Table of Contents

Introduction

The objectives are to create regression models for Actual Temperature and Apparent Temperature using data discussed in a previous report.

Correlations

Before creating Regression Models for the variables of interest in the dataset, we should investigate if there are relationships between the continuous variables in the sample dataset. The correlation between two variables measures the linear relationship between them. The section below aims to investigate the correlation between continuous variables in pairs (the modified sample dataset) and determine if the relationship in the population are statistically significant.

Actual Temperature and Apparent Temperature

From the modified sample dataset obtained from Part I, it was observed that both variables have a normal distribution hence, the Pearson correlation test will be used. Below is the R output

>

```
Pearson's product-moment correlation
data:  Actual Temperature (C) and Apparent Temperature (C)
t = 120.66, df = 198, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.991110 0.994905
sample estimates:
     cor
0.993269
```

- The above output shows the Person correlation coefficient (calculated for the sample data) r = 0.993, which indicates that there is a strong positive relationship between Actual temperature and Apparent temperature.
- At a 95% confidence level, the correlation between the variables in the population is between 0.991 and 0.995.

Additionally, a hypothesis test to test if the correlation between the variables is statistically significant is carried out:

$$H_0: \text{Correlation} = 0$$

$$H_1: \text{Correlation} \neq 0$$

- It shows a p-value < 0.00.1therefore, there is sufficient evidence to reject the null hypothesis at a 5% level of significance as there is overwhelming evidence against H_0.
- This indicates that the test is highly significant and the correlation between the variables is not equals to 0.
- We can conclude that there is a relationship between the variables.

Actual Temperature and Humidity

As observed from Part I, the Humidity variable is not of a normal distribution as it appears to be negatively skewed therefore the Spearman Rho correlation test will be used to investigate any relationship between the two variables.

>

```
Spearman's rank correlation rho
data:  Actual Temperature (C) and Humidity
S = 2032899, p-value = 1.52e-15
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.5247122
```

- The above output shows the Spearman's Rho correlation (calculated for the sample data) r = -0.525, which indicates that there is a negative relationship between Actual Temperature and Humidity.

A hypothesis test to test if the correlation between the variables is statistically significant is carried out:

H_0: Correlation = 0

H_1: Correlation ≠ 0

- It shows a p-value < 0.001 which indicates that there is sufficient evidence to reject the null hypothesis at a 5% level of significance as there is overwhelming evidence against H_0.
- This indicates that the test is highly significant and the correlation between the variables is not equals to 0.
- We can conclude that there is a relationship between the variables.

Actual Temperature and Wind Speed

As observed from Part I, the Wind Speed variable is not of a normal distribution as it appears to be positively skewed therefore the Spearman Rho correlation test will be used to investigate any relationship between the two variables

>

```
        Spearman's rank correlation rho
data:  Actual Temperature (C) and Wind Speed (km/h)
S = 1178745, p-value = 0.1021
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.1159191
```

- The above output shows the Spearman's Rho correlation (calculated for the sample data) r = 0.116,

A hypothesis test to test if the correlation between the variables is statistically significant is carried out:

H_0: Correlation = 0

H_1: Correlation ≠ 0

- It shows a p-value of 0.1021 which indicates that there isn't sufficient evidence to reject the null hypothesis at a 5% level of significance as the p-value is greater than the 5% significance level.
- This indicates that the test is not significant and the correlation between the variables in the population is equals to 0.
- We can conclude that there is no relationship between the variables.

Actual Temperature and Wind Bearing

From Part I, the Wind Bearing variable is not of a normal distribution therefore the Spearman Rho correlation test will be used to investigate the relationship between the variables

>

```
        Spearman's rank correlation rho
data:  Actual Temperature (C) and Wind Bearing (degrees)
S = 1305729, p-value = 0.7713
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.02067843
```

- The above output shows the Spearman's Rho correlation rho (r) = 0.021which indicates there is no relationship between the variables.

A hypothesis test to test if the correlation between the variables is statistically significant is carried out:

$$H_0: \text{Correlation} = 0$$
$$H_1: \text{Correlation} \neq 0$$

- It shows a p-value of 0.771 which means that we accept the null hypothesis as the value is greater than the 5% significance level this indicates overwhelming evidence to accept the null hypothesis.
- This indicates that the test is not significant and the correlation between the variables in the population is equals to 0.
- We can conclude that there is no relationship between the variables.

Actual Temperature and Visibility

As observed from Part I, the Visibility variable is not of a normal distribution therefore the Spearman Rho correlation test will be used to investigate the relationship between the variables

>

```
        Spearman's rank correlation rho
data:  Actual Temperature (C) and Visibility (km)
S = 675303, p-value = 1.128e-13
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.4935103
```

- The above output shows the Spearman's Rho correlation r = 0.494. This indicates a positive relationship between the variables.

A hypothesis test to test if the correlation between the variables is statistically significant is carried out:

$$H_0: \text{Correlation} = 0$$

$$H_1: \text{Correlation} \neq 0$$

- It shows a p-value < 0.001 which indicates that there is sufficient evidence to reject the null hypothesis at a 5% level of significance as there is overwhelming evidence against H_0.
- This indicates that the test is highly significant and the correlation between the variables is not equals to 0.
- We can conclude that there is a relationship between the variables.

Actual Temperature and Pressure

From the modified sample dataset obtained from Part I, the distribution is a normal distribution hence, the Pearson correlation test will be used to investigate the relationship between the variables

>

Pearson's product-moment correlation

data: Actual Temperature (C) and Pressure (millibars)

t = -4.9583, df = 198, p-value = 1.525e-06

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.4503185 -0.2029587

sample estimates:

cor

-0.3323414

- The above output shows the Person correlation coefficient (calculated for the sample data) r = -0.332, which indicates that there is a negative relationship between the variables.
- At a 95% confidence level, the correlation between the variables in the population is between -0.45 and -0.20.

A hypothesis test to test if the correlation between the variables is statistically significant is carried out:

H_0: Correlation = 0

H_1: Correlation $\neq$ 0

- It shows a p-value < 0.001 which indicates that there is sufficient evidence to reject the null hypothesis at a 5% level of significance as there is overwhelming evidence against H_0.
- This indicates that the test is highly significant and the correlation between the variables is not equals to 0.
- We can conclude that there is a relationship between the variables.

Apparent Temperature and Humidity

As observed from Part I, the Humidity variable is not of a normal distribution as it appears to be negatively skewed therefore the Spearman Rho correlation test will be used to investigate any relationship between the two variables

>

Spearman's rank correlation rho

data: Apparent Temperature (C) and Humidity
S = 2004687, p-value = 2.958e-14
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.5035526

- The above output shows the Spearman's Rho correlation r = -0.504. This indicates a negative relationship between the variables.

A hypothesis test to test if the correlation between the variables is statistically significant is carried out:

H_0: Correlation = 0

H_1: Correlation ≠ 0

- It shows a p-value < 0.001 which indicates that there is sufficient evidence to reject the null hypothesis at a 5% level of significance as there is overwhelming evidence against H_0.
- This indicates that the test is highly significant and the correlation between the variables is not equals to 0.
- We can conclude that there is a relationship between the variables.

Apparent Temperature and Wind Speed

As observed from Part I, the Wind Speed variable is not of a normal distribution as it appears to be positively skewed therefore the Spearman Rho correlation test will be used to investigate any relationship between the two variables

>

Spearman's rank correlation rho
data: Apparent Temperature (C) and Wind Speed (km/h)
S = 1244802, p-value = 0.3504
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.06637514

- The above output shows the Spearman's Rho correlation r = 0.07. This indicates there is no relationship between the variables.

A hypothesis test to test if the correlation between the variables is statistically significant is carried out:

$$H_0: \text{Correlation} = 0$$
$$H_1: \text{Correlation} \neq 0$$

- It shows a p-value of 0.35 which means that we accept the null hypothesis as the value is greater than the 5% significance level this indicates overwhelming evidence to accept the null hypothesis.
- This indicates that the test is insignificant and the correlation between the variables in the population is equals to 0.
- We can conclude that there is no relationship between the variables.

Apparent Temperature and Wind Bearing

From Part I, the Wind Bearing variable is not of a normal distribution therefore the Spearman Rho correlation test will be used to investigate the relationship between the variables

>

```
Spearman's rank correlation rho
data:  Apparent Temperature (C) and Wind Bearing (degrees)
S = 1308072, p-value = 0.7903
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.01892135
```

- The above output shows the Spearman's Rho correlation r = 0.02. This indicates there is no relationship between the variables.

A hypothesis test to test if the correlation between the variables is statistically significant is carried out:

$$H_0: \text{Correlation} = 0$$

H_1: Correlation $\neq$ 0

- It shows a p-value of 0.79 which means that we accept the null hypothesis as the value is greater than the 5% significance level this indicates overwhelming evidence to accept the null hypothesis.
- This indicates that the test is not significant and the correlation between the variables in the population is equals to 0.
- We can conclude that there is no relationship between the variables.

Apparent Temperature and Visibility

As observed from Part I, the Visibility variable is not of a normal distribution therefore the Spearman Rho correlation test will be used to investigate the relationship between the variables

>

```
Spearman's rank correlation rho
data:  Apparent Temperature (C) and Visibility (km)
S = 711007, p-value = 3.257e-12
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.4667317
```

- The above output shows the Spearman's Rho correlation r = 0.467. This indicates a positive relationship between the variables.

A hypothesis test to test if the correlation between the variables is statistically significant is carried out:

H_0: Correlation = 0

H_1: Correlation $\neq$ 0

- It shows a p-value < 0.001 which indicates that there is sufficient evidence to reject the null hypothesis at a 5% level of significance as there is overwhelming evidence against H_0.
- This indicates that the test is highly significant and the correlation between the variables is not equals to 0.

- We can conclude that there is a relationship between the variables.

Apparent Temperature and Pressure

From the modified sample dataset obtained from Part I, hence, the Pearson correlation test will be used to investigate the relationship between the variables

>

```
Pearson's product-moment correlation
data:  Apparent Temperature © and Pressure (millibars)
t = -4.6916, df = 198, p-value = 5.048e-06
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.4359106 -0.1857074
sample estimates:
      cor
-0.3162989
```

- The above output shows the Person correlation coefficient (calculated for the sample data) r = -0.316, which indicates that there is a negative relationship between Actual temperature and Pressure.
- At a 95% confidence level, the correlation for the variables in the population is between -0.436 and -0.186.

Additionally, a hypothesis test to test if the correlation between the variables is statistically significant is carried out:

$$H_0: \text{Correlation} = 0$$
$$H_1: \text{Correlation} \neq 0$$

- It shows a p-value < 0.00.1therefore, there is sufficient evidence to reject the null hypothesis at a 5% level of significance as there is overwhelming evidence against H_0.
- This indicates that the test is highly significant and the correlation between the variables is not equals to 0.
- We can conclude that there is a relationship between the variables.

Regression Models for Actual Temperature

In establishing a regression model to predict or forecast the Actual Temperature (which is the dependent variable) we need to use all the other variables as the independent variables and with the use of model selection procedure, select the best model for predicting the variability in Actual Temperature.

The independent variables are:

X_1 = Humidity

X_2 = Visibility

X_3 = Pressure

X_4 = Wind Speed

X_5 = Wind Bearing

X_6 = Modified Summary (categorical variable with four categories; clear, cloudy, foggy, and overcast)

X_7 = Precipitation Type (categorical variable with two categories; rain and snow)

The regression equation is

$$Y = \beta_0 + \beta_{Humidity}X_1 + \beta_{Visibility}X_2 + \beta_{Pressure}X_3 + \beta_{Wind\ Speed}X_4 + \beta_{Wind\ Bearing}X_5 + \beta_{Modified\ Summary}X_6 + \beta_{Precipitation\ Type}X_7$$

Model Selection Procedure

Using R, the coefficients of the regression equation is:

Actual Temperature = 249.211 – 24.397(Humidity) + 0.405(Visibility) – 0.215(Pressure) – 0.343(Wind Speed) – 0.004(Wind Bearing) + 0.771Cloudy + 1.917Foggy + 0.079Overcast – 11.483Snow

- The multiple R^2 value is 66% which indicates that 66% of the variability in Actual Temperature can be explained by all the independent variables.
- The Adjusted R^2 value is 64.4% (for the entire population).
- We can conclude that the model is a reasonable fit.
- However, based on the output of the test (F-statistic) not all the independent variables are statistically relevant to the model.

To achieve a model with relevant independent variables (a parsimonious description of the data), we will use the following model selection procedures.

Forward Selection Model Procedure

Using the forward selection model procedure, the coefficients of the regression equation is:

Actual Temperature = 241.319 – 11.642Snow – 24.748(Humidity) – 0.347(Wind Speed) – 0.206(Pressure) + 0.35(Visibility)

The adjudged relevant or useful independent variables using the forward selection model are Precipitation Type, Humidity, Wind Speed, Pressure, and Visibility.

- Using the forward selection procedure, five out of the seven independent variables are needed to predict the Actual Temperature
- The multiple R^2 value is 65.6% which indicates that about 66% of the variability in Actual Temperature can be explained by the independent variables.
- The Adjusted R^2 value is 64.74% (for the entire population).
- Based on the F-statistic test, a p-value < 0.001 was obtained from the model which indicates that it is statistically significant
- We can conclude that the model is a reasonable fit as it produces a model with fewer independent variables with a similar R^2 value as compared with the model with all independent variables.

Backward Elimination Selection Model Procedure

Using the backward elimination selection model, the coefficients of the regression equation is:

Actual Temperature = 241.319 – 24.748(Humidity) + 0.35(Visibility) – 0.206(Pressure) – 0.347(Wind Speed) – 11.642Snow

- Similar to the forward selection model, the model obtained from the backward elimination selection procedure has a multiple R^2 value of 65.6% which indicates that about 66% of the variability in Actual Temperature can be explained by the independent variables.
- The Adjusted R^2 value is 64.74% (for the entire population).
- Based on the F-statistic test, a p-value < 0.001 was obtained from the model which indicates that it is statistically significant
- We can conclude that the model is a reasonable fit as it produces a model with fewer independent variables with a similar R^2 value as compared with the model with all independent variables.

Stepwise Selection Model Procedure

Using the stepwise selection model, the coefficients of the regression equation is

$$Actual\ Temperature = 241.319 - 11.642Snow - 24.748(Humidity) - 0.347(Wind\ Speed) - 0.206(Pressure) + 0.35(Visibility)$$

- The Stepwise Selection model produce the same coefficients, Multiple R^2 and Adjusted R^2 as both the Forward Selection and Backward Elimination Selection model
- Additionally, based on the F-statistic test, a p-value < 0.001 was obtained from the model which indicates that it is statistically significant
- Therefore, we can conclude that the model is equally a good fit as the two previous models.

Model Evaluation

Because all three model selection methods produced the same result during the model selection procedure any of the model selection procedure can be chosen as the final regression model for predicting Actual Temperature, we can conclude with high confidence that the regression equation for predicting the Actual Temperature is:

$$Actual\ Temperature = 241.319 - 11.642Snow - 24.748(Humidity) - 0.347(Wind\ Speed) - 0.206(Pressure) + 0.35(Visibility)$$

NB: To make a prediction when the "Precipitation type" is rain the regression equation for predicting the Actual Temperature is:

$$Actual\ Temperature_{rain} = 241.319 - 24.748(Humidity) - 0.347(Wind\ Speed) - 0.206(Pressure) + 0.35(Visibility)$$

Furthermore, we can conclude that the model is a reasonable fit as it produces a model with fewer independent variables with a R^2 value of 65.6% which is slightly close to the model with all the independent variables with a R^2 value of 66% as it gives a more parsimonious description of the data. This indicates that about 66% of the variability in Actual Temperature can be explained by the selected model. However, based on the p-values of the independent variables, the relevance of the "Visibility" variable to the model is somewhat questionable. Additionally, the F-statistic test shows a p-value < 0.001, which indicates that the model is statistically significant.

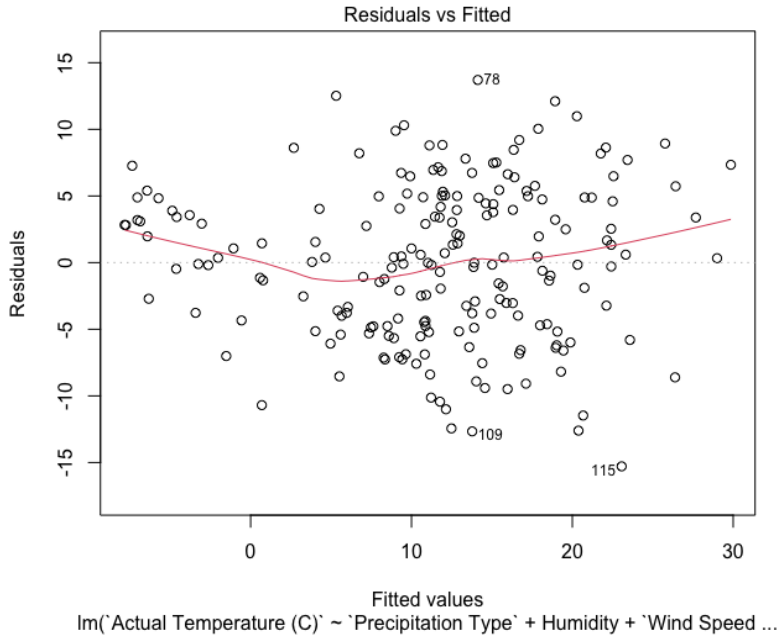
Assumptions

Given that the fitted regression equation is such that the sum of the squared residuals (errors) is a minimum, the residuals are subject to the following assumptions (the assumptions are strictly on the residuals and not on the entire dataset):

- It is normally distributed with a mean of 0
- It has constant variance (i.e., homoscedasticity)

We therefore check the above assumptions of our selected model with the use of a "Residual vs Fitted" Plot and a Q-Q Plot.

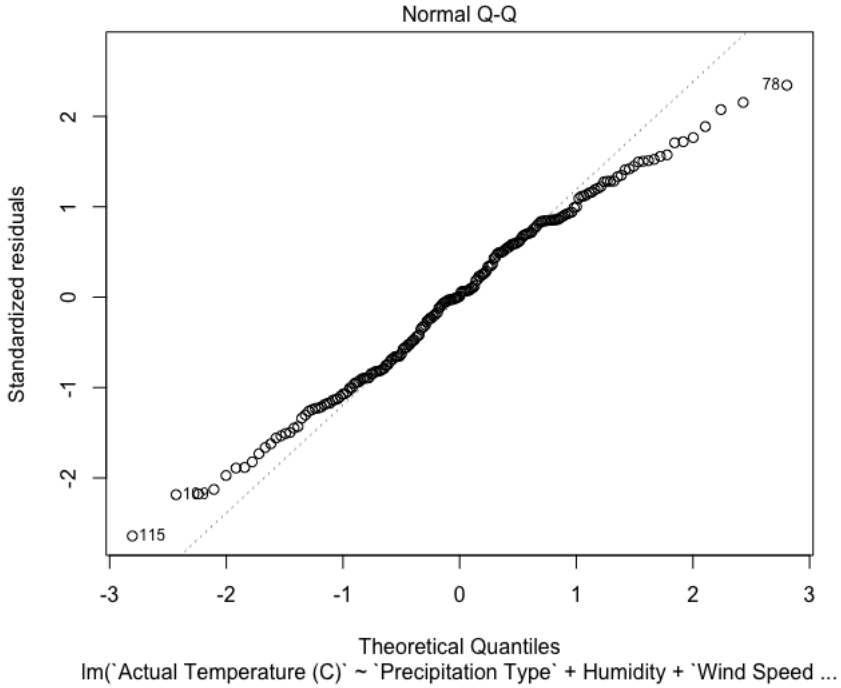
"Residual vs Fitted" Plot



The "Residuals vs Fitted" scatter plot above shows the points of the residual value against the fitted values. From the scatter plot above the following can be observed:

- A random scatter around the mean of 0
- There is an even number of scatter of points from left to right which indicates homoscedasticity (constant variance)
- There are no obvious outliers in the dataset as the points (109 and 115) does not appear to be far removed from other values.

"Q-Q" Plot



The "Normal Q-Q" plot above shows most of the points on a diagonal line which indicates that it is of a normal distribution.

In conclusion, based on the observations from the graph/plot, our assumptions of the regression model are true.

Regression Models for Apparent Temperature

As shown in the previous section, creating a regression model to predict, or forecast the Apparent Temperature (which is the dependent variable) we need to use all the other variables as the independent variables and with the use of model selection procedure, select the best model for predicting the variability in Apparent Temperature.

The independent variables are:

X_1 = Humidity

X_2 = Visibility

X_3 = Pressure

X_4 = Wind Speed

X_5 = Wind Bearing

X_6 = Modified Summary (categorical variable with four categories; clear, cloudy, foggy, and overcast)

X_7 = Precipitation Type (categorical variable with two categories; rain and snow)

The regression equation is

$$Y = \beta_0 + \beta_{Humidity}X_1 + \beta_{Visibility}X_2 + \beta_{Pressure}X_3 + \beta_{Wind\ Speed}X_4 + \beta_{Wind\ Bearing}X_5 + \beta_{Modified\ Summary}X_6 + \beta_{Precipitation\ Type}X_7$$

Model Selection Procedure

Using R, the coefficients of the regression equation is:

Apparent Temperature = 263.404 – 26.257(Humidity) + 0.417(Visibility) – 0.227(Pressure) – 0.461(Wind Speed) – 0.005(Wind Bearing) + 0.917Cloudy + 2.342Foggy + 0.053Overcast – 13.288Snow

- The multiple R^2 value is 63.94% which indicates that about 64% of the variability in Apparent Temperature can be explained by all the independent variables.
- The Adjusted R^2 value is 62.2% (for the entire population).
- We can conclude that the model is a reasonable fit.
- However, based on the output of the test (F-statistic) not all the independent variables are relevant to the model.

Again, to achieve a model with relevant independent variables, we will use the following model selection procedures.

Forward Selection Model Procedure

Using the forward selection model, the coefficients of the regression equation is:

Apparent Temperature = 253.572 – 13.483Snow – 26.691(Humidity) – 0.465(Wind Speed) – 0.216(Pressure) + 0.349(Visibility)

- The multiple R^2 value is 63.5% which indicates that about 64% of the variability in Apparent Temperature can be explained by the independent variables (Precipitation Type, Humidity, Wind Speed, Pressure, and Visibility).
- The Adjusted R^2 value is 62.55% (for the entire population).
- Based on the F-statistic test, a p-value < 0.001 was obtained from the model which indicates that it is statistically significant
- We can conclude that the model is a reasonable fit as it produces a model with fewer independent variables with a similar R^2 value as compared with the model with all independent variables.

Backward Elimination Selection Model Procedure

Using the backward elimination selection model, the coefficients of the regression equation is:

Apparent Temperature = 253.72 – 26.69(Humidity) + 0.349(Visibility) – 0.216(Pressure) – 0.465(Wind Speed) – 13.483Snow

- Similar to the forward selection model, the model obtained using the backward elimination model selection procedure produces a multiple R^2 value of 63.5% which indicates that about 664% of the variability in Apparent Temperature can be explained by the independent variables.
- The Adjusted R^2 value is 62.55% (for the entire population).
- Based on the F-statistic test, a p-value < 0.001 was obtained from the model which indicates that it is statistically significant
- We can conclude that the model is a reasonable fit as it produces a model with fewer independent variables with a similar R^2 value as compared with the model with all independent variables.

Stepwise Selection Model Procedure

Using the stepwise selection model procedure, the coefficients of the regression equation is

Apparent Temperature = 253.72 – 13.483Snow – 26.69(Humidity) – 0.465(Wind Speed) – 0.216(Pressure) + 0.349(Visibility)

- Similar to both the forward selection and the backward elimination model selection procedures, this model also produces a multiple R^2 value of 63.5% which indicates that about 64% of the variability in Apparent Temperature can be explained by the independent variables.
- The Adjusted R^2 value is 62.55% (for the entire population).
- Based on the F-statistic test, a p-value < 0.001 was obtained from the model which indicates that it is statistically significant
- We can conclude that the model is a reasonable fit as it produces a model with fewer independent variables with a similar R^2 value as compared with the model with all independent variables.

Model Evaluation

Because all three model selection methods produced the same result during the model selection procedure any of the model selection procedures can be chosen as the final regression model for predicting Apparent Temperature, we can conclude with high confidence that the regression equation for predicting the Apparent Temperature is:

Apparent Temperature = 253.72 – 13.483Snow – 26.69(Humidity) – 0.465(Wind Speed) – 0.216(Pressure) + 0.349(Visibility)

NB: To make a prediction when the "Precipitation type" is rain the regression equation for predicting the Apparent Temperature is:

Apparent Temperature $_{rain}$ = 253.72 – 26.69(Humidity) – 0.465(Wind Speed) – 0.216(Pressure) + 0.349(Visibility)

Furthermore, we can also conclude that the model is a reasonable fit as it produces a model with fewer independent variables with a R^2 value of 63.5% which is slightly close to the model with all the independent variables with a R^2 value of 63.94% as it gives a more parsimonious description of the data. This indicates that the selected model can explain about 64% of the

variability in Apparent Temperature. However, based on the p-values of the independent variables, the relevance of the "Visibility" variable to the model is somewhat questionable. Additionally, the F-statistic test shows a p-value < 0.001, which indicates that the regression model is statistically significant.

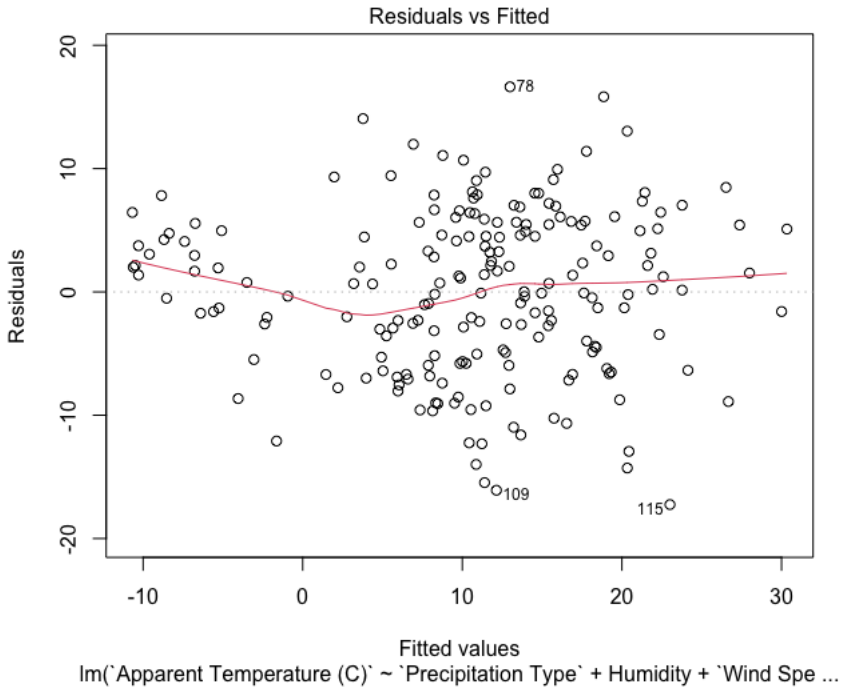
Assumptions

Given that the fitted regression equation is such that the sum of the squared residuals (errors) is a minimum, the residuals are subject to the following assumptions (the assumptions are strictly on the residuals not on the entire dataset):

- It is normally distributed with a mean of 0
- It has constant variance (i.e., homoscedasticity)

We therefore check the above assumptions of our selected model with the use of a "Residual vs Fitted" Plot and a Q-Q Plot.

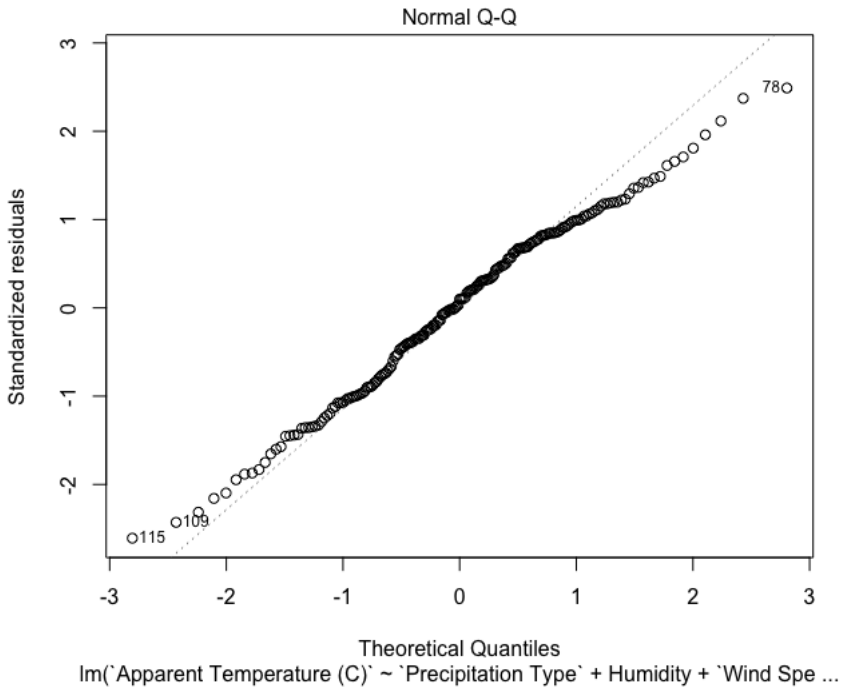
"Residual vs Fitted" Plot



The "Residuals vs Fitted" scatter plot above shows the points of the residual value against the fitted values. From the scatter plot above the following can be observed:

- A random scatter around the mean of 0
- There is an even number of scatter of points from left to right which indicates homoscedasticity
- There are no obvious extreme outliers in the dataset

"Q-Q" Plot



The "Normal Q-Q" plot above shows most of the points on a diagonal line which indicates that it is of a normal distribution

In conclusion, based on the observations from the graph/plot, our assumptions of the regression model are true.

Summary

Decision making based on the investigation and analysis of the sample data obtained from a from a larger database published on Kaggle (www.kaggle.com) that contains weather data from Szeged, a city in Hungary over a 10-year period that spans from 2006 to 2016 which amounts to 96,454 observations. The randomly selected sample data contains 200 observations.

As the main objective of the report is to identify which variables relate to (or causes) Actual Temperature and Apparent Temperature, essentially, a regression model that can forecast the Actual Temperature and the Apparent Temperature is obtained from the sample dataset.

Firstly, we measure the relationships (correlation) between the variables of interest (Actual Temperature and Apparent Temperature) with the other continuous variables in the dataset and test if the relationship between the variables in the population are statistically significant. It was observed that the same set of variables that are related to Actual Temperature are also related to Apparent Temperature.

Next, we create a regression model using all the independent variables in the dataset to forecast the Actual temperature however, using the model selection procedures, the final regression model with fewer independent variables for predicting Actual Temperature obtained is:

$$Actual\ Temperature = 241.319 - 11.642Snow - 24.748(Humidity) - 0.347(Wind\ Speed) - 0.206(Pressure) + 0.35(Visibility)$$

As all three model selection procedures produced the same regression equation for predicting the Actual Temperature with a R^2 value of 65.6% which indicates that about 66% of the variability in Actual Temperature can be explained by the model above.

Similarly, for Apparent Temperature, using model selection procedures, the final regression model with fewer independent variables for predicting Apparent Temperature obtained is:

$$Apparent\ Temperature = 253.72 - 13.483Snow - 26.69(Humidity) - 0.465(Wind\ Speed) - 0.216(Pressure) + 0.349(Visibility)$$

Again, all three model selection procedures produced the same regression equation for predicting the Apparent Temperature with a R^2 value 63.5% which indicates that about 64% of the variability in Apparent Temperature can be explained by the model above.

Lastly, an assumption for the models obtained for both Actual Temperature and Apparent Temperature are checked. The assumptions made are:

- It is normally distributed with a mean of 0 and
- It has constant variance (i.e., homoscedasticity)

The above assumptions about the models obtained for both Actual and Apparent Temperature were confirmed true with the use of a "Residual vs Fitted" and a "Q-Q" plot.

In conclusion, there are many similarities observed between the variables of interest (Actual Temperature and Apparent Temperature) such as the independent variables used in the regression model owing likely to the nature of the variable. The regression models obtained for each of the variables of interest accounts for two-third (2/3) when predicting or forecasting the Actual and Apparent Temperature respectively.