Investigating relation between meteorological factors and forest fire

Forest fire or wild fire is an uncontrolled fire occurring due to nature or human reasons. It occurs in an area of combustible vegetation. A forest fire can go on for days or even weeks and can destroy almost anything that comes in its way, causing damage to the entire vegetation or forest and endangering human lives. Hence it becomes more important to prevent and control forest fires in order to intercept economical and ecological damage.

While human-caused forest fire can be prevented up to a large extent by reducing human interference in the forest fire prone zone and educating human about the causes and effects of forest fire, but it cannot be guaranteed. On the other hand, natural causes responsible for forest fire are impossible to control and prevent. However as Tomas Artes, Ana Cortes, and Tomas Margalef (2016) suggests that a solution to minimize the damage of a natural hazard and to fight against them, is a good prediction of its behavior. Paulo Cortaiz and Anibal Morais (2007) points out that an early detection of forest fire is a key element of triumphant firefighting. If forest fires can be detected in its early stage, then the level of damage caused by it can be reduced. This can be done either by human surveillance or automatic solutions (Cortaiz and Morais, 2007).

Several measures, techniques and models have been used by previous researchers to predict and detect forest fires such as Data Mining techniques to investigate general indicators that could influence forest fire (Zohre Sadat Pourtaghi, Hamid Reza Pourghasemi, Roberta Aretano and Teodoro Semeraro, 2016), Spatial analysis method to examine forest fire risk (Weibin You et al., 2017), several models which makes use of Fire Simulator System and based on genetic algorithms have been developed to predict forest fire propagation (Tomas Artes, Ana Cortes, and Tomas Margalef, 2016), fuzzy logic systems have been proposed by Santiago Garcia-Jimenez et al. (2017).

Meteorological factors are the most influential factors responsible for the fire spread intensity and direction (C. Brun, T. Artes, et al. 2017). These factors can be temperature, wind, rainfall and relative humidity. By analyzing these factors, fatalities can be avoided, as the public can be warned before hand and better fire management decisions can be taken. If the meteorological factors such as wind influences the area affected by the forest fires, then it can be useful to detect how violent the forest fire can be and casualties can be avoided.

A small amount of research (Cortaiz and Morais, 2007) has been dedicated to detecting the spread of forest fires based only on meteorological factors. Hence, in this research report, an analysis of the individual responsibility of these factors influencing forest fires have been tried to detect. The purpose of this research is to investigate if there is any direct influence of one or many meteorological factor with the area burned by forest fire. The meteorological factors considered for the analysis are temperature, wind and relative humidity, as they are known to be the most influential. If the research is successful to detect any kind of relationship then this information can be used by local meteorological stations to predict in advance the chances or intensity of forest fire based on the degree of temperature, wind speed or percentage of relative humidity on that day/ month and make necessary preparations like warning the public, alerting the team of fire fighters to take rapid actions accordingly. This analysis can also be used by researchers to use in neural networks, advanced algorithms or any research related to forest fire.

The hypothesis to be tested for the Forest Fire dataset of the northeast region of Portugal are stated below. Hypothesis 1, 2 and 3 tries to determine the relation of the meteorological factors, wind, temperature and relative humidity(RH) with the month of the year, to find out if these factors changes over the year or not and how much. Hypothesis 4, 5 and 6 investigates if any of these above mentioned factors has direct affect on the area burned.

Hypothesis 1:

H_0 : There is no difference in the wind speed, degrees of temperature and RH over different months of the year.

H_A : There is difference in the wind speed, degrees of temperature and RH over different months of the year.

Hypothesis 2:

H_0 : There is no difference in the degrees of temperature over different months of the year.

H_A : There is difference in the degrees of temperature over different months of the year.

Hypothesis 3:

H_0 : There is no difference in the RH over different months of the year.

H_A : There is difference in the RH over different months of the year.

Hypothesis 4:

H_0 : There is no relationship between area affected and wind speed.

H_A : There is a relationship between area affected and wind speed.

Hypothesis 5:

H_0 : There is no relationship between area affected and temperature.

H_A : There is no relationship between area affected and temperature.

Hypothesis 6:

H_0 : There is no relationship between area affected and relative humidity (RH).

H_A : There is no relationship between area affected and relative humidity (RH).

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 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 selected data variables.

The research analysis will mainly focus on the variables temp, RH, wind, and area. The four meteorological variables will be compared for a relation against the variable area to investigate if these factors are responsible to increase the spread of forest fire. It should be noted that 'rain' can also be a crucial factor to check for the forest fire spread, however the 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, hence do not prove out to be useful for this analysis. Results for descriptive statistics can be seen in Table 2. For the cases where the value for area is zero means that the area burned by forest fire was lower than 1ha/100 = $100m^2$. Zscore was calculated for these variables and the results were saved to the dataset. Standardized scores (value/ standard error) for skewness and kurtosis for all the variables was found to be lying outside the range of -2 and +2. Also, the histograms (Figure 1) clearly shows the skewness of the variables and also Q-Q plots (included in the appendix) confirms that the values are varying from the line and hence the variables should not be treated as normal.

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

		Zscore: area	**Zscore: wind**	**Zscore: temp**	**Zscore: RH**
N	**Valid**	517	517	517	517
	Missing	0	0	0	0
Mean		.0000000	.0000000	.0000000	.0000000
Std. Error of Mean		.04397995	.04397995	.04397995	.04397995
Median		-.1936554	-.0098242	.0707522	-.1402302
Std. Deviation		1.00000000	1.00000000	1.00000000	1.00000000
Skewness		12.847	.571	-.331	.863
Std. Error of Skewness		.107	.107	.107	.107
Kurtosis		194.141	.054	.136	.438
Std. Error of Kurtosis		.214	.214	.214	.214
Minimum		-.20182	-2.01914	-2.87416	-1.79490
Maximum		16.93471	3.00415	2.48179	3.41424

Table 2: Descriptive Statistics for Z-scores of area, wind, temp, rain, RH

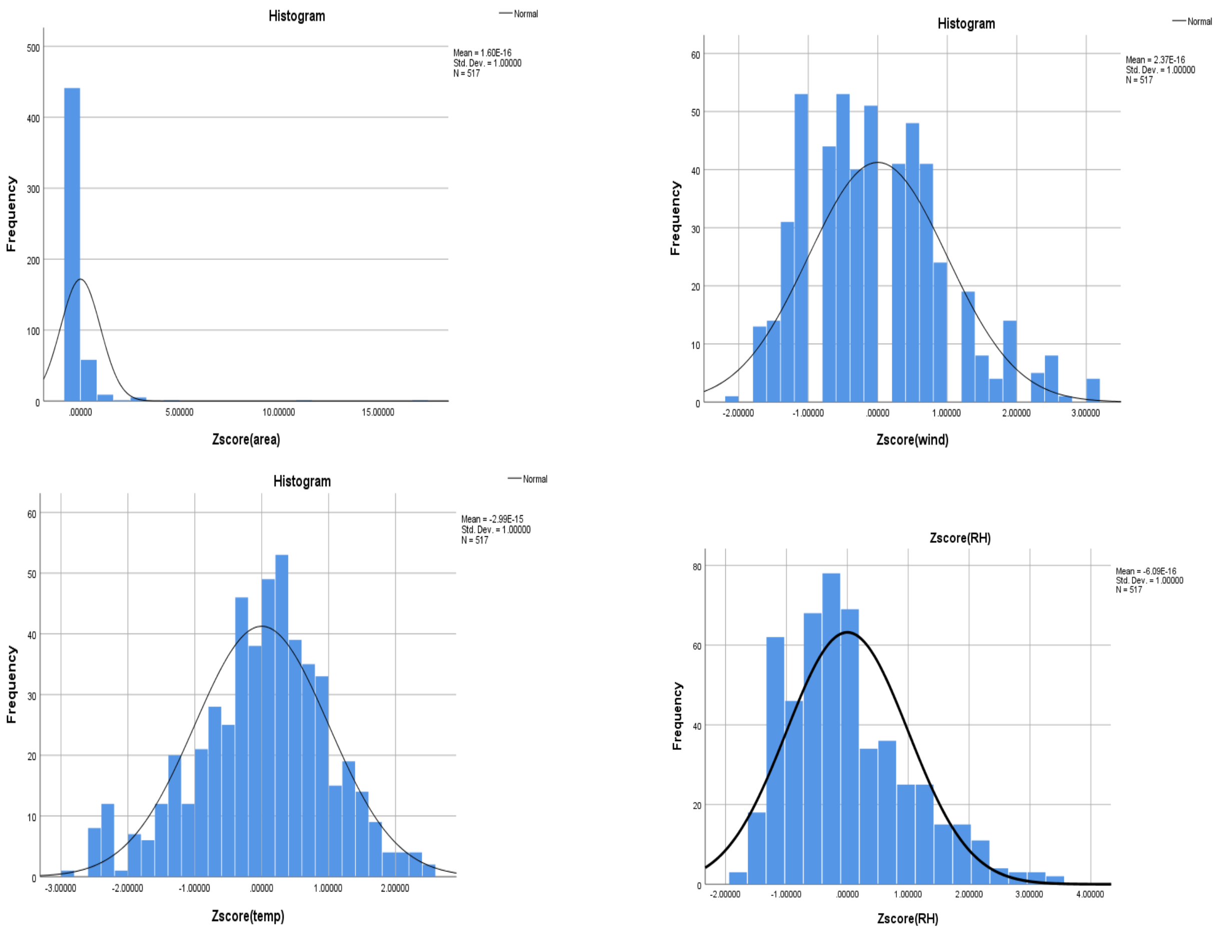


Figure 1: Distribution Plots for Standardized scores of area, wind, temp, RH

Results

After deciding the normality of the data, an attempt was made to transform the data to lessen skewness and revamp symmetry by using Lg10 function and it was found out that the variables area, wind and temp were still skewed (refer to Appendix). However, RH variable was successfully transformed to a normal distribution with standardized skewness of 0.2616 (SE = 0.107) and kurtosis of -1.8785 (SE = 0.214).

Hypothesis 1:

A Kruskal-Wallis H test was conducted to determine if there was any difference in wind speed over twelve different months of the year. As assessed by visual inspection of a box-plot, distribution of wind Z-scores were not similar for all the months. The distributions of Z-scores for wind were statistically significantly different between groups(months) ($H(11) = 55.443$, $p < 0.05$).

Hypothesis 2:

A Kruskal-Wallis H test was conducted to determine if there was any difference in degrees of temperature over twelve different months of the year. As assessed by visual inspection of a box-plot, distribution of temp Z-scores were not similar for all the months. The distributions of Z-scores for temp were statistically significantly different between groups(months) ($H(11) = 208.692$, $p < 0.05$).

Hypothesis 3:

A Kruskal-Wallis H test was conducted to determine if there was any difference in relative humidity percentage over twelve different months of the year. As assessed by visual inspection of a box-plot, distribution of RH_log(transformed values of RH) were not similar for all the months. The distributions of RH_log were statistically significantly different between groups(months) ($H(11) = 31.689$, $p = 0.001$).

Hypothesis 4:

A Spearman's rho test was run to assess the relationship between the Z-scores for wind and area burned by the forest fire. While a very weak positive correlation was observed, it is not statistically significant ($r(515) = 0.053$, $p = 0.227$).

Hypothesis 5:

A Spearman's rho test was run to assess the relationship between the Z-scores for temperature and area burned by the forest fire. While a very weak positive correlation was observed, it is not statistically significant ($r(515) = 0.079$, $p = 0.074$).

Hypothesis 6:

A Spearman's rho test was run to assess the relationship between the transformed variable RH and Z-scores for area burned by the forest fire. While a very very weak negative correlation was observed, it is not statistically significant ($r(515) = - 0.024$, $p = 0.583$).

Please find attached .spv file which contains all details of tests conducted, results and SPSS code.

Discussion

After the analysis presented in the paper, it can be suggested that there is no statistical relationship between the individual meteorological factors chosen (wind, temperature and RH) and the area burned by forest fire. Also, it was confirmed through Kruskal Wallis H test that the variables were not same over the year but varied. Yet this change in temperature, wind speed or relative humidity did not contribute to the spread of forest fire. So, it can be interpreted from the results that none of these factors alone is responsible for the intensity of the forest fire. However, this interpretation paves way to research if two or more of these factors combined influences the forest fire. In further research this possibility can be explored. But the research conducted in this paper may not prove out to be too informative if the area burned by the forest fire very large due to many factors.

As pointed out by Clemens Wastl, Christian Schunk, Michael Leuchner, Gianni B. Pezzatti, Annette Menzel (2012) that currently a key issue in scientific research is climate change. Hence, a limitation with this dataset is that it is a cross-sectional data, collected over a period of only three years. However when detecting forest fires due to meteorological factors, data over a longer time span will prove to be more beneficial to point out the influence of one or more factors, as this will acknowledge the issue of climate change as well.

As forest fire is a disaster not only to the human life but the vegetation and forest covers as well. Every year forest ecosystems are increasingly endangered by fires caused by natural reasons (Zohre Sadat Pourtaghi, Hamid Reza Pourghasemi, Roberta Aretano and Teodoro Semeraro, 2016). Hence it becomes more important to conduct researches which focuses primarily on the natural factors and causes leading to 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

Zohre Sadat Pourtaghi, Hamid Reza Pourghasemi, Roberta Aretano and Teodoro Semeraro (2016). Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques. *Ecological Indicators*, *64*, 72-84. https://doi.org/10.1016/j.ecolind.2015.12.030

Weibin You et al. (2017). Geographical information system-based forest fire risk assessment integrating national forest inventory data and analysis of its spatiotemporal variability. *Ecological Indicators, 77*, 176-184. https://doi.org/10.1016/j.ecolind.2017.01.042

Santiago Garcia-Jimenez, Aranzazu Jurio, et al. (2017). Forest fire detection: A fuzzy system approach based on overlap indices. *Applied Soft Computing*, *52*, 834-842. https://doi.org/10.1016/j.asoc.2016.09.041

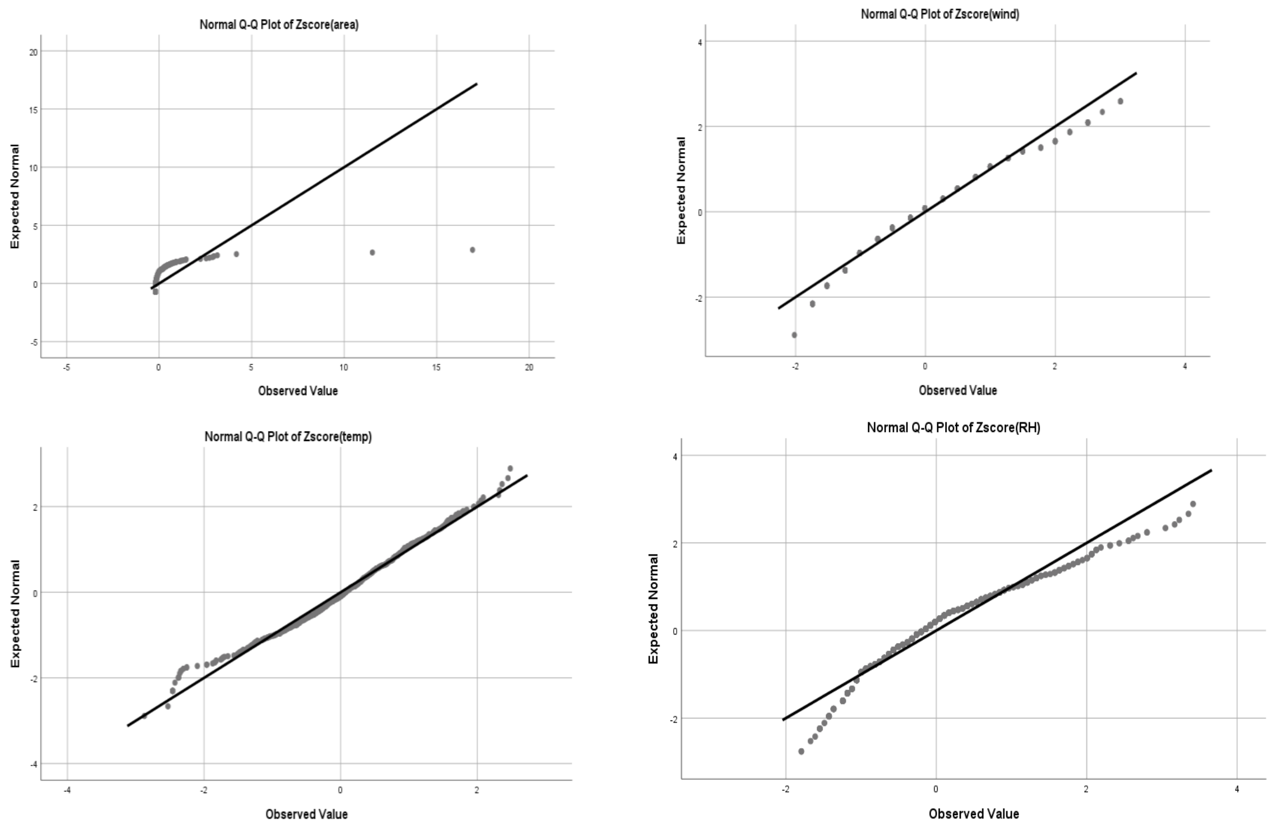
C. Brun, T. Artes, et al. (2017). A High Performance Computing Framework for Continental-Scale Forest Fire Spread Prediction. *Procedia Computer Science*, *108,* 1712-1721. https://doi.org/10.1016/j.procs.2017.05.258

Clemens Wastl, Christian Schunk, Michael Leuchner, Gianni B. Pezzatti, Annette Menzel (2012). Recent climate change: Long-term trends in meteorological forest fire danger in the Alps. *Agricultural and Forest Meteorology*, *162-163*, 1-13. https://doi.org/10.1016/j.agrformet.2012.04.001

Appendix

Descriptive Statistics to check normality

```
EXAMINE VARIABLES=Zarea Zwind Ztemp ZRH
  /PLOT HISTOGRAM NPPLOT
  /COMPARE GROUPS
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.
```



Q-Q Plots for the variables of interest

Transforming Data and examining it to check for normality

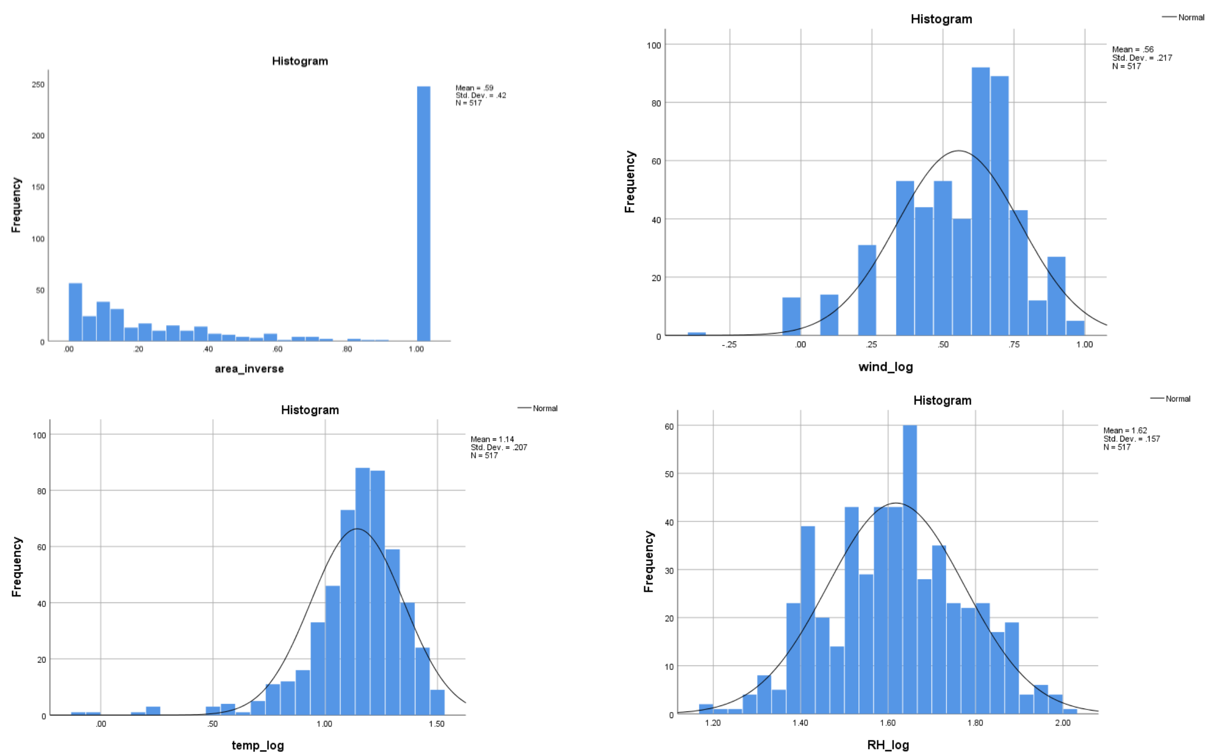
```
COMPUTE area_inverse=1/(area + 1).
EXECUTE.
EXAMINE VARIABLES=area_inverse
  /PLOT HISTOGRAM
  /COMPARE GROUPS
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

COMPUTE wind_log=LG10(wind).
EXECUTE.
```

```
EXAMINE VARIABLES=wind_log
  /PLOT HISTOGRAM
  /COMPARE GROUPS
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

SORT CASES BY temp (D).
COMPUTE temp_log=LG10((33.1+1) - temp).
EXECUTE.
EXAMINE VARIABLES=temp_log
  /PLOT HISTOGRAM
  /COMPARE GROUPS
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

COMPUTE RH_log=LG10(RH).
EXECUTE.
DATASET ACTIVATE DataSet1.
EXAMINE VARIABLES=RH_log
  /PLOT HISTOGRAM
  /COMPARE GROUPS
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.
```



Distribution Plots for the transformed scores of area, wind, temp, RH

Correlations

```
NONPAR CORR
  /VARIABLES=Zwind Zarea
  /PRINT=SPEARMAN TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Correlations

			Zscore(wind)	Zscore(area)
Spearman's rho	Zscore(wind)	Correlation Coefficient	1.000	.053
		Sig. (2-tailed)	.	.227
		N	517	517
	Zscore(area)	Correlation Coefficient	.053	1.000
		Sig. (2-tailed)	.227	.
		N	517	517

```
NONPAR CORR
  /VARIABLES=Ztemp Zarea
  /PRINT=SPEARMAN TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Correlations

			Zscore(temp)	Zscore(area)
Spearman's rho	Zscore(temp)	Correlation Coefficient	1.000	.079
		Sig. (2-tailed)	.	.074
		N	517	517
	Zscore(area)	Correlation Coefficient	.079	1.000
		Sig. (2-tailed)	.074	.
		N	517	517

```
NONPAR CORR
  /VARIABLES=RH_log Zarea
  /PRINT=SPEARMAN TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Correlations

			RH_log	Zscore(area)
Spearman's rho	RH_log	Correlation Coefficient	1.000	-.024
		Sig. (2-tailed)	.	.583
		N	517	517
	Zscore(area)	Correlation Coefficient	-.024	1.000
		Sig. (2-tailed)	.583	.
		N	517	517

Differences

```
*Nonparametric Tests: Independent Samples.
```

```
NPTESTS
  /INDEPENDENT TEST (Zwind Ztemp RH_log) GROUP (month)
  /MISSING SCOPE=ANALYSIS USERMISSING=EXCLUDE
  /CRITERIA ALPHA=0.05  CILEVEL=95.
```

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Zscore(wind) is the same across categories of month.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.
2	The distribution of Zscore(temp) is the same across categories of month.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.
3	The distribution of RH_log is the same across categories of month.	Independent-Samples Kruskal-Wallis Test	.001	Reject the null hypothesis.

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