0.1 Question 0

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 value be high or low.

Homebuyers would be interested in this question and they would want to see the value be low because they want to buy a home for as cheap as possible. The home owners would also be interested in this question and they would want to see the value be high because they want to sell their homes for as expensive as possible. The real estate agents would also be interested in this question and they would want to see a high value in most cases because they are paid in commissions, and they want to earn more money per house sale. They might want it to be low during times when people aren't willing to pay high prices for houses, so that they can make sales.

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.

All of the choices would be unfair. A. It would be unfair for home buyers to be pressured to pay more for the house than what it would sell for because the house was assessed at a higher price than it would sell for. B. It would be unfair for homeowner because when they are selling their houses, people would use the home assesd price in order to bargain with them to lower the price of the house than what they would have paid for. This could potentially lose home buyers who were initially willing to pay a higher price, either not buying the house or paying a much lower price due to the house being assessed at a lower price than it would sell for. C. It would be unfair for the people with expensive properties because their houses would be undervalued and they might have to sell for a lower price than it's accurate value. D. It would be unfair for the people with inexpensive properties because their houses would be undervalued and they might have to sell for a lower price than it's accurate value.

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)

"Racially discriminatory assessments and taxes" were the central problems with the earlier property tax system in Cook County. The system assessed by Joseph Berrios was undervaluing high-priced homes and overvaluing low-priced homes, which shows an underlying racism of the assessor because this system allowed wealthy, white homeowners to pay less in property taxes while less wealthy, non-white homeowners paid more proportionally. In the lecture, this was discussed where prolems like this arise because of human creating bias in the system, which then leads to discrimination in assessments and taxes for minorities. Similar example of this could be Robert Moses'racist bridge that was purposely built low to the ground, so that people of color, who usually transit in buses had to take long ways around.

0.1.4 Question 0e

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

Racism. One possible scenario could have been that wealthy, white property owners had lobbyed or bribed the government workers for their benefits, but it still comes down to the assessor Joseph Berrios racism that targetted non-white property owners to pay disproportionate tax.

0.2 Question 2

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 6

Let's compare the actual parameters (θ_0 and θ_1) from both of our models. As a quick reminder,

for the 1st model,

$$\text{Log Sale Price} = \theta_0 + \theta_1 \cdot (\text{Bedrooms})$$

for the 2nd model,

$$\text{Log Sale Price} = \theta_0 + \theta_1 \cdot (\text{Bedrooms}) + \theta_2 \cdot (\text{Log Building Square Feet})$$

Run the following cell and compare the values of θ_1 from both models. Why does θ_1 change from positive to negative when we introduce an additional feature in our 2nd model?

θ_1 changes from positive to negative because as θ_2 increases, θ_1 decreases as a result of us multiplying by the log sale price, which then returns a negative value when adding both θ_1 and θ_2 together.

In [22]:
```
# Parameters from 1st model
theta0_m1 = linear_model_m1.intercept_
theta1_m1 = linear_model_m1.coef_[0]

# Parameters from 2nd model
theta0_m2 = linear_model_m2.intercept_
theta1_m2, theta2_m2 = linear_model_m2.coef_

print("1st Model\n 0: {}\n 1: {}".format(theta0_m1, theta1_m1))
print("2nd Model\n 0: {}\n 1: {}\n 2: {}".format(theta0_m2, theta1_m2, theta2_m2))
```

```
1st Model
 0: 10.571725401040084
 1: 0.4969197463141442
2nd Model
 0: 1.9339633173823696
 1: -0.030647249803554506
 2: 1.4170991378689644
```

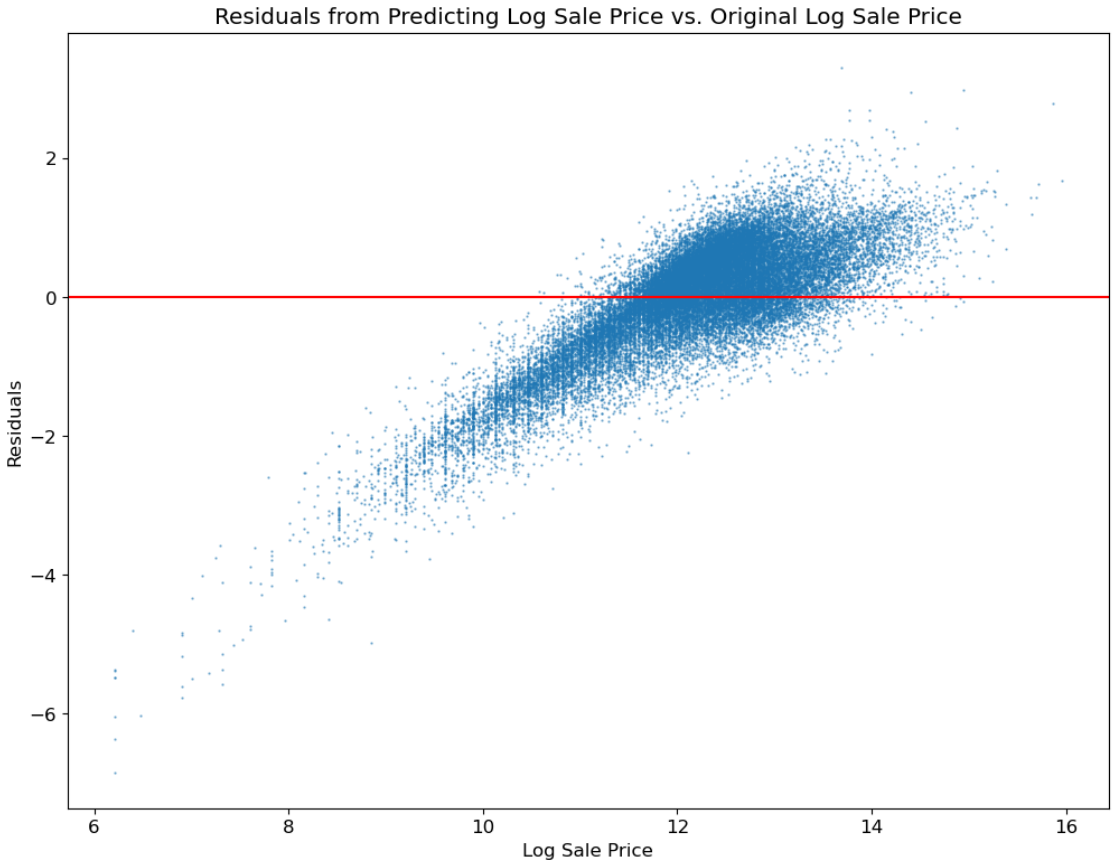
0.4 Question 7

0.4.1 Question 7a

Another way of understanding the performance (and appropriateness) of a model is through a plot of the model 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 **test data**. You should also ensure that the dot size and opacity in the scatter plot are set appropriately to reduce the impact of overplotting.

```
In [29]: plt.scatter(x = y_test_m2, y = y_test_m2 - y_predicted_m2, s = 0.5, alpha = 0.5)
         plt.axhline(y = 0, color = 'red')
         plt.title('Residuals from Predicting Log Sale Price vs. Original Log Sale Price')
         plt.xlabel('Log Sale Price')
         plt.ylabel('Residuals');
```



0.5 Question 9

In building your model in question 8, what different models have you tried? What worked and what did not? Brief discuss your modeling process.

Note: We are looking for a single correct answer. Explain what you did in question 8 and you will get point.

I tried one hot encoding it, but I was too dumb to figure it out, so I had to brute force as many features in as I can in order to get my RMSE down. I picked features I thought would distinguish social classes to help the machine testing.

0.6 Question 10

When evaluating your model, we used root mean squared error. In the context of estimating the value of houses, what does error mean for an individual homeowner? How does it affect them in terms of property taxes?

The error means that the houses were undervalued or overvalued, which means that individual homeowner would be unfairly taxed for their property. This means that for homeowners whose house has been undervalued, they would be paying less than what they should be for property taxes, and for houseowners whose house has been overvalued, they would be paying more than what they should be for property taxes.

In the case of the Cook County Assessor's Office, Chief Data Officer Rob Ross states that fair property tax rates are contingent on whether property values are assessed accurately - that they're valued at what they're worth, relative to properties with similar characteristics. This implies that having a more accurate model results in fairer assessments. The goal of the property assessment process for the CCAO, then, is to be as accurate as possible.

When the use of algorithms and statistical modeling has real-world consequences, we often refer to the idea of fairness as a measurement of how socially responsible our work is. But fairness is incredibly multifaceted: Is a fair model one that minimizes loss - one that generates accurate results? Is it one that utilizes "unbiased" data? Or is fairness a broader goal that takes historical contexts into account?

These approaches to fairness are not mutually exclusive. If we look beyond error functions and technical measures of accuracy, we'd not only consider *individual* cases of fairness, but also what fairness - and justice - means to marginalized communities on a broader scale. We'd ask: What does it mean when homes in predominantly Black and Hispanic communities in Cook County are consistently overvalued, resulting in proportionally higher property taxes? When the white neighborhoods in Cook County are consistently undervalued, resulting in proportionally lower property taxes?

Having "accurate" predictions doesn't necessarily address larger historical trends and inequities, and fairness in property assessments in taxes works beyond the CCAO's valuation model. Disassociating accurate predictions from a fair system is vital to approaching justice at multiple levels. Take Evanston, IL - a suburb in Cook County - as an example of housing equity beyond just improving a property valuation model: Their City Council members recently approved reparations for African American residents.

0.7 Question 11

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

I would define fairness in property assessments and taxes as a property assessment that assesses sole based on features that affect the values of those homes, without any biases that favors or unfavors certain groups of people, whether that be social classes or racial classes. Therefore, people whose houses have been overvalued would not be paying unfair property taxes, and people whose houses have been undervalued would not be paying inaccurate property taxes. Also, there could be another case where it would not be fair for assessments to be unfair and target specific groups of people - when selling houses, those whose houses have been undervalued would have much harder time selling their houses for a fair price, if the assessment was biased and inaccurate, and for those whose houses have been overvalued, it would be unfair for the buyers to be pressured to purchase a house based on an inaccurate and biased overvalued assessment of the house.

0.8 Question 12

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 seems like it sufficiently explains how the residential valuation model works, for those who have technical understanding. Even those who are nontechnical could understand the files if they read through the whole thing and did individual research. For examples, the file not only states all the testing methods that they are using, such as out-of-time-testing, rolling-origin resampling, etc., but also explains what each methods are for those who do not know what they are. I believe, still, it is difficult for nontechnical audiences to understand because the explanations are very centered around professional vocabulary and explanations, but I'm not sure what they could have done differently that could convey a clear explanation to those with technical understanding that also would be easy for nontechnical audiences to understand.