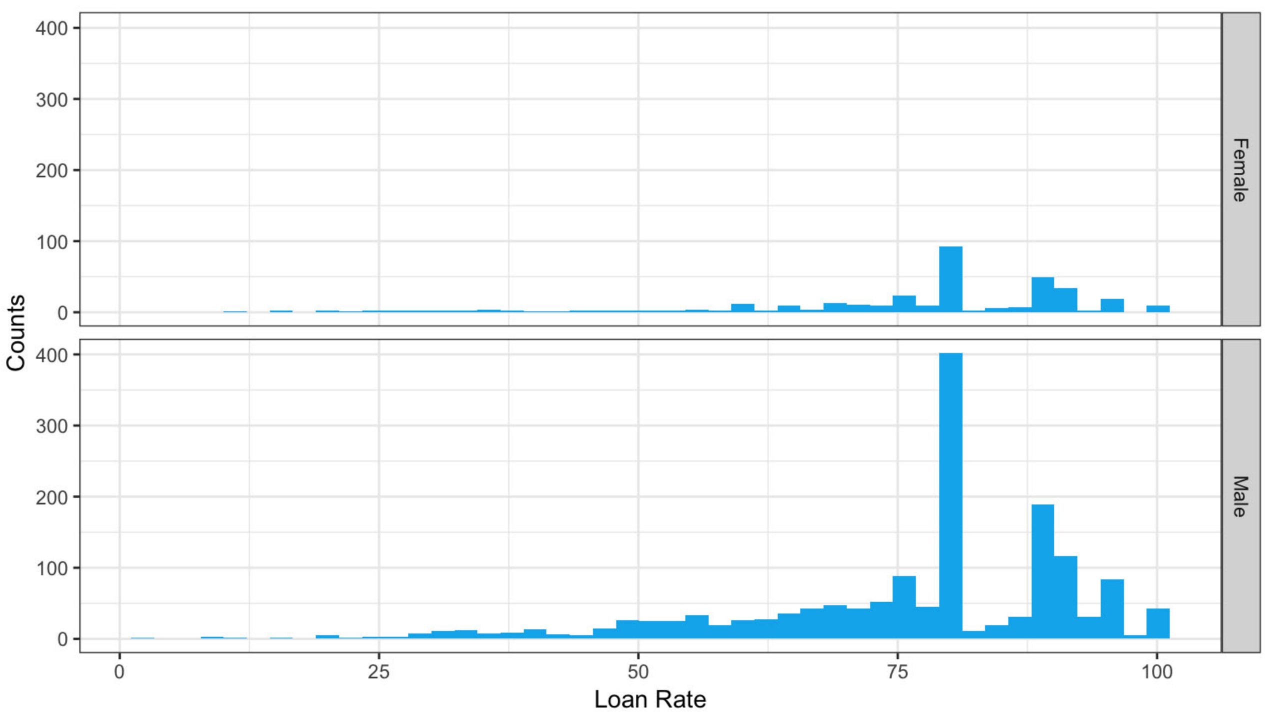



Fig 2. Distribution of Loan Rate by Gender

Loan Rate = Loan Amount / Purchase Price




# 5. Estimating Models

## 5.1 Model considering Race

**(A) Estimate Logit Model**

```
LogitModel = glm(Approve ~ Other_Obligations + Credit_History + Loan_Percentage + Marr
ied + Black + Hispan, data = base , family = "binomial")

summary(LogitModel)
```

```
##
## Call:
## glm(formula = Approve ~ Other_Obligations + Credit_History +
##     Loan_Percentage + Married + Black + Hispan, family = "binomial",
##     data = base)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.8910   0.2433   0.3081   0.3689   2.3525
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)        1.232732   0.684942   1.800 0.071898 .
## Other_Obligations -0.033892   0.010548  -3.213 0.001313 **
## Credit_History1    3.766456   0.220563  17.077  < 2e-16 ***
## Loan_Percentage   -0.015900   0.007009  -2.268 0.023302 *
## Married1           0.481513   0.184651   2.608 0.009116 **
## Black1            -0.869219   0.242665  -3.582 0.000341 ***
## Hispan1           -0.860024   0.323219  -2.661 0.007795 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1431.64  on 1936  degrees of freedom
## Residual deviance:  925.09  on 1930  degrees of freedom
## AIC: 939.09
##
## Number of Fisher Scoring iterations: 6
```

#### The signals of the model suggest that:

1. 
```
Increases in Other Obligation diminish the likelihood of approval of the loan, wh
ich makes sense.
```

2. 
```
Being Black or Hispanic diminish your likelihood of approval of the loan, which d
oes not make sense (your ethnicity does not interfere to your capacity to pay a l
oan), but was expected.
```

### (B) Generate Odds Ratios

```
cbind(exp(coef(LogitModel)), 1)
```

```
##                          [,1] [,2]
## (Intercept)        3.4305891    1
## Other_Obligations  0.9666758    1
## Credit_History1   43.2265786    1
## Loan_Percentage    0.9842253    1
## Married1           1.6185218    1
## Black1             0.4192788    1
## Hispan1            0.4231520    1
```

**The Odd ratios suggest that:**

1. White people had 0.02 times the odds of being approved for a loan as a black person;(p<0.01)
2. White people had 1.02 times the odds of being approved for a loan as a Hispanic person; (p<0.01)
3. One unit of increase in other obligations reduce in 3.33 percentage the odds of being approved for a loan in ;(p<0.01)

**(C) Define prototypical loan applicants (you will need more than 3)**

```
prototype_1 <- data.frame(Other_Obligations = mean(base$Other_Obligations), Married =
"1", Loan_Percentage = mean(base$Loan_Percentage), Credit_History = 1, Black = 0, Hisp
an = 0) %>%
  mutate(Black = as.factor(Black), Hispan = as.factor(Hispan), Credit_History = as.fac
tor(Credit_History))

prototype_2 <- data.frame(Other_Obligations = mean(base$Other_Obligations), Married =
"1", Loan_Percentage = mean(base$Loan_Percentage), Credit_History = 1, Black = 0, Hisp
an = 1) %>%
  mutate(Black = as.factor(Black), Hispan = as.factor(Hispan), Credit_History = as.fac
tor(Credit_History))

prototype_3 <- data.frame(Other_Obligations = mean(base$Other_Obligations), Married =
"1", Loan_Percentage = mean(base$Loan_Percentage), Credit_History = 1, Black = 1, Hisp
an = 0) %>%
  mutate(Black = as.factor(Black), Hispan = as.factor(Hispan), Credit_History = as.fac
tor(Credit_History))

prototype_4 <- data.frame(Other_Obligations = mean(base$Other_Obligations), Married =
"1", Loan_Percentage = mean(base$Loan_Percentage), Credit_History = 0, Black = 0, Hisp
an = 0) %>%
  mutate(Black = as.factor(Black), Hispan = as.factor(Hispan), Credit_History = as.fac
tor(Credit_History))

prototype_5 <- data.frame(Other_Obligations = mean(base$Other_Obligations), Married =
"1", Loan_Percentage = mean(base$Loan_Percentage), Credit_History = 0, Black = 0, Hisp
an = 1) %>%
  mutate(Black = as.factor(Black), Hispan = as.factor(Hispan), Credit_History = as.fac
tor(Credit_History))

prototype_6 <- data.frame(Other_Obligations = mean(base$Other_Obligations), Married =
"1", Loan_Percentage = mean(base$Loan_Percentage), Credit_History = 0, Black = 1, Hisp
an = 0) %>%
  mutate(Black = as.factor(Black), Hispan = as.factor(Hispan), Credit_History = as.fac
tor(Credit_History))

prototype_7 <- data.frame(Other_Obligations = mean(base$Other_Obligations), Married =
"0", Loan_Percentage = mean(base$Loan_Percentage), Credit_History = 0, Black = 0, Hisp
an = 0) %>%
  mutate(Black = as.factor(Black), Hispan = as.factor(Hispan), Credit_History = as.fac
tor(Credit_History))

prototype_8 <- data.frame(Other_Obligations = mean(base$Other_Obligations), Married =
"0", Loan_Percentage = mean(base$Loan_Percentage), Credit_History = 1, Black = 0, Hisp
an = 1) %>%
  mutate(Black = as.factor(Black), Hispan = as.factor(Hispan), Credit_History = as.fac
tor(Credit_History))

prototype_9 <- data.frame(Other_Obligations = mean(base$Other_Obligations), Married =
"0", Loan_Percentage = mean(base$Loan_Percentage), Credit_History = 1, Black = 1, Hisp
an = 0) %>%
  mutate(Black = as.factor(Black), Hispan = as.factor(Hispan), Credit_History = as.fac
tor(Credit_History))
```

```
prototype_10 <- data.frame(Other_Obligations = mean(base$Other_Obligations), Married =
"0", Loan_Percentage = mean(base$Loan_Percentage), Credit_History = 1, Black = 0, Hisp
an = 0) %>%
  mutate(Black = as.factor(Black), Hispan = as.factor(Hispan), Credit_History = as.fac
tor(Credit_History))

prototype_11 <- data.frame(Other_Obligations = mean(base$Other_Obligations), Married =
"0", Loan_Percentage = mean(base$Loan_Percentage), Credit_History = 0, Black = 0, Hisp
an = 1) %>%
  mutate(Black = as.factor(Black), Hispan = as.factor(Hispan), Credit_History = as.fac
tor(Credit_History))

prototype_12 <- data.frame(Other_Obligations = mean(base$Other_Obligations), Married =
"0", Loan_Percentage = mean(base$Loan_Percentage), Credit_History = 0, Black = 1, Hisp
an = 0) %>%
  mutate(Black = as.factor(Black), Hispan = as.factor(Hispan), Credit_History = as.fac
tor(Credit_History))
```

### (D) Predict probabilities for prototypical individuals

```
prototype_1$predictedprob <- round(predict(LogitModel, newdata = prototype_1, type =
"response")*100, digits = 1)

prototype_2$predictedprob <- round(predict(LogitModel, newdata = prototype_2, type =
"response")*100, digits = 1)

prototype_3$predictedprob <- round(predict(LogitModel, newdata = prototype_3, type =
"response")*100, digits = 1)

prototype_4$predictedprob <- round(predict(LogitModel, newdata = prototype_4, type =
"response")*100, digits = 1)

prototype_5$predictedprob <- round(predict(LogitModel, newdata = prototype_5, type =
"response")*100, digits = 1)

prototype_6$predictedprob <- round(predict(LogitModel, newdata = prototype_6, type =
"response")*100, digits = 1)

prototype_7$predictedprob <- round(predict(LogitModel, newdata = prototype_7, type =
"response")*100, digits = 1)

prototype_8$predictedprob <- round(predict(LogitModel, newdata = prototype_8, type =
"response")*100, digits = 1)

prototype_9$predictedprob <- round(predict(LogitModel, newdata = prototype_9, type =
"response")*100, digits = 1)

prototype_10$predictedprob <- round(predict(LogitModel, newdata = prototype_10, type =
"response")*100, digits = 1)

prototype_11$predictedprob <- round(predict(LogitModel, newdata = prototype_11, type =
"response")*100, digits = 1)

prototype_12$predictedprob <- round(predict(LogitModel, newdata = prototype_12, type =
"response")*100, digits = 1)
```

### (D-1) Predict probabilities for prototypical individuals

```
rbind.data.frame(prototype_1, prototype_2, prototype_3, prototype_4, prototype_5,
    prototype_6, prototype_7, prototype_8, prototype_9, prototype_10, prototype_11,
    prototype_12)
```

| Other_Obligations <dbl> | Married <fctr> | Loan_Percentage <dbl> | Credit_History <fctr> | Bl… <fctr> | His… <fctr> | predi |
|---|---|---|---|---|---|---|
| 32.36561 | 1 | 76.07543 | 1 | 0 | 0 | |
| 32.36561 | 1 | 76.07543 | 1 | 0 | 1 | |
| 32.36561 | 1 | 76.07543 | 1 | 1 | 0 | |
| 32.36561 | 1 | 76.07543 | 0 | 0 | 0 | |

| Other_Obligations <dbl> | Married <fctr> | Loan_Percentage <dbl> | Credit_History <fctr> | Bl… <fctr> | His… <fctr> | predi |
|---|---|---|---|---|---|---|
| 32.36561 | 1 | 76.07543 | 0 | 0 | 1 | |
| 32.36561 | 1 | 76.07543 | 0 | 1 | 0 | |
| 32.36561 | 0 | 76.07543 | 0 | 0 | 0 | |
| 32.36561 | 0 | 76.07543 | 1 | 0 | 1 | |
| 32.36561 | 0 | 76.07543 | 1 | 1 | 0 | |
| 32.36561 | 0 | 76.07543 | 1 | 0 | 0 | |

1-10 of 12 rows Previous **1** 2 Next

We can see that White man with % of other obligations has 96% of chances to be approved for a loan. At the same situation, black people have, for the same situation only 90.9% of chances to be approved for a loan and Hispanics have only 91% of chances to be approved for a loan.

### (E) Estimate Probit Model

```
ProbitModel = glm(Approve ~ Other_Obligations + Credit_History + Loan_Percentage + Mar
ried + Black + Hispan, data = base, family = "binomial" (link = "probit"))

summary(ProbitModel)
```

```
##
## Call:
## glm(formula = Approve ~ Other_Obligations + Credit_History +
##     Loan_Percentage + Married + Black + Hispan, family = binomial(link = "probit"),
##     data = base)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.9425   0.2379   0.3068   0.3706   2.2951
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)        0.435076   0.337355   1.290 0.197167
## Other_Obligations -0.016173   0.005446  -2.969 0.002984 **
## Credit_History1    2.169170   0.122830  17.660  < 2e-16 ***
## Loan_Percentage   -0.007393   0.003284  -2.251 0.024368 *
## Married1           0.238606   0.091824   2.599 0.009363 **
## Black1            -0.450122   0.128091  -3.514 0.000441 ***
## Hispan1           -0.444120   0.168624  -2.634 0.008444 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1431.64  on 1936  degrees of freedom
## Residual deviance:  924.98  on 1930  degrees of freedom
## AIC: 938.98
##
## Number of Fisher Scoring iterations: 6
```

### The signals of the model suggest that:

1. ```
   Increases in Other Obligation diminish the likelihood of approval of the loan, wh
   ich makes sense.
   ```
2. ```
   Being Black or Hispanic diminish your likelihood of approval of the loan, which d
   oes not make sense (your ethnicity does not interfere to your capacity to pay a l
   oan), but was expected.
   ```

### (F) Predict probabilities for prototypical individuals

```
prototype_1$predictedprob <- round(predict(ProbitModel, newdata = prototype_1, type =
"response")*100, digits = 1)

prototype_2$predictedprob <- round(predict(ProbitModel, newdata = prototype_2, type =
"response")*100, digits = 1)

prototype_3$predictedprob <- round(predict(ProbitModel, newdata = prototype_3, type =
"response")*100, digits = 1)

prototype_4$predictedprob <- round(predict(ProbitModel, newdata = prototype_4, type =
"response")*100, digits = 1)

prototype_5$predictedprob <- round(predict(ProbitModel, newdata = prototype_5, type =
"response")*100, digits = 1)

prototype_6$predictedprob <- round(predict(ProbitModel, newdata = prototype_6, type =
"response")*100, digits = 1)

prototype_7$predictedprob <- round(predict(ProbitModel, newdata = prototype_7, type =
"response")*100, digits = 1)

prototype_8$predictedprob <- round(predict(ProbitModel, newdata = prototype_8, type =
"response")*100, digits = 1)

prototype_9$predictedprob <- round(predict(ProbitModel, newdata = prototype_9, type =
"response")*100, digits = 1)

prototype_10$predictedprob <- round(predict(ProbitModel, newdata = prototype_10, type
 = "response")*100, digits = 1)

prototype_11$predictedprob <- round(predict(ProbitModel, newdata = prototype_11, type
 = "response")*100, digits = 1)

prototype_12$predictedprob <- round(predict(ProbitModel, newdata = prototype_12, type
 = "response")*100, digits = 1)
```

### (F) Predict probabilities for prototypical individuals

```
rbind.data.frame(prototype_1, prototype_2, prototype_3, prototype_4, prototype_5,
    prototype_6, prototype_7, prototype_8, prototype_9, prototype_10, prototype_11,
    prototype_12)
```

| Other_Obligations <dbl> | Married <fctr> | Loan_Percentage <dbl> | Credit_History <fctr> | Bl… <fctr> | His… <fctr> | predi |
|---|---|---|---|---|---|---|
| 32.36561 | 1 | 76.07543 | 1 | 0 | 0 | |
| 32.36561 | 1 | 76.07543 | 1 | 0 | 1 | |
| 32.36561 | 1 | 76.07543 | 1 | 1 | 0 | |
| 32.36561 | 1 | 76.07543 | 0 | 0 | 0 | |

| Other_Obligations <dbl> | Married <fctr> | Loan_Percentage <dbl> | Credit_History <fctr> | Bl… <fctr> | His… <fctr> | predi |
|---|---|---|---|---|---|---|
| 32.36561 | 1 | 76.07543 | 0 | 0 | 1 | |
| 32.36561 | 1 | 76.07543 | 0 | 1 | 0 | |
| 32.36561 | 0 | 76.07543 | 0 | 0 | 0 | |
| 32.36561 | 0 | 76.07543 | 1 | 0 | 1 | |
| 32.36561 | 0 | 76.07543 | 1 | 1 | 0 | |
| 32.36561 | 0 | 76.07543 | 1 | 0 | 0 | |

1-10 of 12 rows Previous **1** 2 Next

We can see that White man with 32.37% of other obligations has 96.1% of chances to be approved for a loan. At the same situation, black people have, for the same situation only 90.4% of chances to be approved for a loan and Hispanics have only 90.5% of chances to be approved for a loan.
The results are consistent between both models.

# 5.2 Model considering Gender

**(A) Estimate Logit Model**

```
LogitModel = glm(Approve ~ Other_Obligations + Male, data = base,
                 family = "binomial")
summary(LogitModel)
```

```
##
## Call:
## glm(formula = Approve ~ Other_Obligations + Male, family = "binomial",
##     data = base)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6971   0.3912   0.4779   0.5348   1.9165
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)        3.910538   0.337325  11.593  < 2e-16 ***
## Other_Obligations -0.059923   0.008496  -7.053 1.75e-12 ***
## Male1              0.119326   0.176489   0.676    0.499
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1431.6  on 1936  degrees of freedom
## Residual deviance: 1378.3  on 1934  degrees of freedom
## AIC: 1384.3
##
## Number of Fisher Scoring iterations: 5
```

**The signals of the model suggest that:**

1. ```
   Increases in Other Obligation diminish the likelihood of approval of the loan, wh
   ich makes sense.
   ```

2. ```
   Being woman diminish your likelihood of approval of the loan, which does not make
   sense (your gender does not interfere to your capacity to pay a loan), but was ex
   pected.
   ```

### (B) Generate Odds Ratios

```
exp(coef(LogitModel))
```

```
##       (Intercept) Other_Obligations             Male1
##        49.9258045         0.9418374         1.1267374
```

**The Odd ratios suggest that:**

1. Man had 1.13 times the odds of being approved for a loan as a woman;(p<0.01)
2. One unit of increase in other obligations reduce in 5.82 percentage the odds of being approved for a loan in ;(p<0.01)

### (C) Define prototypical loan applicants

```
prototype_woman <- data.frame(Other_Obligations=mean(base$Other_Obligations),Male = 0)
prototype_woman <-prototype_woman %>% mutate(Male = as.factor(Male))
# levels(prototype_woman$Male) <- "Female"
prototype_men <- data.frame(Other_Obligations=mean(base$Other_Obligations),Male = 1)
prototype_men <- prototype_men %>% mutate(Male = as.factor(Male))
# levels(prototype_men$Male) <- "Male"
#Predict probabilities for prototypical individuals
prototype_woman$predictedprob <- round(
  predict (LogitModel,
           newdata = prototype_woman,
           type ="response")*100,
  digits = 1)


prototype_men$predictedprob <-  round(
  predict (LogitModel,
           newdata = prototype_men,
           type ="response")*100,
  digits = 1)
```

#### (C-1) Define prototypical loan applicants - Female

```
prototype_woman
```

| Other_Obligations <dbl> | Male <fctr> | predictedprob <dbl> |
|---:|---|---:|
| 32.36561 | 0 | 87.8 |

1 row

#### (C-2) Define prototypical loan applicants - Male

```
prototype_men
```

| Other_Obligations <dbl> | Male <fctr> | predictedprob <dbl> |
|---:|---|---:|
| 32.36561 | 1 | 89 |

1 row

Both Man and Woman have the similar probabilities of being approved for a loan considering they have 32.37% of other obligation

#### (D) Estimate Probit Model

```
ProbitModel = glm(Approve ~ Other_Obligations + Male , data = base,
                  family = "binomial" (link = "probit"))
summary(ProbitModel)
```

```
##
## Call:
## glm(formula = Approve ~ Other_Obligations + Male, family = binomial(link = "probi
t"),
##     data = base)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7309   0.3967   0.4839   0.5384   1.6647
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)        2.134905   0.177894  12.001  < 2e-16 ***
## Other_Obligations -0.030125   0.004577  -6.582 4.65e-11 ***
## Male1              0.052976   0.094495   0.561    0.575
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1431.6  on 1936  degrees of freedom
## Residual deviance: 1381.7  on 1934  degrees of freedom
## AIC: 1387.7
##
## Number of Fisher Scoring iterations: 5
```

**The signals of the model suggest that:** 1. Increases in Other Obligation diminish the likelihood of approval of the loan, which makes sense.
2. Being woman diminish your likelihood of approval of the loan, which does not make sense (your gender does not interfere to your capacity to pay a loan), but was expected.

#### (E) Predict probabilities for prototypical individuals

```
prototype_woman$predictedprob <- round(
  predict (ProbitModel,
           newdata = prototype_woman,
           type ="response")*100,
  digits = 1)
prototype_men$predictedprob <-  round(
  predict (ProbitModel,
           newdata = prototype_men,
           type ="response")*100,
  digits = 1)
```

#### (E-1) Predict probabilities for prototypical individuals - Female

```
prototype_woman
```

| Other_Obligations | Male | predictedprob |
|---|---|---|
| <dbl> | <fctr> | <dbl> |
| 32.36561 | 0 | 87.7 |

1 row

### (E-2) Predict probabilities for prototypical individuals - Male

```
prototype_men
```

| Other_Obligations | Male | predictedprob |
|---:|---|---:|
| <dbl> | <fctr> | <dbl> |
| 32.36561 | 1 | 88.7 |

1 row

The results are consistent between both models.