11. In the Boston Housing data set, what is the relationship between crime and housing prices? Please support your claims with exploratory analysis conducted in R. Does this relationship make sense? Justify your answer. IE: What are some reasons this relationship makes sense or does not make sense?

If we use a linear model , we can infer from the summary of model that crime rate (‘crim’) in the area is significant with a lower p-value and a negative coefficient .This indicates that the housing price is higher in areas where crime rates are lower .This relationship makes a good point since people prefer to stay in places where there is lesser crime rates.

12. Based on your analysis of the Boston Housing data set, please provide an interpretation for the top 3 strongest absolute correlations. Offer some hypothesis as to why these correlations may be present.

These are the top 3 strongest absolute correlations are

i) rad vs tax= 0.9102282

ii) nox vs dis = - 0.7692301

iii) indus vs nox= 0.7636514

Tax VS index of accessibility to radial highways and full-value property-tax rate are highly correlated as they may indicate the same areas. Tax will be high in the areas where there is closer accessibility to highways.

nitrogen oxides concentration Vs weighted distances to five Boston employment centers are strongly negatively correlated. Lesser the distance between the employment centers more the concentration of nitrogen oxide.

nitric oxides concentration VS proportion of non-retail business acres per town are strongly

correlated . Nitric oxide levels are higher with increase in non-business acres per town.

13. Based on your analysis of the 90th%, 92.5th%, 95th%, 97.5th% and 99th% confidence intervals for the mean of passing yards, you should have noticed that the bounds of the confidence intervals increase as the percentile (90th%, 92.th%, etc) increases. In your own words, explain why this relationship exists.

Yes. The bounds of confidence intervals increases with the percentile increase, this is needed to provide the width (difference between the confidence interval) increase, the more confidence we need, more will be the width, so that the actual value could be captured.

14. Based on your anova of the food calorie counts, please state the null and alternative hypothesis and interpret the results of your anova.

Null Hypothesis - Food types will not affect the calorie count.

Alternative hypotheses – Atleast one of the food type will affect the calorie count.

Results of anova -says that we can reject null hypotheses in favor of alternate hypotheses since the p-values is lower than 0.05 for 95 % Confidence interval.

Karthik Gunasekaran

A20435210