Statistical Analysis

[Enter the name of Student here]

[Enter the name of Institution here]

**Problem**

 A company that makes shoes is facing financial problems and wants to reduce production. It wants that shoe sizes are specified for men and women so that the company makes only the specified shoe sizes for males and females.

**Samples**

 A small sample set of 35 individuals has been provided in the study. It is assumed that this sample adequately represents the whole population.

**Statistical Analysis**

**Analysis of Normality**

 There are various methods used to analyze the normality of any data set. In the data available, we will use the histogram and two tests of normality to assess it. Following are the SPSS outputs for height and shoe size





The above diagrams show that shoe size and height variables are not normally distributed. The height variable is negatively or left skewed whereas the shoe size variable is positively or right skewed. In order to further test the normality, we will discuss the existence of outliers and then apply some tests of normality to accept or reject the underlying hypothesis.

Outliers Analysis for Shoesize



The above diagram shows that there are no outliers in the variable shoe size which means that there is no extra ordinary value in the data set that differ significantly from mean. The shape of histogram is justified by a greater difference between the median value and the top limit of the data i.e. 14.



The above diagram shows the outliers analysis for the variable height. There are no outliers which means that there are no values that differ significantly from the mean or median. In this case, the shape of histogram is depicted in the fact that lowest value is further away from median as compared to the highest value.

|  |
| --- |
| **Tests of Normality** |
|  | Kolmogorov-Smirnova | Shapiro-Wilk |
| Statistic | df | Sig. | Statistic | df | Sig. |
| Height | .118 | 35 | .200\* | .978 | 35 | .677 |
| ShoeSize | .195 | 35 | .002 | .928 | 35 | .055 |
|  |
|  |

The above table shows the tests for normality. The Shapiro Wilk test has been used to assess whether the sample has been obtained from a normal population, and whether the sample is distributed normally on its own as well. The sample size is too small to assess whether it has been taken from a normal population or not. The significance figures in Shapiro Wilk test will only let us reject null hypothesis regarding normality in the sample.

Once we have ascertained that the sample is normally distributed, we will now apply the tests to show if this company can use only one shoe size for both genders and for different heights of people. For gender analysis, we will use the t- test for analysis while for the height scenario, we will use the correlation.

A t- test is based upon several underlying assumptions that are discussed in the following lines:

**Assumption 1**

The dependent variable used in analysis should be continuous in nature which means that it should be measured on the ratio or interval scale. In our example the dependent variable is shoe size which is a continuous variable in a sense that it can take on any value between any two given numbers. A discrete variable on the other hand can take only whole number values. Thus, the first assumption for the t-test has been met.

**Assumption 2**

 The independent variable should be categorical in nature which means that there should be two or more distinct groups in the independent variable under consideration. Our analysis has the independent variable in the name of gender which has two distinct groups namely male and female.

**Assumption 3**

This assumption states that the individual observations included in the analysis should be independent from each other. There should be different participants in each group and there should be no participant who belongs to both groups. This assumption has been fulfilled as this study includes gender as an independent variable. There is no way that an individual can fall into both distinct groups. Another aspect is that subjects on which test has been performed should not be same. If one or more respondents or subjects fall into more than one groups, the paired sample t-test should be used. A situation where this test is applicable would be a weight loss campaign before and after some particular exercise. In this kind of experiment, subjects will be same on which test will be conducted before and after the exercise.

**Assumption 4**

There should be no significant outliers in the data. An outlier is a value that is significantly different from all the other values. Basically, outliers are values in the data which are very much different from all other values. The outliers create the issue of validity of the results of t-test. The outliers have already been discussed in relation to the shoe sizes and heights of the respondents.

**Assumption 5**

 The dependent variable should be approximately normally distributed for each group of the independent variable. The word approximately is used because t-test is quite robust to violation of normality in data. The analysis of normality in the above section has shown that the variable shoe size is approximately normally distributed.

**Assumption 6**

The variances have to be assumed to be equal if independent samples t-test has to be used. The Levene’s test in the t-test allows to check the equality or homogeneity of variances.

|  |
| --- |
| **Group Statistics** |
|  | Gender | N | Mean | Std. Deviation | Std. Error Mean |
| Shoe Size | Male | 17 | 11.294 | 1.8033 | .4374 |
| Female | 18 | 7.111 | 1.1318 | .2668 |

|  |
| --- |
| **Independent Samples Test** |
|  | Levene's Test for Equality of Variances | t-test for Equality of Means |
| F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference |
| Lower | Upper |
| ShoeSize | Equal variances assumed | 3.070 | .089 | 8.270 | 33 | .000 | 4.1830 | .5058 | 3.1540 | 5.2120 |
| Equal variances not assumed |  |  | 8.165 | 26.649 | .000 | 4.1830 | .5123 | 3.1312 | 5.2348 |

The above tables show the results of t-tests that have been run to assess if there are any differences between genders regarding their shoe sizes. There are several figures to be assessed in the above tables. Firstly, lavene’s test has shown a value greater than 0.05 which shows that variances of both groups namely male and female are same. This fulfills the assumption of homogeneity of variances for the t-test. Next important aspect in the above table is the significance column in the above table. The values in this column are less than 0.05 which is the level of significance which means that there is a significant difference between the groups of male and females with their respective shoe sizes which means that the company cannot provide one shoe size for both the genders and it has to provide at least two shoe sizes for the different genders.

After analyzing the relationship between shoe sizes and gender, we will analyze the relationship between shoe sizes and heights of the respondents. Following is a table for cross tabulation between height and shoe size.

|  |
| --- |
| **Height \* ShoeSize Crosstabulation** |
| Count  |
|  | ShoeSize | Total |
| 5.0 | 6.0 | 6.5 | 7.0 | 7.5 | 8.0 | 9.0 | 9.5 | 10.0 | 10.5 | 11.0 | 11.5 | 12.0 | 13.0 | 13.5 | 14.0 |
| Height | 60.0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 62.0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 63.0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 64.0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 65.0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 66.0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 67.0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 68.0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| 69.0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 3 |
| 70.0 | 0 | 0 | 0 | 0 | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 5 |
| 71.0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 3 |
| 72.0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 5 |
| 73.0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 2 |
| 75.0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 76.0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 77.0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Total | 1 | 2 | 4 | 5 | 4 | 1 | 1 | 2 | 2 | 2 | 3 | 1 | 3 | 1 | 1 | 2 | 35 |

A cross tabulation involving height and shoe sizes has been shown above. There is no clear evidence in above table regarding any relationship between the height and shoe size. When we run the correlation analysis between the two variables, we come to know that the correlation coefficient for males is higher than that of females. A bigger shoe is required for men as compared to women. There is a higher demand of shoe size 11-12 in males and 6.5-7.5 in females.

**Conclusion**

 First conclusion is that there is a significant difference between the two genders when they are studied keeping the shoe size in view. There is a positive correlation between heights of people and their shoe sizes. Thus, there is a need of different sizes of shoes made by the company. A shoe size of 11.5 will be suitable for men and 7 will be suitable for women.