0.1 Question 0: Human Context and Ethics

0.1.1 Question 0a

"How much is a house worth?" Who might be interested in an answer to this question? **Please list at least three different parties (people or organizations) and state whether each one has an interest in seeing the housing price to be high or low.**

Families, real estate agencies, and landlords may be interested in how much a house is worth. Families would want to see the housing price from low to high, whereas real estate agencies and landlords may want to see the housing prices from high to low.

0.1.2 Question 0b

Which of the following scenarios strike you as unfair and why? You can choose more than one. There is no single right answer, but you must explain your reasoning.

A. A homeowner whose home is assessed at a higher price than it would sell for.
B. A homeowner whose home is assessed at a lower price than it would sell for.
C. An assessment process that systematically overvalues inexpensive properties and undervalues expensive properties.
D. An assessment process that systematically undervalues inexpensive properties and overvalues expensive properties.

B is unfair because the homeowner wouldn't be making the money they deserve from selling the house, so they would be losing money. C is unfair because they would be overpricing buyers for inexpensive properties and expensive properties wouldn't be sold for their actual worth so it would harm the sellers.

0.1.3 Question 0d

What were the central problems with the earlier property tax system in Cook County as reported by the Chicago Tribune ? And what were the primary causes of these problems? (Note: in addition to reading the paragraph above you will need to watch the lecture to answer this question)

The earlier property tax system was regressive – lower-income households had to pay higher tax rates. One of the main reasons why is because wealthier homeowners could challenge their property tax rate in front of a review board, which lower-income owners couldn't do. This created a greater disparity in the tax rates for different socioeconomic statuses.

0.1.4 Question 0e

In addition to being regressive, how did the property tax system in Cook County place a disproportionate tax burden on non-white property owners?

The property tax system in Cook County placed a disproportionate tax burden on non-white property owners by including race as a factor to determine property value. Non-white owners were more likely to live in lower-value properties due to this. Redlining is a key example of how this takes place and causes problems of inequity.

0.2 Question 2a

Without running any calculation or code, complete the following statement by filling in the blank with one of the comparators below:

$$\geq$$
$$\leq$$
$$=$$

Suppose we quantify the loss on our linear models using MSE (Mean Squared Error). Consider the training loss of the 1st model and the training loss of the 2nd model. We are guaranteed that:

Training Loss of the 1st Model_____Training Loss of the 2nd Model

$$\geq$$

0.3 Question 3b

You should observe that θ_1 changes from positive to negative when we introduce an additional feature in our 2nd model. Provide a reasoning why this may occur. **Hint:** which feature is more useful is predicting `Log Sale Price`?

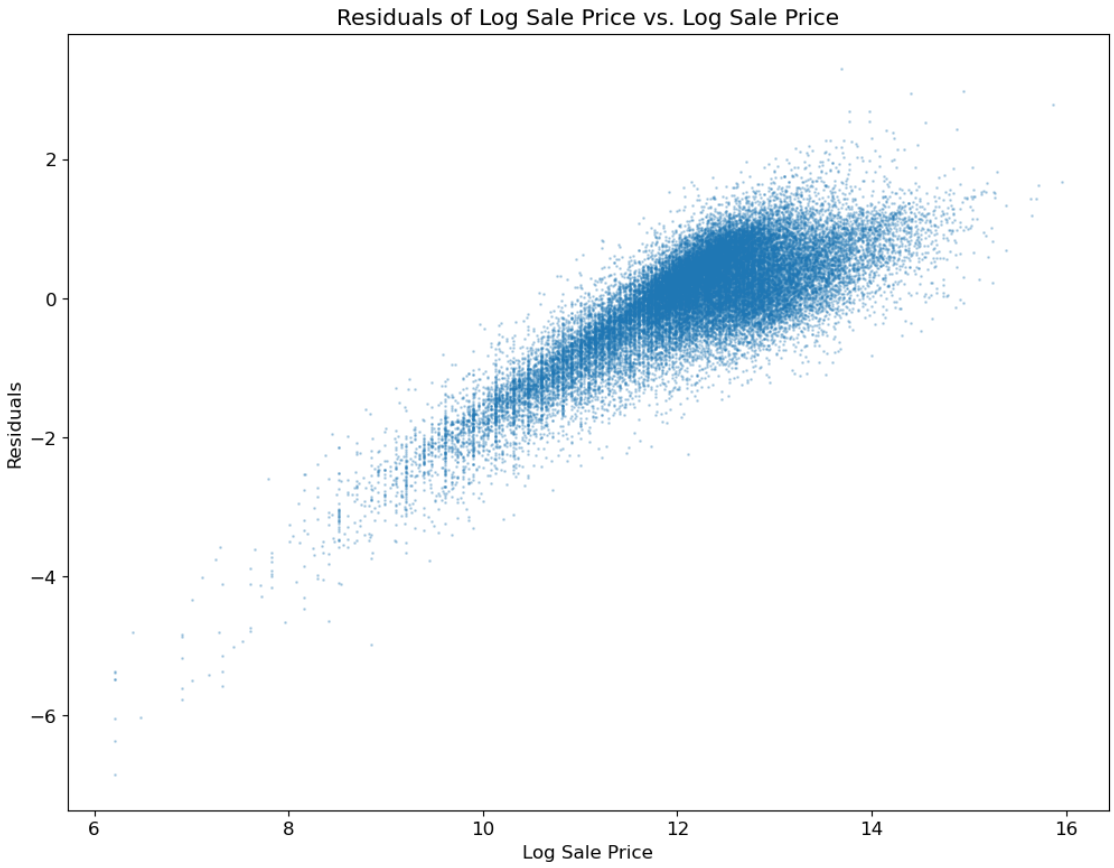
The 2nd model is better fitted because it takes an extra feature into account. As we can see from 3a, the second model has smaller training and validation RMSEs, so it is more useful for predicting Log Sale Price. In both the first and second model, θ_1 is very close to zero, so it probably changed sign because it was more accurate to the data.

0.4 Question 3c

Another way of understanding the performance (and appropriateness) of a model is through a plot of the residuals versus the observations.

In the cell below, use `plt.scatter` to plot the residuals from predicting `Log Sale Price` using **only the 2nd model** against the original `Log Sale Price` for the **validation data**. With a data size this large, it is difficult to avoid overplotting entirely. You should also ensure that the dot size and opacity in the scatter plot are set appropriately to reduce the impact of overplotting as much as possible.

```
In [215]: plt.scatter(y_valid_m2, y_valid_m2-y_predicted_m2, s=1, alpha=0.25)
          plt.xlabel('Log Sale Price')
          plt.ylabel('Residuals')
          plt.title('Residuals of Log Sale Price vs. Log Sale Price');
```



0.5 Question 5

In building your model in question 4, what different models have you tried? What worked and what did not? Brief discuss your modeling process.

Note: We are not looking for a single correct answer. Explain what you did in question 4 and you will get point.

I tested mostly scatterplots and boxplots with several different numerical variables. With features of discrete variables, I used boxplots and with features of continuous variables I used scatterplots. For many of the visualizations, I saw no correlation. However, for the 4 features that I initially chose (Building Square Feet, Estimate (Building), Bedrooms, Garage 1 size), I noticed an association so I decided to keep those variables. After testing with those 4 features, I observed that the RMSE was smaller without Garage 1 Size, so I removed that feature, leaving me with the final 3.

0.6 Question 6 Evaluating Model in Context

0.7 Question 6a

When evaluating your model, we used root mean squared error. In the context of estimating the value of houses, what does residual mean for an individual homeowner? How does it affect them in terms of property taxes? Discuss the cases where residual is positive and negative separately.

Residual represents the difference bewteen how much an individual homeowner is paying and how much they expect to be paying for the property's value. Under the regressiive property tax system, if the residual is posive, this means that homeowners are paying more than they need to, which would cause a decrease in property taxes. If the residual was negative, homeowners would be paying less and that would increase the property tax, which is bad because the property is not being sold for its actual value.

0.8 Question 6b

In your own words, describe how you would define fairness in property assessments and taxes.

Fairness in property assessments and taxes means that more expensive properties should have higher taxes than cheaper properties because people who purchase cheaper properties likely have less money to spend than those purchasing expensive properties. This is essentially what a progressive property tax system is.

0.9 Question 6c

Take a look at the Residential Automated Valuation Model files under the Models subgroup in the CCAO's GitLab. Without directly looking at any code, do you feel that the documentation sufficiently explains how the residential valuation model works? Which part(s) of the documentation might be difficult for nontechnical audiences to understand?

The documentation explains how the residential valuation model works quite thoroughly, however it does not explain the significance of any results. For nontechnical audiences, interpreting the data may be difficult to understand.