****

**Data 602 Project**

Group: 6

Meghana Kompally

Serena Sun

Jannatul Naeema

**Historical Air Quality Analysis**

**Abstract**

The purpose of this report is to apply statistical methods to help understand how various air quality data over the years and how they relate to other related data sets.

**Introduction**

The domain that we will be working with for this project is Environment and Safety. Within this domain, we are trying to investigate various questions pertaining to air quality. The analysis of this data will allow us to draw conclusions about the overall air quality status in Calgary, how on average major greenhouse gasses emissions compare to the global data yearly and how they correlate with Calgary population growth. This will show us if Calgary pollutant levels are similar to that of the rest of the world and give us an idea of the precautions humans should take to protect our respiratory health. It will also help us understand where Calgary stands in its emission levels and allows us to take responsible eco-friendly actions to protect our environment especially due to climate change. We will also be drawing conclusions within the primary data set between the stations from which air quality data was collected from.

**Data Sets and Sources**

The datasets we have chosen for our project is “Historical Air Quality” which was collected by the Calgary Region Airshed Zone and submitted to Alberta Environment and Parks (AEP), WHO Air Quality Database and Calgary- Historical Population Data. This information is publicly available and can be used from the City of Calgary’s Open Data Portal, WHO website and MacroTrends website respectively.

1. Historical Air Quality from Open Calgary Data Portal

URL: <https://data.calgary.ca/Environment/Historical-Air-Quality/uqjm-jxgp/data>

1. WHO Air Quality Database (csv file)

URL: https://www.who.int/data/gho/data/themes/air-pollution/who-air-quality-database

1. Calgary Population Historical Data

URL: https://www.macrotrends.net/cities/20370/calgary/population

**Guiding Questions**

1. Is there any statistical difference between the particulate matter recorded in Calgary vs particulate matter recorded globally in order to gain evidence to support or reject our hypothesis that PM2.5 globally is equal to PM2.5 in Calgary?

This question is posed to see if PM2.5 of Calgary is equal to the World PM2.5 as this allows us to understand if Calgary air quality stands in the same position as the rest of the world.

1. Is there any statistical difference between the Air Quality Index (AQI) measured in the central stations of Calgary and the AQI measured in the Southeast/Northwest and East stations of Calgary.

This question is posed to understand whether Calgary central stations see more pollution compared to the rest of the city as one would expect central calgary to be the busiest hence likely having a higher level of pollutants.

1. Is there any correlation between the major greenhouse gasses -Nitrogen Dioxide (NO2) and Nitric Oxide (NO) and Particulate Matter of 2.5 mass (PM2.5) with the population growth of Calgary over the years? This is to check if Calgary's population is related to the level of greenhouse gas emissions and the amount of air particles in the air.

This question is posed to see whether the level of pollution is correlated to the number of people residing in Calgary as one would expect to see a positive correlation between population growth and pollution levels.

**Statistical Methods by Question**

1. Hypothesis Testing (Two Populations)- Permutation Test and Student T-test was used where the yearly mean PM2.5 for 10 years in Calgary was compared with the yearly mean PM2.5 for 12 years in the world.

Our Hypothesis:

Null Hypothesis, H0 : µCalgary, PM 2.5 - µWorld, PM 2.5 = 0

Alternate Hypothesis, HA : µCalgary, PM 2.5 - µWorld, PM 2.5 ≠ 0

We will run a permutation test so that the normality condition is automatically satisfied and use the resulting p-value and confidence intervals to make the appropriate conclusions. The Student T-test will also be used to check how the empirical p-value differs from the t-test p-value. A normality plot will be plotted to check the normality of the data when using the t-test.

1. Hypothesis Testing (Two Populations)- Permutation Test and Student T-test was used. Here, the data set is divided into two parts, one consisting of data collected from all the Central locations (Calgary Central, Central 2 and Central-Inglewood) and another consisting of the data from the rest of the locations (Calgary Southeast, Northwest and East). Then the yearly mean Air Quality Index (AQI) for each of the two parts was compared against one another.

Our Hypothesis:

Null Hypothesis, H0 : µCentral, Air Quality Index - µNon-Central, Air Quality Index ≤ 0  
Alternate Hypothesis, HA : µCentral, Air Quality Index - µNon-Central, Air Quality Index > 0

We will run a permutation test so that the normality condition is automatically satisfied and use the resulting p-value and confidence intervals to make the appropriate conclusions.The Student T-test will also be used to check whether the empirical p-value differs from the t-test p-value. A normality plot will be plotted to check the normality of the data when using the t-test.

1. Statistical Modeling of Bivariate Data using Linear Regression Analysis is to be used to answer this question. A regression analysis is applied 3 times to see if there is any correlation between Calgary population growth and Nitric Oxide, Nitrogen Dioxide and Particulate matter (2.5).

Hypothesis for Population and PM2.5, Nitrogen Dioxide and Nitric Oxide

H0: The yearly PM2.5 values recorded in Calgary cannot be expressed as a linear function of the population growth in Calgary.

HA : The yearly PM2.5 values recorded in Calgary can be expressed as a linear function of the population growth in Calgary.

H0: The yearly Nitrogen Dioxide values recorded in Calgary cannot be expressed as a linear function of the population growth in Calgary.

HA : The yearly Nitrogen Dioxide values recorded in Calgary can be expressed as a linear function of the population growth in Calgary.

H0: The yearly Nitric Oxide values recorded in Calgary cannot be expressed as a linear function of the population growth in Calgary.

HA : The yearly Nitric Oxide values recorded in Calgary can be expressed as a linear function of the population growth in Calgary.

For all of these regression analysis, we will plot a correlation plot and compute a correlation coefficient to look at the strength of the correlation. Normality and homoscedasticity plots will be plotted to test for normality and uniformity of the data. We will also estimate a linear statistical model for each of the regression analysis.

**Data Analysis**

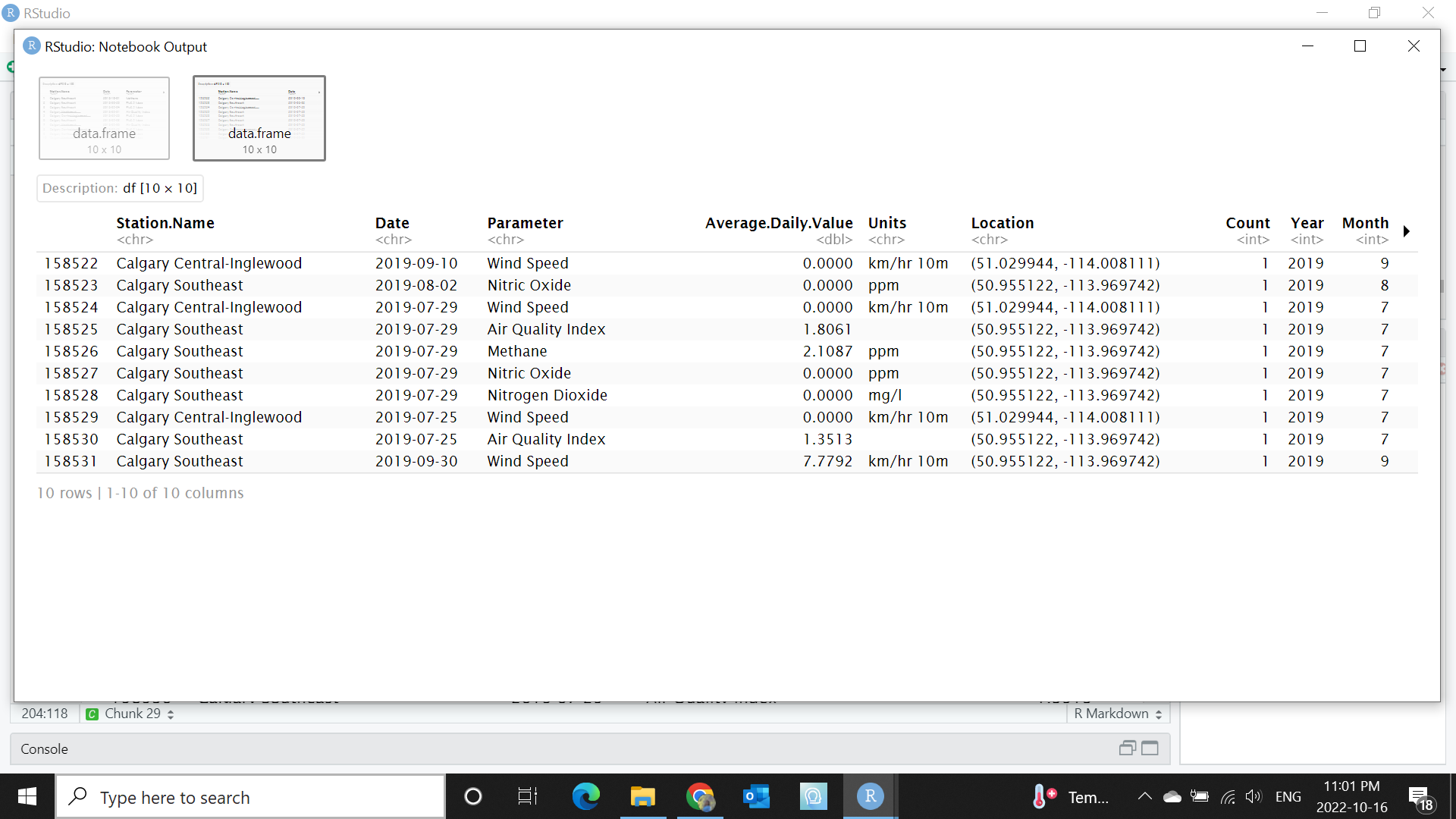
Question A.

Our Hypothesis:

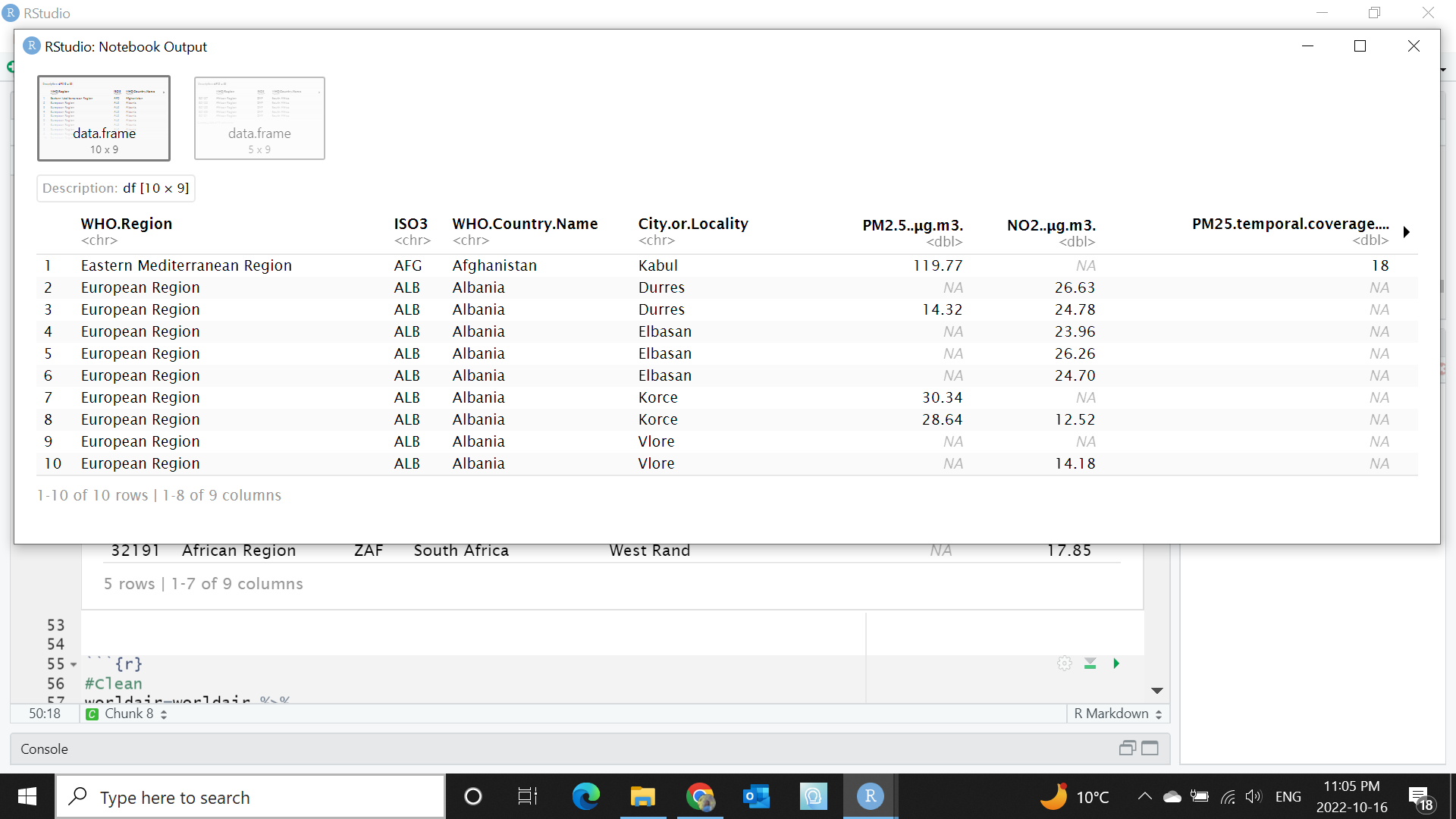
Null Hypothesis, H0 : µCalgary, PM 2.5 - µWorld, PM 2.5 = 0

Alternate Hypothesis, HA : µCalgary, PM 2.5 - µWorld, PM 2.5 ≠ 0

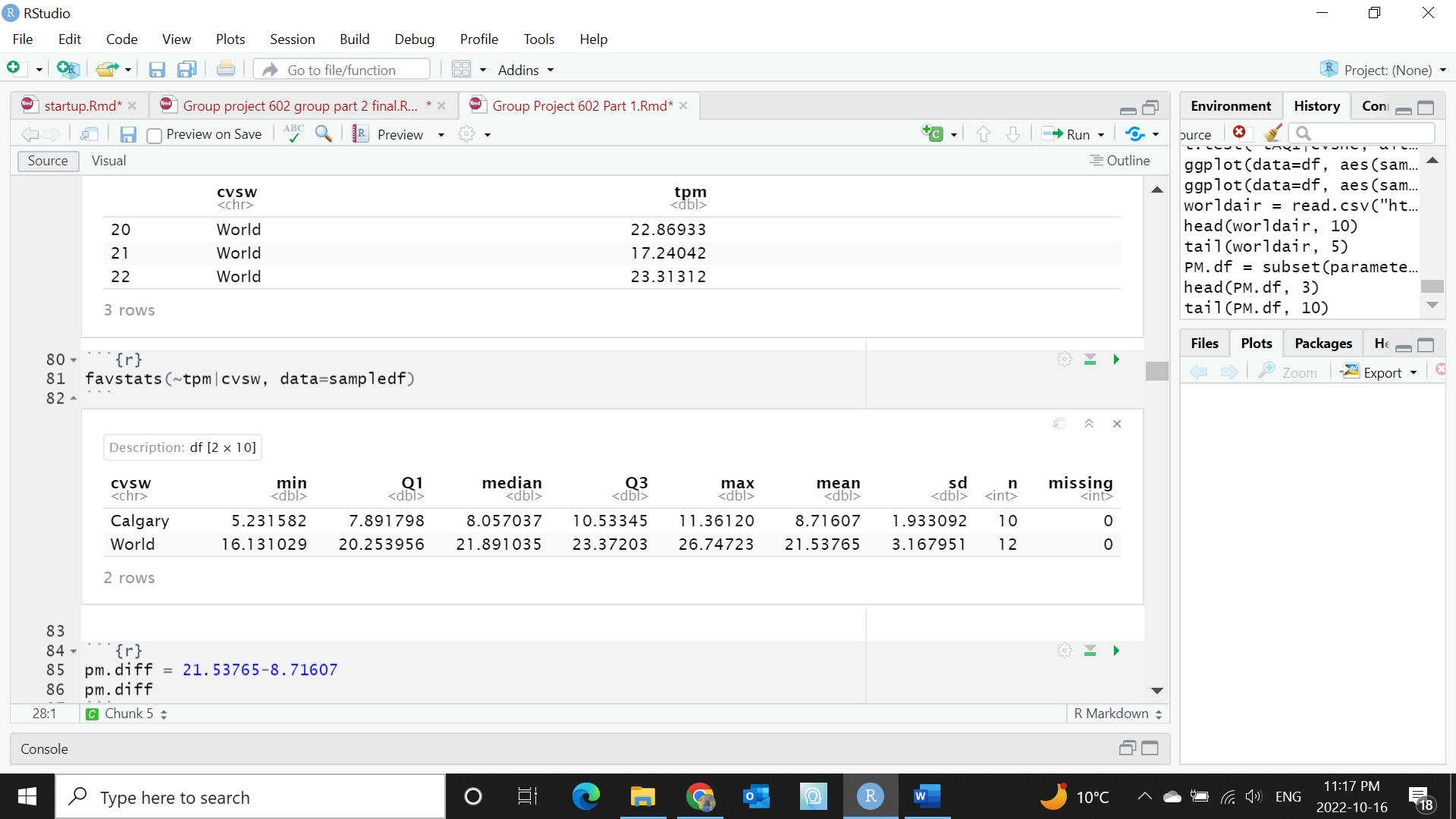
Historical Calgary Air Quality of Calgary Dataset Sample:



World Air Quality Dataset Sample:



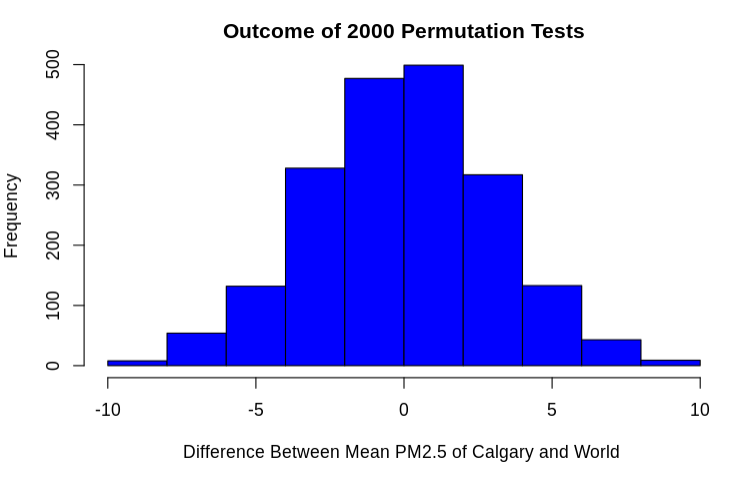
Mean, Median, Standard deviation of PM 2.5 for both Calgary and World Data:



Actual mean difference between PM2.5 in the world and PM2.5 in Calgary:

8.71607-21.53765= -12.82158 μg.m3

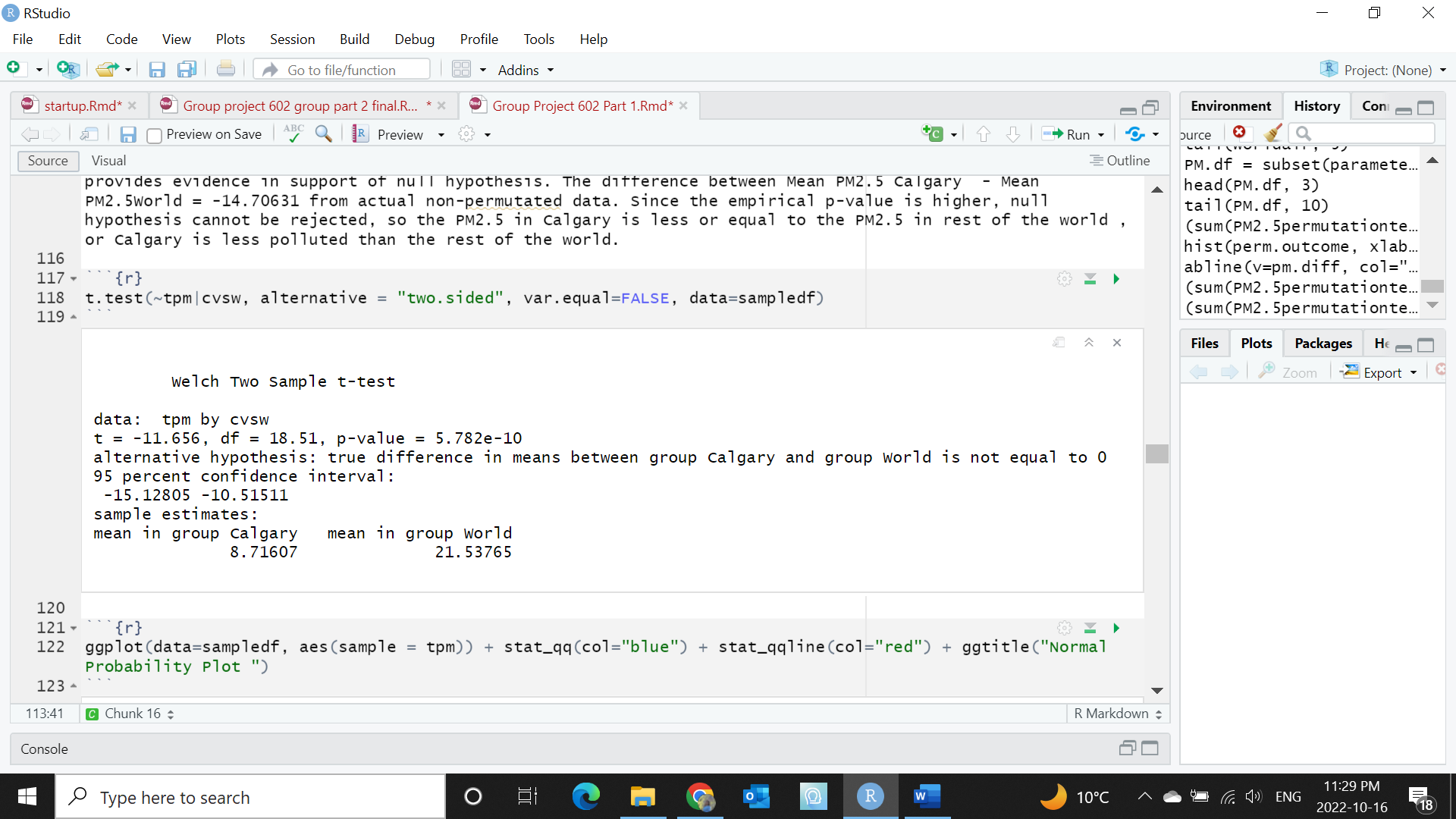
Permutation Test between PM2.5 globally and PM2.5 in Calgary:



Empirical P-Value=0

The empirical P-value was computed to be 0 after running the permutation test 2000 times. Since the P-value is less than 0.05, we find that there is evidence against the null hypothesis hence rejecting it. The difference between Mean PM2.5 Calgary and Mean PM2.5World was equal to -12.82158 μg.m3 from actual non-permuted data. Since this value falls far from the permuted values in the graph above. We can further confirm that the recorded World PM2.5 values are different from Calgary PM2.5 values and that their difference is not equal to 0 and statistically significant.

We also did Student’s T-test to check our hypothesis (alternate statistical method):

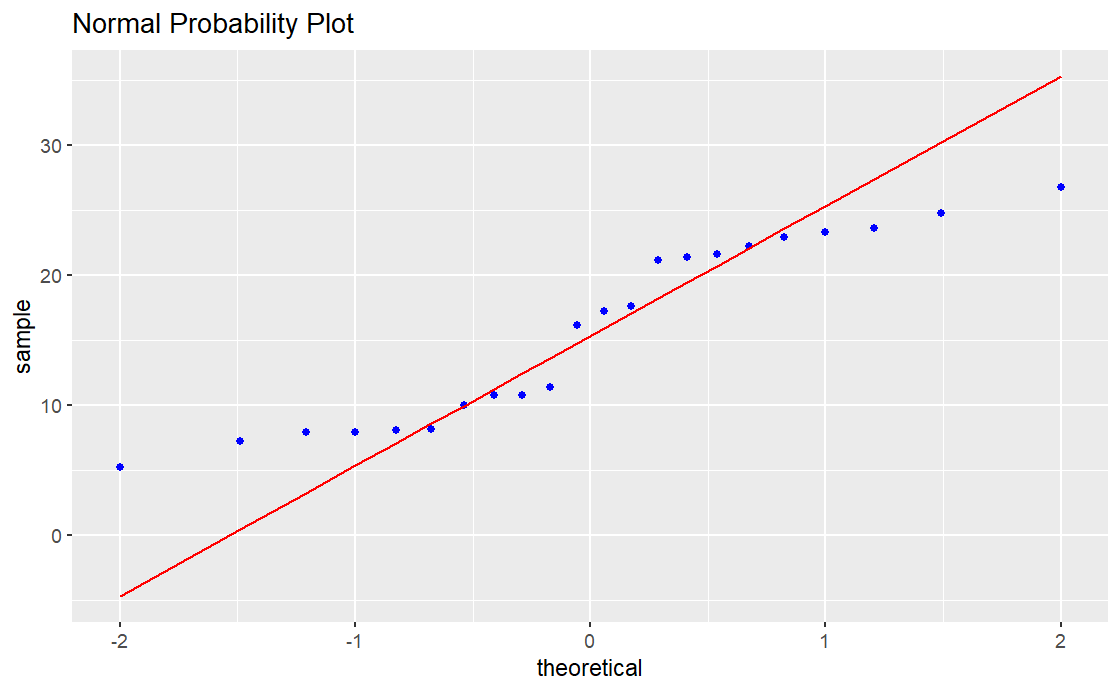


P-value = 5.782e-10 , which too small, so we reject the null hypothesis  
Degree of freedom, df= 18.51  
Tobs = -11.656

95% Confidence Interval:

-15.12805 ≤ PM2.5 ≤ -10.51511

Normal Probability plot to check if the data are in normal distribution



Since most of the points in the normality plot fall on top of the straight line even though there are some trailing edges. It can be concluded that the normality condition is met and satisfied.

Since the normality condition is met, the p-value can be analyzed. Since the computed P-value is less than 0.05, we reject the null hypothesis. Therefore, this t-test further strengthens the claim made in the permutations test that the PM2.5 recorded globally is different from that of PM2.5 recorded locally in the city of Calgary.

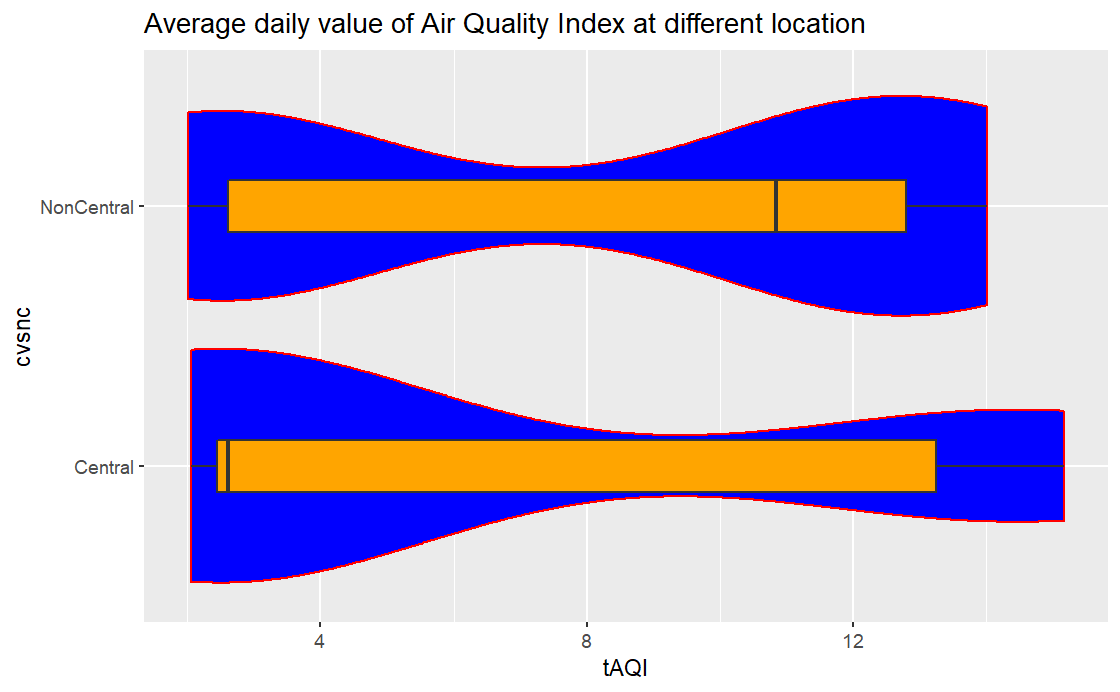
Question B

Our Hypothesis:

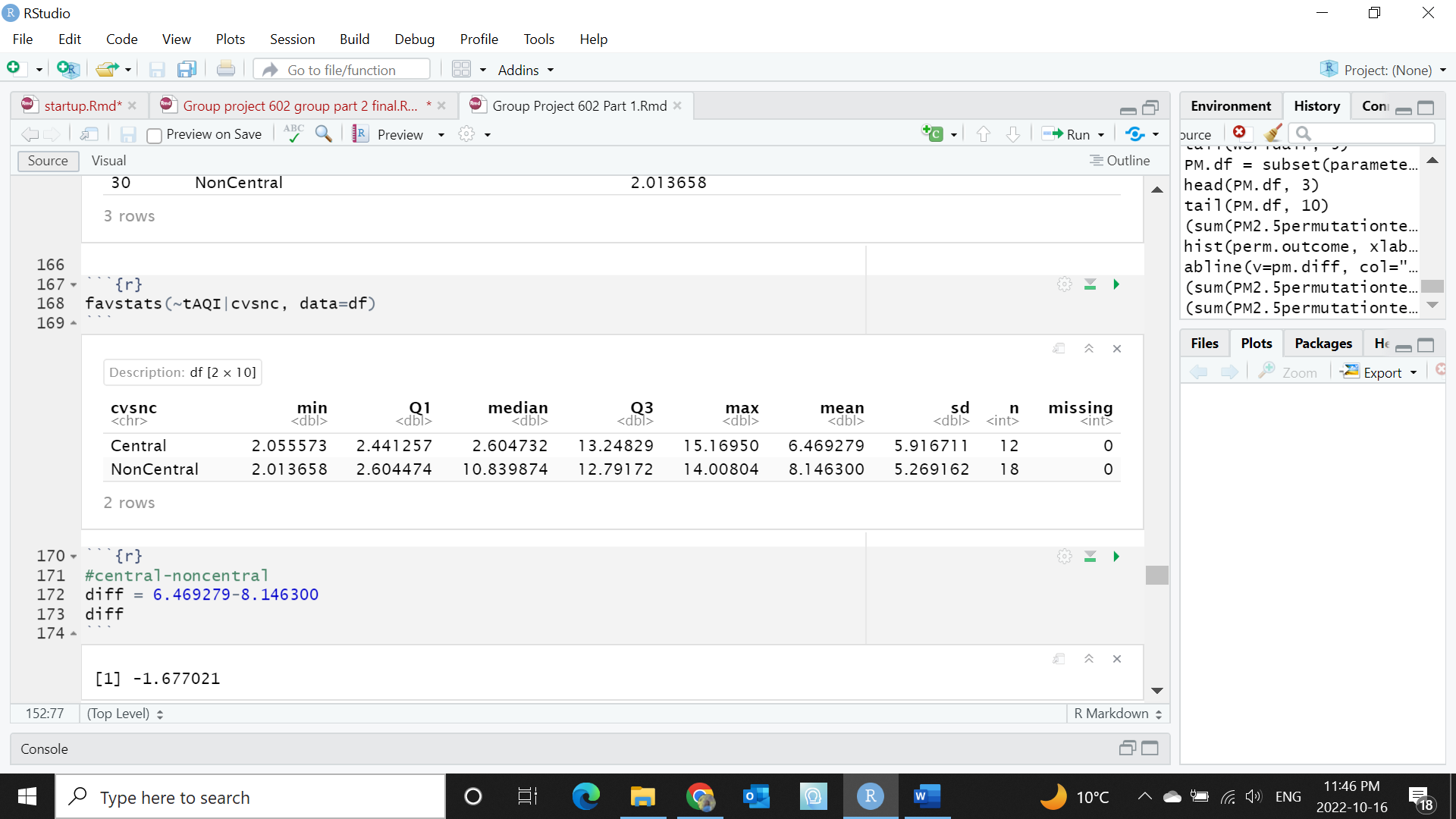
Null Hypothesis, H0 : µCentral, Air Quality Index - µNon-Central, Air Quality Index ≤ 0  
Alternate Hypothesis, HA : µCentral, Air Quality Index - µNon-Central, Air Quality Index > 0

For this question, we are testing whether the Calgary central air quality locations stations see a higher Air Quality Index (AQI) which pertains to low air quality and high risk. This assumption is made as there are many commercial buildings in that area and sees the most amount of activity. Therefore, for this test, our historical air quality data is divided into two categories with Category 1 having central stations (Calgary Central , Calgary Central 2, Calgary Central-Inglewood) and Category 2 having Non-Central Stations (Calgary Southeast, Calgary Northwest, Calgary East).

Here is a visualization of the AQI values in Non Central stations vs. Central stations. Based on the graph, we can see that non central areas see more AQI variability compared to AQI in central areas. The median AQI of non central areas is much higher compared to the median AQI central areas. This graph also goes against our assumption that central AQI would be higher, but we will further test it statistically to see if there is any statistical difference.



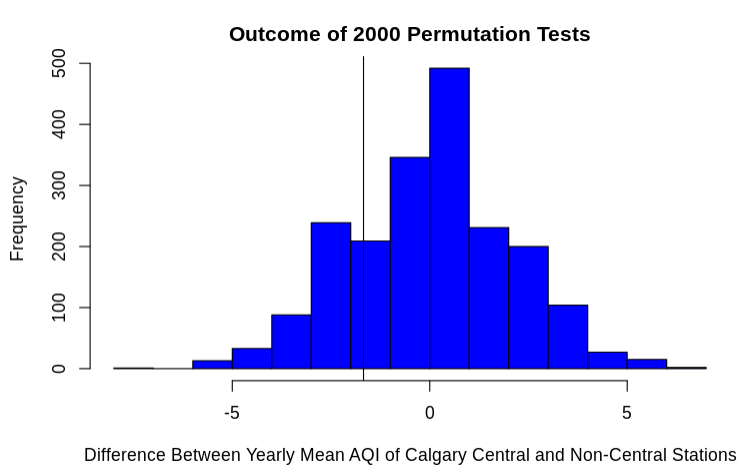
Mean, Median, Standard deviation of Air Quality Index for both Central and Non-central Stations in Calgary :



Actual mean difference between yearly mean AQI in the Calgary central stations and yearly mean AQI in the Calgary non-central stations:

6.469279-8.146300= -1.677021

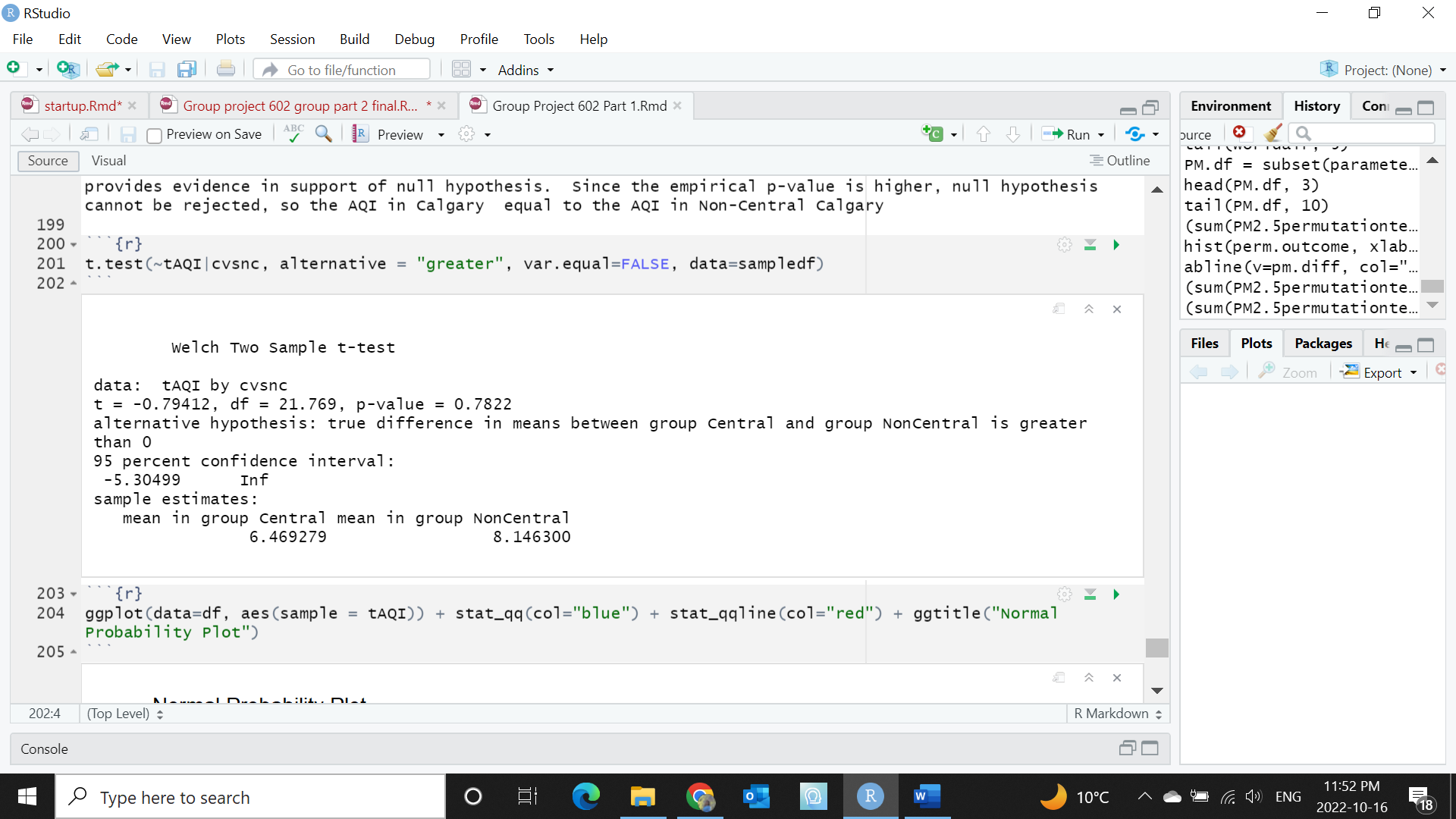
Permutation test for Air Quality Index between central and non central stations in Calgary:



Empirical P-value: 0.795

The empirical P-value was computed to be 0.804 after running the permutation test 2000 times. Since the P-value is more than 0.05, we find that there is no evidence against the null hypothesis hence failing to reject it. The difference between Mean AQI central stations and Mean AQI non-central stations was equal to -1.677021 from actual non-permuted data. Since this value falls within the permuted values in the graph above (black vertical line). We can further confirm that the recorded AQI values from the central stations are in fact less than or equal to the AQI values recorded in the non-central stations. A lower AQI shows that there is a low risk and that the pollution levels are less, therefore the central regions of Calgary are less polluted or as equally polluted as the rest of Calgary. Furthermore, an AQI of around 1 is deemed to be very low risk in more AQI scales.

We also did Student’s T-test to check our hypothesis (alternate statistical method):



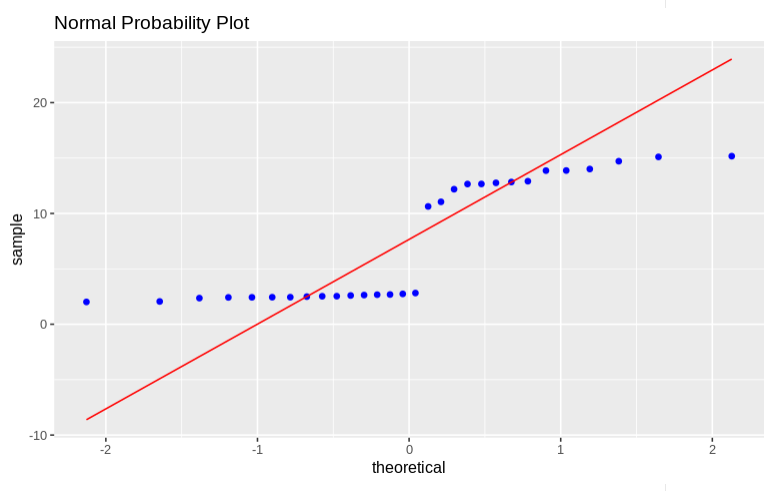
P-value = 0.7822 , Fail to reject the null hypothesis  
Degree of freedom, df= 21.769  
Tobs = -0.79412

95% Confidence Interval:

-5.30499 ≤ Air Quality Index ≤ Inf

The confidence interval captures 0 which means that there is a possibility that the yearly mean AQI captured in the central stations is equal to that of the yearly mean AQI captured in the rest of the stations.

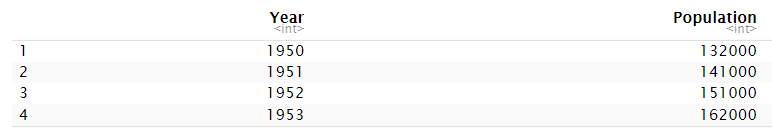
Normal Probability plot to check if the data are in normal distribution



Since most of the points in the normality plot do fall on top of the straight line and have many trailing edges. It can be concluded that the normality condition is not met. Therefore, this t-test analysis is inaccurate and cannot be considered to draw accurate conclusions about the data.

Question C

Calgary Population Growth Data Set Sample:



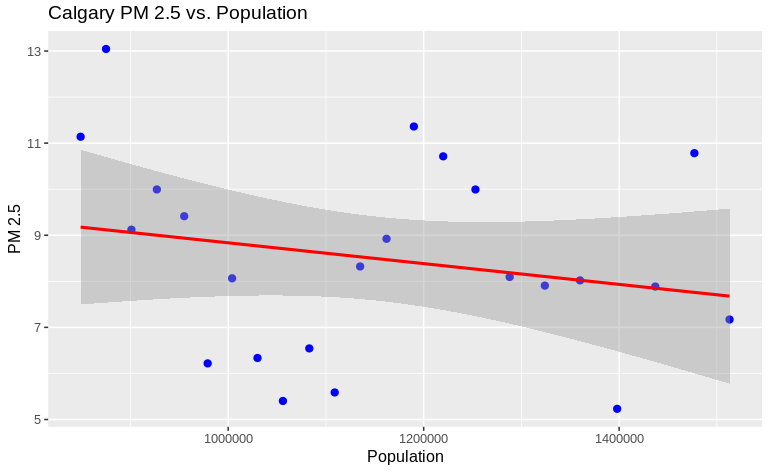
1. Linear Regression Analysis of the recorded PM2.5 and Calgary population growth over the overlapping years.

*Our Hypothesis:*

H0: The yearly PM2.5 values recorded in Calgary cannot be expressed as a linear function of the population growth in Calgary. B =0

HA : The yearly PM2.5 values recorded in Calgary can be expressed as a linear function of the population growth in Calgary. B≠0

A scatterplot of PM2.5 against Calgary population over the years.

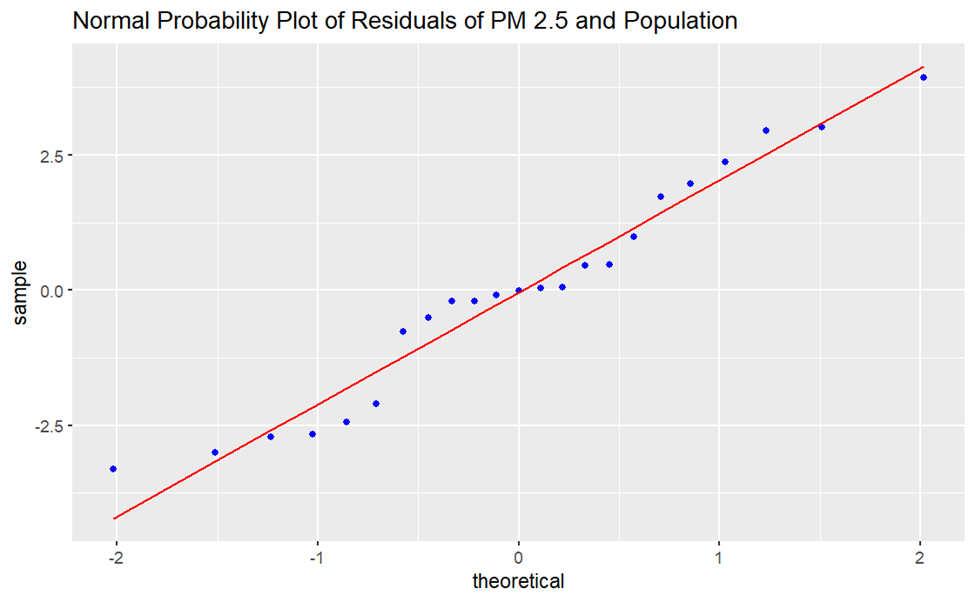


The correlation coefficient was calculated to be -0.2156038, this value and the graph show that there is a slight negative correlation between the yearly PM2.5 and population growth in Calgary meaning the PM2.5 slightly decreases as the population increases.

After further analysis the linear model gives the following equation:

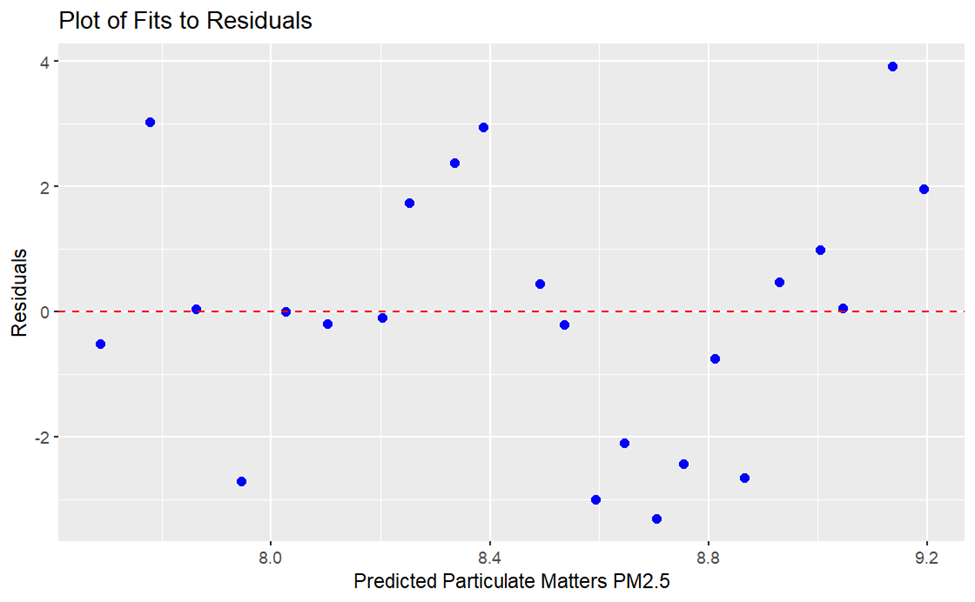
PM 2.5 = A + B \* Population + e  
A= 11.092034603450  
B= -0.000002256207  
PM 2.5 = 11.092034603450 + (-0.000002256207) \* Population + e

Checking for the Normality Condition:



Since most of the points roughly fall along the straight line, we can conclude that the normality condition is met and satisfied.

To inspect the homoscedasticity condition, we plot the fitted values with the residuals:



Since a straight horizontal line is produced and the points are roughly regularly distributed all across the graph. It can be concluded that this data satisfies the condition of homoscedasticity.

Since the conditions are met, the F-value and P-value are analyzed with F-test to draw conclusions from the hypothesis.

F value: 1.0240

P-value(>F): 0.3231

Since the computed P-value is more than 0.05, we fail to reject the null hypothesis. To determine the accuracy of the estimation of our model and how well the model mimics the real world, the coefficient of determination is calculated. This value is . Based on the calculated p-value, we can infer that the population growth in Calgary ***cannot*** be expressed as a linear function of the yearly PM2.5 values recorded in Calgary.

Since, it is determined above that our parameters cannot be expressed as a linear function, the model testing to prove this model is not completed.

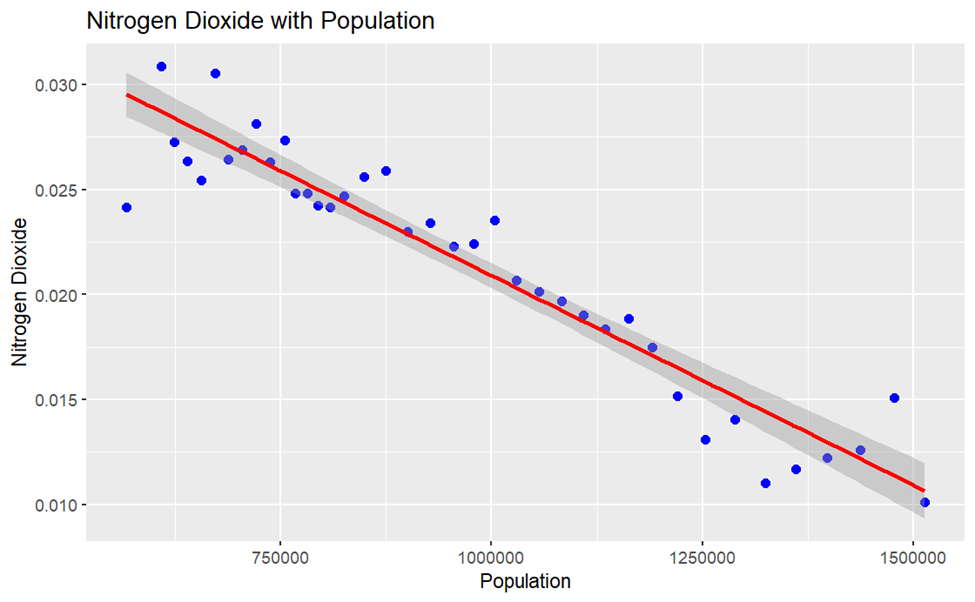
1. Linear Regression Analysis of the recorded Nitrogen Dioxide (mg/l) and Calgary population growth over the overlapping years.

*Our Hypothesis:*

H0: The yearly Nitrogen Dioxide values recorded in Calgary cannot be expressed as a linear function of the population growth in Calgary. B = 0

HA : The yearly Nitrogen Dioxide values recorded in Calgary can be expressed as a linear function of the population growth in Calgary. B≠0

A scatterplot of Nitrogen Dioxide against Calgary population over the years.

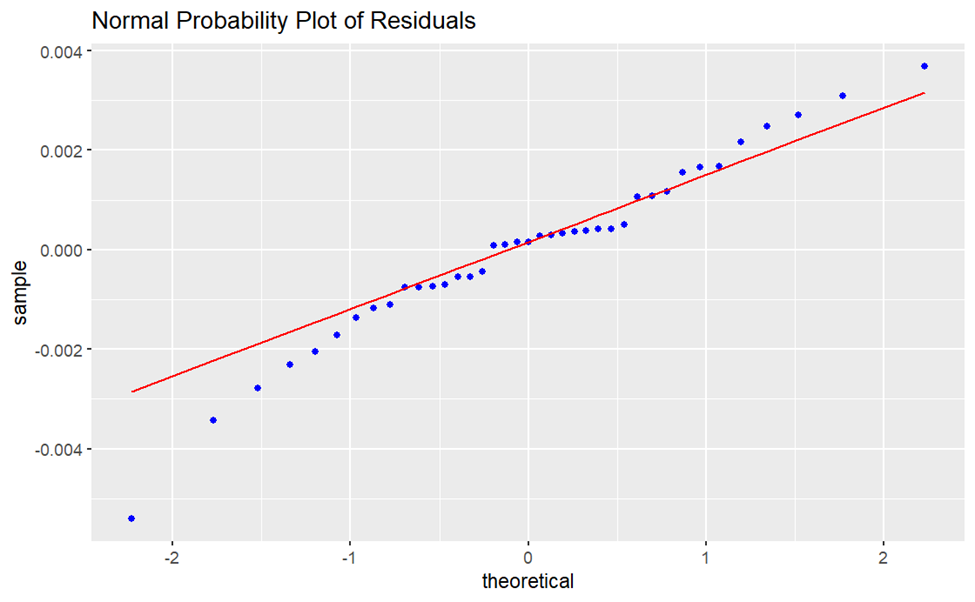


The correlation coefficient was calculated to be r: -0.9497451, this value and the graph show that there is a strong negative correlation between the yearly Nitrogen Dioxide emission and population growth in Calgary which means that Nitrogen Dioxide decreases as the population increases or as the years go by.

After further analysis the linear model gives the following least squares estimation equation:

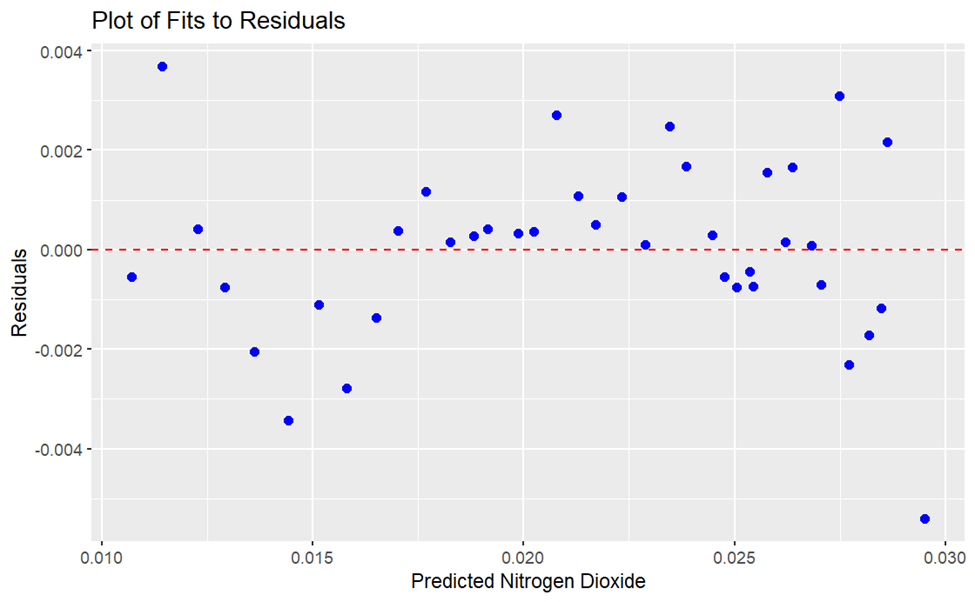
Nitrogen Dioxide = A + B × Population + e  
A= 0.04086983004144  
B= -0.00000001996488  
Nitrogen Dioxide = 0.04086983004144 + (-0.00000001996488) \* Population + e

Checking for the Normality Condition:



Since most of the points roughly fall along the straight line, we can conclude that the normality condition is met and satisfied.

To inspect the homoscedasticity condition, we plot the fitted values with the residuals:



Since a straight horizontal line is produced and the points are roughly regularly distributed all across the graph. It can be concluded that this data satisfies the condition of homoscedasticity.

Since the conditions are met, the F-value and P-value are analyzed with F-test to draw conclusions from the hypothesis.

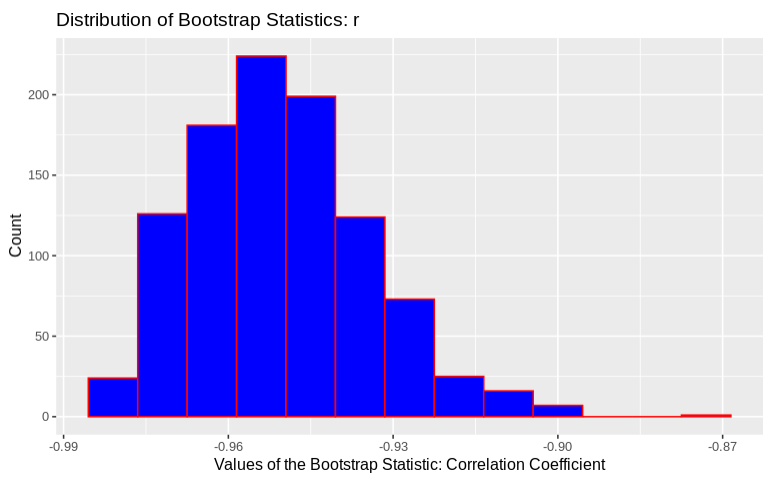
F value: 340.6

P-value(>F): <2e-16

Since the computed P-value is much less than 0.05, we reject the null hypothesis. To determine the accuracy of the estimation of our model and how well the model mimics the real world, the coefficient of determination is calculated. This value is . Based on the calculated p-value, we can infer that the population growth in Calgary ***can*** be expressed as a linear function of the yearly Nitrogen Dioxide values recorded in Calgary.

As it has been shown that these parameters can be expressed as a linear function, we will continue our linear modeling by bootstrapping for r, A and B values to determine how the resulting model (equation) compares to the one produced above.

Bootstrap Distribution of rboot

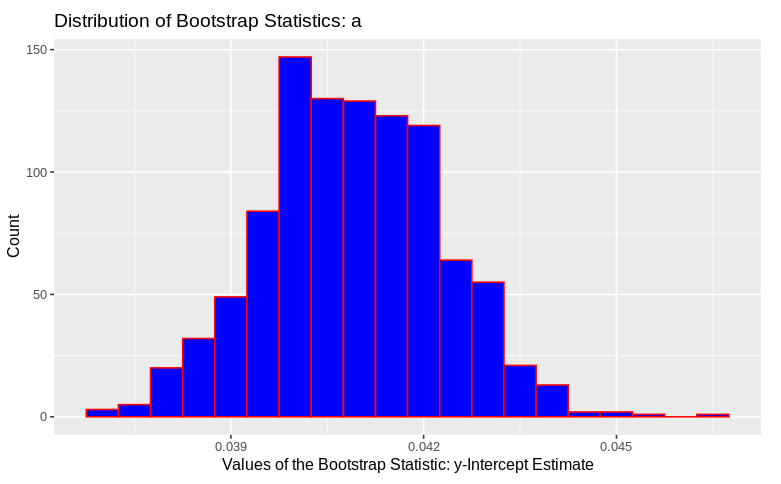


Mean Correlation Coefficient ( rboot) = -0.9504573

95% Confidence Interval: (2.5%, 97.5%)

-0.9763399 r -0.9144689

Bootstrap Distribution of aboot:

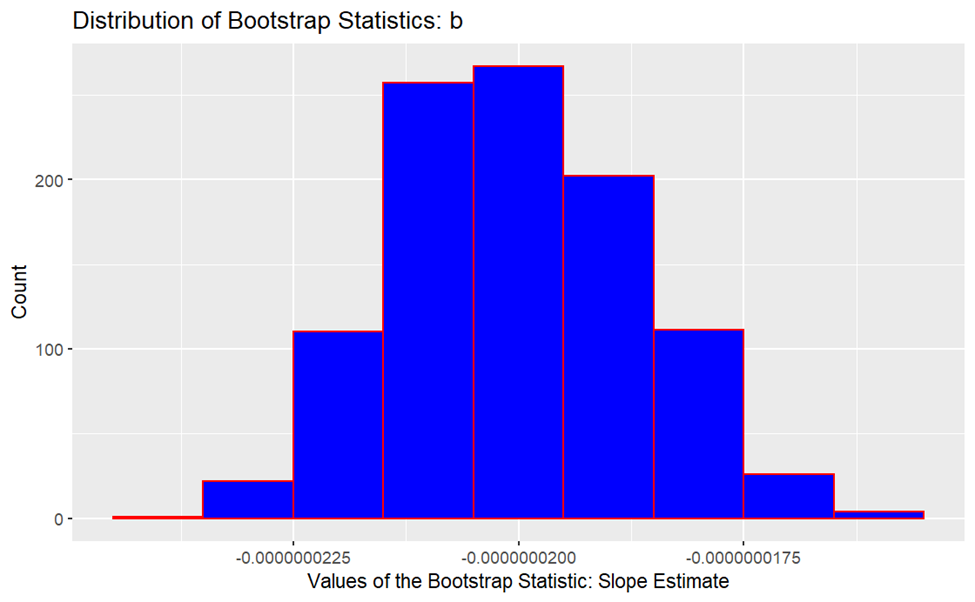


Mean Y-intercept Estimate: 0.04089894

95% Confidence Interval: (2.5% , 97.5%)

0.03821873 A 0.04363896

Bootstrap Distribution of bboot



Mean Slope Estimate: -2.001939×10-08

95% Confidence Interval: (2.5%, 97.5%)

-2.248524×10-08 B -1.736453×10-08

After the completion of bootstrapping the three parameters, we end up with the following bootstrapped equation:

Nitrogen Dioxide = 0.04089894 + (-2.001939×10-08 ) \* Population + e

We can see that the bootstrapped values are very close to our original equation (Least- squares estimation) and fall within the bootstrapped 95% confidence interval.

Nitrogen Dioxide = 0.04086983004144 + (-1.996488×10-08) \* Population + e

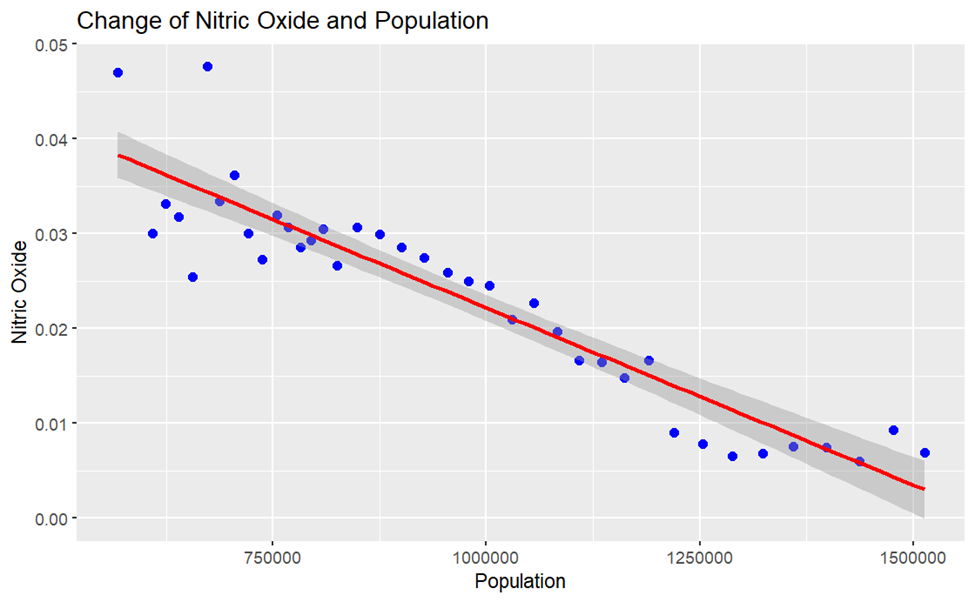
1. Linear Regression Analysis of the recorded Nitric Oxide (ppm) and Calgary population growth over the overlapping years.

*Our Hypothesis:*

H0: The yearly Nitric Oxide values recorded in Calgary cannot be expressed as a linear function of the population growth in Calgary. B = 0

HA : The yearly Nitric Oxide values recorded in Calgary can be expressed as a linear function of the population growth in Calgary.B ≠0

A scatterplot of Nitric Oxide against the Calgary population over the years.

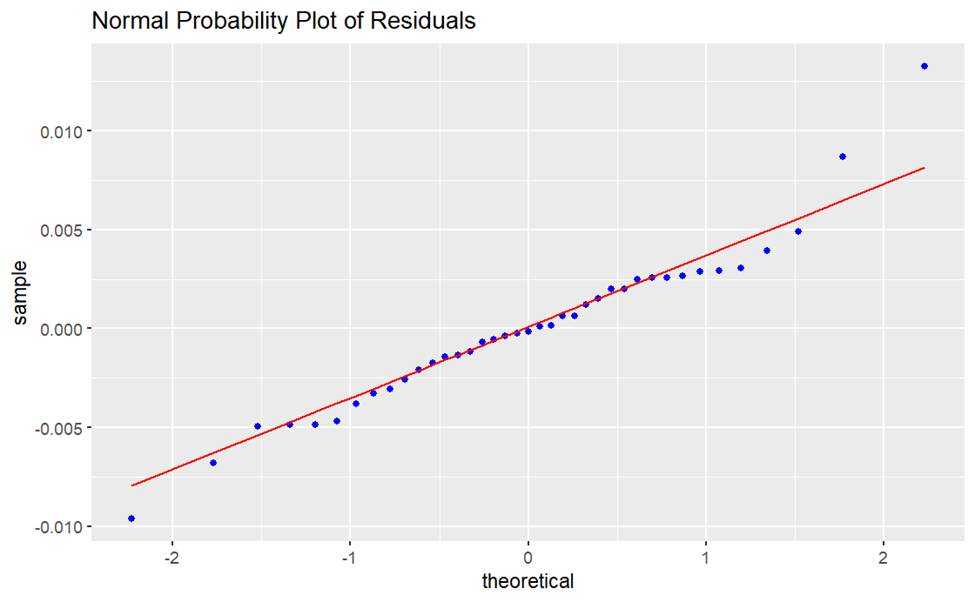


The correlation coefficient was calculated to be r:-0.9275725, this value and the graph show that there is a strong negative correlation between the yearly Nitric Oxide emission and population growth in Calgary which means that Nitric Oxide decreases as the population increases or as the years go by.

After further analysis the linear model gives the following least square estimation equation:

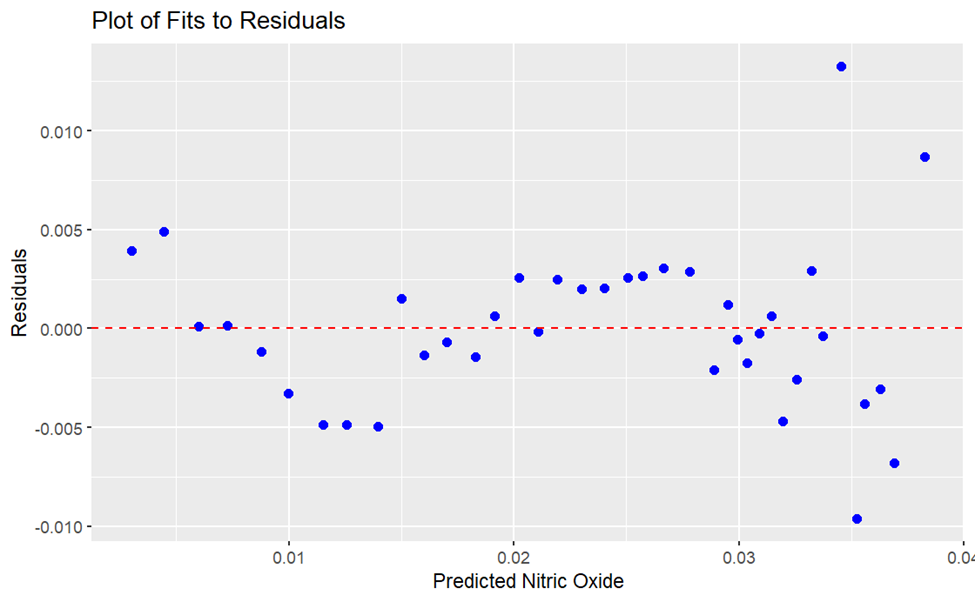
Nitric Oxide = A + B \* Population + e  
A= 0.0595543100516  
B= -0.000002256207  
Nitric Oxide = 0.0595543100516 + (-0.0000000373782 ) \* Population + e

Checking for the Normality Condition:



Since most of the points roughly fall along the straight line, we can conclude that the normality condition is met and satisfied.

To inspect the homoscedasticity condition, we plot the fitted values with the residuals:



Since a straight horizontal line is produced and the points are roughly regularly distributed all across the graph. It can be concluded that this data satisfies the condition of homoscedasticity.

Since the conditions are met, the F-value and P-value are analyzed with F-test to draw conclusions from the hypothesis.

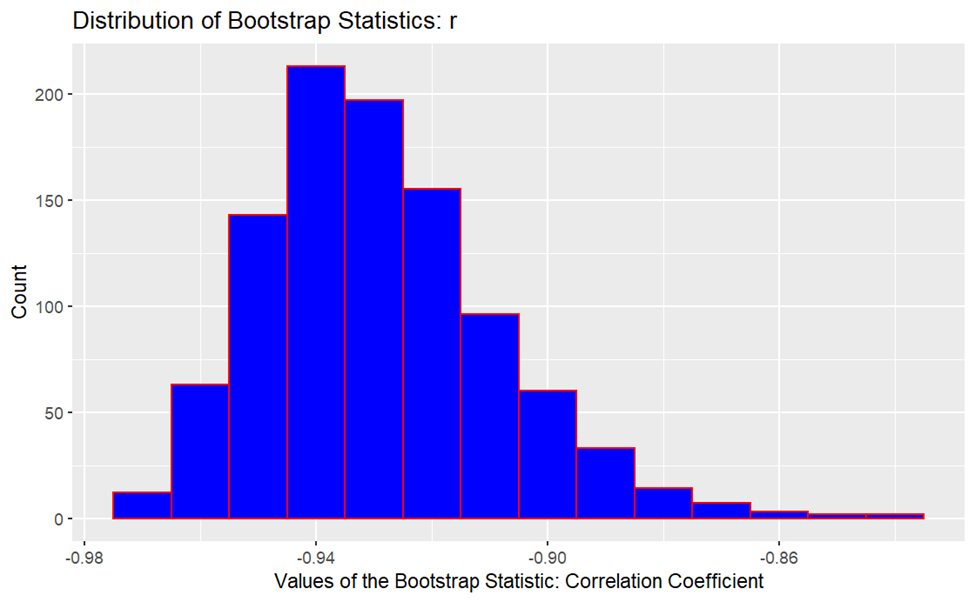
F value: 228

P-value(>F): <2e-16

Since the computed P-value is much less than 0.05, we reject the null hypothesis. To determine the accuracy of the estimation of our model and how well the model mimics the real world, the coefficient of determination is calculated. This value is . Based on the calculated p-value, we can infer that the population growth in Calgary ***can*** be expressed as a linear function of the yearly Nitric Oxide values recorded in Calgary.

As it has been shown that these parameters can be expressed as a linear function, we will continue our linear modeling by bootstrapping for r, A and B values to determine how the resulting model (equation) compares to the one produced above.

Bootstrap Distribution of rboot

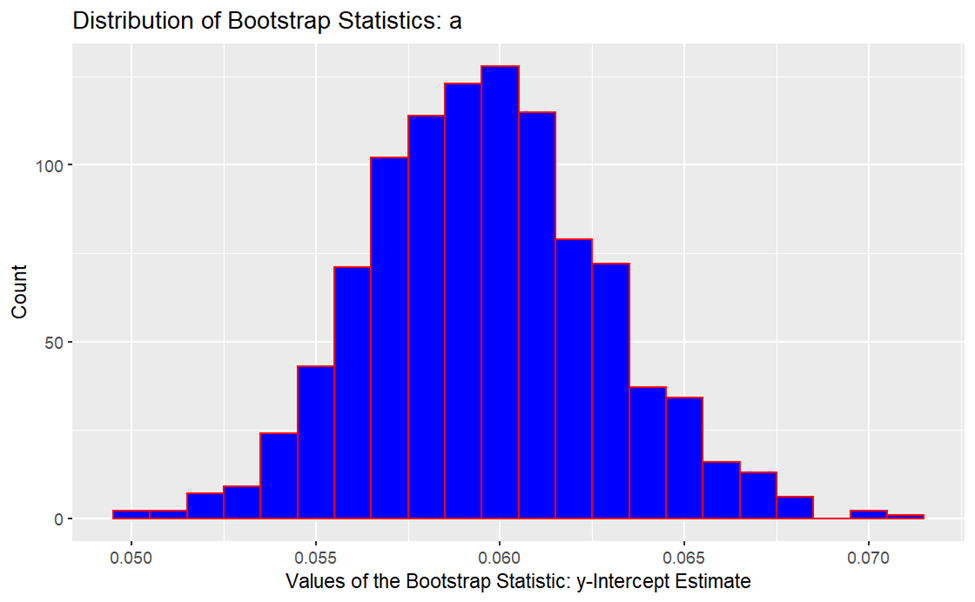


Mean Correlation Coefficient ( rboot): -0.9290379

95% Confidence Interval: (2.5%, 97.5%)

-0.9607345 r -0.8825924

Bootstrap Distribution of aboot

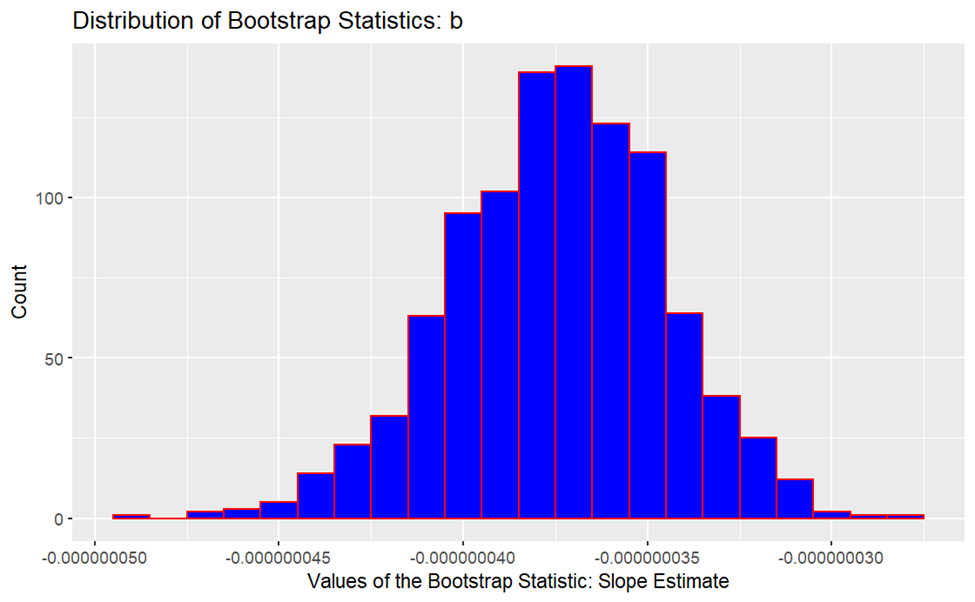


Mean Y-intercept Estimate (A): 0.05962442

95% Confidence Interval: (2.5%, 97.5%)

0.05358987 A 0.06625294

Bootstrap Distribution of bboot



Mean Slope Estimate (B): -3.744776×10-08

95% Confidence Interval: (2.5%, 97.5%)

- 4.350186×10-08 B - 3.198461 ×10-08

After the completion of bootstrapping the three parameters, we end up with the following bootstrapped equation:

Nitric Oxide = 0.05962442 + ( -3.744776×10-08 ) \* Population + e

We can see that the bootstrapped values are very close to our least squares estimation equation (given below) and fall within the bootstrapped 95% confidence interval.

Nitric Oxide = 0.0595543100516 + (-3.73782 × 10-08 ) \* Population + e

**Conclusion**

Guiding Question A:

In this question, we wanted to understand whether the yearly mean of PM2.5 ug/m3 of Calgary is equal to the world PM2.5 values. After performing a permutation test we came to the conclusion that the mean of air pollutant particulate matter PM 2.5 of Calgary is not equal to the PM 2.5 of the rest of the world. We also did a t-test to further confirm our results. Both permutation test and t-test provided support to the alternate hypotheses. This tells us that the PM2.5 levels in Calgary vs the world are significantly different and not related to one another.

In a practical sense, the above results would make sense in that Calgary emissions would be assumed to be relatively low compared to the rest of the world and therefore not equal to one another. For this report, we only wanted to understand if Calgary PM2.5 values were similar to that of the world values, therefore further analysis might be required to understand whether Calgary values are higher or lower than the rest of the world.

Guiding Question B:

In this question, we wanted to explore whether the central air quality stations recorded a higher level of pollution compared to the rest of the stations around Calgary because central Calgary is assumed to see relatively more pollution due to higher human activity. Since the main measure of pollution is AQI, we chose to compare the AQI means between the two station categories. After running the permutation test, the null hypothesis was supported showing that central stations see less than or equal amount of AQI as the rest of the stations. This hypothesis test was also carried out using the t-test, however, it was found that the data was not normally distributed hence failing the normality test. Furthermore, with the number of datapoints being less than 25 (n< 25), we could not use the t-test to further support the null hypothesis.

The results of this test showed that there is no difference between the AQI in central stations vs non-central stations and that the AQI in central Calgary may even be less than or equal to that of the AQI recorded in the rest of the stations. This goes against our assumption that central Calgary may see a higher amount of pollution. The reason behind the result may be because, even though central Calgary, especially the downtown areas sees higher activity, most people prefer to make use of the public transit rather than driving in these areas, which automatically reduces the amount of pollution produced. These areas are also mostly occupied by large commercial buildings that are environmentally cautious which leads to a lower amount of pollution compared to that of individual residential households present in the south, north and east of the city.

Guiding Question C:

In this question, we wanted to check for any correlation between the major greenhouse gasses: Nitrogen Dioxide (NO2) and Nitric Oxide (NO) and Particulate Matter of 2.5 mass (PM2.5) against the population growth of Calgary. If any correlation is observed, we wanted to build a predictive model of the respective pollutants and population.

For PM 2.5 and population regression analysis, we found that there was a weak negative correlation. Based on the calculated p-value, we inferred that the population growth in Calgary ***cannot*** be expressed as a linear function of the yearly PM2.5 values recorded in Calgary. Therefore, further analysis on the predictive model was discontinued.

For Nitrogen Dioxide and population and Nitric Oxide and population regression analysis, we found that there was a strong negative correlation. Based on the calculated p-values for both the cases, we inferred that the population growth in Calgary ***can*** be expressed as a linear function of the yearly Nitrogen Dioxide and Nitric Oxide values recorded in Calgary. Therefore, to further validate the least-squares model, the correlation coefficients as well as the X (B) and Y(A) values were bootstrapped for the estimation of the model. For all the three values, the least square values were very close to the bootstrapped values and fell within the calculated 95% confidence intervals. This was the case for both Nitrogen Dioxide and Nitric Oxide.

The results obtained through the regression analysis makes sense when compared to what is actually happening to the NO2 and NO emissions in Calgary. Although the emissions were predicted to increase as the population level increased, we see that levels of NO2 and NO are actually decreasing in the past two decades due to improved technology at emission sources. This includes personal vehicles and industry. This may also have been due to increased environmental awareness in the public after the industrial revolution in the 1900s. However, despite the decrease, NO2 levels in urban areas are still on the high and exceed the new Canadian Ambient Air Quality Standards which are national standards set up to protect human health and environment.

**Next Steps**

To further analyze our chosen data sets, other parameters within our primary Historical Air Quality data set could be analyzed and visualized. Parameters other than pollutants such as wind speed, temperature, relative humidity etc can be used to create models and analyze the climate/weather patterns in Calgary in the past four decades. Furthermore, our linear models (equations) in question 3 could be used to do some prediction analysis for the future.

**References**

City of Calgary’s Open Data Portal *Historical Air Quality* [online]. Available at:<https://data.calgary.ca/Environment/Historical-Air-Quality/uqjm-jxgp/data>

WHO Air Quality Database (Update 2022), Available at: <https://www.who.int/data/gho/data/themes/air-pollution/who-air-quality-database>

Get the number of rows and columns in R. Available at: <https://www.digitalocean.com/community/tutorials/get-number-of-rows-and-columns-in-r>

Steven R.Hanna, 1989. Confidence limits for air quality model evaluations, as estimated by bootstrap and jackknife resampling methods. Available at:<https://www.sciencedirect.com/science/article/pii/0004698189901613> [Accessed October 4, 2022].

Lee-Ing Tong, Chih-Wei Chang, Shin-En Jin, R.Saminathan, 2012. Quantifying uncertainty of emission estimates in National Greenhouse Gas Inventories using bootstrap confidence intervals. Available at:<https://www.sciencedirect.com/science/article/pii/S1352231012003044> [Accessed October 4, 2022].

W. GeoffreyCobourn, 2010. An enhanced PM2.5 air quality forecast model based on nonlinear regression and back-trajectory concentrations. Available at:<https://www.sciencedirect.com/science/article/pii/S1352231010003821> [Accessed October 4, 2022].

Christopher Kennedy, Julia Steinberger, Barrie Gasson, Yvonne Hansen, Timothy Hillman, Miroslav Havránek, Diane Pataki, Aumnad Phdungsilp, Anu Ramaswami, and Gara Villalba Mendez, 2009. Greenhouse Gas Emissions from Global Cities. Available at:<https://pubs.acs.org/doi/full/10.1021/es900213p> [Accessed October 4, 2022].

Daniel Hoornweg, Lorraine Sugar, and Claudia Lorena Trejos Gómez, 2011. Cities and greenhouse gas emissions: moving forward. Available at:<https://journals.sagepub.com/doi/abs/10.1177/0956247810392270> [Accessed October 4, 2022].

J. Jason West, Steven J. Smith, Raquel A. Silva, Vaishali Naik, Yuqiang Zhang, Zachariah Adelman, Meridith M. Fry, Susan Anenberg, Larry W. Horowitz & Jean-Francois Lamarque, 2013. Co-benefits of mitigating global greenhouse gas emissions for future air quality and human health. Available at:<https://www.nature.com/articles/nclimate2009> [Accessed October 4, 2022].

Tome Eftimov & Peter Korošec, Chapter 4, Introduction to Statistical Analysis. Available at:<https://link.springer.com/chapter/10.1007/978-3-030-96917-2_4> and  [**https://rdcu.be/cWTZx**](https://rdcu.be/cWTZx)[Accessed October 4, 2022].

Paul S Levy and Kristine Stolte, 2000. Statistical methods in public health and epidemiology: a look at the recent past and projections for the next decade. Available at:<https://journals.sagepub.com/doi/abs/10.1177/096228020000900106> [Accessed October 4, 2022].

*Air indicators – nitrogen dioxide*. Alberta.ca. (n.d.). Retrieved October 17, 2022, from https://www.alberta.ca/air-indicators-nitrogen-dioxide.aspx