

Bayesian Multiple Comparisons Procedures for CRD in R Code

Paulo César de Resende Andrade*, Luiz Henrique Cordeiro Rocha, Mariana Mendes da Silva

Instituto de Ciência e Tecnologia, Universidade Federal dos Vales Jequitinhonha e Mucuri, Diamantina, Brasil

Abstract Procedures for multiple comparisons between treatment averages are of great interest in applied research. They are used to compare factor's levels means, since the most popular tests show problems related to ambiguous results and to the control of the type I error rates, besides their performance are worst in heterocedastics and unbalanced cases. This work has as objective to implement two Bayesian alternatives for multiple comparisons proposed by Andrade & Ferreira (2010) for completely randomized design in R code, contemplating the possibility of analyzing homocedastic and heterocedastic cases, with or without balancing. The implementation was illustrated by an example. In addition, the results of the two tests were compared with those of the Tukey test. The implementation was done successfully, allowing more possibility of choice for the user.

Keywords Multiple comparison procedures, Heteroscedasticity, Unbalanced

1. Introduction

A common problem in science and also in industry is the comparison of the means of some treatments of interest, to determine which of these treatments differ from each other, if this difference exists (RAFTER et al, 2002). The most usual way to treat this problem is analysis of variance (ANOVA).

Since the treatments of the fixed-effect test, the global F-test of ANOVA tests the hypothesis of equality between the population means of the treatments compared. If the F test is significant, with more than two treatments being qualitative, multiple comparison procedures (MCP) are then used to test the difference between the treatments (HOCHBERG & TAMHANE, 1987; HSU, 1996; BRETZ et al, 2010).

MCP are statistical procedures that compare two or more means. There is a vast bibliography about them (HINKELMANN & KEMPTHORNE, 1987, HOCHBERG & TAMHANE, 1987; HSU, 1996; BRETZ et al, 2010). All comparisons of treatments are performed when MCP are used. They allow the analysis of differences between means after the conclusion of the experiment to detect possible groups in a set of levels of unstructured factors.

The major problem with these tests is the ambiguity of the results, which makes interpretation difficult (MACHADO et al, 2005). This problem can be circumvented by alternative

methods of clustering (SCOTT & KNOTT, 1974; CALINSKI & CORSTEN, 1985), but they have the disadvantage of being valid only under normality.

A second problem, which does not less important, is the type I error control (HOCHBERG & TAMHANE, 1987; HSU, 1996). A common difference between MCP concerns how to measure type I error rates, which can be either by comparison or by experiment (STEEL & TORRIE, 1980; BENJAMINI & HOCHBERG, 1995). The performance varies with type I error rates and power, making it difficult to decide which MCP to use (DEMIRHAN et al, 2010).

The MCP and the F test require that certain assumptions be satisfied - the samples should be randomly and independently selected; the residues must be normally distributed and the variances must be homogeneous (RAFTER et al, 2002). Because one or more of these assumptions may be violated for a given set of data, it is important to be aware of how this would impact an inferential procedure. The insensitivity of a procedure to one or more violations of its underlying assumptions is called its *robustness*. The first assumption is the least likely to be violated, because it is under the control of the researcher. If violated, neither the MCP nor the F-test are robust. Most of the procedures seem to be robust under moderate departures from normality in that the error rate per experiment will only be slightly higher than specified.

Some MCP have been specifically developed to be used when the variances are not all equal. Many of the proposed procedures control the general risk of type I errors, but have little statistical power. Three procedures that have often been recommended are those developed by Game & Howell (1976), Dunnett C (1980), and Dunnett T3 (1980). These

* Corresponding author:

paulo.andrade@ict.ufvjm.edu.br (Paulo César de Resende Andrade)

Published online at <http://journal.sapub.org/ijps>

Copyright © 2017 Scientific & Academic Publishing. All Rights Reserved

procedures control the overall risk of a type I error experimentally at approximately the level of nominal significance and have the best statistical power between the alternative solutions. Tamhane (1979) proposed two approximated approaches for the multiple comparisons with a control and all-pairwise comparisons when the variances are unequal. Demirhan et al (2010), Ramsey et al (2010) and Ramsey et al (2011) studied the influence of violations of assumptions of normality and homogeneity of variances on the choice of a multiple comparison procedure. Tamhane (1979), Chen & Lee (2011), Li (2012), Shingala & Rajyaguru (2015), Sarmah & Gogoi (2015) have also done studies on multiple comparison procedures for populations with unequal variances.

Bootstrap resampling methods can be used in studies of multiple comparisons of the means of one-factor levels in situations of heterogeneity of variances of normal or non-normal probabilistic models (KESELMAN et al, 2002).

An alternative is the use of Bayesian procedures. A fair number of articles takes into account the problem of multiple comparisons from the Bayesian point of view (DUNCAN, 1965; WALLER & DUNCAN, 1969; BERRY, 1988; GOPALAN & BERRY, 1998; BERRY & HOCHBERG, 1999; SHAFFER, 1999; BRATCHER & HAMILTON, 2005; GELMAN et al, 2012).

Andrade & Ferreira (2010) proposed Bayesian alternatives for multiple comparisons using a methodology based on a posteriori t multivariate distribution, contemplating the possibility of analyzing both cases of homogeneity and heterogeneity of variances, with and without balancing. The proposed alternatives were superior to the other procedures studied, in the simulated examples, because they controlled the type I error and presented a greater power. In addition to having advantages over conventional tests, in the sense that there isn't need for homogeneity of variances and data balancing, that is very significant from a practical point of view. Despite the superiority of the Bayesian alternatives, they weren't implemented which made it difficult to use them.

Free programs, such as the R (R DEVELOPMENT CORE TEAM, 2017) program, are widely used to perform experiments analysis. In addition to being free, R has several packages for the most diverse areas and allows the user to create their own functions. In addition, it receives contributions from researchers from around the world in the form of packages, making a major development of the program and enabling solutions to real problems to be easily found or created by the researcher himself.

Therefore, the objective of this article is to implement in R code, two Bayesian alternatives for multiple comparisons proposed by Andrade & Ferreira (2010), in the context of completely randomized designs, when the validity assumptions are satisfied, as well as when the assumptions are not met.

2. Methodology

A function in code R (R DEVELOPMENT CORE TEAM, 2017) was programmed to perform two Bayesian tests presented by Andrade & Ferreira (2010) for the case of the completely randomized design. This function allows to analyze the experimental data considering the cases of homogeneity and heterogeneity of variances in models with normal distribution, in situations of balancing or not.

For this, a sample of size n of the multivariate t distribution $t(n, \mu, \Sigma, \nu)$ was generated, whose parameters are specified by:

$$\mu = \bar{y} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_k \end{bmatrix}, \quad \Sigma = S_{\bar{y}} = \begin{bmatrix} \frac{s_1^2}{n_1} & 0 & \dots & 0 \\ 0 & \frac{s_2^2}{n_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{s_k^2}{n_k} \end{bmatrix}, \quad \nu = n - k \quad (1)$$

where $n = \sum_{i=1}^k n_i$, k is the number of population means and

Σ is the covariance matrix of the means.

From the posteriori multivariate t distribution, we were generate k chains of means μ_i , using the Monte Carlo method and assuming constant means, vector $\mu = (\mu, \mu, \dots, \mu)^t$, that is, all the same components. Thus, without loss of generality, it was assumed $\mu = 0$ (for all k components), imposing the null hypothesis H_0 in the Bayesian method.

Following the generation of the standardized amplitude of the posteriori, under H_0 , and it was obtained in the posteriori distribution of the averages as follows:

$$q = \frac{\max(\mu_i) - \min(\mu_i)}{\sigma_h} \quad (2)$$

in which σ_h represents the harmonic mean of the variances of the mean k , given by:

$$\sigma_h = \sqrt{\frac{1}{\frac{1}{k} \left(\frac{n_1}{s_1^2} + \frac{n_2}{s_2^2} + \dots + \frac{n_k}{s_k^2} \right)}} \quad (3)$$

to consider the possibility of analyzing both the case of heterogeneous variances and the case of homogeneous variances and under non-equilibrium conditions.

The posteriori densities of the standardized amplitude distribution q and the upper α quantile of this distribution were obtained using a Kernel density estimator of the R program (R DEVELOPMENT CORE TEAM, 2017).

To make the inference about the hypothesis $H_0: \mu_i = \mu_i, i \neq i' = 1, 2, \dots, k$, considering all the pairs, it was obtained

$$\Delta = \sigma_h \cdot q_{\alpha}, \quad (4)$$

for $\alpha\%$ being q_α the upper quantile $100\alpha\%$ of the posterior distribution of $q \in \sigma_h$ obtained from expression (3).

For perform the test *dbayes*, the differences between the pairs of mean and the least significant difference Δ are compared. For any amplitude greater than Δ , the difference is significantly considered non-zero.

For second test, called *pbayes*, a new chain with the lower limits ($LI^{ii'}$) and higher ($LS^{ii'}$) of an interval a posteriori for each pair of averages (μ_i, μ_i') was obtained as follows:

$$\begin{cases} LI^{ii'} = \mu_{ij} - \mu_{i'j} - q_j \sigma_h \\ LS^{ii'} = \mu_{ij} - \mu_{i'j} + q_j \sigma_h \end{cases} \quad (5)$$

As a measure of evidence for or against $H_0: \mu_i = \mu_i'$, we calculated the posterior probability of the intervals containing the value zero. Let $I(0 \in [LI_j^{ii'}, LS_j^{ii'}])$ be the indicator function to verify that the value zero belongs to the interval in the *j*th Monte Carlo sample unit of the a posteriori chain,

$$P(\text{ICB} \supset 0) = \frac{\sum_{j=1}^n I(0 \in [LI_j^{ii'}, LS_j^{ii'}])}{n} \quad (6)$$

After performing the above procedures it was possible to implement a function named *Bayes* (), which should receive the arguments presented in Table 1.

Table 1. Bayes function arguments

Argument	Meaning
N	sample size to be simulated
alpha	significance level
file	file with the experimental data

For proper operation of the *Bayes* () function, in addition to the input arguments, the *rmvnorm* package need to be installed.

The function proposed in this work performs the analysis of variance for the case of completely randomized design; it verifies the fulfillment of the assumptions of homogeneity of variances by the Bartlett test (BARTLETT, 1937) and normality by the Shapiro-Wilk test (SHAPIRO & WILK, 1965); it performs multiple comparisons through the two Bayesian tests.

Table 2. Nitrogen content, in mg, of red clover plants inoculated with combinations of *R. trifolii* and *R. melioli* cultures

Treatments	Repetitions				
	1	2	3	4	5
1	17.7	24.8	27.9	25.2	24.3
2	17.0	19.4	9.1	11.9	15.8
3	20.7	21.0	20.5	18.8	18.6
4	14.3	14.4	11.8	11.6	14.2
5	17.3	19.4	19.1	16.9	20.8

SOURCE: Adapted from Steel & Torrie, p. 141, 1980

The performance of the tests was illustrated with a data set from a CRD experiment with red clover plants: five treatments of different cultures of five nitrogen-fixing bacteria, adapted from Stell & Torrie (1980). The data are shown in Table 2.

In addition, a comparison is made between the results presented by the two tests proposed with test's Tukey, for the same data set.

3. Results

The Bayes function (*N, alpha, file*) receives three entries, *N* is the sample size to be simulated, *alpha* is the significance level and *file* the file with the experimental data.

Initially the analysis of variance was performed and the assumptions of normality and homoscedasticity were verified.

The mean vectors μ (*Yb* in the implemented function), the covariance matrix Σ (*Syb* in the implemented function) and the degree of freedom ν (*nu* in the implemented function) for the multivariate *t* distribution generation were calculated from the experimental data, middle of the *tpostmult* function.

```
tpostmult = function(N,Yb,Syb,nu)
{
  y=rmvnorm(N,sigma=diag(nrow(Syb)))/sqrt(rchisq(N,nu)/nu)
  Sybroot = diag(diag(Syb)^0.5)
  x=y%*%Sybroot+matrix(c(rep(t(Yb),times=N)),N,nrow(Syb),byrow = TRUE)
  return(x)
}
```

By means of the *qpostbayes* function *k k* means chains μ_i were generated, using the Monte Carlo method, imposing the null hypothesis H_0 in the Bayesian method. The generation of the standardized amplitude of the posteriori was performed, under H_0 , from expressions (2) and (3).

```
qpostbayes = function(N,Yb,Syb,nu)
{
  y=rmvnorm(N,sigma=diag(nrow(Syb)))/sqrt(rchisq(N,nu)/nu)
  Sybroot = diag(diag(Syb)^0.5)
  y=y%*%Sybroot
  sigh = (1/(sum(1/diag(Syb))/nrow(Syb)))^0.5
  q = (apply(y,1,max)-apply(y,1,min))/sigh
  q=list(q=q,sigh=sigh)
  return(q)
}
```

The inference about the hypothesis $H_0: \mu_i = \mu_i', i \neq i' = 1, 2, \dots, k$, was made through two Bayesian tests (*dbayes* e *pbayes*).

To test the hypothesis of equality of means by means *dbayes* test, we obtained the least significant difference Δ from the expression (4), delta function in the code.

```
Cadeia=tpostmult(N,mmedias,mvariancia,nu)
var(Cadeia); mvariancia*nu/(nu-2)
apply(Cadeia,2,mean)
q = qpostbayes(N,mmedias,mvariancia,nu)
q$q=sort(q$q)
conf = 1-alfa
delta = (q$q[N*conf])*(q$sigh)
```

The differences were then compared with the delta value. For any amplitude greater than delta, the difference is considered significantly different from zero, that is, there is difference between the treatments of that pair.

```

letras<- character(nlevels(arquivo$trt))
i<- pos<- letra<-1
while(i<=nlevels(arquivo$trt))
{
  ind<- (media.ord[i]-(media.ord[-(1:i)]))<delta
  pos.i<- i+sum(ind)
  if(pos.i>pos)
  {
    letras.vec<- rep("",length(letras))
    letras.vec[i:pos.i]<- letters[letra]
    letras<- paste(letras,letras.vec,sep="")
    pos<- pos.i
    letra<- letra+1
  }
  i<- i+1
}
dif.letras<- c(seq(choose(nlevels(arquivo$trt),2)))
contador=0
for(i in 1:(nlevels(arquivo$trt)-1))
{
  for(j in (i+1):nlevels(arquivo$trt))
  {
    contador=contador+1
    Y=media[i]-media[j]
    Y=abs(Y)
    if(Y<delta)
    {
      dif.letras[contador]="ns"
    }
    else
    {
      dif.letras[contador]="*"
    }
  }
}
}

```

To test the hypothesis of equality of means by means of the *pbayes* test, we used the limits generated from equation (5) and calculated the posterior probability of the intervals containing the value zero, according to equation (6). There isn't difference between the pair of means for probabilities greater than 95% of zero being in the ranges.

```

Cadeial=cbind(Cadeia,as.matrix(q$q,N,1))
comb = choose(nlevels(arquivo$trt),2)
comp1 = c(seq(comb))
comp2 = c(seq(comb))
LI = matrix(nrow=N, ncol=comb)
LS = matrix(nrow=N, ncol=comb)
n = 1
for(i in 1:nlevels(arquivo$trt))
{
  for(k in i:nlevels(arquivo$trt))
  {
    if(i != k)
    {
      LI[,n] = Cadeial[,i] - Cadeial[,k]
      LS[,n] = Cadeial[,i] - Cadeial[,k]
      comp1[n] = i
      comp2[n] = k
      n = n + 1
    }
  }
}
for(j in 1:comb)
{
  for(k in 1:N)
  {
    LI[k,j]= LI[k,j] - delta
    LS[k,j]= LS[k,j] + delta
  }
}
plim=c(seq(choose(nlevels(arquivo$trt),2)))
cont=0
for (i in 1:choose(nlevels(arquivo$trt),2))
{
  cont=0
  for(j in 1:N)
  {
    if(LI[j,i]<=0 && LS[j,i]>=0)
    {
      cont=cont+1
    }
  }
  plim[i]=(cont/N)*100
}

```

The power the *pbayes* test was calculated.

```

diflim<- c(seq(choose(nlevels(arquivo$trt),2)))
for(i in 1:choose(nlevels(arquivo$trt),2))
{
  if(plim[i]>=(100-alfa*100))
  {
    diflim[i]="ns"
  }
  else
  {
    diflim[i]="*"
  }
}
poder = 0
for(i in 1:comb)
{
  poder = poder + plim[i]
}
poder = (poder/comb)

```

The implementation of the function is illustrated by the example cited in materials and methods, using $N = 10000$, $\alpha = 0.05$ and file as the experimental data of Table 2, which should be inserted in R (R DEVELOPMENT CORE TEAM, 2017) according to Figure 1.

```

> arquivo
      trt      y
1    T1 17.7
2    T1 24.8
3    T1 27.9
4    T1 25.2
5    T1 24.3
6    T2 17.0
7    T2 19.4
8    T2  9.1
9    T2 11.9
10   T2 15.8
11   T3 20.7
12   T3 21.0
13   T3 20.5
14   T3 18.8
15   T3 18.6
16   T4 14.3
17   T4 14.4
18   T4 11.8
19   T4 11.6
20   T4 14.2
21   T5 17.3
22   T5 19.4
23   T5 19.1
24   T5 16.9
25   T5 20.8

```

Figure 1. Format in which the data must be inserted in the R

The interface shown to the user for the variance analysis and the normality and homocedasticity assumptions is presented in Figure 2.

```

> Bayes(10000,0.05,arquivo)
Analysis of Variance Table

Response: y
      Df Sum Sq Mean Sq F value    Pr(>F)
trt      4 368.22   92.055    12.41 3.103e-05 ***
Residuals 20 148.36    7.418
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Shapiro-Wilk normality test

data:  ANAVA$res
W = 0.9226, p-value = 0.05872

Bartlett test of homogeneity of variances

data:  ANAVA$res by arquivo$trt
Bartlett's K-squared = 9.5756, df = 4, p-value = 0.04822

```

Figure 2. Interface with variance analysis, Shapiro-Wilk normality test and Bartlett's variance homogeneity test

It is observed that the p-value was lower than alpha (p value = 3.103×10^{-5}), that is, there is a significant difference between the treatments. By means of the Shapiro-Wilk test it is verified that the errors follow a normal or approximately normal distribution (p-value greater than alpha). It is also observed, by the Bartlett test, that the variances are heterogeneous (p-value less than alpha).

The output obtained in the R for the *dbayes* test is shown in Figure 3.

Resultados Teste dbayes

	Trat.	Com_Trat.	Teste_dbayes
1	1	2	*
2	1	3	ns
3	1	4	*
4	1	5	ns
5	2	3	ns
6	2	4	ns
7	2	5	ns
8	3	4	*
9	3	5	ns
10	4	5	ns

	Media	Diferenças
T1	23.98	a
T3	19.92	ab
T5	18.70	abc
T2	14.64	bc
T4	13.26	c

Figure 3. Output for the *dbayes* test, ns indicates non-significant and * significant at 5%. Averages followed by the same letter in the column do not differ by 5% probability by the *dbayes* test

The output for the *pbayes* test with its power in R is shown in Figure 4.

Resultados Teste pbayes

	Trat.	Com_Trat.	Prob.	Teste_pbayes	Teste_pbayes
1	1	2	8.16		*
2	1	3	82.16		*
3	1	4	0.56		*
4	1	5	59.73		*
5	2	3	59.50		*
6	2	4	97.95		ns
7	2	5	79.73		*
8	3	4	13.43		*
9	3	5	100.00		ns
10	4	5	61.48		*

Poder Teste pbayes = 56.27

Figure 4. Output to the *pbayes* test with its power. ns indicates not significant and * significant at 5%

It is observed that the *pbayes* test is more powerful than *dbayes* test since it identifies more significant differences.

The Tukey test was used as a procedure to perform multiple comparisons for the data set under analysis. The comparative results between the two proposed tests and the Tukey test are presented in Figure 5.

	Trat.	Com_Trat.	Teste_dbayes	Teste_pbayes	Tukey
1	1	2	*	*	*
2	1	3	ns	*	ns
3	1	4	*	*	*
4	1	5	ns	*	*
5	2	3	ns	*	*
6	2	4	ns	ns	ns
7	2	5	ns	*	ns
8	3	4	*	*	*
9	3	5	ns	ns	ns
10	4	5	ns	*	*

Figure 5. Comparison between the tests proposed with Tukey

In the example presented, it can be observed that the *pbayes* test shows a greater sensitivity regarding the detection of differences between treatments in relation to the *dbayes* and Tukey tests. The *dbayes* test presented a lower identification in relation to Tukey test, but it is worth noting that since the data do not show homogeneity of variances, the Tukey test result is not reliable.

4. Conclusions

The implementation of the two Bayesian alternatives (*dbayes* and *pbayes* tests) for multiple comparisons proposed by Andrade & Ferreira (2010) was done successfully. The two tests can be performed in software R (R DEVELOPMENT CORE TEAM, 2017), in the context of completely randomized designs, when the validity assumptions are satisfied, as well as when the assumptions are not satisfied.

Selecting an appropriate multiple comparison procedure requires extensive evaluation of the available information on the status of each test. Information on the importance of type I errors, power, computational simplicity, and so on, are extremely important to the selection process. In addition, selecting an appropriate multiple comparison procedure depends on data that conforms to validity assumptions. Routinely selecting a procedure without careful consideration of available and alternative information can severely reduce the reliability and validity of results.

Thus, the implementation of these two tests provides another possibility of choice for the user. The intention is to incorporate the functions developed in an R package and test their performances for other designs and analysis schemes.

REFERENCES

- [1] Andrade, P. C. R.; ferreira, D. F. (2010). Comparações múltiplas bayesianas em modelos normais homocedásticos e heterocedásticos. *Ciência e Agrotecnologia*, Lavras, v.34, n.4, p.845-852, jul./ago.
- [2] Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Statistical Society - Serie A*, v, 60, p.268-282.
- [3] Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate. *J. Roy. Statist. Soc.* 57, 289-300.
- [4] Berry, D.A. (1988). Multiple comparisons, multiple tests, and data dredging: a Bayesian perspective (with discussion). In: BERNARDO, J. M., DEGROOT, M. H., LINDLEY, D. V., SMITH, A. F. M., *Bayesian Statistics*, vol. 3. Oxford: Oxford University Press, p.79-94.
- [5] Berry, D.A.; Hochberg, Y. (1999). Bayesian perspectives on multiple comparisons. *Journal of Statistical Planning and Inference*, 82, p.215-227.
- [6] Bratcher, T., Hamilton, C. (2005). A Bayesian multiple comparison procedure for ranking the means of normally

- distributed data. *Journal of Statistical Planning and Inference* 133, p.23-32.
- [7] Bretz, F.; Hothorn, T.; Westfall, P. (2010). *Multiple Comparisons Using R*. Boca Raton, Florida, USA: Chapman & Hall/CRC Press.
- [8] Calinski, T.; Corsten, L. C. A. (1985). Clustering means in anova by simultaneous testing. *Biometrics*, v.41, n.1, p.39-48.
- [9] Chen, S.Y.; Lee, S.H. (2011). Multiple comparison procedures under heteroscedasticity. *Tamkang Journal of Science and Engineering*, Vol. 14, No. 4, pp. 293-302.
- [10] Demirhan, H.; Dolgun, N. A.; Demirhan, P.; Dolgun, M. O. (2010). Performance of some multiple comparison tests under heteroscedasticity and dependency. *J. Stat. Comput. Simul.*, Essex, v.80, n.10, p.1083-1100.
- [11] Duncan, D.B. (1965). A Bayesian approach to multiple comparisons. *Technometrics*, 7, 171-222.
- [12] Dunnett, C. W. (1980). Pairwise multiple comparisons in the unequal variance case. *Journal of the American Statistical Association*, 75(372), 796-800.
- [13] Game, P.A. and Howell, J.F. (1976). Pairwise Multiple Comparison Procedures with Unequal n's and/or Variances: A Monte-Carlo Study, *Jour. Of Educational Statistics*, 1,113-125.
- [14] Gelman, A.; Hill, J.; Yajima, M. (2012). Why We (Usually) Don't Have to Worry About Multiple Comparisons. *Journal of Research on Educational Effectiveness* 5(2), 189-211.
- [15] Gopalan, R., Berry, D.A. (1998). Bayesian multiple comparisons using dirichlet process priors. *Journal of the American Statistical Association*, 93, 1130-1139.
- [16] Hinkelmann K.; Kempthorne O. (1987). *Design and analysis of experiments*. v. 1. J. Wiley & Sons, New York. 495p.
- [17] Hochberg, Y.; Tamhane, A.C. (1987). *Multiple Comparison Procedure*. Wiley, New York.
- [18] Hsu, J. C. (1996). *Multiple Comparisons: Theory and Methods*. Chapman & Hall, London.
- [19] Keselman, H.J.; Cribbie, R.A.; Wilcox, R.R. (2002). Pairwise Multiple Comparison Tests when Data are Nonnormal. *Educational and Psychological Measurement*, 62: 420-434.
- [20] Li, H. (2012). A multiple comparison procedure for populations with unequal variances. *Journal of Statistical Theory and Applications*. Vol. 11, Number 2, pp. 165-181.
- [21] Machado, A. A., Demétrio, C.G.B., Ferreira, D.F.; Silva, J.G.C. (2005). Estatística experimental: uma abordagem fundamental no planejamento e no uso de recursos computacionais. In: *Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria*, Londrina. Anais, Reunião Brasileira da Sociedade Internacional de Biometria. 290p.
- [22] Rafter, J.; Abell, M.; Braselton, J. (2002). Multiple Comparison Methods for Means. *Society for Industrial and Applied Mathematics review*. Vol.44, N.2, pp.259-278.
- [23] Ramsey, P. H.; Ramsey, P. P.; Barrera, K. (2010). Choosing the best pairwise comparisons of means from non-normal populations, with unequal variances, but equal sample sizes. *J. Stat. Comput. Simul.*, Essex, v.80, n.5, p.595-608.
- [24] Ramsey, P. H.; Barrera, K.; Hachimine-Semprebom, P.; Liu, C-C (2011). Pairwise comparisons of means under realistic non-normality, unequal variances, outliers and equal sample sizes. *J. Stat. Comput. Simul.*, Essex, v.81, n.2, p.125-135.
- [25] R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2017. Disponível em: <http://www.R-project.org>. Acesso em 2017.
- [26] Sarmah, S.; Gogoi, B. (2015). Multiple Comparison Procedures under Equal and unequal population Variances. *International Advanced Research Journal in Science, Engineering and Technology* Vol. 2, Issue 12, pp110-116.
- [27] Scott, A. J.; Knott, M. A. (1974). Cluster analysis method for grouping means in the analysis of variance. *Biometrics*, Washington, v. 30, n. 3, p. 507-512.
- [28] Shaffer, P. J. (1999). A semi-Bayesian study of Duncan's Bayesian multiple comparison procedure. *Journal of Statistical Planning and Inference*, 82, 197-213.
- [29] Shapiro, S. S.; Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, v.52, p. 591-611.
- [30] Shingala, M. C.; Rajyaguru, A. (2015). Comparison of Post Hoc Tests for Unequal Variance. *International Journal of New Technologies in Science and Engineering*, Vol. 2, Issue 5, pp.22-33.
- [31] Steel; R.G.D.; Torrie, J.H. (1980). *Principles and procedures of statistics*. 2ª ed. McGraw-Hill Book, New York. 633p.
- [32] Tamhane, A.C. (1977). Multiple Comparisons in Model I One-Way Anova with Unequal Variances. *Communications in Statistics-Theory and Methods*, A6(1), 15-32.
- [33] Tamhane, A.C. (1979). A Comparison of Procedures for Multiple Comparisons of Means with Unequal Variances. *Journal of the American Statistical Association*, 74, 471-480.
- [34] Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, Washington, v.5, p. 99-114.
- [35] Waller, R.A., Duncan, D.B. (1969). A Bayes rule for the symmetric multiple comparisons problem. *Journal of the American Statistical Association*, 64, 1484-1503.