Comparisons of one sample t-test to some parametric tests

Danjuma Munkaila

Department of Statistics, Kano University of Science and Technology, Wudil
Email: abdulkareem.doki@gmail.com

Abstract: A non-parametric equivalent test to t-test is Wilcoxon rank sum test. The research intends to investigate whether other related parametric test in particular analysis of variance can be successfully applied instead of paired t-test. Two sets of sample of size thirty were drawn from a population of four hundred candidates who sat for Post-UTME after their JAMB examination (UTME). Preliminary analysis was conducted on each set of sample to identify outliers which were excluded for further analysis. The results of the further analysis agreed with one another.

Index Terms: **sample, paired t-test, outliers, variability, analysis of variance, significant difference, level of significance.**

I. INTRODUCTION

Whenever we are confronted with taking a decision about a population based on some representative drawn from the population, then, we are faced with a problem of test of statistical hypotheses which can be accomplished by employing an appropriate statistical tool which may be parametric or non-parametric. Some of these tests involved testing the equality of means of populations.

Usually in practice students t-test is used to test the equality of two population means and the non-parametric equivalent tests are Wilcoxon rank sum test and Mann Whitney test. The research focused on dependent sample test which according to Weiers and Ronald, (1988) can also be termed as paired t- test or matched pair test. This is the appr0priate parametric test to be employed when each sample unit is subjected to two different treatments. According to Dallal, (2007) two measurements are said to be paired when they come from the same sample unit.

Analysis of variance (anova) is a parametric test used to test the equality of several population means. But according to Ronald, (2006) analysis of variance is a statistical technique for comparing two or more population means. Data being used for paired t-test arise in a similar manner with that being used analysis of variance since in each case each sample unit or experimental unit is subjected to treatments. The only difference is that the number of treatments in analysis of variance is higher than that in paired t-test. Therefore with this reason and normality assumptions, the research intends to compare and investigate the agreement or otherwise of the results obtained from paired t-test with one way and two way analysis of variance.

II. RESEARCH QUESTIONS

a. Does the results of paired t-test in each case agreed with that of one way analysis of variance?

b. Does the results of palred t-test in each case agreed with that of two way analysis of variance?

III. METHODS

Two sets of sample of size 30 each were drawn independently by simple random sampling without replacement using a computer random number generator of Minitab version 14 from a population of 400 candidates who obtained their scores in Unified Tertiary Matriculation Examination (UTME) this is labeled as X and Post Unified Tertiary Examination (PUTME) and this is labeled as Y. With the aid of Minitab version 14 a preliminary analysis was conducted on each set of sample and the mean of each set is obtained. The deviations from respective means are obtained to detect outliers which were excluded for further analyses. The two sets of refined

samples were analysed by employing a paired t-test statistic given by $t = \dfrac{\bar{d}}{s.ed}$ and one and two way analysis of variance using Minitab version 14.

IV. RESULTS

Table 1: results of descriptive statistics for the two sets of samples

Sample	Variable	Mean	St. Dev	Maximum	Minimum
1	X,Y	203.70* 112.8	17.16* 58.7	24.5* 248	182* 40
2	X,Y	198.43* 102.93	12.09* 46.21	225* 216	181* 40

The asterisks are the values of variable X

Table 2: results of paired t-test of utme and post-utme scores of sample 1

Variable	Sample size	Mean	St. Dev	T-value	P-value
X	17	205.06	19.52	13.03	0.000
Y	17	104.47	29.04		
X-Y	17	100.59	31.83		

Table: 3 results of one way analysis of variance of sample 1

Source of variation	d.f	Sum of squares	Mean square	F	P
Exam type	1	86003	86003	140.93	0,000
Error	32	19597	612		
Total	33	105600			

Table 4: results of two way analysis of variance of sample 1

Source of variation	d.f	Sum of squares	Mean square	F	P
Candidates	16	11494	718.4	1.42	0.246
Exam	1	86003	86002.9	169.82	0.000

Error	16	8103	506.4		
Total	33	105600			

Table 5: results of paired t-test of sample 2

Variable	Sample size	Mean	St. Dev	T-value	P-value
X	20	203.85	25.03	17.92	0.000
Y	20	86.80	16.68		
X-Y	20	117.05	29.21		

Table 6: results of one way analysis of variance of sample 2

Source of variation	d.f	Sum of squares	Mean square	F	P
Exam type	1	137007	137007	300.04	0,000
Error	38	17352	457		
Total	39	15435			

Table 7: results of two way analysis of variance of sample 2

Source of variation	d.f	Sum of squares	Mean square	F	P
Candidates	19	9248	487	1.14	0.388
Exam	1	137007	137007	321.24	0.000
Error	19	8103	426		
Total	39	154359			

V. DISCUSSION

From the results of table 1, there is relatively high variability in the values of y than in x in the two sets of samples. In sample 1 the standard deviations of the values of y is 58.7 while that of x is 17.16. The variability in the values of y in sample 2 is 46.21 while that of x is 12.09. Based on this results outliers are present in Y than in X. The results of table 2 indicates that the variability in the values of y has been reduced from 58.7 to 29.04 while that of x values has a small increase from 17.16 to 19.52 after the removal of 13 outliers. The results of the paired t-test shows that there is significant difference between the mean performance in utme and post-utme scores

with p-value 0.000 at 0.05 level of significance. The result of one way analysis of variance of table 3 also agreed with that of table 2 with the same p-value at 0.05 level of significance. Results of table 4 of the two- way analysis of variance shows that there is no significant difference in the performance among the candidates, that is the candidates are not significantly difference in their performance with p-value 0.246 but there is significant difference between the mean performance in the two exams with p-value 0.000 at 0.05 percent level of significance in each case. Same conclusion can be derived from the results of table 7 with p-value 0.000 and 0.388 respectively at 0.05 percent level of significance. The results of paired t-test of table 5 indicates that the variability in the values y has been reduced after the removal of outliers from 46.21 to 16.68. It can be concluded that there is a significant difference between the mean performance in two exams with p-value 0.000 at 0.05 percent level of significance. Same conclusion is derived from the results of table 6.

In conclusion it has been shown by this study that one way or two way analysis variance can be employed instead of paired t-test but it has been seen that the two way analysis of variance reveals additional information.

REFERENCES

[1]   Gerald, E.D. (2007). Paired data/paired analyses.emarald.tufts.edu~g/paired.html

[2]   Murray, R.S. and Larry, J.S, (2004). Statistics, McGraw-Hill coporation, Inc. fourth edition.

[3]   Ronald, M.W. (2006). Introduction to business statistics, Thompson higher education 5191 Natorp Boulevard Marson, 0H45040 USA.

[4]   www.statistics.laerd.com

[5]   www.emarald.tufts.edu/~gdallal/paired.html.

[6]   http://www.gla.ac.uk/sums/users/jdbmcdonald/prepost.

[7]   http://wwwa.stattrek.com/hypothesistest/pairedtest.