Jason Melnik

10/27/2021

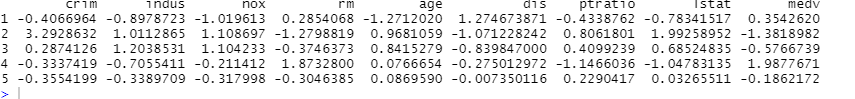
Homework 7

1. The mean is 12.62 and the standard deviation is 7.16 which means we would expect most of the value to be between the ranges 5.46 to 19.78.
   1. 128 24 146 50 147
   2. The graph below shows the scatterplots of NOX, RM, and LSTAT. The communities are as follows with 1 = Black, 2 = Red, 3 = Green, 4 = Cyan, 5 = Blue. This helps us see the amount of communities in each cluster in order from the answer before. Since NOX is nitrogen oxides concentration, RM is average number of rooms per dwelling, and LSTAT is percent of population below the poverty line we can see how each cluster is made. For green the cluster is made for communities that have high NOX, average RM, and average to high LSTAT. The rest can be calculated as well based of the colors and values correlated with them..

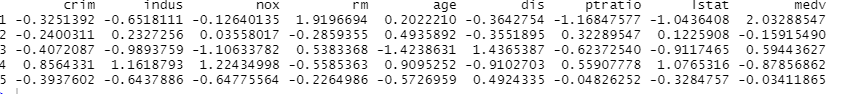
A picture containing map

Description automatically generated

* 1. Cluster Analyzing. The below shows each clusters highest value and lowest value. We can tell from these clusers that cluster 4 is rich communities while cluster 2 is the poorest communities.
     1. Cluster 1 has large value for DIS(1.27) and a low value for AGE(-1.27)
     2. Cluster 2 has a large value for CRIM(3.29) and a low value for MEDV(-1.38)
     3. Cluster 3 has a large value for INDUS(1.20) and a low value for CRIM(-0.84)
     4. Cluster 4 has a large value for MEDV(1.99) and a low value for PTRATIO(-1.15)
     5. Cluster 5 largest value is PTRATIO(0.23) with lowest value CRIM(-0.36)



1. Re-run 2 more times. The below re-runs show that the clusters seem to stay consistent but switch numbers with one another. There is always a rich and poor neighborhood with the stats of having a high MEDV, RM and a low LSTAT, NOX, and CRIM.
   1. The below shows similar results with the rich neighborhood being cluster 1 and the poor neighborhood being cluster 4.



* 1. The below shows similar results with the rich neighborhood being cluster 5 and the poor neighborhood being cluster 2.

A picture containing calendar

Description automatically generated

1. I would choose 6 since its closes as the elbow of the graph. Since 5 is to high and 8 is too low with 7 going up a little. 6 is closes to the desired value for being the k value of this data set.

Chart, scatter chart

Description automatically generated