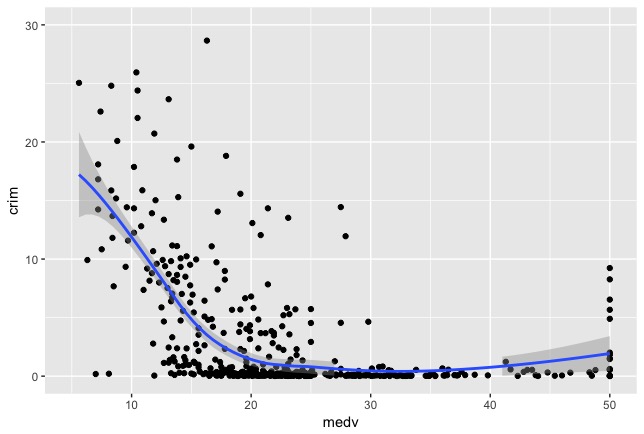
Student: Mikhail Rybalchenko

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?



There is a negative correlation between crime rate in town and median value of housing prices up to $20 thousands. For the houses of value $20k and more, crime rate doesn’t have relationship with housing prices. That make sense, cost of real estate will definitely go down in the areas known for its crime. Safety is one of the questions considered by people looking for a new house, and they will choose more safer areas. As a result, real estate in the areas with higher crime rate will have less demand and lower prices.

On the other hand, there is some level of crime that can’t be assessed without statistics, and in our case it’s appeared to be for the houses valued more than $20K.

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 strongest positive correlation (0.91) is between index of accessibility to radial highways (rad) and full-value property-tax rate per $10,000 (tax). The City of Boston and towns in Massachusetts operate under a property tax classification system which applies different tax rates for residential and commercial properties (<https://www.boston.gov/departments/assessing/how-we-tax-your-property>)

Given that proportion of non-retail business acres per town has a positive relationship with index of accessibility to radial highways (rad) and proportion of residential land zoned for lots over 25,000 sq.ft. has negative correlation, I assume, that strong correlation between rad and tax is a result of lower ratio of residential areas and higher ratio of business areas in the proximity to radial highways, which is reflected in the higher tax rate in the area.

* The negative correlation between ‘nox’ (nitrogen oxides concentration (parts per 10 million) and ‘dis’ (weighted mean of distances to five Boston employment centers

) weighted mean of distances to five Boston employment centers can be a result of manufacturing activities located in the employment centers, which “contribute” to the air pollution in the area.

* The ‘indus’ (proportion of non-retail business acres per town

) and ‘nox’ (nitrogen oxides concentration (parts per 10 million)) positive correlation is also an indicator of manufacturing activities in non-retail business areas.

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.

Graphically, higher confidence level, will increase area under the distribution and give wider confidence interval.

Mathematically, confidence interval can be adjusted by tuning either Z-value (for Normal distribution) or standard error. Z\*SE gives the confidence interval, so with the higher confidence level, Z increases, and we have a wider confidence interval.

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

The null hypothesis H0: Means of calories of all four food types are equal.

Alternative hypothesis H1: Means are not equal.

p-value = 0.0068 which is less than alpha=0.05 (chosen for this test) that means we should reject the null hypothesis H0, which means that for the given data, means are not equal and not coming from the same population.