Optimization of Bayesian multiple comparison tests dbayes and pbayes in R code

Otimização dos testes de comparações múltiplas Bayesianos *dbayes* e *pbayes* em código R

Carlos Henrique Oliveira Ramos[1]; Fernanda Alves Araújo[2];
Paulo César de Resende Andrade[3]

Abstract

The experimental statistic uses multiple comparison procedures (MCP) to verify if there is a difference between the treatments under analysis. However, the presence of unbalanced data and the cases of heterogeneity of variance negatively influence the performance of the most used tests. The *dbayes* and *pbayes* tests were previously implemented in the context of completely randomized designs by one of the authors. These tests are valid for cases where assumptions of variance analysis are met or not, with or without balancing. The objective of this article is to optimize the Bayes function, in R code, that allows the performance of these tests. To validate the optimization, it compared the optimized code with the previous code and used three real situations: one considering all the assumptions, the other two with unbalanced data and with different numbers of treatments. The optimized Bayes function allows the *dbayes* and *pbayes* tests to perform well under conditions of assumption and balancing. These tests can be used satisfactorily in situations of non-compliance with the assumptions. In cases of unbalanced data, with a small number of treatments, the dbayes test presents a result superior to the *Tukey-Kramer* test.

Keywords: Bayesian tests. MCP. Analysis of variance. Completely randomized designs.

Resumo

A estatística experimental utiliza procedimentos de comparações múltiplas (PCM) a fim de verificar se há diferença entre os tratamentos em análise. Entretanto, a presença de dados desbalanceados e casos de heterogeneidade de variâncias influencia negativamente o desempenho dos testes mais utilizados. Os testes *dbayes* e *pbayes* foram implementados anteriormente no contexto dos delineamentos inteiramente casualizados por um dos autores. Esses testes são válidos para casos em que as pressuposições da análise de variância são atendidas ou não, com ou sem balanceamento. O presente artigo tem por objetivo realizar uma otimização da função Bayes, em código R, que permite a realização destes testes. Para validar a otimização, comparou-se o código otimizado com o código anterior e utilizou três situações reais: uma atendendo a todas as pressuposições, as outras duas com dados desbalanceados e com número diferente de tratamentos. A função Bayes otimizada propicia que os testes *dbayes* e *pbayes* tenham bons resultados em condições de atendimento das pressuposições e balanceamento. Estes testes podem ser utilizados satisfatoriamente nas situações de não atendimento das pressuposições. Nos casos de dados desbalanceados, com um pequeno número de tratamentos, o teste dbayes apresenta resultado superior ao teste de *Tukey-Kramer*.

Palavras-chave: Testes Bayesianos. PCM. Análise de variância. Delineamento completamente aleatorizado.

---
[1] Graduando em Ciência e Tecnologia, UFVJM, Diamantina, MG, Brasil.; E-mail: carlosoliveira301098@hotmail.com
[2] Graduanda em Engenharia Mecânica, UFVJM, Diamantina, MG, Brasil; E-mail: fernanda.araujo01@yahoo.com
[3] Prof. Dr., Inst. de Ciência e Tecnologia, UFVJM, Diamantina, MG, Brasil; E-mail: paulo.andrade@ict.ufvjm.edu.br

Ramos, C.H.O.[1]; Araújo, F.A. [2]; Andrade, P.C.R.[3]

Introduction

Experimental Statistics is a daily need of researchers from several areas of knowledge; it is used as an auxiliary tool in decision making. Due this necessity, several methods have been established and incorporated by researchers and practitioners in their work and among these methods, there is the multiple comparison tests, that compare the means of the treatments of interest to determine which of these treatments differ from one another (RAFTER; ABELL; BRASELTON, 2002).

Analysis of variance (ANOVA) was the first method for the analysis of experimental data. It was developed by Ronald Fisher from the 1920 (FISHER, 1973). The F test compares the means in the analysis of variance, but does not designate which means differ from one another. Therefore, multiple comparison procedures (MCP) are used to identify these differences. These are valid when the F test is significant, the nature of the treatments is qualitative and there are more than two treatments, indicating whether or not there is a difference between the treatments. (BRETZ; HOTHORN; WESTFALL, 2010; HOCHBERG; TAMHANE, 1987; HSU, 1996).

Thus, all pairs comparisons of the means of the treatments are performed by the MCPs that analyze the differences between the means after the conclusion of the experiment (BRETZ; HOTHORN; WESTFALL, 2010; HINKELMANN; KEMPTHORNE, 1987; HOCHBERG; TAMHANE, 1987; HSU, 1996). However, it is difficult to choose which MCP to use, since their performance varies in relation to the type I and power error rates (DEMIRHAN *et al.*, 2010).

The use of the F-test depends on the verification of four assumptions to be valid. These assumptions or fundamental assumptions of the analysis of variance are: additivity of the effects allowed in the model, independence, homogeneity of variances and normality of the residues (RAFTER; ABELL; BRASELTON, 2002). If at least one of these hypotheses is not satisfied, the analysis of variance has no validity as a statistical analysis technique and becomes a simple mathematical treatment of the data collected.

However, first two assumptions are less likely to be violated because they are under the control of the researcher. If violated, neither the MCP nor the F-test is robust. Most MCPs appears to be robust at moderate distances from normality, as the error rate per experiment will be only slightly higher than that specified (ANDRADE; ROCHA; SILVA, 2017).

In addition, violation of the hypothesis of homogeneity of variances can affect the performance of the method and compromise the results in different ways. Several studies have been carried out (DUNNETT, 1980; GAME; HOWELL, 1976; TAMHANE, 1977). It is known that heteroscedasticity can affect the inference and may have a direct impact on the conclusions. The presence of variance heterogeneity may also have a serious effect on the validity of the F-test, especially when the sizes of these samples are unbalanced (CHEN; LEE, 2011; DEMIRHAN et al., 2010; KEYES; LEVY, 1997; LI, 2012; O'BRIEN, 1978; RAMSEY; RAMSEY; BARRERA, 2010; RAMSEY; BARRERA; HACHIMINE-SEMPREBOM, 2011; SARMAH; GOGOI, 2015; SHINGALA; RAJYAGURU, 2015; TAMHANE, 1979). Furthermore, an alternative is the use of Bayesian procedures to make multiple comparisons (ANDRADE; FERREIRA, 2010; BERRY, 1988; BERRY; HOCHBERG, 1999; BRATCHER; HAMILTON, 2005; DUNCAN, 1965; GELMAN; HILL; YAJIMA, 2012; GOPALAN; BERRY, 1998; SHAFFER, 1999; WALLER; DUNCAN, 1969).

Andrade, Rocha and Silva (2017) implemented two Bayesian alternatives for multiple comparisons proposed by Andrade and Ferreira (2010), in the context of completely randomized designs using software R (R DEVELOPMENT CORE TEAM, 2019). The methodology was based on the a *posteriori* multivariate *t* distribution, contemplating the possibility of analyzing cases, homogeneity and heterogeneity of variances, with and without balancing.

The objective of the paper is to optimize the Bayes function in order to obtain more reliable results, to extend the tests *dbayes* and *pbayes* for unbalanced data and to improve the man-machine interaction. Also, to compare the optimized function with the previous one and validate the same with real examples in different situations.

Methods

Firstly, to optimize the *Bayes* function, it was necessary to understand the programming logic of this function implemented by Andrade, Rocha and Silva (2017) and the theoretical study carried out by Andrade and Ferreira (2010). The Bayes function was programmed in R code in the context of completely randomized design (CRD). This function is composed of three arguments: sample size to be simulated (*N*), level of significance (*alpha*) and data set to be analysed (*file*).

The *Bayes* function is able to test the normality and the additivity of the residues, the homogeneity of variances, performs the analysis of variance and allows and the comparison of average of qualitative treatments. The averages can be differentiated by the tests of multiple comparisons *dbayes* and *pbayes* (ANDRADE; ROCHA; SILVA, 2017).

For this, a sample of size *n* of the multivariate *t* distribution is generated. The Monte Carlo method is used to generate *k* chains of means based on the distribution a posteriori t multivariate. Later, the a posteriori distribution of the means is obtained through the generation of the standardized amplitude a *posteriori*, under H_0 (ANDRADE; FERREIRA, 2010). Subsequently, used a Kernel density estimator of the R program to obtain densities a *posteriori* of the standardized amplitude distribution *q* and the upper quantile α percentage of that distribution.

According to Andrade and Ferreira (2010), to make the inference about the null hypothesis, considering all pairs, it is necessary to perform the calculation given by equation (1)

$$\Delta = \sigma_h . q_\alpha, \tag{1}$$

in which q_α is the upper quantile $100q_\alpha\%$ of the posterior distribution of *q* and σ_h was obtained according to equation (2)

$$\sigma_h = \sqrt{\frac{1}{\frac{1}{k}\left(\frac{n_1}{s_1^2} + \frac{n_2}{s_2^2} + \ldots + \frac{n_k}{s_k^2}\right)}}. \tag{2}$$

The equation (2) allows analyzing cases with different repetitions numbers and variances.

The *dbayes* test calculates the difference between the pairs of means with the Δ (least significant difference). The hypothesis $H_0 : \mu_i = \mu_{i'}$ is rejected when the modulus of the difference between the pairs is greater than Δ (ANDRADE; FERREIRA, 2010).

The pbayes test calculates probability a posteriori of the intervals containing the value zero. These intervals were obtained using the lower ($LI^{ii'}$) and upper ($LS^{ii'}$) limits of an interval a *posteriori* for each pair of means ($\mu_i, \mu_{i'}$), as shown in equation (3)

$$\begin{cases} LI^{ii'} &= \mu_{ij} - \mu_{i'j} - q_j\sigma_h \\ LS^{ii'} &= \mu_{ij} - \mu_{i'j} + q_j\sigma_h. \end{cases} \tag{3}$$

Treatments are considered equal if zero is contained within this range, otherwise they are considered as different.

Due to the previously described study it was noted that the *Bayes* function did not perform the *dbayes* and *pbayes* tests for the unbalanced data case. In order to so, it was necessary to expand the parameters of the *qpostbayes* (N, Yb, Syb, nu) function and *tposmult* (N, Yb, Syb, nu). According to Andrade and Ferreira (2010) the most important parameters related to the presence or not of the missing data are Yb and Syb. The parameter Yb is a vector whose inputs are the averages of each treatment. When there is different number of repetitions the calculation of each mean differs according to the quantity of plots of each treatment, as shown in equation (4):

$$\sum_{i=1}^{n} \frac{d_i}{n_i}, \tag{4}$$

where in d_i means the treatment data *i* and n_i the number of the treatment repeat *i*.

The parameter Syb is a diagonal matrix. This matrix stores the values of the mean square of the error divided by the number of repetitions of each treatment, denoted by equation (5)

$$\begin{bmatrix} \dfrac{\text{Mean } Sq}{n_1} & 0 & \ldots & 0 \\ 0 & \dfrac{\text{Mean } Sq}{n_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \dfrac{\text{Mean } Sq}{n_k} \end{bmatrix}. \tag{5}$$

The number of repetitions of each treatment is stored in a vector (*nrt*), which it is used to calculate the parameters Yb and Syb. Another parameter that modifies its value with the presence of missing data is the degree of freedom of the residues (*nu*).

In addition, it was observed that the *Bayes* function use several packages, which made it difficult for inexperienced users to interact with the function. Furthermore, the results of the tables show by the *Bayes* function are confusing and misaligned. Therefore, optimizations were needed in these aspects.

Finally, comparing the results of Andrade and Ferreira (2010) and those of the *Bayes* function, differences were observed in the results. These differences occurred because of the divergent calculus of the *delta*, equation (1), and the *pbayes* test regarding to the theory. Therefore, it was necessary to change the programming of both.

The optimization performed in the *Bayes* function was performed in steps: introduction of the method of loading and/or installing packages; enhancement of the delta

65

Ramos, C.H.O.[1]; Araújo, F.A. [2]; Andrade, P.C.R.[3]

calculation and determination of the *pbayes* test limits; insertion of the code that allows the use of unbalanced data and organization of the outputs of the tables in R. The optimization allows performing the tests with greater precision.

Then the optimized code was compared to the code implemented by Andrade, Rocha and Silva (2017). In this case, a dataset of a CRD experiment with red clover plants was used: five treatments of different cultures of five nitrogen-fixing bacteria, the data were adapted from Steel and Torrie (1980). The variations are heterogeneous, that is, not all assumptions have been met.

The performance of the optimized code was illustrated in different situations, three experiments were used. The first deals with blood clotting time taken from 24 animals randomly allocated to four diets (BOX; HUNTER; HUNTER, 2005). This study meets all the assumptions required by the analysis of variance, allowing the comparison of the *dbayes* and *pbayes* tests with the traditional tests: Tukey (TUKEY, 1949), Scott-Knott (SK) (SCOTT; KNOTT, 1974) and Calinski and Corsten (CCF) (CALINSKI; CORSTEN, 1985).

Additionally, the second and third situations present unbalanced data, contemplating numbers of different treatments. The results were compared using the *Tukey-Kramer* test (KRAMER, 1956). The second experiment compares different blankets used to warm patients after a surgical procedure (WESTFALL; RANDALL; RUSSELL 2011). The third study presents the body temperatures of calves that were submitted to different vaccines and then tested to verify their effectiveness (MILLIKEN; JOHNSON, 1943).

Results

Initially, conditional structures were created for the packet-checking functions in R, so that all the packages needed to run the tests were installed and/or loaded. The packages and their purposes are shown in Table 1.

Table 1 – Packages used by the *Bayes* function.

Packages	Objective
lmtest	Performing the Durbin-Watson test
multcomp	Performing multiple comparison tests
mvtnorm	Generation of the data chains
stringr	Formatting tables
dplyr	Ordering data
car	Ordering data

Source: The authors.

Then, the function *qpostbayes* calculate the covariance matrix, the mean vectors and the degree of freedom to generate the multivariate *t* distribution. By means of the *qpostbayes* function were generated k chains of means, using the Monte Carlo method, and the standardized amplitude the *posteriori* (ANDRADE; ROCHA; SILVA, 2017).

In relation to the optimizations and corrections performed for the *pbayes* test, the limits generated from equation (3) were used to test the hypothesis of equality of means. The section corresponding to this step in the optimized code is present in Algorithm 1.

Algorithm 1 - Part of the code for generating upper and lower limits with the comparison between treatments

```
for (i in 1:nlevels (file$trt))
{
  for (j in i:nlevels (file$trt))
  {
   if (i != j)
   {
    LI[,n] = Chain1[,i] − Chain1[,j] − Chain1[,kk] ∗ q$sigh
    LS[,n] = Chain1[,i] − Chain1[,j] + Chain1[,kk] ∗ q$sigh
    comp1[n] = i
    comp2[n] = j
    n = n + 1
   }
  }
}
```

Source: The authors.

For the *dbayes* test, the hypothesis of equality of averages by means of the function *delta* in the code was tested. The absolute value of the difference between the pairs of means of treatments was compared with the *delta*. For any amplitude bigger then *delta*, there is a difference between the treatments compared; otherwise, the treatments are equal.

While the *pbayes* test, the hypothesis of means equality was tested by calculating the probability *posteriori* of the intervals containing the zero value. For this, created a vector that stores the probability values of the test, generating the Algorithm 1 where the symbol $''ns''$ indicates that there is no difference between the means of the treatments and the symbol $''*''$ indicates that there is a significant difference, as the Algorithm 2.

Algorithm 2 - Conditional structure to characterize the comparisons between the treatments using $''*''$ and $''ns''$

```
dif.letters < − c(seq(choose(nlevels(file$trt),2)))
counter=0
for (i in 1:nlevels(file$trt)-1)
  {
    for (j in (i+1):nlevels(file$trt))
    {
      counter=counter+1
      Y=mean[i]- mean[j]
      Y=abs(Y)
      if ( Y < delta)
      {
        dif.letters [counter]=''ns''
      } else
      {
        dif.letters [counter]=''*''
      }
    }
  }
```

Source: The authors.

Algorithm 3 - Fragments of the R code developed for the parameter *Syb*

```
for (i in 1:nlevels(file$trt))
  {
    for (l in 1:nlevels(file$trt))
    {
      if (i= =l)
       {
       mvariance[i,l]=vvariance[i]/nrt[i]
      }
      if (i ! = l)
      {
      mvariance[i,l] = 0
      }
    }
  }
```

Source: The authors.

Another improved point was regarding the output of results to the user. Through the functions to generating tables it was possible to improve the organization of the information, facilitating the visualization and improving the aesthetics of the presentation of the comparisons between the tests.

Table 2 show the Algorithm 2 after the optimization.

Table 2 – Output of the Algorithm 2 comparing *dbayes* and *pbayes* tests after optimization.

Comparison of Tests		
	dbayes_Test	bayes_Test
B – A = = 0	*	ns
C – A = = 0	*	*
C – B = = 0	ns	ns

Source: The authors.

The modification of the parameters *Syb* and *Yb* contributes to the optimization of the *qpostbayes* function, essential for the calculation of the minimum significant difference Δ. Algorithms 3 and 4 show the parameters *Syb* and *Yb*, respectively, in R code. In addition, the vector *nrt* used to the calculation of *Yb* and *Syb* is presented in Algorithm 5, in R code.

Algorithm 4 - Fragments of the R code developed for the parameter *Yb*

```
for (i in 1:nlevels(file$trt))
  {
    mean[i]=0
    sum=0
    for (l in 1:nlevels(file$trt))
    {
    if (file$trt[marker] == file$trt[l])
      {
      if (is.na(file$y[l]))
      {
      }
      else
      {
       sum=sum+file$y[l]
      }
      counter = counter+1
    }
  }
  mean[i] = sum/nrt[i]
  }
```

Source: The authors.

Ramos, C.H.O.[1]; Araújo, F.A.[2]; Andrade, P.C.R.[3]

Algorithm 5 - Fragments of the R code developed for the vector *nrt*

```
for (i in 1:nlevels(file$trt))
  {
    nrt[i]=0
    for (l in marker:length(file$trt))
    {
      if (file$trt[marker] == file$trt[l])
      {
        counter = conter+1
      }
    }
    for (i in 1:length(file$y))
    {
      if (is.na(file$y[i]))
      {
        for (l in 1:nlevels(file$trt))
        {
          if (l = =1)
          {
            if (i<=pos[l])
            {
              nrt[l]=nrt[l]-1
            }
            else
            {
              if ((i>pos[l-1]) && (i<=pos[l]))
              {
                nrt[l]=nrt[l]-1
              }
            }
          }
        }
      }
    }
  }
```

Source: The authors.

For the *Bayes* function $(N, alpha, file)$, the following parameters were used: $N = 10\ 000$, $\alpha = 0.05$ and the data file adapted from Stell & Torrie (1980). Table 3 presents the analysis of the performance of tests, with the results of the present code and those verified by Andrade, Rocha and Silva (2017).

Table 3 – Comparison between the results of the optimized code and that presented by Andrade, Rocha and Silva (2017).

Tests	Old code		Optimized code	
	dbayes	*pbayes*	*dbayes*	*pbayes*
2-1 = = 0	*	*	*	*
3-1 = = 0	ns	*	ns	ns
4-1 = = 0	*	*	*	*
5-1 = = 0	ns	*	ns	ns
3-2 = = 0	ns	*	ns	ns
4-2 = = 0	ns	ns	ns	ns
5-2 = = 0	ns	*	ns	ns
4-3 = = 0	*	*	*	*
5-3 = = 0	ns	ns	ns	ns
5-4 = = 0	ns	*	ns	ns

Source: The authors.

It turns out that after optimizing the code, the *dbayes* tests showed the same sensitivity. However, the results are different. Although it did not directly modify the *dbayes* function, there were differences in the results due to the modification of the *Syb* and *Yb* parameters related to the *delta* calculation. In relation to the *pbayes* test, it identified a smaller number of differences, differing from the results found by Andrade, Rocha and Silva (2017) due to the corrections of the acceptance intervals. However, the current results are in agreement with those presented by Andrade and Ferreira (2010). It is worth noting that the variances are not homogeneous, being a restriction to the use of conventional tests.

Table 4 shows the output, in R, of the code optimized for the data of the experiment presented by Box, Hunter and Hunter (2005). In this table it is shown: the tests performed, their p-value and whether the assumptions were fulfilled or not. The Bartlett test (BARTLETT, 1937), Shapiro-Wilk test (SHAPIRO; WILK, 1965) and Durbin-Watson test (DURBIN; WATSON, 1950) were used to verify the assumptions of homogeneity of variances, normality and independence of residues, respectively. It is observed that all the assumptions were having been attended, so the traditional tests are valid. Table 5 shows the comparison of the *dbayes* and *pbayes* tests with the traditional tests: Tukey, *SK* and *CCF*.

Table 4 – Table of validity of the tests for second study.

Validity of Tests			
	Tests	pvalue_Tests	Results_Tests
Normality	Shapiro	0.48184	Attended
Indep. of the erros	Dubin-watson	0.34431	Attended
homoce-dasticity	Bartlett	0.50755	Attended

Conclusion: The analysis of variance is valid, since all the assumptions have been attended

Source: The authors.

Table 5 – Comparison of *dbayes* and *pbayes* tests with other tests for the set of data presented by Box, Hunter and Hunter (2005).

Comparisons	Tests				
	Tukey	CCF	SK	*dbayes*	*pbayes*
A-B	*	ns	ns	ns	ns
C-A	ns	ns	ns	*	ns
D-A	*	*	*	ns	ns
C-B	*	ns	*	ns	ns
D-B	ns	ns	ns	*	ns
D-C	ns	ns	ns	*	*

Source: The authors.

It was observed that *dbayes* test detected a greater amount of differences between the treatments regarding to *pbayes*, *CCF* and *SK* tests and the same sensitivity as the Tukey test. And note that the *pbayes* test had the same sensitivity as the CCF test. These results are relevant since they show the good sensitivity of the tests, mainly the *dbayes*, for experiments with balanced data and assumption of the analysis of variance attended.

For the second experiment provided by Westfall, Randall and Russell (2011), the pbayes and dbayes tests were compared with the *Tukey-Kramer* test. The results are shown in Table 6.

It is verified that the *dbayes* test presents better result that the *Tukey-Kramer* and *pbayes* tests, showing that there are differences between the means of treatments analyzed. Therefore, the *dbayes* test performs well with data that contains unbalanced data. It is observed that the *pbayes* test did not detect significant differences between treatments b3 e b2, evidencing a lower sensibility in relation to the *Tukey-Kramer* test.

Table 6 – Comparison of the *dbayes* and *pbayes* tests with the *Tukey-Kramer* test for the dataset presented by Westfall, Randall and Russell (2011).

Comparisons	Tests		
	Tukey-Kramer	*dbayes*	*pbayes*
b1-b0	ns	ns	ns
b2-b0	*	*	*
b3-b0	ns	ns	ns
b2-b1	ns	*	ns
b3-b1	ns	ns	ns
b3-b2	*	*	ns

Source: The authors.

Finally, Table 7 shows the results of the *pbayes* and *dbayes* tests for the third experiment, in which the data were provided by Milliken and Johnson (1943). These were also compared to *Tukey-Kramer* test.

Table 7 – Comparison of the *dbayes* and *pbayes* tests with the *Tukey-Kramer* test for the data set presented by Milliken and Johnson (1943).

Treatments	Tests		
	Tukey-Kramer	*dbayes*	*pbayes*
VaccineA–VaccineB	*	*	*
VaccineA–VaccineC	ns	ns	ns
VaccineA–VaccineD	*	*	*
VaccineA–VaccineE	*	*	*
VaccineA–VaccineF	*	*	*
VaccineA–VaccineG	ns	ns	ns
VaccineB–VaccineC	*	*	*
VaccineB–VaccineD	ns	ns	ns
VaccineB–VaccineE	ns	ns	ns
VaccineB–VaccineF	ns	ns	ns
VaccineB–VaccineG	*	*	ns
VaccineC–VaccineD	*	*	*
VaccineC–VaccineE	*	*	*
VaccineC–VaccineF	*	*	*
VaccineC–VaccineG	ns	ns	ns
VaccineD–VaccineE	ns	ns	ns
VaccineD–VaccineF	ns	ns	ns
VaccineD–VaccineG	*	*	*
VaccineE–VaccineF	ns	ns	ns
VaccineE–VaccineG	*	*	ns
VaccineF–VaccineG	*	*	*

Source: The authors.

It is observed that *pbayes* test presented an identical result to the *Tukey-Kramer* result, evidencing a good sensibility in the detection of data with a large number of treatments and with unbalanced data. However, the *pbayes* test identified a smaller number of differences than the other tests, presenting a lower sensitivity.

Conclusion

The optimizations of the *Bayes* function were performed and implemented successfully in R. Positive results were obtained in relation to the output to the user and the reliability of the tests was increased. In addition, there was the expansion of the *Bayes* function, allowing the performance of the *pbayes* and *dbayes* tests for unbalanced data.

The optimized *Bayes* function allowed the *dbayes* and *pbayes* tests to have satisfactory results for the case where traditional tests are valid. In addition, in specific cases, as in the second experiment (unbalanced data with a small number of treatments) the *dbayes* test presents better results than the *Tukey-Kramer* test. It is worth mentioning that the Bayes function is very important in statistical analysis since the *pbayes* and *dbayes* tests are valid with or without the assumptions required by the analysis of variance for both balanced and unbalanced data. Some open questions that should be addressed in future work are the implementations of the function of randomized block design and the extension of these ideas to other experimental arrangements.

References

ANDRADE, P. C. R.; FERREIRA, D. F. Comparações múltiplas bayesianas em modelos normais homocedásticos e heterocedásticos. *Ciência e Agrotecnologia*, Lavras, v. 34, n. 4, p. 845-852, jul./ago. 2010.

ANDRADE, P. C. R.; ROCHA, L. H. C.; SILVA, M. M. Bayesian multiple comparisons procedures for CRD in R. *International Journal of Probability and Statistics*, Toronto, v. 6, n. 3, p. 45-50. 2017.

BARTLETT, M. S. Properties of sufficiency and statistical tests. *Proceedings of the Royal Statistical Society - Serie A*, London, v. 60, p. 268-282, 1937.

BERRY, D. A. 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*. Oxford: Oxford University Press, 1988. v. 3, p. 79-94.

BERRY, D. A.; HOCHBERG, Y. Bayesian perspectives on multiple comparisons. *Journal of Statistical Planning and Inference,* Amsterdam, v. 82, p. 215-227, 1999..

BOX, G. E. P.; HUNTER, J. S.; HUNTER, W. G. *Statistics for experimenters*: Design, innovation and Discovery. [S. l.]: Wiley, 2005. v. 2, p. 133.

BRATCHER, T.; HAMILTON, C. A Bayesian multiple comparison procedure for ranking the means of normally distributed data. *Journal of Statistical Planning and Inference*, Amsterdam, v. 133, p. 23-32, 2005.

BRETZ, F.; HOTHORN, T.; WESTFALL, P. *Multiple comparisons using R*. Boca Raton: Chapman & Hall/CRC Press, 2010.

CALINSKI, T.; CORSTEN, L. C. A. Clustering means in ANOVA by Simultaneous Testing. *Biometrics*, Washington, v. 41, p. 39-48, 1985.

CHEN, S. Y.; LEE, S. H. Multiple comparison procedures under heteroscedasticity. *Tamkang Journal of Science and Engineering*, Taiwan, v. 14, n. 4, p. 293-302, 2011.

DEMIRHAN, H.; DOLGUN, N. A.; DEMIRHAN, P.; DOLGUN, M. O. Performance of some multiple comparison tests under heteroscedasticity and dependency. *Journal of Statistical Computation and Simulation*, New York, v. 80, n. 10, p. 1083-1100, 2010.

DUNCAN, D. B. A Bayesian approach to multiple comparisons. *Technometrics*, Washington, v. 7, p. 171-222, 1965.

DUNNETT, C. W. Pairwise multiple comparisons in the unequal variance case. *Journal of the American Statistical Association*, New York, v. 75, n. 372, p. 796-800, 1980.

DURBIN, J.; WATSON, G. S. Testing for serial correlation in least squares regression I. *Biometrika*, London, v. 37, n. 3-4, p. 409-428, 1950.

FISHER, R. A. *Statistical methods for research workers*. 14. ed. New York: Hafner, 1973.

GAME, P. A.; HOWELL, J. F. Pairwise multiple comparison procedures with unequal n's and/or variances: a Monte-Carlo study. *Journal of Educational Statistics*, Washington, v. 1, p. 113-125, 1976.

GELMAN, A.; HILL, J.; YAJIMA, M. Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, Oakland, v. 5, n. 2, p. 189-211, 2012.

GOPALAN, R.; BERRY, D. A. Bayesian multiple comparisons using dirichlet process priors. *Journal of the American Statistical Association*, New York, v. 93, p. 1130-1139, 1998.

HINKELMANN, K.; KEMPTHORNE, O. *Design and analysis of experiments*. New York: Wiley & Sons, 1987. v. 1.

HOCHBERG, Y.; TAMHANE, A. C. *Multiple comparison procedure*. New York: Wiley, 1987.

HSU, J. C. *Multiple comparisons*: theory and methods. London: Chapman & Hall, 1996.

KEYES, T. K.; LEVY, M. S. Analysis of levene test under design imbalance. *Journal of Educational and Behavioral Statistics*, Washington, v. 22, p. 227-236, 1997.

KRAMER, C. Y. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics*, Washington, v. 12, p. 307-310, 1956.

LI, H. A multiple comparison procedure for populations with unequal variances. *Journal of Statistical Theory and Applications*, Paris, v. 11, n. 2, p. 165-181, 2012.

MILLIKEN, G. A.; JOHNSON, D. E. *Analysis of messy data*. 2nd ed. [*S. l.*]: A Chapman & Hall Book, 1943.

O'BRIEN, R. G. A robust technique for testing heterogeneity of variance effects in factorial design. *Psychometrika*, New York, v. 43, n. 3, p. 327-342, 1978.

R DEVELOPMENT CORE TEAM. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2019. Disponível em: http://www.R-project.org. Acesso em: 9 jan. 2019.

RAFTER, J.; ABELL, M.; BRASELTON, J. Multiple comparison methods for means. *Society for Industrial and Applied Mathematics Review*. Philadelphia, v. 44, n. 2, p. 259-278, 2002.

RAMSEY, P. H.; RAMSEY, P. P.; BARRERA, K. Choosing the best pairwise comparisons of means from nomnormal populations, with unequal variances, but equal sample sizes. *Journal of Statistical Computation and Simulation*, New York, v. 80, n. 5, p. 595-608, 2010.

RAMSEY, P. H.; BARRERA, K.; HACHIMINE-SEMPREBOM, P; LIU, C-C. Pairwise comparisons of means under realistic non-normality, unequal variances, outliers and equal sample sizes. *Journal of Statistical Computation and Simulation*, New York, v. 81, n. 2, p. 125-135, 2011.

SARMAH, S.; GOGOI, B. Multiple comparison procedures under equal and unequal population variances. *International Advanced Research Journal in Science Engineering and Technology*, India, v. 2, n. 12, p. 110-116, 2015.

SCOTT, A. J.; KNOTT, M. A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, Washington, v. 30, n. 3, p. 507-512, 1974.

SHAFFER, P. J. A semi-Bayesian study of Duncan's Bayesian multiple comparison procedure. *Journal of Statistical Planning and Inference*, Amsterdam, v. 82, p. 197-213, 1999.

SHAPIRO, S. S.; WILK, M. B. Analysis of variance test for normality (complete samples). *Biometrika*, Washington, v. 52, p. 591-611, 1965.

SHINGALA, M. C.; RAJYAGURU, A. Comparison of post hoc tests for unequal variance. *International Journal of New Technologies in Science and Engineering*, Delhi, v. 2, n. 5, p. 22-33, 2015.

STEEL, R. G. D.; TORRIE, J. H. Principles and procedures of statistics. 2nd ed. New York: McGraw-Hill Book, 1980.

TAMHANE, A. C. Multiple comparisons in model I one-way anova with unequal variances. *Communications in Statistics-Theory and Methods*, New York, v. A6, n. 1, p. 15-32, 1977.

TAMHANE, A. C. A Comparison of procedures for multiple comparisons of means with unequal variances. *Journal of the American Statistical Association*, New York, v. 74, p. 471-480, 1979.

TUKEY, J. Comparing individual means in the analysis of variance. *Biometrics*, Washington, v. 5, n. 2, p. 99-114, 1949.

WALLER, R. A.; DUNCAN, D. B. A Bayes rule for the symmetric multiple comparisons problem. *Journal of the American Statistical Association*, New York, v. 64, p. 1484-1503, 1969.

Ramos, C.H.O.[1]; Araújo, F.A. [2]; Andrade, P.C.R.[3]

WESTFALL, P. H.; RANDALL, D. T.; RUSSELL, D. W.
Multiple comparisons and multiple tests using SAS. 2nd.
ed. Cary, North Carolina, USA: SAS Institute INC., 2011.