

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.**

The selling party would probably want to price the house high. The buying party would want the price to be low. The last party I can think of is the realtor which would also want it to be low because they are with the buying party. I guess the city government could also be on and would want the price to be high if the property tax is based on the value of the house.

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.

I feel like they are all unfair because a property should be assessed at the value it really is. If it is overvalued then that it unfair for those who own the house. This might cost to them to pay more on property taxes. It is also true for undervalued properties. Where it is really unfair I think is with the last two. If something is systematically assessing the value of properites wrong then it should be corrected to not be unfair toward a certain group.

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 system would overvalue lower income houses and undervalue higher income houses. There was an appeal system, but many of the lower income owners were not able to use it because they did not have the time and resources to go appeal the appraisal. The system was also built on the previous systems that used redlining or race as indicators of a property's value.

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 system was influenced by a system developed by real estate professionals which used race as a factor of valuation. In addition many non-white property owners did not have the time and resources to appeal their appraisal.

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 change 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?

I think that the change is because the Log Building Square Feet is more useful than bedrooms when predicting Log Sale Price. I think that the reason this feature is better for predicting Log Sale Price because they are both the predicting variable and feature are in log versus in nominal terms.

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 [25]: plt.scatter(y= y_valid_m2 - y_predicted_m2, x= y_valid_m2, alpha= .5)
         plt.ylabel('Log Sale Price Residuals')
         plt.xlabel('Log Sale Price')
         plt.title("Log Sale Price Residuals vs. Log Sale Price")
```

```
Out[25]: Text(0.5, 1.0, 'Log Sale Price Residuals 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 looked through every variable to find a relationship with log sale price. Some were very apparent, some not so much. I plotted log sale price with transformed variables. I used the rmse that the autograder gave us. I also tried to make my own and I found that as soon as I use 5 features the error for both the holdout and training sets goes to 0.9. Sometimes the holdout is lower than the training which confused me. I tried to do k fold cross validation I am not sure if I did it right. I tried to check the cv error and adjusting the outlier parameters after looking at the error. In the end I wasn't able to get the prediction on gradscope under the 337k or something like that. Nothing I tried really worked, I am not sure if I implemented the stuff correctly though.

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.

In this context residual is the difference between the actual value of the house and the predicted value of the house. If counties are using this predicted value then the residual should be 0 or very close to 0. If it is not then this causes a lot of problems for homeowners. If their house was over valued then that means that they will pay more in property taxes when they shouldn't. This could also go the other way. If a house is undervalued then the homeowner will pay less in property taxes than they should.

0.8 Question 6b

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

I would say that fairness is being assessed accurately. The assessment should be as close as it can be to the actual value of the house. When it comes to property taxes, I think that homeowner should pay a tax that depends on the value of their house. This is why having an accurate assessment is important. Wealthier homeowners should not be paying less in taxes than less wealthier homeowners. That is not right or fair. There should also not be any discrimination of any kind that affects the assessed value of the house.

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?

Yes, I think that this is documentation explains how it works and their whole process of developing the best model. In addition, they also add a section on ongoing issues, which I think gives them more credibility and allows us to see how maybe they can improve it if these issues are fixed. I think nontechnical audiences will find the modeling section, model training, and choices made sections a bit hard to understand. I feel like these do go in depth, but I feel like you need to have seen a bit of it before to fully understand it.

