SPSS STATISTICAL / DATA ANALYSIS

[Name of Writer]

[Date of Submission]

Table of Content

Introduction

In this report statistical analysis has been conducted via SPSS for the purpose of determining the significance of training on accuracy of malaria detection using virtual microscopy. The data comprises of 10 participants who were divided equally into two groups: control group and trained group. The initial score and final score for both the groups have been assessed in order to find out if there is any statistical significant difference before and after training. In order to test the hypothesis of the undertaken model, the researcher has applied independent sample T-test. Moreover, other supplementary tests have also been applied such as normality test, descriptive statistics, and box plots have been obtained to understand the data in a more comprehensive manner.

Data Analysis

Overview of Data

In this research, the data has been obtained from 10 participants who were bifurcated into two groups, namely: control group and trained group. Control group included the individuals who were not provided training regarding virtual microscopy and trained group included participants who were given proper training after their initial score. The following table gives a summarized view regarding the data undertaken:

		Statistic	**Std. Error**
Initial Score	Mean	13.581	1.64346
	Std. Deviation	5.19708	
Final Score	Mean	56.5	4.2687
	Std. Deviation	13.499	

Table 1: Descriptive Statistics

The above table states the value of mean and standard deviation for initial score and final score for the participants of the study. The mean value in the initial score is recorded to be 13.51 which is deviated by 1.64 units. On the other hand, in the case of final score, the mean score is found to be 56.5 which is much higher than the former. Superficially, it can be stated that overall for both the groups, the final results were significantly better as compared to the initial score. The average of final score is deviated by 4.26 points.

Hypothesis

Following is the hypothesis that is being tested by this report:

H0 = The mean values of experimental group and control group after undergoing training regarding virtual microscopy will be equal

H1 = The mean values of experimental group and control group after undergoing training regarding virtual microscopy are not equal

Normality Test

Normality tests were conducted on the data for the purpose of assessing whether or not the data is normally distributed (Park, 2015). Moreover, the normality was to be determined in order to assess which type of test was to be conducted. The following table shows the results of the normality test:

	Kolmogorov-Smirnova		**Shapiro-Wilk**	
	Statistic	Sig.	Statistic	Sig.
Initial Score	0.233	0.134	0.85	0.057
Final Score	0.202	.200*	0.925	0.399

Table 2: Normality Test

The null hypothesis for this tests is that the data is normally distributed (Norusis, 2011). For both the Kolmogorov-Smirnova and Shapiro-Wilk test, null hypothesis has not been rejected. This indicates that the data is normally distributed. Following is the Q-Q plot for initial score:

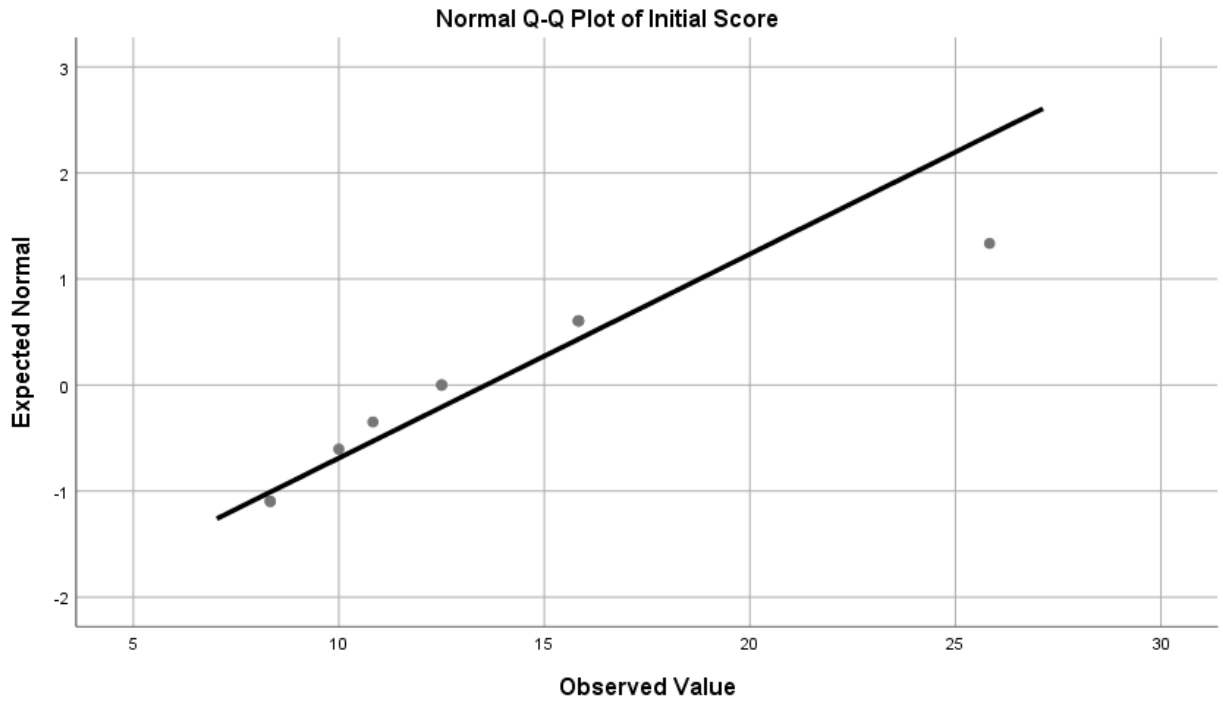

Figure 1: Q-Q Plot for Initial Scores

The above graph shows the observed values for initial score plotted against the expected values. The graph shows an upward trend where most of the values are plotted with in the line except one value which is an outlier. This further validates that the data for initial scores is normally distributed. The following graph shows the Q-Q plot for final scores:

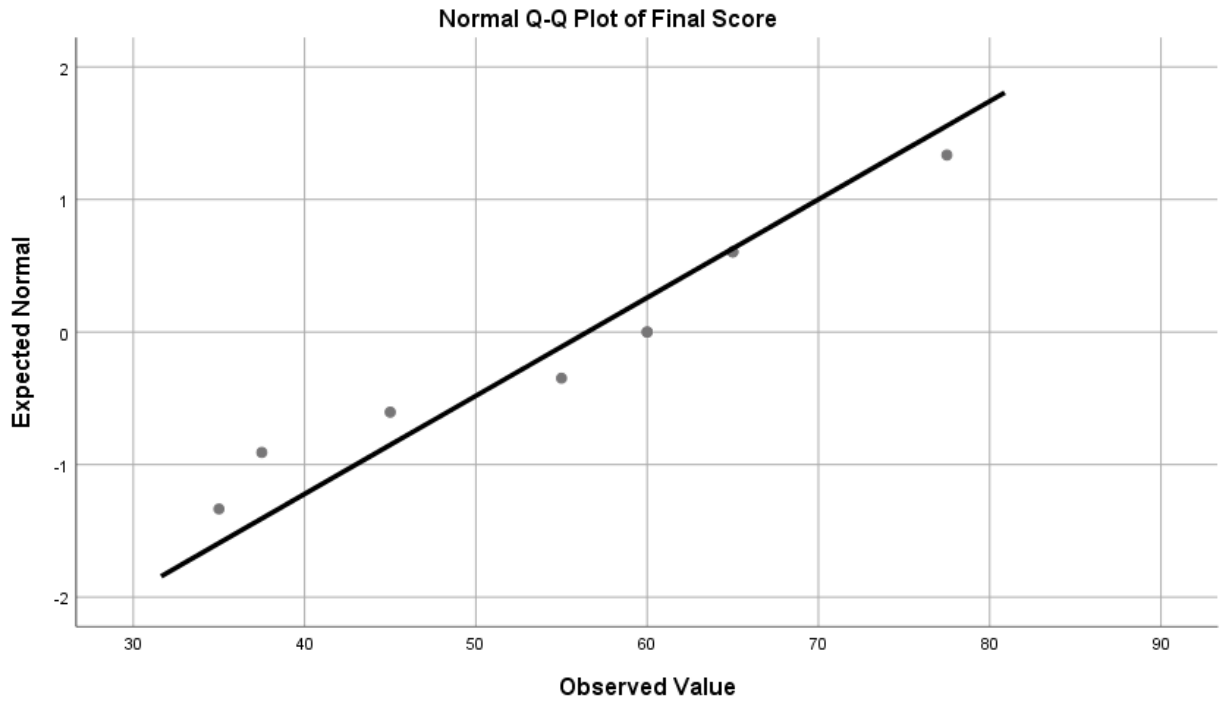

Figure 2: Q-Q Plot for Final Scores

The above graph shows the observed values for final score plotted against the expected values. The graph shows an upward trend where most of the values are plotted with in the line with no outliers. This further corroborates that the data for final scores is normally distributed.

Moreover, from the results of the normality test it can also be said that parametric test can be applied on the model. Hence, in order to test the main hypothesis of this research, independent sample T-test will be used.

Box Plots

For the data under consideration, box plots have been used for the purpose of depicting the groups of numerical data based on the quartiles. The box plots also determine the variability in the data along with the outliers in the data (Kerr, Hall, and Kozub, 2002). The following image shows the box plot for the initial scores of the participants:

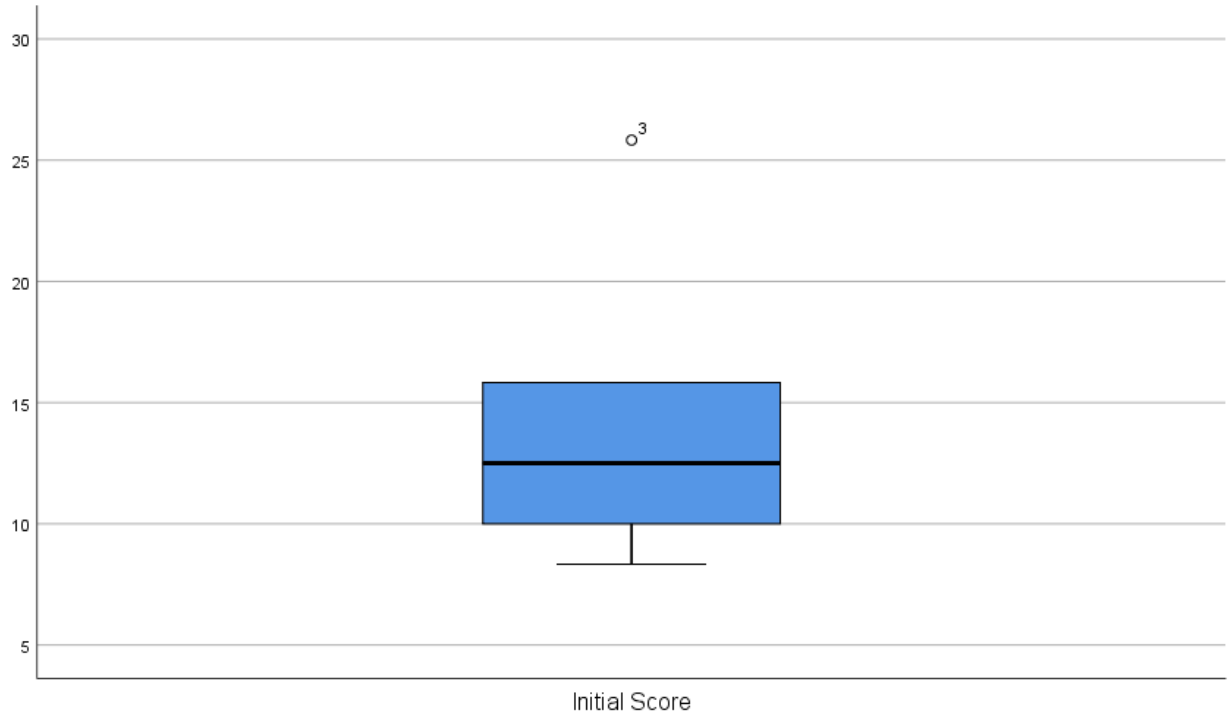

Figure 3: Box Plot for Initial Scores

The box plot shows the minimum value i.e. 8.33 and the maximum value i.e. 15.83 however there is a presence of outlier in the data which is denoted by a small dot superscripted with a 3. This indicates that the score at third number is the outlier in the data i.e. 25.83. It is considered to be an outlier because it is at an abnormal distance from the average values in the initial score. Moreover, the box plot also indicates that majority of the scores are more than the median score. The following image shows the box plot for the final scores of the participants:

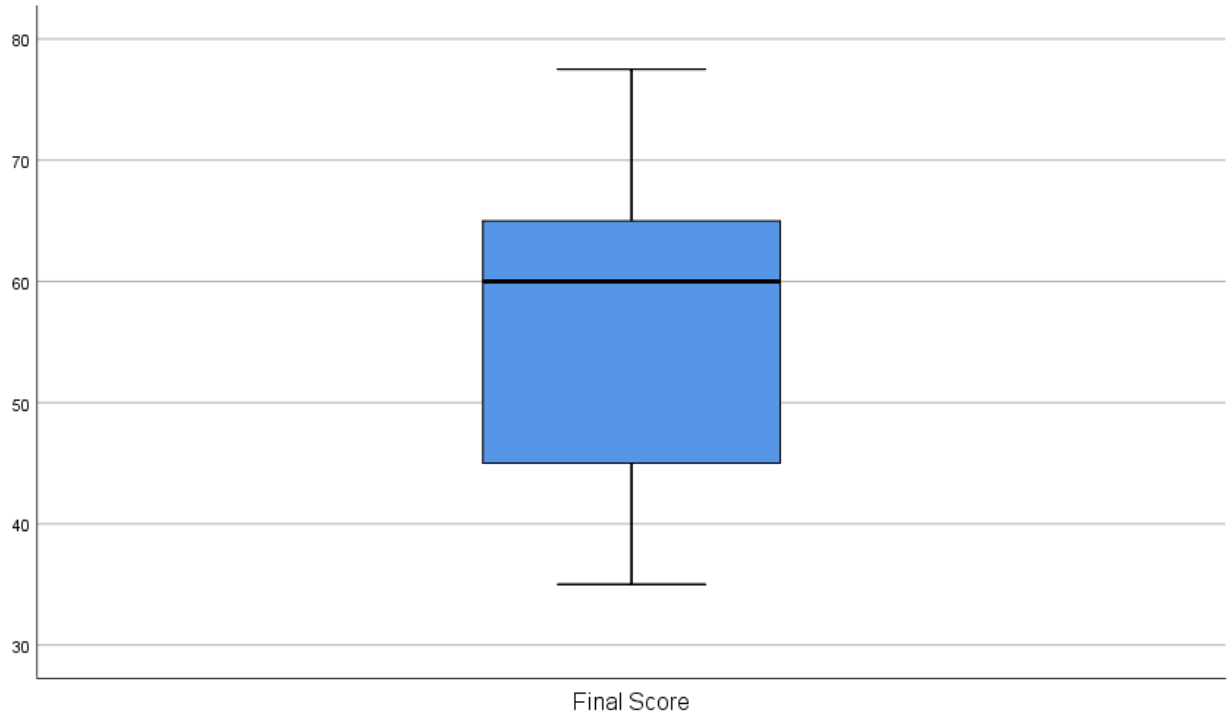

Figure 4: Box Plot for Final Scores

In the case of the final scores, the minimum value is 35 and the maximum value is 77.5. It is also apparent from this box plot that in the case of final scores, there are no outliers. The median value for the final scores appear to be 60 in the above box plot. Whereas, majority of the participants scored less than the median value that is 60. As compared to the box plot for initial scores, it can be stated that there extent of variability is significantly less in the case of final score.

Independent Sample T-test

When the population of two independent groups are compared to see the existence of difference or similarity, independent sample T-test is applied (Allen, Bennett, and Heritage, 2018). In the case of the data set that has been considered, the two independent groups are control group and trained group. The following table shows the group statistics of the model:

	Group	N	Mean	Std. Deviation	Std. Error Mean
Initial Score	Control Group	5	14.664	6.81067	3.04583
	Trained Group	5	12.498	3.3844	1.51355
Final Score	Control Group	5	46.5	10.8397	4.8477
	Trained Group	5	66.5	6.5192	2.9155

Table 3: Group Statistics

The mean values of the initial score indicates that control group had a slightly higher score as compared to the trained group, however, the deviation from the average value was significantly higher in the control group for initial score. The mean values of final score depicts that the scores were improved majorly for both the groups however, the mean score of the trained group was higher i.e. 66.5 as compared to the mean value of control group i.e. 46.5. The deviation in the average value was again higher for the control group. Overall, from the group statistic it can be evaluated that the final score has improved significantly after the intervention applied (training). However, at this stage, the significance of the difference among the mean of two group can be determined with the help of the following table:

Independent Samples Test					
		Levene's Test for Equality of Variances		**t-test for Equality of Means**	
		F	**Sig.**	**t**	**Sig. (2-tailed)**
Initial Score	Equal variances assumed	1.386	0.273	0.637	0.542
	Equal variances not assumed			0.637	0.548
Final Score	Equal variances assumed	2.482	0.154	-3.536	0.008
	Equal variances not assumed			-3.536	0.011

Table 4: Independent Sample T-test

Firstly, in the above table the sig value for Levene's test is given which hypothesize that the population of variances are equal (Marshall and Boggis, 2016). In the case of initial scores of the participants, the sig value for this test is 0.273 which is higher than alpha value at 95% of significance level hence the null hypothesis is accepted stating that population of variances are homogenous or equal. This indicates that in order to test the equality of means, the sig value for 'equal variances' will be undertaken. As per this assumption, the sig value appears to be 0.542 which means that the null hypothesis of equality of means of control group and trained group cannot be rejected.

On the other hand, in the case of final scores of the participants, the sig value for Levene's test is 0.154 which is higher than alpha value at 95% of significance level hence the null hypothesis is accepted stating that population of variances are homogenous or equal. This indicates that in order to test the equality of means, the sig value for 'equal variances' will be undertaken. As per this assumption, the sig value appears to be 0.008 which means that the null hypothesis of equality

of means of control group and trained group is rejected. Henceforth, the results have indicated that final scores for trained group and control group differs significantly.

Conclusion

In this report, statistical analysis and interpretation for comparing the results of quiz undertaken by two groups: control group and trained group have been evaluated using the independent sample T-test. The intervention that was used on the participants was provision of training for virtual microscopy. The results have indicated that there have not been a difference among the initial scores for both the control and trained groups. However, in the case of final scores which were recorded after the provision of training, a statistically significant difference was observed for both the groups. Conclusively, the results have suggested that training is an efficient intervention in improving the accuracy for malaria detection by using the method of virtual microscopy.

References

Allen, P., Bennett, K. and Heritage, B., 2018. *SPSS Statistics: A Practical Guide with Student Resource Access 12 Months*. Cengage AU.

Kerr, A.W., Hall, H.K. and Kozub, S.A., 2002. *Doing statistics with SPSS*. Sage.

Norušis, M.J., 2011. *IBM SPSS statistics 19 guide to data analysis*. Upper Saddle River, New Jersey: Prentice Hall.

Park, H.M., 2015. Univariate analysis and normality test using SAS, Stata, and SPSS.