Chicago Housing

Annelise Escher

5/29/2022

```
rm(list = ls())
library(tidyverse)
library(stargazer)
library(readxl)
library(knitr)
library(lubridate)
library(dplyr)
library(devtools)
library(sf)
library(janitor)
library(MASS)
library(lmtest)
library(sandwich)
knitr::opts_chunk$set(echo = FALSE)
```

Reading in, cleaning, and combining data

Reading in data

Cleaning main datasets

Adding redlining grade and zipcodes to purchase origination dataframe.

Data visualizations

Creating dataframe for visualization by zip code.

```
## # A tibble: 3 x 7
##   zip   loan_total white_percent_avg black_percent_avg latino_percent_avg
##   <chr>      <dbl>             <dbl>             <dbl>              <dbl>
## 1 60614 7683047000             0.791            0.0368             0.0796
## 2 60647 5446853000             0.530            0.0504             0.346
## 3 60657 5118651000             0.812            0.0279             0.0723
## # ... with 2 more variables: asian_percent_avg <dbl>, majority_race <chr>
```

Recreating WBEZ figure. Map of Chicago zip codes color coded by the dollar amount of mortgage loans made to that area from 2012-2018.

```
## pdf
##   2
```

Bar chart of total home mortgage lending to Chicago zip codes, color coding for majority race of neighborhood.

```
## pdf
##   2
```

Creating dataframe of modern home mortgage lending by census tract used for statistical analysis.

```
## # A tibble: 3 x 11
##   census_tract loan_total white_percent_avg black_percent_avg latino_percent_avg
##   <chr>             <dbl>             <dbl>             <dbl>              <dbl>
## 1 17031330100  1747818000             0.482            0.215              0.0611
## 2 17031081800  1010495000             0.826            0.0114             0.0549
## 3 17031833100   835049000             0.606            0.115              0.0892
## # ... with 6 more variables: asian_percent_avg <dbl>, tract_pop <dbl>,
## #   med_family_income <dbl>, avg_loan_amount <dbl>, majority_race <chr>,
## #   id <int>
```

Creating dataframe including redlining scores by census tract for statistical analysis.

```
## [1] 795
```

```
## [1] 742
```

Regression 1: Simple OLS

```
##
## Call:
## lm(formula = loans_millions ~ HRS2010 + tract_pop + med_family_income +
##     avg_loan_amount + majority_race, data = for_regression)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -285.719  -24.549   -2.907   18.785  311.622
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          7.902e+02  3.213e+02   2.460  0.01413 *
## HRS2010              2.306e+01  3.211e+00   7.180 1.71e-12 ***
## tract_pop            1.679e-02  1.164e-03  14.418  < 2e-16 ***
## med_family_income   -1.184e-02  4.139e-03  -2.860  0.00435 **
## avg_loan_amount      1.832e-04  1.384e-05  13.237  < 2e-16 ***
## majority_raceblack  -1.487e+01  6.374e+00  -2.332  0.01995 *
## majority_racelatino -2.671e+01  6.415e+00  -4.163 3.51e-05 ***
## majority_racewhite   7.404e+01  6.172e+00  11.995  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 49.82 on 734 degrees of freedom
## Multiple R-squared:  0.6664, Adjusted R-squared:  0.6632
## F-statistic: 209.5 on 7 and 734 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = loans_millions ~ redline_1930_value + tract_pop +
##     med_family_income + avg_loan_amount + majority_race, data = for_regression)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -284.298  -24.260   -2.334   18.509  315.055
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           7.996e+02  3.229e+02   2.476  0.01349 *
## redline_1930_value    1.992e+01  3.009e+00   6.620 6.91e-11 ***
## tract_pop             1.656e-02  1.167e-03  14.187  < 2e-16 ***
## med_family_income    -1.182e-02  4.160e-03  -2.842  0.00461 **
## avg_loan_amount       1.840e-04  1.390e-05  13.237  < 2e-16 ***
## majority_raceblack   -1.555e+01  6.412e+00  -2.425  0.01556 *
## majority_racelatino -2.683e+01  6.448e+00  -4.161 3.54e-05 ***
## majority_racewhite    7.289e+01  6.191e+00  11.773  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 50.07 on 734 degrees of freedom
## Multiple R-squared:  0.6631, Adjusted R-squared:  0.6599
## F-statistic: 206.4 on 7 and 734 DF,  p-value: < 2.2e-16
```

Regression 2: Past category of HOLC grade as predictor for modern recreated distribution

Creating a new variable showing modern census tracts categorized by the same distribution as the 1930s HOLC grade for chicago.

```
##
##   1   2   3   4
##   4  66 382 290
```

```
##
##           1           2           3           4
## 0.005390836 0.088948787 0.514824798 0.390835580
```

Table comparing number of census tracts in each of the four HOLC grades. This shows us that the top census tracts by level of

```
##
##   1   2   3   4
##   3  61 387 291
```

```
##                     ranking_2010   1   2   3   4
## redline_1930_value
## 1                                  0   1   3   0
## 2                                  0   6  46  14
## 3                                  0  20 222 140
## 4                                  3  34 116 137
```

```
##
## Call:
## lm(formula = ranking_2010 ~ value_3 + value_4, data = for_regression)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3345 -0.3141 -0.3141  0.6655  0.9000
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.10000    0.07528  41.180  < 2e-16 ***
## value_3      0.21414    0.08189   2.615  0.00910 **
## value_4      0.23448    0.08387   2.796  0.00531 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6298 on 739 degrees of freedom
## Multiple R-squared:  0.01086,    Adjusted R-squared:  0.008183
## F-statistic: 4.057 on 2 and 739 DF,  p-value: 0.01769
```

```
##
## Call:
## lm(formula = ranking_2010 ~ HRS2010, data = for_regression)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3675 -0.2895 -0.2755  0.6374  0.8103
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.01187    0.12732  23.656   <2e-16 ***
## HRS2010      0.08892    0.03838   2.316   0.0208 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6306 on 740 degrees of freedom
## Multiple R-squared:  0.007199,   Adjusted R-squared:  0.005858
## F-statistic: 5.366 on 1 and 740 DF,  p-value: 0.0208
```

Regression 3: Linear Probability Model

```
##
## t test of coefficients:
##
##                     Estimate Std. Error t value  Pr(>|t|)
## (Intercept)         1.099142   0.066918  16.425 < 2.2e-16 ***
## redline_1930_value -0.305308   0.018590 -16.424 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## t test of coefficients:
##
```

```
##                    Estimate Std. Error t value  Pr(>|t|)
## (Intercept)         1.698755   0.145169 11.7019 < 2.2e-16 ***
## redline_1930_value -0.359736   0.040676 -8.8439 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## t test of coefficients:
##
##                    Estimate Std. Error t value  Pr(>|t|)
## (Intercept)        -1.797897   0.084950 -21.164 < 2.2e-16 ***
## redline_1930_value  0.665045   0.024034  27.671 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## t test of coefficients:
##
##                    Estimate Std. Error t value  Pr(>|t|)
## (Intercept)         1.099142   0.066918  16.425 < 2.2e-16 ***
## redline_1930_value -0.305308   0.018590 -16.424 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## t test of coefficients:
##
##                    Estimate Std. Error t value  Pr(>|t|)
## (Intercept)         1.698755   0.145169 11.7019 < 2.2e-16 ***
## redline_1930_value -0.359736   0.040676 -8.8439 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## t test of coefficients:
##
##                    Estimate Std. Error t value  Pr(>|t|)
## (Intercept)        -1.797897   0.084950 -21.164 < 2.2e-16 ***
## redline_1930_value  0.665045   0.024034  27.671 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Regression 4: Ordered logit

```
## Call:
## polr(formula = as.factor(ranking_2010) ~ redline_1930_value,
##     data = for_regression, Hess = TRUE)
##
## Coefficients:
##                    Value Std. Error t value
## redline_1930_value 0.3529     0.1118   3.158
##
```

```
## Intercepts:
##     Value   Std. Error t value
## 1|2 -4.3709  0.6779    -6.4481
## 2|3 -1.2323  0.3767    -3.2710
## 3|4  1.5949  0.3751     4.2519
##
## Residual Deviance: 1376.411
## AIC: 1384.411
```

Source: https://stats.oarc.ucla.edu/r/dae/ordinal-logistic-regression/

Regression summaries in stargazer for Latex

Table 1:

	Dependent variable:	
	loans_millions	
	Continous Redline Grade	Categorical Redline Grade
	(1)	(2)
Continous Redline Grade	23.059^{***} (3.211)	
Redline Categories		19.919^{***} (3.009)
Tract Population	0.017^{***} (0.001)	0.017^{***} (0.001)
Average Family Income	-0.012^{***} (0.004)	-0.012^{***} (0.004)
Average Loan Amount	0.0002^{***} (0.00001)	0.0002^{***} (0.00001)
Majority Black	-14.866^{**} (6.374)	-15.547^{**} (6.412)
Majority Latino	-26.707^{***} (6.415)	-26.832^{***} (6.448)
Majority White	74.039^{***} (6.172)	72.890^{***} (6.191)
Observations	742	742
R^2	0.666	0.663
Adjusted R^2	0.663	0.660
Residual Std. Error (df = 734)	49.821	50.068
F Statistic (df = 7; 734)	209.468^{***}	206.377^{***}
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 2:

	Dependent variable:	
	Modern Lending Ranking	
	(1)	(2)
Continous Variable	0.089^{**} (0.038)	
Grade 3		0.214^{***} (0.082)
Grade 4		0.234^{***} (0.084)
Observations	742	742
R^2	0.007	0.011
Adjusted R^2	0.006	0.008
Residual Std. Error	0.631 (df = 740)	0.630 (df = 739)
F Statistic	5.366^{**} (df = 1; 740)	4.057^{**} (df = 2; 739)
Note:		*p<0.1; **p<0.05; ***p<0.01

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Sun, May 29, 2022 - 23:17:31

Table 3:

	Dependent variable:
	ranking_2010
redline_1930_value	0.353***
	(0.112)
Observations	742
Note:	*p<0.1; **p<0.05; ***p<0.01