

# Comparison of Post Hoc Tests for Unequal Variance

*Mital C. Shingala*  
Assistant professor  
Shree J.D.G. Commerce & Shree S.A.S. College of  
Management,  
Varachha  
Surat-395006  
Gujarat  
India  
Email:mitalcs13@yahoo.com

*Dr. Arti Rajyaguru*  
Professor  
Department of Statistics  
Veer Narmad South Gujarat University,  
Udhana Magdalla Road,  
Surat- 395007  
Gujarat  
India  
Email:rajyaguruarti@gmail.com

**Abstract:** One way ANOVA is performed only when the assumption of homogeneity of variance hold. However, it is a robust statistic that can be used when there is a deviation from this assumption. The problem of test any differences in population means when both variances and sample sizes are unequal is called “The Behrens-Fisher problem”. In the literature associated with the problem of Behrens-Fisher, there have been some solutions proposed since the 1930’s. For example, Fisher (1935 and 1941), Welch (1947), Aspin (1948), Cochran and Cox (1950), Qin (1991) and Jing (1995) have proposed different solutions.

In this paper we have compared post hoc tests for unequal population variances for easy application purpose.

**Keywords:** Multiple Comparisons, Post-Hoc Test, Unequal Variance

## 1. INTRODUCTION:

When there are K independent sample from k different populations, the experimenter is interested in testing the omnibus Null hypothesis  $H_0: \mu_1 = \mu_2 = \dots = \mu_k$  and determine a possible alternative. Interest often centers on pairs of means. The risk of a Type I error on a single contrast is known as the per comparison error rate. The complete pair wise contrast together constitutes a family, in that if the omnibus null is true, then all pairs must be equal. Conversely, if any pair is not equal, the omnibus null must be false (Games & Howell, 1976).

Ideally, an ANOVA is performed only when the assumption of homogeneity of variance holds. However, because it is a robust statistic that can be used when there is a deviation from this assumption. When the design involves unequal variances, there are several post hoc procedures, including Games-Howell, Tamhane T2, Dunnett T3 and Dunnett C. None of the tests are exact tests, but the Tamhane T2, Dunnett T3 and Dunnett C are conservative procedures. (De Muth, 2006).

The problem of the comparison means of two populations on the basis of two independent samples from respective populations with unequal variances has been studied for many years and there is a sizable literature. Historically, this problem has come to be known as the problem of Behrens-Fisher. With unequal n of six and up, only the Behrens fisher statistics provided satisfactory control of both the “family wise rate of type 1 errors” and “type 1 error rate on each contrasts”.

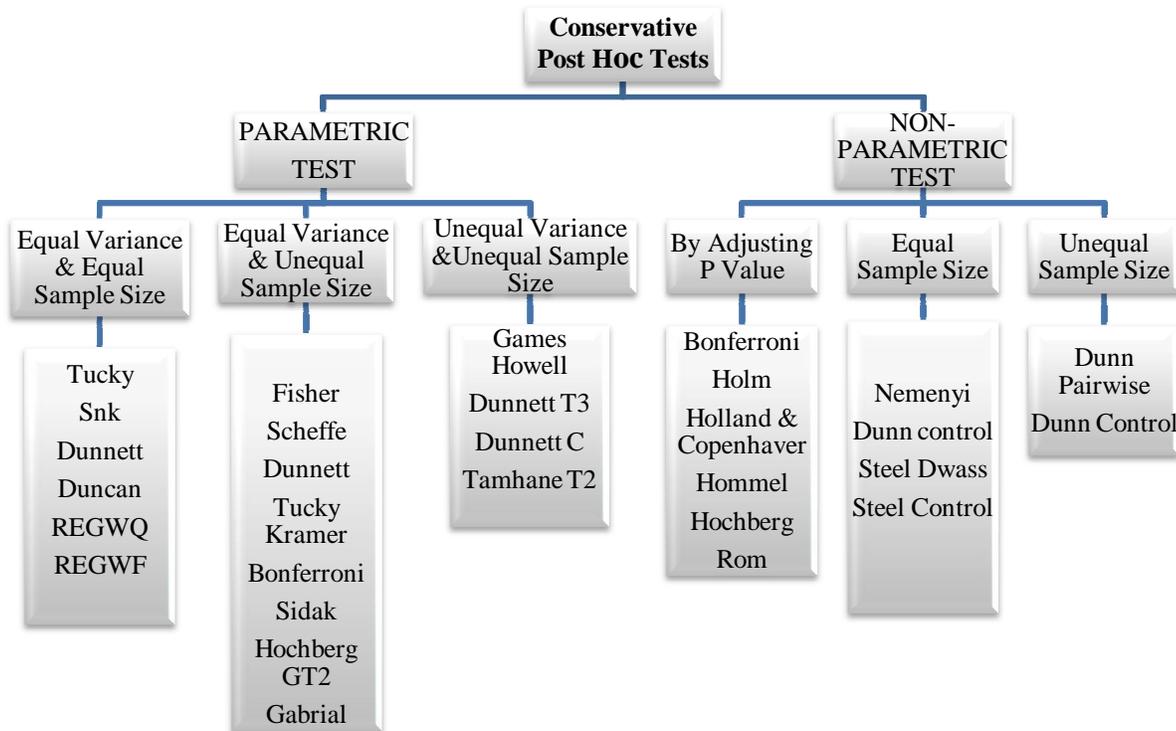
In the literature associated with the Behrens-Fisher problem, there have been some solutions proposed since the 1930's. For example, Fisher (1935 and 1941), Welch (1947), Aspin (1948), Cochran and Cox (1950), Qin (1991) and Jing (1995). Tests are proposed based on these solutions .i.e. Games and Howell (1976) is based on Welch solution to the Behrens-Fisher problem, Dunnett C (1980) is based on Cochran's (1964) solution to Behrens-Fisher problem, Dunnett T3 (1980) is based on Sidak (1967) and Tamhane T2(1980) procedure based on the Welch-type degree of freedoms adjustment in the Dunn Sidak. These procedures control the overall risk of a type I error experiment wise at approximately the nominal significance level and have the best statistical power among the alternatives.

Simulation studied by Dunnett(1980b) found that the Tucky Kramer procedure is sensitive to unequal variances and several approximate solutions for handling unequal variances were proposed, i.e. Games and Howell (1976), Tamhane T2 (1977), Dunnett T3(1980) and Dunnett C (1980) (Wilcox(1987)). When treatment variances are unequal, the pooled estimate of variance cannot be used to calculate the standard error of the comparison. For all pair wise comparisons of means, these methods are used (Day & Queen, 1989).

In this paper, we have considered the comparison and hence application of conservative tests when population variances are different.

## 2. CONSERVATIVE TESTS:

Conservative tests can be classified on the basis of their assumptions of variance and sample sizes as follows:



Hypothesis is same for all the tests as given below.

Ho:  $\mu_i = \mu_j \forall i \neq j = 1, 2, \dots, k$  Vs.  $H_1: \mu_i \neq \mu_j$  for at least for one pair.

Along with the assumption related to variances and sample size mentioned in the above figure, other common assumptions are

1. The observations in each group are independent.
2. The observations are normally distributed.
3. It requires a significant  $F$  for ANOVA.

## 2.1 GAMES HOWELL TEST (1976):

The Games-Howell (GH) procedure is an extension of the Tucky Kramer test. GH method gives the best performance for pair wise comparisons. It may be too liberal when sample sizes are small and therefore recommend when the sample sizes are greater than five (Toothaker, 1991). Along with the assumptions given earlier, in each group we must have at least 6 observations. If the omnibus null is true, then all pairs must be equal. Conversely, if any pairs are not equal, the omnibus null must be false (Games & Howell, 1976).

The Welch correction is a widely used solution for Behrens Fisher problem of testing the difference between the mean of two populations when the variances are unequal. This test is based on the Welch corrections to degree of freedoms with the  $t$ -test and uses the studentized range statistic.

This method is based on the Studentized range distribution. This method use formula for Welch's approximate degree of freedom to obtain approximate confidence interval for the difference between two means based on student's  $t$  test.

The Welch correction was designed to provide a valid  $t$ -test in the presence of unequal population variances. It involves using a corrected number of degrees of freedom  $\nu$  to assess the significance of the  $t$ -statistic computed as usual.  $\nu$  is the next smaller integer of the value obtained from the following equation:

$$\nu = \frac{\left( \frac{s_i^2/n_i + s_j^2/n_j}{\frac{s_i^2/n_i}{n_i - 1} + \frac{s_j^2/n_j}{n_j - 1}} \right)^2}{\dots(1)}$$

Where,

$s_i^2$  and  $s_j^2$  are the sample variances for the  $i^{\text{th}}$  and  $j^{\text{th}}$  groups.

$n_i$  and  $n_j$  are the respective sample sizes from population  $i$  and  $j$ .

The degree of freedom compared is usually not integer, in which case the next smaller integer should be used.

**Test Statistics:**

$$t_w = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}}} \quad \dots(2)$$

Where,

$\bar{x}_i$  and  $\bar{x}_j$  are the two sample means

$s_i^2$  and  $s_j^2$  are the sample variances for the  $i^{\text{th}}$  and  $j^{\text{th}}$  groups.

$n_i$  and  $n_j$  are the respective sample sizes from population  $i$  and  $j$ .

**Critical Value:**

$$q_{\alpha,k,v} / \sqrt{2} \quad \dots(3)$$

Where ,  $v$  is as mentioned in (1).

$k$ =Number of treatments.

$q_{\alpha,k,v}$  denotes the upper  $\alpha$  point of the studentized range distribution with parameters  $k$  and  $v$ .

The degree of freedom compared is usually not integer, in which case the next smaller integer should be used.

Reject the null hypothesis if  $t_w \geq q_{\alpha,k,v} / \sqrt{2}$  and do not reject  $H_0$  otherwise.

Several different critical values of the studentized range table will be needed, since  $v$  will tend to vary with the different pair wise comparisons. This method defines a different critical value for each pair wise comparison and determined by the variances and numbers of observations in each group under comparison (Armstrong, 2006). Confidence interval width will be different due to the different standard errors and different  $v$  (Games & Howell, 1976).

**Advantages:**

1. When statistical power is to be considered the Games-Howell procedure consistently provides narrower confidence limits.
2. It keeps the experiment wise error rate under control when we have unequal size group with unequal variances.
3. This method is robust to non normality (Day & Quinn, 1989).

**Disadvantages:**

1. This method is relatively liberal and can be too liberal when sample size is small ( $n < 15$ ) (Toothaker, 1991; Dunnett, 1980b).
2. It can exceed the nominal  $\alpha$  level when the degree of freedom is small (Wilcox, 1987; Dunnett 1980b; Rafter et al., 2002).

3. Games Howell procedure yield family wise type 1 error rate greater than the specified level alpha (Tamhane, 1979).
4. This method gives narrower confidence intervals (higher power) at the expense of occasionally excessive Type I FWER.
5. Games-Howell procedure may be slightly liberal when population variances are equal and conservative as the variances differed (Dunnett, 1980; De Muth, 2006).
6. This procedure is somewhat cumbersome to perform by hand because the degree of freedom must be recomputed, and a new critical value obtained for each comparison.

## 2.2 TAMHANE T2 TEST (1977):

T2 procedure proposed by Tamhane employs Sidak's (1967) multiplicative inequality in conjunction with the Welch approximate solution. Tamhane T2 procedure can be applied to any linear contrast, but only for pair wise differences. It is a pair wise procedure. Tamhane indicates that the Games-Howell procedure can be slightly liberal and therefore Tamhane advocate a procedure known as a Tamhane T2 method. He found inflated type I error rates for several situations but no clear pattern was identified (Olejnik & Lee, 1990).

This method is based on the student t-distribution. It uses Sidak test to set the alpha level and Welch procedure to determine degrees of freedom.

Test statistics is the same as (2).

### **Critical Value:**

$$t_{\gamma, \hat{\nu}_{ij}} \quad \dots(4)$$

$$\text{Where } \gamma = 1 - (1 - \alpha)^{1/k^*} \quad \dots(5)$$

$\hat{\nu}_{ij}$  is the same as Games Howell method.

$t_{\gamma, \hat{\nu}_{ij}}$  is the two sided  $\gamma$  point of student t distribution with  $\hat{\nu}_{ij}$  degree of freedom.

Reject the null hypothesis if  $t_w \geq t_{\gamma, \hat{\nu}_{ij}}$ , and do not reject  $H_0$  otherwise.

### **Advantages:**

1. This method can be applied to any linear contrast (Olejnik & Lee, 1990).
2. It is appropriate when variances are unequal and/or sample sizes are different (De Muth, 2006).

### **Disadvantages:**

1. It is conservative (conservative since it is based on Sidak's multiplicative inequality (Tamhane, 1979)).

### 2.3 DUNNETT T3 TEST (1980):

It is modification of Tamhane T2 multiple comparison test. It is used for pair-wise comparisons. It is priori multiple comparison tests used to determine whether the mean of a control condition differs from that of two or more experimental conditions in ANOVA (Olejnik & Lee, 1990). This test is based on the Studentized maximum modulus. Dunnett T3 procedure is used when the sample sizes are small i.e.  $n < 50$  (Hochberg & Tamhane, 1987). It is used if the degree of freedom is small. This method is based upon Sidak's (1967) uncorrelated-t inequality, which is sharper than Sidak's (1967) multiplicative inequality.

Test statistics is the same as (2)

**Critical Value:**

$$SMM_{\alpha, k^*, \nu_{ij}} \dots (6)$$

Where,

$k^*$  is the number of comparisons

$\nu_{ij}$  is same as Games Howell method.

SMM is the upper  $\alpha$  point of the Studentized Maximum Modulus distribution of  $k^*$  uncorrelated normal variates with degree of freedom  $\nu_{ij}$ .

Reject the null hypothesis if  $t_W \geq M_{\alpha, k^*, \nu_{ij}}$ , and do not reject  $H_0$  otherwise.

**Advantages:**

1. This method keeps very tight Type I error control (Field, 2009; Rafter et al, 2002, Stolone, 1981; Wilcox, 1987).
2. It is used for pair-wise comparisons.
3. It can be used for equal and unequal group sizes where the variances are unequal (Olejnik & Lee, 1990).

**Disadvantages:**

1. As more comparisons are made, the probability of type 1 error increases.
2. This method is conservative for unequal sample size.

### 2.4 DUNNETT C TEST (1980) :

Dunnett C procedure is recommended when the sample sizes are large. It is used for pair-wise comparisons, recommended if the degrees of freedom are large. It is based on a weighted average of two studentized range distribution. This method is analog of Cochran methods for the case  $k=2$ .

Test statistics is the same as (2).

**Critical Value :**

$$Q / \sqrt{2} \dots (7)$$

Where  $Q=A/B \dots (8)$

$$A = \left( q_{\alpha, k, \nu_i} \frac{s_i^2}{n_i} + q_{\alpha, k, \nu_j} \frac{s_j^2}{n_j} \right) \dots (9)$$

$$B = \left( \frac{s_i^2}{n_i} + \frac{s_j^2}{n_j} \right) \dots(10)$$

$$v_i = n_i - 1 \quad v_j = n_j - 1$$

Reject the null hypothesis if  $t_w \geq Q/\sqrt{2}$ , and do not reject  $H_0$  otherwise.

It is the weighted average of student's t proposed by Cochran (1964) as an approximate solution to the Behrens Fisher problem for  $k=2$  groups. The C procedure, which is a natural extension of Cochran's method to  $k > 2$ , does not appear to have been previously suggested for use in pair wise multiple comparisons.

#### **Advantages:**

Along with all the benefits /advantages of Dunnett T3,

1. This method is robust to non normality (Day & Queen, 1989).
2. It claims to achieve Error rates  $\leq \alpha$  in the infinite degree of freedom case.

#### **Disadvantages:**

1. Like Dunnett T3, This test takes account of the increased probability of making a Type I error the more comparisons that are made (Toothaker, 1991).
2. This method is conservative for unequal sample size.

### **3. COMPARISON:**

The methods discussed above are compared by taking based on different aspects like Confidence Interval, different situations and Conservatism:

#### **Confidence Interval:**

Dunnett T3 procedure is based on tighter inequality than Tamhane T2 method, so Dunnett T3 procedure always has shorter confidence interval than Tamhane T2 procedure, whereas Dunnett C method has shorter Confidence Interval than Dunnett T3 method for large degree of freedom, but longer Confidence Interval for small degrees of freedom. The critical value obtained by GH method is less than the Tamhane T2 procedure with equality holding if  $k=2$ . Thus GH method uniformly dominates Tamhane T2 procedure. The critical value obtained by GH method is smaller than Dunnett T3 method for  $k > 2$ , the Confidence interval obtained using the GH procedure are uniformly shorter than Dunnett T3 procedure. The GH intervals are always shorter than the Dunnett C and Dunnett T3 procedure (Rafter et al., 2002; Hochberg & Tamhane, 1987). GH procedure having shorter confidence intervals than Tamhane T2 procedure (Tamhane, 1979).

#### **Different situation:**

Stoline (1981) prefers Dunnett C procedure to the GH procedure, but Dunnett suggests that when the observations are at least 50 in each group, the GH procedure does not exceed the nominal  $\alpha$ , and is closer to the nominal  $\alpha$  level than Dunnett C. Comparing the Dunnett C procedure with the Dunnett T3 procedures, Dunnett T3 procedures gives better performance for pair wise contrasts when the degrees of freedom are small, while Dunnett C gives better performance when the degrees of freedom are moderately

large (Hochberg & Tamhane, 1987; Wilcox, 1987; Dunnett, 1980b). Dunnett T3 is generally better than Dunnett C or Games Howell procedure when degrees of freedom are less than or equal to 50 (Wilcox, 1987). Games Howell procedure seems slightly preferable to Dunnett C procedure for large degree of freedom when statistical power has been considered.

### **Conservatism:**

For small degrees of freedom, the Dunnett C procedure becomes more conservative than Tamhane T2 and Dunnett T3 (Dunnett, 1980b). Dunnett C, Dunnett T3 and Tamhane T2 procedures are found to be always conservative. Games Howell procedure is more conservative than Dunnett T3 procedure when sample sizes are small but the opposite is true when sample sizes are large. With infinite degrees of freedom Dunnett C and Games Howell procedure become identical (Dunnett, 1980b; Olejnik & Lee, 1990;). Thus the C procedure can also claim to achieve error rates  $\leq \alpha$  for infinite degrees of freedom. Games-Howell is slightly less conservative than Tamhane T2 and can be liberal when sample size is small. For larger samples that are approximately equal in size, Tamhane T2 procedure is more conservative than Dunnett T3 method where as Dunnett T3 method is more conservative than Dunnett C method for large samples, while Dunnett C is more conservative for small sample (De Muth, 2006; Field, 2009). Dunnett (1980b) has shown that for equal sample size the Dunnett T3 and Dunnett C procedures are both conservative, even under conditions of extreme variance heterogeneity. Dunnett C and Dunnett T3 procedures control the family wise Type 1 error rate. Dunnett T3 or Dunnett C procedure is better choice than the Games Howell procedure if control of the family wise type 1 error rate is especially important. Dunnett T3 is to be preferred over Tamhane T2 since it approaches more closely the nominal error rate  $\alpha$  without becoming liberal (Dunnett, 1980b).

### **Simulation Study:**

The marks obtained by BBA students in their semester 1, Semester 2, Semester 3 and Semester 4 have been taken as data. The researcher want to check whether there is a significant difference between the marks obtained in any of the semesters or not. The marks of total 406 students have been taken for that purpose. We checked the Normality by Kolmogorov-Smirnov Test and found that observations are normally distributed. We have also checked the variance Homogeneity by Levene statistics at 5% level of significance and found that all the populations have not the same variance.

We have used ANOVA to check whether there is significant difference in semester wise marks obtained by students or not. Welch statistics at 5 % level of significance is used for this purpose and found that there is at least one significant difference in semester wise marks obtained by students.

Hence Post Hoc test is used to check where this differences are. Here in our data is normal but the all populations are not having the same variance, so we have to use the method like Games-Howell, Tamhane T2 or Dunnett T3.

### Multiple Comparisons

Dependent Variable: marks obtained by students							
			Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
<b>Tamhane T2</b>	Sem1	Sem2	-27.618*	8.822	0.012	-51.05	-4.18
		Sem3	-23.345	8.9	0.054	-46.94	0.25
		Sem4	-44.148*	8.396	0.000	-66.48	-21.81
	Sem 2	Sem3	4.273	8.804	0.997	-19.12	27.66
		Sem4	-16.53	8.294	0.256	-38.65	5.59
	Sem3	Sem4	-20.803	8.377	0.081	-43.09	1.49
<b>Dunnett T3</b>	Sem1	Sem2	-27.618*	8.822	0.012	-51.04	-4.19
		Sem3	-23.345	8.9	0.054	-46.94	0.25
		Sem4	-44.148*	8.396	0.000	-66.47	-21.82
	Sem2	Sem3	4.273	8.804	0.997	-19.11	27.65
		Sem4	-16.53	8.294	0.254	-38.64	5.58
	Sem3	Sem4	-20.803	8.377	0.08	-43.08	1.48
<b>Games-Howell</b>	Sem1	Sem2	-27.618*	8.822	0.011	-50.47	-4.77
		Sem3	-23.345*	8.9	0.045	-46.36	-0.33
		Sem4	-44.148*	8.396	0.000	-65.92	-22.38
	Sem 2	Sem3	4.273	8.804	0.962	-18.53	27.08
		Sem4	-16.53	8.294	0.195	-38.09	5.03
	Sem 3	Sem4	-20.803	8.377	0.066	-42.53	0.92
<b>Dunnett C</b>	Sem1	Sem2	-27.618*	8.822		-50.66	-4.58
		Sem3	-23.345*	8.9		-46.51	-0.18
		Sem4	-44.148*	8.396		-66.14	-22.16
	Sem 2	Sem3	4.273	8.804		-18.72	27.27
		Sem4	-16.53	8.294		-38.34	5.28
	Sem 3	Sem4	-20.803	8.377		-42.75	1.15

\*. The mean difference is significant at the 0.05 level.

P value can not be obtained by Dunnett C in SPSS.

Table:1

### CONCLUSION:

From this table we conclude that by Tamhane procedure, there is a significant difference in Marks between Sem1 & Sem 2 and between Sem 1 & Sem 4. By Dunnett T3 procedure, there is a significant difference in Marks between Sem1 & Sem 2 and between Sem 1 & Sem 4. By Games Howell procedure, there is a significant difference in Marks between Sem1 & Sem 2; between Sem 1 & Sem 3; between sem 1 & Sem4. By Dunnett C procedure, there is a significant difference in Marks between Sem1 & Sem 2; between Sem 1 & Sem 3; between sem 1 & Sem4. From this analysis we can see that Confidence Interval

of Tamhane T2 procedure is higher than Dunnett T3, Dunnett C and Games Howell procedure. In general, Confidence interval can be arranged by low to high Games Howell, Dunnett C, Dunnett T3, Tamhane T2.

**Length of Confidence Intervals**

Tamhane T2 & Dunnett T3	Tamhane T2 & Games Howell	Tamhane T2 & Dunnett C	Dunnett T3 & Games Howell	Dunnett T3 & Dunnett C	Games Howell & Dunnett C
0.02	1.17	0.79	1.15	0.77	-0.38
0	1.16	0.86	1.16	0.86	-0.3
0.02	1.13	0.69	1.11	0.67	-0.44
0.02	1.17	0.79	1.15	0.77	-0.38
0.02	1.12	0.62	1.1	0.6	-0.5
0.02	1.13	0.68	1.11	0.66	-0.45

Table:2

From this analysis, we can also conclude that Confidence Interval of Tamhane T2 is higher than Dunnett T3, Confidence Interval of Tamhane T2 is greater than Games Howell, Confidence Interval of Tamhane T2 is higher than Dunnett C (It should be noted that it is smaller if sample sizes are small), Confidence Interval of Dunnett T3 is higher than Games Howell, Confidence Interval of Dunnett T3 is higher than Dunnett C (It should be noted that it is smaller if sample sizes are small), Confidence Interval of Dunnett C is higher than Games Howell. GH intervals are always shorter than Dunnett C and Dunnett T3 and Tamhane T2 procedure. This findings are supporting to the theory discussed in Section 3(Comparison).

**Ratio**

For the purpose of comparison, a wide variety of ratios of the length of the GH interval to those of the Tamhane T2, Dunnett T3 and Dunnett C method are calculated for 95% and 99 % Confidence Interval has been done and it is reported in the following table.

95 % CI			99 % CI		
Range of R(GH/T2) values	Range of R(GH/T3) values	Range of R(GH/C) values	Range of R(GH/T2) values	Range of R(GH/T3) values	Range of R(GH/C) values
0.975037337	0.975453575	0.991753472	0.989676041	0.98967604	0.98756661
0.975418521	0.975418521	0.993524714	0.98974359	0.99009376	0.99009376
0.97470338	0.975139978	0.989995452	0.989548339	0.98973306	0.98495263
0.974989312	0.97540633	0.991737334	0.989659476	0.98983595	0.9875467
0.974683544	0.975124378	0.988537368	0.989457831	0.98945783	0.98242991
0.97465231	0.975089767	0.989749431	0.989528796	0.98971386	0.98474135

Table:3

The ratio of the confidence interval is  $< 1$ , then Games Howell method is better than Tamhane T2. Clearly, the improvement of the Games Howell method over the Tamhane T2 is rather significant and same as for Dunnett T3 and Dunnett C also. The Confidence Intervals produced by GH method are significantly narrower than those produced by Tamhane T2, Dunnett T3 and Dunnett C.

Tests discussed in this paper are summarized in the following table for ready reference.

Test	Test statistic	Critical Value	Sample Large/small	Sample size Equal/unequal
Games Howell (1976)	$\frac{\bar{x}_i - \bar{x}_j}{\sqrt{\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}}}$	$q_{\alpha,k,v} / \sqrt{2}$	>5	Unequal
Tamhane T2 (1977)		$t_{\gamma, v_{ij}}$ $\gamma = 1 - (1 - \alpha)^{1/k^*}$	Small	Unequal
Dunnett T3 (1980)		$SMM_{\alpha, k^*, v_{ij}}$	Small(n<15)	Equal/Unequal
Dunnett C (1980)		$Q / \sqrt{2}$ Where $Q=A/B$ $A = (q_{\alpha, J, v_i} \frac{s_i^2}{n_i} + q_{\alpha, J, v_j} \frac{s_j^2}{n_j})$ $B = (\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j})$	Large(n>50)	Equal/Unequal

**REFERENCES:**

1. Armstrong, R. A. & Hilton, A. (2006). Post Hoc ANOVA Tests. Microbiologist. Vol 6. pp 34-36.
2. Dunnett, C. W. (1980b). Pair wise multiple comparisons in the unequal variance case. Journal of the American Statistical Association. Vol 75. pp 796-800.
3. Day, R.W. and Quinn, G.P. (1989). Comparisons of treatments after an analysis of variance in ecology. Ecological Monographs, Vol 59, No 4. pp 433-463.
4. De Muth, J. (2006). Basic Statistics and Pharmaceutical Statistical Applications. Second Edition. CRC Press. 2<sup>nd</sup> edition.
5. Field, A. (2009). Discovering statistics using SPSS. 3<sup>rd</sup> ed. London: Sage.
6. Games, P. A., and Howell, J. F. (1976). Pair wise multiple comparison procedures with unequal n's and/or variances. Journal of Educational Statistics, Vol 1. Pp 13-125.
7. Hochberg, Y. & Tamhane, A.C. (1987). Multiple Comparisons Procedures. New York: John Wiley.

8. Rafter, J., Abell, M. & Braselton, J. (2002). Multiple Comparison Methods for Means. Society for Industrial and Applied Mathematics review. Vol.44 No.2 .pp.259-278.
9. Stoline, M. R. (1981). The status of multiple comparisons: simultaneous estimation of all pair wise comparisons in one-way ANOVA designs. The American Statistician. Vol 35, pp 134-141.
10. Toothaker, L. (1991). Multiple comparisons for researchers. Newbury Park. CA: Sage.
11. Tamhane, A. C. (1979). A comparison of procedures for multiple comparisons of means with unequal variances. Journal of the American Statistical Association. Vol 74. No 366, pp 471-480.
12. Olejnik, S., & Lee, J. (1990). Multiple Comparison Procedures when Population Variances Differ. Paper presented at the Annual Meeting of the American Educational Research Association .Boston, MA, April 16-20.
13. Wilcox, R. R. (1987). New Statistical Procedures for the Social Sciences: Modern Solutions to Basic Problems. Psychology Press.
14. Zar, J. H. (1999). Bio statistical Analysis, 4th ed. Upper Saddle River, NJ. Prentice-Hall.