



TESTE DE SCHEFFÉ PARA FATORIAIS TRIPLO COM INTERAÇÃO TRIPLA SIGNIFICATIVA

Beatriz de Oliveira Rodrigues¹, Felipe de Melo Taveira², Flávio Bittencourt³,
Adriana Dias⁴

¹ Graduanda em Ciência da Computação, Universidade Federal de Alfenas, a14006@bcc.unifal-mg.edu.br

² Graduado em Ciência da Computação, Universidade Federal de Alfenas,
felipe.melo.taveira@gmail.com

³ Doutor em Engenharia Agrícola, Departamento de Estatística, Universidade Federal de Alfenas,
flavio.bittencourt@unifal-mg.edu.br

⁴ Doutora em Estatística e Experimentação Agropecuária, Departamento de Estatística, Universidade Federal de Alfenas, adriana.dias@unifal-mg.edu.br

Resumo: A análise de variância é uma técnica estatística muito utilizada nos diferentes ramos do conhecimento e aplicada em diversas situações, sendo mais comum a comparação de dois ou mais tratamentos. Em um experimento com estrutura fatorial tripla, ou seja, cuja análise engloba três variáveis independentes, os tratamentos são formados pela combinação dos níveis de cada variável. Desta forma, um tratamento é formado pela combinação do nível i de uma variável combinado com o nível j da segunda variável combinado com o nível k da terceira variável. Quando a interação tripla é significativa e há interesse em testar algum contraste de médias de um fator dentro dos níveis dos outros fatores, o número de operações matemáticas é extenso, mas poderia ser facilitado com o uso de uma rotina computacional. Um dos testes que poderia ser utilizado para testar um contraste de médias de interesse é o teste de Scheffé, que apresenta algumas vantagens em relação aos demais testes, como a de poder comparar um contraste de médias (combinação linear de médias) e a possibilidade de definir as comparações de médias a serem realizadas após a análise do experimento. Este trabalho apresenta como objetivo a apresentação de uma rotina computacional em linguagem R para a execução do teste de Scheffé para testar contrastes de médias em experimentos em estrutura fatorial tripla com interação tripla significativa.

Palavras-chave: Médias; Contrastes; Desdobramento.

Área do Conhecimento: Ciências Exatas e da Terra.

1 INTRODUÇÃO

Em comparação com outros testes estatísticos, o teste de Scheffé apresenta um poder estatístico muito baixo quando utilizado para contrastes que envolvem apenas duas médias. Ainda assim, é um teste de grande importância na análise de dados, pois, ao contrário de outros testes, permite testar um contraste com mais de duas médias, e, neste caso o poder estatístico é maior.

Na análise de variância de um fatorial tripla a complexidade de cálculos manuais torna o teste de Scheffé demasiadamente trabalhoso, o que, na maioria das vezes, o inviabiliza. Porém, com o uso de programas computacionais que permitem programação de uma rotina, os cálculos tornam-se mais fáceis de serem realizados, e assim, o pesquisador poderá concentrar seus esforços apenas nas definições dos contrastes de interesse. Um programa computacional, ou mais especificamente, uma linguagem de programação, mundialmente conhecido e de distribuição livre que pode ser utilizado neste caso é o programa R (R CORE TEAM, 2017).

O programa R (R CORE TEAM, 2017) é de domínio público, compatível com os sistemas operacionais *Windows*, *Mac* e *Linux*, podendo ser facilmente instalado em quaisquer desses sistemas operacionais sem exigir muito conhecimento na área. Além disso, há um número significativo de pessoas investindo o seu tempo no desenvolvimento de “pacotes estatísticos”, a fim de facilitar a vida dos que necessitam realizar análises estatísticas por meio deste programa.

Dois pacotes, *agricolae* (MENDIBURU, 2015) e *asbio* (AHO, 2016), disponíveis em <https://www.r-project.org/>, realizam o teste de Scheffé para testar contrastes que abordam duas médias



em experimentos sem estrutura fatorial, ou seja, não permitem explorar a infinidade de possibilidades que o teste apresenta.

Desta forma o presente trabalho tem por objetivo apresentar uma rotina computacional em linguagem R para a execução do teste de Scheffé para testar contrastes de médias em experimentos em estrutura fatorial tripla com interação tripla significativa.

2 REFERENCIAL TEÓRICO

Uma análise de variância para experimentos em estrutura fatorial tripla (ou *three-way ANOVA*) é um método estatístico que permite determinar o efeito de três fatores, Fator A, Fator B e Fator C, os quais representam cada qual uma variável independente, sob uma variável resposta ou variável dependente. Cada fator apresenta categorias, também conhecidas como níveis, o Fator A pode ter i níveis, B, j níveis e C, k níveis de modo que o nível i do Fator A é representado como A_i , o nível j do Fator B como B_j e o nível k do Fator C como C_k e a combinação destes níveis dos diferentes fatores forma os tratamentos (VIEIRA, 1999).

Em muitas situações, após realizar a análise de variância, o pesquisador não tem interesse em comparar pares de médias, e sim comparar um grupo de médias, por exemplo, a média do tratamento 1 versus as médias dos tratamentos 2 e 3. Estas comparações são realizadas por meio de testes *post hoc* em que se verifica se o valor do contraste (uma combinação linear das médias observadas dos tratamentos) assume um valor menor ou maior/igual do que uma diferença mínima significativa (DMS) do teste utilizado. A maioria desses testes realiza comparações de pares de médias, isto é, só é possível comparar a média de um tratamento com a de outro tratamento. O teste de Scheffé é o único que permite comparar um contraste de médias de tratamentos que possui mais do que duas médias.

Autores como Quinn (2002), Storck *et al.* (2006), Gomes (2009) e Banzatto e Kronka (2013) ressaltam que o teste de Scheffé quando utilizado para comparação de duas médias é mais conservador do que o teste de Tukey, sendo, portanto, mais apropriado para comparação de grupos de médias. Um teste conservador é aquele que rejeita a hipótese de igualdade de médias com baixa probabilidade. Segundo Storck *et al.* (2006) e Banzatto e Kronka (2013), o teste de Scheffé permite a comparação de todo e qualquer contraste de médias, não exigindo ortogonalidade entre os contrastes estudados, sendo necessário apenas a significância do teste F na análise de variância.

Um contraste é definido como uma combinação linear das médias verdadeiras dos tratamentos do tipo $Y = \sum_{i=1}^k c_i \mu_i$, porém, experimentalmente, se trabalha com a estimativa de um contraste de médias amostrais dos tratamentos do tipo $\hat{Y} = \sum_{i=1}^k c_i m_i$, sendo c_i os coeficientes e m_i as médias observadas dos tratamentos, de modo que a soma algébrica dos coeficientes seja nula, isto é, $\sum_{i=1}^k c_i = 0$. Um algoritmo para a realização do teste de Scheffé pode ser encontrado em Banzatto e Kronka (2013). Inicia-se com a definição de um contraste qualquer de interesse do tipo $Y = \sum_{i=1}^k c_i \mu_i$; em seguida, calcula-se a estimativa deste contraste baseando-se nos valores das médias observadas, ou seja, $\hat{Y} = \sum_{i=1}^k c_i m_i$; calcula-se a estimativa da variância da estimativa do contraste adotando o quadrado médio do resíduo (QMRes) obtido na análise de variância, $\hat{V}(\hat{Y}) = QMRes \sum_{i=1}^k \frac{c_i^2}{n_i}$, em que n_i representa o número de repetições de cada tratamento, e por fim, calcula-se o valor da estatística de

teste, $S = \sqrt{(I - 1)\hat{V}(\hat{Y})F}$, em que I representa o número de tratamentos e F o valor do quantil da distribuição F conforme o nível de significância e graus de liberdade de tratamentos e do resíduo. Para a conclusão, ou seja, para verificar se o contraste entre as médias ou grupo de médias é significativo em um nível de probabilidade, deve-se observar se $|\hat{Y}| > S$.

Entretanto, para realizar um teste de Scheffé após uma análise de variância de um fatorial triplo é necessário modificar a estatística de teste, isto é, ampliar o algoritmo descrito acima. Isto porque uma análise de variância de um fatorial triplo apresenta as seguintes fontes de variação: (a) Fator A; (b) Fator B; (c) Fator C; (d) Int.AxB; (e) Int.AxC; (f) Int.BxC; (g) Int.AxBxC. Para cada fonte de variação, há uma hipótese nula a ser testada e de acordo com a não rejeição da hipótese nula um caminho deve ser adotado para completar a análise dos dados. Neste trabalho será considerada a situação em que a interação tripla é significativa, de modo que, de acordo com Gomes (2009) e Banzatto e Kronka (2013) é necessário o desdobramento da interação para se estudar os contrastes de médias de interesse, quais sejam, testar as médias dos níveis do Fator A dentro dos níveis dos Fatores B e C, testar as médias dos níveis do Fator B dentro dos níveis dos Fatores A e C e testar as médias dos níveis do Fator C dentro dos níveis dos Fatores A e B.



No programa R (R CORE TEAM, 2017) rotinas computacionais para análise de dados são fáceis de serem implementadas, isto porque o programa é uma linguagem de programação e um IDE (do inglês *Integrated Development Environment* ou Ambiente Integrado de Desenvolvimento) que é utilizado para cálculos estatísticos e para a confecção de gráficos, principalmente. Foi desenvolvido em 1993 por Ross Ihaka e Robert Gentleman do departamento de Estatística da Universidade de Auckland, Nova Zelândia que aliado a um IDE permite a manipulação de dados, geração de gráficos e a realização de cálculos variados. O R é facilmente extensível por meio de funções e extensões e a comunidade R é reconhecida pelos seus contributos ativos em termos de pacotes (bibliotecas de funções). Muitas das funções padrão do R são escritas no próprio R, o que torna fácil para os usuários seguir as escolhas algorítmicas desenvolvidas. Para tarefas computacionais intensivas, os códigos C, C++, e Fortran podem ser ligados e chamados durante a execução. Usuários experientes podem escrever código C ou Java11 para manipular diretamente objetos R. Como em muitas linguagens, o R suporta matrizes aritméticas e a estrutura de dados inclui escalares, vetores, matrizes, quadros de dados e listas. O Programa está disponível gratuitamente para instalação, podendo ser baixado em <https://www.r-project.org/> (R CORE TEAM, 2017).

3 MATERIAL E MÉTODOS

A implementação da rotina capaz de executar o teste de Scheffé para experimentos em estrutura fatorial tripla com interação tripla significativa foi baseada em rotinas computacionais apresentadas por Taveira *et al.* (2016), Taveira *et al.* (2017), Rodrigues *et al.* (2017) e outras rotinas disponíveis para o R. Além disso, foram construídas estatísticas de teste para serem utilizadas no contraste de interesse para se estudar o desdobramento dos níveis dos fatores considerando-se a interação tripla significativa.

4 RESULTADOS E DISCUSSÃO

Primeiramente serão apresentadas as estatísticas de teste elaboradas para se testarem contrastes considerando, respectivamente, os desdobramentos dos níveis do Fator A dentro dos níveis de B e C, dos níveis do Fator B dentro dos níveis de A e C e dos níveis do Fator C dentro dos níveis de A e B:

- A dentro de B_jC_k: $S = \sqrt{(I - 1)F_{tab} \frac{QMR}{L} \sum_i^I a_i^2}$
- B dentro de A_iC_k: $S = \sqrt{(J - 1)F_{tab} \frac{QMR}{L} \sum_i^J b_j^2}$
- C dentro de A_iB_j: $S = \sqrt{(K - 1)F_{tab} \frac{QMR}{L} \sum_i^K c_k^2}$

A execução da rotina elaborada permite realizar o teste de Scheffé para comparação de contrastes de interesse por meio da linha de comando:

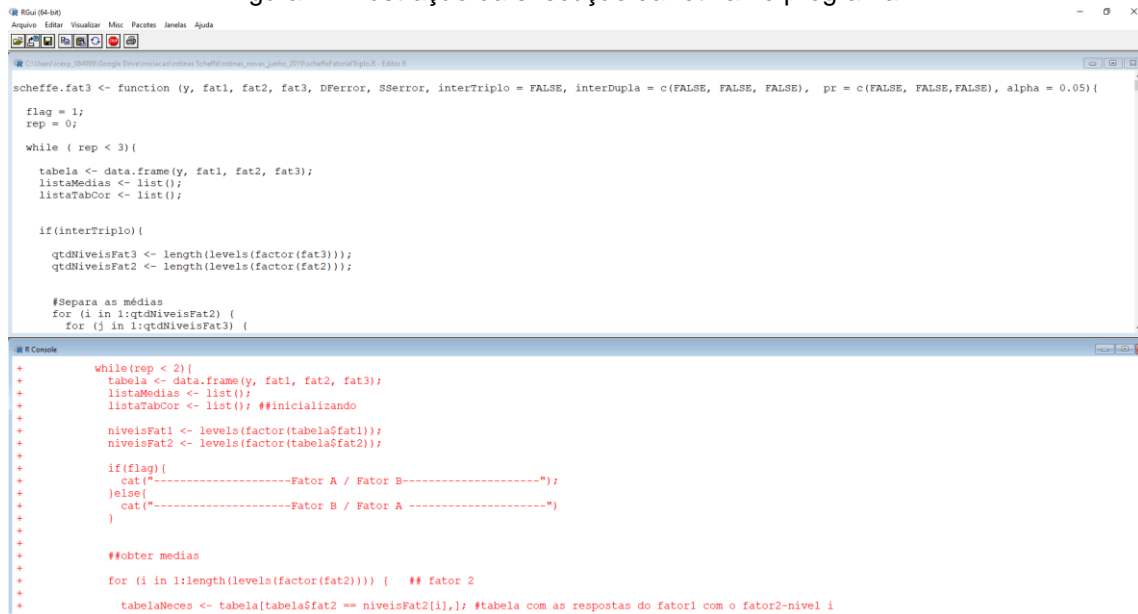
```
scheffe.fat3(y, fat1, fat2, fat3, DFerror, SSerror, interTriplo = FALSE,
interDupla = c(FALSE, FALSE, FALSE), pr = c(FALSE, FALSE, FALSE), alpha = 0.05)
```

Em que: *y* é o vetor que contém a variável resposta; *fat1*, *fat2* e *fat3* são os vetores das variáveis independentes; *DFerror* e *SSerror* são os graus de liberdade e a soma de quadrados do resíduo obtidos da análise de variância; *interTriplo*, *interDupla* e *pr* são os argumentos que devem ser modificados conforme a análise de variância, deixar *FALSE* quando não for significativo ou modificar por *TRUE* caso contrário; *alpha*: o nível de significância. O *default* da linha de comando considera inicialmente todas as fontes de variação como *FALSE*, as quais devem ser atualizadas conforme o resultado da análise de variância e, caso a interação tripla seja significativa, isto é *TRUE*, a rotina permite apenas a execução da interação tripla para diferentes contrastes de interesse. Além disso, o pesquisador pode observar os desdobramentos e as respectivas médias e o campo para declarar o contraste de interesse na tela *Data Editor*.

A Figura 1 representa a captura da tela durante a execução da rotina no programa R:



Figura 1 – Ilustração da execução da rotina no programa R



```
scheffe.fat3 <- function (y, fat1, fat2, fat3, DError, SError, interTriplo = FALSE, interDupla = c(FALSE, FALSE, FALSE), pr = c(FALSE, FALSE, FALSE), alpha = 0.05) {
  flag = 1;
  rep = 0;
  while ( rep < 3) {
    tabela <- data.frame(y, fat1, fat2, fat3);
    listaMedias <- list();
    listaTabCor <- list();

    if(interTriplo){
      qtdNiveisFat3 <- length(levels(factor(fat3)));
      qtdNiveisFat2 <- length(levels(factor(fat2)));

      #Separa as médias
      for (i in 1:qtdNiveisFat2) {
        for (j in 1:qtdNiveisFat3) {

```

```
+ while(rep < 2){
+   tabela <- data.frame(y, fat1, fat2, fat3);
+   listaMedias <- list();
+   listaTabCor <- list(); ##inicializando
+   niveisFat1 <- levels(factor(tabela$fat1));
+   niveisFat2 <- levels(factor(tabela$fat2));
+
+   if(flag){
+     cat("-----Fator A / Fator B-----");
+   } else{
+     cat("-----Fator B / Fator A -----")
+   }
+
+   ##obter medias
+   for (i in 1:length(levels(factor(fat2)))) { ## fator 2
+     tabelaNeces <- tabela[tabela$fat2 == niveisFat2[i],]; #tabela com as respostas do fator1 com o fator2-nivel i

```

Na tela capturada nota-se duas divisões, sendo a primeira a tela onde foi confeccionada a rotina e a segunda a tela que mostra a execução da rotina. A partir desse ponto faz-se a análise de variância dos dados que serão analisados e toma-se os valores dos graus de liberdade e soma de quadrados do resíduo. Considerando um conjunto de dados fictícios e fazendo-se a análise de variância dos dados obteve-se o seguinte resultado:

Figura 2 – Resultado da análise de variância de um conjunto de dados

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|----------------|----|---------|---------|---------|--------------|
| Bloco | 3 | 23002 | 7667 | 0.576 | 0.635019 |
| fat1 | 2 | 280979 | 140490 | 10.550 | 0.000287 *** |
| fat2 | 1 | 1810799 | 1810799 | 135.975 | 3.02e-13 *** |
| fat3 | 1 | 4709401 | 4709401 | 353.633 | < 2e-16 *** |
| fat1:fat2 | 2 | 101204 | 50602 | 3.800 | 0.032725 * |
| fat1:fat3 | 2 | 144534 | 72267 | 5.427 | 0.009172 ** |
| fat2:fat3 | 1 | 722016 | 722016 | 54.217 | 1.87e-08 *** |
| fat1:fat2:fat3 | 2 | 96633 | 48316 | 3.628 | 0.037647 * |
| Residuals | 33 | 439467 | 13317 | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |

Ao considerar um nível de significância igual a 5%, observa-se na Figura 2 que a interação tripla foi significativa ($p=0,03765$). Aplicando-se a rotina com o objetivo de se estudar algum contraste de interesse, abrem-se as seguintes telas no programa R, uma após a outra, conforme exibem as Figuras 3, 4 e 5.

Figura 3 – Desdobramento das médias dos níveis do Fator A dentro dos níveis dos Fatores B e C

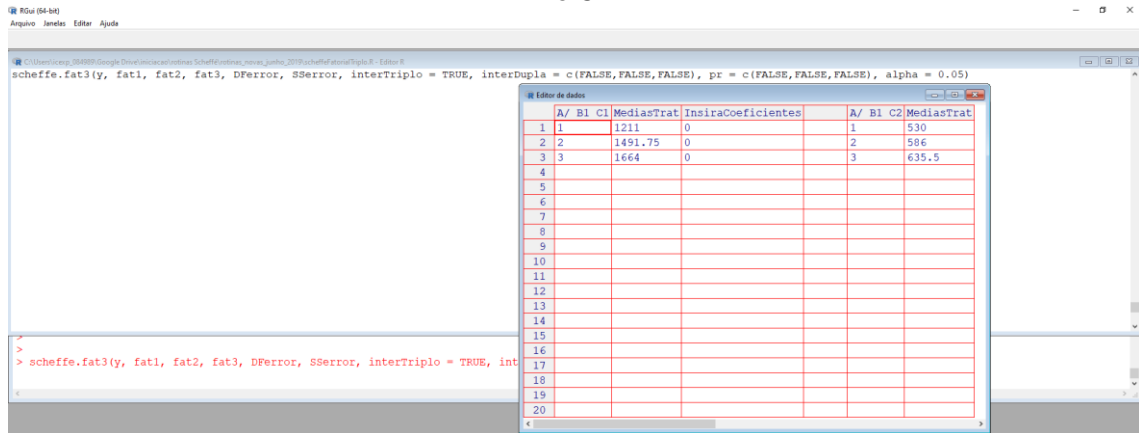


Figura 4 – Desdobramento das médias dos níveis do Fator B dentro dos níveis dos Fatores A e C

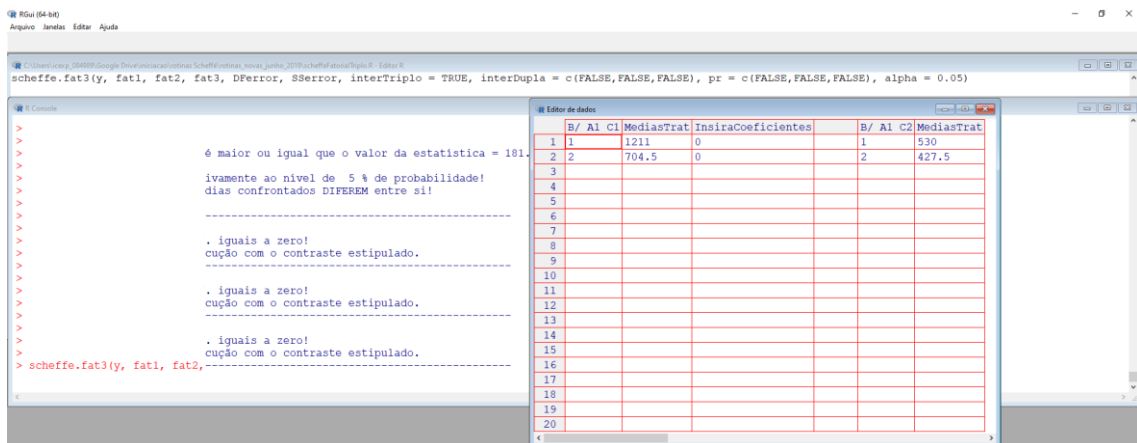
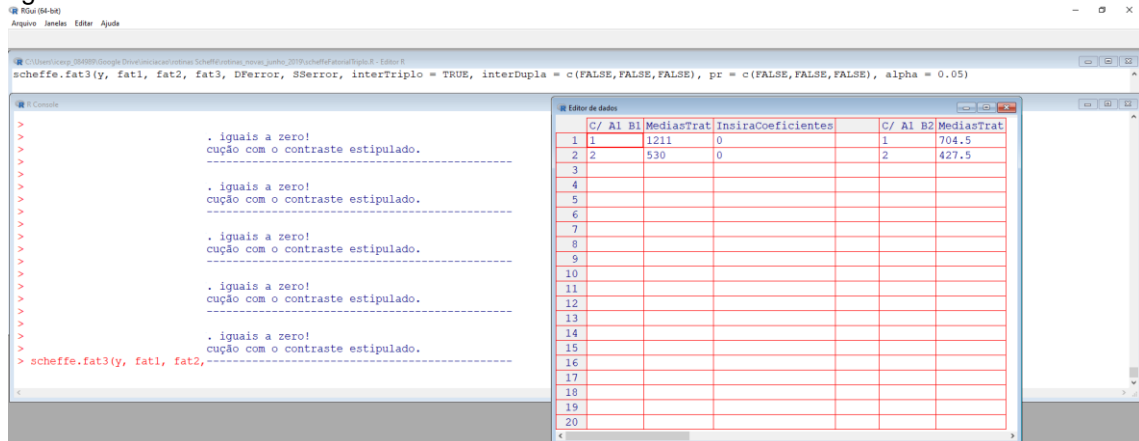
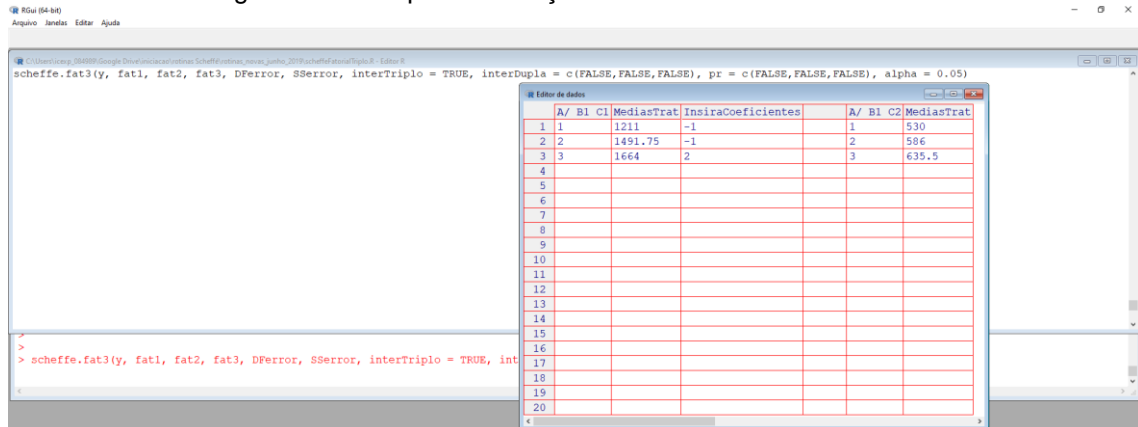


Figura 5 – Desdobramento das médias dos níveis do Fator C dentro dos níveis dos Fatores A e B



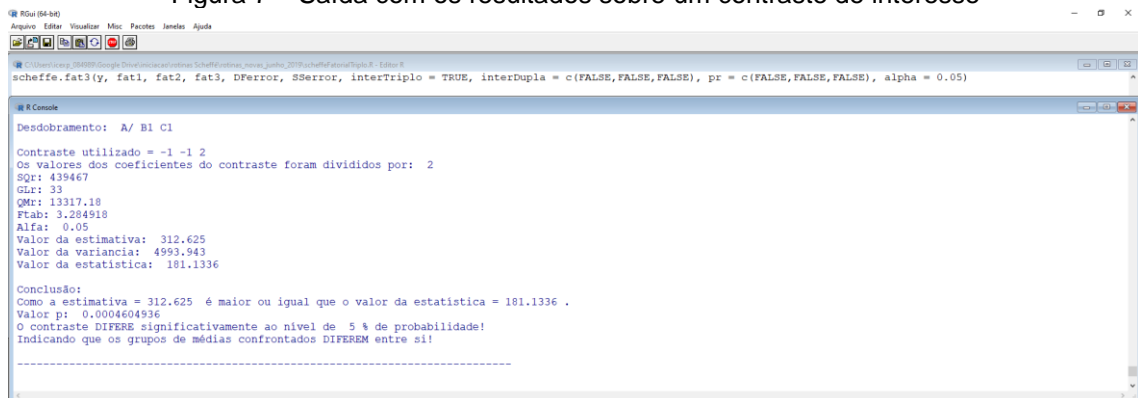
As Figuras 3, 4 e 5 representam a tela *Editor* do programa R que se abre ao executar a linha de comando para a inserção do(s) contraste(s) de interesse. Nota-se que a inserção do(s) contraste(s) de interesse poderá ser realizada em cada fator dentro dos níveis dos outros fatores, os quais podem ser digitados dentro de cada tela e que horizontalmente todos os desdobramentos irão aparecendo. A Figura 6 representa a declaração de um contraste de interesse para comparar a média obtida do nível 3 do Fator A dentro de B1C1 versus as médias dos níveis 1 e 2 do Fator A dentro de B1C1, isto é, a combinação linear $\hat{Y} = 2m_{(A3/B1C1)} - m_{(A1/B1C1)} - m_{(A2/B1C1)}$.

Figura 6 – Exemplo de inserção de um contraste de interesse



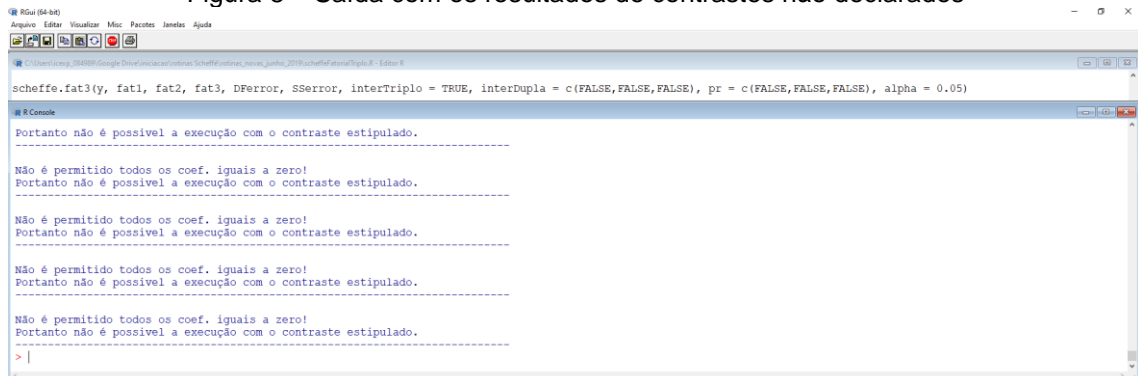
Observa-se que a inserção do contraste de interesse é feita na coluna *InsiraCoeficientes* e, embora não se tenha exemplificado, pode-se inserir valores fracionários neste campo. Demais contrastes de interesse podem ser realizados quantas vezes se desejar para os diferentes desdobramentos e para um mesmo desdobramento é necessário repetir a execução da linha de comando. Declarado um determinado contraste e fechada a janela *Editor* o programa realiza os cálculos e apresenta a seguinte saída, representada na Figura 7.

Figura 7 – Saída com os resultados sobre um contraste de interesse



A saída apresenta informações sobre o contraste realizado, exibindo-se em qual desdobramento foi realizado o teste de Scheffé, estatísticas de interesse e, principalmente, o valor-p do teste e a conclusão. Por outro lado, quando não se insere algum contraste no desdobramento das médias de algum fator, a mensagem declarada pela rotina está apresentada na Figura 8.

Figura 8 – Saída com os resultados de contrastes não declarados





Desta forma, quando os coeficientes não são modificados por valores válidos para um contraste a mensagem apresentada relata o fato.

5 CONCLUSÃO

A rotina executa facilmente o teste de Scheffé para se estudar as médias dos desdobramentos dos níveis de um fator dentro dos níveis de outros dois fatores quando a interação tripla é significativa. Após a finalização da rotina espera-se realizar o teste de Scheffé nas diferentes situações conforme o resultado da análise de variância.

6 REFERÊNCIAS

AHO, K. **asbio: A Collection of Statistical Tools for Biologists**. [S.l.], 2016. R package version 1.3-1. Disponível em: <http://CRAN.R-project.org/package=asbio>.

BANZATTO, D. A.; KRONKA, S. N. **Experimentação Agrícola**. 4. ed. Jaboticabal: FUNEP, 2013.

GOMES, F. P. **Curso de Estatística Experimental**. 15. ed. São Paulo: FEALQ, 2009.

MENDIBURU, F. de. **agricolae: Statistical Procedures for Agricultural Research**. [S.l.], 2015. R package version 1.2-3. Disponível em: <http://CRAN.R-project.org/package=agricolae>.

QUINN, G. P. **Experimental Design and data analysis for biologists**. Cambridge: Cambridge University Press, 2002.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2017. Disponível em: <http://www.R-project.org/>.

RODRIGUES, B. de O.; TAVEIRA, F. de M.; BITTENCOURT, F.; DIAS, A. Teste Scheffé no R para experimentos em parcela subdividida. In: XV ENCONTRO MINEIRO DE ESTATÍSTICA. **Anais...** Belo Horizonte: UFMG, 2017. p. 118.

STORCK, L.; GARCIA, D. C.; LOPES, S. J.; ESTEFANEL, V. **Experimentação vegetal**. 2. ed. Santa Maria: Ed. UFSM, 2006.

TAVEIRA, F. de M.; BITTENCOURT, F.; DIAS, A.; NOGUEIRA, D. A. Uma rotina em R para o teste Scheffé envolvendo três ou mais médias. In: XIV ENCONTRO MINEIRO DE ESTATÍSTICA. **Anais...** Varginha: UNIFAL-MG, 2016. p. 170.

TAVEIRA, F. de M.; RODRIGUES, B. de O.; BITTENCOURT, F.; DIAS, A. Teste Scheffé em fatorial duplo cruzado: uma rotina em R. In: XV ENCONTRO MINEIRO DE ESTATÍSTICA. **Anais...** Belo Horizonte: UFMG, 2017. p. 77.

VIEIRA, S. **Estatística Experimental**. 2. ed. São Paulo: Editora Atlas, 1999.