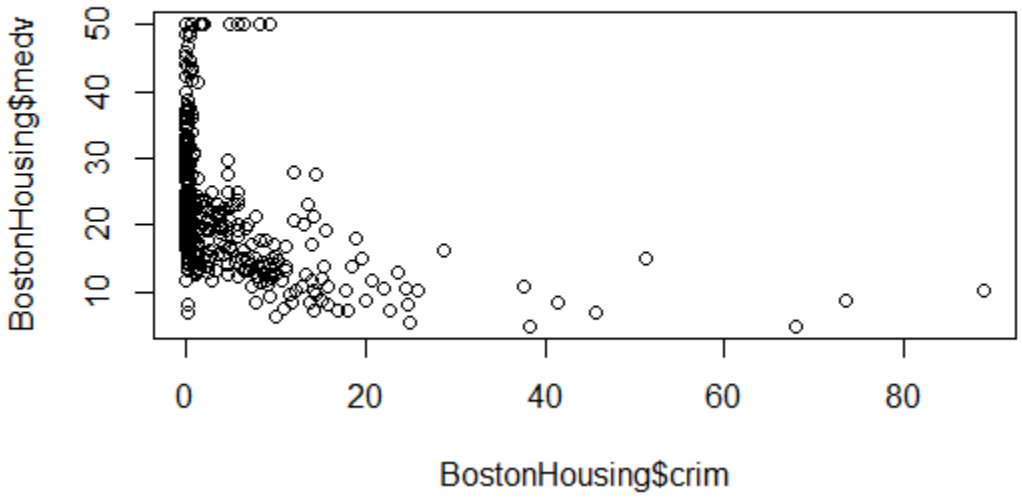
Arinjay Jain(**A20447307**)

CSP/MATH 571 Data Preparation and Analysis **Instructor: Adam McElhinney**
Homework 2, part 2 of 2

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 an exponential model, we can infer that the median value (medv) is higher in areas where the crime rate (crim) is lower. The graph above typically indicates a decay or disassociation exponential model. This relationship makes sense as people want to live in an area where the rate of crime is low.

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.

The top 3 strongest absolute correlations

1. rad vs tax= 0.9102282
2. nox vs dis = - 0.77
3. indus vs nox= 0.76

Arinjay Jain(**A20447307**)

1. rad vs tax= 0.9102282

Rad: index of accessibility to radial highways

Tax: full-value property-tax rate per USD 10,000

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

2. nox vs dis = - 0.77

Nox: nitric oxides concentration (parts per 110 million)

Dis: weighted distances to five Boston employment centres

Nitrogen oxides concentration Vs weighted distances to five Boston employment centers is strongly negatively correlated. Lesser the distance between the Boston employment centers more the concentration of nitrogen oxide. That means if we have more employment in the same area (Lesser distance between them) then we have a high chance of Air pollution due to a high percentage of Nitrogen oxide.

3. indus vs nox= 0.76

Indus: proportion of non-retail business acres per town

Nox: nitric oxides concentration (parts per 110 million)

Nitrogen oxides concentration VS proportion of non-retail business acres per town are strongly correlated. Nitric oxide levels are higher with an increase in non-business acres per town. That means if we have more big Industries in the same area then we have a high chance of Air pollution due to the high percentage of Nitrogen oxide.

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.

Ans. Yes, this is true confidence intervals width is increasing with increasing percentile value. If we increase the percentile that means we want to analyse more data from the population. If you want to analyse more data that means we want wider width of confidence intervals (more distance between the confidence interval) so that we can get more actual data and lesser chance to reject the Null hypothesis.

Arinjay Jain(**A20447307**)

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- Any Food types (Food1, Food2, Food3, and Food4) will not affect the calorie count.

Alternative hypothesis- At least one of the food types will affect the calorie count.

Result: from ANOVA test we can say our P-value (Pr (>F)) is smaller than alpha (0.05) for 95 % Confidence interval. So we can reject null hypotheses in favor of alternative hypotheses.