**ONE SAMPLE TEST**

Table 1.0, Summary statistics for tannin concentration on 30 trees (parametric)

|  |  |
| --- | --- |
| **Standard Error Mean** | 1.31430092 |
| **Variance** | 51.8216069 |
| **Sample size** | 30 |

|  |  |  |
| --- | --- | --- |
| **Test** | **Statistic** | **P value** |
| **Student’s t** | -6.500562 | Pr> |t| <.0001 |
| **Sign M** | -11 | Pr>= |M| <.0001 |
| **Sign Rank S** | -207.5 | Pr>= |S| <.0001 |

Table 1.1, Test for Location

|  |  |  |
| --- | --- | --- |
| **Test** | **Statistic** | **p Value** |
| **Shapiro-Wilk W** | 0.973597 | Pr<W 0.6415 |
| **Kurtosis** | -0.8829405 | **-** |
| **skewness** | -0.1257882 | **-** |

Table 1.2, Test for Normality

|  |  |
| --- | --- |
| **Range** | 26.92000 |
| **Interquartile Range** | 10.50600 |
| **median** | -9.27300 |
| **25th & 75th percentile** | -13.6850 & -3.1790 |

Table 1.3, Summary statistics for tannin concentration on 30 trees (non-parametric)

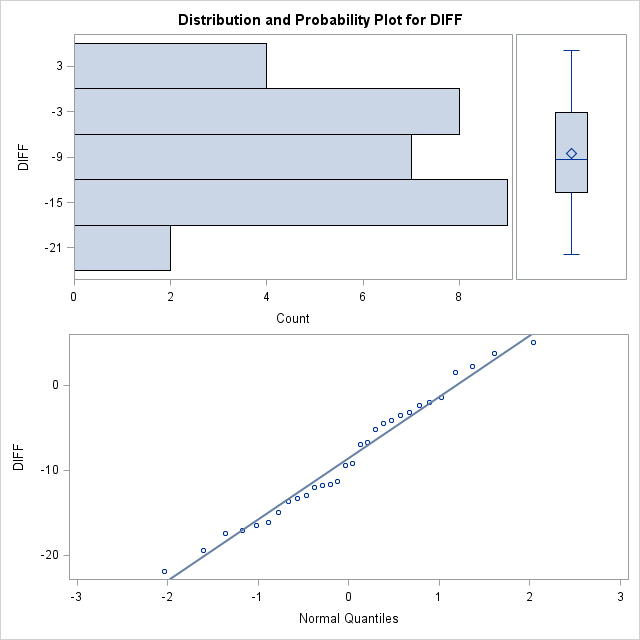


Figure 1.0, Graphical representation of normality

**ASSUMPTIONS/CONCLUSIONS**

**Table 1.0,** based on the above output from summary statistics, standard error of the mean gave rough estimate of the intervals in the population mean is likely to fall. This shows that the larger the sample size, the smaller the sample standard error. Like the standard deviation, 1.96 to obtain an estimate of where 95% pf the population sample means are expected to fall in the theoretical sampling distribution can multiply it. Additionally, variance of 51.8216069 indicates that the data somehow spread out from the mean this because variance measures how far set of data spread, zero variance indicate that all data values are identical and non-zero variances are positive.

**Table 1.1 and figure 1.0,** from the results generate above by these representations there are several assumptions that should be made and met. The normality in our particular case indicates that p value is not close to zero that is it doesn’t come close to having a bell-shaped distribution with mean 0 and standard deviation 1. Additionally, measures like skewness and kurtosis was derived to test those assumptions with skewness and kurtosis must fall between ±2 and ±7 respectively. In addition, wilk’s test above did not have any significance to meet the assumption of normality

**Table 1.3,** referring to this table variability in the output was measured by these non-parametric measures. Focusing on Signed Rank S test with p value less than 0.0001, which is a bit close to normality. Conclusion derived from parametric test does not show much difference from the non-parametric test. Upper (75th) and lower (25th) interquartile range values showed there is statistical dispersion in the data set.

**Method/Results,** the one sample t-test is a statistical procedure used to determine whether a process with a specific mean could have generated a sample of observations. The null hypothesis and the alternative hypothesis assumes that some difference exists between the true mean and the comparison value whereas the null hypothesis assumes that no difference exists. The purpose of this test is to determine if the null hypothesis should be rejected given the 30 trees before and after the herbivore attack. For the parametric procedure, the one sample test makes several assumptions. Although t-test are sometimes robust, it always helps to evaluate the degree of deviation from assumptions like the normality, independence and level of measurement.

**TWO SAMPLE TEST**

The primary t-test output below from **Table 2.0** shows sample statistics and confidence intervals for each group as well as for the difference. The confidence intervals in the results below for the food type do not overlap, meaning that the food type means are significantly different.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **food** | **N** | **Mean** | **Std Dev** | **Std Err** | **95% CL** | **95% CL** |
| **D** | 17 | 0.8881 | 0.1486 | 0.0360 | 0.8117 | 0.1106 |
| **L** | 10 | 1.3544 | 0.4615 | 0.1459 | 1.0242 | 0.3175 |
| **Diff (1-2)** |  | -0.4663 | 0.3013 | 0.1201 |  |  |

Table 2.0, summary statistics for two sample test (Liver powder vs powered dog food)

The sample is not large enough to even apply central limit theorem and assume normally distributed sample means. However, if this were not the case, the test for normality using distribution would not be feasible. The next assumptions to examine is the homogeneity with the t-test analysis. Looking at the **Table 2.1** below, I f the test is significant, it showed the evidence that the group variance is different (in other words, the equality of variances assumption has not been met). In this situation, the p value is low, so you can conclude that the assumption of equal group variance is not reasonably appropriate.

Table 2.1, Equality of Variances

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Method** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **Folded F** | 9 | 16 | 9.65 | 0.0001 |

Furthermore, the two-test report; the difference between them is in the degrees of freedom. The first (pooled) is interpreted when the assumption of equal group variances has been met. The second (Satterthwaite) is interpreted when the assumption of equal group variance has not been met also. From **Table 2.1,** the p value associated with the t test for equal variances is less than the alpha of 0.05; so, concluded by rejecting the null hypothesis.

Table 2.2, T test for Equal and Unequal Variances

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Method** | **Variances** | **DF** | **t Value** | **Pr > |t|** |
| **Pooled** | Equal | 25 | -3.88 | 0.0007 |
| **Satterthwaite** | Unequal | 10.109 | -3.10 | 0.0111 |

Formula 1.0**,** standard effect size , where is sample mean, SD is standard deviation

Formula 1.1,Anti-log transformationmean = **10^’**, where ^ equivalent to “raised to”

Formula 1.2, Interval bounds = **10^ (’**

Table 2.3, Antilog transformed summary statistics for two sample test (Liver powder vs dog food)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **food** | **N** | **Mean** | **Std Dev** | **Std Err** | **Upper** | **Lower** |
| **D** | 17 | 7.728585 | 0.1486 | 0.0360 | 8.14283601 | 7.3354086 |
| **L** | 10 | 22.615177 | 0.4615 | 0.1459 | 24.6151322 | 20.777716 |
| **Diff (1-2)** |  | -0.4663 | 0.3013 | 0.1201 |  |  |

Table 2.4, Equality of variances (Antilog transformed)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Method** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **Folded F** | 9 | 16 | 9.65 | 0.0001 |

Table 2.5, T test for Equal and Unequal Variances (Antilog transformed)

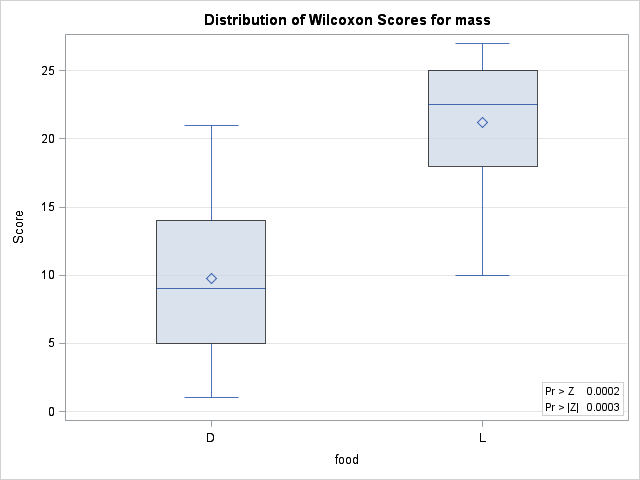
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Method** | **Variances** | **DF** | **t Value** | **Pr > |t|** |
| **Pooled** | Equal | 25 | -3.88 | 0.0007 |
| **Satterthwaite** | Unequal | 10.109 | -3.10 | 0.0111 |

The log transformation is most appropriate transformation from several observation and evaluation been made above and even the results from above could re define most factors. The log transformation makes highly skewed distribution less skewed and it was valuable for both making the patterns in my data set more interpretable and help meet assumptions of inferential statics made above. **Table 2.3 –** **2.5** shows results of how log transformation can lead to homogenous variance. The comparison of the means of the log transformed values is actually the comparison of geometric means by using **Formula1.1.**

Table 3.0, Wilcoxon Two-Sample Test

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **food** | **N** | **Sum of Scores** | **Expected Under H0** | **Std Dev Under H0** | **Mean Score** |
| **D** | 17 | 166.0 | 238.0 | 19.916492 | 9.764706 |
| **L** | 10 | 212.0 | 140.0 | 19.916492 | 21.200000 |

Figure 3.0, Box Plot of Wilcoxon Scores



**Table 3.0**, shows the results of Wilcoxon analysis. The Wilcoxon two sample test statistics equals 212, which is the sum of the Wilcoxon scores for the smaller sample. The sum is greater 166, which is expected value under the null hypothesis of no difference between the two samples, Dog food (D) and Liver (L).

**Figure 3.0**, this displays the box plot of Wilcoxon scores classified by food, which corresponds to the Wilcoxon analysis in **Table 3.0**.

**SAS OUTPUT**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **food** | **N** | **Mean** | **Std Error Mean** | **Student's t(p value)** | **Signed Rank S(p value)** | **Shapiro-Wilk W (p value)** | **Variance** |
| **D** | 17 | 0.88805882 | 0.03602945 | Pr > |t| <0.0001 | Pr >= |S|<0.0001 | Pr < W 0.1542 | 0.02206806 |
| **L** | 10 | 1.3544 | 0.14594848 | Pr > |t| <0.0001 | Pr >= |S|<0.0001 | Pr < W 0.0097 | 0.21301 |

Table 4.0 Summary statistics of univariate (Mass)

Table 4.0 Summary statistics of univariate (LOGM)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **food** | **N** | **Mean** | **Std Error Mean** | **Student's t(p value)** | **Signed Rank S(p value)** | **Shapiro-Wilk W (p value)** | **Variance** |
| **D** | 17 | 0.0569828 | 0.01695615 | Pr > |t| 0.0040 | Pr >= |S| 0.0032 | 0.4355 | 0.00488769 |
| **L** | 10 | 0.1132456 | 0.04036226 | Pr > |t|   |  | | --- | | 0.0205 | | Pr >= |S|0.0098 | Pr < W   |  | | --- | | 0.1392 | | 0.01629 |

From **Table 4.0, 4.1** the log transformation was adapted in my report analysis by using to make the data confirm to normality. Quite often when the distribution of the data set is non-normal, transformations of data are applied to make the data as normal as possible and thus increase the validity of the associated statistical analyses. Comparatively the log transformation is, arguably the most appropriate as per the out values in the tables above and helps the conformity to normality very easy