Congratulations! You passed!

Next Item

1 / 1 point

1. Suppose you are conducting an inferential analysis and are using a formal model to specify your approach. What are the parameters in a formal model?

- They are summary statistics that are derived from the dataset.
- They are numerical quantities that represent features of the population.

Correct

- They are numerical quantities that represent features of the dataset.

1 / 1 point

2. Your team is interested in studying the relationship between the price of a house in a city and the number of square feet in the house. To do this, you manage to obtain a representative sample of data from 200 homes in the city and fit a linear regression model with home price as the outcome and square footage as the key predictor.

What is the parameter that you are trying to estimate in this problem?

- The expected change in home price associated with a change in square footage across all homes in the city.

Correct

- The average home price in that city.
- The correlation between home price and square footage in the dataset.
- The average square footage of a home in that city.

1 / 1 point

3. You and your team are interested in predicting home prices in a city based on a variety of features of the house (e.g. square footage, acreage, number of bedrooms, etc.). One of your analyst suggests this may be difficult because we don't understand very well the causal relationship between house features and the price of the house. He asks whether the team should proceed with the analysis.

How might you respond to this?

- The analysis can proceed because in a prediction analysis, we care less about the exact mechanisms underlying the relationships between the features and the outcome, as long as we can build a model with good prediction skill.

Correct

- The analysis can move forward after we do an associational analysis of square footage and house price.
- The analysis cannot go forward until we develop a better understanding of the causal relationships between the house features and the house price.

1 / 1 point

4. You and your team are tasked with helping to determine whether an incremental new feature should be added to an existing product. The question is whether the adding the new feature will increase the number of units sold of the product. The data you have available are from *other* products the company sells that include this same feature.

Because your company is a global company, it is able to release new features to different parts of the world at different times. You determine that a simple analysis that could be done would be to look across all the company's markets and compare units sold between those markets that have the new feature and those that don't adjusting for potentially confounding differences between the various markets.

An analyst on your team builds a prediction model from the data and notes to you that the inclusion of the new feature only explains about 2% of the variation in the number of units sold (i.e. it is a weak predictor) and that there are many other factors that better predict the number of units sold.

What conclusion can you draw from this?

- It is not clear yet what to conclude from this analysis because a prediction analysis was done rather than an associational analysis.

Correct
An associational analysis is need to assess the relationship between the inclusion of the feature and the outcome. For example, at this point, we do not even know if the relationship between the feature and the outcome is positive or negative.

- The new feature likely does not increase unit sales because it is not a strong predictor of unit sales in the model.
- The new feature will increase unit sales because 2% is actually a substantial amount of variation to explain (i.e. the analyst is incorrect in her interpretation).

1 / 1 point

5. You and your team are tasked with helping to determine whether an incremental new feature should be added to an existing product. The question is whether the adding the new feature will increase the number of units sold of the product. The data you have available are from *other* products the company sells that include this same feature.

Because your company is a global company, it is able to release new features to different parts of the world at different times. You determine that a simple analysis that could be done would be to look across all the company's markets and compare units sold between those markets that have the new feature and those that don't adjusting for potentially confounding differences between the various markets.

An analyst on your team conducts an associational analysis and fits a number of different

models to the data. Across the primary and secondary models, she notes that the estimate of the change in the number of units sold associated with the inclusion of the new feature ranges from an increase of 20% to an increase of 21%, which everyone agrees is a narrow range, and the uncertainty associated with those estimates is roughly the same across the models.

Which model should be chosen as the final model, to be used in future stages of the data analysis?

- [x] The model with the fewest parameters, because parsimony usually allows for a simpler explanation of what is going on in the data or the population.

Correct

- [ ] The model with the largest number of parameters because it will have better predictive skill.
- [ ] The model that produces an estimate of the association that is in the middle of the range of estimates.
- [ ] The model with the largest estimated association because the company prefers a larger increase in sales.