# Lending Mortage Analysis - Modeling

Exploratory Data Analysis

*Marjorie Blanco, Joe Thompson, Haodi Tu*

```
##
## Descriptive statistics
## ============================================================
## Statistic   N   Mean  St. Dev. Min  Pctl(25) Pctl(75)  Max
## ------------------------------------------------------------
## GDLIN     1,888 0.91    0.28    0      1        1       1
## OBRAT     1,888 32.36   8.06   0.00  28.00    37.00   95.00
## BLACK     1,888 0.10    0.30    0      0        0       1
## HISPAN    1,888 0.06    0.23    0      0        0       1
## APPROVE   1,888 0.88    0.33    0      1        1       1
## LOANPRC   1,888 75.44  16.51    2     69.3     89.6    100
## ------------------------------------------------------------
```

```
##
## Descriptive statistics - non-Hispanic White
## ============================================================
## Statistic   N   Mean  St. Dev. Min  Pctl(25) Pctl(75)  Max
## ------------------------------------------------------------
## GDLIN     1,597 0.94    0.24    0      1        1       1
## OBRAT     1,597 31.99   7.98   0.00  27.80    36.50   95.00
## APPROVE   1,597 0.91    0.28    0      1        1       1
## LOANPRC   1,597 74.08  16.87    2     67.1     87.4    100
## ------------------------------------------------------------
```

```
##
## Descriptive statistics - non-Hispanic Black
## ============================================================
## Statistic  N  Mean  St. Dev.  Min  Pctl(25) Pctl(75)  Max
## ------------------------------------------------------------
## GDLIN     187 0.72    0.45     0      0        1       1
## OBRAT     187 34.95   7.98   5.60   31.00    38.85   63.00
## HISPAN    187 0.00    0.00     0      0        0       0
## APPROVE   187 0.66    0.47     0      0        1       1
## LOANPRC   187 82.43  12.48   28.99  79.97    90.13   99.76
## ------------------------------------------------------------
```

```
##
## Descriptive statistics - Hispanic
## ============================================================
## Statistic  N  Mean  St. Dev.  Min  Pctl(25) Pctl(75)  Max
## ------------------------------------------------------------
## GDLIN     104 0.87    0.34     0      1        1       1
## OBRAT     104 33.41   8.52   14.60  29.00    38.05   62.00
## BLACK     104 0.00    0.00     0      0        0       0
## APPROVE   104 0.78    0.42     0      1        1       1
## LOANPRC   104 83.78  10.81    40      80      90.2    96
## ------------------------------------------------------------
```

```
##  MARRIED  GDLIN        OBRAT       BLACK    HISPAN   MALE     APPROVE
##  .:   2   0: 163   Min.   : 0.00   0:1701   0:1784   0: 352   0: 228
##  0: 638   1:1725   1st Qu.:28.00   1: 187   1: 104   1:1536   1:1660
##  1:1248            Median :33.00
##                    Mean   :32.36
##                    3rd Qu.:37.00
##                    Max.   :95.00
##     LOANPRC
##  Min.   : 2.105
##  1st Qu.:69.262
##  Median :80.000
##  Mean   :75.442
##  3rd Qu.:89.613
##  Max.   :99.760

## $Hispanic
##  MARRIED GDLIN      OBRAT       MALE   APPROVE    LOANPRC
##  .: 1    0:14   Min.   :14.60   0:20   0:23    Min.   :40.09
##  0:30    1:90   1st Qu.:29.00   1:84   1:81    1st Qu.:80.00
##  1:73           Median :33.00                  Median :88.85
##                 Mean   :33.41                  Mean   :83.78
##                 3rd Qu.:38.05                  3rd Qu.:90.22
##                 Max.   :62.00                  Max.   :95.65
##                  RACE
##  Hispanic          :104
##  non-Hispanic Black:  0
##  non-Hispanic White:  0
##
##
##
##
## $`non-Hispanic Black`
##  MARRIED GDLIN       OBRAT       MALE    APPROVE    LOANPRC
##  .:  0   0: 52   Min.   : 5.60   0: 48   0: 63   Min.   :28.99
##  0: 74   1:135   1st Qu.:31.00   1:139   1:124   1st Qu.:79.97
##  1:113           Median :35.00                   Median :82.35
##                  Mean   :34.95                   Mean   :82.43
##                  3rd Qu.:38.85                   3rd Qu.:90.13
##                  Max.   :63.00                   Max.   :99.76
##                  RACE
##  Hispanic          :  0
##  non-Hispanic Black:187
##  non-Hispanic White:  0
##
##
##
##
## $`non-Hispanic White`
##  MARRIED  GDLIN        OBRAT       MALE     APPROVE     LOANPRC
##  .:   1   0:  97   Min.   : 0.00   0: 284   0: 142   Min.   : 2.105
##  0: 534   1:1500   1st Qu.:27.80   1:1313   1:1455   1st Qu.:67.123
##  1:1062            Median :32.60                    Median :79.769
##                    Mean   :31.99                    Mean   :74.081
##                    3rd Qu.:36.50                    3rd Qu.:87.429
##                    Max.   :95.00                    Max.   :99.518
##                   RACE
##  Hispanic          :   0
```

```
##  non-Hispanic Black:   0
##  non-Hispanic White:1597
##
##
##
```

## Data Discription

The data set includes the following variables:

- APPROVE = 1 if mortgage loan was approved, = 0 otherwise
- GDLIN = 1 if credit history meets guidelines, = 0 otherwise
- LOANPRC = loan amount/purchase price
- OBRAT = other obligations as a percent of total income
- MALE = 1 if male, = 0 otherwise
- MARRIED = 1 if married, = 0 otherwise
- BLACK = 1 if black, = 0 otherwise
- HISPAN = 1 if Hispanic, = 0 otherwise

## Logistic Model

$$log(p/1-p) = \beta_0 + \beta_1 * GDLIN + \beta_2 * OBRAT + \beta_3 * BLACK + \beta_4 * HISPAN + \beta_5 * LOANPRC$$

```
## # weights:  7 (6 variable)
## initial  value 1308.661877
## iter  10 value 451.283143
## final  value 451.261471
## converged
```

```
## 'log Lik.' -451.2615 (df=6)
```

```
##
## Logistic Regression
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                                APPROVE
## -----------------------------------------------
## GDLIN1                        3.737***
##                                (0.221)
##
## OBRAT                         -0.031***
##                                (0.011)
##
## BLACK1                        -0.917***
##                                (0.246)
##
## HISPAN1                       -0.827**
##                                (0.324)
##
## LOANPRC                       -0.017**
##                                (0.007)
##
## Constant                       1.533**
```

```
##                             (0.699)
##
## ------------------------------------------
## Observations                 1,888
## Log Likelihood             -451.261
## Akaike Inf. Crit.           914.523
## ==========================================
## Note:             *p<0.1; **p<0.05; ***p<0.01

##
## Odds Ratio
## ==========================================
##                       Dependent variable:
##                   -------------------------
##                             APPROVE
## ------------------------------------------
## GDLIN1                     41.961***
##                             (0.221)
##
## OBRAT                      0.969***
##                             (0.011)
##
## BLACK1                      0.400
##                             (0.246)
##
## HISPAN1                     0.438
##                             (0.324)
##
## LOANPRC                    0.983***
##                             (0.007)
##
## Constant                   4.631***
##                             (0.699)
##
## ------------------------------------------
## Observations                 1,888
## Log Likelihood             -451.261
## Akaike Inf. Crit.           914.523
## ==========================================
## Note:             *p<0.1; **p<0.05; ***p<0.01

##
## ==========================================
##                       Dependent variable:
##                   -------------------------
##                             APPROVE
## ------------------------------------------
## GDLIN1                     -3.737***
##                             (0.221)
##
## OBRAT                      0.031***
##                             (0.011)
##
## BLACK1                     0.917***
##                             (0.246)
##
```

```
## HISPAN1                      0.827**
##                              (0.324)
##
## LOANPRC                      0.017**
##                              (0.007)
##
## Constant                    -1.533**
##                              (0.699)
##
## ------------------------------------------
## Akaike Inf. Crit.            914.523
## ==========================================
## Note:             *p<0.1; **p<0.05; ***p<0.01
```

Controlling for other variables, applicants who meet the credit guidelines have an estimate of 41.9612143 times higher odds of loan approval than the applicants who do not meet the credit guidelines. We find that after controlling for all other variables, odds of loan approval for White applicants are 2.46 times and 3.24 times greater in comparison to Blacks and Hispanics respectively.

For every one unit change in `OBRAT`, the log odds of loan approval (versus non loan approval) decreases by 0.0312188.

For every one unit change in `LOANPRC`, the log odds of loan approval (versus non loan approval) decreases by 0.0167011.

The log odds of loan approval for applicants that meet credit guidelines increases by 3.7367457.

The log odds of loan approval for Black applicants decreases by 0.9171449.

The log odds of loan approval for Hispanic applicants decreases by 0.8266367.

For example, for a black person whose credit history meets guidline (GDLIN = 1), loan amount price is 100 (LOANPRC = 100) and other obligations as a percent of total income is none (OBRAT = 0), the log odds of loan approval is 93.6%

**CIs using profiled log-likelihood**

```
##                   2.5 %       97.5 %
## (Intercept)  0.19416207  2.937219042
## GDLIN1       3.31229466  4.182112724
## OBRAT       -0.05333480 -0.009301146
## BLACK1      -1.39135036 -0.426283068
## HISPAN1     -1.43764172 -0.165924717
## LOANPRC     -0.03156216 -0.002916465
```

**CIs using standard errors**

```
##                   2.5 %       97.5 %
## (Intercept)  0.16272694  2.902760759
## GDLIN1       3.30270694  4.170784503
## OBRAT       -0.05325850 -0.009179045
## BLACK1      -1.39893109 -0.435358718
## HISPAN1     -1.46068910 -0.192584237
## LOANPRC     -0.03101509 -0.002387147
```

**Odds ratios only**

```
## (Intercept)      GDLIN1       OBRAT      BLACK1     HISPAN1     LOANPRC
##   4.6308658  41.9612143   0.9692635   0.3996585   0.4375183   0.9834376
```

**Odds ratios and 95% CI**

```
##                     OR      2.5 %     97.5 %
## (Intercept)  4.6308658  1.2142931 18.8633152
## GDLIN1      41.9612143 27.4480373 65.5040992
## OBRAT        0.9692635  0.9480625  0.9907420
## BLACK1       0.3996585  0.2487392  0.6529315
## HISPAN1      0.4375183  0.2374872  0.8471100
## LOANPRC      0.9834376  0.9689307  0.9970878
```

| GDLIN | OBRAT | BLACK | HISPAN | LOANPRC | fit | PredictedProb |
|---|---|---|---|---|---|---|
| 0 | 32.35767 | 1 | 0 | 75.44245 | -1.6545412 | 0.1604961 |
| 1 | 32.35767 | 1 | 0 | 75.44245 | 2.0822046 | 0.8891615 |
| 0 | 32.35767 | 0 | 1 | 75.44245 | -1.5640329 | 0.1730687 |
| 1 | 32.35767 | 0 | 1 | 75.44245 | 2.1727128 | 0.8977722 |
| 0 | 32.35767 | 0 | 0 | 75.44245 | -0.7373963 | 0.3235738 |
| 1 | 32.35767 | 0 | 0 | 75.44245 | 2.9993495 | 0.9525447 |

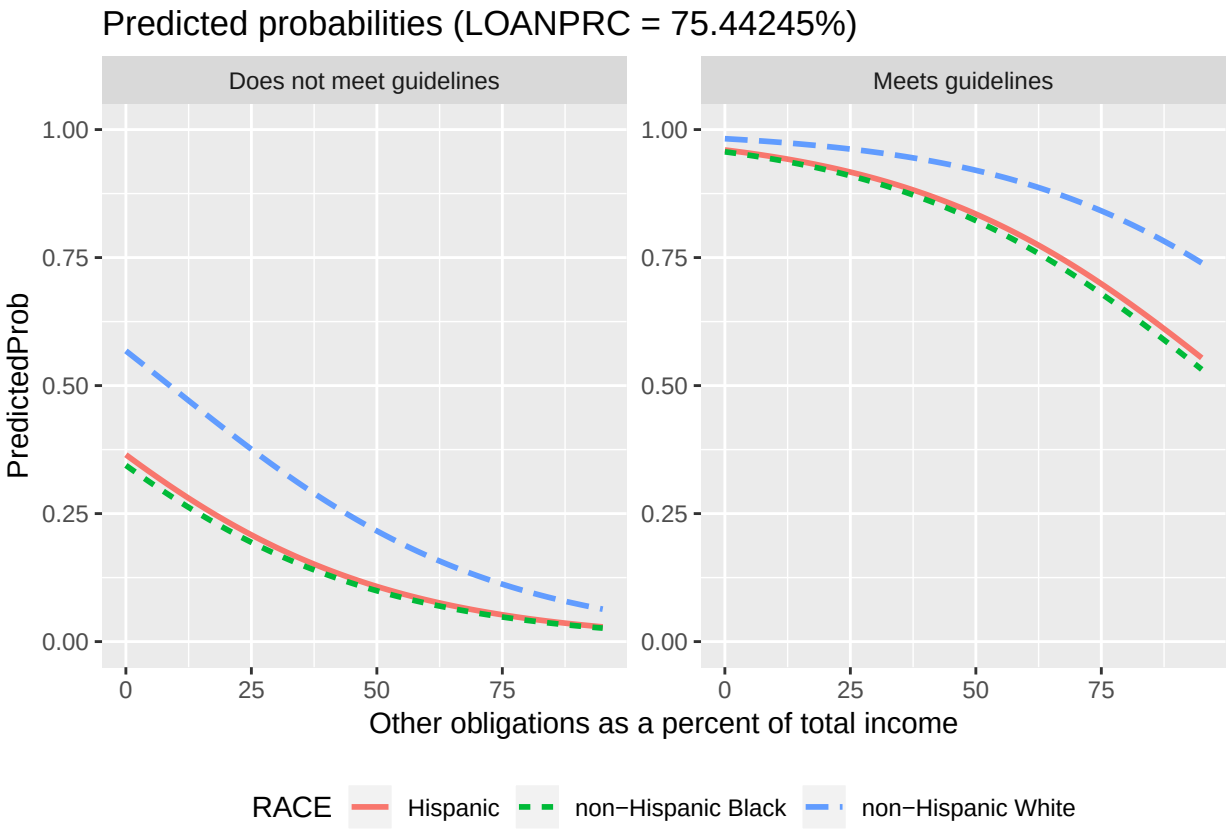
```
## [1] 0.1630777
```

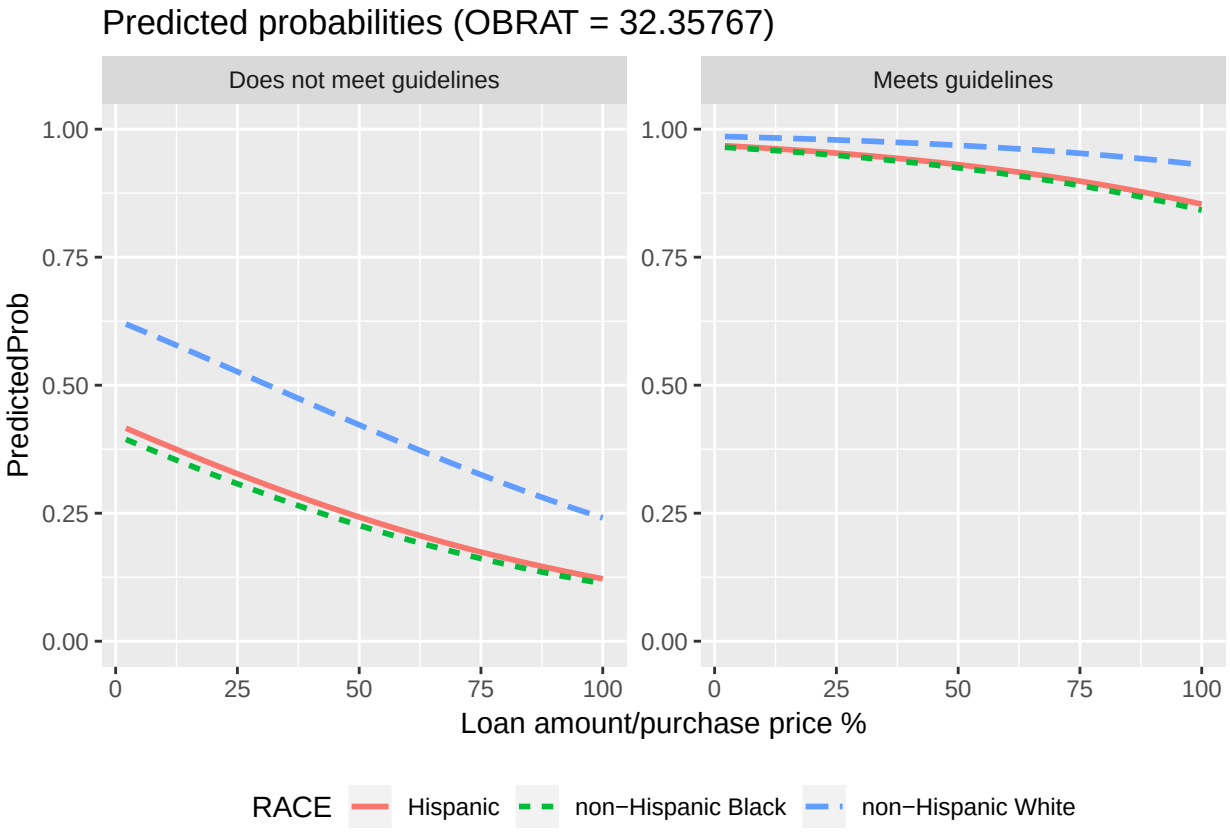
```
## [1] 0.0633832
```

This table represents the predicted probability of some prototypical individuals across different races from Logit model. This table shows that Black and Hispanic applicants are less likely to receive loan approval in comparison to White applicants.

We observe that White applicants who meet the credit guidelines with DTI and LTV evaluated at the mean of data set have approximately 16.3% point higher probability of loan approval than similar Black applicants.

We observe that White applicants who do not meet the credit guidelines with DTI and LTV evaluated at the mean of data set have approximately 6.3% point higher probability of loan approval than similar Black applicants.


Predicted probabilities (LOANPRC = 75.44245%)
Does not meet guidelines
Meets guidelines
PredictedProb
1.00
0.75
0.50
0.25
0.00
0
25
50
75
Other obligations as a percent of total income
RACE
Hispanic
non−Hispanic Black
non−Hispanic White


Predicted probabilities (OBRAT = 32.35767)
Does not meet guidelines
Meets guidelines
PredictedProb
1.00
0.75
0.50
0.25
0.00
0
25
50
75
100
Loan amount/purchase price %
RACE
Hispanic
non−Hispanic Black
non−Hispanic White

**Probit Model**

$$Probit(approve) = \beta_0 + \beta_1 * GDLIN + \beta_2 * OBRAT + \beta_3 * BLACK + \beta_4 * HISPAN + \beta_5 * LOANPRC$$

```
##
## Call:
## glm(formula = APPROVE ~ GDLIN + OBRAT + BLACK + HISPAN + LOANPRC,
##     family = binomial(link = "probit"), data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.8688   0.2496   0.3144   0.3604   2.1669
##
## Coefficients:
##              Estimate Std. Error z value             Pr(>|z|)
## (Intercept)  0.583275   0.341346   1.709             0.087497 .
## GDLIN1       2.161541   0.124153  17.410 < 0.0000000000000002 ***
## OBRAT       -0.014974   0.005721  -2.617             0.008858 **
## BLACK1      -0.473416   0.129376  -3.659             0.000253 ***
## HISPAN1     -0.422091   0.168680  -2.502             0.012338 *
## LOANPRC     -0.007808   0.003404  -2.294             0.021781 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1391.24  on 1887  degrees of freedom
## Residual deviance:  902.36  on 1882  degrees of freedom
## AIC: 914.36
##
## Number of Fisher Scoring iterations: 6

## 'log Lik.' -451.1823 (df=6)

##
## Probit Regression
## ================================================
##                         Dependent variable:
##                     ---------------------------
##                                APPROVE
## ------------------------------------------------
## GDLIN1                        2.162***
##                                (0.124)
##
## OBRAT                         -0.015***
##                                (0.006)
##
## BLACK1                        -0.473***
##                                (0.129)
##
## HISPAN1                       -0.422**
##                                (0.169)
##
## LOANPRC                       -0.008**
##                                (0.003)
##
```

```
## Constant                    0.583*
##                             (0.341)
##
## ------------------------------------------
## Observations                 1,888
## Log Likelihood             -451.182
## Akaike Inf. Crit.           914.365
## ==========================================
## Note:             *p<0.1; **p<0.05; ***p<0.01
```

For every one unit change in `OBRAT`, the probit odds of loan approval (versus non loan approval) decreases by 0.0149739.

For every one unit change in `LOANPRC`, the probit odds of loan approval (versus non loan approval) decreases by 0.0078082.

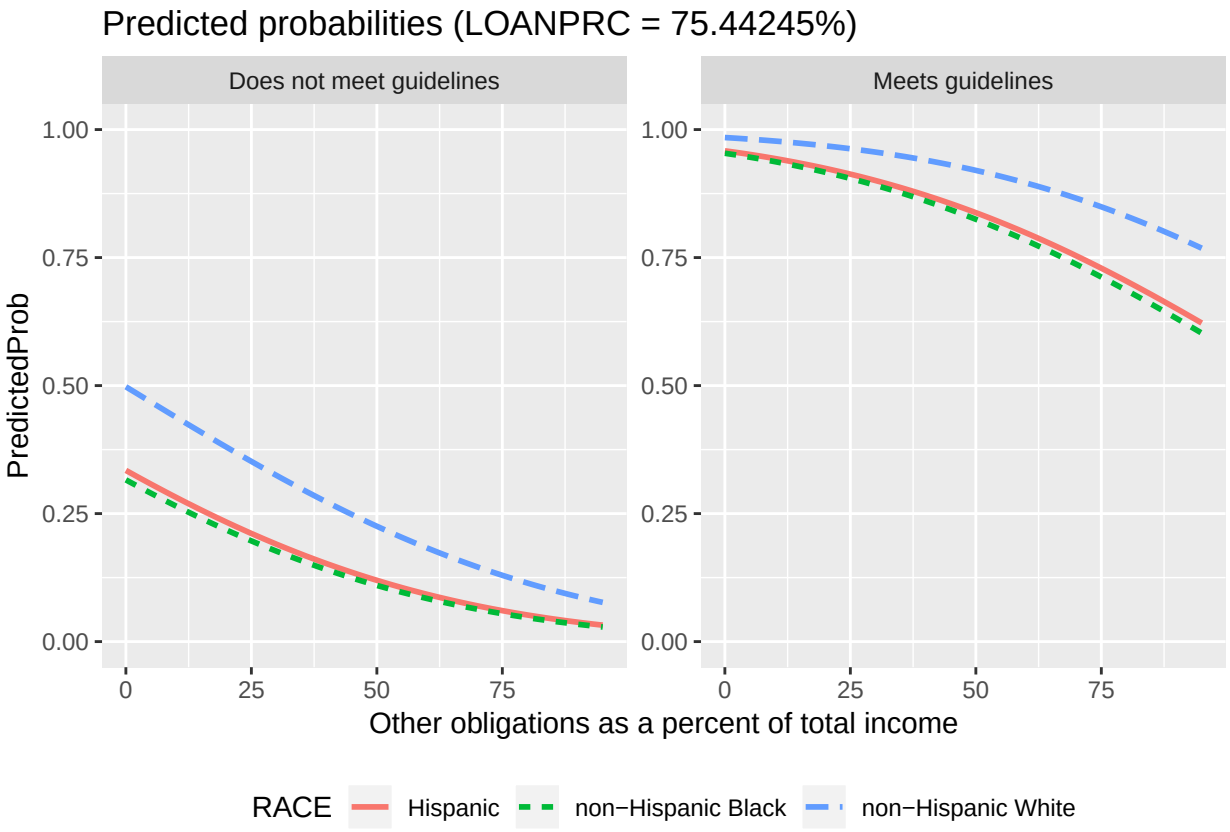
The probit odds of loan approval for applicants that meet credit guidelines increases by 2.1615407.

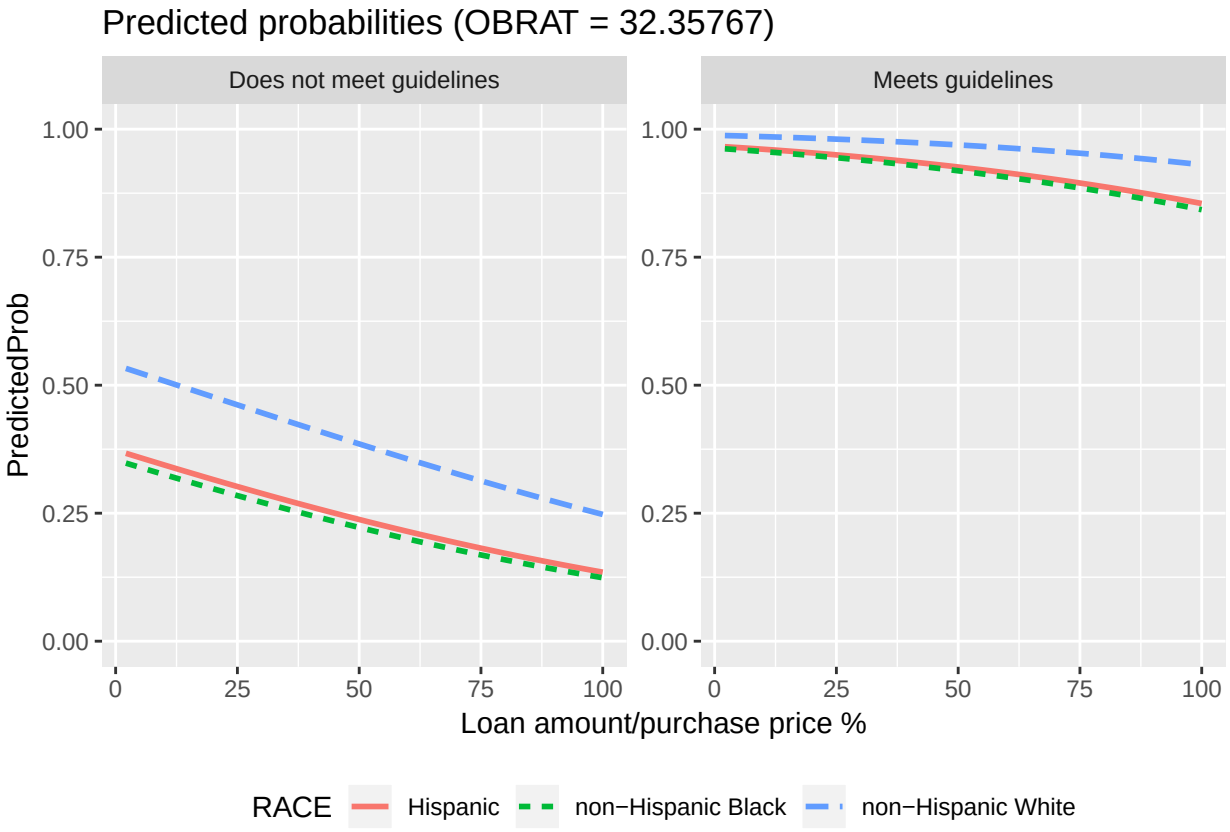
The probit odds of loan approval for Black applicants decreases by 0.4734163.

The probit odds of loan approval for Hispanic applicants decreases by 0.4220909.

For example, for a black person whose credit history meets guidline (GDLIN = 1), loan amount price is 100 (LOANPRC = 100) and other obligations as a percent of total income is none (OBRAT = 0), the probit odds of loan approval is 93.2%

| GDLIN | OBRAT | BLACK | HISPAN | LOANPRC | fit | fit |
|---|---|---|---|---|---|---|
| 0 | 32.35767 | 1 | 0 | 75.44245 | -0.9637321 | 0.1675901 |
| 1 | 32.35767 | 1 | 0 | 75.44245 | 1.1978087 | 0.8845042 |
| 0 | 32.35767 | 0 | 1 | 75.44245 | -0.9124067 | 0.1807773 |
| 1 | 32.35767 | 0 | 1 | 75.44245 | 1.2491340 | 0.8941920 |
| 0 | 32.35767 | 0 | 0 | 75.44245 | -0.4903158 | 0.3119552 |
| 1 | 32.35767 | 0 | 0 | 75.44245 | 1.6712249 | 0.9526614 |


Predicted probabilities (LOANPRC = 75.44245%)
Does not meet guidelines
Meets guidelines
PredictedProb
1.00
0.75
0.50
0.25
0.00
0
25
50
75
Other obligations as a percent of total income
RACE
Hispanic
non−Hispanic Black
non−Hispanic White


Predicted probabilities (OBRAT = 32.35767)
Does not meet guidelines
Meets guidelines
PredictedProb
1.00
0.75
0.50
0.25
0.00
0
25
50
75
100
Loan amount/purchase price %
RACE
Hispanic
non−Hispanic Black
non−Hispanic White

## Probablities Comparison

| GDLIN | OBRAT | BLACK | HISPAN | LOANPRC | LogitProb | ProbitProb |
|---|---|---|---|---|---|---|
| 0 | 32.35767 | 1 | 0 | 75.44245 | 0.1604961 | 0.1675901 |
| 1 | 32.35767 | 1 | 0 | 75.44245 | 0.8891615 | 0.8845042 |
| 0 | 32.35767 | 0 | 1 | 75.44245 | 0.1730687 | 0.1807773 |
| 1 | 32.35767 | 0 | 1 | 75.44245 | 0.8977722 | 0.8941920 |
| 0 | 32.35767 | 0 | 0 | 75.44245 | 0.3235738 | 0.3119552 |
| 1 | 32.35767 | 0 | 0 | 75.44245 | 0.9525447 | 0.9526614 |