Investigating factors that influence forest fires

Forest fires are a natural or human caused uncontrolled fire occurring in an area of combustible vegetation. Due to this disaster, each year a large area of forest is destroyed all around the world, causing human suffering, ecological and economical damage, and affects forest preservation (Cortaiz and Morais, 2007). Hence, early detection or prediction of behavior of disaster can help to fight against them by minimizing the damages (Tomas Artes, Ana Cortes, and Tomas Margalef, 2016).

This report is an extension of the previous report on "Investigating relation between meteorological factors and forest fire" (Refer to report D17124379PSICAPart1.pdf). In the previous report, an attempt was made to detect if natural or meteorological factors such as temperature, wind and relative humidity contribute to forest fires. It was observed that none of these factors alone are enough to cause forest fires or contribute to the spread of forest fires. However, these factors collectively can lead to forest fires and its spread. Hence, this report will attempt to determine if these factors collectively influence forest fires or not. Factors like fuel moisture codes which are influenced by natural factors such as Fine Fuel Moisture Code(FFMC), Duff Moisture Code(DMC), Drought Code(DC) and Initial Spread Index(ISI), are also considered in this report.

The purpose of this research is to investigate if two or more factors mentioned above causes forest fire or its spread. The technique used to achieve this will be Multiple Linear Regression. Through various regression models, attempts will be made to predict values of the continuous dependent variable "area" from several predictors as mentioned above. If this research is successful to detect a number of factors that causes forest fire or its spread, then this information can be used to alert team of fire fighters, warning public in advance and take any other necessary steps to fight and reduce damage caused by it.

Methodology

The dataset used in this analysis was collected from January 2000 to December 2003 and was built from two sources by Paulo Cortaiz and Anibal Morais (2007). In one database, records such as, date, time, spatial location, kind of vegetation, burned area were entered on a daily basis by the inspector who was in charge of Montesinho park, Portugal fire occurrences. Another database was composed by

Braganca Polytechnic Institute. In this a number of weather observations were noted within every 30 minute period by the local meteorological station. The data from the above two databases was then integrated by Cortez and Morais into a single dataset with 517 cases. Table 1 displays the list of data variables selected by Cortez and Morais.

Attribute Description	
X	x-axis coordinate: 1 to 9
Y	y-axis coordinate: 1 to 9
month	January to December
day	Monday to Sunday
FFMC	Fine Fuel Moisture Code: 18.7 to 96.20
DMC	Duff Moisture Code: 1.1 to 291.3
DC	Drought Code: 7.9 to 860.6
ISI	Initial Spread Index: 0.0 to 56.10
temp	temperature (in ◦C): 2.2 to 33.30
RH	relative humidity (in %): 15.0 to 100
wind	wind speed (in km/h): 0.40 to 9.40
rain	rain (in mm/m2): 0.0 to 9.40
area	Total burned area (in ha): 0.0 to 6.4

Table 1: Description of Variables of Interest for Analysis

Variables of Interest

Independent variables:

- **FFMC (continuous variable):** influences ignition and fire spread
- **DMC & DC (continuous variables):** contributes to depth of burn and overall fire intensity
- **ISI (continuous variable):** score that coordinates with fire velocity spread
- **temp, RH, wind (continuous variables):** meteorological variables. Another meteorological variable 'rain' is not considered for this research as data contained under rain variable depicts that either no rain or very little rain is experienced in the north eastern region of Portugal during January 2000 to December 2003
- **Season (nominal variable) :** this nominal variable was created from 'Month' variable. This was done to reduce the number of categories from 12 months

to 4 seasons of the year. Basis for grouping months under a season is taken from the article "Seasons in Portugal" and are coded as below:

March to May - Spring (1)
Jun - Aug - Summer (2)
Sept to Nov - Autumn (3)
Dec to Feb - Winter (4)

These categories may be re-coded as dummy variables as the need be for regression models.

Dependent variables:

- **Area (continuous variable):** For the cases where the value for area is zero means that the area burned by forest fire was lower than 1ha/100 = $100m^2$. This dependent variable is very skewed towards 0.0, as small fires were observed in Portugal during the time this data was collected. Hence, these cases cannot be ignored.

Tests for Normality

Results for descriptive statistics can be seen in Table 2.
Zscore was calculated for these variables and the results were saved to the dataset. Looking at the Standardized scores (value/ standard error) for skewness and kurtosis for all the variables in Table 2 it was found that the scores were lying outside the range of -2 and +2. Also, the histograms (included in the appendix) clearly shows the skewness of the variables and hence the variables should not be treated as normal.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
FFMC	517	18.7	96.2	90.645	5.5201	30.472	-6.576	.107	67.066	.214
DMC	517	1.1	291.3	110.872	64.0465	4101.952	.547	.107	.205	.214
DC	517	7.9	860.6	547.940	248.0662	61536.835	-1.100	.107	-.245	.214
ISI	517	.0	56.1	9.022	4.5595	20.789	2.536	.107	21.458	.214
Temp	517	2.2	33.3	18.889	5.8066	33.717	-.331	.107	.136	.214
RH	517	15	100	44.29	16.317	266.260	.863	.107	.438	.214
Wind	517	.4	9.4	4.018	1.7917	3.210	.571	.107	.054	.214
Area	517	.00	1090.84	12.8473	63.65582	4052.063	12.847	.107	194.141	.214
Valid N (listwise)	517									

Table 2: Descriptive Statistics for all the variables of interest

After observing the normality of the data, an attempt was made to transform the data to lessen skewness and revamp symmetry by using Lg10 function for positively skewed variables such as DMC, RH and inverse of variable was carried out if the variable was strongly skewed such as for Area and FFMC variable. The skewness of the variables could not be completely eradicated from the variables except in case of the variable RH (refer to Appendix for descriptive statistics) with standardized skewness of 0.2616 (SE = 0.107) and kurtosis of -1.8785 (SE = 0.214).

Throughout the research the statistical significance level used as a cut-off is 0.05.

Baseline Model

For this model independent variables chosen are temperature, relative humidity, wind and season. As stated above in the 'Introduction' section, in the last research no correlation was found between meteorological factors (temperature, relative humidity and wind) alone with the area affected by forest fires. This model will try to find out if these natural factors collectively contributes to the spread of forest fires. An important factor taken into consideration for this model other than meteorological factors is 'Season'.

Before going ahead with the regression model, an independent samples test was carried out to investigate if there is any variance for the considered meteorological factors across seasons of the years.

Model with Fuel Codes as predictors

The independent variables considered for this model are FFMC, DMC and DC. The reason to choose these variables for the analysis is that fuel codes are affected by natural factors mentioned above and influences forest fire (Cortaiz and Morais, 2007). A Spearman's rho test was run to assess the relationship between these variables and area affected by forest fire (refer to Appendix).

Model with Duff Moisture Code(DMC) and Drought Code(DC)

The independent variables used in this model are DMC and DC. These two variables contributes to the depth of burn and overall fire intensity. As in this data set, the area spread is not huge, hence it can be a good idea to consider the fire intensity and it's affect in the particular area. FFMC is dropped as a factor for this model as it influences ignition and does not contribute to the overall fire intensity.

Results

Baseline Model

The hypothesis for the baseline model is stated as:

H_0 : Area affected by forest fire cannot be predicted by temperature, relative humidity, wind and season taken together.

H_A : Area affected by forest fire can be predicted by temperature, relative humidity, wind and season taken together.

It can be seen from the test summary table below, that distribution of temperature, relative humidity and wind is not the same across all the seasons ($p < 0.05$). Hence, considering these factors together to predict area affected by forest fires is a good idea.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Temp is the same across categories of Seasons.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.
2	The distribution of RH_log is the same across categories of Seasons.	Independent-Samples Kruskal-Wallis Test	.004	Reject the null hypothesis.
3	The distribution of Wind is the same across categories of Seasons.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Regression Results

Variable	Coefficient	Std. Error	t-value	Pr(>|t|)
Intercept	-.462	.192	-2.411	<.05

Temp	.151	.064	2.345	<.05
RH	-.042	.055	-.756	>.05
Wind	-.090	.045	-1.999	<.05
Spring	.603	.219	2.759	<.05
Summer	.562	.215	2.616	<.05
Autumn	.367	.204	1.799	>.05

Summary Statistics	**Value**
f-statistic	2.808
Degrees of freedom	516
p-value	<.05
Adjusted R-square	.021

A multiple linear regression analysis was conducted to determine if season of the year and meteorological factors temperature, relative humidity and wind could predict the area affected by forest fires.

In order to include the seasons of the year in the regression model it was recorded into three variables season_spring (1 for spring, 0 for summer, autumn and winter), season_summer(1 for summer, 0 for spring, autumn and winter) and season_autumn(1 for autumn, 0 for summer, spring and winter).

Examination of the histogram, normal P-P plot of standardised residuals and the scatterplot of the dependent variable(refer to appendix), area, and standardised residuals showed that no outliers existed (100% within limits of -1.96 to +1.96 and none with Cook's distance >1s).

Tests to see if the data met the assumption of collinearity indicated that multicollinearity was a concern (Spring, Tolerance = .36, VIF = 2.77; Summer, Tolerance = .16, VIF = 6.027, Autumn, Tolerance= .19, VIF= 5.072). The scatterplot of standardised residuals showed that the data did not meet the assumptions of homogeneity of variance and linearity. The data also did not meet the assumption of non-zero variances of the predictors.

Hence, **this model fails to reject the null hypothesis**. This means a combination of temperature, relative humidity, wind and season cannot predict the area affected by forest fire.

Model with Fuel Codes as predictors

The hypothesis for this model is stated as:

H_0 : Area affected by forest fire cannot be predicted by FFMC, DMC and DC taken together.

H_A : Area affected by forest fire can be predicted by FFMC, DMC and DC taken together.

As can be observed from the results of Spearman's rho test (refer to appendix), there is a very very weak negative correlation between FFMC and Area with no statistical significance($r(515) = -.025$, $p > 0.05$). There is a very weak negative correlation between DMC and Area with no statistical significance($r(515) = -.072$, $p > 0.05$). And, there is a very weak positive correlation between DC and Area with no statistical significance($r(515) = .062$, $p > 0.05$).

Regression Results

Variable	**Coefficient**	**Std. Error**	**t-value**	**Pr(>|t|)**
Intercept	-.462	.044	.000	>.05
FFMC	-.035	.049	-.709	>.05
DMC	.016	.065	.239	>.05
DC	.094	.060	1.561	>.05

Summary Statistics	**Value**
f-statistic	1.494
Degrees of freedom	516

p-value	>.05
Adjusted R-square	.003

A multiple linear regression analysis was conducted to determine if fuel codes FFMC, DMC and DC could predict the area affected by forest fires.

Examination of the histogram, normal P-P plot of standardised residuals and the scatterplot of the dependent variable(refer to appendix), area, and standardised residuals showed that no outliers existed (100% within limits of -1.96 to +1.96 and none with Cook's distance >1s).

Examination for multicollinearity showed that the tolerance and variance influence factor measures were within acceptable levels (tolerance > 0.4, VIF < 2.5). The scatterplot of standardised residuals showed that the data did not meet the assumptions of homogeneity of variance and linearity. The data also did not meet the assumption of non-zero variances of the predictors.

Hence, **this model fails to reject the null hypothesis**. This means a combination of fuel codes FFMC, DMC and DC cannot predict the area affected by forest fire.

Model with Duff Moisture Code and Drought Code

The hypothesis for this model is stated as:

H_0 : Area affected by forest fire cannot be predicted by DMC and DC taken together.

H_A : Area affected by forest fire can be predicted by DMC and DC taken together.

Regression Results

Variable	Coefficient	Std. Error	t-value	$Pr(>\lvert t\rvert)$
Intercept	-5.272E-16	.044	.000	>.05
DMC	.004	.059	-.072	>.05
DC	.025	.059	1.445	>.05

Summary Statistics	Value
f-statistic	1.992
Degrees of freedom	516
p-value	>.05
Adjusted R-square	.004

A multiple linear regression analysis was conducted to determine if fuel codes DMC and DC could predict the area affected by forest fires.

Examination of the histogram, normal P-P plot of standardised residuals and the scatterplot of the dependent variable(refer to appendix), area, and standardised residuals showed that no outliers existed (100% within limits of -1.96 to +1.96 and none with Cook's distance >1s).

Examination for multicollinearity showed that the tolerance and variance influence factor measures were within acceptable levels (tolerance > 0.4, VIF < 2.5). The scatterplot of standardised residuals showed that the data did not meet the assumptions of homogeneity of variance and linearity. The data also did not meet the assumption of non-zero variances of the predictors.

Hence, **this model fails to reject the null hypothesis**. This means a combination of fuel codes DMC and DC cannot predict the area affected by forest fire.

Discussion

The regression models presented above tried to come up with the factors that could influence the area affected by forest fires. However, it was observed that the data did not meet the assumptions required to build a regression model. Some of the assumptions that were violated were random normal distribution of errors, homoscedasticity, linearity, non zero variances. Hence, the models built did not prove out to be good fits for the data. In future, some other models such as support vector machines, random forests can be considered for the analysis of factors influencing forest fires.

One important issue with the original set of data was that the target variable 'Area' consisted of almost 50% values close to zero, as small forest fires might have occurred during the time of data collection. Due to this ignoring these values would have affected the data. This extreme skewness could have been the reason affecting the data as whole when fitting in the regression models.

In future, reasons for small forest fires can be examined such as, what factors are preventing spread of forest fires, even if the forest fire was small but how much was the intensity of it that is how much damage did it cause. For this data such as, casualties caused, number and condition of trees and shrubs burnt can be collected. Moreover, factors such as type of vegetation, human causes, time of fire etcetera can also be considered to further explore and determine the cause and spread of forest fires.

References:

P. Cortez and A. Morais (2007). A Data Mining Approach to Predict Forest Fires using Meteorological Data. Forest Fires. Retrieved from http://archive.ics.uci.edu/ml/datasets/Forest+Fires

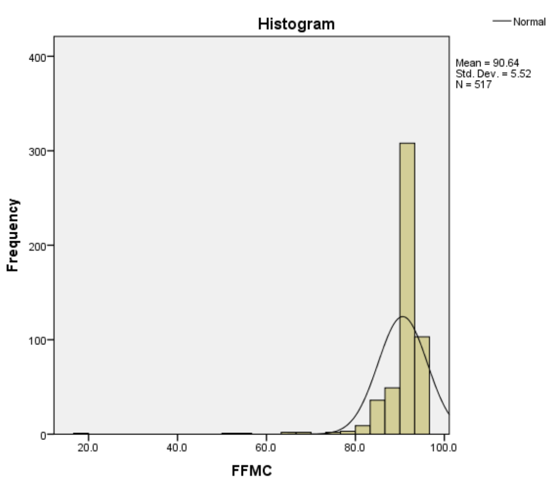
Tomàs Artés, Ana Cortés, Tomàs Margalef (2016). Large Forest Fire Spread Prediction: Data and Computational Science. *Procedia Computer Science, 80*, 909-918. https://doi.org/10.1016/j.procs.2016.05.330

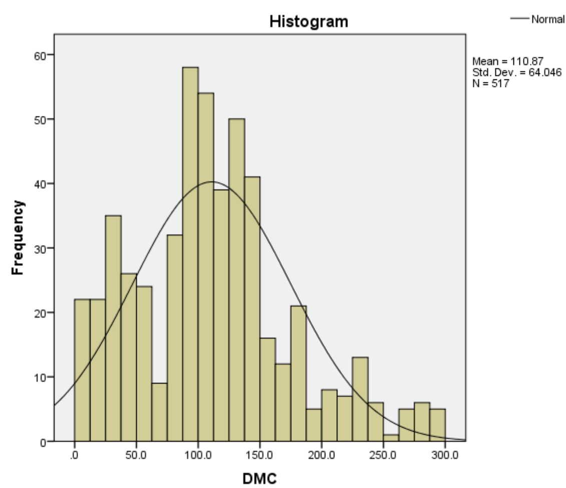
P. Cortez and A. Morais (2007). A Data Mining Approach to Predict Forest Fires using Meteorological Data. *In J. Neves, M. F. Santos and J. Machado Eds., New Trends in Artificial Intelligence, Proceedings of the 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence, December, Guimarães, Portugal, 512-523*. http://www3.dsi.uminho.pt/pcortez/fires.pdf

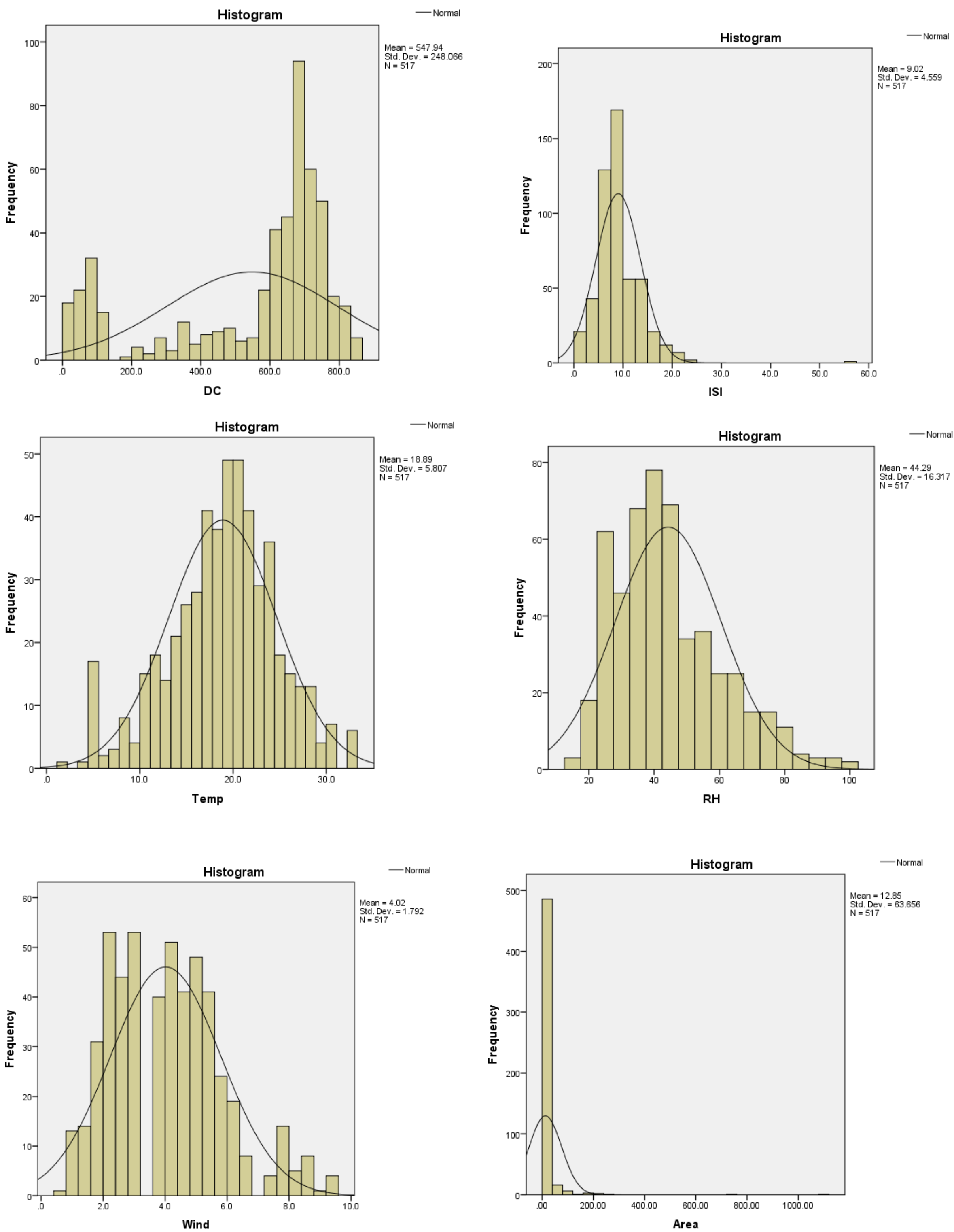
Seasons of the Year (2016). Seasons in Portugal. Retrieved from https://seasonsyear.com/Portugal

Appendix

Descriptive Statistics to check normality







Histograms for the variables of interest

Transforming Data and examining it to check for normality

```
COMPUTE FFMC_inverse=1/ ((96.2+1) - FFMC).
```

```
EXECUTE.
COMPUTE DMC_log=LG10(DMC).
EXECUTE.
COMPUTE DC_log=LG10((860.6+1)-DC).
EXECUTE.
COMPUTE ISI_log=LG10(ISI).
EXECUTE.
COMPUTE Temp_log=LG10((33.1+1) - Temp).
EXECUTE.
COMPUTE RH_log=LG10(RH).
EXECUTE.
COMPUTE Wind_log=LG10(Wind).
EXECUTE.
COMPUTE Area_inverse=1/(Area+1).
EXECUTE.

DESCRIPTIVES VARIABLES=FFMC_inverse DMC_log DC_log ISI_log Temp_log RH_log
Wind_log Area_inverse
  /SAVE
  /STATISTICS=MEAN STDDEV VARIANCE MIN MAX KURTOSIS SKEWNESS.
```

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
FFMC_inverse	517	.01	1.00	.2125	.14244	.020	2.896	.107	10.903	.214
DMC_log	517	.04	2.46	1.9319	.38236	.146	-1.693	.107	3.485	.214
DC_log	517	.00	2.93	2.3593	.37396	.140	-.957	.107	4.309	.214
ISI_log	517	-1.00	1.75	.8950	.26050	.068	-1.955	.107	8.677	.214
Temp_log	517	-.10	1.50	1.1425	.20730	.043	-1.784	.107	6.413	.214
RH_log	517	1.18	2.00	1.6181	.15686	.025	.028	.107	-.402	.214
Wind_log	517	-.40	.97	.5556	.21685	.047	-.715	.107	.717	.214
Area_inverse	517	.00	1.00	.5875	.41958	.176	-.167	.107	-1.800	.214
Valid N (listwise)	517									

Correlations

```
NONPAR CORR
  /VARIABLES=FFMC Area
  /PRINT=SPEARMAN TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Correlations

			FFMC	Area
Spearman's rho	FFMC	Correlation Coefficient	1.000	-.025
		Sig. (2-tailed)	.	.566
		N	517	517
	Area	Correlation Coefficient	-.025	1.000
		Sig. (2-tailed)	.566	.
		N	517	517

```
NONPAR CORR
  /VARIABLES=DMC Area
  /PRINT=SPEARMAN TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Correlations

			DMC	Area
Spearman's rho	DMC	Correlation Coefficient	1.000	-.072
		Sig. (2-tailed)	.	.102
		N	517	517
	Area	Correlation Coefficient	-.072	1.000
		Sig. (2-tailed)	.102	.
		N	517	517

```
NONPAR CORR
  /VARIABLES=DC Area
  /PRINT=SPEARMAN TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Correlations

			DC	Area
Spearman's rho	DC	Correlation Coefficient	1.000	.062
		Sig. (2-tailed)	.	.162
		N	517	517
	Area	Correlation Coefficient	.062	1.000
		Sig. (2-tailed)	.162	.
		N	517	517

```
NONPAR CORR
  /VARIABLES=ISI Area
  /PRINT=SPEARMAN TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Correlations

			ISI	Area
Spearman's rho	ISI	Correlation Coefficient	1.000	-.012
		Sig. (2-tailed)	.	.777
		N	517	517
	Area	Correlation Coefficient	-.012	1.000
		Sig. (2-tailed)	.777	.

N	517	517

```
*Nonparametric Tests: Independent Samples.
NPTESTS
  /INDEPENDENT TEST (ZArea_inverse) GROUP (Seasons)
  /MISSING SCOPE=ANALYSIS USERMISSING=EXCLUDE
  /CRITERIA ALPHA=0.05  CILEVEL=95.
```

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Area is the same across categories of Seasons.	Independent-Samples Kruskal-Wallis Test	.064	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Baseline Model

Recoding Dummy Variable

```
RECODE Season (1=1) (2=0) (3=0) (4=0) INTO Season_Spring.
VARIABLE LABELS  Season_Spring 'Spring'.
EXECUTE.
RECODE Season (1=0) (2=1) (3=0) (4=0) INTO Season_Summer.
VARIABLE LABELS  Season_Summer 'Summer'.
EXECUTE.
RECODE Season (1=0) (2=0) (3=1) (4=0) INTO Season_Autumn.
VARIABLE LABELS  Season_Autumn 'Autumn'.
EXECUTE.
```

Checking for difference

```
*Nonparametric Tests: Independent Samples.
NPTESTS
  /INDEPENDENT TEST (ZTemp_log ZRH_log ZWind_log) GROUP (Seasons)
  /MISSING SCOPE=ANALYSIS USERMISSING=EXCLUDE
  /CRITERIA ALPHA=0.05  CILEVEL=95.
```

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Temp is the same across categories of Seasons.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.
2	The distribution of RH_log is the same across categories of Seasons.	Independent-Samples Kruskal-Wallis Test	.004	Reject the null hypothesis.
3	The distribution of Wind is the same across categories of Seasons.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

```
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT ZArea_inverse
  /METHOD=ENTER ZTemp_log ZRH_log ZWind_log Season_Spring Season_Summer
Season_Autumn
  /PARTIALPLOT ALL
  /SCATTERPLOT=(*ZRESID ,*ZPRED)
  /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)
  /CASEWISE PLOT(ZRESID) OUTLIERS(3)
  /SAVE PRED COOK LEVER ZRESID SRESID SDRESID.
```

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	16.504	6	2.751	2.808	.011[b]
	Residual	499.496	510	.979		
	Total	516.000	516			

a. Dependent Variable: Area

b. Predictors: (Constant), Autumn, Temp, Wind, Spring, RH_log, Summer

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.179[a]	.032	.021	.98964853	.056

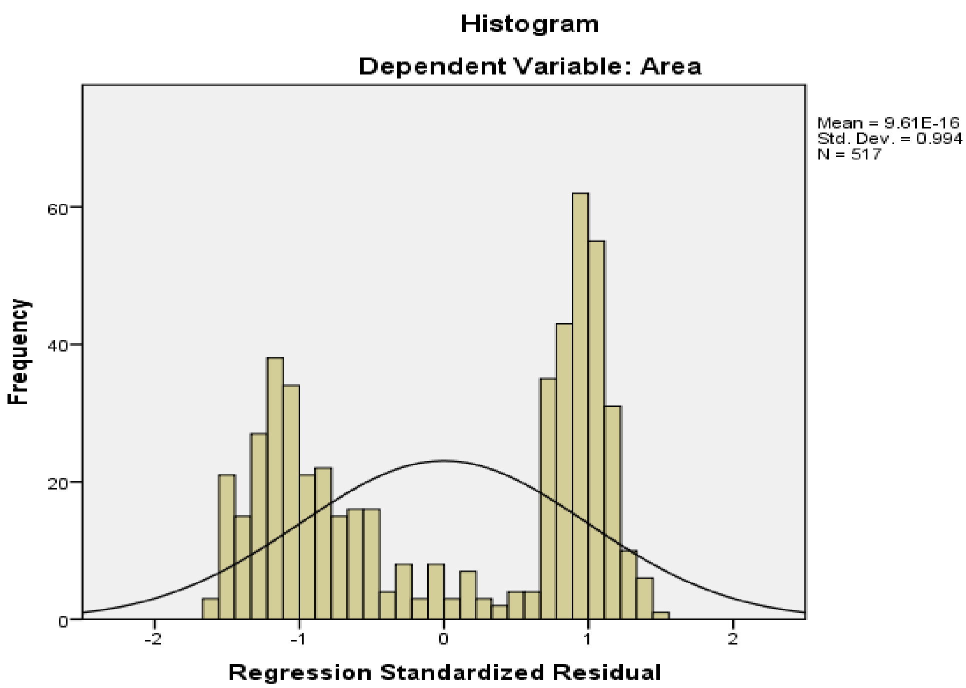
a. Predictors: (Constant), Autumn, Temp, Wind, Spring, RH_log, Summer

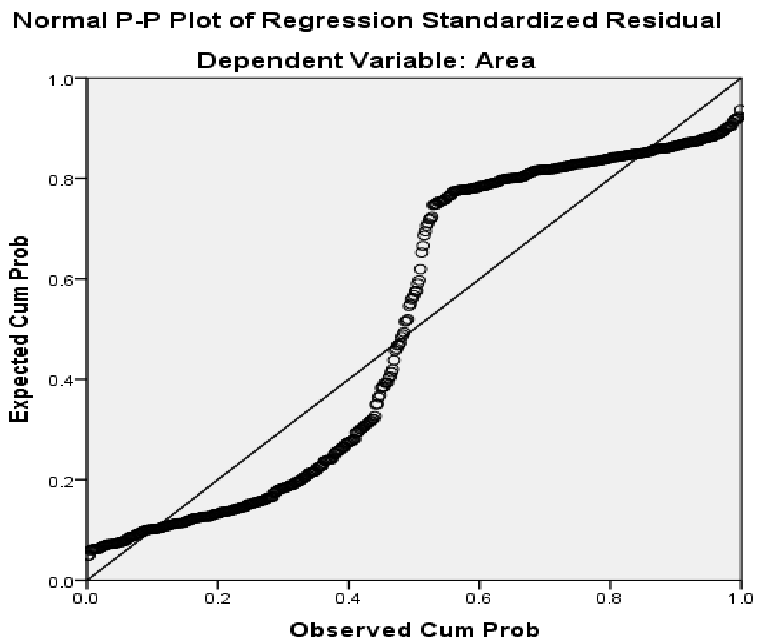
b. Dependent Variable: Area

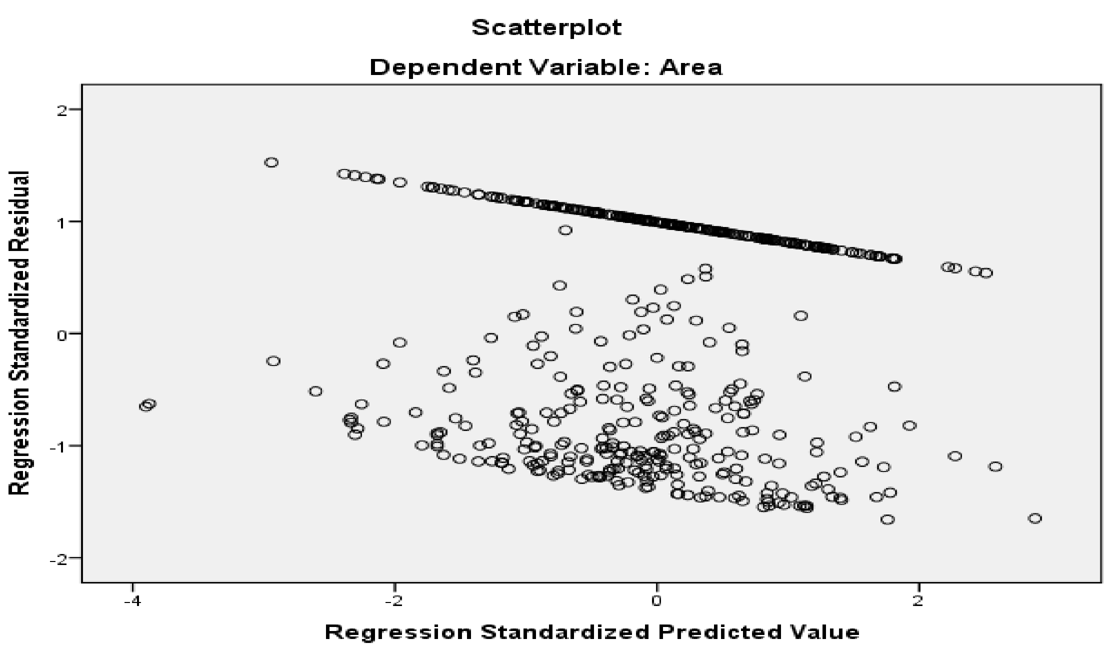
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-.462	.192		-2.411	.016	-.839	-.086					
	Temp	.151	.064	.151	2.345	.019	.024	.277	.077	.103	.102	.459	2.180
	RH_log	-.042	.055	-.042	-.756	.450	-.150	.067	.022	-.033	-.033	.627	1.596
	Wind	-.090	.045	-.090	-1.999	.046	-.178	-.002	-.059	-.088	-.087	.946	1.057
	Spring	.603	.219	.200	2.759	.006	.174	1.033	.090	.121	.120	.360	2.774
	Summer	.562	.215	.280	2.616	.009	.140	.984	.024	.115	.114	.166	6.027
	Autumn	.367	.204	.177	1.799	.073	-.034	.767	-.050	.079	.078	.197	5.072

a. Dependent Variable: Area







Model with Fuel codes as predictors

```
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT ZArea_inverse
  /METHOD=ENTER ZFFMC_inverse ZDMC_log ZDC_log
```

```
/PARTIALPLOT ALL
/SCATTERPLOT=(*ZRESID ,*ZPRED)
/RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)
/CASEWISE PLOT(ZRESID) OUTLIERS(3)
/SAVE PRED COOK LEVER ZRESID SRESID SDRESID.
```

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.469	3	1.490	1.494	.215[b]
	Residual	511.531	513	.997		
	Total	516.000	516			

a. Dependent Variable: Area

b. Predictors: (Constant), DC, FFMC, DMC

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.093[a]	.009	.003	.99856764	.018

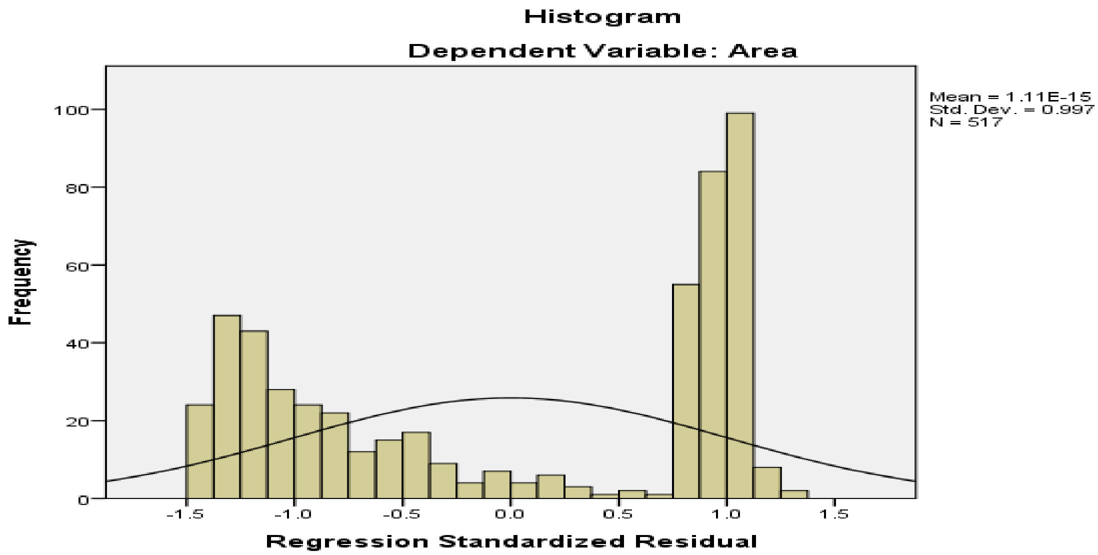
a. Predictors: (Constant), DC, FFMC, DMC

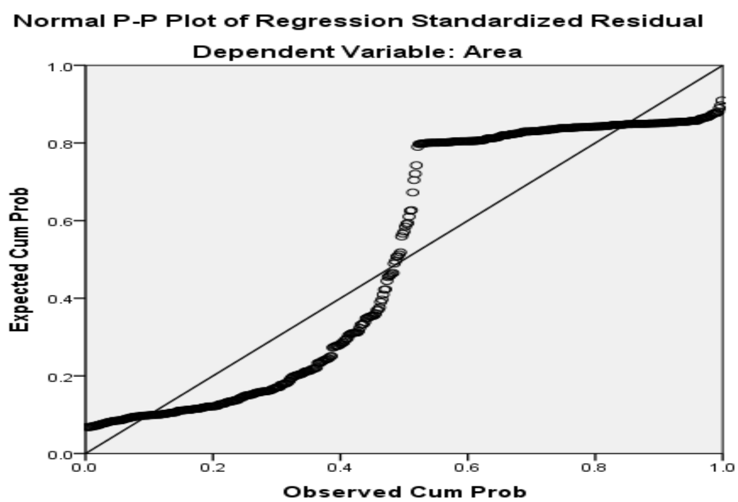
b. Dependent Variable: Area

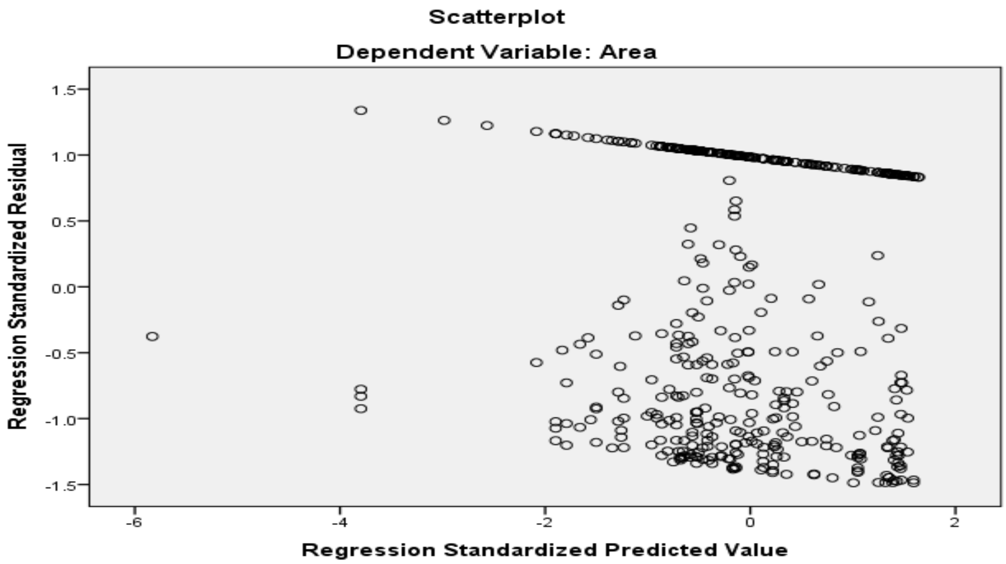
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-5.327E-16	.044		.000	1.000	-.086	.086					
	FFMC	-.035	.049	-.035	-.709	.479	-.131	.062	-.040	-.031	-.031	.805	1.243
	DMC	.016	.065	.016	.239	.811	-.112	.143	-.060	.011	.010	.457	2.187
	DC	.094	.060	.094	1.561	.119	-.024	.212	.088	.069	.069	.535	1.867

a. Dependent Variable: Area


Histogram
Dependent Variable: Area
Mean = 1.11E-15
Std. Dev. = 0.997
N = 517
Frequency
0
20
40
60
80
100
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
Regression Standardized Residual


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: Area
Expected Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0
Observed Cum Prob



Model with DMC and DC as predictors

```
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT ZArea_inverse
  /METHOD=ENTER ZDMC_log ZDC_log
  /PARTIALPLOT ALL
  /SCATTERPLOT=(*ZRESID ,*ZPRED)
  /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)
  /CASEWISE PLOT(ZRESID) OUTLIERS(3)
  /SAVE PRED COOK LEVER ZRESID SRESID SDRESID.
```

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3.968	2	1.984	1.992	.138[b]
	Residual	512.032	514	.996		
	Total	516.000	516			

a. Dependent Variable: Area

b. Predictors: (Constant), DC, DMC

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.088[a]	.008	.004	.99808385	.015

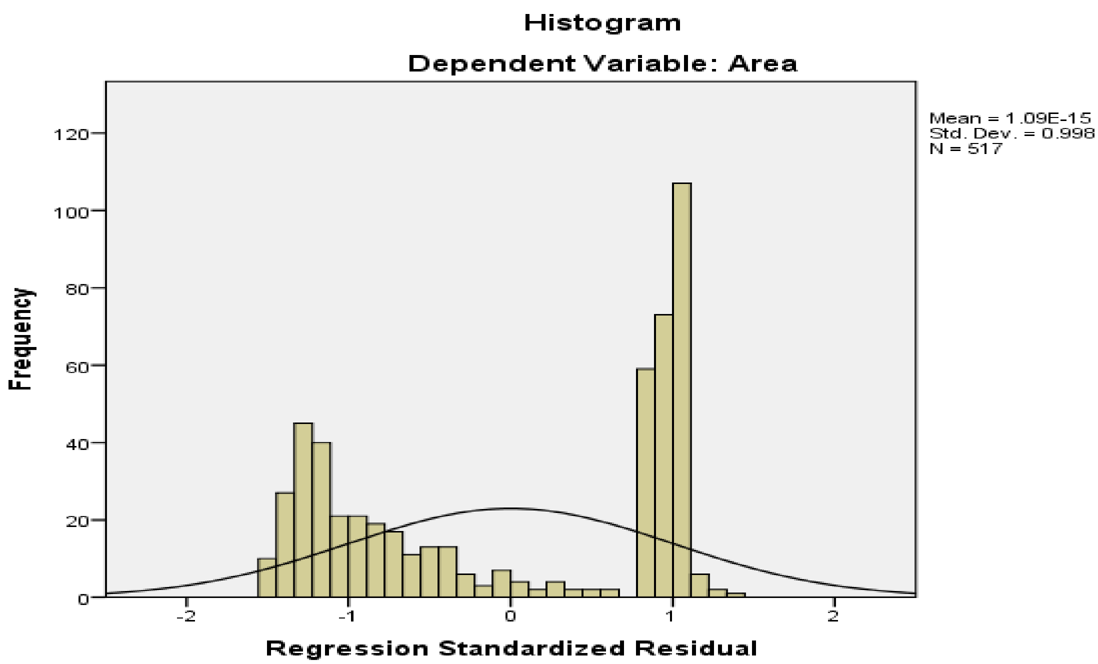
a. Predictors: (Constant), DC, DMC

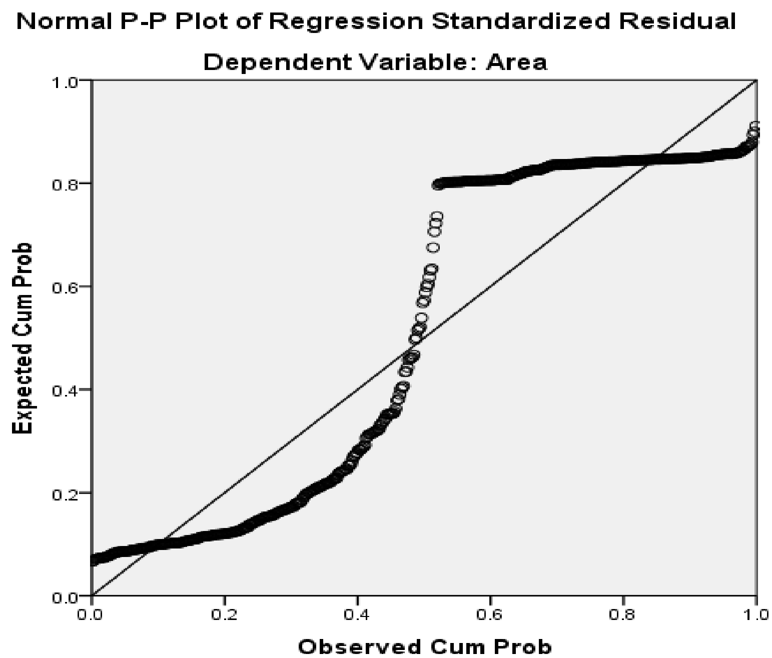
b. Dependent Variable: Area

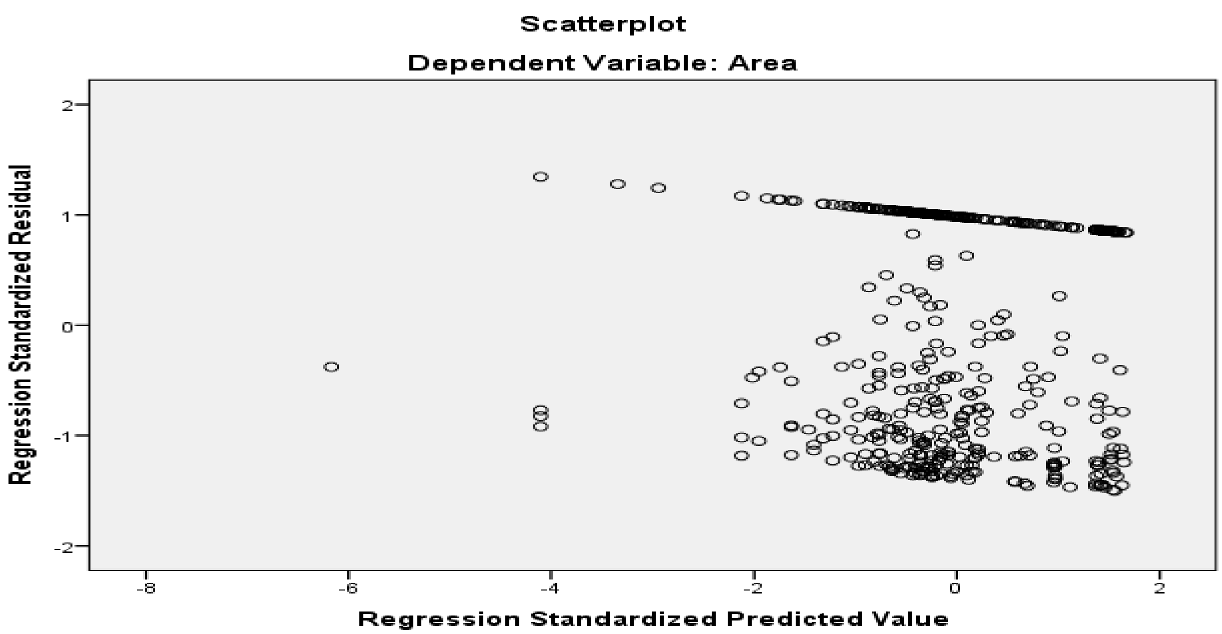
Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound	Correlations Zero-order	Correlations Partial	Correlations Part	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	-5.272E-16	.044		.000	1.000	-.086	.086					
	DMC	-.004	.059	-.004	-.072	.942	-.120	.111	-.060	-.003	-.003	.560	1.784
	DC	.085	.059	.085	1.445	.149	-.030	.200	.088	.064	.063	.560	1.784

a. Dependent Variable: Area




Normal P-P Plot of Regression Standardized Residual
Dependent Variable: Area
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0


Scatterplot
Dependent Variable: Area
Regression Standardized Residual
Regression Standardized Predicted Value
-8
-6
-4
-2
0
2
-1
1