Exploring ONI as a Factor in Severe and Non-Severe Cyclones

By Jackson Downs

Overview

The topic of predicting severe and non-severe cyclones is of great importance to the Australian Community. This is because around 85% of Australians live within 50 km of the sea[1]. If we can successfully predict when a cyclone is occurring, we can safely evacuate the community with time to spare. Additionally, if we can predict whether the cyclones are severe or not, we can make sure we aren't wasting time and resources for evacuating people when the storm isn't as dangerous as we believe.

In this introduction, we believe it's also important to identify what makes a cyclone severe versus non-severe. If the sustained wind force reaches hurricane forces of at least 118 km/h, the system is defined as a severe tropical cyclone[2]. Other words to describe cyclones may be hurricane or typhoons. For the difference between severe and non-severe cyclones, we can look at the wind force.

In my review for past writing on Australian cyclones, I came across a few interesting articles. One looked at parametric wind fields, which is a long the basis of what we're looking for (Article on Google Scholar: Mapping the Wind Hazard of Global Tropical Cyclones with Parametric Wind Field Models by Considering the Effects of Local Factors: Looks at wind fields of historical tropical cyclones.). However, the article didn't focus specifically on Australian cyclones, which is what I was looking for.

As the previous academic article mentioned in its abstract, there are very few articles on Australian cyclone factors. Additionally, I couldn't find any past academic articles talking about the comparison between Australian cyclones and the Ocean Nino Index, or ONI.

Due to the lack of previous work with ONI being a factor in Australian cyclones, I will conduct an analysis to see if I can find a model with ONI that can successfully predict severe and non-severe cyclones. This will be important to a majority of the Australian community concerned about cyclones' affecting their daily lives and the Australian government in charge of notifying and evacuating citizens.

Data Characteristics

Before we begin our analysis, we first want to explore our data

```
dim(cyclones)

## [1] 37  8
```

[1] Australia State of the Environment Report. "Coasts." *Australia State of the Environment Report*, Australia State of the Environment Report, 25 Nov. 2021, soe.environment.gov.au/theme/coasts.

[2] "Severe Wind." *Severe Wind | Geoscience Australia*, Australian Government, www.ga.gov.au/scientific-topics/community-safety/severe-wind#:~:text=A%20tropical%20cyclone%20is%20a%20low-pressure%20system%20which,the%20world%20they%20are%20called%20hurricanes%20or%20typhoons.

The dimensions of our data are 37 columns and 8 rows. This may seem like a small amount of data, but each row is actually one full year's worth of data. For example, the first row of data is 1969. Therefore, we have 37 years' worth of data to work with, which is enough sample size. Additionally, the data is time series because it's measured over the course of time[3].

Variables

Excluded from Analysis

In our dataset, we will have some variables that aren't particularly useful to our analysis but could be helpful in understanding the data as a whole.

Total Cyclones

The total cyclones column details the sum of non-severe and severe cyclones each year. After doing some preliminary research, I was interested to see if our data matched up with what was expected from cyclones. On average, I found that the total number of cyclones on average is 13[4]. We can see how that compares to our code.

```
summary(cyclones$Total)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##       5      10      12      12      15      20
```

The mean and median have 12 average cyclones per year. This is consistent with common research shows with average cyclones per year. While Total Cyclones won't be important to our analysis, being consistent with research is important, because it means that severe and non-severe cyclones are probably consistent too.

Year

The year column details the year for which the data from the rest of the row is found. As we can see from the list below, we start in the year 1969 and go to the year 2005.

```
list(cyclones$Year)

## [[1]]
##  [1] 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983
```

---

[3] "Time Series - Statista Definition." *Statista Encyclopedia*, www.statista.com/statistics-glossary/definition/409/time_series/#:~:text=Definition%20Time%20series%20A%20time%20series%20is%20a,years%20is%20an%20example%20for%20a%20time%20series.

[4] Lazar, Nick. "Australia Resources." *Australia 101 - Australian Cyclones*, australia101.com/australia/climate-in-australia/australian-cyclones/#:~:text=An%20averagte%20of%2013%20cyclones%20occur%20in%20Australia,Australia%20is%20the%20most%20cyclone-prone%20region%20in%20Australa.

```
## [16] 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998
## [31] 1999 2000 2001 2002 2003 2004 2005
```

This column is great if we want to look at one individual row of data for an outlier. However, it isn't useful for our analysis.

Variables Included In Our Analysis

```
f <- function(x) {
  ans <- c(Mean = mean(x),
           SD = sd(x),
           Median = median(x),
           Min = min(x),
           Max = max(x))
  return(ans)
}

View(round(t(apply(cyclones[,(2:8)], 2, f)),2))
```

	Mean	SD	Median	Min	Max
Severe	5.46	2.26	5.0	3.0	11.0
NonSevere	6.54	3.52	6.0	0.0	15.0
Total	12.00	3.70	12.0	5.0	20.0
JFM	0.06	0.93	0.2	-1.7	2.0
AMJ	0.05	0.61	0.1	-0.9	1.2
JAS	0.07	0.78	0.1	-1.3	2.0
OND	0.06	1.13	-0.1	-2.0	2.5

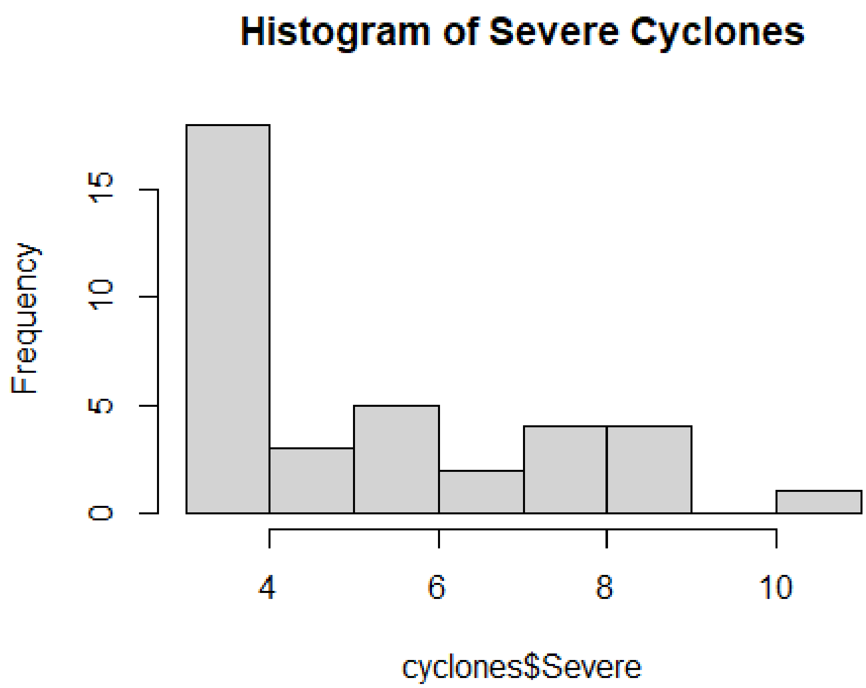
Target Variables

Next, we can see the variables we are trying to predict.

Severe Cyclones

In the introduction, I determined the criteria for a severe cyclone. In this section, I'll talk more about the summary statistics. We can see from the Mean of 5.46 and Median of 5 that data should be relatively distributed. However, visualizing it with the histogram leads to a different story.

```
hist(cyclones$Severe, main = "Histogram of Severe Cyclones")
```



From these histograms, there's two important things to note. First, most years have had four severe cyclones or less. This is why a majority of our data is skewed to the left. Secondly, there seems to be one year with more than ten cyclones that could be an outlier. To see if that's the case, we run an outlier test on that point:

```
#Severe Outlier
quantile(cyclones$Severe, probs = c(0.25,0.75))

## 25% 75%
##   4   7

#Now we can calculate our IQR
IQR = 7-4
Lower_Bound = 4 - (IQR*1.5)
Upper_Bound = 7 + (IQR*1.5)
```

```
Total_Outlier_Range = c(Lower_Bound, Upper_Bound)
Total_Outlier_Range

## [1] -0.5 11.5
```

The range for severe cyclones is from -0.5 to 11.5. As we can see from our chart, the maximum value of severe cyclones is 11. While it may appear, there was an outlier after examining the histogram, the outlier test shows that isn't the case.

Non-Severe Cyclones

Similarly, to Severe cyclones, I'll use this section to talk about non-severe cyclones' summary statistics. In observing the chart, we can see that mean and median are only 0.54 apart, but there is a minimum of 0 and a maximum of 15. Along with a standard deviation of 3.52, this indicates to me that the data is more spread out, and that the maximum of 15 could be a possible outlier.

In Figure 1 of the Appendix, we created a histogram for Non-Severe cyclones, and we can see that data looks relatively normally distributed. While it may be a bit skewed left, there doesn't seem to be any reason for concern.

However, in Figure 2 in the Appendix, I decided to run the Outlier Test just to see if all the rows of data were inside the range. Similarly, to Severe cyclones, I got a range of -2 to 11.5. This means that the maximum of 15 is considered an outlier according to our Outlier Test. To take a further look, we check what rows have more than 11.5 non-severe cyclones

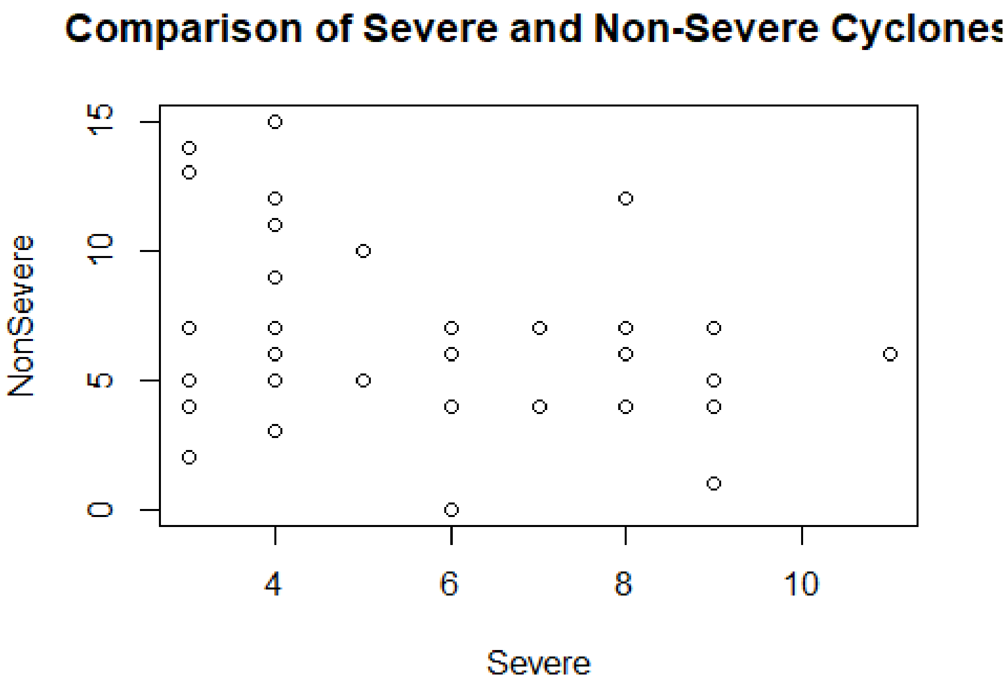
```
which(cyclones$NonSevere > 11.5)

## [1]  2  5  6  7 15
```

In our output, we get five rows of data with Non-Severe cyclones over 11.5. This is concerning that we have so many rows of data considered to be outliers. We won't remove the rows just because it appears the data has outliers. However, it is something to note if these rows affect our analysis later in the paper.

Comparison between Severe and Non-Severe Cyclones

With two target variables, I believe it would be important to compare severe and non-severe cyclones. To start, we can see that non-severe cyclones have a larger mean, median and standard deviation than severe cyclones. From this, we can see that non-severe cyclones occur more often but are also more variable due to the larger standard deviation. Additionally, there's more Non-Severe Cyclones to predict data on, so we can create a more accurate model. Additionally, I decided to plot the two target variables against each other to see any trends:

```
plot(cyclones$Severe, cyclones$NonSevere, main = "Comparison of Severe and Non-Severe
Cyclones", xlab = "Severe", ylab = "NonSevere" )
```



As we can see, the data doesn't seem to indicate a pattern, but there does seem to be a few points in the top left. This would be years with a lot of non-severe cyclones and not a lot of severe cyclones, which is possible after observing the two outliers in non-severe cyclones. Lastly, in Figure 3 in the Appendix, I decided to see if the two variables were correlated and got a value of -0.24. This indicates the values aren't really correlated which is good for our analysis. Overall, this plot doesn't indicate any reasons for concerns

Predictors:

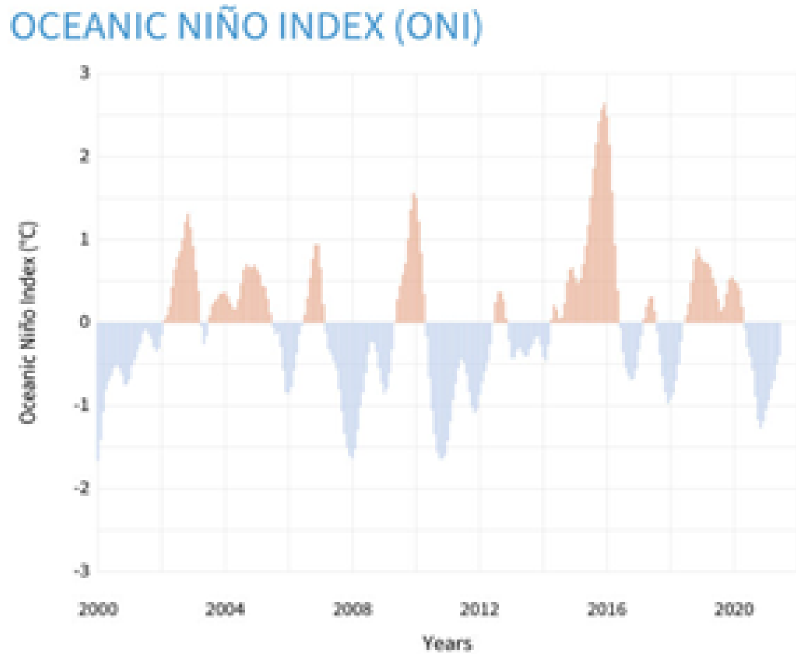
Finally, we can look at our variables trying to predict severe and non-severe cyclones.

JFM:

JFM is the Oceanic Nino Index, ONI, averaged from January to March. Before proceeding, I believe it's important to identify what the Oceanic Nino Index is. ONI looks at the average temperature anomaly in the surface waters of the east-central tropical Pacific[5]. In simpler terms, ONI is the difference in temperature from the average in water around Australia. While much of the past literature looked at wind as a factor for cyclones, we'll be looking at water temperature.

[5] Lindsey, Rebecca. "Climate Variability: Oceanic Niño Index." *Climate Variability: Oceanic Niño Index | NOAA Climate.gov*, www.climate.gov/news-features/understanding-climate/climate-variability-oceanic-ni%C3%B1o-index#:~:text=The%20Oceanic%20Ni%C3%B1o%20Index%20%28ONI%29%20is%20NOAA%27s%20primary,the%20east-central%20tropical%20Pacific%2C%20near%20the%20International%20Dateline.

A value of -0.5 is considered low, while a value of 0.5 is considered to be high. From the same source, there's a good chart as well to help visual ONI from year to year.



For the variable JFM, we will be looking at the ONI averaged from January to March. We can now get into the summary statistics. We can see the mean is about 0.06, and the median is 0.2, safely inside the normal range of -0.5 to 0.5. Additionally, the standard deviation of 0.93 isn't too different to our other observations. In Figure 4 of the Appendix, we can visualize JFM with a histogram. From this histogram, we can see that data looks normally distributed and there isn't too much to worry about.

AMJ

AMJ is the Oceanic Nino Index from April to June. For an explanation on the Oceanic Nino Index, you can refer to the previous section where I give a more in-depth description. We can see from the chart that the mean and median are only separated by 0.05. Additionally, AMJ has the smallest standard deviation of all predictor variables along with a minimum of only -0.9 and maximum of 1.2. Unsurprisingly, the histogram of AMJ, Figure 5 in the Appendix, shows that data is extremely condensed like the summary statistics show. Overall, the data looks normal and ready for analysis.

JAS

JAS is the Oceanic Nino Index from July to September. To look at the summary statistics, we can see that there's only a difference of 0.03 between the mean and median. One thing to note is that a maximum of 2 may be a bit high considering the standard deviation of only 0.78. In Figure 6 of the Appendix, we can see a histogram of JAS. The data looks relatively normal, but there could be an outlier with the maximum. Due to this, we run an outlier test for JAS.

As we can see from Figure 7, the range for outliers in JAS extends from -1.75 to 1.85. This would mean that our maximum of 2 is an outlier. We can use the which() function to see what row that is.

```
which(cyclones$JAS > 1.85)

## [1] 29
```

We can see that column 29, or the year 1997, has a JAS of 2. While this is an outlier, it’s only something to note for now as it isn't the same row that had an abnormal amount of non-severe cyclones.

OND

OND is the Oceanic Nino Index from October to December. For the summary statistics, we can see the median is negative, but the difference from the mean is only 0.07. Additionally, the standard deviation is the clear largest of any of the predictor variables. This means it’s unsurprising that we have a notably low minimum and high maximum. To further understand the data, we created a histogram for OND in the Appendix with Figure 8.

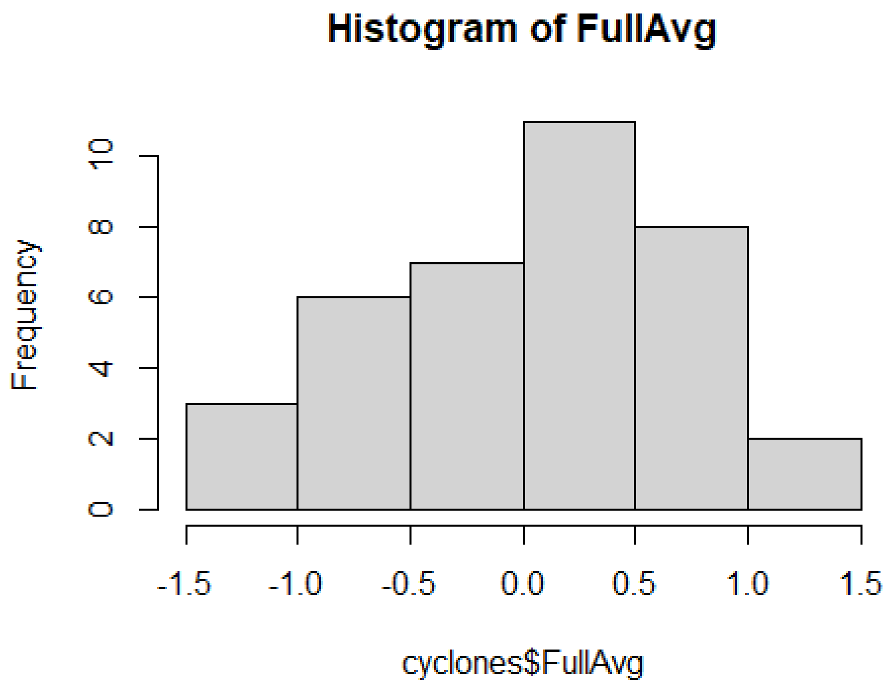
The data does look relatively spread out, and the end of the right side of the histogram could be a possible outlier. Due to this, we run an outlier test for OND in Figure 9 of the Appendix.

For our outlier test, we get a range of -3.45 to 3.35. This means that the minimum and maximum are within our range of data and there’s no outliers for OND.

FullAvg

FullAvg is the average of the four previous variables: JFM, AMJ, JAS and OND. This is a variable created by me, not provided. I did this to see if ONI is significant regardless of season. To explore the summary statistics, we can see the median and mean are close. Additionally, the standard deviation, minimum and maximum look normal as well. In the code below, we can see the histogram for FullAvg.

```
#Histogram for FullAvg
hist(cyclones$FullAvg, main = "Histogram of FullAvg")
```



As we can see, the data looks relatively normal with a very slight left skew. However, it looks perfectly acceptable for analysis.

Relationships

Prior to our analysis, it'll be important to know if any of our variables are correlated with each other. Due to this, the code below will be a chart comparing variables correlations with each other.

	Total	JFM	AMJ	JAS	OND
Total	1.00000000	-0.081859692	-0.4616937	-0.6717537	-0.682480411
JFM	-0.08185969	1.000000000	0.6770606	0.1343538	0.005864462
AMJ	-0.46169372	0.677060592	1.0000000	0.7042454	0.596064937
JAS	-0.67175366	0.134353758	0.7042454	1.0000000	0.943124589
OND	-0.68248041	0.005864462	0.5960649	0.9431246	1.000000000

```
View(cor(cyclones[4:8]))
```

I consider a strong correlation to be around 0.65, so any variables above that we'd consider strongly correlated. Additionally, I decided to leave FullAvg out since it was created from the other four predictors. Therefore, it would naturally have strong correlations with them.

For the four variables, we can see that three sets have strong correlations: JFM & AMJ, AMJ & JAS, JAS & OND. These variables do share one commonality in that they border each other. For the two sets of variables that don't border each other, JFM & JAS and AMJ & OND, the correlation wasn't as strong with 0.13 and 0.6 respectively.

From this, we can conclude that it may be unlikely we see many models with more than one variable due to a lot of the variables having high correlations with each other.

Model Creation and Selection

We have reached the model creation stage. However, we have two target variables, so I will split the model creation stage into two parts. Firstly, I will begin with the non-severe model selection. After coming to a decision on a model or no model at all, I will proceed to the severe model stage. Both sections will have the same steps for model analysis, but with different models due to the fact we are predicting different target variables.

If you want to skip directly to what model I choose or if I choose a model at all, you can scroll down to the "Non-Severe Final Model Choice" on page 24.

Non-Severe Model Selection

Before we go into the tests in assessing our model, we can begin by looking at the model's preliminary statistics.

	JAS Mod	OND Mod	Full Avg Mod
Variables Included	JAS	OND	FullAvg

Number of Predictors	1	1	1
Number of Precise Predictors	1	1	1
AIC	180.61	183.89	181.97
Residual Variance	45.422	48.702	46.781

Appendix Figure 10: Summary of JAS Model
Appendix Figure 11: Summary of OND Model
Appendix Figure 12: Summary of FullAvg Model

Variable Included and Number of Predictors

The first two rows are the variable names and number of predictors. Since each model has one variable, the model is named after the variable. Additionally, the number of predictors shows that each model has only one predictor.

For the number of predictors, it's just the number of variables in the model.

Precise Predictors

A precise predictor is identified as a predictor variable where the standard error is less than half of the coefficient; R identifies these with stars next to the p-value in the summaries (which can be found in Appendix Figures 10-12). To unpack the statistical jargon, a precise predictor is a predictor variable which we can be confident is predicting the target variable (Non-Severe) precisely. Essentially, the more precise predictors you have, the better.

We can see from our table that each model has one precise predictor. Since every model only has one predictor, it means that each model's only predictor is precise. While this is good to see, it doesn't separate any of the models apart.

AIC

In its simplest form, AIC is a mathematical method for estimating how well predictors fit a model[6]. To be more concise, the lower your AIC is for your model the better. However, how do we identify whether differences in AIC are actually significant? The chart below shows us what a significant difference in AIC looks like[7].

Difference : Model A and B	Decision (assuming $A < B$)
$0.0 < \text{difference} \leq 2.5$	No difference in models
$2.5 < \text{difference} \leq 6.0$	Prefer A if $n > 256$
$6.0 < \text{difference} \leq 9.0$	Prefer A if $n > 64$
$10.0 < \text{difference}$	Prefer A

If you remember from the data characteristics section, we had 37 rows of data, so n = 37. This means that if there is an AIC difference of more than ten between any of the models, that's a significant difference. The AICs are part of the summaries of a model, which can be found in Figures 10-12 in the Appendix section. However, I did list them in the table. As you can see from the table, there isn't a difference of more than three between any of the AIC's. Therefore, we can determine that AIC won't make a difference in our model selection process

Residual Deviance

Next, we have residual deviance. Residual Deviance tells you how well the target variable (Non-Severe) can be responded to by x predictor variable[8]. Additionally, the closer to 0, the better for residual deviance. One thing to note is the more variables you have, the lower your residual deviance is because it will be more representative of your data.

As we can see from the chart, the residual deviance for the JAS model is 45.4, FullAvg Model is 46.8 and the OND is 48.7. Since all the models have one variable, we can determine the JAS model is most representative of the data. However, the difference in residual deviance is only three between all the models, so it doesn't seem to be an indicator.

---

[6] Bevans, Rebecca. "An Introduction to the Akaike Information Criterion." *Scribbr*, 18 June 2021, www.scribbr.com/statistics/akaike-information-criterion/.

[7] Hilbe, J.M. 2009. Logistic Regression Models. Boca Raton, FL: Chapman & Hall/CRC.

[8] Zach. "How to Interpret Null & Residual Deviance (with Examples)." *Statology*, 1 Sept. 2021, www.statology.org/null-residual-deviance/#:~:text=The%20residual%20deviance%20tells%20us%20how%20well%20the,as%3A%20X2%20%3D%20Null%20deviance%20%E2%80%93%20Residual%20deviance.

Non-Severe Model Test Assessment

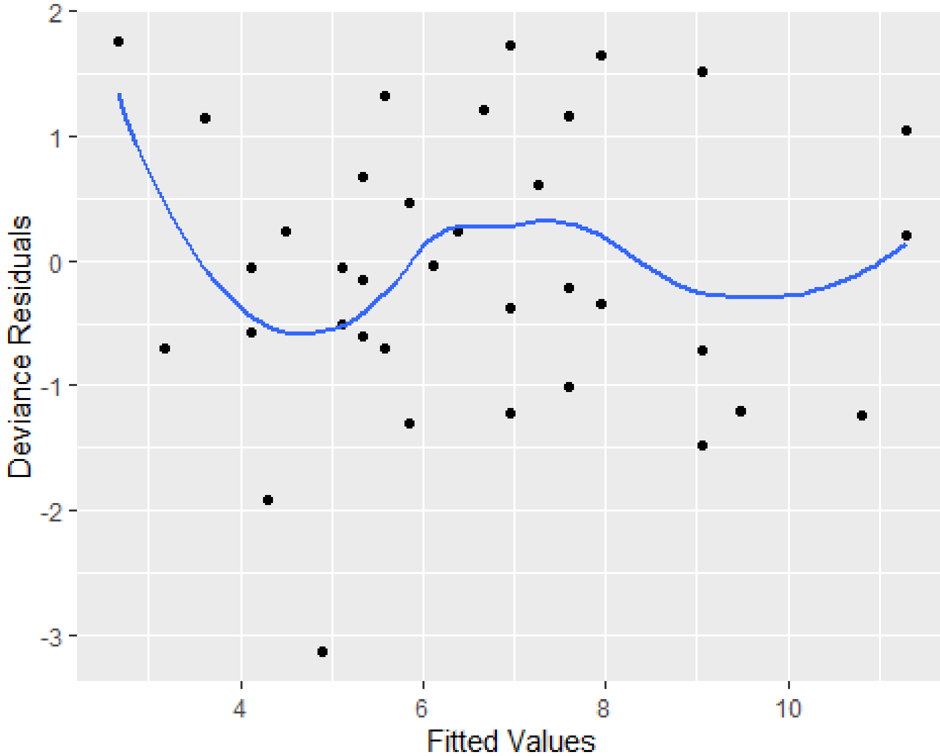
Moving on, we can see how each of these models performs in specific tests.

	JAS Mod	OND Mod	Full Avg Mod
Fitted Values vs Deviance Residuals	Not flat	Not flat	Relatively flat
Fitted Values vs Quantile Residuals	Not flat	Partially flat	Partially flat
Predictor vs Quantile Residuals	Not flat	Not flat	Partially flat
Working Response vs Linear Predictor	Failed	Failed	Mostly Failed
QQ-Plot for Deviance Residuals	Passed	Passed	Passed
QQ Plot for Quantile Residuals	Passed	Passed	Passed
Term Plot	Failed	Failed	Failed
Overdispersion (1 is ideal)	1.19	1.27	1.28

Fitted Values Vs Deviance Residuals

Our first test is comparing a model's fitted values to their deviance residuals. Like most residual plots, a good model will have the points close to 0. In the case of this model, I created a blue line for each plot to show the slope and see if it centers around 0. The focus for this test is to see if the blue line centers around 0. In an attempt not to clog the paper with code, I'll show the plot for the JAS model. In the Appendix, Figure 13 will be the plot for the OND model and Figure 15 will be the plot for the FullAvg Model.

```
cyclones$JAS.mu <- predict(JAS.mod, type = "response")
cyclones$JAS.rD <- resid(JAS.mod, type = "deviance")

p <- ggplot(data = cyclones,
            mapping = aes(x = JAS.mu,
                          y = JAS.rD))
p <- p + geom_point() + geom_smooth(se = FALSE)
p <- p + labs(x = "Fitted Values",
              y = "Deviance Residuals")
p
```



For the JAS Model, we can see that the blue line centers around 0 but isn't particularly flat. This is a cause for concern and is a reason I'd say the JAS model doesn't pass this test. Additionally, there appears to be a Deviance residual point below 3 that could skew the data.

```
which(cyclones$JAS.rD < -3)
```

```
## [1] 26
```

We can see that the row is 26, or the year 1994. We haven't had a concern about this point thus far, but it is important to note that this could be skewing the data.

As for the other two models in the Appendix, the OND Model (Figure 13) and the FullAvg Model (Figure 15), we can touch on their plots briefly. For the OND Model, the blue line is like the JAS model in that it centers around 0 but is relatively bumpy. Additionally, we can see in the plot that there is a deviance residual below -3. In Figure 14 of the Appendix, we confirm that it's row 26; the same row that we worried was skewing the JAS Model plot

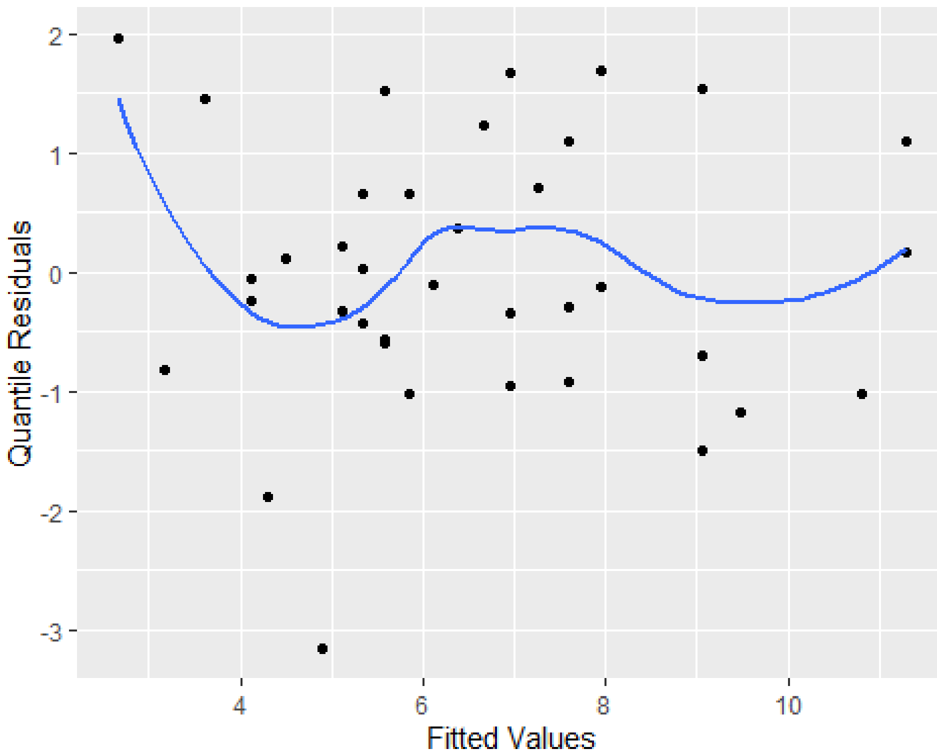
For the FullAvg Model, the line looks to be much flatter than the previous two models, which is a very good sign. Additionally, in Figure 16, we can see that the deviance residual below -3 is for row 26, the same as the JAS and OND models.

Overall, the FullAvg Model performs the best on this test. However, the line is still quite bumpy and there is a possible outlier in row 26 with a very low deviance residual.

Fitted Values vs Quantile Residuals

Our next test is comparing a model's fitted values to its residuals. Like the last section, a good model will have the majority of its points close to 0. To visualize this, we added a blue line to the plot to visualize the slope. A good blue line would be relatively flat across 0. Once again, I will show the code and plot for the JAS model here and the code for the other two models in the Appendix. However, I will give explanations of all the plots

```
#Fitted Values vs Quantile Residuals - JAS Mod
library(statmod)
cyclones$JAS.rQ <- qresid(JAS.mod)

p <- ggplot(data = cyclones,
            mapping = aes(x = JAS.mu,
                          y = JAS.rQ))
p <- p + geom_point() + geom_smooth(se = FALSE)
p <- p + labs(x = "Fitted Values",
              y = "Quantile Residuals")
p
```



In looking at this plot, we can see that the line is extremely bumpy. While it is around 0, I would not consider the line flat. Therefore, I would say it does not pass the test. Additionally, there seems to be an outlier below -2.5 for Quantile Residuals that is skewing the data

```
#Which JAS.rQ is less than -2
which(cyclones$JAS.rQ < -2)

## [1] 26
```

As we can see, it's the same row 26 that was a possible outlier in our first test. We can assume that this point could be affecting the data, but we will leave it in our analysis as I don't believe the graph would greatly change it. Additionally, I won't keep mentioning this row 26 of data as to not be repetitive.

In Figure 17 of the Appendix, we can see the Fitted Values vs Quantile Residuals plot for the OND Model. In comparison to the JAS model, the blue line is much less bumpy. However, I wouldn't consider it flat due dip around 5 of Fitted Values and the two bumps after that.

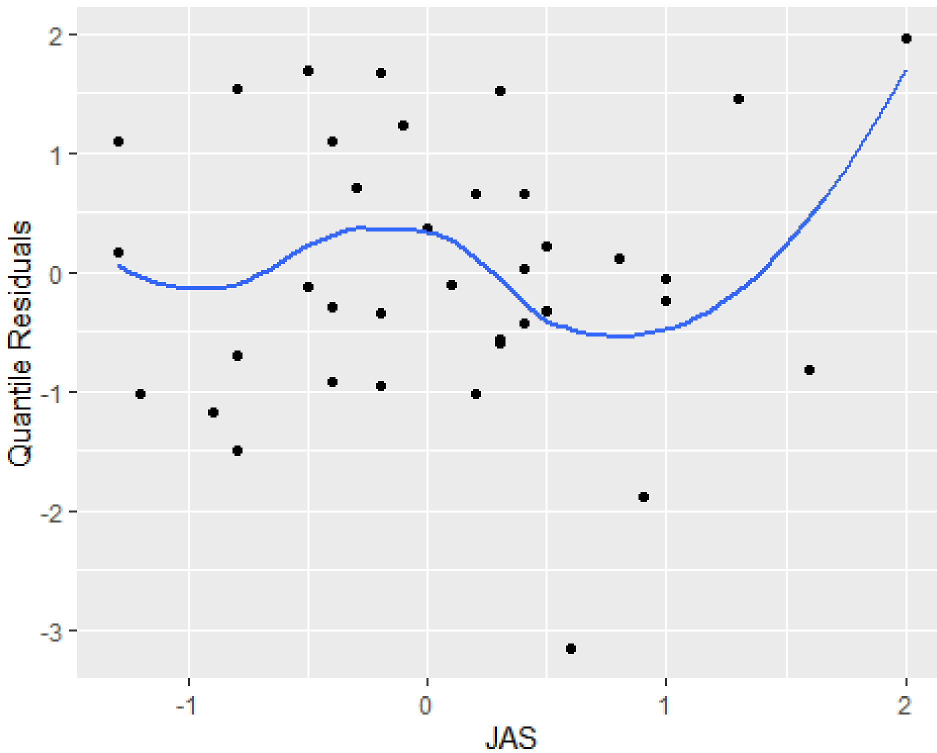
In Figure 18 of the Appendix, we can see the Fitted Values vs Quantile Residuals plot for the FullAvg Model. Similar to the OND plot, the FullAvg plot is much flatter than the JAS plot. However, there is a bump in the center that stops it from being relatively flat.

Overall, none of these models pass this test with flying colors, but the OND model and FullAvg model has plots that are partially flat.

Predictor Versus Quantile Residuals

In addition to how our fit values do first Quantile residuals, we also want to see how our predictors due versus them. Since all our models are one variable, this is relatively simple. We can run that one variable against its quantile residuals. Like the previous two plots, an indication of a good predictor is when the points are gathered around 0. This means that the blue line we've added in would ideally be flat on 0. Also like the last two sections, I'll show the code for the JAS model here, and the other two models in the Appendix. However, I will explain all the plots and whether they pass the test.

```
#Code for JAS vs JAS.rQ
p <- ggplot(data = cyclones,
            mapping = aes(x = JAS,
                          y = JAS.rQ))
p <- p + geom_point() + geom_smooth(se = FALSE)
p <- p + labs(x = "JAS",
              y = "Quantile Residuals")
p
```



As we can see from the plot, the blue line is not flat across 0. At the beginning, there doesn't seem to be a big bump. However, the real concern comes from the end where the blue line swings significantly upward. Additionally, it seems to be angled at the point above 2 for Quantile Residuals.

```
#Which JAS.rQ point is above 2
which(cyclones$JAS.rQ > 2)

## 29
```

The point in question is row 29, or the year 1997 in our data. That row may be familiar because I referred to it earlier in the paper when I said the JAS for 1997 failed the outlier test. This could be a possible reason why a model with JAS wouldn't be plausible. However, that blue line is still very bumpy for the JAS model, and it fails this test.

In Figure 19 of the Appendix, we created the plot comparing OND and its Quantile Residuals. Like JAS, I would not consider the blue line flat. While the line is relatively flat in the beginning, it has a big upswing at the end that ruins any form of flatness.

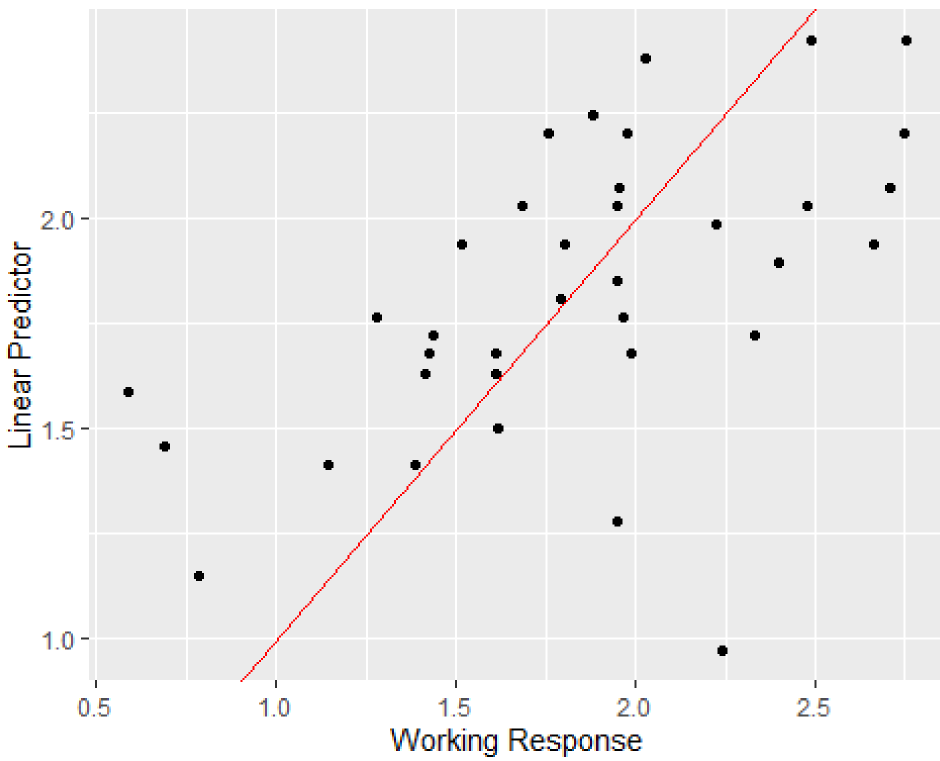
In Figure 20 of the Appendix, we created the plot comparing FullAvg and its Quantile Residuals. While the line is flatter than the previous two, we've explored, there is still a relatively big bump at the start that would stop him from passing his test.

Overall, I feel that none of the three models pass this test as none of the plots had the points closely gathered around 0.

Working Response vs Linear Predictor

In this section, we will be checking to see if our link function, the systematic part of our function, is reasonable. In simpler terms, an indication of a good link function would be if the points were gathered around the red line we created. To start, we can look at the plot for this for the JAS model.

```
#Working Response vs Linear Predictor - JAS model

cyclones$JAS.eta <- predict(JAS.mod, type = "link")
cyclones$JAS.rW <- resid(JAS.mod, type = "working")

p <- ggplot(data = cyclones,
            mapping = aes(x = JAS.eta + JAS.rW,
                          y = JAS.eta))
p <- p + geom_point()
p <- p + labs(x = "Working Response",
              y = "Linear Predictor")
p <- p + geom_abline(intercept = 0, slope = 1,
                     color = "red")
p
```



As we can see from the plot, the points do not appear to be gathered around the red line. Instead, they seem to be randomly scattered with only a slight sense of direction toward the red line.

In Figure 21, we look at the same plot for our OND model. Like the JAS model, the points don't seem to be gathered around the red line. While a few clusters of points are next to the red line, the rest seem to be randomly scattered with no direction.

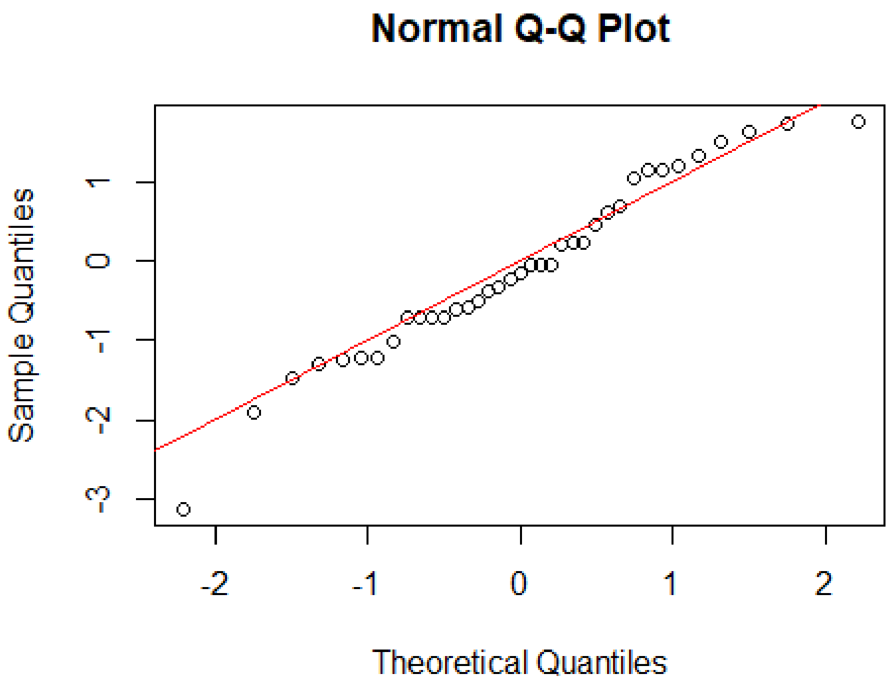
In Figure 22, we look at the plot for our FullAvg model. Like the previous two models, the points don't seem to be gathered along the red line. While they do seem closer than the previous two models, I would not say that the FullAvg model has passed this test.

Overall, like the last test, I would say that none of the models passed this test. It is a major concern that none of the link functions seem reasonable.

Q-Q Plot for Deviance Residuals

In our next test, we're going to look at our model's error and distribution. We can do this by creating a Q-Q Plot for our Deviance Residuals and Quantile Residuals (the next section). A good model will have the points in the Q-Q plot follow the red line we created. In our plot below, we can see the Q-Q plot for the JAS model.

```
#Q-Q Plot for Deviance Residuals for JAS model
qqnorm(cyclones$JAS.rD)
abline(0, 1, col = "red")
```



In this plot, we can see that the points mostly follow the red line. While they do tail off a bit toward the end, I'm confident to say the JAS model passes this test.

In Figure 23 of the Appendix, we look at the Q-Q Plots of the Deviance Residuals of the OND Model. From the plot, we can see that the points stray a bit more from the line than the JAS model. However, I would say the points follow the red line well enough, so the OND model passes the test.

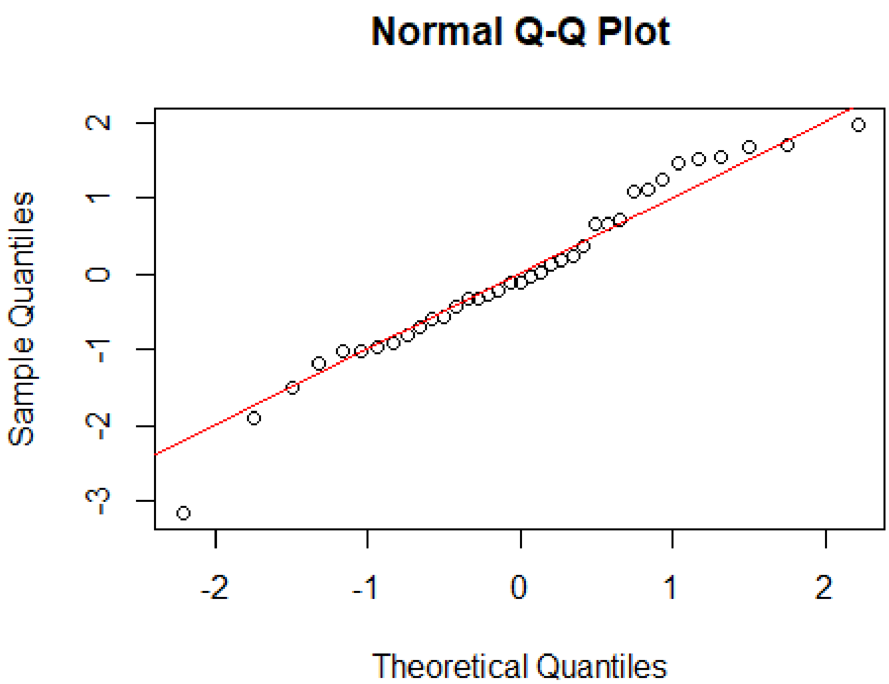
Lastly, Figure 24 of the Appendix has the Q-Q Plots for the Deviance Residuals of the FullAvg Model. We can see from this plot that it also passes this test. Almost all the points, except for the three at the end and one at the beginning, are hugged to the red line.

Overall, all three models pass this test. This is a good sign for the models but makes it difficult to separate which is better or if any of them are useful.

Q-Q Plot for Quantile Residuals

Like the last section, we are going to create a Q-Q Plot for each model but use the quantile residuals instead of the deviance residuals. The indications for a good model are the same as most of the points must be hugging the red line. Firstly, we can look at the Q-Q Plot comparing the Quantile Residuals for the JAS Model.

```
#Q-Q Plot for Quantile Residuals for JAS Mod
qqnorm(cyclones$JAS.rQ)
abline(0, 1, col = "red")
```



While some points stray toward the right end of the line, I would say the JAS model passes this test. Most of the points are on the line or right next to it, so that indicates this model has passed the test.

In Figure 25 of the Appendix, we can find the Q-Q Plot for the Quantile Residuals of the OND Model. While the points do stray a bit more from the red line than in the JAS plot, I would still say the OND model passes this test. This is because we can see a clear pattern from the points, and it follows the line.

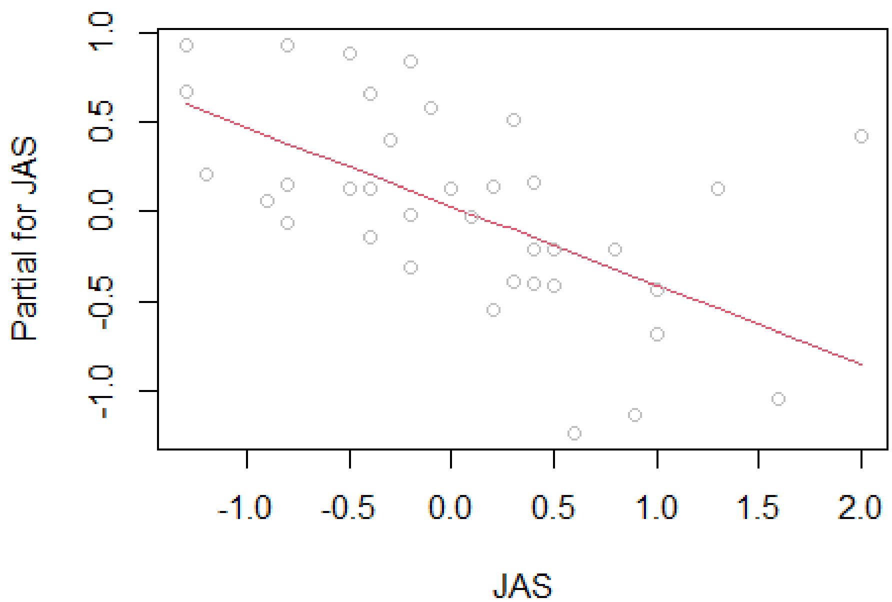
In Figure 26 of the Appendix, we can find the Q-Q Plot for the Quantile Residuals of the FullAvg Model. While the three points at the end are off the red line, most of the points follow the line enough for me to say it passes the test.

To conclude, all three models once again pass this Q-Q plot test. While this is great for our models, it doesn't help in separating which is better.

Term Plots

Our next test will be us visualizing a term plot. The question we are trying to answer here is: Did we put the variables in our model on the right scale? If we did, the grey points from the plot would gather around the red line we created. To start, we can see if the JAS model passed the test.

```
#Term Plot for JAS Mod
termplot(JAS.mod, partial.resid = TRUE, terms = "JAS")
```



As we can see, the points have no correlation with the red line at all. Instead, they are scattered throughout; this means that the JAS model fails this test.

In Figure 27 of the Appendix, we can see the term plot for the OND model. Like the JAS model, the OND model fails this test as the points are nowhere near the red line.

In Figure 27 of the Appendix, we can see the term plot for the FullAvg model. While some points are close to the red line, the overwhelming number of points are far from the line. Therefore, the FullAvg model fails this test.

To conclude, we can see that all three models fail this test as none of their points gather around the red line. Overall, this doesn't provide any reason to choose any of the models.

Overdispersion

Our last test is looking if there's overdispersion in any of our models. Overdispersion is when we see more variation in our model than expected or predicted[9]. The ideal overdispersion is 1, and any overdispersion within 0.9 to 1.1 is acceptable as well. You can see the overdispersion in the original chart, but the code below shows how I calculated each overdispersion.

```
#Overdispersion for Non-Severe Models

OD_JAS <- (phi <- sum(resid(JAS.mod, type="pearson")^2) / (37-2)) #n = 37, 2 parameters

OD_OND <- (phi <- sum(resid(OND.mod, type="pearson")^2) / (37-2)) #n = 37, 2 parameters

OD_FullAvg <- (phi <- sum(resid(FullAvg.mod, type="pearson")^2) / (37-2)) #n = 37, 2
parameters

OD_JAS

## [1] 1.188664

OD_OND

## [1] 1.267941

OD_FullAvg

## [1] 1.280022
```

As we can see from the overdispersion calculations, all three models have an overdispersion of 1.19 or above. This is particularly concerning and could steer us clear from choosing any of the models

Non-Severe Final Model Choice

After running all these tests and comparing all three models, I decided to select none of the models. I believe that none of the models are good at predicting non-severe cyclones. The main reason for this was the model statistics and performances on the tests. While all the models had similar AIC's and Residual Deviances, those numbers weren't

[9] "7.2.2 - Overdispersion: Stat 504." *PennState: Statistics Online Courses*, online.stat.psu.edu/stat504/lesson/7/7.2/7.2.2.

particularly impressive. Additionally, all three models failed most of the tests. Outside having a precise predictor and passing the Q-Q Plot tests, these models didn't seem like reasonable models. Instead of choosing the best of these three, I feel that none of our predictor variables predict non-severe cyclones well.

Severe Model Selection

The format will be identical to the last section with preliminary model statistics and performances on tests for models predicting severe cyclones. However, these sections will be shorter as I won't be re-explaining each statistic or test as to not be repetitive. If you are confused on a statistic or test and its significance, you can refer to the Non-Severe section. Additionally, if you want to skip to the model selection you can go to the "Severe Final Model Choice" section on page 30.

	Naïve Model	Only OND
Variables Included	N/A	OND
Number of Predictors	0	1
Precise Predictors	0	0
AIC	163.37	163.5
Residual Deviance	32.22	30.343

Appendix Figure 29: Summary of Naive Model
Appendix Figure 30: Summary of OND Model

Variable Included and Number of Predictors

The first two rows are the variable names and number of predictors. For the Naive Model, you may be a bit confused as there's zero variables included. Why would I include a model with zero variables in our analysis? This model was one of the better performers in predicting severe cyclones. To be simpler, random guessing was better than most of our variables at predicting severe cyclones, which could be an indication that we might not choose any models at all. Due to this, the Naive Model has zero variables, and the predictor variables are replaced with a 1 as you can see in the summary in Figure 29 in the Appendix.

For the OND Model, it's more straightforward. The model includes one predictor, and it's OND, which the model is named after. The summary for this model can be found in the Appendix in Figure 30.

Precise Predictors

For precise predictors, we can see the Naive model has no precise predictors. This is unsurprising because the model has no predictors in the first place.

For the OND model, there are also no precise predictors. This is concerning as the only predictor in the model, OND, is not precise in predicting severe cyclones. Due to this, neither of these models seem reasonable to select for now.

AIC

If you remember from the non-severe section, there was a chart to determine significant differences in AIC. For convenience, I decided to show the chart again.

Difference : Model A and B	Decision (assuming $A < B$)
$0.0 <$ difference ≤ 2.5	No difference in models
$2.5 <$ difference ≤ 6.0	Prefer A if $n > 256$
$6.0 <$ difference ≤ 9.0	Prefer A if $n > 64$
$10.0 <$ difference	Prefer A

If you remember from the data characteristics section, we had 37 rows of data, so n = 37. This means that if there is an AIC difference of more than ten between any of the models, that's a significant difference. The difference between the AIC is 0.13, which is extremely small. This means that AIC shows no significant difference between these models.

Residual Deviance

Lastly, we can use our chart to see that the residual deviance for our Naive Model is 32.22 and 30.34 for our OND Model. It's unsurprising that the OND Model has a smaller residual deviance as it includes a predictor, while the Naive Model does not. However, the difference of 1.88 isn't too significant with residual deviance. Due to this, residual deviance doesn't seem to be an indicator in deciding which model is better.

Severe Model Test Assessment

Moving on, we can see how each of these models performs in specific tests.

	Naive Mod	OND Mod
Fitted Values vs Deviance Residuals	Fails	Partially flat
Fitted Values vs Quantile Residuals	Fails	Partially Flat
Predictor vs Quantile Residuals	Fails	Partially Flat
Working Response vs Linear Predictor	Fails	Fails
QQ-Plot for Deviance Residuals	Partially Passed	Passed
QQ Plot for Quantile Residuals	Passed	Passed
Term Plot	N/A	Fails
Overdispersion	0.93	0.89

Fitted Values vs Deviance Residuals

In Figure 31 in the Appendix, we can see the plot for Fitted Values versus Deviance Residuals for our Naive Model. Since there's no predictors for the Naive Model, they all have the same fitted values at 5.46. This easily fails the test as the points aren't flat across 0.

In Figure 32 of the Appendix, we can see the plot for the Fitted Values versus Deviance Residuals for our OND model. The blue line is flat for most of the plot but does seem to tail off at the end. Additionally, there doesn't seem to be any outlier points like our non-severe data had.

Overall, the OND model partially passes this test since the line is mostly flat; the Naive Model easily fails this test. It is concerning that neither of these models had a flat blue line across 0 though.

Fitted Values vs Quantile Residuals

In Figure 33 of the Appendix, we can see the plot for Fitted Values versus Quantile Residuals for our Naive Model. Like the last plot and Figure 31, the Quantile Residuals create a vertical line as they all have the same fitted value. Due to this, the Naive Model easily fails this test.

In Figure 34 of the Appendix, we can see the plot for Fitted Values versus Quantile Residuals for our OND Model. This plot is also like the Fitted Values vs Deviance Residuals plot as it's mostly flat until a bump at the end. Once again, I would say the OND Model partially passes this test.

Overall, neither model fully passes this test which is disappointing. However, there is some promise with the OND model's first two results.

Predictor vs Quantile Residuals

In Figure 35 of the Appendix, you can see a plot comparing 1, the predictor of the Naive model, and the Quantile Residuals of the Naive Model. You can begin to see a theme with the Naive Model's test plots as this is a vertical line across 1. This is no surprise because every predictor is 1. Like the past two tests, the Naive Model fails this test.

In Figure 36 of the Appendix, you can see a plot comparing OND, the predictor of the OND model, and the OND Model's Quantile Residuals. The blue line in this plot seems to be a backward version of the last two, with a bump at the front and it being relatively flat after. For this reason, I would say this model partially passes this test.

Overall, the Naive Model fails this test and the OND model partially passes this test. At this rate, I don't believe either of the models are suitable for predicting severe cyclones, but we still have more tests to run.

Working Response vs Linear Predictor

In Figure 37 of the Appendix, we can see the Working Response versus Linear Predictor for the Naive Model. From the plot, you can see the points create a horizontal line where the linear predictor equals 1.697. Since the points aren't remotely gathered around the red line, the Naive Model fails this test.

In Figure 38 of the Appendix, we can see the Working Response versus Linear Predictor for the OND Model. We can see from the plot that the points are randomly scattered throughout with no correlation to the red line. Due to this, the OND model fails this test.

Overall, both models easily failed this test as neither of their points gathered around the red line. This is extremely concerning and an indication we don't have the correct link function.

Q-Q Plot for Deviance Residuals

In Figure 39 of the Appendix, we create a Q-Q Plot for the Naive Model's Deviance Residuals. In this plot, it does appear that the points are close to the red line. However, you can see they aren't gathered around the red line. For this reason, I'll say that the Naive Model only partially passes this test.

In Figure 40 of the Appendix, we create a Q-Q Plot for the OND Model's Deviance Residuals. We can see from this plot that the points are gathered around the red line, even if they aren't exactly on it. For this reason, the OND Model passes this test.

Overall, this test provided some separation in seeing why the OND Model is better than the Naive Model.

Q-Q Plot for the Quantile Residuals

In Figure 41 of the Appendix, we can see the Q-Q Plot for the Naive Model's Quantile Residuals. In this plot, the points are gathered around the red line. Due to this, the Naive Model passes this test.

In Figure 42 of the Appendix, we can see the Q-Q Plot for the OND Model's Quantile Residuals. Like the Naive Model, the OND Model passes this test as the points are gathered along the red line.

Overall, both models passed this test. While it's good to see both of our models pass this test, it doesn't help in separating which model is better.

Term Plot

For the Naive Model, we were unable to create a term plot. This is because there were no predictors, we were able to identify as terms, hence why we put N/A in the term plot section for the Naive Model.

In Figure 43 of the Appendix, we create the term plot for the OND model. We can see from the plot that the grey points don't have any correlation and aren't gathered around the red line. Therefore, the OND model fails this test.

Overall, this test did not help either model's case for being selected.

Overdispersion

For our last test, we look at a model's overdispersion. If you remember from the non-severe section, the ideal dispersion would be between 0.9 and 1.1.

```
#Overdispersion for Severe Models

OD_Naive <- (phi <- sum(resid(Naive.Mod, type="pearson")^2) / (37-1)) #n = 37, 1
parameter

OD_OND2 <- (phi <- sum(resid(OND.Mod, type="pearson")^2) / (37-2)) #n = 37, 2 parameters


OD_Naive

## [1] 0.9320682

OD_OND2

## [1] 0.8887248
```

As you can see, the Naive Model's dispersion is 0.93, which is close enough to 1 to be reasonable. However, the OND Model's dispersion is 0.89, which is a bit more concerning. Overall, this would give me reason to select the Naive Model over the OND Model.

Severe Final Model Choice

After all these tests, I've decided to select once again none of the models. First, we can look at the Naive Model. The Naive Model never really had a chance to be selected due to the fact it included none of the predictors we were interested in and failed many of the tests we ran. However, I included because some of its statistics like overdispersion, AIC and residual deviance were as good as any of the models with predictors.

This leads into why I didn't select the OND model. While it didn't fail most of the tests like the Naive Model, its statistics were as good as a model with no predictors. This indicates that none of the five predictors we explored were significant in predicting severe cyclones

Summary and Concluding Remarks

From my analysis of severe and non-severe cyclones, we can see that ONI doesn't appear to influence cyclones. From the four variables we were provided and one we created, it appeared that water temperature and season aren't significant in predicting non-severe and severe cyclones. For non-severe cyclones, we had some promising models from the preliminary statistics. However, the models didn't perform well on most of our tests. Meanwhile for severe cyclones, most of our models were outperformed by the Naive Model, which included no variables at all.

In the future, there are some other variables I'd be interested in exploring. In my introduction, I explained the difference between a severe and non-severe cyclone and how it had to do wind speed. In future analysis, I'd like to look at variables that have an effect on wind speed. Additionally, I think it would be interesting to explore hurricanes or cyclones in other areas of the world and see if there are any similarities. For example, are there variables significant in predicting hurricanes in the Gulf of Mexico? If there are, they may be applicable to cyclones around Australia as well.

Overall, predicting cyclones in the Australian region is critical to saving lives and spending relief effort money in a smart manner. While none of our five predictors were useful, we hope to explore more data and find variables in the future that are significant.

Appendix

Figure 1: Histogram for Non-Severe Cyclones

```
#Figure 1 Appendix
hist(cyclones$NonSevere, main = "Histogram of NonSevere Cyclones")
```

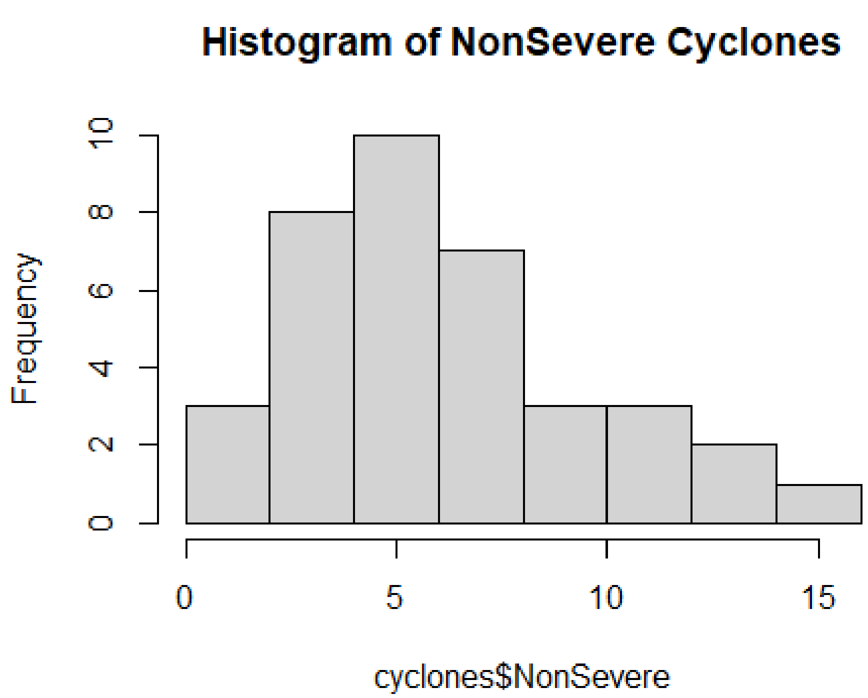


Figure 2: Outlier Test for Non-Severe Cyclones

```
#Figure 2 Appendix
#NonSevere Outlier
quantile(cyclones$NonSevere, probs = c(0.25,0.75))

## 25% 75%
##   4   7

#Now we can calculate our IQR
IQR2 = 7-4
Lower_Bound2 = 4 - (IQR2*1.5)
Upper_Bound2 = 7 + (IQR2*1.5)
Total_Outlier_Range2 = c(Lower_Bound2, Upper_Bound2)
Total_Outlier_Range2

## [1] -0.5 11.5
```

Figure 3: Correlation between Severe and Non-Severe Cyclones

```
#Figure 3 of Appendix
cor(cyclones$Severe, cyclones$NonSevere)

## [1] -0.2352758
```

Figure 4: Histogram for JFM

```
#Figure 4 of Appendix
hist(cyclones$JFM, main = "Histogram of JFM")
```

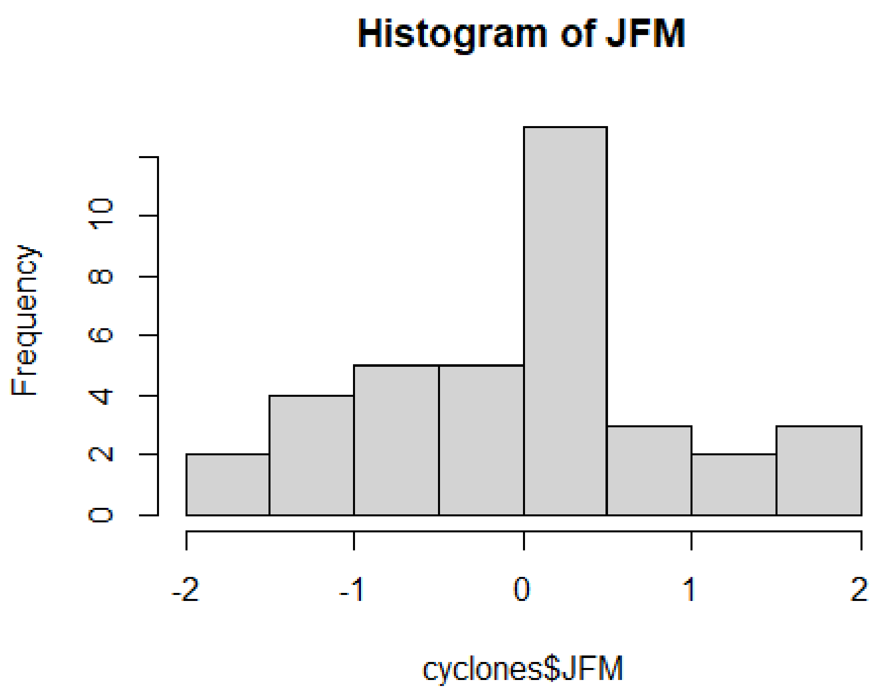


Figure 5: Histogram for AMJ

```
#Figure 5 of Appendix
hist(cyclones$AMJ, main = "Histogram of AMJ")
```

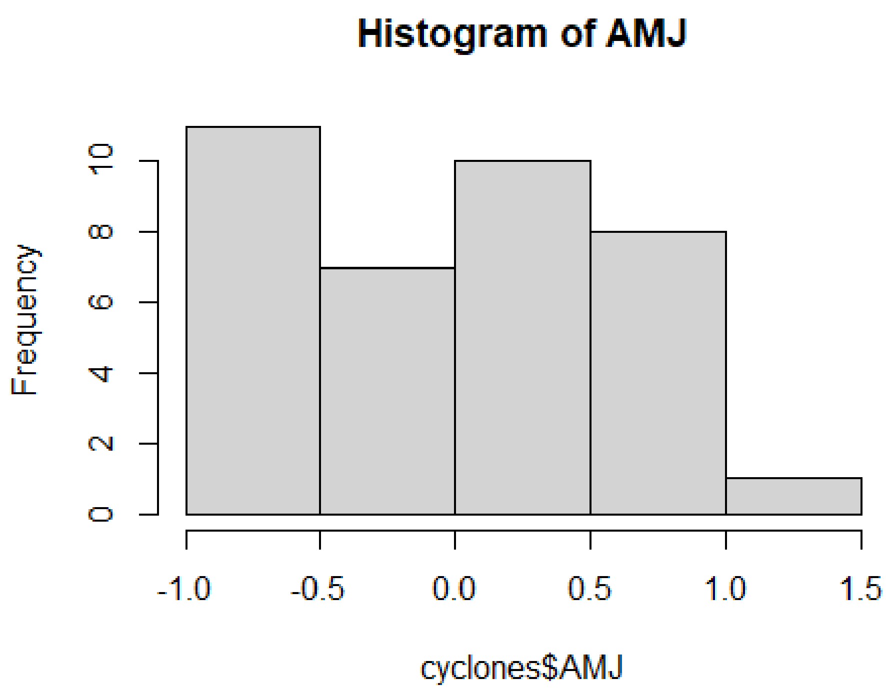


Figure 6: Histogram for JAS

```
#Figure 6 of Appendix
hist(cyclones$JAS, main = "Histogram of JAS")
```

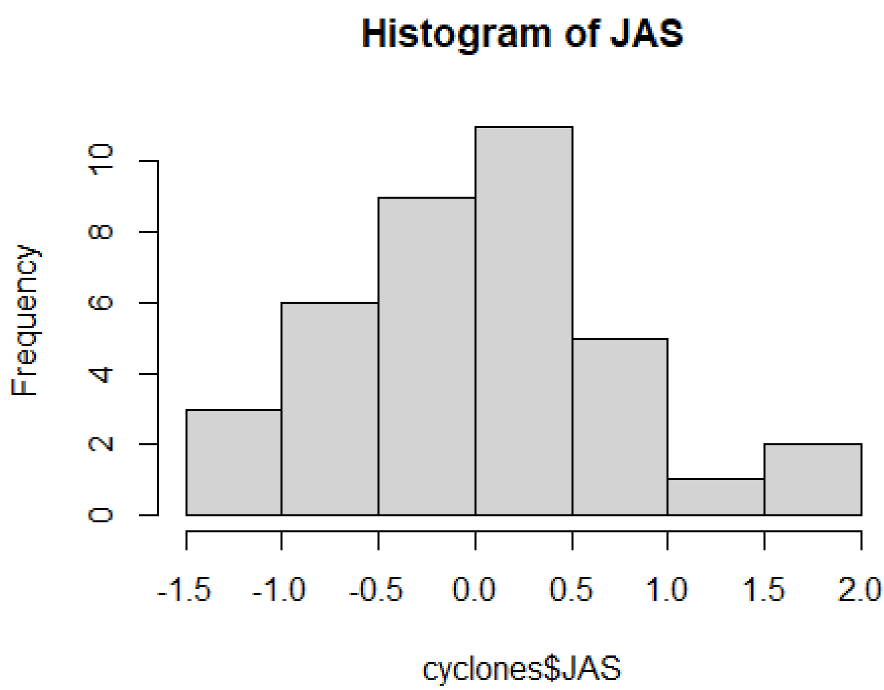


Figure 7: Outlier Test for JAS

```
#Figure 7 Appendix
#JAS Outlier
quantile(cyclones$JAS, probs = c(0.25,0.75))

##  25%  75%
## -0.4  0.5

#Now we can calculate our IQR
IQR3 = 0.5 - (-0.4)
Lower_Bound3 = -0.4 - (IQR3*1.5)
Upper_Bound3 = 0.5 + (IQR3*1.5)
Total_Outlier_Range3 = c(Lower_Bound3, Upper_Bound3)
Total_Outlier_Range3

## [1] -1.75  1.85
```

Figure 8: Histogram for OND

```
#Figure 8 of Appendix
hist(cyclones$OND, main = "Histogram of OND")
```

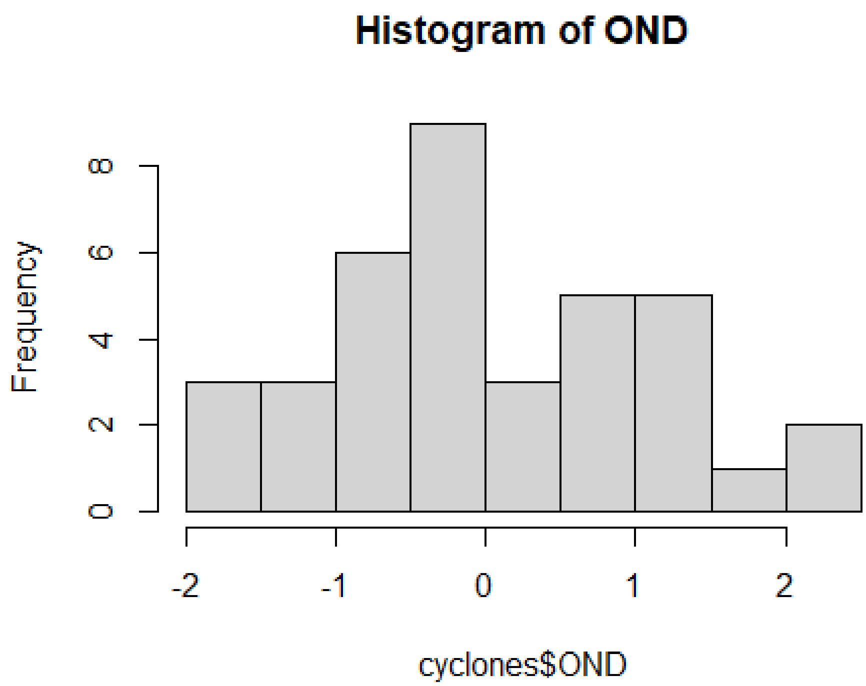


Figure 9: Outlier Test for OND

```
#Figure 9 Appendix
#JAS Outlier
quantile(cyclones$OND, probs = c(0.25,0.75))

##  25%  75% 
## -0.9  0.8

#Now we can calculate our IQR
IQR4 = 0.8 - (-0.9)
Lower_Bound4 = -0.9 - (IQR4*1.5)
Upper_Bound4 = 0.8 + (IQR4*1.5)
Total_Outlier_Range4 = c(Lower_Bound4, Upper_Bound4)
Total_Outlier_Range4

## [1] -3.45  3.35
```

Figure 10: Summary for JAS Mod (Non-Severe)

```
#Figure 10 for Appendix
JAS.mod <- glm(NonSevere ~ JAS,
             data = cyclones,
             family = poisson(link = "log"))
summary(JAS.mod)

##
## Call:
## glm(formula = NonSevere ~ JAS, family = poisson(link = "log"),
##     data = cyclones)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.1298  -0.7080  -0.1520   0.6818   1.7640
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.85235    0.06621  27.977  < 2e-16 ***
## JAS         -0.43929    0.08826  -4.977 6.46e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 71.218  on 36  degrees of freedom
## Residual deviance: 45.422  on 35  degrees of freedom
## AIC: 180.61
##
## Number of Fisher Scoring iterations: 4
```

Figure 11: Summary of OND Mod (Non-Severe)

```
#Figure 11 of Appendix
OND.mod <- glm(NonSevere ~ OND,
             data = cyclones,
             family = poisson(link = "log"))
summary(OND.mod)

##
## Call:
## glm(formula = NonSevere ~ OND, family = poisson(link = "log"),
##     data = cyclones)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
```

```
## -3.00913  -0.97562  -0.07957   0.83940   1.84923
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.84752    0.06637  27.836  < 2e-16 ***
## OND         -0.28113    0.06031  -4.662 3.14e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 71.218  on 36  degrees of freedom
## Residual deviance: 48.702  on 35  degrees of freedom
## AIC: 183.89
##
## Number of Fisher Scoring iterations: 4
```

Figure 12: Summary of FullAvg Mod (Non-Severe)

```
#Figure 12 for Appendix
FullAvg.mod <- glm(NonSevere ~ FullAvg,
             data = cyclones,
             family = poisson(link = "log"))
summary(FullAvg.mod)
```

```
##
## Call:
## glm(formula = NonSevere ~ FullAvg, family = poisson(link = "log"),
##     data = cyclones)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -3.07919  -0.74284  -0.06899   0.77310   2.59830
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.85601    0.06596  28.138  < 2e-16 ***
## FullAvg     -0.47972    0.09783  -4.904  9.4e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 71.218  on 36  degrees of freedom
## Residual deviance: 46.781  on 35  degrees of freedom
## AIC: 181.97
```

```
##
## Number of Fisher Scoring iterations: 4
```

Figure 13: Fitted Values vs Deviance Residuals Plot for OND Mod (Non-Severe)

```
#Figure 13 in Appendix
cyclones$OND.mu <- predict(OND.mod, type = "response")
cyclones$OND.rD <- resid(OND.mod, type = "deviance")

p <- ggplot(data = cyclones,
            mapping = aes(x = OND.mu,
                          y = OND.rD))
p <- p + geom_point() + geom_smooth(se = FALSE)
p <- p + labs(x = "Fitted Values",
              y = "Deviance Residuals")
p
```

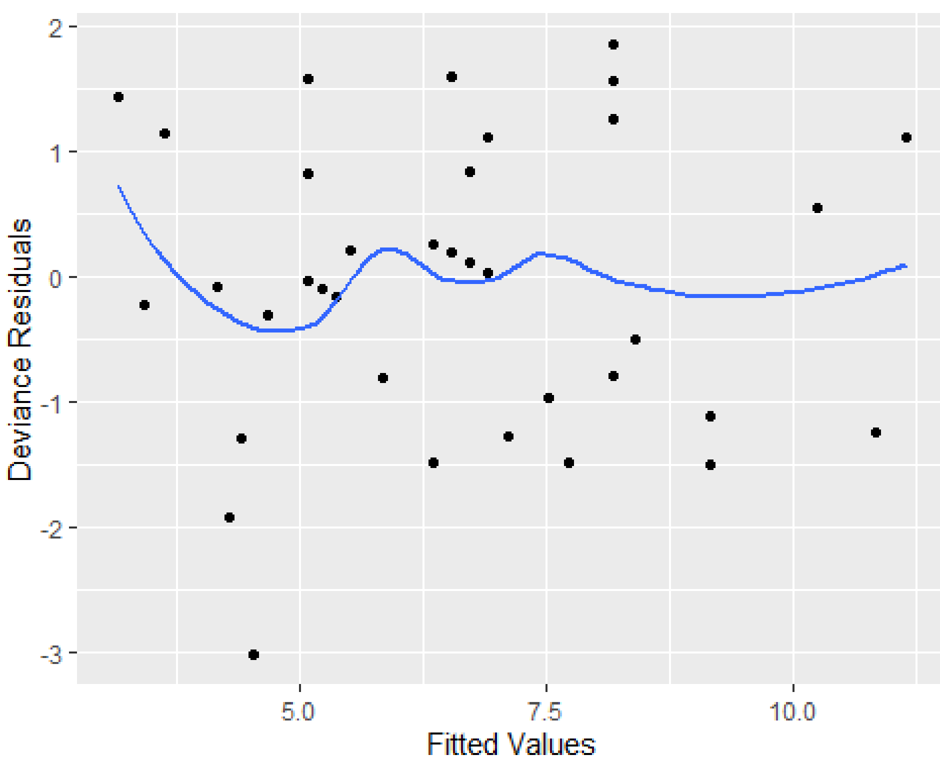


Figure 14: Which Residual Deviance is below -3, OND plot

```
#Figure 14 in the Appendix
which(cyclones$OND.rD < -3)

## [1] 26
```

Figure 15: Fitted Values vs Deviance Residuals Plot for FullAvg Mod (Non-Severe)

```
#Figure 15 in Appendix
cyclones$FullAvg.mu <- predict(FullAvg.mod, type = "response")
cyclones$FullAvg.rD <- resid(FullAvg.mod, type = "deviance")

p <- ggplot(data = cyclones,
            mapping = aes(x = FullAvg.mu,
                          y = FullAvg.rD))
p <- p + geom_point() + geom_smooth(se = FALSE)
p <- p + labs(x = "Fitted Values",
              y = "Deviance Residuals")
p

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

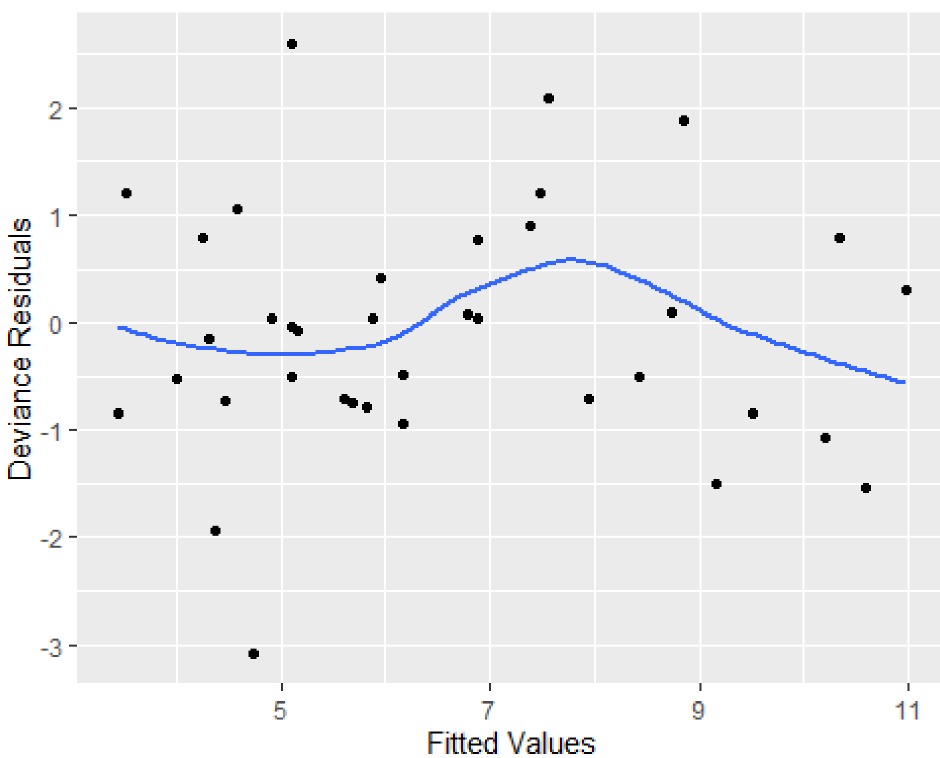


Figure 16: Which Residual Deviance is below -3, FullAvg plot

```
#Figure 16 in the Appendix
which(cyclones$FullAvg.rD < -3)

## [1] 26
```

Figure 17: Fitted Values vs Quantile Residuals plot for OND plot (Non-Severe)

```
#Figure 17 of the Appendix
library(statmod)
cyclones$OND.rQ <- qresid(OND.mod)

p <- ggplot(data = cyclones,
            mapping = aes(x = OND.mu,
                          y = OND.rQ))
p <- p + geom_point() + geom_smooth(se = FALSE)
p <- p + labs(x = "Fitted Values",
              y = "Quantile Residuals")
p
```

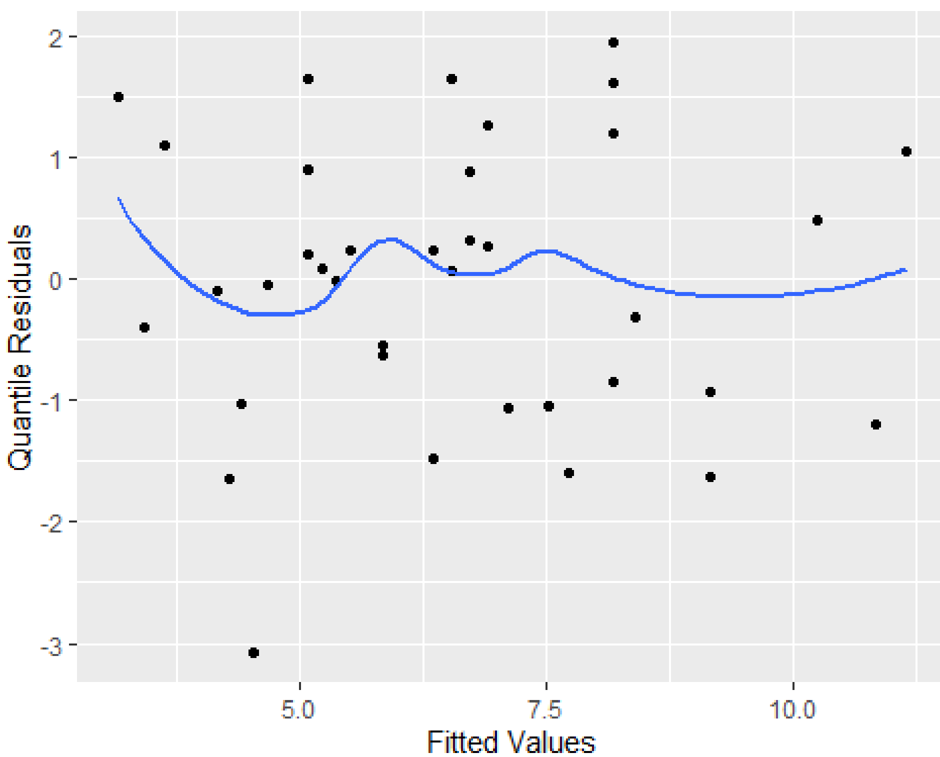


Figure 18: Fitted Values vs Quantile Residuals plot for FullAvg plot (Non-Severe)

```
#Figure 18 of the Appendix
library(statmod)
cyclones$FullAvg.rQ <- qresid(FullAvg.mod)

p <- ggplot(data = cyclones,
            mapping = aes(x = FullAvg.mu,
                          y = FullAvg.rQ))
p <- p + geom_point() + geom_smooth(se = FALSE)
p <- p + labs(x = "Fitted Values",
              y = "Quantile Residuals")
p
```

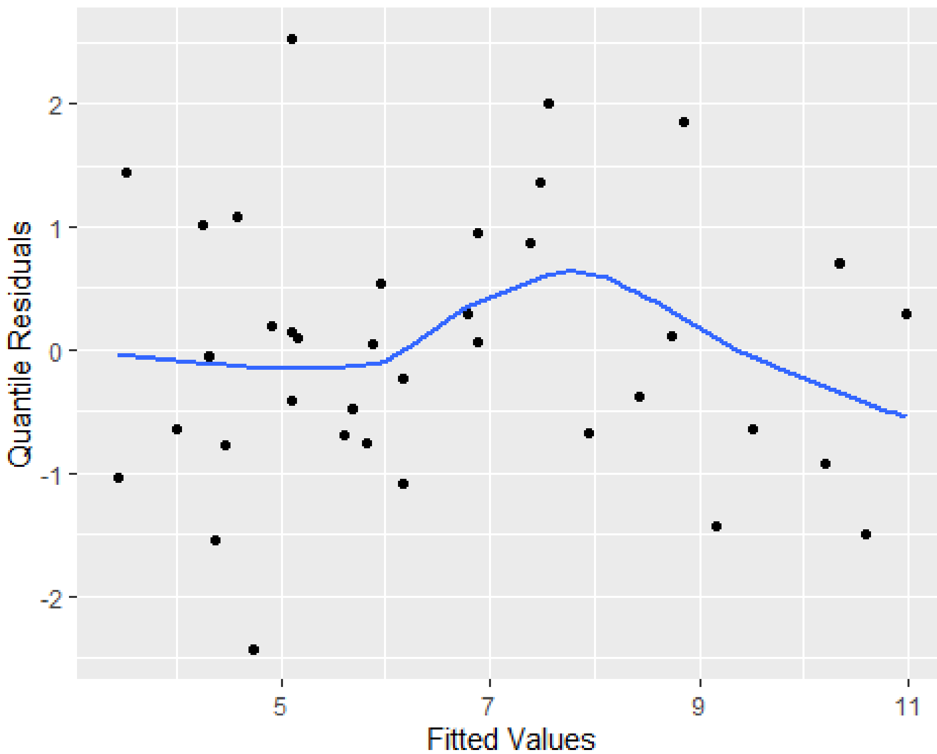


Figure 19: OND vs Quantile Residuals plot (Non-Severe)

```
#Figure 19 of the Appendix
p <- ggplot(data = cyclones,
            mapping = aes(x = OND,
                          y = OND.rQ))
p <- p + geom_point() + geom_smooth(se = FALSE)
p <- p + labs(x = "OND",
```

```
              y = "Quantile Residuals")
p
```

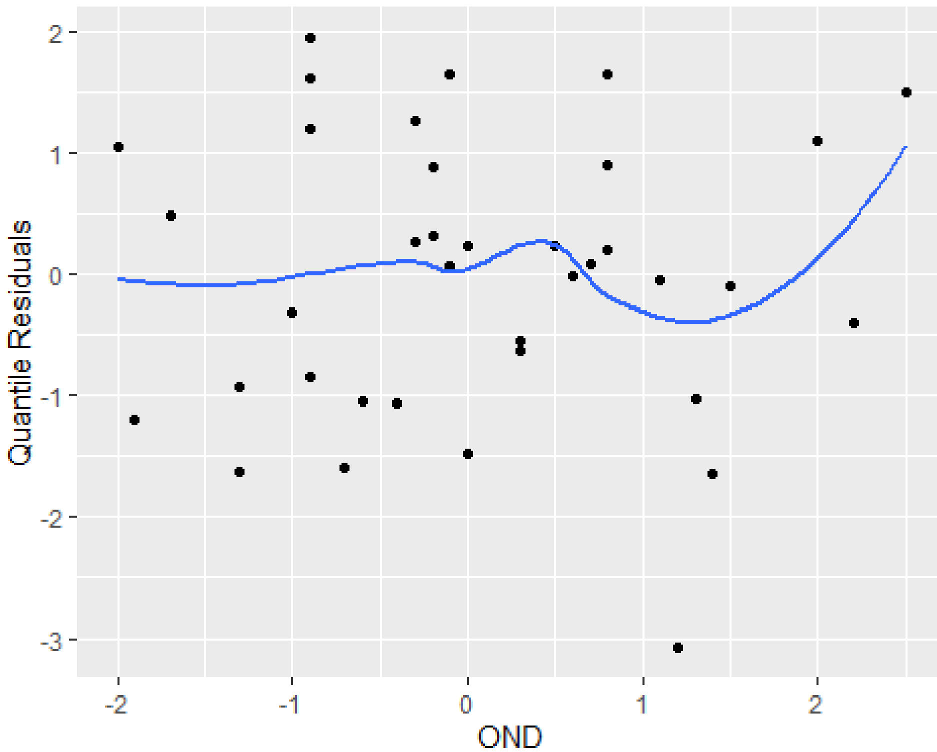


Figure 20: FullAvg vs Quantile Residuals plot (Non-Severe)

```
#Figure 20 of the Appendix
p <- ggplot(data = cyclones,
            mapping = aes(x = FullAvg,
                          y = FullAvg.rQ))
p <- p + geom_point() + geom_smooth(se = FALSE)
p <- p + labs(x = "FullAvg",
              y = "Quantile Residuals")
p
```

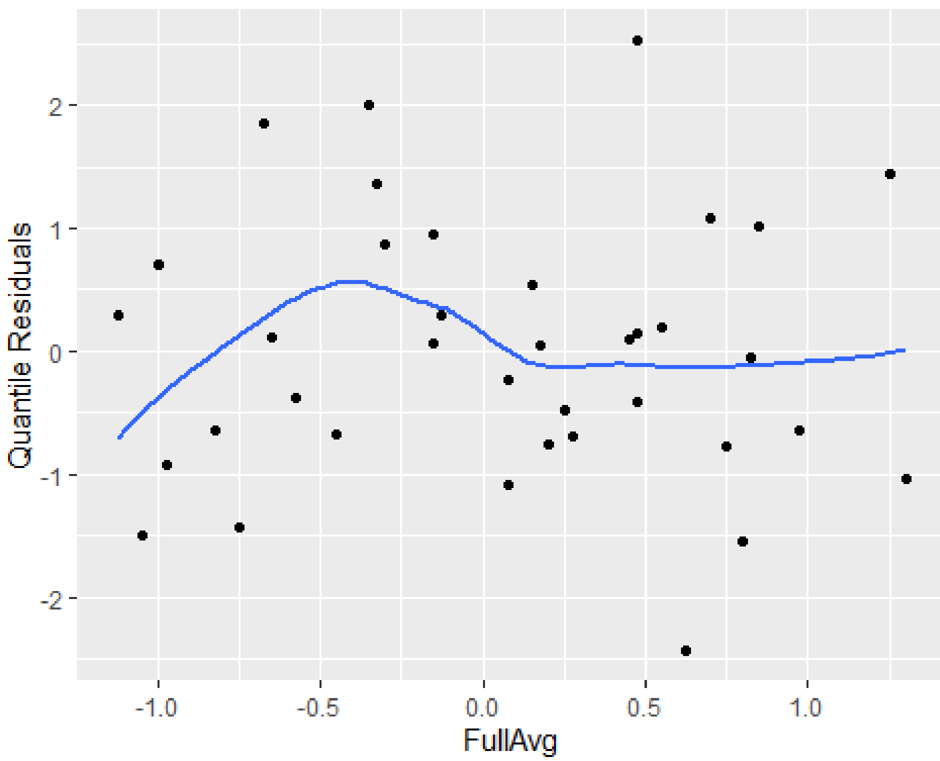


Figure 21: Working Response vs Linear Predictor plot for OND (Non-Severe)

```
#Figure 21 in the Appendix

cyclones$OND.eta <- predict(OND.mod, type = "link")
cyclones$OND.rW <- resid(OND.mod, type = "working")

p <- ggplot(data = cyclones,
            mapping = aes(x = OND.eta + OND.rW,
                          y = OND.eta))
p <- p + geom_point()
p <- p + labs(x = "Working Response",
              y = "Linear Predictor")
p <- p + geom_abline(intercept = 0, slope = 1,
                     color = "red")
p
```

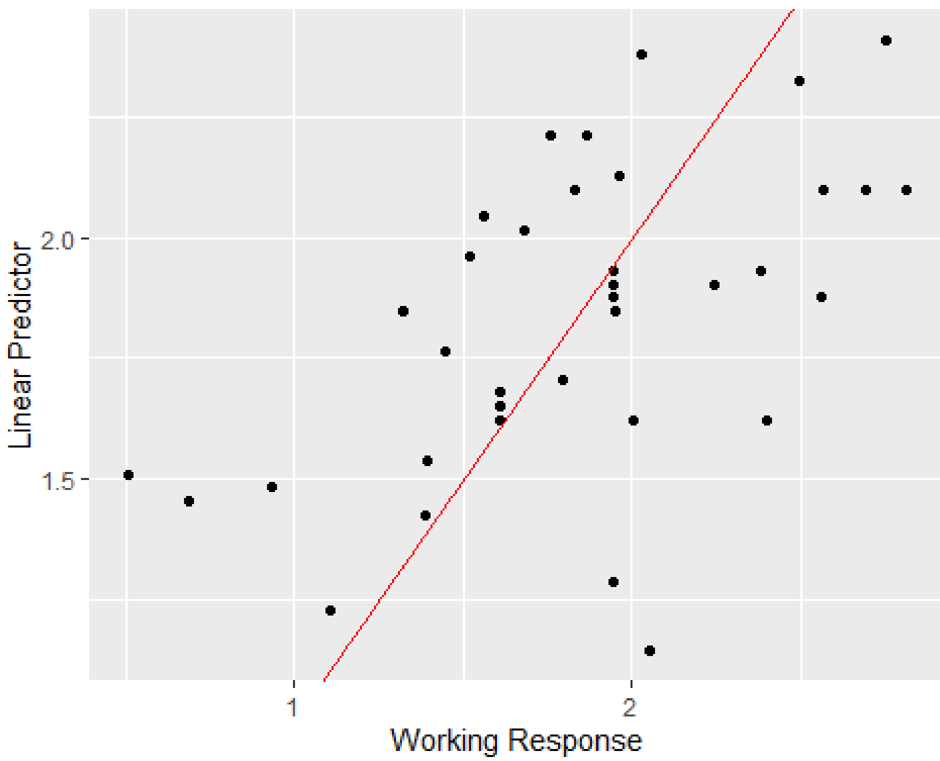


Figure 22: Working Response vs Linear Predictor plot for FullAvg (Non-Severe)

```
#Figure 22 in the Appendix

cyclones$FullAvg.eta <- predict(FullAvg.mod, type = "link")
cyclones$FullAvg.rW <- resid(FullAvg.mod, type = "working")

p <- ggplot(data = cyclones,
            mapping = aes(x = FullAvg.eta + FullAvg.rW,
                          y = FullAvg.eta))
p <- p + geom_point()
p <- p + labs(x = "Working Response",
              y = "Linear Predictor")
p <- p + geom_abline(intercept = 0, slope = 1,
                     color = "red")
p
```

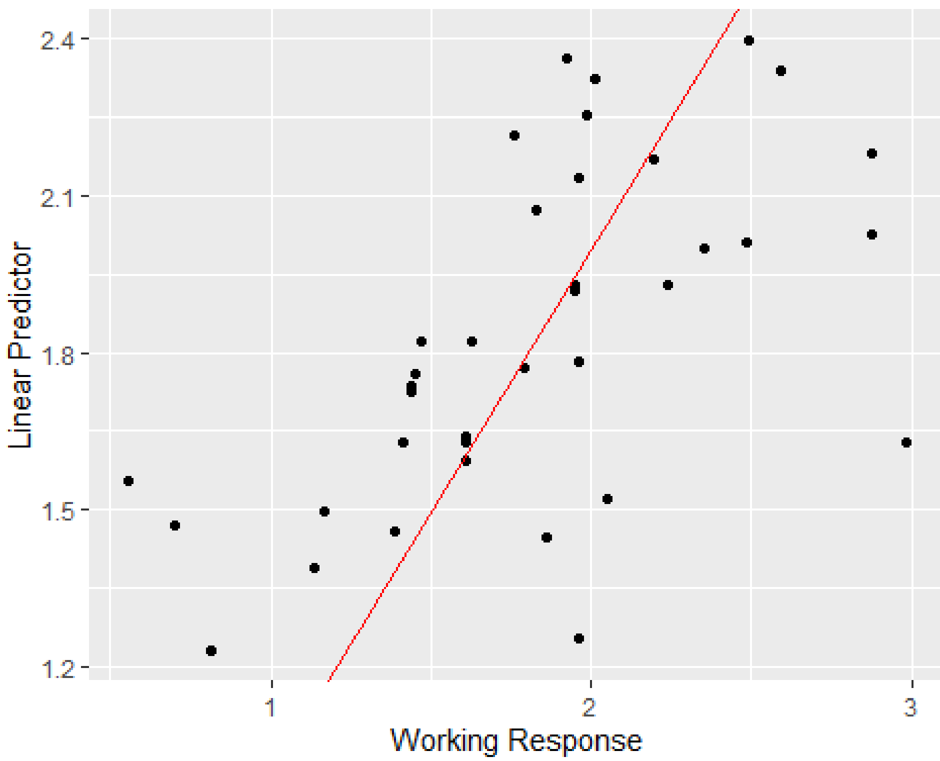


Figure 23: Q-Q Plot for Deviance Residuals of OND Model (Non-Severe)

```
#Figure 23 of the Appendix
qqnorm(cyclones$OND.rD)
abline(0, 1, col = "red")
```

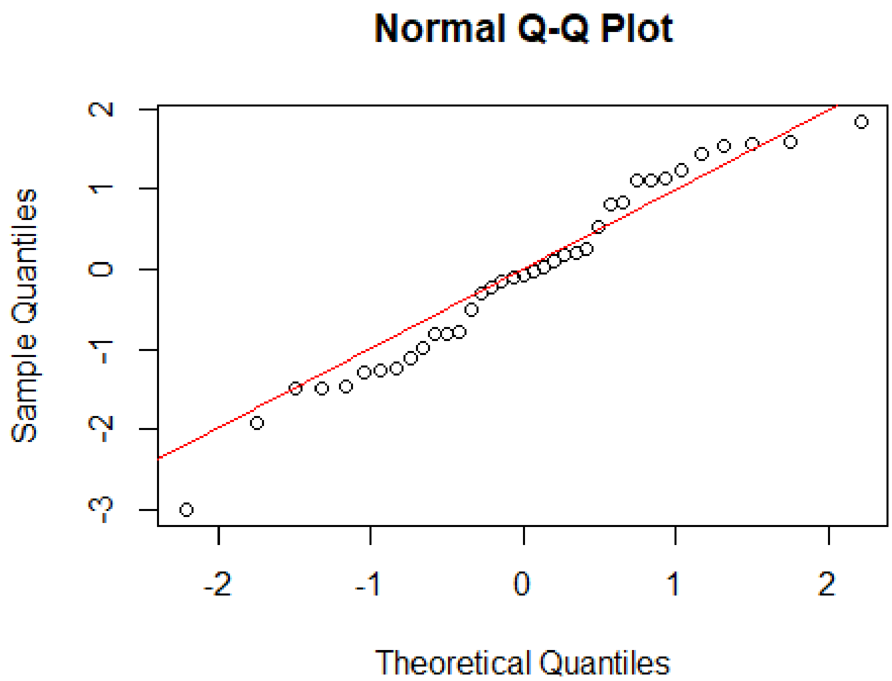


Figure 24: Q-Q Plot for Deviance Residuals of FullAvg Model (Non-Severe)

```
#Figure 24 of the Appendix
qqnorm(cyclones$FullAvg.rD)
abline(0, 1, col = "red")
```

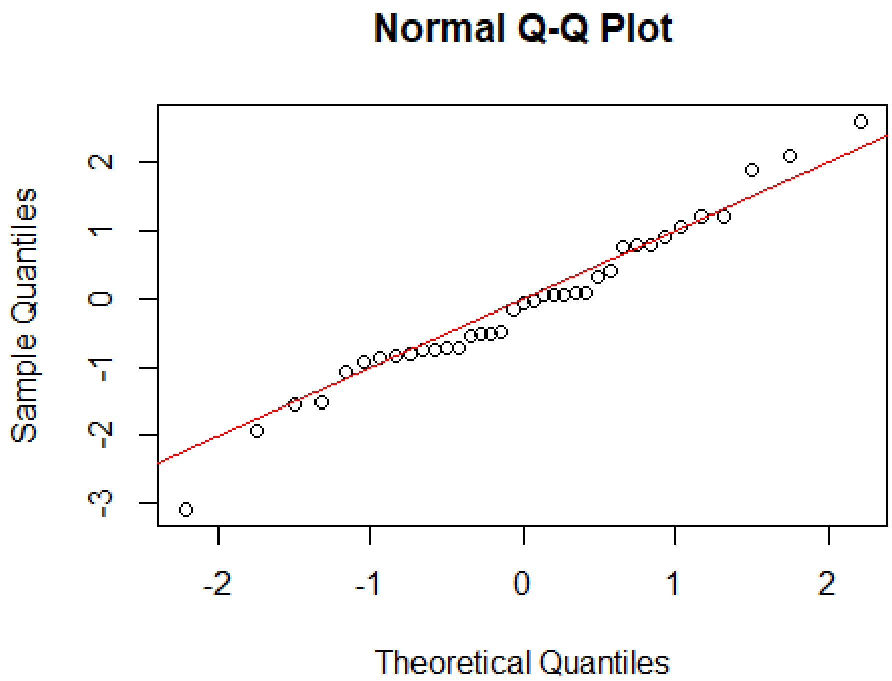


Figure 25: Q-Q Plot for Quantile Residuals of OND Model (Non-Severe)

```
#Figure 25 for the Appendix
qqnorm(cyclones$OND.rQ)
abline(0, 1, col = "red")
```

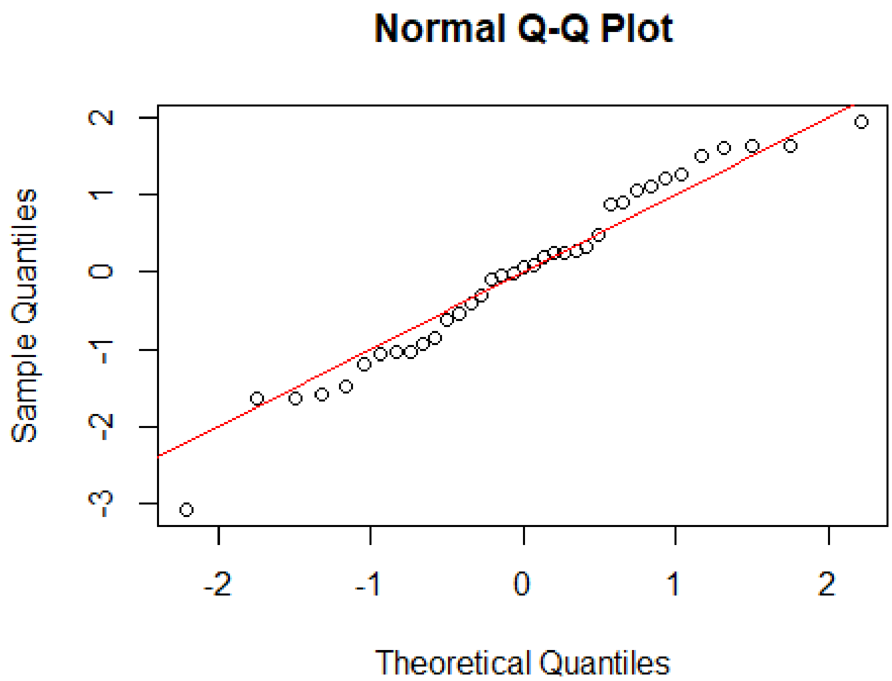


Figure 26: Q-Q Plot for Quantile Residuals of FullAvg Model (Non-Severe)

```
#Figure 26 for the Appendix
qqnorm(cyclones$FullAvg.rQ)
abline(0, 1, col = "red")
```

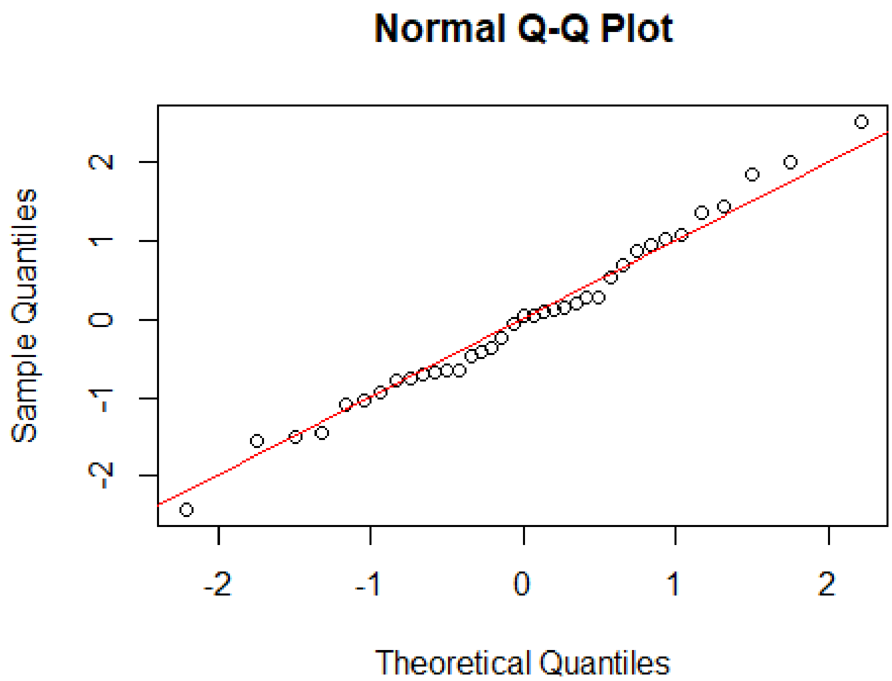


Figure 27: Term Plot for OND Model (Non-Severe)

```
#Figure 27 of the Appendix
termplot(OND.mod, partial.resid = TRUE, terms = "OND")
```

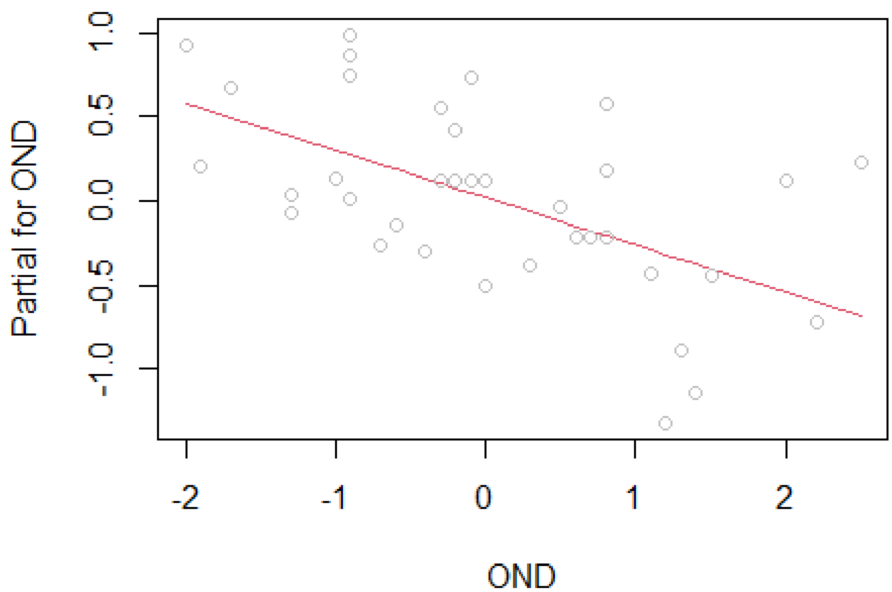


Figure 28: Term Plot for FullAvg Model (Non-Severe)

```
#Figure 28 of the Appendix
termplot(FullAvg.mod, partial.resid = TRUE, terms = "FullAvg")
```

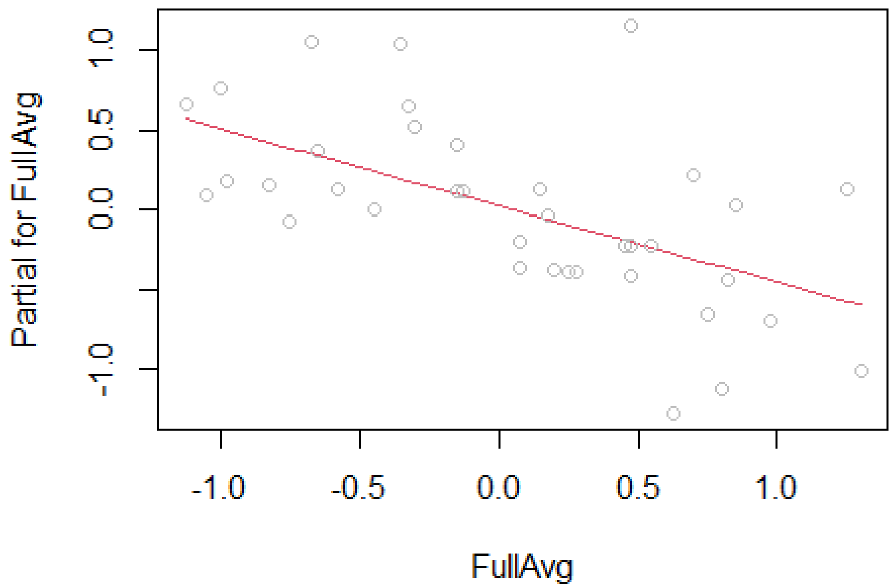


Figure 29: Summary of Naive Mod (Severe)

```
#Figure 29 in the Appendix
Naive.Mod <- glm(Severe ~ 1,
              data = cyclones,
              family = poisson(link = "log"))
summary(Naive.Mod)

##
## Call:
## glm(formula = Severe ~ 1, family = poisson(link = "log"), data = cyclones)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1517  -0.6561  -0.1995   0.6315   2.0811
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.69735    0.07036   24.12   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
```

```
##
##     Null deviance: 32.22  on 36  degrees of freedom
## Residual deviance: 32.22  on 36  degrees of freedom
## AIC: 163.37
##
## Number of Fisher Scoring iterations: 4
```

Figure 30: Summary of OND Mod (Severe)

```
#Figure 30 in the Appendix
OND.Mod <- glm(Severe ~ OND,
             data = cyclones,
             family = poisson(link = "log"))
summary(OND.Mod)
```

```
##
## Call:
## glm(formula = Severe ~ OND, family = poisson(link = "log"), data = cyclones)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3229  -0.8811  -0.1917   0.6931   1.8683
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.69808    0.07042  24.114   <2e-16 ***
## OND         -0.08698    0.06375  -1.364    0.172
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 32.220  on 36  degrees of freedom
## Residual deviance: 30.343  on 35  degrees of freedom
## AIC: 163.5
##
## Number of Fisher Scoring iterations: 4
```

Figure 31: Fitted Values vs Deviance Residuals Plot for Naive Model (Severe)

```
#Figure 31 in Appendix
cyclones$Naive.mu <- predict(Naive.Mod, type = "response")
cyclones$Naive.rD <- resid(Naive.Mod, type = "deviance")
```

```
p <- ggplot(data = cyclones,
            mapping = aes(x = Naive.mu,
                          y = Naive.rD))
p <- p + geom_point() + geom_smooth(se = FALSE)
p <- p + labs(x = "Fitted Values",
              y = "Deviance Residuals")
p
```

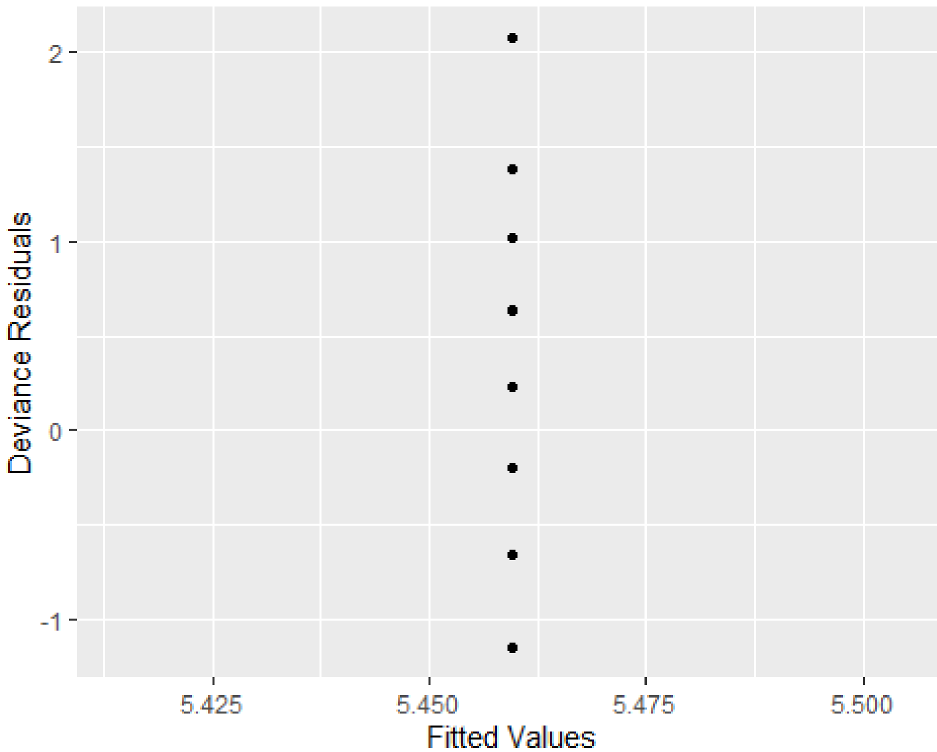


Figure 32: Fitted Values vs Deviance Residuals Plot for OND Model (Severe)

```
#Figure 32 in Appendix
cyclones$OND2.mu <- predict(OND.Mod, type = "response")
cyclones$OND2.rD <- resid(OND.Mod, type = "deviance")

p <- ggplot(data = cyclones,
            mapping = aes(x = OND2.mu,
                          y = OND2.rD))
p <- p + geom_point() + geom_smooth(se = FALSE)
p <- p + labs(x = "Fitted Values",
              y = "Deviance Residuals")
p
```

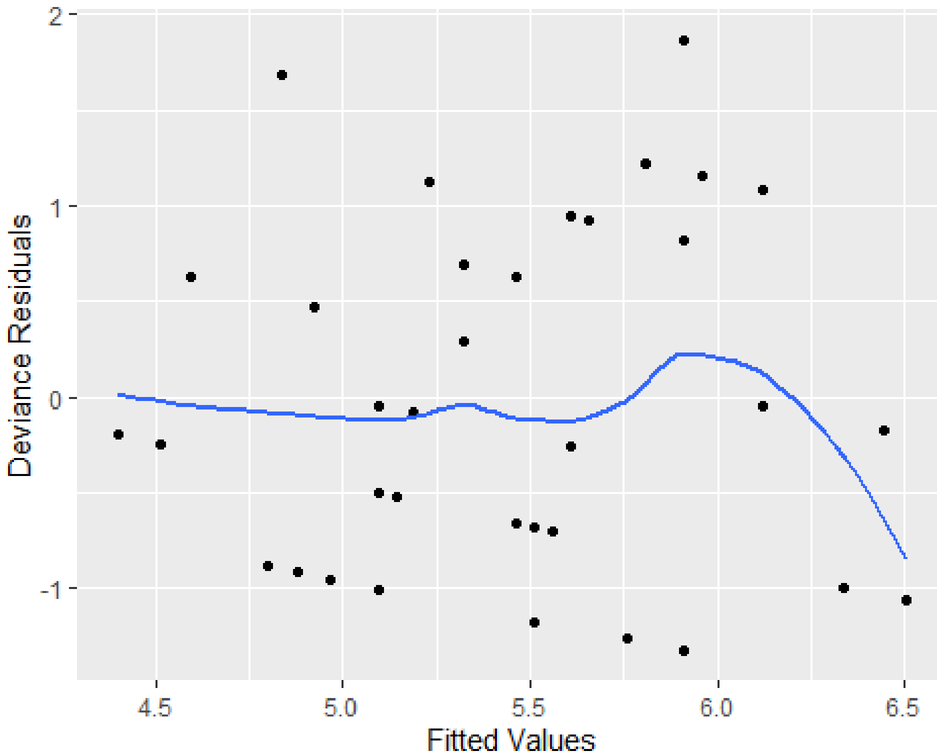


Figure 33: Fitted Values vs Quantile Residuals Plot for Naive Model (Severe)

```
#Figure 33 of the Appendix
cyclones$Naive.rQ <- qresid(Naive.Mod)

p <- ggplot(data = cyclones,
            mapping = aes(x = Naive.mu,
                          y = Naive.rQ))
p <- p + geom_point() + geom_smooth(se = FALSE)
p <- p + labs(x = "Fitted Values",
              y = "Quantile Residuals")
p
```

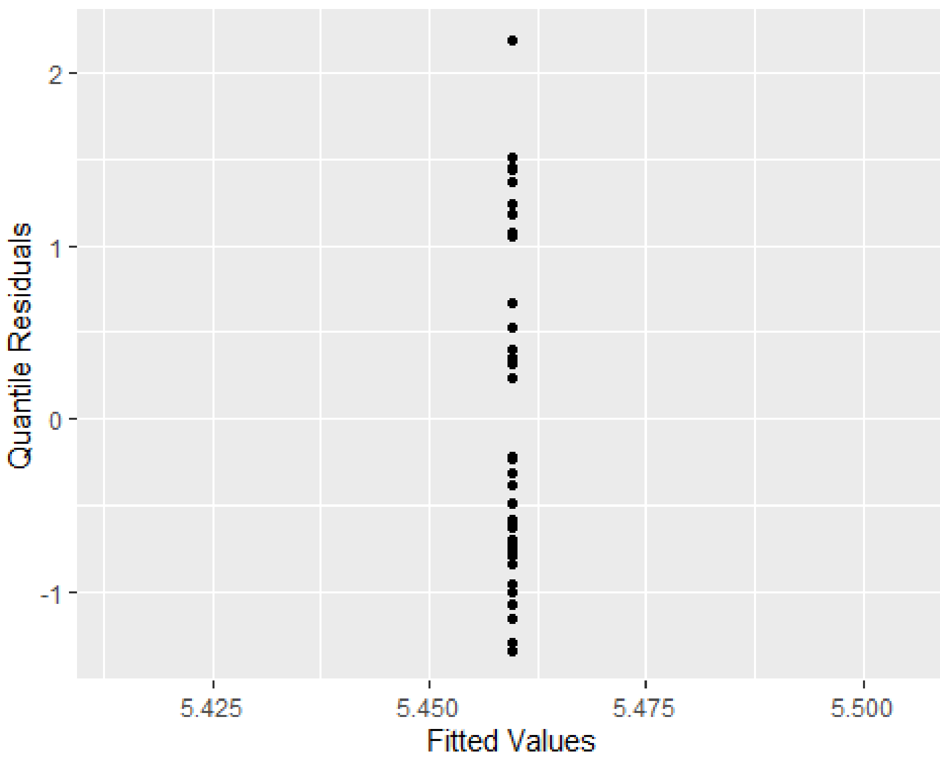


Figure 34: Fitted Values vs Quantile Residuals Plot for OND Model (Severe)

```
#Figure 34 of the Appendix
cyclones$OND2.rQ <- qresid(OND.Mod)

p <- ggplot(data = cyclones,
            mapping = aes(x = OND2.mu,
                          y = OND2.rQ))
p <- p + geom_point() + geom_smooth(se = FALSE)
p <- p + labs(x = "Fitted Values",
              y = "Quantile Residuals")
p
```

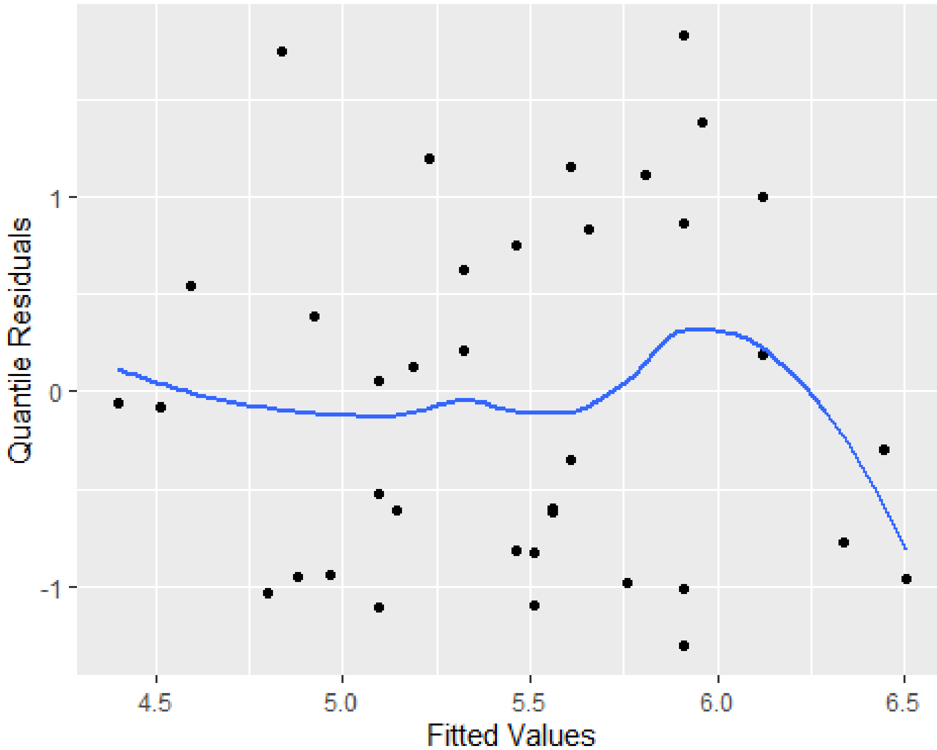


Figure 35: 1 vs Quantile Residuals Plot (Severe)

```
#Figure 35 of the Appendix
p <- ggplot(data = cyclones,
            mapping = aes(x = 1,
                          y = Naive.rQ))
p <- p + geom_point() + geom_smooth(se = FALSE)
p <- p + labs(x = "1",
              y = "Quantile Residuals")
p
```

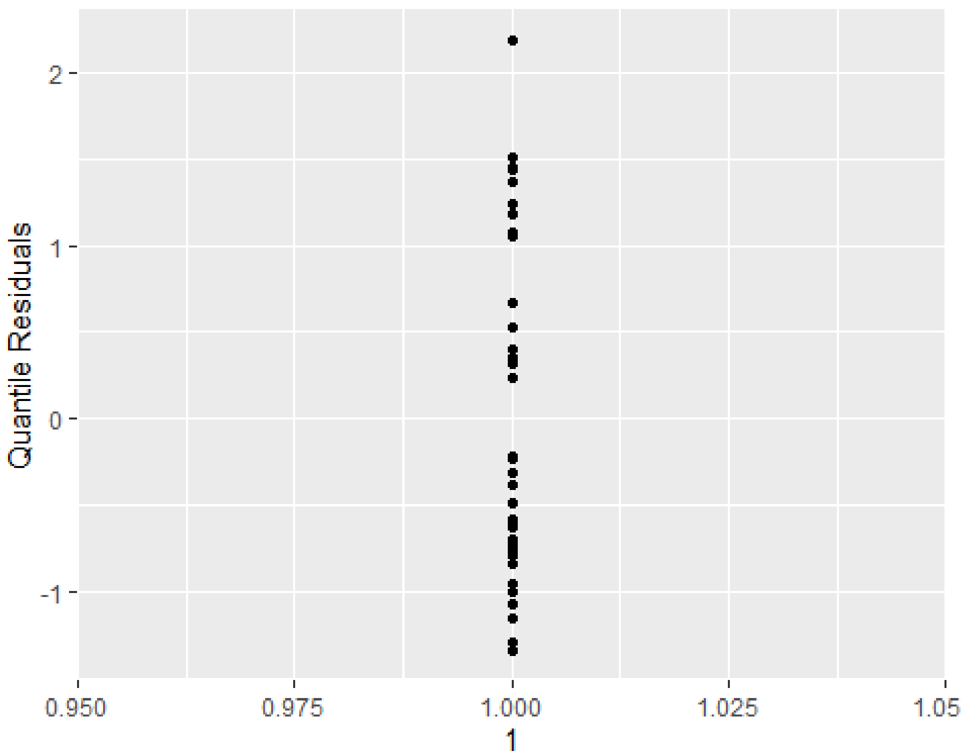


Figure 36: OND vs Quantile Residuals Plot (Severe)

```
#Figure 36 of the Appendix
p <- ggplot(data = cyclones,
            mapping = aes(x = OND,
                          y = OND2.rQ))
p <- p + geom_point() + geom_smooth(se = FALSE)
p <- p + labs(x = "OND",
              y = "Quantile Residuals")
p
```

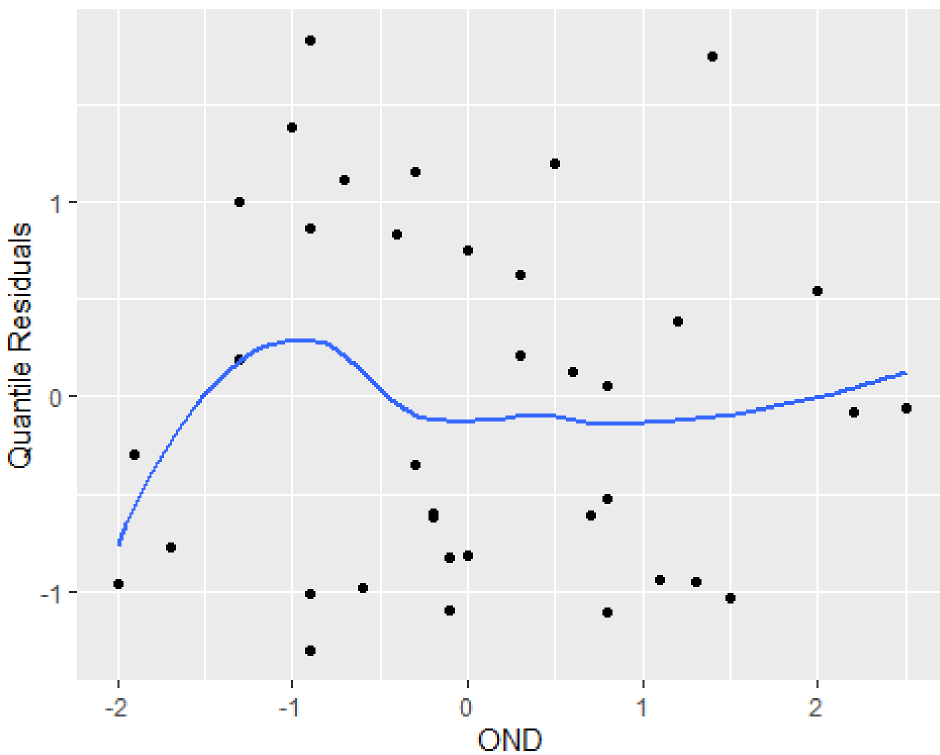


Figure 37: Working Response vs Linear Predictor Plot for Naive Model (Severe)

```
#Figure 37 in the Appendix

cyclones$Naive.eta <- predict(Naive.Mod, type = "link")
cyclones$Naive.rW <- resid(Naive.Mod, type = "working")

p <- ggplot(data = cyclones,
            mapping = aes(x = Naive.eta + Naive.rW,
                          y = Naive.eta))
p <- p + geom_point()
p <- p + labs(x = "Working Response",
              y = "Linear Predictor")
p <- p + geom_abline(intercept = 0, slope = 1,
                     color = "red")
p
```

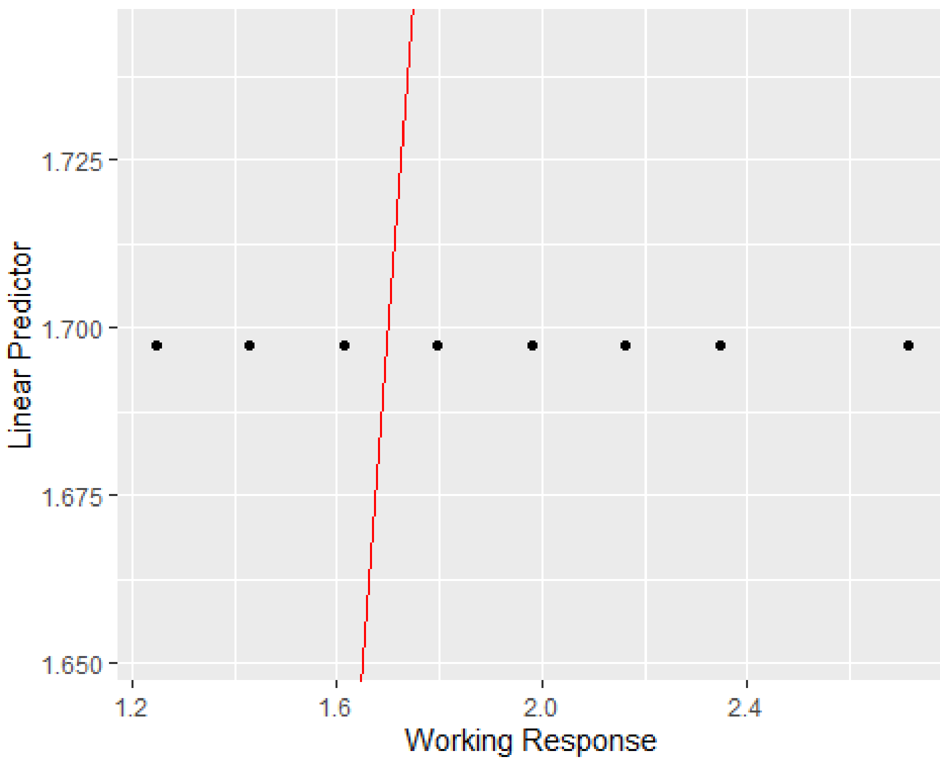


Figure 38: Working Response vs Linear Predictor Plot for OND Model (Severe)

```
#Figure 38 in the Appendix

cyclones$OND2.eta <- predict(OND.Mod, type = "link")
cyclones$OND2.rW <- resid(OND.Mod, type = "working")

p <- ggplot(data = cyclones,
            mapping = aes(x = OND2.eta + OND2.rW,
                          y = OND2.eta))
p <- p + geom_point()
p <- p + labs(x = "Working Response",
              y = "Linear Predictor")
p <- p + geom_abline(intercept = 0, slope = 1,
                     color = "red")
p
```

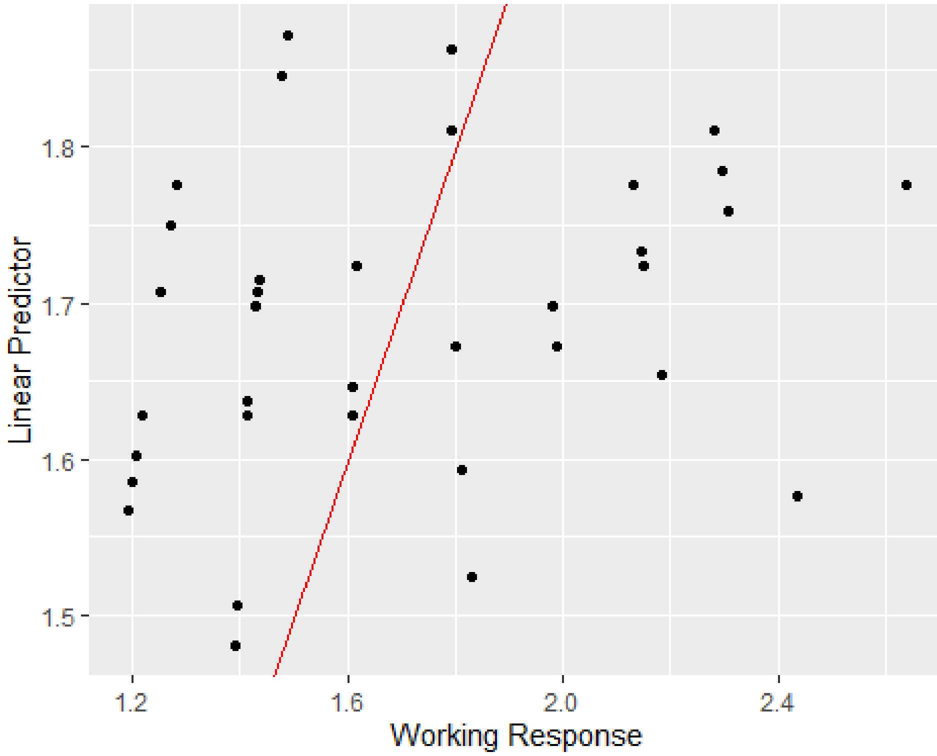


Figure 39: Q-Q Plot for Deviance Residuals of Naive Model (Severe)

```
#Figure 39 of the Appendix
qqnorm(cyclones$Naive.rD)
abline(0, 1, col = "red")
```

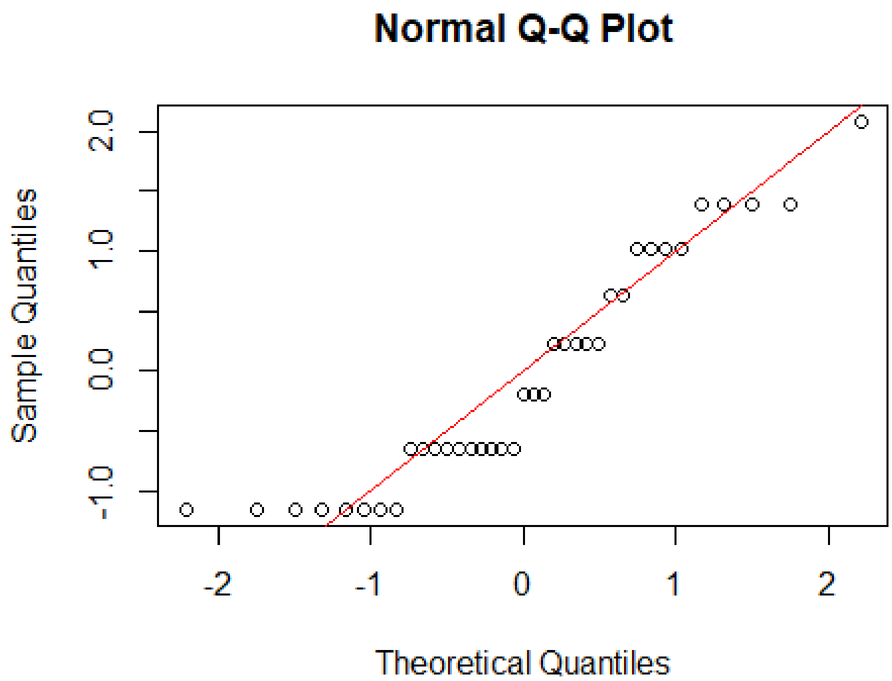


Figure 40: Q-Q Plot for Deviance Residuals of OND Model (Severe)

```
#Figure 40 of the Appendix
qqnorm(cyclones$OND2.rD)
abline(0, 1, col = "red")
```

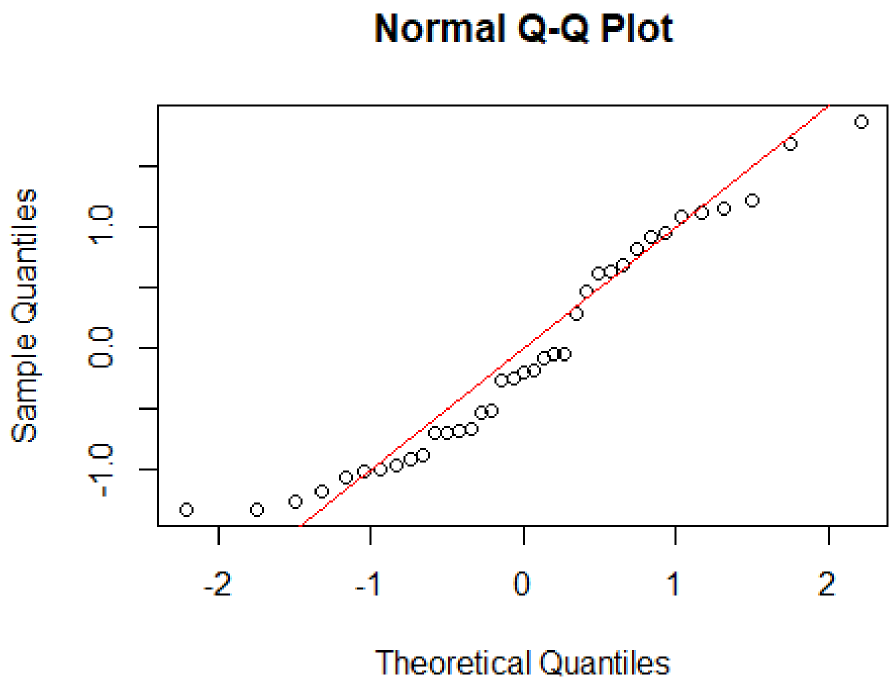


Figure 41: Q-Q Plot for Quantile Residuals of Naive Model (Severe)

```
#Figure 41 of the Appendix
qqnorm(cyclones$Naive.rQ)
abline(0, 1, col = "red")
```

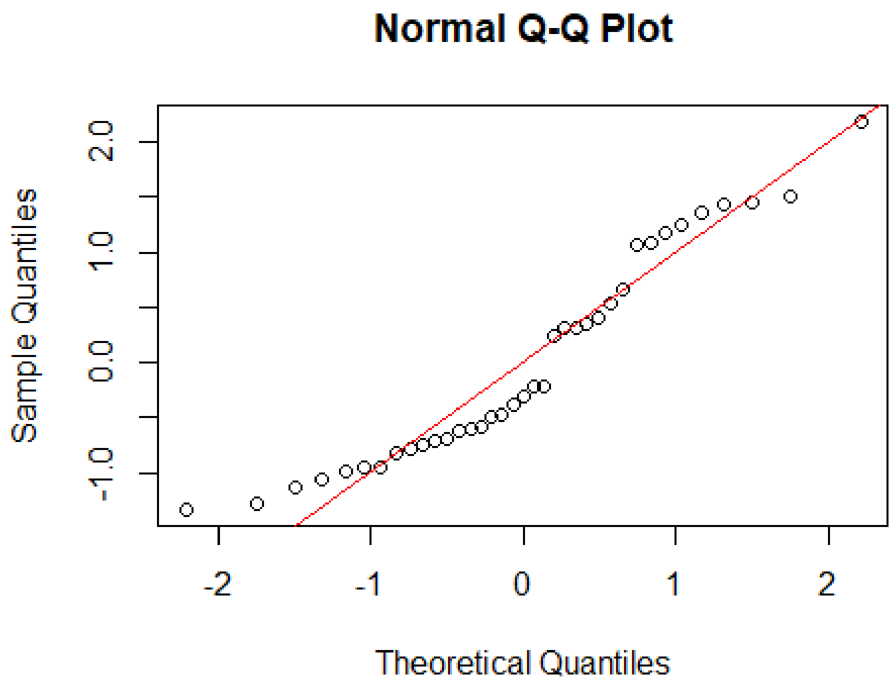


Figure 42: Q-Q Plot for Quantile Residuals of OND Model (Severe)

```
#Figure 42 of the Appendix
qqnorm(cyclones$OND2.rQ)
abline(0, 1, col = "red")
```

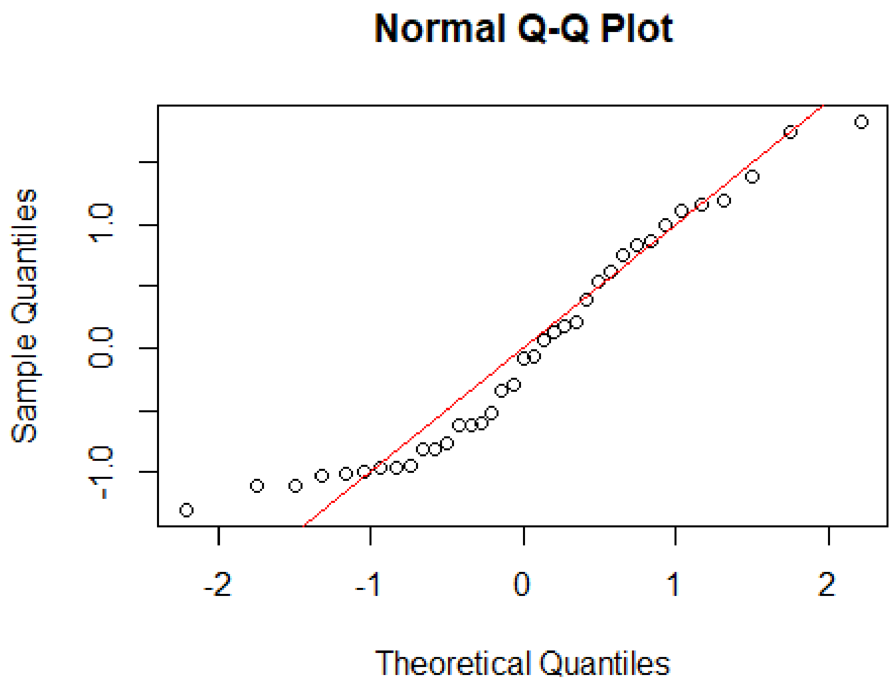
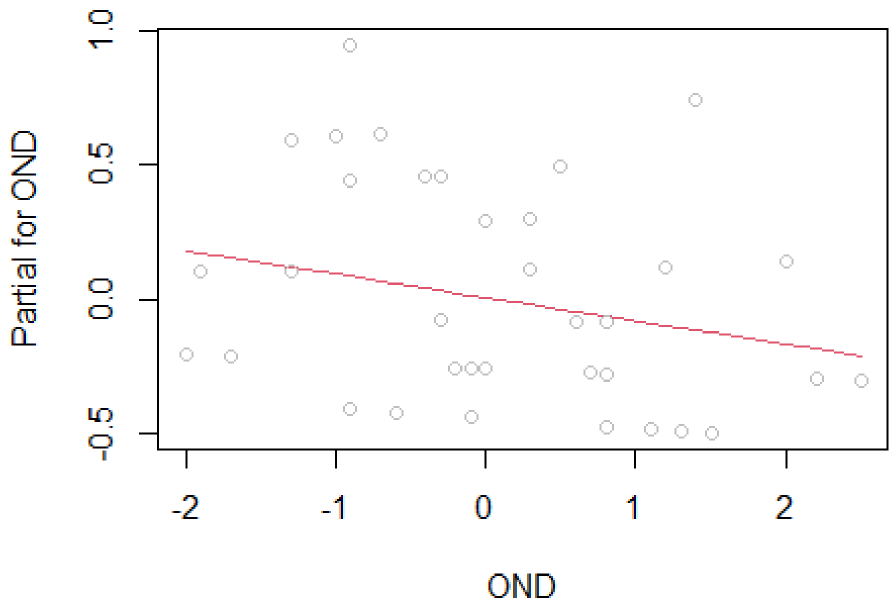


Figure 43: Term Plot for OND Model (Severe)

```
#Figure 43 of the Appendix
termplot(OND.Mod, partial.resid = TRUE, terms = "OND")
```


Partial for OND
1.0
0.5
0.0
-0.5
-2
-1
0
1
2
OND