Final Project Report for ELE: 581

Name: Muhammad Enayetur Rahman
Student ID: 100635221
Data set: Airfoil Self-Noise Data Set

Part A:

Total 6 attributes are:

1. Frequency, in Hertz
2. Angle of attack, in degrees
3. Chord length in meters
4. Free-stream velocity, in meters per second
5. Suction side displacement thickness, in meters

And the final output is:

6. Scaled sound pressure level, in meters

Here is the basic statistical summary for each attribute. Attribute 'Sound level' would be my *final output regression attribute*.
In the summary, for each attribute minimum value, 1st Quartile, Median value, Mean value, 3rd Quartile value and Maximum value are given using R's command:

- *Summary(dataframe$attribute)*

Histogram of each attribute is found using R's command

- *hist(dataframe$attribute)*

Summary and Histogram of each attribute:

Attribute: Frequency

frequency
Min. : 200
1st Qu.: 800
Median : 1600
Mean : 2886
3rd Qu.: 4000
Max. : 20000

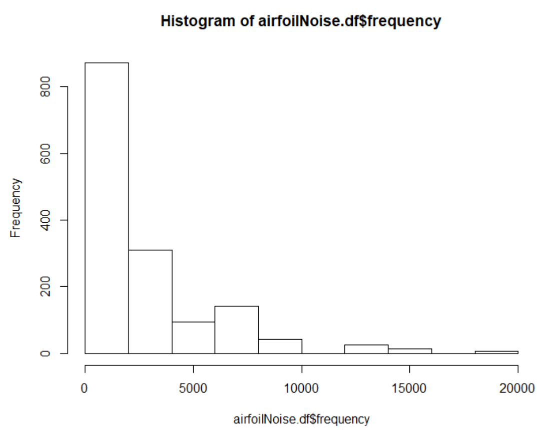


Figure 1: Histogram of Attribute Frequency

Attribute: Angle of attack

angle_of_attack
Min. :0.000
1st Qu.:2.000
Median :5.400
Mean :6.782
3rd Qu.:9.900
Max. :22.200

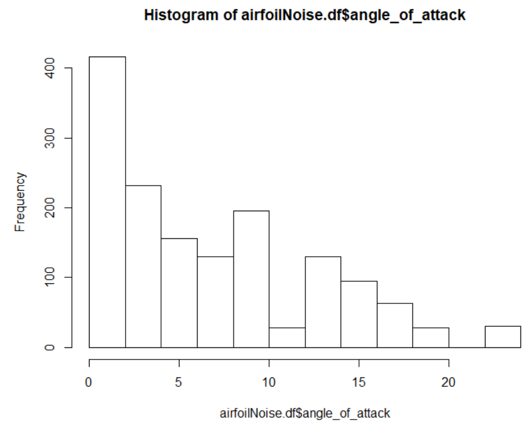


Figure 2: Histogram of Attribute Angle of Attack

Attribute: Chord Length

chord_length
Min. : 0.0254
1st Qu.: 0.0508
Median : 0.1016
Mean : 0.1365
3rd Qu.: 0.2286
Max. : 0.3048

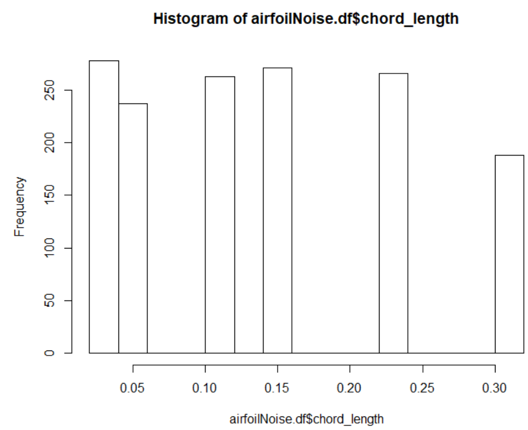


Figure 3: Histogram of Attribute Chord Length

Attribute: Free Stream Velocity

free_stream_velocity
Min. : 31.70
1st Qu.: 39.60
Median : 39.60
Mean : 50.86
3rd Qu.: 71.30
Max. : 71.30

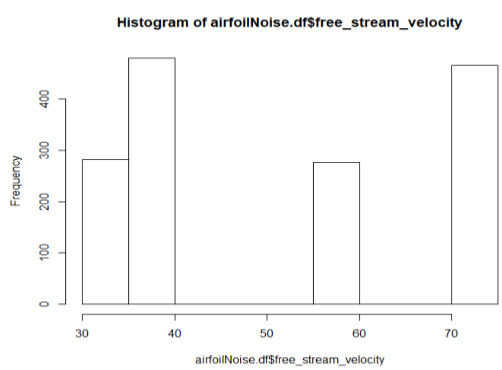


Figure 4: Histogram of Attribute Free stream velocity

Attribute: Displacement Thickness

displacement_thickness
Min. : 0.0004007
1st Qu.: 0.0025351
Median : 0.0049574
Mean : 0.0111399
3rd Qu.: 0.0155759
Max. : 0.0584113

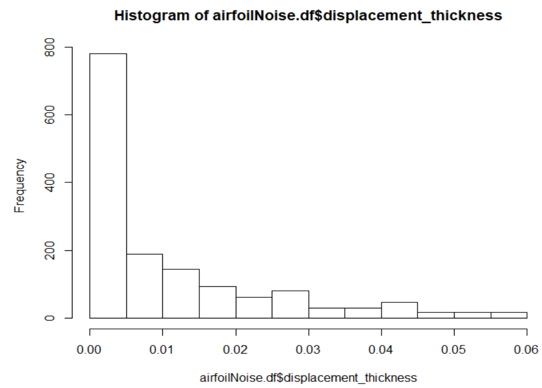


Figure 5: Histogram of Attribute Displacement Thickness

Attribute: Sound Level

sound_level
Min. : 103.4
1st Qu.: 120.2
Median : 125.7
Mean : 124.8
3rd Qu.: 130.0
Max. : 141.0

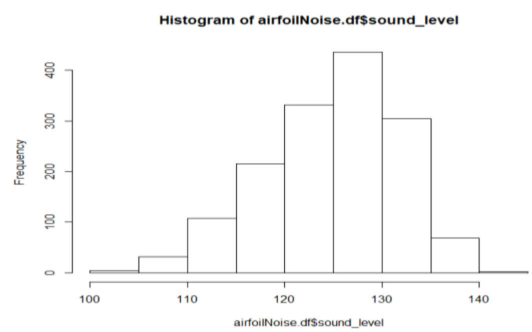


Figure 6: Histogram of Attribute Sound Level

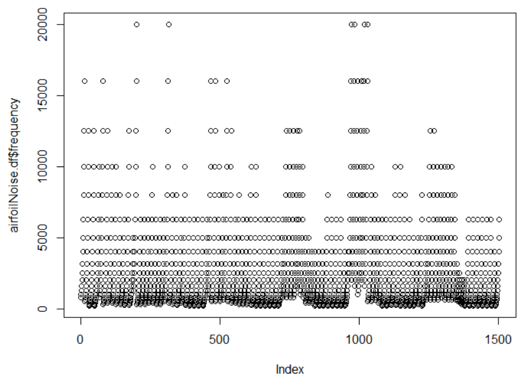
Graphs of each independent variable:

Graph of each attribute is found using R's command

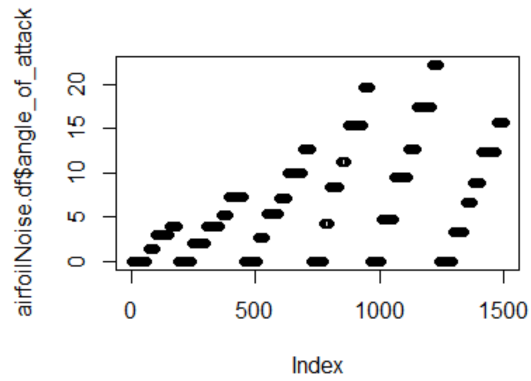
- *Plot(dataframe$attribute)*

The graphs of all of the attributes are given below:

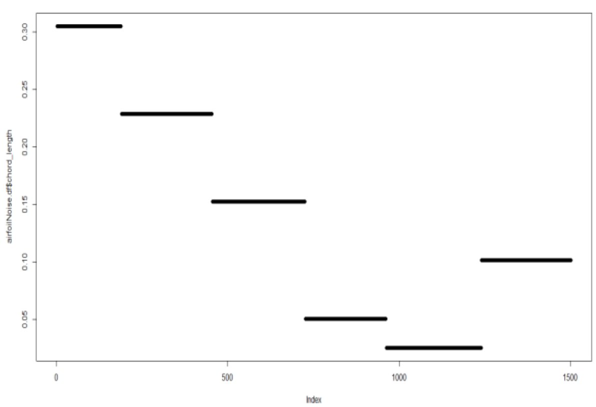
Frequency



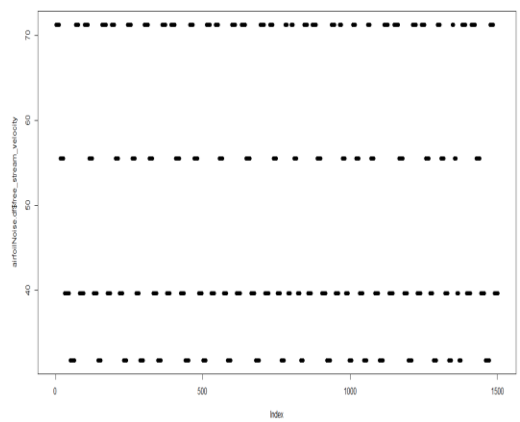
Angle of Attack



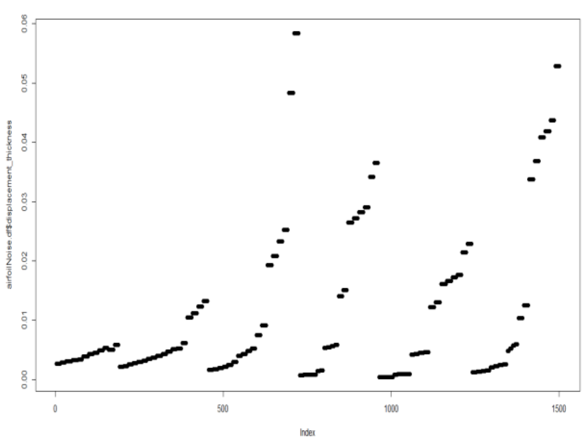
Chord Length



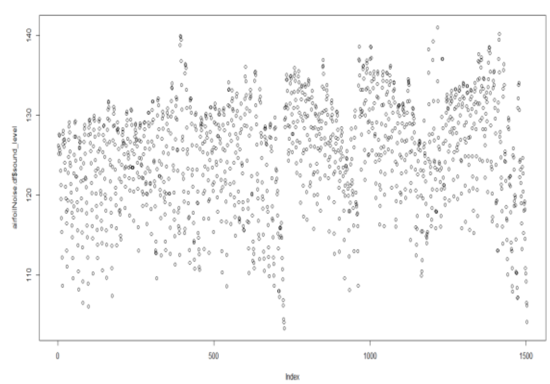
Free stream velocity



Displacement Thickness



Sound level



Part B:

For “Linear” kernel the following table information is used: Gamma = 0.2.

ID	Kernel	Cost	Epsilon	Root mean Squared Value (RMSE)
1	Linear	0.001	0.1	5.366429
2	Linear	0.01	0.1	4.83299
3	Linear	0.01	0.4	4.817667
4	Linear	0.1	0.1	4.872883
5	Linear	1	0.1	4.876625
6	Linear	1	0.5	4.816702
7	Linear	1	0.6	4.811977
8	Linear	10	0.1	4.876962
9	Linear	10	0.5	4.816691
10	Linear	100	.1	4.876561
11	Linear	100	.5	4.816847

Code attached (File: regression_linear.R)

For polynomial Kernel: Gamma = 0.2

ID	Kernel	Cost	Degree	Epsilon	Root mean Squared Value (RMSE)
1	Polynomial	10	2	0.1	5.627001
2	Polynomial	10	2	0.8	5.545119
3	Polynomial	10	3	0.1	4.400711
4	Polynomial	10	3	0.3	4.394331
5	Polynomial	100	2	0.1	5.633465
6	Polynomial	100	3	0.1	4.399498
7	Polynomial	100	3	0.3	4.392198
8	Polynomial	100	3	0.9	4.488404

Code attached (File: regression_poly.R)

For **Radial** kernel: Gamma = 0.2

ID	Kernel	Cost	Epsilon	Root mean Squared Value (RMSE)
1	Radial	1	0.1	3.087114
2	Radial	1	0.9	3.81452
3	Radial	10	0.1	2.57082
4	Radial	10	0.5	2.798251
5	Radial	10	0.9	3.572354
6	Radial	100	0.1	2.168457

Code attached (File: regression_radial.R)

Two best models are from Radial Kernel: Considering other kernel's complexity and Root Mean Squared Error (RMSE).

ID	Kernel	Cost	Epsilon	Root mean Squared Value (RMSE)
1	Radial	1	0.1	3.087114
3	Radial	10	0.1	2.57082

Part C:

Using the Bootstrap, hold-out method with my top two models and 95% confidence interval, the lower and upper bound errors are given as follows:

Best Models (Radial)	**Lower Bound Error**	**Upper Bound Error**
Model ID: 1 (Kernel=Radial, Cost=1, Epsilon=0.1)	2.901717	3.585302
Model ID: 3 (Kernel=Radial, Cost=10, Epsilon=0.1)	2.465413	3.138644

Code attached (File: regression_bootstrap.R)

1. For a total of 1000 Bootstrap samples with the 95% error confidence interval, the lower bound is **2.5th percentile** which is: **25th value** of the corresponding error array and the upper bound is **97.5th percentile** which is: **975th** value of the corresponding error array.
   For Model ID: 1, the 95% confidence interval is: [2.901717, 3.585302].
   For Model ID: 6, the 95% confidence interval is: [2.465413, 3.138644]

2. Two models are: f_{D1} = [kernel = Radial, C = 1, Epsilon = 0.1] with 95% confidence interval **[2.901717, 3.585302]**, RMSE = 3.087114 and f_{D2} = [kernel = Radial, C = 10] with the 95% confidence interval **[2.465413, 3.138644]**, RMSE = 2.57082

   we found: model ID: 3's confidence interval is completely overlapped with the model ID: 1's confidence interval. As a result, the performance of these two models are **not significantly** different.

   To select the best models among these two, it is needed to see the other parameters besides RMSE. So, complexity is considered here. As a result, model ID: 1 is which has Cost = 1, Epsilon = 0.1 is much less complex than model ID: 3, which is Cost = 100, Epsilon = 0.1. So, model ID: 1 ie. **fD1 = [kernel = Radial, cost = 1, Epsilon = 0.1]** is selected.