Correlation and Regression Analysis

Reginald Mapp

Columbia Southern University

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

The Sun Coast Remediation data represents a mean of 5.657 for microns, standard deviation of 0.2556, median and mode of 6 and 8 respectively and count of 103. The data fulfils the assumptions for parametric statistical procedures including skewness and kurtosis, and assumptions of normality (McCluskey, Mb, Frca, Ghaaliq, & Mb, 2007). Normality means the distribution is bell shaped.

**Correlation: Descriptive Statistics and Assumption Testing**

**Frequency distribution table.**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  | *1* | *4* | *11* |
| 1 | 1 |  |  |
| 4 | -0.11041 | 1 |  |
| 11 | 0.094073 | -0.71954 | 1 |

**Measure of central tendency.** R2 for the data is 0.0096

**Evaluation.** Correlation analysis demonstrates that there is no correlation between the variables considering a correlation nearing 0(in essence 0.094).

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

**Frequency distribution table.**

|  |  |
| --- | --- |
|  |  |
| *Regression Statistics* |
| Multiple R | 0.937994 |
| R Square | 0.879833 |
| Adjusted R Square | 0.879286 |
| Standard Error | 160.4165 |
| Observations | 222 |

**Descriptive statistics table.**

|  |
| --- |
| *1985.119* |
|  |  |
| Mean | 589.727 |
| Standard Error | 30.98806 |
| Median | 507.705 |
| Mode | 234 |
| Standard Deviation | 461.7116 |
| Sample Variance | 213177.6 |
| Kurtosis | 0.388974 |
| Skewness | 0.925083 |
| Range | 2251.404 |
| Minimum | 20.456 |
| Maximum | 2271.86 |
| Sum | 130919.4 |
| Count | 222 |

**Evaluation.**

From the analysis, the correlation coefficient is *multiple R* which is 0.938 indicating the strength of the linear relationship between the variables (McCluskey, Mb, Frca, Ghaaliq, & Mb, 2007). A correlation coefficient of 0.938 means that there is a strong positive relationship between safety training expenditure and lost time hours.

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

**Frequency distribution table**

|  |  |
| --- | --- |
|  |  |
| *Regression Statistics* |
| Multiple R | 0.063026 |
| R Square | 0.003972 |
| Adjusted R Square | -2.5E-05 |
| Standard Error | 568.7967 |
| Observations | 1502 |

**Descriptive statistics table.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |
| Regression | 6 | 1928953 | 321492.1 | 0.993702 | 0.427877 |  |
| Residual | 1495 | 4.84E+08 | 323529.7 |  |  |  |
| Total | 1501 | 4.86E+08 |   |   |   |  |

**Evaluation.** From the analysis, multiple R is correlation coefficient of 0. 063 meaning there is a weak linear relationship between the independent variable and the dependent variable (McCluskey, Mb, Frca, Ghaaliq, & Mb, 2007). The data fulfils the assumptions of parametric statistical testing.

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

**Frequency distribution table.**

|  |
| --- |
| t-Test: Paired Two Sample for Means |
|  |  |  |
|  | *Variable 1* | *Variable 2* |
| Mean | 32.85714 | 33.28571 |
| Variance | 150.4583 | 155.5 |
| Observations | 49 | 49 |
| Pearson Correlation | 0.992236 |  |
| Hypothesized Mean Difference | 0 |  |
| Df | 48 |  |
| t Stat | -1.9298 |  |
| P(T<=t) one-tail | 0.029776 |  |
| t Critical one-tail | 1.677224 |  |
| P(T<=t) two-tail | 0.059553 |  |
| t Critical two-tail | 2.010635 |   |

**Evaluation.** The output from the analysis indicates that the mean for variable 1 and variable 2 is 32.857 and 33.286 respectively. In the variance row, the data shows that the variance values are not equal but they are close enough hence the assumption of equal variances can be made. The p-value of the statistics is a significant part of t-tests. If the p-value<the significance level, that means the means are statistically significant. From the data, the p–value<significance level of 0.95 which means the means are statistically significant (McCluskey, Mb, Frca, Ghaaliq, & Mb, 2007). The data fulfils the assumption of parametric statistical testing.

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

**Frequency distribution table.**

|  |
| --- |
| t-Test: Two-Sample Assuming Unequal Variances |
|  |  |  |
|  | *Variable 1* | *Variable 2* |
| Mean | 32.85714 | 33.28571 |
| Variance | 150.4583 | 155.5 |
| Observations | 49 | 49 |
| Hypothesized Mean Difference | 0 |  |
| Df | 96 |  |
| t Stat | -0.17151 |  |
| P(T<=t) one-tail | 0.432092 |  |
| t Critical one-tail | 1.660881 |  |
| P(T<=t) two-tail | 0.864184 |  |
| t Critical two-tail | 1.984984 |   |

**Evaluation**

Statistical significance is based on the p-value. Since p-value refers to probability of test results based on null hypothesis, a low value means that there is a low probability of obtaining values observed in case the null hypothesis is true (McCluskey, Mb, Frca, Ghaaliq, & Mb, 2007). From the p-value of 0.432, the null hypothesis is rejected.

**ANOVA: Descriptive Statistics and Assumption Testing**

**Frequency distribution table.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ANOVA: Single Factor |  |  |  |  |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Column 1 | 20 | 178 | 8.9 | 9.357895 |  |  |
| Column 2 | 20 | 182 | 9.1 | 3.042105 |  |  |
| Column 3 | 20 | 140 | 7 | 6.631579 |  |  |
| Column 4 | 20 | 108 | 5.4 | 1.410526 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 182.8 | 3 | 60.93333 | **11.9231** | 0.006 | **2.724944** |
| Within Groups | 388.4 | 76 | 5.110526 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 571.2 | 79 |   |   |   |   |
| ANOVA |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |
| Regression | 6 | 1928953 | 321492.1 | 0.993702 | 0.427877 |  |
| Residual | 1495 | 4.84E+08 | 323529.7 |  |  |  |
| Total | 1501 | 4.86E+08 |   |   |   |  |

**Evaluation.**

The data analysis demonstrates that the F value is greater compared to F-critical value at 0.05 alpha level. There is evidence to reject the null hypothesis and conclude that the samples have significantly different mean. Further, the p-value< alpha level hence we reject the Null Hypothesis (McCluskey, Mb, Frca, Ghaaliq, & Mb, 2007).

References

McCluskey, A., Mb, B., Frca, C., Ghaaliq, A., & Mb, L. (2007). Statistics IV: Interpreting the results of statistical tests. In *Continuing Education in Anaesthesia Critical Care and Pain*.