Development of New LSD Formula when Numbers of Observations Are Unequal

Ali A. Al-Fahham

Basic Science Department, Faculty of Nursing, University of Kufa, Najaf, Iraq
Email: aliaz.mahdi@uokufa.edu.iq

Abstract

In the case of comparison between samples with unequal replications $[n_1 \# n_2 \# n_3 \# \cdots \# n_k]$, the statistician should use LSD (Least Significant Difference) many times to achieve these comparisons, this method consumes more time and effort. The purpose of this research is to find out a new method to do easy and reliable comparison between population means when the experiment or data involving different numbers of replications. In order to do this, a new formula was designed; the aim of this new formula is to develop the multiple comparisons into one step instead of doing it with many steps that consume more effort and time. In this work, the researcher used the mean of the replication number in the LDS formula instead of doing the calculation many times for each two means. The results of this study found that the new formula give similar results for multiple comparisons with minimal effort. It was concluded that the new formula (named Al-fahham's Formula) achieves the same results with less time and efforts. It was recommended that Al-fahham's Formula can be approved in the statistical textbooks and procedures.

Keywords

ANOVA, Statistics, LSD Formula, Unequal Observations, F Test

1. Introduction

The analysis of variance ANOVA is the parametric method for identify whether differences occur significantly in an experiment including more than conditions. It is an example of simultaneous inference in which multiple comparisons are achieved and investigate only two-tailed hypotheses. The null hypothesis indicates that differences are not found between the populations represented by these conditions [1] [2].

In general, for any ANOVA with k levels, the null hypothesis is: $H_0: = \mu_1 = \mu_2$

$= \cdots = \mu_k$, so that the alternative hypothesis would be H$_1$: $\neq \mu_1 \neq \mu_2 \neq \cdots \neq \mu_k$.

When F test is significant, it reveals only that somewhere among the means of the samples at least two of them differ significantly. However, F test does not specify which specific means shows the significant difference. Thus, if F test for the multiple treatment study is significant, we will indicate that there are significant differences among the means of samples or populations, but we can't identify where they are or how much the power of the difference [2].

Therefore, when F test is significant we make a second procedure, called post hoc comparisons. Post hoc comparisons are similar to t-tests in which we do comparisons between all possible pairs of means for the studied factor, one pair comparison at a time, to indicate which means differ significantly. For analyzing the differences between means, the statistician often do specific comparisons, the most common of which is comparing two means, *i.e.* what we called "pairwise comparisons" [3]. Seven post hoc comparisons methods are used: 1) least significance difference (LSD) 2) preliminary one-way ANOVA F-test is done as pairwise comparisons by the (LSD), 3) Bonferroni inequality method, 4) Tukey Studentized range, 5) Duncan multiple range, 6) Newman-Keuls multiple range, and 7) Scheffè F-projection method [4].

The first pairwise comparison method was invented by Ronald Aylmer Fisher (1890-1962), who tried to indicate which treatments had a significant different effect in the ANOVA test, he establish his work on the least squares of the standard deviations of the treatment means with equal replications [5].

LSD is well-known methods of simultaneous inference. The main idea of this method is to be used in the comparisons between the populations as pairs. It is then used to continue in a two-way analysis of variance, indicating that the null hypothesis has already been rejected [6].

The main advantage of the LSD is to calculate the smallest significant difference between two means; it helps to determine the populations that are statistically different in means. If the overall F test is significant, LSD, a procedure analogous to ordinary Student's t test, is used to test any pair of means. If the difference between two population means is more than the LSD, then those means differ significantly at the a specific level of confidence, only one LSD value is required to test all the possible comparisons between the population means when replications are equal [4].

In the case of different numbers of observations for treatment or samples $[\,n_1 \# n_2 \# n_3 \# \cdots \# n_k\,]$, a different LSD must be calculated for each comparison involving different numbers of replications and resulting in different values of LSD for each two means, the statistician should use LSD many times to achieve the pairwise comparisons, as follows [7]:

$$\text{LSD} = t\alpha, df \sqrt{\frac{MSE}{n_1} + \frac{MSE}{n_2}} \tag{1}$$

MSE: Mean Square of Error.

2. The Aim of the Study

This study aims to reduce the time and effort for the statisticians and researchers when doing multiple comparisons for treatment with different replications or different number of observations. This is performed by using one LSD value for all comparisons, instead of calculating LSD many times.

3. The Suggested Development

If the difference between numbers of observations are not very large, statistician can use the mean number of observations (*i.e.* the mean of replications of treatment) to calculate LSD value, so that we can use the formula one time only, as follows:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

$$H_1 : \mu_1 \neq \mu_2 \neq \cdots \neq \mu_k$$

$$\text{LSD}_d = t\alpha, df \sqrt{\frac{2MSE}{\bar{n}}} \tag{2}$$

LSD_d = LSD for different number of replication

$$\bar{n} = \frac{n_1 + n_2 + \cdots + n_k}{k} \tag{3}$$

$\bar{n}$ = mean of samples obsevations, Where: $n_1 \# n_2 \# n_3 \# \cdots \# n_k$

$t\alpha$ = critical value of t for a definite level of confidence

df = degree of freedom

$$MSE \left(\text{Mean Square of error} \right) = \sum_{j=i}^{k} \sum_{i=1}^{n_i} \left(x_{ij} - \bar{x} \right)^2 \tag{4}$$

where:

k = number of groups, n = objects in group, i = row, j = column, x = group mean, $\bar{x}$ = grand mean

So that:

If $\mu_1 - \mu_2 \geq \text{LSD}_d$ so that reject null hypothesis and accept alternative, the conclusion is that the means of the two populations are different.

3.1. The Proposed Name

Alfahham's Formula

3.2. The Proposed Abbreviation

LSD_d: stands for: Least Significant Difference for different replications

4. Applications of the New LSD_d Formula

In the following examples, we will explain that using of the new LSD_d formula noticeably reduce the steps of the pairwise comparisons of samples with unequal replications, with achieving similar results.

4.1. Example (1)

Table 1 shows the Hb values of three different samples, it confirms that the suggested formula achieves the comparisons with minimal steps, as the classical method performs the comparison in three steps, while the new method does that in one step. According to Table 2, there is a high significant difference between groups.

The Classical Method

1) Male vs Children: test for difference at $\alpha = 0.01$

$$\text{LSD} = t\alpha, df \sqrt{\frac{MSE}{n_1} + \frac{MSE}{n_2}} = 2.291\sqrt{\frac{1.754}{5} + \frac{1.754}{8}}$$

$$\text{LSD} = 1.72$$

*mean difference = 16.5 − 13.4 = 3.1 > LSD

2) Male vs Female

$$\text{LSD} = t\alpha, df \sqrt{\frac{MSE}{n_1} + \frac{MSE}{n_2}} = 2.291\sqrt{\frac{1.754}{5} + \frac{1.754}{6}}$$

$$\text{LSD} = 1.83$$

*mean difference = 13.4 − 9.75 = 3.65 > LSD

3) Female vs Children

$$\text{LSD} = t\alpha, df \sqrt{\frac{MSE}{n_1} + \frac{MSE}{n_2}} = 2.291\sqrt{\frac{1.754}{6} + \frac{1.754}{8}}$$

$$\text{LSD} = 2.08$$

Table 1. Hb values of three different samples (children, female, and male).

	Children	Female	male	
	14	11	13	
	16	10	15	
	17	11	12	
	18	9	12	
	16	8	15	
	16	9.5		
	17			
	18			
n	8	6	5	$N = 19$
$\bar{x}$	16.50	9.75	13.40	$Y = 13.55$

Table 2. ANOVA table for Hb values of the three different samples.

Source of Variance	SS	df	F		P value
Treatment	156.372	2	78.1862	44.56	<0.001
Error	28.075	16	1.7547		
Total	184.447	18			

*mean difference = 16.5 − 9.75 = 6.75 > LSD

The New method (by using Al-Fahham's Formula) $\alpha = 0.01$

$$\overline{n} = \frac{5+6+8}{3} = 6.33$$

$$\text{LSD}_d = t\alpha, df \sqrt{\frac{2MSE}{\overline{n}}} = 2.291\sqrt{\frac{2 \times 1.754}{6.33}}$$

$$\text{LSD}_d = 1.7$$

*mean difference [6.75 or 3.65 or 3.1] > LSD.

4.2. Example (2)

Examples2: Table 3 shows the Hb values of three different samples, this example confirms that the new suggested formula achieves the similar results as the other methods (*i.e.* Tukey method):

According to Table 4 and Tukey method, there is a high significant different between the three samples that have different replications.

The New method

Al-fahham's Formula: $\alpha = 0.01$

$$\overline{n} = \frac{3+5+8}{3} = 5.33$$

$$\text{LSD}_d = t\alpha, df \sqrt{\frac{2MSE}{\overline{n}}} = 2.291\sqrt{\frac{2 \times 1.193}{5.33}} = 1.44$$

Table 3. Hb values of three different samples (children, female and male).

	Children	Female	male	
	14	11	13	
	16	10	12	
	17	11	13	
	18	9		
	16	9.9		
	16			
	17			
	18			
n	8	5	3	$N = 16$
$\overline{x}$	16.5	10.18	12.333	$Y = 13.004$

Table 4. ANOVA table for Hb values of the three different samples.

Source of Variance	SS	df	F		P value
Treatment	130.245	2	65.1224	54.57	<0.001
Error	15.515	13	1.1934		
Total	145.759	15			

*comparisons:

male vs female, male vs children, female vs children:

$[2.15, 4.17, 6.32] > \text{LSD}_d \,(1.44)$

According to our suggested method, there is significant different between the three samples that have different replications.

5. Conclusion

This research confirmed that the new LSD formula named Al-fahham's Formula (LSDd) achieves the same results in the pairwise comparisons as do the other methods with less time and efforts.

Recommendations

Al-fahham's Formula (LSD_d) can be approved in the textbooks of statistics and used in the statistical procedures.

References

[1]   Hothorn, T., Bretz, F. and Westfall, P. (2008) A Problem in Statistical Analysis: Simultaneous Inference. *Biometrical Journal*, **50**, 346-363. https://doi.org/10.1002/bimj.200810425

[2]   Heiman, G.W. (2011) Basic Statistics for the Behavioral Sciences. 6th Edition, Wadsworth/Cengage, Belmont, 293.

[3]   Seaman, M.A., Levin, J.R. and Serlin, R.C. (1991) New Developments in Pairwise Multiple Comparisons Some Powerful and Practicable Procedures. *Psychological Bulletin*, **110**, 577-586. https://doi.org/10.1037/0033-2909.110.3.577

[4]   Dodge, Y. and Thomas, D.R. (1980) On the Performance of Non-Parametric and Normal Theory Multiple Comparison Procedures. *Sankhya*, **B42**, 11-27.

[5]   Yates, F. and Mather, K. (1963) Ronald Aylmer Fisher: 1890-1962. *Biographical Memoirs of Fellows of the Royal Society*, **9**, 91-129. https://doi.org/10.1098/rsbm.1963.0006

[6]   Miller Jr., R.G. (1981) Simultaneous Statistical Inference. 2nd Edition, Springer, Berlin, Heidelberg, New York. https://doi.org/10.1007/978-1-4613-8122-8

[7]   Dodge, Y. (2008) Least Significant Difference Test. In: Dodge, Y., Ed., *The Concise Encyclopedia of Statistics*, Springer, New York, 302-304.