

Modification of Fisher-Hayter Test

Open
Access

Friday Zinzendoff Okwonu^{1,2}, Nor Aishah Ahad^{2,*}

¹ Department of Mathematics, Faculty of Science, Delta State University, P.M.B.1, Abraka, Delta State, Nigeria

² School of Quantitative Sciences, College of Arts and Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia

ABSTRACT

Pair wise comparison are applied to answer which group differs when the null hypothesis in ANOVA test is rejected. This study modified the conventional Fisher-Hayter test to test the significant difference between the means when ANOVA test is significant. The conventional Fisher-Hayter test use MSE due to error while the modified Fisher-Hayter test use MSE due to treatment with different degrees of freedom. The result from modified Fisher-Hayter test are then compared with Fisher least significant difference, conventional Fisher-Hayter test and Tukey Honest Significant Different test. The result reveals that the modified Fisher-Hayter test showed comparable performance compared to the conventional Fisher-Hayter test and can be used as alternative of the pairwise comparison procedures.

Keywords:

Body mass index, Fisher-Hayter Test,
Fisher's Least Significance Difference

Copyright © 2020 PENERBIT AKADEMIA BARU - All rights reserved

1. Introduction

The analysis of variance (ANOVA) is probably the most frequently applied of all statistical analyses to analyze data. ANOVA is used widely in many areas of research, such as medicine, education, sociology, psychology, economics, industry and commerce. ANOVA was designed to test for the differences in means for two or more independent groups. ANOVA procedure culminate in an assessment of the ratio of two variances based on a pertinent F-distribution and this quickly became known as a F-test [1,2]. ANOVA F value indicates the rejection or acceptance of the null hypothesis based on the statement of the equality of the means [3]. Suppose the test is not significant, a situation whereby the computed F value is less than the F critical value, in such case no further analysis is needed. On the other hand, if the test is significant, it implies that the means are not equal; hence further analysis is required to determine the group difference [4-5]. However, ANOVA does not show the means whose group differs [6-7]. Two means comparison is frequently considered otherwise referred to as "paired comparisons". Fisher coined the term in 1935. This comparison procedure was later named after Fisher and now known as "Fisher's least significant difference (FLSD)" test. "Pair wise comparison", "multiple comparison" or "post-hoc test" are applied to

* Corresponding author.

E-mail address: Nor Aishah Ahad (aishah@uum.edu.my)

answer which group differs when the null hypothesis is rejected based on the ANOVA computed F value [8].

There are several methods for performing pair wise comparison, such as the Tukey method, the Fisher's least significance difference, Bonferroni method, Dunnett method, Scheffé's test, and so on. This paper will focus on FLSD, Fisher-Hayter and Tukey test. When ANOVA test is non-significant, the FLSD and the Tukey test are meaningless. The FLSD does not control the Type 1 error. In general, whenever the sample sizes are equal the least significance difference is computed once [9-12]. As a consequence, a revised version of the FLSD test has been proposed by Hayter (and is known as the Fisher-Hayter procedure) where the modified least significant difference (MLSD) is used instead of the least significant difference. The Fisher-Hayter procedure is more conservative than the FLSD, but more powerful than the Tukey approach [5].

In this paper, we modify the existing Fisher-Hayter test to have alternative test method that expected to behave similar or better than the conventional Fisher-Hayter test. In this paper, FLSD, Fisher-Hayter (FH) test and Tukey Honest Significant Different (THSD) test are compared with the modified fisher-Hayter (MFH) test to investigate whether the MFH test can identify significant and non-significant difference at various alpha level. Therefore, this study was conducted to investigate if body mass index determines the carrying capacity of the subject under-consideration. This study provides answers to the following questions. Do under-weight pairs have the same carrying capacity or different capacity? Do normal weight pairs have the same carrying capacity or different capacity? Do over-weight pairs have the same carrying capacity or different capacity?

The rest of this paper is organized as follows: Section Two discusses the different methods while Section Three consist of data collection and analysis. Conclusion is mentioned in Section Four.

2. Methodology

2.1 Fisher's Least Significance Difference (FLSD)

The Fisher's least significance difference (FLSD) is coined on the basis that the analysis of variance approach rejects the equality of the sample means [5]. However, the Fisher's least significance difference is applied to investigate "where" the difference of means emanated from [13-14]. The mean squared due to error (MSE) is computed as follows;

$$\beta = \frac{\sum_{i=1}^k (n_i - 1) s_i^2}{\sum_{i=1}^k n_i - k}, \quad (1)$$

$$s_i^2 = \frac{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{n_i - 1}, \quad \bar{x}_i = \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i}.$$

where n_i is the sample size, k denotes measurement classification and β denotes mean square due to error (MSE), s_i^2 and \bar{x}_i are the respective sample variances and sample means, respectively.

Therefore the test statistic is $\bar{x}_i - \bar{x}_l$ and the null hypothesis is rejected at a specified α level of significance if $|\bar{x}_i - \bar{x}_l| \geq \varepsilon$, where $\varepsilon = t_{\alpha/2} \sqrt{\beta(\vartheta + \tau)}$ is the least significance difference $\vartheta = \frac{1}{n_1}$ and $\tau = \frac{1}{n_2}$, $m = n_1 + n_2$. The rule implies that if the absolute value of the mean difference is greater than ε at a specified α level we reject the null hypothesis. Observe that $t_{\alpha/2}$ in the computation of the least significance difference involves the t distribution with $\sum_{i=1}^k n_i - k$ degrees of freedom. In general, the FLSD increases the Type 1 error because it does not accommodate adjustment for several comparisons.

2.2 The Fisher-Hayter (FH) Test

The Fisher-Hayter test is a modification of the least significant difference coined by Fisher in 1935 [6]. The performance of this approach is compared with the FLSD [15-18]. This test procedure is stated as follows;

$$F - H = Q(\alpha)_{(k-1)} \sqrt{\frac{\gamma}{m}}, \tag{2}$$

where $\gamma = MSE$ and $Q(\alpha)_{(k-1)}$ denotes the alpha level of significant based on the Studentized range distribution for the range $k - 1$, and $v = m - k$ degrees of freedom [11]. The test decision is based on the following condition

$$|\bar{x}_i - \bar{x}_l| \geq F - H. \tag{3}$$

The test is significant if the above condition is true and the contrary is also true.

2.3 Modified Fisher –Hayter (MFH) Test

This test procedure modifies the Fisher–Hayter test approach by substituting the MSE due to error by MSE due to treatment (MSTR). This procedure is stated as follows

$$|\bar{x}_i - \bar{x}_j| \geq MFH \tag{4}$$

where $MFH = Q_{\alpha, N-1, (M-N)} \sqrt{\frac{MSTR}{N_i + 2(\frac{N_j}{2})}}$, and $Q_{\alpha, N-1, (M-N)}$ denotes the studentized range distribution with range $N - 1$ and $M - N$ degrees of freedom.

2.4 Tukey Honest Significant Difference (THSD) Test

The Tukey approach is applied to determine whether the existence of mean difference between pairs of means are significant based on computed Tukey value. This technique applies the “Studentized range distribution” [10] to search for significance difference among the sample means. The Tukey technique involve $(1 - \alpha)$ probability such that

$$\bar{x}_i - \bar{x}_l - Q(\alpha)_{n, n(m-1)} \leq \bar{x}_n - \bar{x}_m + Q(\alpha)_{n, n(m-1)} \frac{\gamma}{\sqrt{m}}, \tag{5}$$

where $\gamma = \sqrt{MSE}$, $n < m$. This can be summarized as follows:

$$\bar{x}_i - \bar{x}_l \geq \rho, \tag{6}$$

where the cutoff is $\rho = Q(\alpha)_{n, n(m-1)} \frac{\gamma}{\sqrt{m}}$. Suppose the mean difference is greater than ρ , this implies that the mean difference is significant. On the other hand, if

$$\bar{x}_i - \bar{x}_l < \rho, \tag{7}$$

it implies that it is not significant. The Tukey technique is used to determine whether the mean difference between the group’s understudy is significant or not. The merit of this approach is based on the fact that all assumptions on the pair-wise comparisons are considered [10-12]. Generally, the post hoc test discussed so far requires $z = \frac{n(n-1)}{2}$ comparisons.

2.5 Hypothesis for the study

H₀: all participants have the same carrying capacity based on BMI

H₁: not all participants have the same carrying capacity based on BMI

3. Data Collection and Analysis

This study investigates the time six laborers can offload blocks from local wooden ferry at the Kurutie Jetty of the Niger Delta creek of the Facados river to a building site. From source 1,500 blocks were loaded into the ferry. After a day journey on the creek it arrived its destination (Kurutie). The weight and height of the six laborers were taken, see Table 1. They gave the information below voluntarily and approved the entire process. We recorded the number of blocks each participant carried from the ferry to the building site. The distance from the ferry at the jetty to the building site is about 200 meters. All participants used wheelbarrows of the same quality and size. For every hour, we recorded the number of blocks each participant carried, see Table 2. X,Y,Z,L,Q and S were used to represents participants names.

Table 1
 Participant’s Body Mass Index

| Participants | X | Y | Z | L | Q | S |
|-------------------------|------|------|------|------|------|------|
| Weight (kg) | 68 | 62 | 53 | 78 | 66 | 83 |
| Height (m) | 1.7 | 1.5 | 1.6 | 1.8 | 1.7 | 1.8 |
| BMI(kg/m ²) | 23.5 | 27.6 | 20.7 | 24.1 | 22.8 | 25.6 |

In Table 1, Y and S are overweight while X, Z, L and Q are normal weight. The six hours exercise is contained in Table 2. This table contains the number of blocks each participant carried per hour.

Table 2
 Block Offloading Data

| X | Y | Z | L | Q | S | Total |
|------------|------------|------------|------------|------------|------------|--------------|
| 44 | 47 | 32 | 36 | 42 | 33 | 234 |
| 50 | 45 | 18 | 43 | 38 | 42 | 236 |
| 41 | 47 | 25 | 40 | 49 | 26 | 228 |
| 46 | 33 | 30 | 32 | 35 | 19 | 195 |
| 30 | 53 | 40 | 46 | 36 | 44 | 249 |
| 42 | 39 | 33 | 41 | 45 | 39 | 239 |
| 253 | 264 | 178 | 238 | 245 | 203 | 1,381 |

After offloading, 119 blocks were broken to pieces. The analysis of variance for the above information is contained in the ANOVA table.

Table 3
 ANOVA Table for Block Offloading

| Source of variation | Sum of square | DF | Mean square | F |
|---------------------|---------------|----|-------------|-------------|
| Treatments | 901.14 | 5 | 180.23 | 3.58 |
| Errors | 1511.17 | 30 | 50.37 | |
| Total | 2412.31 | 35 | | |

****Table F value (F=2.53); DF: degrees of freedom**

Since the computed F value is greater than the table F critical value, this implies that there exist significance differences between the group means. This conclusion allows for further investigations. Based on the significant nature of the test, the following post hoc tests were used to investigate the differences. To determine the number of pair-wise comparison test, we apply $\frac{n(n-1)}{2}$.

Table 4
 Difference between means for different groups and the LSD value at $\alpha = 0.05$ (8.37)

| | X | Y | Z | L | Q | S |
|---|---|--------|--------|--------|--------|--------|
| X | 0 | 1.83ns | 12.5s | 2.5ns | 1.34ns | 8.34ns |
| Y | | 0 | 14.33s | 4.33ns | 3.17ns | 10.17s |
| Z | | | 0 | 10.0s | 11.16s | 4.16ns |
| L | | | | 0 | 1.16ns | 5.84ns |
| Q | | | | | 0 | 7.0ns |
| S | | | | | | 0 |

ns: non significant (+) and s: significant(#)

Table 5
 Difference between means for different groups and the LSD value at $\alpha = 0.01$ (10.07)

| | X | Y | Z | L | Q | S |
|---|---|--------|--------|--------|--------|--------|
| X | 0 | 1.83ns | 12.5s | 2.5ns | 1.34ns | 8.34ns |
| Y | | 0 | 14.33s | 4.33ns | 3.17ns | 10.17s |
| Z | | | 0 | 10.0ns | 11.16s | 4.16ns |
| L | | | | 0 | 1.16ns | 5.84ns |
| Q | | | | | 0 | 7.0ns |
| S | | | | | | 0 |

ns: non significant (+) and s: significant(#)

Table 6
 Difference between means for different groups and the F-H Test at $\alpha = 0.05$ (11.88)

| | X | Y | Z | L | Q | S |
|---|---|--------|--------|--------|---------|---------|
| X | 0 | 1.83ns | 12.5s | 2.5ns | 1.34ns | 8.34ns |
| Y | | 0 | 14.33s | 4.33ns | 3.17ns | 10.17ns |
| Z | | | 0 | 10.0ns | 11.16ns | 4.16ns |
| L | | | | 0 | 1.16ns | 5.84ns |
| Q | | | | | 0 | 7.0ns |
| S | | | | | | 0 |

ns: non significant (+) and s: significant(#)

Table 7

Difference between means for different groups and the F-H Test at $\alpha = 0.01$ (14.63)

| | X | Y | Z | L | Q | S |
|---|---|--------|---------|--------|---------|---------|
| X | 0 | 1.83ns | 12.5ns | 2.5ns | 1.34ns | 8.34ns |
| Y | | 0 | 14.33ns | 4.33ns | 3.17ns | 10.17ns |
| Z | | | 0 | 10.0ns | 11.16ns | 4.16ns |
| L | | | | 0 | 1.16ns | 5.84ns |
| Q | | | | | 0 | 7.0ns |
| S | | | | | | 0 |

ns: non significant (+) and s: significant(#)

Table 8

Difference between means for different groups and the MFH Testat $\alpha = 0.05$ (11.78)

| | X | Y | Z | L | Q | S |
|---|---|--------|--------|--------|---------|---------|
| X | 0 | 1.83ns | 12.5s | 2.5ns | 1.34ns | 8.34ns |
| Y | | 0 | 14.33s | 4.33ns | 3.17ns | 10.17ns |
| Z | | | 0 | 10.0ns | 11.16ns | 4.16ns |
| L | | | | 0 | 1.16ns | 5.84ns |
| Q | | | | | 0 | 7.0ns |
| S | | | | | | 0 |

ns: non significant (+) and s: significant(#)

Table 9

Difference between means for different groups and the MFH Test at $\alpha = 0.01$ (14.36)

| | X | Y | Z | L | Q | S |
|---|---|--------|---------|--------|---------|---------|
| X | 0 | 1.83ns | 12.5ns | 2.5ns | 1.34ns | 8.34ns |
| Y | | 0 | 14.33ns | 4.33ns | 3.17ns | 10.17ns |
| Z | | | 0 | 10.0ns | 11.16ns | 4.16ns |
| L | | | | 0 | 1.16ns | 5.84ns |
| Q | | | | | 0 | 7.0ns |
| S | | | | | | 0 |

ns: non significant (+) and s: significant(#)

Table 10

Difference between means for different groups and the THSD Test at $\alpha = 0.05$ (11.68)

| | X | Y | Z | L | Q | S |
|---|---|--------|--------|--------|---------|---------|
| X | 0 | 1.83ns | 12.5s | 2.5ns | 1.34ns | 8.34ns |
| Y | | 0 | 14.33s | 4.33ns | 3.17ns | 10.17ns |
| Z | | | 0 | 10.0ns | 11.16ns | 4.16ns |
| L | | | | 0 | 1.16ns | 5.84ns |
| Q | | | | | 0 | 7.0ns |
| S | | | | | | 0 |

ns: non significant (+) and s: significant(#)

Table 11

Difference between means for different groups and the THSD Test at $\alpha = 0.01$ (13.01)

| | X | Y | Z | L | Q | S |
|---|---|--------|--------|--------|---------|---------|
| X | 0 | 1.83ns | 12.5ns | 2.5ns | 1.34ns | 8.34ns |
| Y | | 0 | 14.33s | 4.33ns | 3.17ns | 10.17ns |
| Z | | | 0 | 10.0ns | 11.16ns | 4.16ns |
| L | | | | 0 | 1.16ns | 5.84ns |
| Q | | | | | 0 | 7.0ns |
| S | | | | | | 0 |

ns: non significant (+) and s: significant(#)

Table 12
 Summarized performance analyses of methods

| Methods | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|
| FLSD | F-H | MFH | THSD |
| 5%(10+) / (5#) | 5%(13+) / (2#) | 5%(13+) / (2#) | 5%(13+) / (2#) |
| 1%(11+) / (4#) | 1%(15+) / (#) | 1%(15+) / (#) | 1%(14+) / (1#) |

Number of non-significant pairs (+) and significant pairs(#)

Table 12 shows different α values and the number of significant and non-significant results by the methods. The modified Fisher-Hayter test and the Fisher-Hayter test performed comparable at different alpha values. The study shows that there is significant difference between the mean blocks carried by each participant. The analysis indicates that non-significant and significant difference revealed by each test procedure is related to the carrying capacity of each participant. On the other hand, the difference of the BMI implies that there is no significant difference between participants. Hence BMI cannot be applied to determine carrying capacity of the pairs. On the other hand, mean difference indicates the strength of the participating pairs. The study shows that all the participants have different carrying capacity irrespective of their body mass index. In general, a significant pair indicates strong economic interest.

4. Conclusions

This study modified the conventional Fisher-Hayter test to test the significant difference between the means when ANOVA test is significant. The conventional Fisher-Hayter test use MSE due to error while the modified Fisher-Hayter test use MSE due to treatment with different degrees of freedom. The result reveals that the modified Fisher-Hayter test showed comparable performance compared to the conventional Fisher-Hayter test and can be used as alternative of the pairwise comparison procedures.

Acknowledgement

The Authors are thankful to Delta State University, Abraka and Universiti Utara Malaysia for their kind support to carry out this research work.

References

- [1] Kim, Tae Kyun. "Understanding one-way ANOVA using conceptual figures." *Korean journal of anesthesiology* 70, no. 1 (2017): 22.
<https://doi.org/10.4097/kjae.2017.70.1.22>
- [2] Rutherford, Andrew. *ANOVA and ANCOVA: a GLM approach*. John Wiley & Sons, 2011.
<https://doi.org/10.1002/9781118491683>
- [3] Keppel, Geoffrey. *Design and analysis: A researcher's handbook*. Prentice-Hall, Inc, 1991.
- [4] Newsom USP 634 Data Analysis I Spring, 2013.
- [5] Williams, Lynne J., and Herve Abdi. "Fisher's least significant difference (LSD) test." *Encyclopedia of research design* 218 (2010): 840-853.
- [6] Richter, S. J. , and McCan, M. H. "Using the Tukey Kramer Test as the omnibus test in the Hyter-Fisher procedure." *British Journal of Mathematical and Statistical Psychology* 65, no. 1 (2012): 499-510.
<https://doi.org/10.1111/j.2044-8317.2012.02041.x>
- [7] Lee, Sangseok, and Dong Kyu Lee. "What is the proper way to apply the multiple comparison test?." *Korean journal of anesthesiology* 71, no. 5 (2018): 353.
<https://doi.org/10.4097/kja.d.18.00242>
- [8] Chen, Shi-Yi, Zhe Feng, and Xiaolian Yi. "A general introduction to adjustment for multiple comparisons." *Journal of thoracic disease* 9, no. 6 (2017): 1725.
<https://doi.org/10.21037/jtd.2017.05.34>

-
- [9] Anderson D. R., Sweeney D. J., and Williams T. A. *Statistics for Business and Economics*. Thompson Higher Education, 2008.
- [10] Devore, Jay L. *Probability and Statistics for Engineering and the Sciences*. Cengage learning, 2011.
- [11] Hayter, Anthony J. "The maximum familywise error rate of Fisher's least significant difference test." *Journal of the American Statistical Association* 81, no. 396 (1986): 1000-1004.
<https://doi.org/10.1080/01621459.1986.10478364>
- [12] Seaman, Michael A., Joel R. Levin, and Ronald C. Serlin. "New developments in pairwise multiple comparisons: some powerful and practicable procedures." *Psychological Bulletin* 110, no. 3 (1991): 577.
<https://doi.org/10.1037/0033-2909.110.3.577>
- [13] Meier, Ulrich. "A note on the power of Fisher's least significant difference procedure." *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry* 5, no. 4 (2006): 253-263.
<https://doi.org/10.1002/pst.210>
- [14] Ruxton, Graeme D., and Guy Beauchamp. "Time for some a priori thinking about post hoc testing." *Behavioral ecology* 19, no. 3 (2008): 690-693.
<https://doi.org/10.1093/beheco/arn020>
- [15] Wang, Bushi, and Xiping Cui. "An improved uniformly more powerful exact Fisher–Hayter pairwise comparisons procedure." *Biometrical Journal* 59, no. 4 (2017): 767-775.
<https://doi.org/10.1002/bimj.201500265>
- [16] Welkowitz, Joan, Barry H. Cohen, and Robert B. Ewen. *Introductory statistics for the behavioral sciences*. John Wiley & Sons, 2006.
- [17] LeBlanc, David C. *Statistics: concepts and applications for science*. Jones & Bartlett Learning, 2004.
- [18] Toothaker, Larry E. *Multiple comparison procedures*. No. 89. Sage, 1993.
<https://doi.org/10.4135/9781412985178>