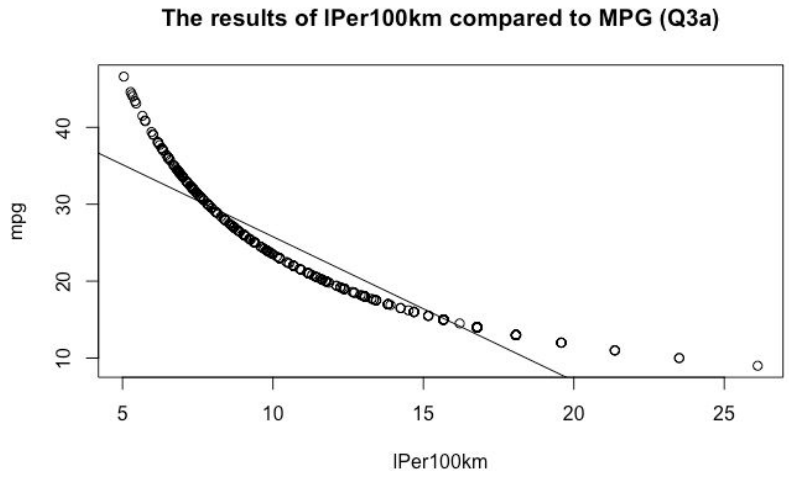
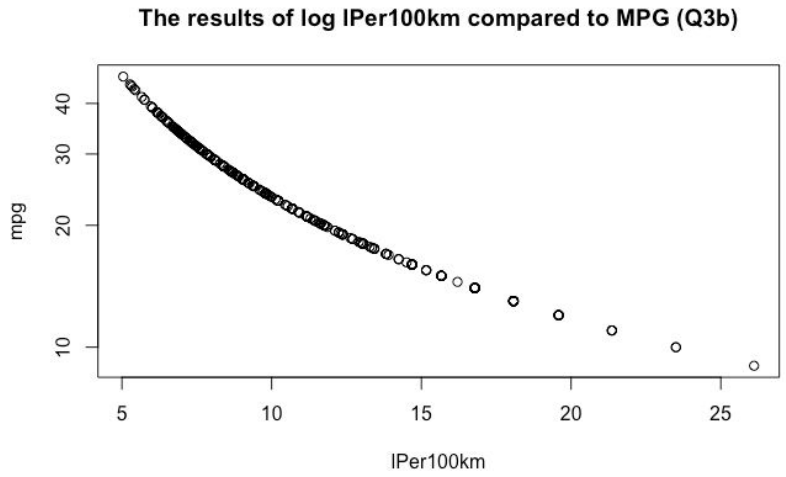
Regression & Association Rules

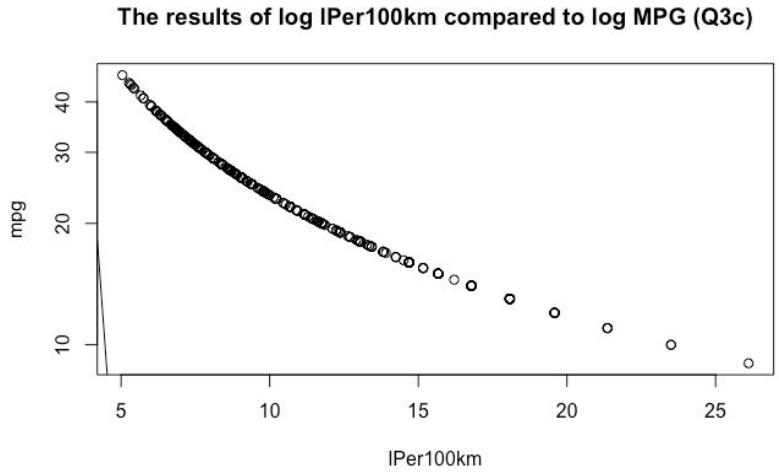
1. I used linear regression to predict how the weight will affect the miles per gallon. The slope in my graph comparing mpg to weight, corresponds to the predictor variable, weight, by showing a downwards slope. The interpretation of this means as the weight of the vehicle gets larger the mpg gets lower. If I were to explain the significance of this finding to someone, I would suggest to them that if they were to purchase a larger vehicle then they may have to consider spending more money on gas, since the miles per gallon, or in other words, distance that they would get would potentially be lower than when compared to a lighter weighted vehicle.

2. When you change to the logarithm of mileage and compare the results to the previous results we see a difference in the slope. The slope in the previous graph is a steady line Y = -0.0076766*X + 46.3173644 however this logarithm of mileage graph does not show a slope because Y = -3.512e-04*X + 4.145e+00 which won't be shown on the graph.

3.







4. The results of the logarithm of lp100km with the mpg is perfectly linear. Meaning there is no difference between x and y, lPer100km and mpg, making the slope -22.64 and intercept to 76.90.

Missing Items & Association Rules

5.

a. The support of the transaction would be 1/4 of A=>B, as there is only one transaction that is the value 1, for both. The confidence is 1/2 of A=>B since there are only two instances of A and only one of A => B. (1/2)/(2/4) is the lift of A=>B meaning that lift is 1 and therefore A and B are

two independent occurrences and we cannot say that A influences the appearance of B in any way.

b. The support of the transaction in this scenario would be 1 / 4 of A => B as the total is still n=4. The confidence in this case is 1 / 3 of A => B since there are three cases of A. The lift is (1 / 3) / (2 / 4) is 0.66 meaning that there's a chance that A influences the appearance of B in that if A is present then B most likely will **not** be present with it.

c. The support of the transaction is this scenario would be 1 / 5 of A => B as the total is n=5. The confidence 1 / 3 of A => B since there are three cases of A. The lift is
lift: (1 / 3) / (2 / 5) is 0.83 meaning that A does influence the appearance of B in that if A does appear there isn't a chance that B will appear still.

Congressional Voting

6. The missing values from the UCI webpage are due to "voted present"", "voted present to avoid conflict of interest", and "did not vote or otherwise make a position known". This can happen for example if the voter may not have been informed fully about the topic or they may have a conflict of interest (n.a., 2019).

In this dataset an missing value is given a '?' as a value in the table. Azevedo et al. state that there are two ways to deal with missing data, one way is to delete incomplete observations entirely. Another way is to replace any missing values with an estimated guessed value based on previous data. However both these ways affect the conclusions that can be gathered from the data.

In my opinion on what should be done in this case, since each vote is tied to an individual person, I believe it is best to leave the ? as a third variable. We cannot give a value to that specific person, because it would indicate we have voted for them without their approval. However we cannot delete this person from the table as they have voted for other categories and that would skew the data.

7. Looks like there was a pattern found in the V7 columns that when there was a number of voters
who voted 'no', they would also vote 'no' for column V6.
Voters who also voted 'no' on V7 would tend to vote 'no' on V14
Voters who voted 'no' on V7 would vote 'yes' on V10.
Voters who voted 'yes' on V12 would tend to be democrats.

If we change the confidence threshold to .75 we see V7 'no' voters tend to vote 'no' on V15 instead of V6 but still V14. We still have voters who voted 'yes' on V12 would tend to be democrats. V12 'yes' voters would vote 'no' on V5.

If we change the confidence threshold to .85 we see voters vote 'no' on V7 also vote 'no' on V6 and V13 but 'yes' on V8. We still have voters who voted 'yes' on V12 would tend to be democrats.

It seems like the association rule of who voted 'yes' on V12 tend to be democrats has always been assumed in the various confidence levels.

Bibliography

Abstention. (2019, December 25). Retrieved March 7, 2020, from https://en.wikipedia.org/wiki/Abstention

Azevedo, C., Proença, H., Salgado, C. M., & Vieira, S. M. (2016, September 10). Missing Data. Retrieved March 7, 2020, from https://link.springer.com/chapter/10.1007/978-3-319-43742-2_13