

Answer Key

The mortgage foreclosure crisis has been in the news lately. Changes in the mortgage market since the middle of the 1990s has lead to both more risky loans and to increasing geographic concentration. Specifically, the subprime mortgages (loans with relatively high mortgage rates and higher upfront fees to compensate for associated greater risk) have become associated with increasing foreclosures in particular regions. Dan Immergluck¹ wanted to know why some metropolitan areas had higher concentrations of these loans than others. Therefore, he created a regression model to examine factors that could predict the share of subprime loans of all home loans in 2003 in 103 U.S. metropolitan areas with populations greater than 500,000.

Table 1 presents the variables he included in his model, their means and standard deviations, the OLS regression coefficients for each factor, the standard errors, and the levels of significance.

Table 1. Results of model predicting share of 2003 home purchase loans made by subprime lenders in U.S. metropolitan areas with populations over 500,000 in 2000.

	Mean	Std. dev.	B	Std. error	Signif.
Constant			1.107	0.270	0.000
Proportion of purchase loans made by subprime lenders, 1997	0.111	0.058	0.594	0.353	0.096
Median home purchase loan size, 1997	0.033	0.019	-8.320×10^{-7}	5.449×10^{-7}	0.130
Proportion buyers who were Black, 1997	0.060	0.046	0.164	0.253	0.518
Proportion buyers who were Hispanic, 1997	0.068	0.113	0.086	0.057	0.135
Proportion residents with college degree or more, 2000	0.262	0.060	-0.315	0.134	0.021
Proportion of housing units built before 1950, 2000	0.207	0.127	0.042	0.046	0.356
Ratio of median loan size to FHA limit, 1997	0.381	0.067	0.057	0.120	0.635
Ratio of median buyer income to median loan size, 1997	0.536	0.078	-3.186	0.844	0.000
Ratio of median buyer income to median loan size, 1997, squared	0.293	0.086	2.463	0.693	0.001
1997 unemployment rate	0.051	0.023	0.178	0.520	0.732
Change in unemployment rate, 1997-2003	0.015	0.018	0.458	0.445	0.307
Ratio of median purchase loan size, 1997-2003	0.448	0.187	0.121	0.060	0.047
Change in proportion of buyers who were Black, 1997-2003	-0.005	0.015	0.809	1.701	0.636
Change in proportion of buyers who were Hispanic, 1997-2003	0.018	0.023	1.281	0.896	0.156
<i>N</i> = 103					
<i>R</i> ² = .672					

Please answer the following questions to help you think through his analysis.

¹ Immergluck, Dan. 2008. "From the exotic to the subprime: excessive mortgage market risk and foreclosures." *Journal of the American Planning Association* 74, 1: 59-76.

1. Interpret the estimated regression coefficients for “proportion residents with college degree or more, 2000” (10 points)

Each additional percentage point of residents with a college degree or more is associated with a decrease of 0.315 in the share of 2003 home purchase loans that were made by subprime lenders in U.S. metropolitan areas with populations over 500,000 in 2000.

2. Is the estimated regression coefficient for “proportion residents with college degree or more, 2000” significantly different from zero at a 5 percent level? Make sure to formally set up and test the hypotheses. HINT: You may be able to use information in the table to save you some steps. (10 points)

Hypothesis test for one coefficient:

Set up hypothesis: $H_0: \beta_5 = 0$ $H_a: \beta_5 \neq 0$

Decision rule: **If $|t| > t_\alpha$ then reject the null hypothesis. This is a two-sided test, so the critical value $t_\alpha = 1.99$ with 88 d.f. ($n-k-1 = 103-14-1 = 88$)**

Calculate test statistic: $t = \frac{\hat{\beta}_5 - \beta_{H_0}}{SE_{\hat{\beta}_5}} = \frac{-0.315}{0.134} = -2.35$

Since $|t| > t_\alpha$ (that is, $2.35 > 1.99$), we can reject the null hypothesis and conclude that, yes, the estimated coefficient for “proportion residents with college degree or more, 2000” is significantly different from zero at a 5% level.

3. Construct and interpret 95 percent confidence interval around the estimated regression coefficient for “proportion residents with college degree or more, 2000.” (10 points)

A 95% confidence interval around the true coefficient is calculated as:

$$\hat{\beta}_5 \pm t_c * SE_{\hat{\beta}_5}$$

So, the 95% confidence interval around the coefficient for “proportion residents with college degree or more, 2000” is:

$$-0.315 \pm 1.99 * 0.134 = [-0.582, -0.048]$$

We can say with 95% certainty that a one percentage point increase in residents with a college degree or more in the year 2000 is associated with between a 0.048 and 0.582 decrease in the share of 2003 home purchase loans that were made by subprime lenders (again, in U.S. metropolitan areas with populations over 500,000 in 2000).

4. How strong is the evidence of a relationship between the share of home purchase loans made by subprime lenders in 2003 and education levels (measured by the proportion of residents with a college degree)? Explain how you come to your conclusions. (5 points)

To assess the strength of the evidence for a relationship between a particular variable and the dependent variable, we can use the p-value. In this case, the p-value = 0.021, which tells us that there is only a 2% chance of falsely rejecting the null hypothesis that there is no linear relationship between our measure of education levels and the share of home purchase loans that were made by subprime lenders in 2003. (Equivalently, we can say that we would expect to see a coefficient of -0.315, or one even further away from 0, 2% of the time if the true coefficient were really 0.)

5. Interpret the R^2 for the model. (5 points)

R^2 for this regression tells us that the variables included in the model explain 67.2% of the variation in the share of home purchase loans made by subprime lenders in 2003.

6. Conduct a hypothesis test about whether this model has significant explanatory power. Make sure to formally set up your hypotheses. (10 points)

To test whether the model has significant explanatory power, we conduct a global F-test:

H_0 : all of the coefficients in the model equal zero (i.e. $\beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0$)

H_a : at least one coefficient does not equal zero

Decision rule: if the critical F-value is greater than the calculated F-statistic, we reject the null hypothesis.

$k = 14$ degrees of freedom for the numerator

$n - (k + 1) = 103 - (14 + 1) = 88$ degrees of freedom for the denominator

$F_{\text{critical}, \alpha=0.05} = 1.8$

$$F = \frac{R^2/k}{(1-R^2)/(n-(k+1))} \text{ or, equivalently, } = \frac{R}{(1-R)} * \frac{n-(k+1)}{k}$$

$$= \frac{0.672/14}{(1-0.672)/88} = 12.878$$

$F > F_{\text{critical}, \alpha=0.05}$, so we can reject the null hypothesis and conclude that, yes, our model has significant explanatory power.

7. Given your interpretation of the R^2 (from question 5) and the results of your hypothesis test about the explanatory power of the model (from question 6), do you think this is a good model of the share of 2003 home purchase loans that are subprime? Why or why not? Write a sentence or two stating your specific reasons. (10 points)

With a relatively high R^2 of .672, we know that the factors accounted for in this model explain over two thirds of the variation in a city's percentage of home purchase loans that are made by subprime lenders. Combined with the result of our F-test, which tells us that our model is significantly better at explaining variation in percentage of loans that are subprime than using the mean share of loans that are subprime, we could reasonably conclude that this is a good model.

One could also have argued that this is not necessarily a good model, since the dependent variable (share of loans that are subprime) may be causing changes in our independent variables (e.g. if an increased presence of subprime lenders in a city were to increase the number of Black or Hispanic homebuyers, by providing more access to credit for Black and Hispanic households than prime lenders). If this were the case, R^2 might be overstating how much variation in the share of loans that are subprime is explained by the variables in our model.

8. The estimated regression coefficients for Proportion of buyers who were Black or Hispanics are not significant at the 10% level. Provide at least two possible explanations for this finding. (10 points)

It is possible that race is not an important factor in explaining the share of loans that are subprime. This would be the case for example if:

- **Blacks/Hispanics had access to the same types of loans than Whites (either because there were no disparities or because there are effective programs in place to address previous disparities)**

But it is also possible that race and subprime loans are correlated and yet the model does not capture it. This would be the case for instance if:

- **other variables in the model, which are correlated with race, are capturing a similar relationship. For example, the proportion of buyers who were Black/Hispanic might be positively correlated with the proportion of residents with a college degree, or negatively correlated with the unemployment rate. If that were the case, and all those variables were included in the model, the coefficients of those other variables would capture some of the relationship between the share of loans that were subprime and the proportion of buyers who were Black/Hispanic due to multicollinearity and they would understate the relationship between proportion Black/Hispanic and share of subprime loans; or,**

- the proportion of buyers who are Black AND Hispanic is relatively large (compared to Black not Hispanic and to Hispanic but not Black), the two variables will be highly correlated and it would be difficult to statistically identify the impact of either one while controlling for the other; or,
 - the relationship is not linear; or,
 - there were very few Blacks/Hispanics taking mortgage loans (because most of them are renting, for instance). In this case, the variable might not have been significant because of the limited number of observations in the sample.
9. What is the share of subprime mortgages in 2003 in the average metropolitan area larger than 500,000 in a population with a relatively high rate of college degrees of 36%, like the Seattle Metro area? What about the average Metro area of similar size that has a lower rate of college degrees in the population, like Grand Rapids, MI, with 25%. (10 points)

We evaluate the equation mean of all the independent variables, except for the proportion of college degrees, which we set at .36 for Seattle, WA metro area and at .25 for the Grant Rapids, MI metro.

For the Seattle Metro, when $x_5=.36$

$$\begin{aligned} \hat{y} &= \hat{\beta}_0 + \hat{\beta}_1 * x_1 + \hat{\beta}_2 * \bar{x}_2 + \hat{\beta}_3 * \bar{x}_3 + \hat{\beta}_4 * \bar{x}_4 + \hat{\beta}_5 * .36 + \hat{\beta}_6 * \bar{x}_6 + \hat{\beta}_7 * \bar{x}_7 + \\ &\hat{\beta}_8 * x_8 + \hat{\beta}_9 * \bar{x}_9 + \hat{\beta}_{10} * \bar{x}_{10} + \hat{\beta}_{11} * \bar{x}_{11} + \hat{\beta}_{12} * \bar{x}_{12} + \hat{\beta}_{13} * \bar{x}_{13} + \hat{\beta}_{14} * \bar{x}_{14} \\ &= 1.107 + (0.594)(0.111)+(-8.32 \times 10^{-10})(.033)+(0.164)(0.06)+(0.086)(0.068)+ \\ &\quad (-0.315)(0.36) + (0.42)(0.207)+(0.057)(0.381)+(-3.186)(0.536) + (2.463)(0.293) + \\ &\quad (0.178)(0.051)+(0.458)(0.015)+(0.121)(0.448)+(0.809)(-0.005)+(1.281)(0.018) \\ &=.208765 \text{ or } 21\% \end{aligned}$$

and for the Grand Rapids, MI metro when $x_5 = .25$:

$$\begin{aligned} \hat{y} &= \hat{\beta}_0 + \hat{\beta}_1 * x_1 + \hat{\beta}_2 * \bar{x}_2 + \hat{\beta}_3 * \bar{x}_3 + \hat{\beta}_4 * \bar{x}_4 + \hat{\beta}_5 * .25 + \hat{\beta}_6 * \bar{x}_6 + \hat{\beta}_7 * \bar{x}_7 + \\ &\hat{\beta}_8 * x_8 + \hat{\beta}_9 * \bar{x}_9 + \hat{\beta}_{10} * \bar{x}_{10} + \hat{\beta}_{11} * \bar{x}_{11} + \hat{\beta}_{12} * \bar{x}_{12} + \hat{\beta}_{13} * \bar{x}_{13} + \hat{\beta}_{14} * \bar{x}_{14} \\ &= 1.107 + (0.594)(0.111)+(-8.32 \times 10^{-10})(.033)+(0.164)(0.06)+(0.086)(0.068)+ \\ &\quad (-0.315)(0.25) + (0.42)(0.207)+(0.057)(0.381)+(-3.186)(0.536) + (2.463)(0.293) + \\ &\quad (0.178)(0.051)+(0.458)(0.015)+(0.121)(0.448)+(0.809)(-0.005)+(1.281)(0.018) \\ &=.242415 \text{ or } 24\% \end{aligned}$$

10. Write a paragraph explaining your results to a non-technical policy audience. In your response, talk about (1) what factors are associated with the concentration of subprime

loans in certain metropolitan areas, (2) one factor you thought might have been important but is not, and (3) the interpretations of the answers you gave in 1-9 above. (20 points)

The results of this study indicate that cities with higher rates of subprime mortgages tend to have lower education levels (as measured by percentages of residents with college degrees), lower incomes relative to loan size, and higher growth in median loan size between 1997 to 2003. Each additional percentage point of residents with a college degree or more is associated with a decrease of 0.315 in the share of 2003 home purchase loans that were made by subprime lenders in U.S. metropolitan areas with populations over 500,000 in 2000.² In fact, a one percentage point increase in residents with a college degree or more in the year 2000 is associated with between a 0.048 and 0.582 decrease in the share of 2003 home purchase loans that were made by subprime lenders.³ 67 percent of the variation in the share of mortgages that are subprime can be explained by characteristics of the market in 1997:

- the proportion of loans made by subprime lenders
 - the median home purchase loan size
 - the proportion buyers who were Black
 - the proportion buyers who were Hispanic
 - the ratio of median loan size to FHA limit
 - the ratio of median buyer income to median loan size
 - the unemployment rate
- plus
- the change in the unemployment rate from 1997 to 2003,
 - the ratio of median purchase price to loan size from 1997-2003,
 - the change in the proportion of buyers who were Black between 1997 & 2003
- and
- the change in the proportion of buyers who were Hispanic between 1997 and 2003.

In total, these factors are better at predicting the share of subprime loans than the average share in 2003.

We would expect that an average metro area of 500,000 or more in population with a relatively high share of college degrees in 1997 (say 36%) would have 21% of its loan market as subprime in 2003, while the average metro area with a lower share of college degrees (25%) in 1997 would have about 24% of its market that were subprime in 2003.

One may have thought that proportion of residents who are Black or Hispanic would have some relationship with the share of subprime loans in a metropolitan area, but it is possible that other factors in the model are decreasing their impact or that the sample size is too small to detect an effect.

² Significant at a .05 level, p-value=.02

³ 95% confidence interval