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.

1. Buyers: This group of people would like to see that the value is low so that they can purchase a house for cheaper.
2. Sellers: This group of people would like to see higher prices so that they can sell their house for a higher price.
3. Homeowners: This group of people would like to see lower values so that they pay less taxes on their property.

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 scenarios are unfair as there is some imbalance on one side or the other. For a home that is assessed higher than sale price is hurts the owner who should not have to pay as much tax since their house's price is overinflated. For a homeowner whose house is lower than sale price it means that if they want to sell they would not recieve the fair market compensation. In the case that overvalues inexpensive properties and undervalues expensive properties you get people who now pay more taxes than they should (inexpensive properties) and others would pay less taxes. Finally, in the inverse of the previous situation richer houses now pay more taxes than they should and cheaper houses now sell for less which cheats them or fair compensation when they sell the property.

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)

According to the Chicago Tribune a major problem was that black households were often being overvalued which put them at risk of being kicked out of their own homes because they would not be able to afford the property taxes.This was caused by a combination of a complicated tax system which made the use of a lawyer neccessary along with methods that overvalued all the properties. These in combination made it unfair as wealthier people who were often white could pay lawyers to argue about their property value and reduce it which often led them to paying less taxes compared to black counterparts.

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?

White property owners were able to afford lawyers which could argue for lower property values which in turn made it so they had to pay less taxes. Non-white owners who tend to be lower income could not afford the same services so the property value given to them, which was often inflated, was the one they had to pay taxes on.

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?

```
In [23]: # 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
```

We can assume that it changes signs because the weighting of the variable is not as impactful as our other variable that we have introduced and infact detracts from the total cost as the number of bedrooms increase.

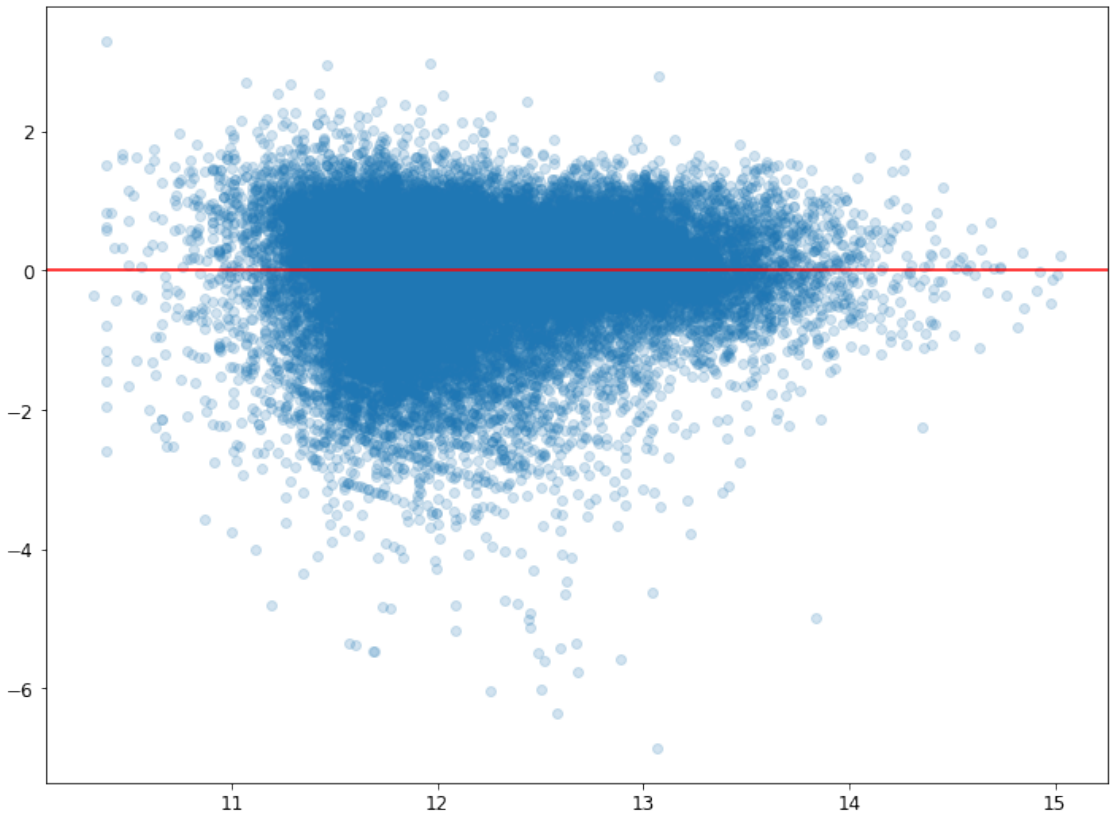
0.4 Question 7

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 [24]:
```
plt.scatter(y_predicted_m2, y_test_m2 - y_predicted_m2, alpha=0.2)
plt.axhline(y = 0, color='r');
```



0.5 Question 9

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 lower the rmse the closer we were to an appropriate value of that persons house. If the rmse is high then we did not properly estimate the house cost according to the actual data and if our estimates were implemented that would mean that they would not pay a fair tax rate either because we undervalued their homes or overvalued them by a significant margin.

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.

0.6 Question 10

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

In a simple definition fairness in property assessments and taxes would mean that people's houses are assessed as close to the current market value of their house and as a result they would pay an amount of property taxes based on a percentage of that houses fair value. In addition fair assessment would also include fair recourse where the barrier to fix any mistakes is as low as possible so that lower income people are not cheated on the value of their property, this would probably look like a simpler process to report and problems instead of heavily relying on lawyers. Finally, it is important that people who are affected by poor property assesments are compensated and could possibly be in the form of a tax break based on the mistaken calculation plus some interest.

0.7 Question 11

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 first thing that I would look for as a nontechnical person would be how they calculate the housing prices, which is not something that is immediatly apparent in the documentation and feels hidden in the standard operating procedures(SOP). When looking in SOP there is a lot of technical equations that most nontechnical people would probably not understand making it less accessable. However, if you have a little bit of background working with technical things they go into great detail and are deliberate with how they define things. Overall, the documentation seems more skewed towards people with more technical backgrounds as it would require a lot of dedication for a single person to try to understand the language and equations they use from scratch.