Sun Coast Remediation Research Project

Reginald Mapp

Columbia Southern Unversity

**Data Analysis: Descriptive Statistics and Assumption Testing**

Sun Coast provides relief solutions to marketing and government organizations, where they clean harmful substances from water and soils. Sun coast generally cares about health safety, wellbeing, worker remuneration costs, and also likely enduring legal actions from any injury occurring or any sickness associated with work duty.

It is a quantitative study; it connects research businesses and their uses in this real world. Sun coast statistical tools remediations are;

Table of contents- This is a tool used in creating tables.

Executive Summary- The executive will come here, and the executive should be composed after the project is complete.

Project completion includes the following terms used below.

Safety training effectiveness- It is usually done for each contract agreement given to the sun-coast.

Sound level exposure- It is generally noisy since there is a couple of heavy machinery being used in the clients' operations and remediation also.

New employee training- All employees participate in new safety and health training, upon which the employee is tested on their knowledge.

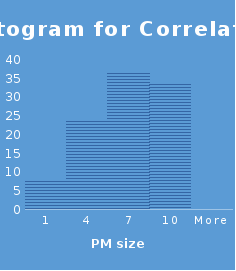
Return on investment- It offers four lines of service to customers like monitoring of air, remediation of soil, reclamation of water, and training about health and safety.

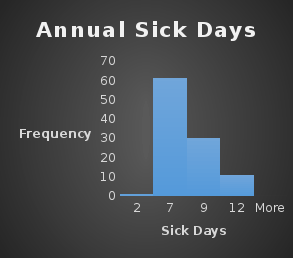
**Correlation: Descriptive Statistics and Assumption Testing**

**Frequency distribution table.**

|  |  |
| --- | --- |
| micro-range | Frequency |
| 0-1 | 8 |
| 2-4 | 24 |
| 5-7 | 37 |
| 8-10 | 34 |
|  |  |
|  |  |
| sick-days employee | Frequency |
| 0 to 2 | 1 |
| 4 to 7 | 61 |
| 8 to 9 | 30 |
| 10 to 12 | 11 |

**Histogram**

****

****

**Descriptive statistics table**

**Descriptive Statistics Table**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| microns |  |  | sick day |  |
|  |  |  |  |  |
| Mean | 5.65728155 |  | Mean | 7.126214 |
| Standard Error | 0.25560014 |  | Standard Error | 0.186484 |
| Median |  |  | Median |  |
| Mode |  |  | Mode |  |
| Standard Deviation | 2.59405814 |  | Standard Deviation | 1.892605 |
| Sample Variance | 6.72913764 |  | Sample Variance | 3.581953 |
| Kurtosis | -0.8521619 |  | Kurtosis | 0.124923 |
| Skewness | -0.37325713 |  | Skewness | 0.14225 |
| Range | 9.8 |  | Range | 10 |
| Minimum | 0.2 |  | Minimum |  |
| Maximum | 10 |  | Maximum | 12 |
| Sum | 582.7 |  | Sum | 734 |
| Count | 103 |  | Count | 103 |
| Largest(1) | 10 |  | Largest(1) | 12 |
| Smallest(1) | 0.2 |  | Smallest(1) |  |
| Confidence Level (95.0%) | 0.50698167 |  | Confidence Level (95.0%) | 0.36989 |

**Measurement scale**

Ordinal

Data are ranked and ordered without actually establishing the degree of variation between them.

**The measure of central tendency.**

Mean

**Evaluation.**

The above statistics have indifferences as the test static of the sample data to that of the average population were different.

The assumptions in the parametric testing were not met as there were indifferences in the results under a 95 percent confidence interval. First, there were differences in the data, which led to differences in the measures of central tendency. For instance, the mean of the data for microns and sick days as projected by that of 5.65 and that of 7.12, respectively. Despite having similar counts that were also a difference that arose between the highest and lowest number in the data provided. Additionally, parameters in the test static for the two populations gave contrastive results. Thus, the assumptions in the parametric testing remained unmet.

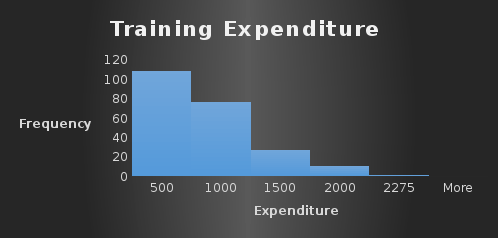
**Simple Regression: Descriptive Statistics and Assumption Testing**

**Frequency distribution table**

|  |  |
| --- | --- |
| Expenditure | Frequency |
| 20-500 | 108 |
| 501-1000 | 76 |
| 1001-1500 | 27 |
| 1501-2000 | 11 |
| 2001-2500 | 1 |

|  |  |
| --- | --- |
| Time | Frequency |
| 0-50 | 6 |
| 51-100 | 26 |
| 101-200 | 98 |
| 201-300 | 85 |
| 301-400 | 8 |

**Histogram**





**Descriptive statistics table**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| safety training expenditure |  |  | lost time hours |  |
|  |  |  |  |  |
| Mean | 595.9843812 |  | Mean | 188.0045 |
| Standard Error | 31.4770075 |  | Standard Error | 4.803089 |
| Median | 507.772 |  | Median | 190 |
| Mode | 234 |  | Mode | 190 |
| Standard Deviation | 470.0519613 |  | Standard Deviation | 71.72542 |
| Sample Variance | 220948.8463 |  | Sample Variance | 5144.536 |
| Kurtosis | 0.444080195 |  | Kurtosis | -0.50122 |
| Skewness | 0.951331922 |  | Skewness | -0.08198 |
| Range | 2251.404 |  | Range | 350 |
| Minimum | 20.456 |  | Minimum | 10 |
| Maximum | 2271.86 |  | Maximum | 360 |
| Sum | 132904.517 |  | Sum | 41925 |
| Count | 223 |  | Count | 223 |
| Largest(1) | 2271.86 |  | Largest(1) | 360 |
| Smallest(1) | 20.456 |  | Smallest(1) | 10 |
| Confidence Level (95.0%) | 62.03197147 |  | Confidence Level (95.0%) | 9.465484 |

**Measurement scale**

Nominal

There is a measurement of scale in which the numbers serve as labels only to classify the object.

**The measure of central tendency.**

Median

**Evaluation.**

The p-value for both training expenditures and the lost time hours is exceedingly high.

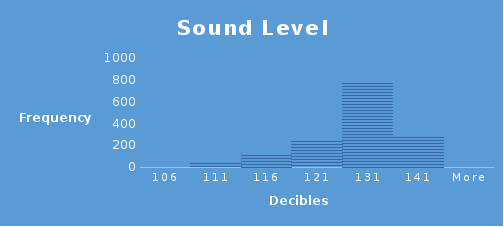
The assumptions for parametric testing in the study prove to be met, as expressed by the test statistic. First, there is a huge difference that emerges in the data between the training expenditure and the lost time hours. Findings from the statistical test indicate that the p-value in both the training expenditure and lost time hours is exceeding high. However, an analysis of the data indicates that there lost time hours has a smaller confidence interval as opposed to that of training expenditure. Thus, the statistical tests prove the assumptions as there is a great difference that emerges in the two data sets.

**Multiple Regressions: Descriptive Statistics and Assumption Testing**

**Frequency distribution table**

|  |  |
| --- | --- |
| Decibel | Frequency |
| 100-106 |  |
| 107-111 | 51 |
| 112-116 | 126 |
| 117-121 | 249 |
| 122-131 | 786 |
| 132-141 | 287 |

**Histogram.**



**Descriptive statistics table.**

|  |  |
| --- | --- |
| Decibel |  |
|  |  |
| Mean | 124.8359 |
| Standard Error | 0.177945 |
| Median | 125.721 |
| Mode | 127.315 |
| Standard Deviation | 6.898657 |
| Sample Variance | 47.59146 |
| Kurtosis | -0.31419 |
| Skewness | -0.41895 |
| Range | 37.607 |
| Minimum | 103.38 |
| Maximum | 140.987 |
| Sum | 187628.4 |
| Count | 1503 |

**Measurement scale.**

Interval

This is because the difference between the two variables id meaningful and equal, and the presence of 0 is arbitrary.

**The measure of central tendency.**

Mean

**Evaluation.**

There is no direct relation between the variables.

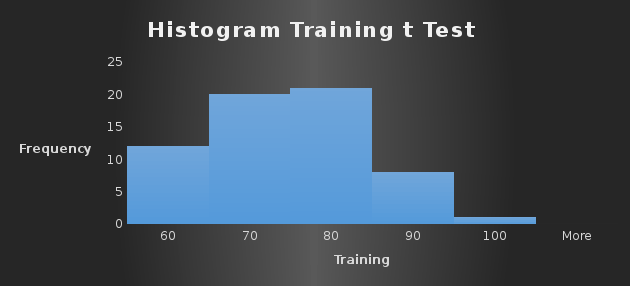
The assumptions for parametric testing were unmet, as it is evident that it is no relationship between the variables. In such circumstances, there is a null hypothesis for each variable, an indication that the variables do not fit in the multiple regression equation. Since the variable does not have any relations, there remains a standard error in the data. Since the null hypothesis was untrue, there is less probability of obtaining a test statistic based on the data provided. This is because there are two variables at the expense of three. This makes the parametric assumptions remain unmet as there is no clear relationship.

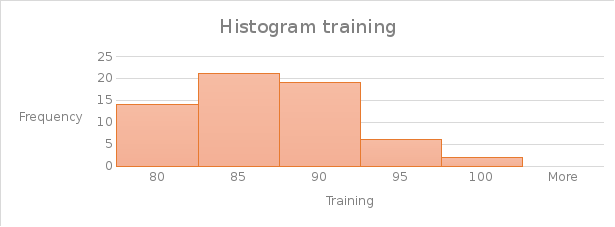
**Independent Samples *t*-Test: Descriptive Statistics and Assumption Testing**

**Frequency distribution table**

|  |  |  |
| --- | --- | --- |
| Group A Training | Frequency | |
| 49-60 | 12 | |
| 61-70 | 20 | |
| 71-80 | 21 | |
| 81-90 | 8 | |
| 91-100 | 1 | |
|  | |
| Group B Training | | Frequency | |
| 74-80 | | 14 | |
| 81-85 | | 21 | |
| 86-90 | | 19 | |
| 91-95 | | 5 | |
| 96-100 | | 3 | |

**Histogram**



****

**Descriptive statistics table**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Prior Training |  |  | Revised Training |  |
|  |  |  |  |  |
| Mean | 69.79032 |  | Mean | 84.77419 |
| Standard Error | 1.402788 |  | Standard Error | 0.659479 |
| Median | 70 |  | Median | 85 |
| Mode | 80 |  | Mode | 85 |
| Standard Deviation | 11.04556 |  | Standard Deviation | 5.192742 |
| Sample Variance | 122.0045 |  | Sample Variance | 26.96457 |
| Kurtosis | -0.77668 |  | Kurtosis | -0.35254 |
| Skewness | -0.0868 |  | Skewness | 0.144085 |
| Range | 41 |  | Range | 22 |
| Minimum | 50 |  | Minimum | 75 |
| Maximum | 91 |  | Maximum | 97 |
| Sum | 4327 |  | Sum | 5256 |
| Count | 62 |  | Count | 62 |
| Largest(1) | 91 |  | Largest(1) | 97 |
| Smallest(1) | 50 |  | Smallest(1) | 75 |
| Confidence Level (95.0%) | 2.805048 |  | Confidence Level (95.0%) | 1.31871 |

**Measurement scale.**

Interval

This is because the difference between the two variables id meaningful and equal, and the presence of 0 is arbitrary.

**The measure of central tendency.**

Mean

**Evaluation.**

There is an indirect relationship between the sample data and the normal population.

The assumptions were met. Statically test indicates that the probability test is lower than the p-value. For instance, in the first data, the p-value is 2.8, whereas the second data has a p-value of 1.31. The p value is greater than 0. This indicates that there is an indirect relationship between the data, as evidenced by the p-value. The dependent variables were normally distributed. Additionally, there are two groups that are independent of each other such as the test scores for the revised training and that of prior training. Therefore, there is an indirect relationship between the data provided.

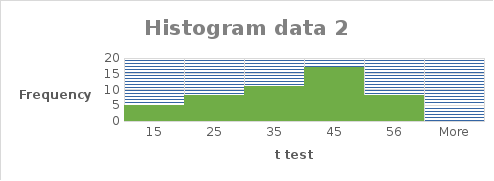
**Dependent Samples (Paired-Samples) *t*-Test: Descriptive Statistics and Assumption Testing**

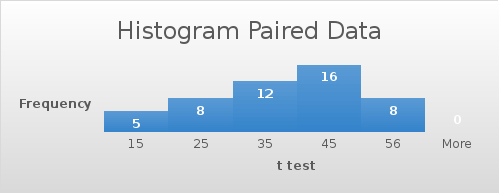
**Frequency distribution table**

|  |  |
| --- | --- |
| Exposure | Frequency |
| 5-15 | 5 |
| 16-25 | 8 |
| 26-35 | 12 |
| 36-45 | 16 |
| 46-56 | 8 |

|  |  |
| --- | --- |
| Exposure | Frequency |
| 5-15 | 5 |
| 16-25 | 8 |
| 26-35 | 11 |
| 36-45 | 17 |
| 46-56 | 8 |

**Histogram.**



****

**Descriptive statistics table**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pre-Exposure μg/dL |  |  | Post-Exposure μg/dL |  |
|  |  |  |  |  |
| Mean | 32.8571429 |  | Mean | 33.28571 |
| Standard Error | 1.75230655 |  | Standard Error | 1.781423 |
| Median | 35 |  | Median | 36 |
| Mode | 36 |  | Mode | 38 |
| Standard Deviation | 12.2661458 |  | Standard Deviation | 12.46996 |
| Sample Variance | 150.458333 |  | Sample Variance | 155.5 |
| Kurtosis | -0.57603713 |  | Kurtosis | -0.65421 |
| Skewness | -0.42510965 |  | Skewness | -0.48363 |
| Range | 50 |  | Range | 50 |
| Minimum |  |  | Minimum |  |
| Maximum | 56 |  | Maximum | 56 |
| Sum | 1610 |  | Sum | 1631 |
| Count | 49 |  | Count | 49 |
| Largest(1) | 56 |  | Largest(1) | 56 |
| Smallest(1) |  |  | Smallest(1) |  |
| Confidence Level (95.0%) | 3.52324845 |  | Confidence Level (95.0%) | 3.581792 |

**Measurement scale.**

Interval.

This is because the difference between the two variables id meaningful and equal, and the presence of 0 is arbitrary.

**The measure of central tendency.**

Mean

**Evaluation.**

The null hypothesis is accepted as the null hypothesis is greater than 0.

The assumptions for parametric testing were met. This is because the data consisted of a dependent variable that was continuous on a ratio basis. Additionally, the observations of the data collected were independent of one another. This is irrespective of the fact that dependent variables were normally distributed. A comparison of the two means indicates that there is a statistical difference between the mean. In the data present, the difference between the mean is 0.4285671. An evaluation of the statistical test indicates that the t-test is greater than the calculated test. We see that the differences between the observed t-test and the calculated t-test lead to the acceptance of the null hypothesis.

**ANOVA: Descriptive Statistics and Assumption Testing**

**Frequency distribution table.**

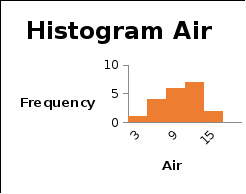
|  |  |
| --- | --- |
| Air | Frequency |
| 1-3 | 1 |
| 4-6 | 4 |
| 7-9 | 6 |
| 10-12 | 7 |
| 13-15 | 2 |

|  |  |
| --- | --- |
| Soil | Frequency |
| 5-7 | 3 |
| 8-10 | 13 |
| 11-13 | 4 |

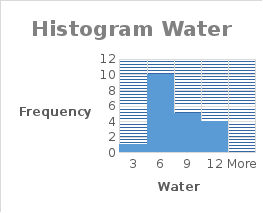
|  |  |
| --- | --- |
| Water | Frequency |
| 1-3 | 1 |
| 4-6 | 10 |
| 7-9 | 5 |
| 10-12 | 4 |

|  |  |
| --- | --- |
| Training | Frequency |
| 1-3 | 1 |
| 4-6 | 16 |
| 7-9 | 3 |

**Histogram.**

****





**Descriptive statistics table**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A = Air |  |  | B = Soil |  |
|  |  |  |  |  |
| Mean | 8.9 |  | Mean | 9.1 |
| Standard Error | 0.684028 |  | Standard Error | 0.390007 |
| Median |  |  | Median |  |
| Mode | 11 |  | Mode |  |
| Standard Deviation | 3.059068 |  | Standard Deviation | 1.744163 |
| Sample Variance | 9.357895 |  | Sample Variance | 3.042105 |
| Kurtosis | -0.6283 |  | Kurtosis | 0.11923 |
| Skewness | -0.36085 |  | Skewness | 0.492002 |
| Range | 11 |  | Range |  |
| Minimum |  |  | Minimum |  |
| Maximum | 14 |  | Maximum | 13 |
| Sum | 178 |  | Sum | 182 |
| Count | 20 |  | Count | 20 |
| Largest(1) | 14 |  | Largest(1) | 13 |
| Smallest(1) |  |  | Smallest(1) |  |
| Confidence Level(95.0%) | 1.431688 |  | Confidence Level(95.0%) | 0.816294 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| C = Water |  |  | D = Training |  |
|  |  |  |  |  |
| Mean |  |  | Mean | 5.4 |
| Standard Error | 0.575829 |  | Standard Error | 0.265568 |
| Median |  |  | Median |  |
| Mode |  |  | Mode |  |
| Standard Deviation | 2.575185 |  | Standard Deviation | 1.187656 |
| Sample Variance | 6.631579 |  | Sample Variance | 1.410526 |
| Kurtosis | -0.23752 |  | Kurtosis | 0.253747 |
| Skewness | 0.760206 |  | Skewness | 0.159183 |
| Range |  |  | Range |  |
| Minimum |  |  | Minimum |  |
| Maximum | 12 |  | Maximum |  |
| Sum | 140 |  | Sum | 108 |
| Count | 20 |  | Count | 20 |
| Largest(1) | 12 |  | Largest(1) |  |
| Smallest(1) |  |  | Smallest(1) |  |
| Confidence Level (95.0%) | 1.205224 |  | Confidence Level (95.0%) | 0.55584 |

**Measurement scale**

**Ratio**

The ration is quantitative in nature and allows comparison of intervals and differences.

**The measure of central tendency.**

Mean.

Mean is used to measure the central tendency

**Evaluation.**

The means are not equal as they base on different sets of data.

Based on the data provided, the assumptions that can be derived are those for normality, equal variance, and that of independent errors. From the data, there is an interaction of the variables with no restrictions. The parametric assumptions in this scenario would relate to the parameters on the population distribution upon which data is drawn. Additionally, a non-parametric test would refer to that which makes no such assumptions. This leads to a normal distribution, homogeneity of the variances, multiple groups which relate to the same variance as well as linearity on the independent relationships. Thus, the assumptions define the type of variance.

References

Judd, C. M., McClelland, G. H., & Ryan, C. S. (2017). *Data analysis: A model comparison approach to regression, ANOVA, and beyond*. Routledge.

Blaikie, N. (2003). *Analyzing quantitative data: From description to explanation*. Sage.