

## 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.

Potential Buyer: interested in seeing the value to be low. Landlord: interested in seeing the value to be high  
Housing Agent: interested in seeing the value to be high



### 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.
- D. In this situation, the range of price of market is enlarged, and it is unfair that it makes rich people even richer, and poor people even poorer.



### 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)

*Type your answer here, replacing this text.*



#### 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?

*Type your answer here, replacing this text.*





## 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 ( $\theta_0$  and  $\theta_1$ ) 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  $\theta_1$  from both models. Why does  $\theta_1$  change from positive to negative when we introduce an additional feature in our 2nd model?

```
In [41]: # 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
```

In the 1st model, the number of bedrooms are indicating the building square feet indirectly; In 2nd model, since we introduced the parameter of Log Building Square Feet, the correlation of Price and Building Square Feet is not represented by number of bedrooms anymore.



## 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 [42]: fig = plt.figure(figsize=(8,8))
         plt.scatter(y_test_m2,y_predicted_m2,s=5,alpha=0.5)
         plt.xlabel('Original Log Sale Price')
         plt.ylabel('Predicted Log Sale Price')
         plt.title('Original Log Sale Price vs Predicted Log Sale Price')
```

```
Out[42]: Text(0.5, 1.0, 'Original Log Sale Price vs Predicted Log Sale Price')
```



## 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 error means the difference of the real price of their house and the estimated price given by authorities. If their house is overestimated in price, they are going to pay more tax, and vice versa.





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.

A fair model should accurately estimate the value of property, and unbiased from irrelevant factors like the race of homeowner. With well-considered consideration about the features which are directly relevant to the price of property, and discarding those factors that are not.



## 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 documentation includes many technical terms like L1/L2 regularization, K-folds.... Without prior knowledge about data science, audience may not understand what they are, and thus causing obstacles for understanding the mechanism of model.

