Week 5 Assignment

Linear Regression

1. Predict 2013 home prices using state information only. Answer these questions using all of the training data available.

a. What is the intercept? What does it correspond to? Ans: 6.766e+04. The mean value of the response variable (2007 home price) when all of the predictor variables(id, zip, state, county, poverty) in the model are equal to zero

b. How do you get this information from your regression? Ans:

housing\_data\_model <- lm(price2007 ~ . , data = housetrain\_data)

c. Based on your regression coefficients, what states have the most and least expensive average homes? Ans: NY & CA were most expensive on average. KS, ND & WY are the lowest

d. How do you get this information from your regression? Ans: It’s clear when you look at the linear model data. You can get these numbers also by tabulating(table) data per-state. The numbers are highest consistently for CA. KS, ND & WY are the lowest and this data is found using the same methods.

e. What is the average price of homes in that state? Ans: CA mean price: 611432.06

f. How do you get this information from your regression?

Ans: agg\_tbl <- dataFrame %>% group\_by(state) %>%

summarise(mean\_price=mean(price2007),

.groups = 'drop')

g. Please upload your R code for the above questions.

2. Predict 2013 home prices from state and county information.

a. What US counties have the highest and lowest regression coefficients? Why?

Highest – Pitkin: 857920.2646 Lowest – calaveras: -352624.3847

The national home prices for 2013 increased with Pitkin being the highest having a downward effect on home prices in Calaveras County. The only correlation I could find was a slight correlation between poverty levels and national home prices; Calaveras being brought to the lowest number.

b. Please upload your R code for the above question