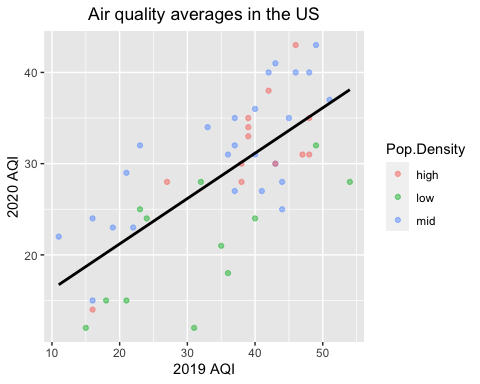
Final Project

Daniel Bostwick

5/2/2021

Did the closure of non-essential businesses affect the change in air quality during the year 2020? If so why? Were there any factors that led to these results?

#### Linear Regression model on AQI from 2019 to 2020

* I’m interested in seeing some correlation between the average air quality from 2019 and the average air quality from 2020.
* There is a positive correlation between the AQI’s. Correlation coefficient equal to 0.702. 

#### Linear Regression Test. Lets see some summary statistics for the graph.

* The average of the AQI from 2019 was 35.44 give or take 11.23.
* The average of the AQI from 2020 was 28.44 give or take 7.96.
* The r.m.s. error is about 5.669.
* There seems to be some good correlation between the averages of the two AQIs.
* However, this doesn’t tell us much about why the data is correlated or how this pertains to non-essential business closures.

#### Testing for independence between population size and the status of business closures nationwide using a Chi-Squared Distribution.

##### Null Hypothesis:

* Population size and the status of business closures are independent of each other and these results will be due to chance variation.

##### Alternative Hypothesis:

* Population size and the status of business closures are dependent of each other and these results will not be due to chance variation.

##### Testing for Chi-Squared Statistic and the p-value for the data

High.Pop Mid.Pop Low.Pop Col.Total  
Closed 5 6 2 13  
Some Open Limit 4 3 3 10  
Some Open 0 2 1 3  
All Open Limit 1 3 2 6  
All Open 3 8 3 14  
Never Closed 0 3 2 5  
7 13 25 13 51

High.Pop Mid.Pop Low.Pop  
Closed 3.3137255 6.372549 3.3137255  
Some Open Limit 2.5490196 4.901961 2.5490196  
Some Open 0.7647059 1.470588 0.7647059  
All Open Limit 1.5294118 2.941176 1.5294118  
All Open 3.5686275 6.862745 3.5686275  
Never Closed 1.2745098 2.450980 1.2745098

Pearson's Chi-squared test  
  
data: a.chi.test  
X-squared = 6.5815, df = 10, p-value = 0.7643

* The Chi-Squared statistic is 6.581 with a p-value of 0.9682
* I double checked the expected values by hand and manually input the chi-squared statistic and p-value, they were right.

##### Conclusion

The p-value is 0.7643 which is well above the 0.05 threshold, so we must not reject the null and say that the population size and business closures are most likely independent of each other.

##### Testing for chance variation between the two averages against the total average.

* The average of the AQI from the simple random sample for states that fully or mostly closed down is 28 give or take 6.856.
* The average of the AQI from the simple random sample for states that DID NOT fully or mostly close down is 27.75 give or take 8.771
* The average of the AQI from 2020 was 28.44 give or take 7.956.

#### Looking for correlation between Non-Essential Business Closures and 2020 Average AQI.

* Were the AQI levels for the states that fully closed or at least closed down MOST non-essential businesses lower than the national average? (8 states chosen randomly that meet the criteria)
* Were the AQI levels for the states that DID NOT fully closed or at least closed down MOST non-essential businesses greater than the national average? (8 states chosen randomly that meet the criteria)

##### After running a T-Test for both simple random samples, I found that both p-values were well above the 0.05 threshold and that they were most likely not due to chance variation.