

**Universidade de São Paulo  
Escola Superior de Agricultura ‘Luiz de Queiroz’**

**Um estudo sobre testes de comparação de médias e sua aplicação  
condicional a um teste F global significativo na análise de variância**

**Josiane Rodrigues**

Tese apresentada para obtenção do título de Doutora em  
Ciências. Área de concentração: Estatística e  
Experimentação Agronômica

**Piracicaba  
2015**

Josiane Rodrigues  
Licenciada em Matemática

**Um estudo sobre testes de comparação de médias e sua aplicação condicional a um teste  
F global significativo na análise de variância**  
versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientadora:  
Profa. Dra. **SÔNIA MARIA DE STEFANO PIEDADE**

Tese apresentada para obtenção do título de Doutora em  
Ciências. Área de concentração: Estatística e  
Experimentação Agronômica

**Piracicaba**  
**2015**

**Dados Internacionais de Catalogação na Publicação  
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Rodrigues, Josiane

Um estudo sobre testes de comparação de médias e sua aplicação condicional a um teste F global significativo na análise de variância / Josiane Rodrigues. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2015.

164 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. Simulação 2. Monte Carlo 3. Nível de significância por comparação 4. Nível de significância por experimento 5. Poder I. Título

CDD 519.535  
R696

**“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”**

## DEDICATÓRIA

**A Deus.**

*Aos meus pais,  
**Antonio e Sonia**, pelo amor e por tudo que fizeram  
pela minha formação, pessoal e profissional.*

*Aos meus avós,  
**Antonio (in memorian), Dalva, Gilda e Oswaldo**,  
pelo carinho e exemplo de vida.*

*A minha irmã,  
**Cristiane**, pela eterna amizade e incentivo.*

*Ao meu namorado,  
**Emanuel**, pelo cuidado, dedicação e paciência.*



## AGRADECIMENTOS

Agradeço primeiramente a Deus, pela minha vida e saúde.

Aos meus pais Antonio e Sonia, por todo amor e dedicação ao longo de todos esses anos.

Aos meus avós Oswaldo, Antonio (*in memorian*), Gilda e Dalva, por todo carinho e exemplo de vida.

À minha irmã Cristiane, pela amizade e incentivo.

Ao meu querido Emanuel, pela dedicação, incentivo, carinho e paciência. Obrigada por estar comigo em todos os momentos.

À Profa. Dra. Sônia Maria De Stefano Piedade, pela amizade e por toda orientação ao longo deste trabalho, que contribuíram para o meu crescimento e formação acadêmica.

Ao Prof. Dr. Idemauro Antonio Rodrigues de Lara, pela amizade e por todas as sugestões e colaboração dispensadas ao trabalho.

Ao Prof. Dr. Antonio Carlos Simões Pião e ao Prof. Dr. José Silvio Govone, por todas as contribuições e sugestões realizadas, que contribuíram para o enriquecimento da pesquisa.

Aos professores do Programa de Pós-Graduação em Estatística e Experimentação Agronômica, pela formação.

Aos funcionários do Departamento de Ciências Exatas da ESALQ/USP, pela amizade e auxílio permanente.

A todos os meus amigos da ESALQ, pelo auxílio e companheirismo.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro durante a pesquisa.

A todas as pessoas que direta ou indiretamente contribuíram para a realização deste trabalho.



*“Comece fazendo o que é necessário,  
depois o que é possível, e de repente você  
estará fazendo o impossível.”*

*São Francisco de Assis*





## SUMÁRIO

RESUMO .....	11
ABSTRACT .....	13
1 INTRODUÇÃO .....	15
2 REVISÃO BIBLIOGRÁFICA.....	19
2.1 Experimentos com um fator: a análise de variância.....	19
2.1.1 Análise de variância no delineamento inteiramente casualizado .....	20
2.1.2 Análise de variância no delineamento casualizado em blocos .....	27
2.1.3 Análise de variância no delineamento quadrado latino .....	32
2.1.4 Dados desbalanceados.....	35
2.1.5 Pressuposições da análise de variância: a análise de resíduos .....	36
2.2 Testes de comparação de médias .....	45
2.2.1 Taxas de erro nos testes de comparação de médias.....	47
2.2.2 Teste DMS de Fisher.....	49
2.2.3 Teste de Tukey .....	50
2.2.4 Teste de Duncan.....	51
2.2.5 Teste de Student-Newman-Keuls.....	52
2.2.6 Teste de Dunnett.....	53
2.2.7 Teste de Scheffé.....	54
2.2.8 Teste <i>t</i> (Contrastes ortogonais).....	54
2.2.9 Teste de Bonferroni.....	56
2.2.10 Escolha do procedimento adequado.....	57
2.3 Região de inferência do teste <i>F</i> global versus região de inferência do teste de Tukey .....	58
3 MATERIAL E MÉTODOS.....	63
3.1 Comportamento dos testes de comparação de médias em relação as situações de divergência.....	63
3.2 Níveis de significância por comparação e por experimento.....	64
3.3 Poder .....	66
4 RESULTADOS.....	69
4.1 Experimentos com resultados contraditórios .....	69
4.2 Comportamento dos testes de comparação de médias em relação as situações de divergência.....	74
4.3 Níveis de significância por comparação e por experimento.....	78
4.4 Poder .....	86
5 DISCUSSÃO.....	91
6 CONSIDERAÇÕES FINAIS.....	95
REFERÊNCIAS .....	97
BIBLIOGRAFIA CONSULTADA.....	103
APÊNDICES .....	105
ANEXOS.....	119



## RESUMO

### **Um estudo sobre testes de comparação de médias e sua aplicação condicional a um teste $F$ global significativo na análise de variância**

Neste trabalho, por meio de experimentos simulados no delineamento inteiramente casualizado pelo método de Monte Carlo, inicialmente comparou-se o resultado do teste  $F$  global da análise de variância com o resultado de cada um dos seguintes procedimentos para a comparação de médias de tratamentos entre si duas a duas: Tukey, Duncan, DMS de Fisher, Student-Newman-Keuls e Scheffé. Ainda, os níveis de significância por comparação e por experimento, condicionais e incondicionais a um resultado preliminar significativo do teste  $F$  global, foram estimados para cada um dos testes de médias por meio do mesmo processo de simulação. Por fim, o poder condicional e incondicional de cada um dos procedimentos também foi estimado. Neste estudo notou-se a presença das duas situações contraditórias possíveis ao confrontar o resultado do teste  $F$  global com o resultado de cada um dos testes de médias analisados. No entanto, a presença dessas situações parece não estar ligada ao desvio de ao menos uma das pressuposições da análise. Notou-se, todavia, que testes que apresentaram maior incidência da situação de um teste  $F$  não significativo e ao menos um contraste entre médias significativo, apresentaram também estimativa para o nível de significância por experimento muito acima do nível nominal de 5% declarado. Além disso, testes com maior incidência da situação de um teste  $F$  global significativo e nenhum contraste entre médias significativo mostraram-se menos poderosos que os demais procedimentos avaliados. Verificou-se também que a aplicação condicional dos testes de comparações múltiplas a um resultado preliminar significativo do teste  $F$  global pode levar a um nível de significância real maior do que o nível nominal declarado. Essa aplicação condicional pode, ainda, levar a um pequeno aumento no poder dos testes em experimentos em que o número de tratamentos e a diferença real entre médias adjacentes são baixos.

Palavras-chave: Simulação; Monte Carlo; Nível de significância por comparação; Nível de significância por experimento; Poder



## ABSTRACT

### **A study of mean comparison tests and their conditional application to a significant overall $F$ -test on analysis of variance**

In this work, through experiments simulated in completely randomized design by Monte Carlo's method, initially the result of the overall  $F$ -test on analysis of variance was compared with the result of each one of the following multiple comparison procedures: Tukey, Duncan, Fisher's least significant difference, Student-Newman-Keuls and Scheffé. Furthermore, the comparisonwise and experimentwise error rates, conditional and unconditional to a significant result of the overall  $F$ -test, were estimated for each one of the tests analyzed by the same simulation process. Finally, the conditional and unconditional power of each procedure was also estimated. In this study it was noticed the presence of the two possible conflicting situations to confront the overall  $F$ -test result with the result of each multiple comparison test analyzed. However, the presence of these situations doesn't seem to be linked to deviation at least one of the assumptions of the analysis. It was noted, however, tests that had a higher incidence of the situation with not significant  $F$ -test and at least one significant contrast between means, also showed estimate for experimentwise error rate much bigger than the nominal level of 5% declared. In addition, tests with higher incidence of the situation with significant overall  $F$ -test and no significant contrast between means, were less powerful than the other evaluated procedures. It was also found that the conditional application of multiple comparison tests to a significant result of overall  $F$ -test can lead to real significance level greater than the nominal level declared. This conditional application can also lead to a small increase in the power of tests in experiments in which the number of treatments and the real difference between adjacent means are small.

Keywords: Simulation; Monte Carlo; Comparisonwise error tare; Experimentwise error rate; Power



# 1 INTRODUÇÃO

Um problema muito comum na área da ciência assim também como na área da indústria é a comparação das médias de alguns tratamentos de interesse, para determinar quais desses tratamentos diferem entre si, caso essa diferença exista (RAFTER; ABELL; BRASELTON, 2002). De acordo com Girardi, Cargnelutti Filho e Storck (2009), a maneira mais usual para tratar esse problema é a análise de variância. Desde que desenvolvida essa técnica tornou-se uma importante ferramenta de pesquisa estatística dos cientistas e seu uso dispersou-se, rapidamente, para outras ciências (DEMÉTRIO, 1978).

Sendo os tratamentos do ensaio de efeito fixo, o teste  $F$  global da análise de variância testa a hipótese de igualdade entre as médias populacionais dos tratamentos  $H_0: \mu_1 = \mu_2 = \dots = \mu_a$ , a determinado nível de significância. Caso o teste  $F$  seja significativo, sendo os tratamentos com mais de dois níveis e de natureza qualitativa, então a aplicação de testes de comparação de médias é feita, com o intuito de investigar eventuais diferenças entre pares de médias específicos ou combinações lineares dessas médias.

Conforme Hochberg e Tamhane (1987), os fundamentos dos testes de médias foram estabelecidos no final da década de 1940 e início da década de 1950, principalmente por David Duncan, Samarendra Nath Roy, Henry Scheffé e John Tukey, embora algumas das ideias sobre o assunto tenham aparecido muito mais cedo nos trabalhos de Ronald Aylmer Fisher, William Sealy Gosset (Student), dentre outros.

Segundo Nogueira (2007), os métodos de comparação de médias têm merecido a atenção de muitos pesquisadores e, embora a literatura sobre o assunto seja vasta, o campo ainda se encontra em aberto para futuras pesquisas. O que se tem hoje é um número elevado de procedimentos para a comparação de médias de tratamentos, cada um deles fundamentado num conjunto de suposições que os tornam específicos para determinados fins.

De acordo com um estudo realizado, o único método estatístico aplicado com mais frequência do que os testes de médias é o teste  $F$  global da análise de variância (MEAD; PIKE, 1975 apud HSU, 1996). No entanto, enquanto testes de comparação de médias ocupam o segundo lugar em frequência de utilização, eles talvez ocupem o primeiro lugar no que diz respeito a sua frequência de abusos (HSU, 1996). Nos levantamentos realizados por Petersen (1977), Cardellino e Siewerdt (1992), Santos, Moreira e Beltrão (1998), Bezerra Neto, Nunes e Negreiros (2002), Bertoldo et al. (2008a) e Bertoldo et al. (2008b), por exemplo, constatou-



se que muitas vezes esses testes são utilizados de forma indiscriminada, o que pode resultar em informações incompletas ou errôneas acerca dos tratamentos investigados.

Um dos dilemas envolvidos com os testes de médias é a sua aplicação condicional a um resultado preliminar significativo do teste  $F$  global na análise de variância (HSU, 1996). De acordo com Cardellino e Siewerdt (1992), a aplicação condicional dos testes a um resultado significativo do teste  $F$  é uma questão polêmica e responsável por dividir a opinião de muitos estudiosos, e que deve, portanto, ser melhor investigada.

No entanto, não é possível desvincular esse estudo dos possíveis erros que podem ser cometidos num teste de hipótese, afinal, ao analisar a hipótese para um contraste de médias, o teste, seja ele aplicado apenas mediante um resultado significativo do teste  $F$  global ou não, possui probabilidades para os erros tipo I e II.

Entretanto, segundo Girardi, Cargnelutti Filho e Storck (2009), há diferentes formas de avaliar a taxa de erro tipo I para um procedimento de comparação de médias, tais como o erro tipo I por comparação e o erro tipo I por experimento, por exemplo, o que acaba dificultando a avaliação do desempenho desses procedimentos de uma maneira geral. Se experimentos simulados sem o efeito de tratamento são submetidos a testes de comparações múltiplas das médias (comparações duas a duas), a taxa de erro por comparação pode ser estimada por meio da razão entre o número de inferências errôneas (quantidade de vezes que o teste não aceitou  $H_0: \mu_i = \mu_{i'}$ ) e o número total de inferências realizadas. Por sua vez, a taxa de erro por experimento pode ser estimada por meio da razão entre o número de experimentos com no mínimo uma inferência errônea e o número total de experimentos simulados (GIRARDI; CARGNELUTTI FILHO; STORCK, 2009).

Outro aspecto que dificulta a avaliação desses procedimentos está relacionado ao poder dos testes, uma vez que existem diferentes configurações para as médias populacionais dos tratamentos comparados, como citado por Einot e Gabriel (1975), Rafter, Abell e Braselton (2002) e Girardi, Cargnelutti Filho e Storck (2009). A primeira delas, conhecida como configuração da amplitude mínima, é o equivalente a separar as médias em dois grupos: as primeiras das médias iguais entre si, mas diferente da segunda metade das médias, também iguais entre si ( $\mu_1 = \mu_2 = \dots = \mu_{a/2} < \mu_{(a/2)+1} = \dots = \mu_{a-1} = \mu_a$ ). Em outra configuração das médias, a da amplitude máxima, a primeira das médias é a menor delas, a última média é a maior, e as médias intermediárias são superiores a menor, inferiores a maior, e não diferem entre si ( $\mu_1 < \mu_2 = \mu_3 = \dots = \mu_{a-2} = \mu_{a-1} < \mu_a$ ). Por fim, a configuração igualmente espaçada atribui para cada média um incremento fixo em relação a média anterior ( $\mu_1 < \mu_2 < \mu_3 < \dots < \mu_{a-2} < \mu_{a-1} < \mu_a$ ). Segundo Girardi, Cargnelutti Filho e Storck

(2009), a configuração igualmente espaçada é ideal para avaliar o poder dos testes, uma vez que as médias, duas a duas, apresentam uma diferença que é testada por meio do procedimento em avaliação.

Diversos trabalhos têm sido realizados para avaliar procedimentos de comparação de médias no que diz respeito ao controle das taxas de erro tipo I e/ou também quanto ao poder do teste, como pode ser verificado em Boardman e Moffitt (1971), Carmer e Swanson (1971, 1973), Thomas (1974), Bernhardson (1975), Einot e Gabriel (1975), Perecin e Barbosa (1988), Silva, Ferreira e Bearzoti (1999), Conagin e Gomes (2004), Conagin e Barbin (2006), Conagin, Barbin e Demétrio (2008), Girardi, Cargnelutti Filho e Storck (2009) e Sousa, Lira Junior e Ferreira (2012), por exemplo. Todavia, o estudo desses conceitos aliado a um resultado preliminar significativo do teste  $F$  global ou não carece de estudos adicionais.

Pretende-se nesse trabalho, a partir da análise de experimentos com um fator simulados no delineamento inteiramente casualizado pelo método de Monte Carlo:

- i) Avaliar os testes Tukey, Duncan, DMS de Fisher, Student-Newman-Keuls e Scheffé, cada um deles para a comparação de médias de tratamentos entre si duas a duas, no que diz respeito a sua concordância com o resultado do teste  $F$  global, assim também como verificar se eventuais contradições podem estar associadas ao desvio de ao menos uma das pressuposições da análise, em cenários formados por combinações entre o número de tratamentos, o número de repetições e o coeficiente de variação dos ensaios;
- ii) Estudar e comparar cada um dos testes citados no que se refere ao controle das taxas de erro tipo I condicionais e incondicionais a um resultado significativo do teste  $F$  global na análise de variância, nos mesmos cenários simulados em que os resultados dos testes foram confrontados com o resultado do teste  $F$ ;
- iii) Avaliar e comparar cada um dos procedimentos de comparação de médias quanto ao seu poder, condicional e incondicional a um resultado significativo do teste  $F$ , em cenários simulados formados pela combinação entre o número de tratamentos e o número de erro padrão da média.



## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 Experimentos com um fator: a análise de variância

Um problema comumente encontrado em diversas áreas da ciência é a comparação dos efeitos de alguns tratamentos de interesse, para determinar quais desses produzem resultados médios diferentes entre si (GIRARDI; CARGNELUTTI FILHO; STORCK, 2009).

Rafter, Abell e Braselton (2002) citam um exemplo em que um fabricante deseja examinar o efeito de diferentes embalagens sobre a venda de um determinado produto. Segundo os autores, uma forma razoável de proceder nesse caso é selecionar um grupo de lojas com volume de vendas comparáveis e, de forma aleatória e independente, fazer com que cada uma das lojas trabalhe com um dos tipos de embalagem a serem testados. Ainda, é assumido que outras eventuais condições que possam vir a afetar as vendas do produto, tais como preço, espaço na prateleira, esforços promocionais, dentre outros, são as mesmas em todas as lojas consideradas na pesquisa.

Quando a fase da coleta de dados terminar, não é possível apenas observar as médias de vendas do produto, obtidas com cada uma das embalagens, e afirmar com clareza se elas são ou não iguais entre si, afinal, as semelhanças (ou diferenças) das médias tanto podem ser explicadas pelos tratamentos serem de fato iguais (ou diferentes) entre si quanto aos seus efeitos médios, assim também como por simples acaso, ou seja, devido a variações inerentes a cada uma das lojas e que não foram controladas durante a pesquisa.

Um método comum para a investigação dessa questão é a análise de variância, procedimento este que compara a variação devida aos tratamentos com a variação devida ao acaso ou resíduo. Embora seja um método que exige o cálculo de variâncias, na realidade ele compara as médias dos tratamentos testados no experimento quando estes são de efeito fixo.

Cabe aqui uma breve discussão sobre efeitos fixos e aleatórios. Num experimento, o experimentador pode escolher especificamente alguns tratamentos de seu interesse para que sejam comparados entre si. Nesse caso, as conclusões obtidas por ele estarão restritas aos tratamentos avaliados na pesquisa, ou seja, não podem ser estendidas a tratamentos similares que não foram diretamente considerados na análise. Diz-se então que os tratamentos são de efeito fixo, e o principal objetivo da análise de variância é a comparação das médias dos tratamentos, bem como a estimação dos parâmetros do modelo estatístico em questão.

Por outro lado, os tratamentos da pesquisa podem ser uma amostra aleatória de uma população de tratamentos. Nesse caso, as conclusões obtidas no ensaio, adquiridas com base na

amostra considerada, podem ser então estendidas para todos os tratamentos da população, mesmo que eles não tenham sido diretamente especificados na análise. Os tratamentos então são considerados de efeito aleatório, e o principal objetivo da análise de variância é a estimação dos componentes de variância do modelo.

O teste básico utilizado para comparar estimativas de variâncias na análise de variância foi proposto por Ronald Aylmer Fisher, e é conhecido como teste  $z$  de Fisher. Atualmente, o teste foi substituído pelo seu equivalente teste  $F$  de Snedecor que, segundo Hsu (1996), é o procedimento estatístico mais utilizado por pesquisadores em suas análises de dados.

A seguir será explanado de que forma proceder a análise de variância de experimentos com um único fator em cada um dos seguintes delineamentos experimentais: inteiramente casualizado, casualizado em blocos e quadrado latino.

### 2.1.1 Análise de variância no delineamento inteiramente casualizado

Quando se dispõe de unidades experimentais tão similares quanto possível para a realização de um experimento, tais como um terreno uniforme ou uma ninhada de animais homogêneos, por exemplo, os tratamentos devem ser designados às unidades experimentais por puro e simples sorteio. A esse delineamento experimental dá-se o nome de inteiramente casualizado.

Seja um experimento com um fator realizado segundo esse delineamento em que  $a$  tratamentos foram aplicados um mesmo número  $r$  de vezes, somando um total de  $n$  observações. Os dados obtidos no ensaio podem ser apresentados como na Tabela 2.1.

Tabela 2.1 – Esquema de um experimento com um fator realizado no delineamento inteiramente casualizado

Tratamento	Repetição			
	1	2	...	$r$
1	$y_{11}$	$y_{12}$	...	$y_{1r}$
2	$y_{21}$	$y_{22}$	...	$y_{2r}$
...	...	...	...	...
$a$	$y_{a1}$	$y_{a2}$	...	$y_{ar}$

O modelo de médias referente a esse experimento é escrito da seguinte maneira:

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad (2.1)$$

com  $i = 1, \dots, a$  e  $j = 1, \dots, r$ , em que  $y_{ij}$  é a resposta do  $i$ -ésimo tratamento em sua  $j$ -ésima repetição,  $\mu_i$  é a média populacional referente ao  $i$ -ésimo tratamento, e  $\varepsilon_{ij}$  é o erro aleatório, que incorpora todas as fontes de variabilidade que não foram controladas durante o ensaio. O erro é suposto normal e independentemente distribuído com média zero e variância  $\sigma^2$ .

Uma maneira alternativa de escrever um modelo para os dados é definir

$$\mu_i = \mu + \tau_i, \quad (2.2)$$

de modo que a eq. (2.1) possa ser reescrita da seguinte forma:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}. \quad (2.3)$$

Nesta forma do modelo,  $\mu$  é o parâmetro comum a todos os tratamentos, de efeito fixo e chamado de média geral, e  $\tau_i$  é o parâmetro referente ao efeito do  $i$ -ésimo tratamento. A eq. (2.3) é usualmente conhecida como modelo de efeitos.

Este último modelo descreve duas situações diferentes no que diz respeito aos efeitos de tratamentos. Num primeiro caso, o parâmetro  $\tau_i$  é de efeito fixo, ou seja, o experimentador escolheu especificamente alguns tratamentos de seu interesse para serem testados no experimento. O modelo estatístico também é dito fixo, e o intuito da análise de variância é comparar as médias dos tratamentos entre si, bem como estimar os parâmetros do modelo, ou seja,  $\mu, \tau_i$  e  $\sigma^2$ .

Num segundo caso, o parâmetro  $\tau_i$  é de efeito aleatório, ou seja, os tratamentos do ensaio representam uma amostra oriunda de uma população de tratamentos. O modelo estatístico é considerado aleatório e o objetivo da análise de variância é testar hipóteses sobre a variabilidade desses efeitos, bem como estimar essa variação.

No presente estudo trabalhar-se-á com a análise de variância no caso de tratamentos de efeito fixo apenas e, portanto, o objetivo será testar a hipótese de igualdade entre as médias populacionais dos tratamentos comparados. Para testar essa hipótese, assume-se, como já colocado, que os erros do modelo seguem uma distribuição normal de média zero e variância  $\sigma^2$ , e essas condições implicam que

$$y_{ij} \sim N(\mu + \tau_i, \sigma^2),$$

e as observações são mutuamente independentes entre si.

A hipótese de interesse, também conhecida como hipótese de homogeneidade, supõe que não há o efeito de tratamentos. Essa hipótese, para o modelo de médias, pode ser representada da seguinte maneira:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_a. \quad (2.4)$$

No modelo de efeitos, decompõe-se a média populacional  $\mu_i$ , referente ao  $i$ -ésimo tratamento, em dois componentes, como em (2.2) e, de um modo geral, costuma-se pensar em  $\mu$  como uma média total das observações de modo que

$$\frac{\sum_i \mu_i}{a} = \mu,$$

e esta definição implica que

$$\sum_i \tau_i = 0.$$

Assim sendo, os efeitos de tratamentos podem ser pensados como desvios da média geral. Consequentemente, uma maneira alternativa de escrever a hipótese de homogeneidade, agora em termos desses efeitos, é da forma:

$$H_0: \tau_1 = \tau_2 = \dots = \tau_a = 0. \quad (2.5)$$

Para testar a hipótese de igualdade entre as médias populacionais dos tratamentos, seja como em (2.4) ou (2.5), faz-se necessário realizar a análise de variância dos dados obtidos no ensaio. A ideia do procedimento é decompor a variabilidade total das observações entre todas as causas de variação do experimento.

A soma de quadrados total corrigida é usada como uma medida de variabilidade geral das observações. Ela é definida da seguinte maneira:

$$SQ_{Total} = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2,$$

em que  $\bar{y}_{..}$  é a média geral dos dados obtidos no ensaio. De fato, dividindo-se a soma de quadrados total corrigida pelo número de graus de liberdade das observações, ou seja,  $(n - 1)$ , obtém-se a variância amostral dos dados, a qual representa uma medida padrão de variabilidade.

A soma de quadrados total corrigida pode, ainda, ser escrita da seguinte forma:

$$SQ_{Total} = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = \sum_i \sum_j [(\bar{y}_i - \bar{y}_{..}) + (y_{ij} - \bar{y}_i)]^2,$$

em que  $\bar{y}_i$  representa a média amostral do  $i$ -ésimo tratamento, ou seja,

$$SQTotal = r \sum_i (\bar{y}_i - \bar{y}_{..})^2 + \sum_i \sum_j (y_{ij} - \bar{y}_i)^2 + 2 \sum_i \sum_j (\bar{y}_i - \bar{y}_{..})(y_{ij} - \bar{y}_i). \quad (2.6)$$

No entanto, o produto cruzado da eq. (2.6) é nulo, afinal,

$$\begin{aligned} \sum_i \sum_j (\bar{y}_i - \bar{y}_{..})(y_{ij} - \bar{y}_i) &= \sum_i (\bar{y}_i - \bar{y}_{..}) \sum_j (y_{ij} - \bar{y}_i) = \sum_i (\bar{y}_i - \bar{y}_{..})(y_i - r\bar{y}_i) \\ &= \sum_i (\bar{y}_i - \bar{y}_{..}) \left( y_i - r \frac{y_i}{r} \right) = 0, \end{aligned} \quad (2.7)$$

sendo  $y_i$  o total associado ao  $i$ -ésimo tratamento e, portanto,

$$SQTotal = r \sum_i (\bar{y}_i - \bar{y}_{..})^2 + \sum_i \sum_j (y_{ij} - \bar{y}_i)^2. \quad (2.8)$$

A eq. (2.8) indica que a variabilidade total dos dados, tal como medido pela soma de quadrados total corrigida, pode ser repartida em uma soma dos quadrados das diferenças entre as médias dos tratamentos e a média geral, mais uma soma dos quadrados das diferenças das observações dentro de tratamentos a partir da média do tratamento.

As diferenças entre as médias dos tratamentos observados e a média geral é uma medida das diferenças entre as médias dos tratamentos, ao passo que as diferenças das observações dentro de um tratamento a partir da média desse tratamento podem ser devidas apenas ao erro aleatório (MONTGOMERY, 2001). Assim, pode-se escrever a eq. (2.8) simbolicamente como

$$SQTotal = SQTrat + SQRes,$$

em que  $SQTrat$  representa a soma de quadrados de tratamentos e  $SQRes$  a soma de quadrados do resíduo. Como tem-se um total de  $n$  observações, há  $(n - 1)$  graus de liberdade associados ao total. Os tratamentos e o resíduo, por sua vez, possuem  $(a - 1)$  e  $(n - a)$  graus de liberdade associados, respectivamente.

As quantias

$$QMTrat = \frac{SQTrat}{a - 1}$$

e

$$QMRes = \frac{SQRes}{n - a}$$



representam os quadrados médios de tratamentos e do resíduo, respectivamente. O valor esperado do quadrado médio de tratamentos é obtido a partir dos seguintes cálculos:

$$\begin{aligned}
E(QMTrat) &= E \left[ \frac{r \sum_i (\bar{y}_i - \bar{y}_{..})^2}{a-1} \right] = \frac{r}{a-1} E \left[ \sum_i (\bar{y}_i - \bar{y}_{..})^2 \right] \\
&= \frac{r}{a-1} E \left[ \sum_i (\bar{y}_i^2 - 2\bar{y}_i \bar{y}_{..} + \bar{y}_{..}^2) \right] = \frac{r}{a-1} E \left( \sum_i \bar{y}_i^2 - 2\bar{y}_{..} \sum_i \bar{y}_i + a\bar{y}_{..}^2 \right) \\
&= \frac{r}{a-1} E \left( \sum_i \bar{y}_i^2 - 2a\bar{y}_{..}^2 + a\bar{y}_{..}^2 \right) = \frac{r}{a-1} E \left( \sum_i \bar{y}_i^2 - a\bar{y}_{..}^2 \right) \\
&= \frac{r}{a-1} \left\{ E \left[ \sum_i \left( \frac{y_i}{r} \right)^2 \right] - E \left[ a \left( \frac{y_{..}}{ar} \right)^2 \right] \right\} = \frac{r}{a-1} \left[ \frac{1}{r^2} E \left( \sum_i y_i^2 \right) - \frac{1}{ar^2} E(y_{..}^2) \right] \\
&= \frac{r}{a-1} \left[ \frac{1}{r^2} \sum_i E(y_i^2) - \frac{1}{ar^2} E(y_{..}^2) \right] \\
&= \frac{r}{a-1} \left\{ \frac{1}{r^2} \sum_i E \left[ \left( r\mu + r\tau_i + \sum_j \varepsilon_{ij} \right)^2 \right] \right. \\
&\quad \left. - \frac{1}{ar^2} E \left[ \sum_i \sum_j (\mu + \tau_i + \varepsilon_{ij}) \right]^2 \right\} \\
&= \frac{r}{a-1} \left\{ \frac{1}{r^2} \sum_i E \left[ r^2\mu^2 + r^2\tau_i^2 + \left( \sum_j \varepsilon_{ij} \right)^2 + dp \right] \right. \\
&\quad \left. - \frac{1}{ar^2} E \left( ar\mu + r \sum_i \tau_i + \sum_i \sum_j \varepsilon_{ij} \right)^2 \right\} \\
&= \frac{r}{a-1} \left\{ \frac{1}{r^2} \sum_i (r^2\mu^2 + r^2\tau_i^2 + r\sigma^2) \right. \\
&\quad \left. - \frac{1}{ar^2} E \left[ a^2r^2\mu^2 + \left( \sum_i \sum_j \varepsilon_{ij} \right)^2 + 2ar\mu \sum_i \sum_j \varepsilon_{ij} \right] \right\} \\
&= \frac{r}{a-1} \left( a\mu^2 + \sum_i \tau_i^2 + \frac{a\sigma^2}{r} - a\mu^2 - \frac{\sigma^2}{r} \right) = \sigma^2 + \frac{r \sum_i \tau_i^2}{a-1},
\end{aligned}$$

em que  $y_{..}$  representa a soma de todas as observações do ensaio e  $dp$  os duplos produtos do produto notável em questão.

De forma similar é possível obter o valor esperado do quadrado médio do resíduo:

$$\begin{aligned}
E(QMRes) &= E \left[ \frac{\sum_i \sum_j (y_{ij} - \bar{y}_i)^2}{n - a} \right] = \frac{1}{n - a} E \left[ \sum_i \sum_j (y_{ij} - \bar{y}_i)^2 \right] \\
&= \frac{1}{n - a} E \left[ \sum_i \sum_j (y_{ij}^2 - 2y_{ij}\bar{y}_i + \bar{y}_i^2) \right] \\
&= \frac{1}{n - a} E \left( \sum_i \sum_j y_{ij}^2 - 2r \sum_i \bar{y}_i^2 + r \sum_i \bar{y}_i^2 \right) \\
&= \frac{1}{n - a} E \left( \sum_i \sum_j y_{ij}^2 - \frac{1}{r} \sum_i y_i^2 \right) \\
&= \frac{1}{n - a} \left[ E \left( \sum_i \sum_j y_{ij}^2 \right) - \frac{1}{r} E \left( \sum_i y_i^2 \right) \right] \\
&= \frac{1}{n - a} \left[ \sum_i \sum_j E(y_{ij}^2) - \frac{1}{r} \sum_i E(y_i^2) \right] \\
&= \frac{1}{n - a} \left\{ \sum_i \sum_j E[(\mu + \tau_i + \varepsilon_{ij})^2] - \frac{1}{r} \sum_i E \left[ \left( r\mu + r\tau_i + \sum_j \varepsilon_{ij} \right)^2 \right] \right\} \\
&= \frac{1}{n - a} \left\{ \sum_i \sum_j E(\mu^2 + \tau_i^2 + \varepsilon_{ij}^2 + dp) \right. \\
&\quad \left. - \frac{1}{r} \sum_i E \left[ r^2\mu^2 + r^2\tau_i^2 + \left( \sum_j \varepsilon_{ij} \right)^2 + dp \right] \right\} \\
&= \frac{1}{n - a} \left[ \sum_i \sum_j (\mu^2 + \tau_i^2 + \sigma^2) - \frac{1}{r} \sum_i (r^2\mu^2 + r^2\tau_i^2 + r\sigma^2) \right] \\
&= \frac{1}{n - a} \left[ ar\mu^2 + r \sum_i \tau_i^2 + ar\sigma^2 - \frac{1}{r} (ar^2\mu^2 + r^2 \sum_i \tau_i^2 + ar\sigma^2) \right] \\
&= \frac{1}{n - a} \left( ar\mu^2 + r \sum_i \tau_i^2 + ar\sigma^2 - ar\mu^2 - r \sum_i \tau_i^2 - a\sigma^2 \right) \\
&= \frac{1}{n - a} \sigma^2 (n - a) = \sigma^2.
\end{aligned}$$

Assim sendo, o quadrado médio do resíduo estima a variância  $\sigma^2$  e, se não existe diferença entre as médias dos tratamentos, o que implica  $\tau_i = 0, \forall i = 1, \dots, a$ , então o quadrado médio de tratamentos também estima essa variância. No entanto, se as médias dos tratamentos diferem entre si, o valor esperado do quadrado médio de tratamentos é maior que  $\sigma^2$ . Logo, um teste de hipótese comparando  $QMTrat$  e  $QMRes$  pode ser realizado para verificar se existem ou não evidências de diferença significativa entre as médias populacionais dos tratamentos comparados.

A  $SQTotal$  é uma soma de quadrados em variáveis aleatórias normalmente distribuídas. Por conseguinte, pode-se mostrar que  $SQTotal/\sigma^2$  segue uma distribuição qui-quadrado com  $(n - 1)$  graus de liberdade. Além disso, pode-se demonstrar que  $SQRes/\sigma^2$  também segue uma distribuição qui-quadrado, agora com  $(n - a)$  graus de liberdade e, por fim,  $SQTrat/\sigma^2$  segue uma qui-quadrado com  $(a - 1)$  graus de liberdade se a hipótese de homogeneidade é verdadeira. O Teorema 2.1 é útil para estabelecer a independência entre  $SQTrat/\sigma^2$  e  $SQRes/\sigma^2$ .

Teorema 2.1 (Teorema de Cochran) – Sejam  $Z_1, Z_2, \dots, Z_\theta$  variáveis aleatórias independentes e identicamente distribuídas com distribuição normal padrão e seja

$$\sum_{i=1}^{\theta} Z_i^2 = Q_1 + Q_2 + \dots + Q_g,$$

em que  $g \leq \theta$  e  $Q_i$  possui  $\theta_i$  graus de liberdade ( $i = 1, \dots, g$ ). Então  $Q_1, Q_2, \dots, Q_g$  são variáveis aleatórias independentes que seguem distribuição qui-quadrado com  $\theta_1, \theta_2, \dots, \theta_g$  graus de liberdade, respectivamente, se e somente se

$$\theta = \theta_1 + \theta_2 + \dots + \theta_g.$$

Como os graus de liberdade de tratamentos mais os do resíduo somam  $(n - 1)$ , o qual representa o número de graus de liberdade das observações, então o teorema de Cochran implica que  $SQTrat/\sigma^2$  e  $SQRes/\sigma^2$  são variáveis aleatórias independentes que seguem distribuição qui-quadrado. Além do mais, se a hipótese de igualdade entre as médias populacionais dos tratamentos é verdadeira, a razão

$$F_0 = \frac{SQTrat/\sigma^2 / a - 1}{SQRes/\sigma^2 / n - a} = \frac{QMTrat}{QMRes} \quad (2.9)$$

possui distribuição  $F$  de Snedecor com  $(a - 1)$  e  $(n - a)$  graus de liberdade. A expressão em (2.9) representa o teste estatístico para a hipótese de homogeneidade como em (2.4) ou (2.5).

Sob a hipótese alternativa, de que ao menos um contraste entre médias populacionais de tratamentos difere de zero, o valor esperado do numerador da estatística em (2.9) é maior que o valor esperado do denominador. Logo, há evidências para não aceitar a hipótese de igualdade entre as médias populacionais dos tratamentos comparados se

$$F_0 > F_{(\alpha; a-1, n-a)},$$

em que  $F_{(\alpha; a-1, n-a)}$  é o valor tabelado de  $F$  de Snedecor ao nível de significância  $\alpha$ , segundo  $(a - 1)$  e  $(n - a)$  graus de liberdade.

De forma mais simplificada, pode-se escrever a soma de quadrados total corrigida e a soma de quadrados de tratamentos como

$$SQTotal = \sum_i \sum_j y_{ij}^2 - \frac{y_{..}^2}{n}$$

e

$$SQTrat = \frac{\sum_i y_{i.}^2}{r} - \frac{y_{..}^2}{n},$$

respectivamente. A soma de quadrados referente ao resíduo é, então, obtida pela seguinte diferença:

$$SQRes = SQTotal - SQTrat.$$

Todo o procedimento para a análise de variância pode ser resumido como na Tabela 2.2.

Tabela 2.2 – Análise de variância de um experimento com um fator realizado no delineamento inteiramente casualizado

Causas de variação	Graus de liberdade	Somas de quadrados	Quadrados médios	$F$
Tratamentos	$a - 1$	$SQTrat = \frac{\sum_i y_{i.}^2}{r} - \frac{y_{..}^2}{n}$	$\frac{SQTrat}{a - 1}$	$F_0 = \frac{QMTrat}{QMRes}$
Resíduo	$n - a$	$SQRes = SQTotal - SQTrat$	$\frac{SQRes}{n - a}$	
Total	$n - 1$	$SQTotal = \sum_i \sum_j y_{ij}^2 - \frac{y_{..}^2}{n}$		

### 2.1.2 Análise de variância no delineamento casualizado em blocos

Quando o pesquisador dispõe de unidades experimentais homogêneas para comparar os tratamentos de seu interesse, deve sortear os tratamentos às unidades sem quaisquer restrições, realizando, portanto, um experimento no delineamento inteiramente casualizado. No entanto, essa situação nem sempre é encontrada na prática. Muitas vezes o pesquisador dispõe de pequenos grupos de unidades homogêneas, mas nenhum desses grupos tem unidades em número suficiente para realizar o experimento. Faz então um experimento no delineamento

casualizado em blocos, em que todos os tratamentos são sorteados a cada um dos blocos, de modo que cada tratamento apareça em cada bloco uma única vez.

Para ilustrar esse tipo de delineamento pode-se citar, por exemplo, um experimento cujo interesse seja estudar o efeito de diferentes rações no ganho de peso em suínos, e, para tanto, dispõe-se de porcos de mesma raça mas de pesos diferentes. Para aplicar os tratamentos, o pesquisador deve então escolher para cada bloco porcos de mesmo peso e, dentro de cada bloco, os tratamentos devem ser sorteados às unidades experimentais inteiramente ao acaso.

Para entender a lógica da análise de variância num experimento com um fator realizado segundo o delineamento casualizado em blocos, considerar-se-á um ensaio em que a área experimental foi dividida em  $b$  blocos para a comparação de  $a$  tratamentos de interesse, totalizando  $n$  observações. O quadro do experimento pode ser visualizado na Tabela 2.3.

Tabela 2.3 – Esquema de um experimento com um fator realizado no delineamento casualizado em blocos

Tratamento	Bloco			
	1	2	...	$b$
1	$y_{11}$	$y_{12}$	...	$y_{1b}$
2	$y_{21}$	$y_{22}$	...	$y_{2b}$
...	...	...	...	...
$a$	$y_{a1}$	$y_{a2}$	...	$y_{ab}$

O modelo estatístico mais tradicional referente a esse experimento é o modelo de efeitos, representado da seguinte forma:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}, \quad (2.10)$$

com  $i = 1, \dots, a$  e  $j = 1, \dots, b$ , em que  $y_{ij}$  é a resposta do  $i$ -ésimo tratamento no  $j$ -ésimo bloco,  $\mu$  é a média geral, de efeito fixo,  $\tau_i$  é o efeito do  $i$ -ésimo tratamento,  $\beta_j$  é o efeito do  $j$ -ésimo bloco e  $\varepsilon_{ij}$  é o erro experimental, suposto aleatório, normal e independentemente distribuído com média zero e variância  $\sigma^2$ .

Inicialmente considerar-se-á tratamentos e blocos de efeito fixo. Consequentemente, os efeitos de tratamentos e blocos serão considerados como desvios da média geral, de modo que

$$\sum_i \tau_i = \sum_j \beta_j = 0.$$

Sendo os tratamentos de efeito fixo, o interesse está em testar a igualdade das médias populacionais dos tratamentos comparados. Assim, para o caso do modelo de efeitos em (2.10), a hipótese de interesse pode ser representada como em (2.5).

Também é possível usar o modelo de médias para descrever um experimento realizado segundo esse delineamento, da seguinte maneira:

$$y_{ij} = \mu_{ij} + \varepsilon_{ij}, \quad (2.11)$$

com  $i = 1, \dots, a$  e  $j = 1, \dots, b$ , em que

$$\mu_{ij} = \mu + \tau_i + \beta_j.$$

Ainda, como a média do  $i$ -ésimo tratamento

$$\mu_i = \frac{1}{r} \sum_j (\mu + \tau_i + \beta_j) = \mu + \tau_i,$$

uma maneira alternativa para descrever a hipótese de interesse, agora em termos das médias de tratamentos, é como em (2.4).

A soma de quadrados total corrigida pode ser expressa como:

$$\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = \sum_i \sum_j [(\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})]^2, \quad (2.12)$$

em que  $\bar{y}_{..}$ ,  $\bar{y}_{i.}$  e  $\bar{y}_{.j}$  representam as médias geral, do  $i$ -ésimo tratamento e do  $j$ -ésimo bloco, respectivamente.

Expandindo o lado direito da eq. (2.12) obtém-se

$$\begin{aligned} \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 &= b \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 + a \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2 \\ &+ \sum_i \sum_j (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 + 2 \sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..})(\bar{y}_{.j} - \bar{y}_{..}) \\ &\quad + 2 \sum_i \sum_j (\bar{y}_{.j} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}) \\ &\quad + 2 \sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}). \end{aligned} \quad (2.13)$$

É possível provar que os três produtos cruzados da eq. (2.13) são nulos, como foi feito na eq. (2.7). Logo,

$$\begin{aligned} \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 &= b \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 + a \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2 \\ &\quad + \sum_i \sum_j (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 \end{aligned} \quad (2.14)$$

representa uma partição da soma de quadrados total corrigida. Expressando a eq. (2.14) simbolicamente tem-se que

$$SQTotal = SQTrat + SQBlocos + SQRes. \quad (2.15)$$

Como há um total de  $n$  observações, a soma de quadrados total possui  $(n - 1)$  graus de liberdade. Existem ainda  $a$  tratamentos e  $b$  blocos e, portanto, as somas de quadrados de tratamentos e de blocos possuem  $(a - 1)$  e  $(b - 1)$  graus de liberdade, respectivamente. A soma de quadrados do resíduo, por sua vez, pode ser dada pela diferença entre a soma de quadrados total e as somas de quadrados de tratamentos e de blocos. Logo, ela possui  $ab - 1 - (a - 1) - (b - 1) = (a - 1)(b - 1)$  graus de liberdade.

Além do mais, como os graus de liberdade do lado direito da eq. (2.15) somam igual aos graus de liberdade do lado esquerdo da equação, fazendo as suposições usuais sobre os erros do modelo, o Teorema 2.1 garante que  $SQTrat/\sigma^2$ ,  $SQBlocos/\sigma^2$  e  $SQRes/\sigma^2$  são variáveis aleatórias independentes com distribuição qui-quadrado.

É possível demonstrar que os valores esperados dos quadrados médios de cada um dos componentes da análise de variância, sendo tratamentos e blocos de efeito fixo, são

$$E(QMTrat) = \sigma^2 + \frac{b \sum_i \tau_i^2}{a - 1},$$

$$E(QMBlocos) = \sigma^2 + \frac{a \sum_j \beta_j^2}{b - 1}$$

e

$$E(QMRes) = \sigma^2.$$

Para testar a hipótese de igualdade entre as médias populacionais dos tratamentos comparados, faz-se uso da estatística

$$F_0 = \frac{QMTrat}{QMRes}, \quad (2.16)$$

que possui distribuição  $F$  de Snedecor com  $(a - 1)$  e  $(a - 1)(b - 1)$  graus de liberdade, se a hipótese de homogeneidade é verdadeira. A região crítica é representada pela cauda superior da distribuição  $F$  e, portanto, há evidências para não aceitar a hipótese nula se

$$F_0 > F_{(\alpha; (a-1), (a-1)(b-1))},$$

em que  $F_{(\alpha; (a-1), (a-1)(b-1))}$  é o valor tabelado de  $F$  de Snedecor ao nível de significância  $\alpha$ , segundo  $(a - 1)$  e  $(a - 1)(b - 1)$  graus de liberdade.

De forma mais simples, pode-se ainda escrever a soma de quadrados total corrigida, a soma de quadrados de blocos e a soma de quadrados de tratamentos como

$$SQ_{Total} = \sum_i \sum_j y_{ij}^2 - \frac{y_{..}^2}{n},$$

$$SQ_{Blocos} = \frac{\sum_j y_{.j}^2}{a} - \frac{y_{..}^2}{n},$$

e

$$SQ_{Trat} = \frac{\sum_i y_{i.}^2}{b} - \frac{y_{..}^2}{n},$$

em que  $y_{.j}$  e  $y_{i.}$  representam a soma de todas as observações do experimento, do  $j$ -ésimo bloco e do  $i$ -ésimo tratamento, respectivamente. A soma de quadrados referente ao resíduo, por sua vez, pode ser obtida pela diferença

$$SQ_{Res} = SQ_{Total} - SQ_{Blocos} - SQ_{Trat}.$$

Todo o procedimento para a análise de variância do experimento pode ser resumido como na Tabela 2.4.

Tabela 2.4 – Análise de variância de um experimento com um fator realizado no delineamento casualizado em blocos

Causas de variação	Graus de liberdade	Somas de quadrados	Quadrados médios	$F$
Blocos	$b - 1$	$SQ_{Blocos} = \frac{\sum_j y_{.j}^2}{a} - \frac{y_{..}^2}{n}$	$\frac{SQ_{Blocos}}{b - 1}$	$F_0$
Tratamentos	$a - 1$	$SQ_{Trat} = \frac{\sum_i y_{i.}^2}{b} - \frac{y_{..}^2}{n}$	$\frac{SQ_{Trat}}{a - 1}$	
Resíduo	$(a - 1)(b - 1)$	$SQ_{Res}$ (por diferença)	$\frac{SQ_{Res}}{(a - 1)(b - 1)}$	
Total	$n - 1$	$SQ_{Total} = \sum_i \sum_j y_{ij}^2 - \frac{y_{..}^2}{n}$		

Embora o procedimento para a análise de variância tenha sido realizado considerando blocos de efeito fixo, o mesmo processo pode ser realizado se estes são considerados de efeito aleatório, o que é muitas vezes o caso. A diferença é que o valor esperado do quadrado médio de tratamentos fica da forma



$$E(QMBlocos) = \sigma^2 + a\sigma_B^2,$$

em que  $\sigma_B^2$  é o componente de variância do efeito de blocos. Ainda assim, o valor esperado do quadrado médio de tratamentos é livre de qualquer efeito de blocos, e o teste para a hipótese de homogeneidade é feito da maneira usual.

### 2.1.3 Análise de variância no delineamento quadrado latino

O delineamento quadrado latino é usado para eliminar duas fontes de variabilidade, isto é, sistematicamente permite a construção de blocos de duas maneiras distintas ou, em outras palavras, em duas direções: uns constituindo as chamadas linhas e outros as chamadas colunas. Como exemplo, pode-se citar um experimento em que o pesquisador deseja comparar três tipos de rações no ganho de peso em suínos e, para tanto, dispõe de animais de diferentes raças e pesos. Para realizar o experimento, o pesquisador pode então organizar blocos de animais de mesma raça (em coluna, por exemplo) e também blocos de animais de mesmo peso (em linha).

Uma vez construídos os quadrados latinos, faz-se necessário sortear os tratamentos, de modo que todos os tratamentos apareçam em cada uma das linhas e também em cada uma das colunas uma única vez. Assim sendo, linhas, colunas e tratamentos devem ser iguais em número, o que torna os experimentos em quadrado latino não muito comuns na prática, devido as restrições desse delineamento.

Para compreender a lógica da análise de variância num experimento com um fator realizado de acordo com o delineamento quadrado latino, considerar-se-á um experimento em que a área experimental foi dividida em  $a$  linhas e  $a$  colunas, para a comparação de  $a$  tratamentos de interesse. O total de unidades experimentais do ensaio será representado pela letra  $n$ .

O modelo estatístico mais comum para representar um experimento realizado de acordo com o delineamento quadrado latino é o modelo de efeitos, representado da forma:

$$y_{ijk} = \mu + \tau_i + l_j + c_k + \varepsilon_{ijk}, \quad (2.17)$$

com  $i, j, k = 1, \dots, a$ , em que  $y_{ijk}$  é a resposta do  $i$ -ésimo tratamento aplicado à  $j$ -ésima linha e  $k$ -ésima coluna,  $\mu$  é a média geral, de efeito fixo,  $\tau_i$  é o efeito do  $i$ -ésimo tratamento,  $l_j$  é o efeito da  $j$ -ésima linha,  $c_k$  é o efeito da  $k$ -ésima coluna e  $\varepsilon_{ijk}$  é o erro experimental, suposto aleatório, normal e independentemente distribuído com média zero e variância  $\sigma^2$ .

Inicialmente considerar-se-á tratamentos, linhas e colunas como sendo de efeito fixo e, portanto, tem-se que

$$\sum_i \tau_i = \sum_j l_j = \sum_k c_k = 0,$$

uma vez que os mesmos são considerados como desvios da média geral.

Como tratamentos são considerados de efeito fixo, o objetivo principal da análise de variância é a comparação das médias de tratamentos entre si. Assim, para o caso do modelo de efeitos como em (2.17), a hipótese de interesse é dada como na expressão (2.5).

Também é possível utilizar o modelo de médias para descrever um experimento realizado segundo o delineamento quadrado latino, da seguinte forma:

$$y_{ijk} = \mu_{ijk} + \varepsilon_{ijk}, \quad (2.18)$$

com  $i, j, k = 1, \dots, a$ , em que

$$\mu_{ijk} = \mu + \tau_i + l_j + c_k.$$

Ainda, como a média do  $i$ -ésimo tratamento

$$\mu_i = \frac{1}{a} \sum_j \sum_k (\mu + \tau_i + l_j + c_k) = \mu + \tau_i,$$

outra maneira para descrever a hipótese de interesse, agora em termos das médias de tratamentos, é como na expressão (2.4).

A soma de quadrados total corrigida pode ser expressa como

$$\begin{aligned} \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{...})^2 &= \sum_i \sum_j \sum_k [(\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{..k} - \bar{y}_{...}) \\ &\quad + (y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}_{...})]^2, \end{aligned} \quad (2.19)$$

em que  $\bar{y}_{...}$ ,  $\bar{y}_{i..}$ ,  $\bar{y}_{.j.}$  e  $\bar{y}_{..k}$  representam as médias geral, do  $i$ -ésimo tratamento, da  $j$ -ésima linha e da  $k$ -ésima coluna, respectivamente.

Expandindo o lado direito da eq. (2.19) obtém-se

$$\begin{aligned} \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{...})^2 &= a \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 + a \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2 + a \sum_k (\bar{y}_{..k} - \bar{y}_{...})^2 \\ &\quad + \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}_{...})^2 + 2 \sum_i \sum_j \sum_k (\bar{y}_{i..} - \bar{y}_{...})(\bar{y}_{.j.} - \bar{y}_{...}) \end{aligned}$$

$$\begin{aligned}
& + 2 \sum_i \sum_j \sum_k (\bar{y}_{i..} - \bar{y}_{...})(\bar{y}_{..k} - \bar{y}_{...}) + 2 \sum_i \sum_j \sum_k (\bar{y}_{.j.} - \bar{y}_{...})(\bar{y}_{..k} - \bar{y}_{...}) \\
& \quad + 2 \sum_i \sum_j \sum_k (\bar{y}_{i..} - \bar{y}_{...})(y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}_{...}) \\
& \quad + 2 \sum_i \sum_j \sum_k (\bar{y}_{.j.} - \bar{y}_{...})(y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}_{...}) \\
& \quad + 2 \sum_i \sum_j \sum_k (\bar{y}_{..k} - \bar{y}_{...})(y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}_{...}). \tag{2.20}
\end{aligned}$$

É possível provar que os seis produtos cruzados da eq. (2.20) são nulos, como foi feito na eq. (2.7). Assim,

$$\begin{aligned}
\sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{...})^2 &= a \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 + a \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2 + a \sum_k (\bar{y}_{..k} - \bar{y}_{...})^2 \\
& \quad + \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}_{...})^2. \tag{2.21}
\end{aligned}$$

Expressando a eq. (2.21) simbolicamente tem-se que

$$SQTotal = SQTrat + SQLinhas + SQColunas + SQRes. \tag{2.22}$$

É possível demonstrar que os valores esperados dos quadrados médios de cada um dos componentes da análise, sendo tratamentos, linhas e colunas de efeito fixo, são

$$E(QMTrat) = \sigma^2 + \frac{a \sum_i \tau_i^2}{a - 1},$$

$$E(QMLinhas) = \sigma^2 + \frac{a \sum_j l_j^2}{a - 1},$$

$$E(QMColunas) = \sigma^2 + \frac{a \sum_k c_k^2}{a - 1}$$

e

$$E(QMRes) = \sigma^2.$$

Sob as suposições usuais de que os erros são variáveis aleatórias independentes, normalmente distribuídas e com mesma variância, cada uma das somas de quadrados do lado direito da eq. (2.22) é, mediante a divisão por  $\sigma^2$ , uma variável aleatória independente com distribuição qui-quadrado. A estatística apropriada para testar a hipótese de que não há o efeito

de tratamentos é como na eq. (2.16) que, sob a hipótese de homogeneidade, possui distribuição  $F$  com  $(a - 1)$  e  $(a - 2)(a - 1)$  graus de liberdade.

O procedimento completo para a análise de variância está apresentado na Tabela 2.5, sendo  $y_{i..}$ ,  $y_{.j.}$ ,  $y_{..k}$  e  $y_{...}$  os totais do  $i$ -ésimo tratamento, da  $j$ -ésima linha, da  $k$ -ésima coluna e das observações, respectivamente.

Tabela 2.5 – Análise de variância de um experimento com um fator realizado no delineamento quadrado latino

Causas de variação	Graus de liberdade	Somas de quadrados	Quadrados médios	$F$
Linhas	$a - 1$	$SQ_{Linhas} = \frac{\sum_j y_{.j.}^2}{a} - \frac{y_{...}^2}{n}$	$\frac{SQ_{Linhas}}{a - 1}$	$F_0$
Colunas	$a - 1$	$SQ_{Colunas} = \frac{\sum_k y_{..k}^2}{a} - \frac{y_{...}^2}{n}$	$\frac{SQ_{Colunas}}{a - 1}$	
Tratamentos	$a - 1$	$SQ_{Trat} = \frac{\sum_i y_{i..}^2}{a} - \frac{y_{...}^2}{n}$	$\frac{SQ_{Trat}}{a - 1}$	
Resíduo	$(a - 2)(a - 1)$	$SQ_{Res}$ (por subtração)	$\frac{SQ_{Res}}{(a - 2)(a - 1)}$	
Total	$n - 1$	$SQ_{Total} = \sum_{ijk} y_{ijk}^2 - \frac{y_{...}^2}{n}$		

Se linhas e/ou colunas são consideradas de efeito aleatório, o teste da hipótese de homogeneidade é realizado da maneira usual, com a diferença de que os valores esperados para linhas e colunas ficam da forma

$$E(QMLinhas) = \sigma^2 + a\sigma_L^2$$

e

$$E(QMColunas) = \sigma^2 + a\sigma_C^2,$$

em que  $\sigma_L^2$  e  $\sigma_C^2$  são os componentes de variância dos efeitos de linhas e colunas, respectivamente. Ainda assim, o valor esperado do quadrado médio de tratamentos é livre do efeito de linhas e colunas, e o teste para a hipótese de homogeneidade é feito da maneira tradicional.

#### 2.1.4 Dados desbalanceados

Sabe-se que, na prática, nem sempre é possível realizar um experimento em que todos os tratamentos tenham um mesmo número de repetições, o que muitas vezes pode ser explicado

por questões de orçamento, perda de unidade experimental no decorrer do experimento, ou mesmo por falta de material. Para contornar esse problema, duas alternativas podem ser utilizadas: a técnica da imputação dos dados faltantes ou a análise de variância dos dados desbalanceados.

Segundo Nunes, Klück e Fachel (2009), desde a década de 1980 surgiram técnicas estatísticas que envolvem a imputação dos dados faltantes. Essas técnicas, de acordo com Nunes (2007), têm por objetivo completar os conjuntos de dados e possibilitar a análise usual com dados completos.

As técnicas de imputação se dividem em dois casos: imputação simples ou única e imputação múltipla. A imputação simples ou única ocorre quando os dados faltantes são substituídos uma única vez (ENGELS; DIEHR, 2003), e então o conjunto de dados é analisado da maneira usual, usando-se procedimentos para dados completos. Já a imputação múltipla ocorre quando, para cada dado faltante, são imputados vários valores ao invés de um (RUBIN, 1987). Assim sendo, vários bancos de dados completos são obtidos, e cada conjunto de dados também é analisado da forma tradicional. Ao final das análises obtém-se, além das estimativas pontuais dos parâmetros por meio das médias das múltiplas imputações realizadas, a precisão das estimativas a partir do que é gerado pela incerteza dos dados (NUNES, 2007).

Atualmente é possível, ainda, devido ao desenvolvimento metodológico e computacional direcionado para a análise de dados desbalanceados, a realização da análise de variância sem que haja a necessidade da estimação dos valores perdidos no decorrer do ensaio, apenas mediante pequenos ajustes que devem ser feitos nas somas de quadrados que compõe a análise. Mais detalhes sobre o assunto podem ser encontrados em Iemma (1995a, 1995b).

De acordo com Montgomery (2001), há duas vantagens na escolha de um delineamento balanceado. Em primeiro lugar, a estatística para o teste da hipótese de igualdade entre as médias populacionais dos tratamentos é relativamente insensível a pequenos desvios da suposição de homogeneidade de variâncias se o número de repetições para tratamentos é o mesmo. Este não é o caso quando se trabalha com dados desbalanceados. Em segundo lugar, o poder do teste  $F$  global é maximizado se as amostras para tratamentos são de igual tamanho.

### **2.1.5 Pressuposições da análise de variância: a análise de resíduos**

A decomposição da variação das observações por meio de uma análise de variância é um procedimento puramente algébrico. No entanto, o uso da decomposição para testar formalmente eventuais diferenças que possam existir entre as médias dos tratamentos requer

que certos pressupostos sejam satisfeitos (MONTGOMERY, 2001). Especificamente, estas pressuposições são que as observações são descritas adequadamente pelo modelo, e que os erros são normalmente e independentemente distribuídos com média zero e variância comum,  $\sigma^2$ . Se cada uma delas é válida, o procedimento da análise de variância é o teste exato para a hipótese de homogeneidade entre as médias populacionais dos tratamentos comparados.

Na prática, no entanto, essas pressuposições nem sempre são totalmente satisfeitas. Consequentemente, é imprudente proceder a análise de variância sem antes checar cada uma delas. Violações das pressuposições e da adequação do modelo podem ser facilmente investigadas por meio da análise dos resíduos do modelo.

Anscombe e Tukey (1963) definem resíduo como a diferença entre o valor observado e o valor ajustado de uma observação, ou seja, o mesmo pode ser definido como o valor observado do erro. No delineamento inteiramente casualizado, por exemplo, o resíduo da  $j$ -ésima observação do  $i$ -ésimo tratamento é definido como

$$e_{ij} = y_{ij} - \hat{y}_{ij},$$

em que  $y_{ij}$  é a resposta obtida com o  $i$ -ésimo tratamento em sua  $j$ -ésima repetição e  $\hat{y}_{ij}$  é o seu valor predito, também conhecido como valor ajustado.

Para saber se cada uma das pressuposições da análise está satisfeita, convém verificar os seguintes itens: a presença de dados discrepantes, se os erros são independentes, se a variância dos mesmos é constante, se a sua distribuição é normal e se o modelo é aditivo. De acordo com Parente (1984), o procedimento usual para checar cada uma dessas pressuposições consiste na execução da análise gráfica dos resíduos. Em seguida, para a confirmação ou não da análise gráfica, podem ser aplicados testes estatísticos. Cada um desses procedimentos será discutido a seguir.

### 2.1.5.1 Dados discrepantes

Há várias formas para se detectar a presença de dados discrepantes (*outliers*) num conjunto de dados. Uma das verificações pode ser feita por meio da análise gráfica dos resíduos do modelo, afinal, dados discrepantes são observações que possuem um resíduo muito grande ou muito pequeno em comparação com a maioria dos resíduos do gráfico.

Neter e Wasserman (1974), Seber (1977), Box, Hunter Jr. e Hunter (1978), Draper e Smith (1998), Dean e Voss (1999), Montgomery (2001), dentre outros, aconselham realizar a

análise gráfica dos resíduos com base nos resíduos padronizados do modelo pois, nesse caso, o valor discrepante pode ser facilmente detectado.

Para encontrar os resíduos padronizados, é necessário dividir cada um dos resíduos pela raiz quadrada do quadrado médio residual obtido na análise de variância. Considerando ainda um experimento realizado segundo o delineamento inteiramente casualizado, os resíduos padronizados são definidos como:

$$z_{ij} = \frac{e_{ij}}{\sqrt{QMRes}}$$

Se  $\varepsilon_{ij} \sim N(0, \sigma^2)$ , os resíduos padronizados devem seguir aproximadamente uma distribuição normal padrão. Assim sendo, cerca de 68% dos resíduos padronizados devem estar no intervalo entre -1 e 1 e cerca de 95% dos mesmos devem estar no intervalo delimitado pelos valores -2 e 2. Por sua vez, valores que se localizarem fora do intervalo que vai de -3 a 3 devem ser investigados.

A presença de um ou mais pontos discrepantes pode afetar seriamente a análise de variância, pois pode levar à ausência de normalidade dos erros, bem como à heterogeneidade de variâncias, uma vez que uma observação discrepante altera de forma significativa a variância do grupo onde ele é detectado. Portanto, quando um deles é localizado, uma investigação mais cuidadosa deve ser realizada (PARENTE, 1984).

Geralmente, um dado discrepante pode ser devido a algum erro de medição ou registro. Nesse caso, o procedimento mais comum consiste em descartar a observação. No entanto, se esse não é o caso, as circunstâncias experimentais em torno desta unidade experimental devem ser cuidadosamente estudadas.

De acordo com Parente (1984), quando se suspeita que o dado discrepante é uma observação correta de um fenômeno pouco conhecido, então ele pode ser tomado com um peso reduzido ou, ainda, eliminado da análise e discutido isoladamente. Outro procedimento possível seria transformar todos os dados do ensaio e aplicar a análise de variância usual ou, então, usar testes não-paramétricos. Em todo caso, deve-se ter o cuidado para não rejeitar ou descartar uma observação a menos que se tenha motivos estatísticos razoáveis para isso.

### **2.1.5.2 Independência**

Uma pressuposição para a realização da análise de variância, como já colocado, é a de que os erros são variáveis aleatórias independentes. De acordo com Demétrio (1978), esse

pressuposto pode ser assegurado por um dos princípios básicos da experimentação que é a casualização.

A não independência dos erros, no entanto, é um problema potencialmente grave que se pode ter numa análise de variância, afinal, além dos resultados da mesma ficarem comprometidos, a dependência dos erros é difícil de ser corrigida (MONTGOMERY, 2001).

Segundo Dean e Voss (1999), a causa mais provável de não independência entre os erros do modelo é a similaridade entre unidades experimentais juntas no tempo ou no espaço. Portanto, para checar a independência dos erros, basta construir um gráfico dos resíduos padronizados contra a ordem em que as observações foram coletadas, seja no tempo ou no espaço. Se a pressuposição estiver sendo satisfeita, os pontos do gráfico ficarão dispostos aleatoriamente em torno do valor zero, sem nenhum padrão aparente. Caso contrário, as observações mostrarão clara correlação com a sua ordem de coleta.

Se uma tendência linear no tempo for notada no gráfico, por exemplo, pode ser possível adicionar termos ao modelo estatístico que possam representar esse efeito temporal (DEAN; VOSS, 1999). No caso de um experimento conduzido no delineamento inteiramente casualizado, por exemplo, em que essa tendência entre os erros é clara, um modelo mais complexo que pode ser adotado é da forma

$$y_{ij} = \mu + \tau_i + \gamma t_{ij} + \varepsilon_{ij},$$

em que  $t_{ij}$  representa o tempo em que a observação foi coletada e  $\gamma$  é um parâmetro de tendência temporal linear, e que deve ser estimado.

De acordo com Dean e Voss (1999), se a análise de variância é conduzida mesmo quando os erros se mostram variáveis aleatórias dependentes, o verdadeiro nível de significância ao qual os testes de hipóteses são submetidos pode ficar muito superior do que o nível de significância nominal imposto para a aplicação dos métodos, e os níveis de confiança e poder dos testes podem ficar deflacionados.

### **2.1.5.3 Homogeneidade de variâncias**

Depois de verificada a independência entre os erros do modelo, o próximo passo é então averiguar se a variância dos erros é constante.

Uma forma prática de verificar essa pressuposição é construindo um gráfico dos resíduos padronizados contra os valores ajustados do modelo. Se a pressuposição de homogeneidade de variâncias estiver sendo de fato satisfeita, os pontos devem se distribuir



aleatoriamente em torno do zero, sem nenhum padrão aparente. Caso contrário, os pontos tendem a apresentar alguma tendência no gráfico então construído.

Trata-se, porém, de uma análise exploratória apenas. Se o gráfico apresentar indícios de heterogeneidade de variâncias, um teste estatístico para confirmar essa suspeita pode ser aplicado. Dentre eles pode-se citar o teste de Levene, por exemplo, o qual se mostra bastante eficiente mesmo quando a hipótese de normalidade dos erros não é aceita.

No teste de Levene, tem-se como hipóteses

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_a^2;$$

$$H_1: \text{a afirmação acima não é verdadeira para pelo menos um } \sigma_i^2$$

em que  $a$  é o número de tratamentos do ensaio, e  $\sigma_i^2$  é a respectiva variância associada ao  $i$ -ésimo tratamento.

O procedimento, proposto por Levene (1960), consiste em fazer uma modificação no conjunto de dados original. Considerando ainda um ensaio realizado segundo o delineamento inteiramente casualizado, o processo consiste em realizar a seguinte transformação:

$$l_{ij} = |y_{ij} - \bar{y}_i|,$$

em que  $i = 1, \dots, a$ ,  $j = 1, \dots, r$ ,  $l_{ij}$  representa o dado após ele ter sido transformado,  $y_{ij}$  representa o dado original e  $\bar{y}_i$  representa a média do  $i$ -ésimo tratamento para os dados originais.

Após esse procedimento, faz-se então a análise de variância dos dados já transformados, e a estatística para o teste da hipótese de homocedasticidade é, então, da seguinte forma:

$$F = \frac{r(n-a) \sum_i (\bar{l}_i - \bar{l}_{..})^2}{(a-1) \sum_i \sum_j (l_{ij} - \bar{l}_i)^2}, \quad (2.23)$$

sendo  $n$  o número total de observações do ensaio,  $\bar{l}_i$  a média amostral do  $i$ -ésimo tratamento para os dados transformados, e  $\bar{l}_{..}$  a média geral dos dados modificados.

Há evidências para não aceitar a hipótese de homogeneidade de variâncias se o valor obtido por meio da expressão (2.23) for maior que  $F_{(\alpha; a-1, n-a)}$ , em que este é o valor tabelado de  $F$  de Snedecor com  $(a-1)$  e  $(n-a)$  graus de liberdade, respectivamente, ao nível de significância  $\alpha$ .

Sendo confirmada a suspeita de variâncias heterogêneas, é possível ainda assim realizar a análise dos dados, contanto que os dados sejam ponderados pelo inverso das variâncias ou seja feita uma transformação dos mesmos, que torne as variâncias homogêneas. Nos casos em

que se conhece a distribuição teórica das observações, esta informação deve ser utilizada no momento de determinar a melhor transformação a ser realizada (MONTGOMERY, 2001). Se as observações seguem uma distribuição de Poisson, por exemplo, a transformação raiz quadrada  $y_{ij}^* = \sqrt{y_{ij}}$  ou  $y_{ij}^* = \sqrt{1 + y_{ij}}$  deve ser utilizada. Se os dados seguem uma distribuição log-normal, a transformação logarítmica  $y_{ij}^* = \log y_{ij}$  é apropriada. Para dados binomiais expressos como frações, a transformação arco seno  $y_{ij}^* = \arcsen\sqrt{y_{ij}}$  é que deve ser empregada.

De acordo com Dean e Voss (1999) e Montgomery (2001), se a pressuposição de homocedasticidade é ligeiramente violada, o teste  $F$  global é pouco afetado em modelos balanceados de efeitos fixos. No entanto, em caso de dados desbalanceados ou em casos em que uma variância é consideravelmente maior que as demais, o problema é mais grave. Especificamente, se os tratamentos que possuem as maiores variâncias possuem os menores tamanhos amostrais, o nível de significância real dos testes de hipóteses fica maior do que o nível nominal declarado. Por outro lado, se os tratamentos com maiores variâncias possuem também os maiores tamanhos amostrais, então o nível de significância real fica menor que o nível nominal.

No entanto, segundo Demétrio (1978), o teste  $F$  global da análise de variância é o menos afetado com a falta de homogeneidade de variâncias. As comparações entre as médias dos tratamentos, porém, feitas por meio da aplicação dos testes de médias, utilizando-se do quadrado médio residual, apresentarão seus resultados distorcidos caso essa pressuposição esteja violada. Assim sendo, é necessário realizar uma transformação para estabilizar as variâncias.

#### **2.1.5.4 Distribuição normal**

Outra pressuposição da análise de variância é a de que os erros possuem distribuição normal ou, ao menos, aproximadamente normal.

Uma forma de verificar essa pressuposição é construir um histograma dos resíduos padronizados. Se os erros de fato seguem uma distribuição normal com média zero e variância  $\sigma^2$ , o gráfico deve ser simétrico com relação ao zero e possuir o formato de um “sino”. No entanto, se o número de observações do experimento é pequeno, flutuações no gráfico podem ocorrer com facilidade, ou seja, pequenos desvios não necessariamente implicam em violações graves dessa pressuposição (MONTGOMERY, 2001).

Outro procedimento bastante usual para verificar essa pressuposição é construir um gráfico normal de probabilidades, dos resíduos padronizados contra os escores de uma distribuição normal padrão. Caso os pontos se localizem, aproximadamente, em torno da reta identidade, então aparentemente a pressuposição de normalidade está sendo satisfeita. Caso contrário, há evidências de que a mesma está sendo violada e, para confirmar essa suspeita, pode-se fazer uso de um teste estatístico, tal como o de Shapiro-Wilk ou o de Kolmogorov-Smirnov, por exemplo.

Em ambos os testes, tem-se como hipóteses:

$H_0$ : *Os dados seguem uma distribuição normal;*

$H_1$ : *Os dados não seguem uma distribuição normal.*

Para realizar o teste de Shapiro-Wilk, deve-se proceder da seguinte maneira: seja  $x_1, x_2, \dots, x_n$  uma amostra aleatória de observações, e seja  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$  uma estatística de ordem. A estatística para o teste da hipótese de normalidade é da forma:

$$W = \frac{\varphi^2}{\sum_{m=1}^n (x_{(m)} - \bar{x}_{(\cdot)})^2},$$

com  $\bar{x}_{(\cdot)} = \sum_m x_{(m)}/n$ . A constante  $\varphi$ , por sua vez, é obtida da seguinte maneira:

$$\varphi = \sum_{m=1}^{n/2} w_{n-m+1} (x_{(n-m+1)} - x_{(m)})$$

se  $n$  é par, e

$$\varphi = \sum_{m=1}^{(n+1)/2} w_{n-m+1} (x_{(n-m+1)} - x_{(m)})$$

se  $n$  é ímpar, em que  $w_{n-m+1}$  dependem dos valores da estatística de ordem e estão tabelados em Shapiro e Wilk (1965).

Há evidências para não aceitar a hipótese de normalidade se  $W < W_{(\alpha,n)}$ , em que  $W_{(\alpha,n)}$  é o valor crítico da distribuição da estatística  $W$ .

Já o teste de Kolmogorov-Smirnov, originalmente proposto na década de 1930 em artigos de Kolmogorov (1933) e Smirnov (1936, 1937) observa a máxima diferença absoluta entre a função de distribuição acumulada da normal e a função de distribuição empírica dos dados. Seja  $x_1, x_2, \dots, x_n$  uma amostra aleatória de uma população com função de distribuição desconhecida. A estatística para o teste da hipótese de normalidade é

$$D = \sup_x |F(x) - F_n(x)|,$$

em que  $F(x)$  é a função de distribuição acumulada da normal e  $F_n(x)$  representa a função de distribuição empírica do conjunto de dados. Existem evidências para não aceitar a hipótese de normalidade se  $D > D_{(\alpha,n)}$ , em que  $D_{(\alpha,n)}$  é o valor crítico da distribuição da estatística  $D$ .

Se a suspeita de não-normalidade for confirmada, ainda assim a análise de variância pode ser realizada, mediante uma transformação no conjunto de dados, tal como a de Box e Cox (1964), por exemplo, que sugeriram uma alteração no conjunto de dados da forma  $y^* = y^\lambda$ . O procedimento consiste na obtenção de vários valores de  $\lambda$  e na geração de uma análise de variância para cada transformação, da seguinte forma

$$y^\lambda = \begin{cases} \frac{y^\lambda - 1}{\lambda \dot{y}^{\lambda-1}} & \lambda \neq 0; \\ \dot{y} \ln y & \lambda = 0, \end{cases}$$

em que  $\dot{y} = \ln^{-1}[(1/n) \sum \ln y]$  é a média geométrica das observações. A estimativa de máxima verossimilhança de  $\lambda$  é o valor para o qual a soma de quadrados do resíduo da análise de variância é mínima.

De qualquer forma, em modelos de efeitos fixos, segundo Montgomery (2001), o teste  $F$  global da análise de variância, bem como procedimentos de comparação de médias, são robustos a pequenas transgressões da normalidade e, portanto, essas violações são comuns e não costumam afetar os resultados das análises. Já os modelos de efeitos aleatórios são mais severamente afetados com a fuga dessa pressuposição.

### 2.1.5.5 Aditividade

A pressuposição de aditividade do modelo não está mais relacionada aos erros experimentais, mas sim à estrutura do modelo, e indica que os efeitos de tratamentos e ambientais devem se somar, ou seja, que não deve haver interação entre eles. Esse pressuposto pode ser checado por meio do teste de não-aditividade proposto por Tukey (1949), na própria análise de variância.

O teste originalmente foi apresentado para experimentos realizados no delineamento casualizado em blocos, mas também pode ser aplicado para experimentos no delineamento quadrado latino, com algumas pequenas modificações. Nos experimentos em blocos, é verificado se tratamentos e blocos são aditivos e, nos quadrados latinos, se tratamentos, linhas e colunas são aditivos.

Tendo como exemplo um experimento realizado no delineamento casualizado em blocos, se os efeitos são aditivos, tem-se que o modelo de efeitos, tal como na eq. (2.10), pode ser utilizado para descrever os dados do ensaio. Já se os efeitos não são aditivos, o modelo adequado é da forma:

$$y_{ij} = \mu + \tau_i + \beta_j + \eta \tau_i \beta_j + \varepsilon_{ij},$$

com  $i = 1, \dots, a$  e  $j = 1, \dots, b$ , sendo  $a$  e  $b$  o número de tratamentos e de blocos do ensaio, respectivamente.

O teste de Tukey para não-aditividade verifica as hipóteses:

$$H_0: \eta = 0 \text{ (modelo aditivo);}$$

$$H_1: \eta \neq 0 \text{ (modelo não - aditivo).}$$

Na análise de variância, a interação entre tratamentos e blocos só será representada pelo resíduo se o modelo for aditivo. Caso contrário, essa interação não é a medida adequada da variação residual. A soma de quadrados para não-aditividade nada mais é, portanto, do que uma interação dos efeitos lineares de tratamentos e blocos (DEMÉTRIO, 1978).

Se a soma de quadrados para não-aditividade for significativa, isso indica que a diferença entre tratamentos não é constante para os diversos blocos, como deveria ser se os efeitos fossem aditivos (DEMÉTRIO, 1978). Essa soma de quadrados é definida como:

$$SQ_{\text{NãoAditividade}} = \frac{[\sum_i \sum_j y_{ij} (\bar{y}_{i.} - \bar{y}_{..}) (\bar{y}_{.j} - \bar{y}_{..})]^2}{\sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..})^2 (\bar{y}_{.j} - \bar{y}_{..})^2},$$

sendo  $\bar{y}_{i.}$ ,  $\bar{y}_{.j}$  e  $\bar{y}_{..}$  as médias do  $i$ -ésimo tratamento, do  $j$ -ésimo bloco e das observações, respectivamente.

A análise de variância com a aplicação do teste de Tukey para não-aditividade tem a forma da Tabela 2.6.

Tabela 2.6 – Análise de variância com a aplicação do teste de Tukey para não-aditividade num experimento com um fator realizado no delineamento casualizado em blocos

Causas de variação	Graus de liberdade	Somas de quadrados
Blocos	$b - 1$	$SQBlocos$
Tratamentos	$a - 1$	$SQTrat$
Não-aditividade	1	$SQNãoAditividade$
Novo Resíduo	$(a - 1)(b - 1) - 1$	$SQNovoRes = SQRes - SQNãoAditividade$
Total	$n - 1$	$SQTotal$

A não-aditividade pode, então, ser testada pelo teste  $F$  global:

$$F = \frac{QM\text{N}\text{ãoAditividade}}{QM\text{NovoRes}},$$

com 1 e  $(pq - p - q)$  graus de liberdade.

Quando a não-aditividade for significativa, há evidências para não aceitar a hipótese de nulidade, ou seja, o modelo apropriado não é aditivo. De acordo com Demétrio (1978), se essa não-aditividade for devida a interação dos efeitos principais, deve-se fazer uma transformação no conjunto de dados original. Já se ela for devida a observações discrepantes, faz-se a análise considerando-as como unidades experimentais perdidas.

A presença de não-aditividade nos dados pode levar a heterocedasticidade dos erros e, conseqüentemente, a utilização da variância residual pode conduzir a estimativas viciadas dos efeitos de tratamentos e dar níveis de significância falsos para os testes de hipóteses (DEMÉTRIO, 1978).

## 2.2 Testes de comparação de médias

Se na condução da análise de variância para um modelo estatístico em que os tratamentos são de efeito fixo o teste  $F$  global apresenta resultado significativo, conclui-se então que existem evidências de diferenças entre as médias populacionais dos tratamentos comparados e, segundo Chew (1976), o passo seguinte da pesquisa é então determinar quais as médias de tratamentos que diferem entre si, a fim de que possa(m) ser então indicado(s) o(s) melhor(es) tratamento(s), dentro das expectativas da pesquisa.

De acordo com Nogueira (2007), existem dois procedimentos possíveis nessa análise continuada dos dados: a aplicação da análise de regressão ou a aplicação de um teste de comparação de médias. A análise de regressão deve ser aplicada quando os tratamentos envolvidos no ensaio são quantitativos e, os testes de médias, quando os mesmos são qualitativos.

No presente trabalho, o interesse está em estudar os testes de comparação de médias apenas e, portanto, não será discutida aqui a análise de regressão. No entanto, mais detalhes sobre o assunto podem ser encontrados nos trabalhos de Draper e Smith (1998), Kutner et al. (2004), Montgomery, Peck e Vining (2006), dentre outros.

Os testes de comparação de médias são, de acordo com Hsu (1996), o segundo procedimento estatístico mais utilizado por pesquisadores em suas análises de dados, perdendo

apenas para o teste  $F$  global na análise de variância. No entanto, se são o segundo em uso, certamente são o primeiro em mau uso.

Os procedimentos de comparação de médias, de um modo geral, permitem testar hipóteses do tipo:

$$H_0: Y = 0;$$

$$H_1: Y \neq 0,$$

em que  $Y$  representa uma função linear das médias populacionais dos tratamentos comparados, da seguinte forma:

$$Y = \sum_{i=1}^a c_i \mu_i,$$

sendo  $a$  o número de tratamentos envolvidos no experimento,  $c_i$  o coeficiente associado a  $\mu_i$ , podendo assumir valores positivos, negativos ou até mesmo zero, e  $\mu_i$  representa a média populacional referente ao  $i$ -ésimo tratamento.

Impondo a restrição de que

$$\sum_{i=1}^a c_i = 0,$$

a função linear  $Y$  denomina-se um contraste.

Um estimador não viciado de  $Y$  é representado da seguinte maneira:

$$\hat{Y} = \sum_{i=1}^a c_i \bar{y}_i, \quad (2.24)$$

sendo  $\bar{y}_i$  a média amostral associada ao  $i$ -ésimo tratamento.

A variância de  $\hat{Y}$ ,  $V(\hat{Y})$ , é obtida a partir dos seguintes cálculos:

$$V(\hat{Y}) = V\left(\sum_{i=1}^a c_i \bar{y}_i\right) = \sum_{i=1}^a c_i^2 V(\bar{y}_i) = \sum_{i=1}^a c_i^2 \frac{\sigma_i^2}{r_i},$$

em que  $\sigma_i^2$  é a variância do  $i$ -ésimo tratamento e  $r_i$  o seu respectivo número de repetições.

Supondo a condição de homogeneidade de variâncias, tem-se que:

$$V(\hat{Y}) = \sigma^2 \sum_{i=1}^a \frac{c_i^2}{r_i}.$$

No entanto, a variância  $\sigma^2$  é muitas vezes desconhecida, afinal, na maioria dos casos, dispõe-se apenas de uma amostra. Portanto, uma estimativa da variância da estimativa do contraste pode ser obtida da seguinte maneira:

$$\hat{V}(\hat{Y}) = s^2 \sum_{i=1}^a \frac{c_i^2}{r_i}, \quad (2.25)$$

em que  $s^2$  é um estimador não viciado de  $\sigma^2$ , representado pelo quadrado médio residual e, portanto:

$$\hat{V}(\hat{Y}) = QMResíduo \sum_{i=1}^a \frac{c_i^2}{r_i}.$$

A seguir serão apresentadas as possíveis taxas de erro de um teste para a comparação de médias de tratamentos e, em sequência, alguns desses procedimentos comumente empregados por pesquisadores na análise estatística de dados.

### 2.2.1 Taxas de erro nos testes de comparação de médias

A aplicação de um teste de comparação de médias usualmente consiste na realização de uma sequência de testes de hipóteses, chamada de família de testes, sobre um conjunto de combinações lineares das médias populacionais dos tratamentos comparados.

A esses testes de hipóteses, assim como em qualquer outro, estão associados erros de decisão. Se uma dessas hipóteses for verdadeira, por exemplo, e se ela não for aceita, então comete-se um erro tipo I. Ao contrário, se uma dessas hipóteses for falsa, e se ainda assim ela for aceita, então comete-se um erro tipo II.

Qualquer teste para decidir entre duas hipóteses  $H_0$  e  $H_1$  é avaliado em termos das probabilidades dos dois tipos de erro. Se, por exemplo, a decisão é não aceitar  $H_0$ , então a probabilidade de erro tipo I deve ser muito pequena. Denota-se essa probabilidade pela letra grega  $\alpha$  e, usando o operador de probabilidade  $P$ , é possível defini-la da seguinte forma:

$$\alpha = P(\text{Erro tipo I}) = P(\text{Rejeitar } H_0 \mid H_0 \text{ é verdadeira}).$$

Se, ao contrário, a decisão é aceitar  $H_0$ , então a probabilidade de erro tipo II deve ser muito baixa. Essa probabilidade é denotada pela letra grega  $\beta$  e, usando o operador de probabilidade  $P$ , é possível defini-la como:

$$\beta = P(\text{Erro tipo II}) = P(\text{Aceitar } H_0 \mid H_0 \text{ é falsa}).$$



Outro conceito que vale a pena discutir aqui é o de poder do teste, que também tem como objetivo conhecer o quanto o teste de hipótese controla o erro tipo II, ou qual a probabilidade de não aceitar a hipótese nula sendo ela realmente falsa. Em termos do operador de probabilidade, o poder de um teste estatístico ( $\pi$ ) é definido como:

$$\pi = 1 - \beta = P(\text{Rejeitar } H_0 \mid H_0 \text{ é falsa}).$$

O procedimento geral num teste de hipótese consiste em especificar um valor para a probabilidade  $\alpha$  de ocorrência de erro tipo I, também conhecida como nível de significância do teste e, então, seguir com os procedimentos de modo que a probabilidade de erro tipo II tenha um valor adequadamente pequeno ou, em outras palavras, de modo que o poder do teste seja alto.

De acordo com Rafter, Abell e Braselton (2002), pode-se citar a presença de três níveis de significância ou, em outras palavras, três taxas de erro tipo I em testes estatísticos para a comparação de médias de tratamentos. A primeira delas é chamada de taxa de erro por comparação (*comparisonwise error rate*), a segunda é denominada de taxa de erro por experimento (*experimentwise error rate*) e, a terceira, de taxa de erro por família (*per-family error rate*).

Quando num experimento cada um dos vários testes de hipóteses para a comparação das médias dos tratamentos é feito com o mesmo nível de significância, então este é chamado de nível de significância por comparação,  $\alpha_c$ , e corresponde a probabilidade de incorretamente não aceitar cada uma das hipóteses nulas que compõe a família dos testes realizados. Já o nível de significância por experimento,  $\alpha_e$ , diz respeito a probabilidade de que em pelo menos um dos testes de hipóteses que compõe a família das comparações realizadas ocorra o erro tipo I, ou seja, corresponde a probabilidade de não aceitação, de forma incorreta, de pelo menos uma das hipóteses nulas que compõe a família dos testes.

Por fim, a taxa de erro por família (*TEPF*) não representa uma probabilidade, tal como o nível de significância por comparação e o nível de significância por experimento, mas corresponde ao valor esperado de erros na família dos testes realizados. Por exemplo, supondo que a hipótese de homogeneidade é verdadeira, se cada um dos  $c = a(a - 1)/2$  testes para comparação das médias de tratamentos duas a duas é realizado com uma probabilidade de erro tipo I de  $\alpha/c$ , o valor esperado de erros tipo I é  $c \times (\alpha/c) = \alpha$ . Assim, quando  $\mu_1 = \mu_2 = \dots = \mu_a$ , a taxa de erro por família é exatamente igual a  $\alpha$ . Para qualquer outra configuração das médias, entretanto, essa taxa é inferior a esse valor. Uma relação entre esses três tipos de taxa de erro é dada por:

$$\alpha_c \leq \alpha_e \leq TEPF.$$

Vários estudos de simulação têm sido realizados para avaliar o desempenho de alguns procedimentos de comparação de médias com relação aos erros tipo I e/ou quanto ao poder do teste, em diferentes cenários formados pela combinação de diversas variáveis. De acordo com Conagin e Gomes (2004) esses estudos são de fundamental importância, pois auxiliam o pesquisador na escolha do teste mais adequado a ser utilizado em sua pesquisa, afinal, segundo Machado et al. (2005), a escolha apropriada depende das qualidades estatísticas desse método, e esta qualidade também é função do tipo de erro que é controlado e da forma como estes erros são controlados.

### 2.2.2 Teste DMS de Fisher

O teste da diferença mínima significativa de Fisher, ou simplesmente teste DMS de Fisher, foi proposto para comparar todo e qualquer contraste entre duas médias de tratamentos. Assim sendo, as hipóteses a serem testadas são da forma:

$$H_0: \mu_i - \mu_{i'} = 0;$$

$$H_1: \mu_i - \mu_{i'} \neq 0,$$

sendo  $\mu_i$  e  $\mu_{i'}$  as médias populacionais associadas ao  $i$ -ésimo e  $i'$ -ésimo tratamento, respectivamente, com  $i, i' = 1, \dots, a$ , em que  $a$  é o número de tratamentos envolvidos no experimento e  $i \neq i'$ .

No procedimento proposto por Fisher (1935), há evidências para não aceitar a hipótese nula de igualdade entre as médias populacionais dos tratamentos comparados, a determinado nível de significância, sempre que a diferença, em módulo, entre as estimativas das duas médias for maior que a diferença mínima significativa (DMS) sugerida pelo autor para a aplicação do teste, ou seja, sempre que

$$|\bar{y}_i - \bar{y}_{i'}| > DMS,$$

em que  $\bar{y}_i$  e  $\bar{y}_{i'}$  representam as médias amostrais associadas ao  $i$ -ésimo e ao  $i'$ -ésimo tratamento, respectivamente.

No caso de dados balanceados, a diferença mínima significativa para a aplicação do teste é da forma:

$$DMS = t_{(\alpha/2;v)} \sqrt{\frac{2s^2}{r}},$$

em que  $t_{(\alpha/2;v)}$  é o valor tabelado de  $t$  de Student a um nível de significância  $\alpha$  e segundo  $v$  graus de liberdade associados ao resíduo,  $s^2$  é o quadrado médio residual e  $r$  é o número de repetições para tratamentos.

No caso de dados desbalanceados, porém, tem-se que a diferença mínima significativa é da seguinte maneira:

$$DMS = t_{(\alpha/2;v)} \sqrt{s^2 \left( \frac{1}{r_i} + \frac{1}{r_{i'}} \right)},$$

em que  $r_i$  e  $r_{i'}$  representam o número de repetições do  $i$ -ésimo e  $i'$ -ésimo tratamento, respectivamente.

Muitas vezes a aplicação do teste DMS de Fisher é precedida por um resultado significativo do teste  $F$  global na análise de variância. No entanto, se o teste  $F$  for não significativo, nenhuma comparação de médias é feita, eliminando assim a possibilidade de se cometer o erro tipo I. Este procedimento é denominado teste DMS de Fisher protegido. A diferença mínima significativa para aplicação do método é como no teste DMS de Fisher se o teste  $F$  global é significativo, e

$$DMS = \infty$$

se o teste  $F$  é não significativo.

### 2.2.3 Teste de Tukey

O teste proposto por Tukey (1953) também compara todo e qualquer contraste entre duas médias de tratamentos. Na aplicação do método, existem evidências para não aceitar a hipótese nula de igualdade entre as médias populacionais de dois tratamentos, a determinado nível de significância, sempre que a diferença, em módulo, entre as estimativas das duas médias comparadas for maior que a diferença mínima significativa ( $T_\alpha$ ) proposta para aplicação do teste, ou seja, quando

$$|\bar{y}_i - \bar{y}_{i'}| > T_\alpha,$$

em que  $\bar{y}_i$  e  $\bar{y}_{i'}$  representam as médias amostrais associadas ao  $i$ -ésimo e  $i'$ -ésimo tratamento, respectivamente.

No caso de dados balanceados, a diferença mínima significativa para a aplicação do teste é da forma:

$$T_\alpha = q_{(\alpha;a,\nu)} \sqrt{\frac{s^2}{r}}$$

em que  $q_{(\alpha;a,\nu)}$  é a amplitude total estudentizada, encontrada nas tabelas do teste de Tukey, com nível de significância  $\alpha$ ,  $a$  tratamentos envolvidos e  $\nu$  graus de liberdade associados ao resíduo,  $s^2$  é o quadrado médio residual e  $r$  é o número de repetições dos tratamentos.

Caso o número de repetições para tratamentos não seja o mesmo, no entanto, o teste, também conhecido nesse caso como procedimento de Tukey-Kramer, tem como diferença mínima significativa

$$T_\alpha = q_{(\alpha;a,\nu)} \sqrt{\frac{1}{2} \left( \frac{1}{r_i} + \frac{1}{r_{i'}} \right) s^2},$$

em que  $r_i$  e  $r_{i'}$  representam o número de repetições do  $i$ -ésimo e  $i'$ -ésimo tratamento, respectivamente.

#### 2.2.4 Teste de Duncan

O teste de Duncan (1955) também permite testar todo e qualquer contraste entre duas médias de tratamentos. No entanto, diferentemente dos testes de Tukey e DMS de Fisher, o teste de Duncan utiliza do cálculo de múltiplas diferenças mínimas significantes.

Para aplicar o método, o primeiro passo consiste em ordenar as médias amostrais obtidas para cada um dos tratamentos, seja em ordem crescente ou decrescente. Calcula-se então a diferença mínima significativa para comparar a maior das médias com a menor delas. A comparação entre elas corresponde a um intervalo que abrange todas as  $a$  médias. Caso a diferença entre elas não seja considerada significativa, o teste se encerra, caso contrário calcula-se então outra diferença mínima significativa, agora para comparar médias em um intervalo abrangendo  $(a - 1)$  médias, e assim por diante. Sempre que duas médias não são consideradas estatisticamente diferentes, não se pode testar as diferenças entre médias que estão no intervalo delimitado por aquelas duas médias.

No caso de mesmo número de repetições para tratamentos, a diferença mínima significativa ( $R_p$ ) proposta para aplicação do teste é da forma:

$$R_p = z_{(\alpha;p,\nu)} \sqrt{\frac{s^2}{r}}$$

sendo  $z_{(\alpha;p,\nu)}$  a amplitude total estudentizada obtida nas tabelas do teste de Duncan, ao nível de significância  $\alpha$ , para  $p$  médias abrangidas pelo intervalo delimitado pelas duas médias comparadas,  $p = 2, \dots, a$ , e  $\nu$  graus de liberdade associados ao resíduo,  $s^2$  é o quadrado médio residual e  $r$  é o número de repetições para tratamentos.

Caso o número de repetições para tratamentos não seja o mesmo, no entanto, a diferença mínima significativa proposta para a aplicação do teste é da seguinte maneira:

$$R_p = z_{(\alpha;p,\nu)} \sqrt{\frac{1}{2} \left( \frac{1}{r_i} + \frac{1}{r_{i'}} \right) s^2},$$

em que  $r_i$  e  $r_{i'}$  representam o número de repetições do  $i$ -ésimo e  $i'$ -ésimo tratamento, respectivamente.

Há evidências para não aceitar a hipótese nula de igualdade entre as duas médias populacionais de tratamentos comparadas, ao nível de significância estabelecido, quando

$$|\bar{y}_i - \bar{y}_{i'}| > R_p,$$

em que  $\bar{y}_i$  e  $\bar{y}_{i'}$  representam as médias amostrais associadas ao  $i$ -ésimo e  $i'$ -ésimo tratamento, respectivamente.

### 2.2.5 Teste de Student-Newman-Keuls

O teste de Student-Newman-Keuls (SNK) foi desenvolvido por Gossett (1927), Newman (1939) e Keuls (1952). Assim como o teste de Duncan, o teste de SNK utiliza de múltiplas diferenças mínimas significantes para a comparação das médias de tratamentos entre si duas a duas, de acordo com o posicionamento das mesmas após serem ordenadas, seja em ordem crescente ou decrescente.

A diferença entre ambos os métodos está no fato de que ao invés de buscar o valor de  $z_{(\alpha;p,\nu)}$  nas tabelas do teste de Duncan, busca-se agora a amplitude total estudentizada  $q_{(\alpha;p,\nu)}$  obtida nas tabelas do teste de Tukey.

### 2.2.6 Teste de Dunnett

O método proposto por Dunnett (1964) é recomendado quando se pretende comparar as médias dos tratamentos apenas com a média do tratamento controle. Assim sendo, as hipóteses a se testar são da forma

$$H_0: \mu_i - \mu_c = 0;$$

$$H_0: \mu_i - \mu_c \neq 0,$$

em que  $\mu_c$  representa a média populacional do tratamento controle e  $\mu_i$  a média populacional do  $i$ -ésimo tratamento, para  $i = 1, \dots, (a - 1)$ .

Há evidências para não aceitar a hipótese nula de igualdade entre as médias populacionais dos dois tratamentos comparados, a determinado nível de significância, sempre que a diferença, em módulo, entre as estimativas das duas médias for maior que a diferença mínima significativa ( $D_\alpha$ ) proposta para aplicação do teste, ou seja, sempre que

$$|\bar{y}_i - \bar{y}_c| > D_\alpha,$$

em que  $\bar{y}_i$  e  $\bar{y}_c$  representam as médias amostrais associadas ao  $i$ -ésimo tratamento e ao tratamento controle, respectivamente.

No caso de dados balanceados, a diferença mínima significativa proposta para a aplicação do teste é da seguinte forma:

$$D_\alpha = d_{(\alpha; a-1, \nu)} \sqrt{\frac{2s^2}{r}},$$

sendo  $d_{(\alpha; a-1, \nu)}$  o valor tabelado para o teste de Dunnett, obtido ao nível de significância  $\alpha$ ,  $(a - 1)$  tratamentos excluindo o controle e  $\nu$  graus de liberdade associados ao resíduo,  $s^2$  é o quadrado médio residual e  $r$  é o número de repetições para os tratamentos.

No caso de número desigual de repetições para os tratamentos comparados, a diferença mínima significativa pode ser escrita como

$$D_\alpha = d_{(\alpha; a-1, \nu)} \sqrt{s^2 \left( \frac{1}{r_i} + \frac{1}{r_c} \right)},$$

sendo  $r_i$  e  $r_c$  o número de repetições do  $i$ -ésimo tratamento e do tratamento controle, respectivamente.

### 2.2.7 Teste de Scheffé

O teste proposto por Scheffé (1953) permite testar todo e qualquer contraste entre médias de tratamentos, mesmo que este seja escolhido “a posteriori”, ou seja, depois de uma análise prévia dos dados. Assim sendo, o método permite testar hipóteses da forma:

$$H_0: \sum_{i=1}^a c_i \mu_i = 0;$$

$$H_1: \sum_{i=1}^a c_i \mu_i \neq 0,$$

em que  $\mu_i$  representa a média populacional referente ao  $i$ -ésimo tratamento,  $i = 1, \dots, a$ , e  $c_i$  é o coeficiente associado à média  $\mu_i$ .

Para testar cada um dos contrastes, faz-se necessário o cálculo da seguinte estatística:

$$S_\alpha = \sqrt{(a-1)\hat{V}(\hat{Y})F_{(\alpha;a-1,\nu)}},$$

em que  $a$  é o número de tratamentos do ensaio,  $\hat{V}(\hat{Y})$  é a estimativa da variância da estimativa do contraste em questão, como na eq. (2.25), e  $F_{(\alpha;a-1,\nu)}$  é o valor tabelado de  $F$  de Snedecor ao nível de significância  $\alpha$ , segundo  $(a-1)$  e  $\nu$  graus de liberdade associados a tratamentos e ao resíduo, respectivamente.

Há evidências para não aceitar a hipótese de nulidade, a determinado nível de significância, quando

$$|\hat{Y}| > S_\alpha,$$

sendo  $\hat{Y}$  como na expressão (2.24).

### 2.2.8 Teste $t$ (Contrastes ortogonais)

Como requisitos para aplicação do teste  $t$ , um aprimoramento do teste DMS de Fisher, tem-se os seguintes: 1) os contrastes devem ser escolhidos “a priori”, ou seja, antes de serem examinados os dados. Essa primeira exigência se dá devido ao fato de que quando os contrastes são escolhidos depois de uma análise prévia dos dados do experimento, o pesquisador tende a escolher testar contrastes que envolvam as médias de tratamentos que mais diferem entre si, o que pode estar acontecendo por simples acaso; 2) os contrastes devem ser no máximo igual ao

número de graus de liberdade para tratamentos; 3) os contrastes devem ser ortogonais entre si, o que faz com que a variação de cada contraste seja independente da variação do outro.

Diz-se que dois contrastes

$$Y_1 = c_1\mu_1 + c_2\mu_2 + \dots + c_a\mu_a$$

e

$$Y_2 = c'_1\mu_1 + c'_2\mu_2 + \dots + c'_a\mu_a,$$

em que  $\mu_i$  representa a média populacional do  $i$ -ésimo tratamento e  $c_i$  o seu respectivo coeficiente, com  $i = 1, \dots, a$ , são ortogonais entre si se

$$\frac{c_1c'_1}{r_1} + \frac{c_2c'_2}{r_2} + \dots + \frac{c_ac'_a}{r_a} = 0,$$

em que  $r_i$  representa o número de repetições associado ao  $i$ -ésimo tratamento.

Caso o número de repetições para tratamentos seja o mesmo, no entanto, a condição de ortogonalidade entre os contrastes  $Y_1$  e  $Y_2$  se resume ao fato de que

$$c_1c'_1 + c_2c'_2 + \dots + c_ac'_a = 0.$$

As hipóteses a serem verificadas pelo teste  $t$  são como as acuradas pelo procedimento de Scheffé. No entanto, para testar cada um dos contrastes pré-estabelecidos, faz-se necessário o cálculo da seguinte estatística:

$$t = \frac{\hat{Y} - 0}{\sqrt{\hat{V}(\hat{Y})}}$$

em que  $\hat{Y}$  representa a estimativa do contraste em questão, tal como na expressão (2.24), e  $\hat{V}(\hat{Y})$  a estimativa da sua respectiva variância, como em (2.25).

Há evidências para não aceitar a hipótese de nulidade, a determinado nível de significância, quando

$$|t| > t_{(\alpha/2; \nu)},$$

sendo  $t_{(\alpha/2; \nu)}$  o valor tabelado de  $t$  de Student ao nível de significância  $\alpha$  e segundo  $\nu$  graus de liberdade associados ao resíduo.

No entanto, de acordo com Barbin (2003), esse mesmo estudo de contrastes realizado pelo teste  $t$  pode ser realizado, de modo talvez mais elegante, pelo teste  $F$  na própria análise de variância. Para tanto, faz-se necessário decompor os  $(a - 1)$  graus de liberdade associados a



tratamentos, bem como a sua respectiva soma de quadrados, entre cada um dos  $(a - 1)$  contrastes ortogonais, de forma que cada um desses contrastes fique com um grau de liberdade associado, e a sua respectiva soma de quadrados seja da forma:

$$SQY = \frac{(\sum_{i=1}^a r_i c_i \bar{y}_i)^2}{\sum_{i=1}^a r_i c_i^2}.$$

Para testar cada um dos contrastes, faz-se necessário o cálculo da estatística:

$$F = \frac{QMY}{QMRes} = \frac{SQY/1}{QMRes} = \frac{SQY}{QMRes}.$$

Pelo teste  $F$ , há evidências para não aceitar a hipótese nula de que o contraste é igual a zero sempre que

$$F > F_{(\alpha;1,\nu)},$$

em que  $F_{(\alpha;1,\nu)}$  é o valor tabelado de  $F$  de Snedecor ao nível de significância  $\alpha$ , segundo um grau de liberdade associado a tratamentos e  $\nu$  graus de liberdade associados ao resíduo.

Pode-se dizer que o teste  $t$  nada mais é do que uma segunda maneira de usar o teste  $F$ , afinal, o quadrado de uma variável aleatória  $t$ , com  $\nu$  graus de liberdade, é uma variável aleatória  $F$ , com graus de liberdade 1 e  $\nu$ , para o numerador e o denominador, respectivamente.

### 2.2.9 Teste de Bonferroni

O procedimento proposto por Bonferroni (1935, 1936) nada mais é do que um aperfeiçoamento do teste  $t$ . Enquanto neste último a probabilidade de que ao menos um dos contrastes seja significativo por simples acaso é de aproximadamente  $h\alpha$ , sendo  $h$  o número de contrastes a se testar e  $\alpha$  o nível de significância nominal estabelecido para aplicação do método, o nível de significância que passa a ser considerado para testar qualquer um dos contrastes pelo teste de Bonferroni é

$$\alpha_c = \frac{\alpha}{h}.$$

Assim sendo, o nível de significância por experimento passa a ser de aproximadamente  $h\alpha_c = h\alpha/h = \alpha$ , e o problema do teste  $t$  fica então superado.

No entanto, é necessário tomar cuidado com a aplicação do teste. Quanto maior o número de contrastes que se deseja testar, mais rigoroso o procedimento se torna, no sentido de

que o nível de significância por comparação fica muito baixo. Além disso, o teste exige que os contrastes devem ser escolhidos antes de serem examinados os dados. Mas diferentemente do teste *t*, o método de Bonferroni pode ser aplicado a qualquer número de contrastes, sem que haja a exigência de ortogonalidade nem de um número máximo pré-determinado de contrastes.

### 2.2.10 Escolha do procedimento adequado

Devido a variedade de testes de comparação de médias disponíveis e da dificuldade encontrada pelos pesquisadores a respeito de qual o melhor procedimento a ser utilizado, Cardellino e Siewerdt (1992) sugeriram uma chave para determinar o melhor teste de acordo com o tipo de dados disponíveis e também de acordo com os objetivos da pesquisa, de forma a minimizar os equívocos decorrentes da escolha equivocada do teste, o que pode levar o pesquisador a obter conclusões não fidedignas a respeito dos tratamentos investigados, bem como a fazer inferências errôneas sobre eles (Tabela 2.7).

Tabela 2.7 – Chave para escolha do teste de comparação de médias

1. Tratamentos são níveis de um fator quantitativo	Sim	.....2
	Não	.....3
2. Ajustar uma equação de regressão	Sim	.....Regressão
	Não	.....Williams <sup>1</sup>
3. Contrastes definidos pela inspeção dos dados	Sim	.....Scheffé
	Não	.....4
4. Contrastes ortogonais	Sim	.....Teste <i>t</i> ou teste <i>F</i>
	Não	.....5
5. Comparar todos os pares de médias	Sim	.....DMS de Fisher, Tukey, Duncan ou SNK
	Não	.....6
6. Comparação somente com o controle	Sim	.....Dunnett
	Não	.....Bonferroni

Fonte: Adaptado de Cardellino e Siewerdt (1992).

<sup>1</sup> O teste de Williams consiste na comparação de diferentes níveis quantitativos com um nível controle. O objetivo desse teste é detectar qual a menor dose que difere significativamente da dose controle (CARDELLINO; SIEWERDT, 1992, p.993).

### 2.3 Região de inferência do teste $F$ global versus região de inferência do teste de Tukey

De acordo com Hsu (1996), a aplicação condicional de testes de comparação de médias a um resultado preliminar significativo do teste  $F$  global é uma prática infeliz, baseada nas crenças equivocadas de que, segundo ele, uma “proteção mágica” é envolvida fazendo-se primeiro um teste de igualdade entre as médias populacionais dos tratamentos comparados e constatando-se evidências de diferença entre elas, e, ainda, na de que nenhum resultado útil é encontrado se a hipótese nula de homogeneidade é aceita.

Ainda, de acordo com o autor, a exigência prévia de um resultado significativo do teste  $F$  global para aplicação de um teste de médias altera o nível de significância real ao qual o teste de médias é submetido, que se torna maior do que o nível nominal declarado, devido à condicionalidade imposta. Como exemplo a esse fato, o autor cita o trabalho de Olshen (1973), o qual também se mostra contra a exigência prévia de um resultado significativo do teste  $F$  para aplicação de um teste de comparação de médias de tratamentos.

Olshen (1973) provou que se o teste de Scheffé é aplicado apenas mediante um resultado significativo do teste  $F$  global, então o nível de significância real ao qual o teste é submetido fica alterado, maior do que o nível de significância nominal imposto. Em outras palavras, de acordo com Hsu (1996), dentre os experimentos em que testes de comparação de médias são efetivamente realizados, a proporção de experimentos para os quais algumas das afirmações feitas estão equivocadas pode ser muito mais elevada do que a taxa de erro nominal ao qual o método é submetido.

Hsu (1996) coloca ainda um exemplo em que ilustra que é possível obter um teste  $F$  significativo e nenhum contraste entre duas médias de tratamentos significativo pelo método de Scheffé, e, segundo o autor, o mesmo pode ocorrer para qualquer outro método que não este quando aplicado diante da não aceitação da hipótese de homogeneidade entre as médias populacionais de tratamentos. Reciprocamente, de acordo com ele, é possível que o teste  $F$  aceite a hipótese de homogeneidade mas um método de comparação detecte diferenças significativas entre as médias comparadas, afinal, todos os testes estatísticos possuem probabilidades de erros associadas e, assim, é possível ocorrer divergências entre o resultado do teste  $F$  global e o resultado dos testes de médias aplicados posteriormente.

Para ilustrar, o autor apresenta um exemplo a demonstrar que a região de inferência do teste  $F$  global não é necessariamente a mesma que a região de inferência de um teste de comparação de médias. Para tanto, supôs-se três tratamentos de interesse, cada um deles

repetido um mesmo número  $r$  de vezes, e o teste de comparação de médias como sendo o de Tukey. A região de aceitação da hipótese de homogeneidade pelo teste  $F$  global, ao nível de significância  $\alpha$ , é da seguinte forma:

$$(\bar{y}_1 - \bar{y})^2 + (\bar{y}_2 - \bar{y})^2 + (\bar{y}_3 - \bar{y})^2 \leq F_{(\alpha;2,\nu)} \hat{\sigma}^2 \left( \frac{2}{r} \right),$$

a qual representa um cilindro infinito com uma seção transversal circular centrada no vetor  $\{(\delta, \delta, \delta) : \delta \in \mathbb{R}\}$ , como na Figura 2.1, em que  $\bar{y}_i$  representa a média amostral associada ao  $i$ -ésimo tratamento,  $\bar{y}$  representa a estimativa da média geral,  $F_{(\alpha;2,\nu)}$  é o valor tabelado de  $F$  de Snedecor ao nível de significância  $\alpha$  e segundo 2 graus de liberdade associados à tratamentos e  $\nu$  graus de liberdade associados ao resíduo, e  $\hat{\sigma}^2$  representa a estimativa obtida para a variância residual, considerada aqui igual para todos os tratamentos investigados, a qual pode ser dada pelo quadrado médio do resíduo obtido na análise de variância.

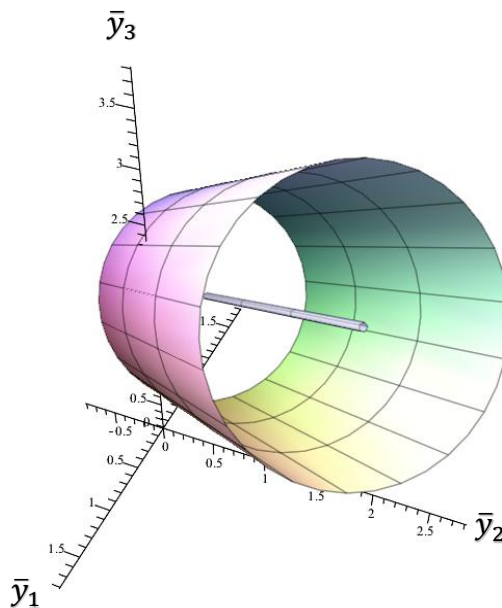


Figura 2.1 – Região de aceitação da hipótese nula de homogeneidade entre as médias populacionais dos três tratamentos comparados pelo teste  $F$  global

Para chegar à região obtida, basta realizar o seguinte processo: na análise de variância do experimento realizado nas condições descritas anteriormente, vão existir evidências para aceitar a hipótese de homogeneidade entre as médias populacionais dos três tratamentos comparados, ao nível de significância estabelecido, se

$$F_0 = \frac{QMTrat}{QMRes} \leq F_{(\alpha;2,\nu)}.$$

É sabido que o quadrado médio residual é um estimador não viciado para  $\sigma^2$  e, ainda, que

$$QMTrat = \frac{SQTrat}{glTrat} = \frac{r[(\bar{y}_1 - \bar{y})^2 + (\bar{y}_2 - \bar{y})^2 + (\bar{y}_3 - \bar{y})^2]}{2},$$

ou seja, há evidências para aceitar a hipótese de homogeneidade se

$$\frac{r[(\bar{y}_1 - \bar{y})^2 + (\bar{y}_2 - \bar{y})^2 + (\bar{y}_3 - \bar{y})^2]}{2} \leq \frac{F_{(\alpha; 2, \nu)} \hat{\sigma}^2}{2}.$$

Desenvolvendo a inequação acima, tem-se que a região de aceitação da hipótese de homogeneidade entre as médias populacionais dos três tratamentos comparados pelo teste  $F$ , ao nível de significância  $\alpha$ , é exatamente como a colocada pelo autor.

Já os intervalos de confiança de Tukey, ao nível de confiança de  $100(1 - \alpha)\%$ , para os contrastes de interesse, ou seja,  $\mu_1 - \mu_2$ ,  $\mu_1 - \mu_3$  e  $\mu_2 - \mu_3$ , irão todos conter o zero se e somente se:

$$\max |\bar{y}_i - \bar{y}_j|_{i \neq j} \leq q_{(\alpha; 3, \nu)} \sqrt{\frac{QMResíduo}{r}},$$

ou seja, se o máximo das diferenças, em módulo, entre as estimativas das médias comparadas entre si duas a duas não for superior a diferença mínima significante calculada pelo método, em que  $q_{(\alpha; 3, \nu)}$  é a amplitude total estudentizada para o teste de Tukey com nível de significância  $\alpha$ , 3 tratamentos envolvidos e  $\nu$  graus de liberdade associados ao resíduo, sendo que a região obtida representa um cilindro infinito com uma seção transversal hexagonal também centrada no vetor  $\{(\delta, \delta, \delta): \delta \in \mathbb{R}\}$ , como na Figura 2.2.

Uma visão da sobreposição dos cilindros, paralela ao vetor  $\{(\delta, \delta, \delta): \delta \in \mathbb{R}\}$ , está apresentada pela Figura 2.3. Nessa sobreposição fica evidente que se  $(\bar{y}_1, \bar{y}_2, \bar{y}_3)$  é uma “esquina” do hexágono, fora do círculo, então a hipótese nula de homogeneidade  $H_0: \mu_1 = \mu_2 = \mu_3$  não é aceita pelo teste  $F$  global, mas os intervalos de confiança para  $\mu_1 - \mu_2$ ,  $\mu_1 - \mu_3$  e  $\mu_2 - \mu_3$  todos conterão o zero. Por outro lado, se  $(\bar{y}_1, \bar{y}_2, \bar{y}_3)$  está no círculo mas fora do hexágono, então a hipótese nula de homogeneidade será aceita pelo teste  $F$ , mas ao menos um dos intervalos de confiança para os contrastes considerados não conterá o zero.

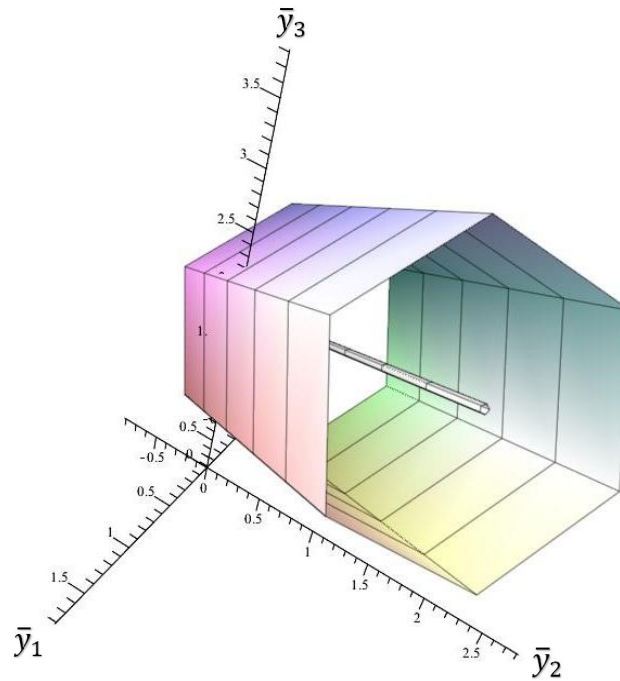


Figura 2.2 – Região de aceitação das três hipóteses de homogeneidade entre as médias populacionais de tratamentos quando comparadas duas a duas pelo teste de Tukey

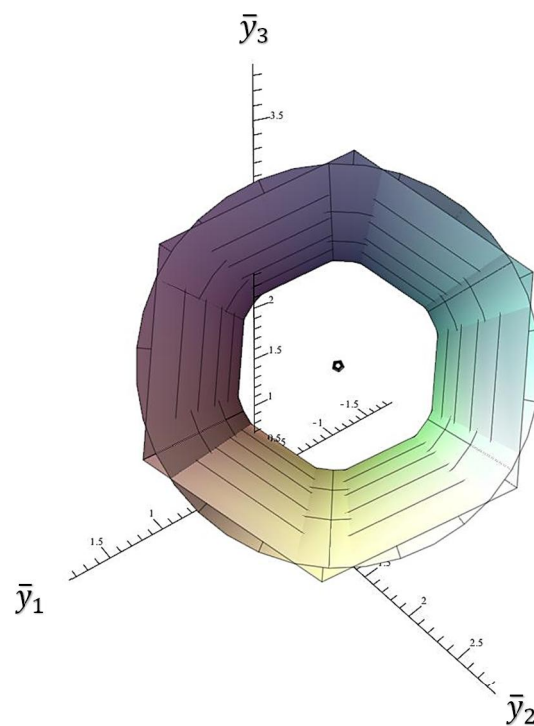


Figura 2.3 – Região de inferência do teste  $F$  versus região de inferência do teste de Tukey

Em suma, o que o autor busca dizer é que considerar procedimentos de comparação de médias apenas se o teste  $F$  global na análise de variância não aceita a hipótese nula de homogeneidade entre as médias populacionais dos tratamentos comparados é um equívoco e deve, portanto, ser reconsiderada.

### 3 MATERIAL E MÉTODOS

#### 3.1 Comportamento dos testes de comparação de médias em relação as situações de divergência

Para alcançar o primeiro objetivo proposto, comparou-se o resultado do teste  $F$  global da análise de variância com o resultado de cada um dos seguintes testes para comparação de médias de tratamentos entre si duas a duas: Tukey, Duncan, DMS de Fisher, SNK e Scheffé.

Foram simulados, via método de Monte Carlo, 256000 experimentos, 2000 para cada cenário, sendo estes num total de 128 casos, formados pela combinação dos seguintes fatores: 3, 5, 7 e 9 tratamentos, 3, 4, 10 e 20 repetições, coeficiente de variação de 1%, 5%, 10% e 20% e duas combinações dos efeitos de tratamentos. Os experimentos foram simulados de acordo com o modelo estatístico referente ao delineamento inteiramente casualizado, da seguinte forma:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij},$$

em que  $y_{ij}$  representa o valor simulado da resposta obtida no  $i$ -ésimo tratamento em sua  $j$ -ésima repetição,  $\mu$  é a média geral, arbitrada como sendo 100,  $\tau_i$  é o efeito fixo do  $i$ -ésimo tratamento ( $\sum_i \tau_i = 0$ ), e  $\varepsilon_{ij}$  é o erro aleatório, gerado independentemente com distribuição normal de média zero e desvio padrão ( $\sigma$ ) variando conforme o coeficiente de variação desejado.

Nesse contexto, para cada um dos experimentos simulados, foram exploradas as seguintes possibilidades para cada um dos testes de comparações múltiplas:

- Situação 1: teste  $F$  não significativo e nenhum dos contrastes entre duas médias de tratamentos significativo;
- Situação 2: teste  $F$  global não significativo e ao menos um contraste entre duas médias significativo;
- Situação 3: teste  $F$  significativo e ao menos um dos contrastes entre duas médias de tratamentos também significativo;
- Situação 4: teste  $F$  global significativo e nenhum contraste entre duas médias significativo.

As situações 2 e 4 foram classificadas como sendo de divergência, pois apontam resultados contraditórios entre o teste  $F$  global e o teste de comparação múltipla em avaliação.



O objetivo desse estudo inicial foi verificar o comportamento de cada um dos procedimentos de comparação de médias com relação a essas duas situações.

De acordo com Montgomery (2001), a aplicação válida dos testes de significância na análise de variância requer que certos pressupostos sejam satisfeitos: os erros do modelo devem ser normalmente e independentemente distribuídos com média zero e variância comum,  $\sigma^2$ . Ainda, Anscombe e Tukey (1963) mostraram que os testes para verificar homogeneidade de variâncias e normalidade dos erros, por exemplo, são sensíveis à presença de dados discrepantes e, portanto, qualquer desses procedimentos pode ser usado para verificar a presença de *outliers* em conjuntos de dados.

Assim sendo, os testes de Shapiro-Wilk e Levene foram implementados para verificar as pressuposições de normalidade e homocedasticidade dos erros, respectivamente, além da ausência de dados discrepantes, em cada um dos experimentos simulados. A independência dos erros, por sua vez, ficou garantida pelo princípio da casualização, implementado durante o processo de simulação. Essa análise foi realizada com o intuito de verificar se o aparecimento de situações de divergência entre o resultado de cada um dos procedimentos de comparação de médias e o resultado do teste  $F$  global pode estar associado ao desvio de ao menos uma das pressuposições da análise.

Tomou-se o cuidado de, para todos os cenários, realizar as análises com base na mesma semente aleatória, de modo a assegurar que possíveis diferenças não ocorressem devido ao erro aleatório do processo de simulação, mas sim devido às diferenças entre os testes de comparação de médias considerados. Ainda, o nível de significância nominal adotado em todos os casos foi o de 5%. Para simular os dados experimentais e realizar as análises estatísticas foi desenvolvido um algoritmo usando o *software* R (2013).

### **3.2 Níveis de significância por comparação e por experimento**

Tendo-se realizado essa análise inicial, as taxas de erro tipo I por comparação e por experimento foram então estimadas empiricamente usando do mesmo processo de simulação, para os mesmos cinco procedimentos de comparação de médias considerados no estudo. Ainda, essas taxas de erro foram consideradas de duas formas distintas: i) quando o procedimento de comparação múltipla foi aplicado independentemente do resultado do teste  $F$  global; ii) quando o mesmo foi aplicado apenas se o teste  $F$  da análise de variância foi significativo.

Foram então simulados mais 128000 experimentos (64 cenários x 2000 experimentos por cenário), agora sem considerar o efeito de tratamento, ou seja, em situação de nulidade

completa ( $\tau_1 = \tau_2 = \dots = \tau_a = 0$ ). Os cenários foram formados pela combinação dos seguintes fatores: 3, 5, 7 e 9 tratamentos, 3, 4, 10 e 20 repetições e coeficiente de variação de 1%, 5%, 10% e 20%.

Em cada um dos cenários simulados foram então estimados os níveis de significância por comparação condicional e incondicional e também os níveis de significância por experimento condicional e incondicional de cada um dos seguintes procedimentos para a comparação de médias duas a duas: Tukey, Duncan, DMS de Fisher, SNK e Scheffé.

Boardman e Moffitt (1971) definiram o nível de significância por comparação ( $\alpha_c$ ) como sendo a proporção do número de inferências errôneas (concluir que  $\mu_i \neq \mu_{i'}$  quando  $\mu_i = \mu_{i'}$ ), dividido pelo número total de comparações realizadas. Assim, tomando o cenário demonstrativo de  $a = 3$ ,  $r = 3$  e  $CV = 1\%$ , por exemplo, o nível de significância por comparação incondicional pode ser estimado por meio da razão entre o número total de inferências errôneas e o número total de inferências (nesse caso, 2000 experimentos x 3 contrastes por experimento = 6000 contrastes).

Já quando os testes de comparação de médias são aplicados apenas mediante um resultado preliminar significativo do teste  $F$  global, o nível de significância por comparação, agora denominado condicional ( $\alpha_1$ ), pode ser estimado empiricamente tomando o número total de erros tipo I cometidos nas comparações das  $a$  médias duas a duas, dentre os experimentos que apresentaram resultado significativo para o teste  $F$ , e o número total de inferências realizadas (BERNHARDSON, 1975).

Ainda, uma segunda maneira de estimar o nível de significância por comparação condicional ( $\alpha_2$ ) é tomar o número total de inferências errôneas dentro dos experimentos com resultado significativo para o teste  $F$  global e dividi-lo pelo número total de inferências realizadas dentro desses ensaios (BERNHARDSON, 1975).

Boardman e Moffitt (1971) definiram ainda o nível de significância por experimento ( $\alpha_e$ ) como sendo a proporção do número de experimentos com no mínimo uma inferência errônea ( $\mu_i \neq \mu_{i'}$  quando  $\mu_i = \mu_{i'}$ ) dividido pelo número total de experimentos. Assim, tomando o mesmo cenário demonstrativo de  $a = 3$ ,  $r = 3$  e  $CV = 1\%$ , o nível de significância por experimento incondicional pode ser estimado por meio da razão entre o número de ensaios com no mínimo uma inferência errada entre as 3 testadas e o número total de experimentos (nesse caso, 2000).

O nível de significância por experimento para a combinação do teste  $F$  da análise de variância e os testes de comparação de médias considerados, ao qual chamar-se-á nível de

significância por experimento condicional ( $\alpha_3$ ), pode ser estimado tomando o número total de ensaios que apresentaram tanto um teste  $F$  global significativo e no mínimo uma comparação resultando em erro tipo I, dividido pelo número total de experimentos (BERNHARDSON, 1975).

O mesmo nível de significância pode ainda ser estimado baseado apenas nos experimentos que apresentaram um resultado significativo para o teste  $F$  global. Este nível ( $\alpha_4$ ) pode ser então estimado dividindo o número total de ensaios que apresentaram tanto um teste  $F$  significativo e no mínimo uma comparação derivando em erro tipo I, pelo número total de ensaios dentre os 2000 com o teste  $F$  global significativo (BERNHARDSON, 1975).

De forma a verificar se cada um dos níveis de significância estimados diferiu do nível de significância nominal estabelecido ( $\alpha = 5\%$ ), foi utilizado o limite inferior de 0,0375 e o limite superior de 0,0625, obtidos com base no intervalo de confiança (IC) exato de 95% para uma proporção  $\hat{p} = 0,05$ , expresso da seguinte maneira:

$$IC = \hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{2000}},$$

em que  $z_{\alpha/2}$  é o valor tabelado da distribuição normal padrão, ao nível de significância  $\alpha$ . Deste modo, níveis de significância dentro desse intervalo não foram considerados como diferentes do nível nominal estabelecido.

### 3.3 Poder

Por fim, o poder dos cinco procedimentos de comparação de médias considerados foi então avaliado pelo estudo de simulações de Monte Carlo. Em cada um dos cenários a configuração igualmente espaçada das  $a$  médias populacionais dos tratamentos comparados, citada por Einot e Gabriel (1975), Rafter, Abell e Braselton (2002) e Girardi, Cargnelutti Filho e Storck (2009), foi utilizada para o estudo do poder dos testes ( $\mu_1 < \mu_2 < \dots < \mu_{a-1} < \mu_a$ ).

Essa configuração das médias dos tratamentos define que todos os contrastes para a comparação de médias duas a duas apresentam uma diferença que é testada por meio do procedimento de comparação de médias em avaliação. Assim sendo, ela é adequada para avaliar o poder do teste, uma vez que toda e qualquer diferença acusada como significativa pelo procedimento deve-se as diferenças de fato existentes entre as médias populacionais dos tratamentos (GIRARDI; CARGNELUTTI FILHO; STORCK, 2009).

Para esse estudo do poder foram então simulados mais 32000 experimentos (16 cenários x 2000 experimentos por cenário), com diferentes efeitos de tratamentos definidos segundo a configuração igualmente espaçada, que atribui para cada média um incremento fixo (número de erro padrão da média) em relação a média anterior ( $\tau_1 < \tau_2 < \dots < \tau_{a-1} < \tau_a$ ). Dessa forma buscou-se criar um ambiente propício para o estudo do poder de cada um dos cinco procedimentos de comparação de médias considerados, pois a situação de diferença entre as médias dos tratamentos foi simulada para ser verdadeira, respeitando a restrição de que  $\sum_i \tau_i = 0$ .

Os 16 cenários foram formados pela combinação entre o número de tratamentos ( $a = 3, 5, 7$  e  $9$ ) e o número de erro padrão da média definido por  $\sigma_\mu = \frac{\sigma}{\sqrt{r}}$  ( $\sigma_\mu = 1, 2, 4$  e  $8$ ) mantendo-se fixo  $r = 20$  e  $CV = 5\%$ . A fixação desses dois fatores se deve ao fato dos efeitos de tratamentos estarem vinculados ao erro padrão da média, o que não altera o poder em diferentes precisões experimentais (SILVA; FERREIRA; BEARZOTI, 1999).

O poder incondicional e condicional de cada um dos cinco testes foi então estimado em cada um dos cenários simulados. Tomando o cenário demonstrativo de  $a = 3$ ,  $\sigma_\mu = 1$ ,  $r = 20$  e  $CV = 5\%$ , por exemplo, o erro padrão da média é  $\sigma_\mu = \frac{5}{\sqrt{20}} \cong 1,12$  e os efeitos de tratamentos com base em  $1\sigma_\mu$  são  $\tau_1 = -1,12$ ,  $\tau_2 = 0$  e  $\tau_3 = 1,12$ . O poder incondicional ( $\pi$ ) do teste foi então estimado pela razão entre o número total de inferências corretas nos 2000 experimentos (concluir que  $\mu_i \neq \mu_j$ , quando de fato  $\mu_i \neq \mu_j$ ) e o número total de inferências (nesse caso, 2000 experimentos x 3 contrastes por experimento = 6000 contrastes).

O poder condicional ( $\pi_1$ ) de cada um dos procedimentos, por sua vez, pode ser estimado em cada um dos cenários simulados por meio da razão entre o número total de inferências corretas nos experimentos que apresentaram resultado significativo para o teste  $F$  global e o número total de inferências realizadas nos 2000 experimentos. O mesmo poder pode, ainda, ser estimado apenas com base nos ensaios com resultado significativo para o teste  $F$ . Esse poder condicional ( $\pi_2$ ) é determinado a partir do número total de inferências corretas dentre os experimentos com teste  $F$  global significativa na análise de variância dividido pelo número total de inferências realizadas dentro desses ensaios.

De acordo com Girardi, Cargnelutti Filho e Storck (2009) é importante que procedimentos de comparação de médias apresentem níveis de significância próximos ao nível de significância nominal pré-estabelecido e que o poder dos testes seja alto (o mais próximo possível de 1), o que será investigado no decorrer do trabalho. Ainda, com o estudo dos níveis de significância e respectivo poder de cada um dos testes de médias considerados buscar-se-á

verificar se existe alguma relação entre o aparecimento de alguma das situações de divergência ao comparar os resultados desses procedimentos com o resultado do teste  $F$  global e o fato dos mesmos não controlarem alguma das taxas de erro tipo I ou possuírem um baixo poder. Além disso, esse estudo permitirá concluir se a aplicação condicional dos testes altera o nível de significância nominal ao qual o método é submetido, bem como se modifica o poder dos testes.

## 4 RESULTADOS

### 4.1 Experimentos com resultados contraditórios

Para mostrar que pode ocorrer a existência de contradição entre o resultado do teste  $F$  global na análise de variância e os resultados de procedimentos de comparação de médias, em particular os que foram considerados no presente estudo, segue na Tabela 4.1 os dados de um experimento simulado pelo método de Monte Carlo no delineamento inteiramente casualizado, em que foram considerados 5 tratamentos, cada um deles com 4 repetições e coeficiente de variação de 5%.

Tabela 4.1 – Resultados de um experimento simulado pelo método de Monte Carlo no delineamento inteiramente casualizado, com 5 tratamentos, 4 repetições e coeficiente de variação de 5%

Repetição	Tratamento				
	V1	V2	V3	V4	V5
1	127	112	93	104	97
2	114	120	111	113	104
3	90	109	91	100	106
4	105	133	85	111	94

Antes de proceder à análise de variância dos dados do ensaio, inicialmente foi realizada uma análise gráfica com base nos resíduos padronizados do modelo estatístico referente ao delineamento inteiramente casualizado, de forma a verificar se cada uma das pressuposições para a realização do procedimento estava sendo atendida (Figura 4.1).

Ainda que os gráficos indiquem que as pressuposições da análise de variância estejam satisfeitas, os testes de Levene e Shapiro-Wilk foram aplicados para verificar as pressuposições de homocedasticidade e normalidade dos erros, respectivamente. Como ambos os procedimentos são sensíveis à presença de dados discrepantes (ANSCOMBE; TUKEY, 1963), eles também foram usados para verificar a presença de *outliers* no conjunto de dados. A independência, por sua vez, ficou garantida pelo princípio da casualização (DEMÉTRIO, 1978), implementada durante o processo de simulação. Os resultados encontrados confirmam as suspeitas adquiridas com base na análise gráfica, ao nível de significância de 5% (Tabela 4.2).

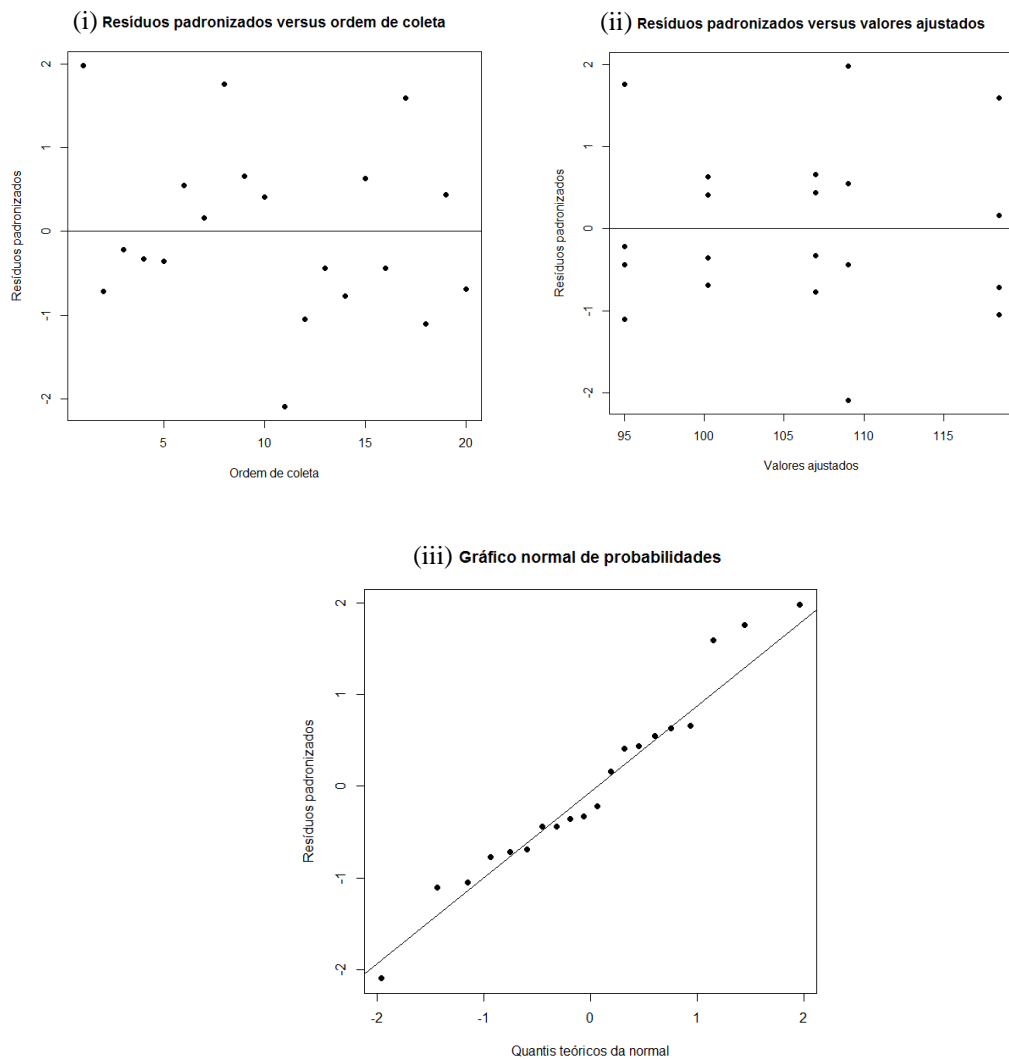


Figura 4.1 – Gráficos para verificar (i) Independência dos erros e presença de dados discrepantes; (ii) Homogeneidade de variâncias; (iii) Normalidade dos erros

Tabela 4.2 – Resultados obtidos com a aplicação dos testes de Levene e Shapiro-Wilk para os dados da Tabela 4.1, ao nível de significância de 5%

Teste	Valor da estatística	Valor de $p$
Levene	0,8246	0,5297
Shapiro-Wilk	0,9613	0,5702

A análise de variância referente aos dados do experimento, exibida na Tabela 4.3, apresenta o teste  $F$  global não significativo ao nível de 5%, ou seja, não há evidências para não aceitar a hipótese nula de homogeneidade entre as médias populacionais dos tratamentos comparados.

Tabela 4.3 – Análise de variância referente aos dados da Tabela 4.1, ao nível de significância de 5%

Causas de variação	Graus de liberdade	Somas de quadrados	Quadrados médios	$F$	Valor de $p$
Tratamentos	3	0,05	0,01667	0,392	0,0579
Resíduo	12	0,51	0,04250		
Total	15	0,56			

Ainda assim, cada um dos seguintes testes para a comparação das médias de tratamentos entre si duas a duas foi aplicado: Tukey, Duncan, DMS de Fisher, SNK e Scheffé. Os resultados obtidos com a aplicação dos procedimentos, ainda ao nível de 5%, estão apresentados na Tabela 4.4.

Tabela 4.4 – Resultados obtidos a partir da aplicação dos testes de médias para os dados da Tabela 4.1

Médias amostrais	Tukey	Duncan	DMS de Fisher	SNK	Scheffé
V2 – 118,5	a	a	a	a	a
V1 – 109	ab	ab	ab	ab	a
V4 – 107	ab	ab	ab	ab	a
V5 – 100,2	ab	b	b	ab	a
V3 – 95	b	b	b	b	a

Médias seguidas pela mesma letra não diferem entre si, ao nível de significância de 5%.

É possível notar que os testes de Tukey, Duncan, DMS de Fisher e SNK identificaram diferenças significativas entre um, dois, dois e um par de médias, respectivamente, indo contra o resultado obtido a partir da aplicação do teste  $F$  global, de que todas as médias populacionais de tratamentos não diferem entre si, ao mesmo nível de significância em que foram aplicados os procedimentos para a comparação das médias.

O teste de Scheffé, ao contrário, não encontrou diferença que pudesse ser considerada significativa entre as médias dos tratamentos quando comparadas duas a duas. No entanto, talvez algum contraste que não foi testado poderia ser significativo nesse caso, uma vez que o método é mais geral que os demais e pode ser utilizado para testar contrastes mais complicados, que não sejam aqueles para a comparação das médias pareadas.

Os resultados de um segundo experimento simulado pelo método de Monte Carlo no delineamento inteiramente casualizado, ainda com 5 tratamentos, cada um deles com 4 repetições, e coeficiente de variação de 10%, estão apresentados na Tabela 4.5.



Tabela 4.5 – Resultados de um experimento simulado pelo método de Monte Carlo no delineamento inteiramente casualizado, com 5 tratamentos, 4 repetições e coeficiente de variação de 10%

Repetição	Tratamento				
	V1	V2	V3	V4	V5
1	102	92	88	106	101
2	108	104	84	114	93
3	102	100	103	108	83
4	91	110	92	100	93

Assim como no caso do primeiro experimento simulado, as pressuposições de ausência de dados discrepantes, independência, homocedasticidade e normalidade dos erros foram checadas antes de proceder à análise de variância dos dados do ensaio (Figura 4.2 e Tabela 4.6).

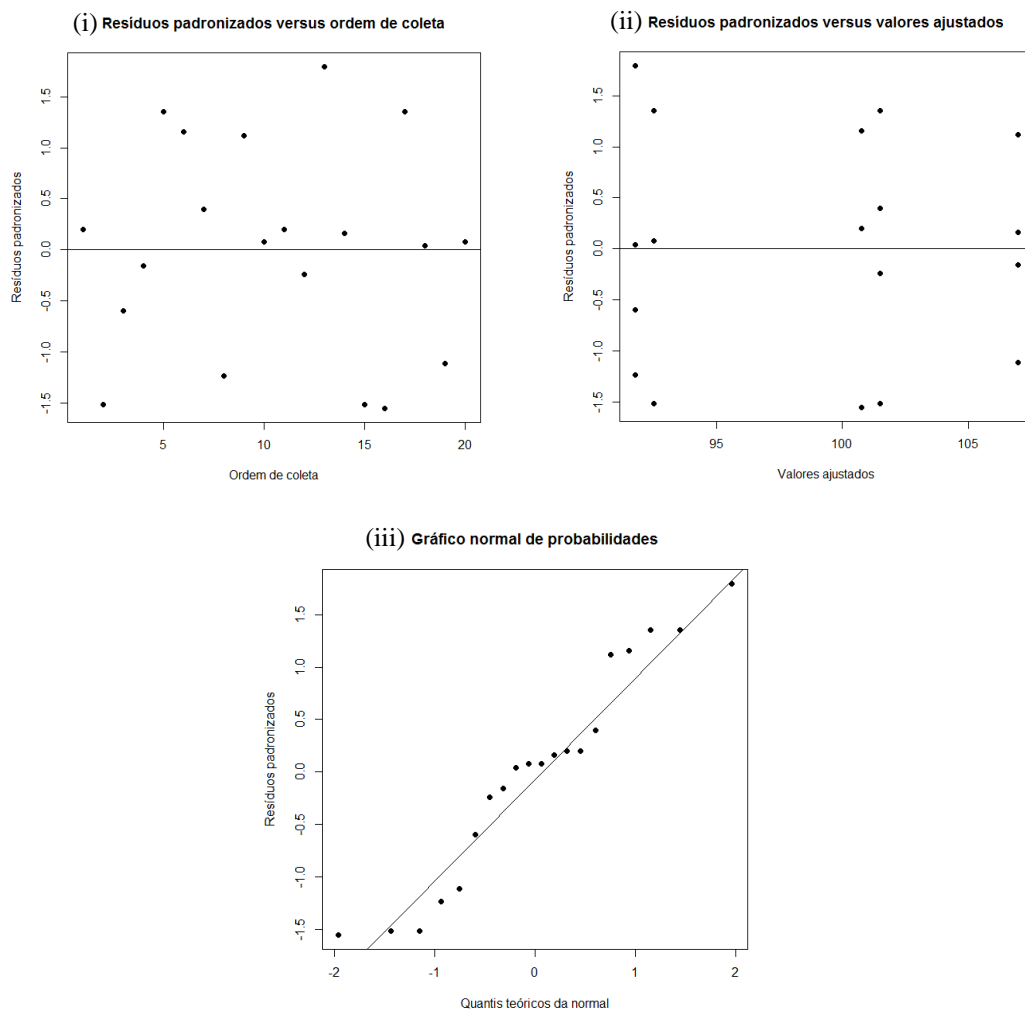


Figura 4.2 – Gráficos para verificar (i) Independência dos erros e presença de dados discrepantes; (ii) Homogeneidade de variâncias; (iii) Normalidade dos erros

Tabela 4.6 – Resultados da aplicação dos testes de Levene e Shapiro-Wilk para os dados da Tabela 4.5, ao nível de significância de 5%

Teste	Valor da estatística	Valor de $p$
Levene	0,1084	0,9777
Shapiro-Wilk	0,9367	0,2078

Como pode ser verificado, cada uma das pressuposições da análise de variância está sendo satisfeita e, portanto, a mesma foi aplicada sem a necessidade de se realizar transformações no conjunto de dados original (Tabela 4.7).

Tabela 4.7 – Análise de variância referente aos dados da Tabela 4.5

Causas de variação	Graus de liberdade	Somas de quadrados	Quadrados médios	$F$	Valor de $p$
Tratamentos	4	670,7	167,67	3,202*	0,0435*
Resíduo	15	785,5	52,37		
Total	19	1456,2			

\* Significativo ao nível de 5%.

Ao nível de significância de 5%, o teste  $F$  global foi significativo e, portanto, há evidências para não aceitar a hipótese nula de igualdade entre as médias populacionais dos tratamentos comparados. Os testes de Tukey, Duncan, DMS de Fisher, SNK e Scheffé para a comparação das médias dos tratamentos duas a duas foram então aplicados de modo a confrontar os resultados obtidos em cada um deles com o resultado do teste  $F$  (Tabela 4.8).

Tabela 4.8 – Resultados obtidos a partir da aplicação dos testes de médias para os dados da Tabela 4.5

Médias amostrais	Tukey	Duncan	DMS de Fisher	SNK	Scheffé
V4 – 107	A	a	a	a	a
V2 – 101,5	A	ab	ab	a	a
V1 – 100,8	A	ab	ab	a	a
V5 – 92,5	A	b	b	a	a
V3 – 91,75	A	b	b	a	a

Médias seguidas pela mesma letra não diferem entre si, ao nível de significância de 5%.

Os testes de Tukey, SNK e Scheffé não encontraram diferença que pudesse ser considerada significativa ao comparar as médias de tratamentos entre si, indo contra o resultado do teste  $F$  global na análise de variância, que não aceitou a hipótese de igualdade entre as médias

populacionais dos tratamentos, ao mesmo nível de probabilidade. Os testes de Duncan e DMS de Fisher, ao contrário, detectaram diferenças significativas entre dois pares de médias, mostrando assim coerência com o resultado do teste  $F$ .

Com base nos experimentos simulados é possível afirmar que podem ocorrer situações em que o teste  $F$  global da análise de variância apresente resultado não significativo e, ainda assim, alguns testes de médias detectem diferenças significativas entre médias de tratamentos quando comparadas entre si e, também, pode ocorrer o caso em que o teste  $F$  global apresente resultado significativo e alguns testes de médias aplicados posteriormente não encontrem diferenças entre as médias populacionais dos tratamentos se comparadas duas a duas, ao mesmo nível de probabilidade em que foi realizado o teste  $F$ , e em condição de ausência de pontos discrepantes, independência, homogeneidade de variâncias e normalidade dos erros.

#### **4.2 Comportamento dos testes de comparação de médias em relação as situações de divergência**

Por meio do estudo de simulações de Monte Carlo, o objetivo da presente seção foi comparar o resultado do teste  $F$  global da análise de variância com o resultado de cada um dos seguintes testes para comparação de médias de tratamentos entre si duas a duas: Tukey, Duncan, DMS de Fisher, SNK e Scheffé, bem como verificar se a existência de contradição entre o resultado do teste  $F$  e os resultados desses procedimentos pode estar associada ao desvio de ao menos uma das pressuposições da análise de variância.

Nesse contexto, a análise foi conduzida da seguinte forma: para cada um dos testes de comparações múltiplas foi verificada a respectiva porcentagem de experimentos, por cenário, que apresentaram as situações de divergência 2 ou 4 (ANEXO A a ANEXO H). Uma análise gráfica foi então realizada para facilitar a visualização dos resultados obtidos (ANEXO D1 a ANEXO G2). Essa análise permitiu verificar se o número de experimentos com resultados contraditórios foi ou não significativo em cada um dos cenários, bem como possibilitou compreender se a variação de cada um dos fatores que compõe os mesmos, ou seja, número de tratamentos, número de repetições e coeficiente de variação dos ensaios, afetou ou não o número de experimentos com resultados contraditórios encontrados.

Ainda, para cada um dos testes, foi apurada a porcentagem de experimentos, por cenário, dentre os que exibiram a situação de divergência 2 ou 4, que apresentaram resultado não significativo e significativo para os testes de Levene e Shapiro-Wilk, respectivamente, ao nível de significância de 5% (ANEXO I a ANEXO P).

Para o procedimento de Tukey, ambas as situações de contradição, ou seja, 2 e 4, estiveram presentes nos resultados, mas apenas em um pequeno número dos experimentos simulados na grande maioria dos cenários. À medida que o número de tratamentos dos cenários aumentou, pode-se notar que o número de experimentos com resultados contraditórios também sofreu acréscimos, mas ainda se manteve baixo na maioria dos casos (ANEXOS D1 e G1). Já o número de repetições para tratamentos, quando alterado, mostrou não provocar mudanças que pudessem ser consideradas expressivas nos resultados (ANEXOS E1 e H1).

Por sua vez, o aumento do coeficiente de variação de 1% para 5% levou ao acréscimo no número de experimentos nas situações de divergência 2 ou 4. Quando este passou de 5% para 10%, houve uma diminuição no número de experimentos com esses resultados contraditórios e, quando este coeficiente passou de 10% para 20%, esse número decresceu ainda mais (ANEXOS F1 e I1). Na maioria dos cenários simulados, entretanto, o coeficiente de variação de 1% foi o que levou ao menor número de ensaios por cenário com alguma das situações de divergência.

Para o teste de Duncan, também ambas as situações de divergência foram encontradas nos resultados obtidos. A situação 2 sempre apareceu em um número bem maior de experimentos por cenário do que o encontrado para o método de Tukey. Ainda ficou evidente que, conforme o número de tratamentos dos cenários aumentou, maior o número de experimentos que apresentou esse resultado, chegando a atingir 865 (43,25%) dos experimentos num dos cenários simulados (ANEXO J1).

Por sua vez, o número de repetições para tratamentos, quando alterado, não levou a alterações significativas nos resultados para a situação de divergência 2 (ANEXO K1).

Já a situação de divergência 4 não se mostrou presente em nenhum dos cenários simulados com 3 e 5 tratamentos, e em apenas um experimento dos cenários com 7 tratamentos. Já para o caso com 9 tratamentos, essa situação apareceu num número bastante baixo de experimentos por cenário, numa porcentagem sempre muito próxima de zero (ANEXO M1).

Por sua vez, quando o número de repetições para tratamentos subiu de 3 para 4, houve uma pequena diminuição no número de ensaios por cenário com o resultado contraditório 4 e, quando as repetições passaram de 4 para 10, esse número apresentou um pequeno aumento, ficando pouco maior quando as repetições passaram para 20 (ANEXO N1). Esse padrão pode ser notado em apenas alguns dos cenários simulados. Na maioria deles, porém, a alteração do número de repetições para tratamentos não levou a alterações significativas nos resultados.

Com relação ao coeficiente de variação, situação semelhante à do teste de Tukey foi encontrada, seja para a situação de divergência 2 ou 4 (ANEXOS L1 e O1).

Para o teste DMS de Fisher, as duas situações de discordância também estiveram presentes nos resultados. A situação 2 sempre ocorreu em um número maior de experimentos por cenário do que o encontrado para o teste de Duncan e, ainda, esse número também aumentou à medida que o número de tratamentos dos cenários cresceu, chegando a aparecer em 1212 (60,60%) dos ensaios em um dos cenários simulados (ANEXO P1).

Já a situação de divergência 4 só foi encontrada em experimentos com 9 tratamentos, sempre em um número baixo dos ensaios, numa percentagem muito próxima de zero, e em menor número de experimentos por cenário do que o encontrado para o procedimento de Duncan (ANEXO S1). Pode-se concluir que o acréscimo no número de tratamentos parece ter conduzido ao aumento do número de experimentos com esse resultado contraditório.

Já a variação do número de repetições para tratamentos não levou a alterações expressivas no número de experimentos por cenário com a situação contraditória 2 (ANEXO Q1). Para a situação 4, entretanto, quando o número de repetições passou de 3 para 4, ocorreu uma pequena diminuição no número de experimentos com essa situação de divergência, que ficou maior quando o número de repetições passou para 10, e voltou a diminuir quando este subiu para 20 (ANEXO T1). No entanto, mais uma vez esse padrão foi notado em apenas alguns dos cenários simulados, sendo que na maioria deles a variação desse fator não provocou alterações expressivas nos resultados.

Com relação ao coeficiente de variação, situação semelhante à dos demais testes até então foi encontrada (ANEXOS R1 e U1).

Para o teste de SNK, também ambas as situações de divergência estiveram presentes nos resultados. A situação de divergência 2 apareceu em número pouco maior de experimentos por cenário que o encontrado para o método de Tukey, mas em número muito menor que os encontrados para os testes de Duncan e DMS de Fisher.

Já a situação contraditória 4 apareceu em número menor de experimentos por cenário que o encontrado para o método de Tukey, com exceção para os casos com 9 tratamentos e, na maioria dos cenários, em valores maiores que os obtidos para os testes de Duncan e DMS de Fisher. Pode-se notar, assim como para os demais testes, que à medida que o número de tratamentos dos cenários aumentou, isso provocou um acréscimo no número de ensaios com as situações de divergência 2 ou 4 (ANEXOS V1 e Y1).

Já o número de repetições para tratamentos mostrou não alterar de forma significativa os resultados obtidos (ANEXOS W1 e Z1). Com relação ao coeficiente de variação dos ensaios, situação similar a dos demais testes estudados foi encontrada, seja para a situação de divergência 2 ou 4 (ANEXOS X1 e A2).

Por fim, para o teste de Scheffé, a situação 2 apareceu em número menor de ensaios por cenário que os obtidos para cada um dos métodos anteriores, sendo ele muitas vezes nulo, o que ocorreu em todos os cenários simulados com 9 tratamentos. Conforme o número de tratamentos aumentou de 3 até 7, o número de experimentos por cenário com a situação de divergência 2 também cresceu, mas sempre se manteve muito próximo de zero na maioria dos casos. Quando o número de tratamentos subiu para 9, porém, nenhum dos cenários apresentou experimentos com essa situação de divergência, ou seja, quando o número de tratamentos foi de 7 para 9, o número de ensaios com esse resultado zerou (ANEXO B2).

Quando ocorreu a variação do número de repetições para tratamentos, entretanto, nenhum padrão foi notado nos resultados (ANEXO C2). Já o coeficiente de variação, conforme cresceu, provocou um decréscimo no número de ensaios por cenário com essa situação de divergência (ANEXO D2).

Já a situação 4 foi encontrada em número maior de experimentos por cenário do que os obtidos para cada um dos demais testes, mas na maioria dos casos em percentualidade ainda pequena. O número de tratamentos, conforme aumentou, levou também a um acréscimo no número de ensaios na situação de divergência 4 (ANEXO E2).

O número de repetições dos tratamentos, por sua vez, quando alterado mostrou não influenciar os resultados de forma que pudesse ser considerada significativa (ANEXO F2) e, o coeficiente de variação, quando passou de 1% para 5%, levou ao aumento do número de experimentos por cenário na situação de divergência 4. Quando este passou de 5% para 10%, no entanto, provocou diminuição desse número, que ficou ainda menor quando o coeficiente de variação foi de 20% (ANEXO G2). Ainda assim, o coeficiente de variação de 1% foi o que levou aos menores resultados.

Os ANEXOS de I a P apresentam, para cada um dos cenários simulados, a respectiva porcentagem de experimentos (dentre os que exibiram as situações de divergência 2 ou 4) que mostraram resultado não significativo e significativo para os testes de Levene e Shapiro-Wilk, respectivamente, para cada um dos procedimentos de Tukey, Duncan, DMS de Fisher, SNK e Scheffé. Assim, no ANEXO I, tomando o cenário demonstrativo de 3 tratamentos e 3 repetições por tratamento, por exemplo, tem-se que 100% dos experimentos simulados neste cenário e que exibiram as situações de divergência 2 ou 4 para o teste de Tukey, apresentaram resultado não significativo para o teste de Levene, enquanto que 91,67% desses experimentos demonstraram resultado não significativo para o teste de Shapiro-Wilk, ao nível de 5% de probabilidade.

Com base nos resultados percebeu-se que, dentro de cada um dos cenários simulados, apenas uma pequena porcentagem dos experimentos que apresentaram as situações de

divergência 2 ou 4 tiveram resultado significativo para ao menos um dos testes de Shapiro-Wilk ou Levene, o que leva a considerar que os resultados contraditórios parecem não estar ligados ao desvio de ao menos uma das pressuposições da análise.

### 4.3 Níveis de significância por comparação e por experimento

Os resultados para os níveis de significância por comparação e por experimento estão apresentados do ANEXO Q ao ANEXO B1.

Com relação aos níveis de significância por comparação pode-se notar o seguinte comportamento dos resultados: as taxas incondicionais  $\alpha_c$  foram sempre iguais as taxas condicionais  $\alpha_1$  para o procedimento de Scheffé, em todos os cenários simulados, enquanto houve uma tendência para as taxas  $\alpha_c$  serem ligeiramente maiores que  $\alpha_1$  para cada um dos demais testes considerados no estudo. Por sua vez, as taxas condicionais  $\alpha_2$  foram sempre maiores que as taxas  $\alpha_c$  nos cinco procedimentos de comparação de médias avaliados, em cada um dos cenários simulados. As diferenças entre essas duas últimas taxas de erro se mostraram grandes quando o número de tratamentos do cenário foi pequeno, e passaram a diminuir à medida que o número de tratamentos dos ensaios aumentou.

No que diz respeito aos níveis de significância por experimento, bem como ocorreu para os níveis de significância por comparação, as taxas incondicionais  $\alpha_e$  foram sempre iguais as taxas condicionais  $\alpha_3$  para o procedimento de Scheffé, em todos os cenários simulados, enquanto que para cada um dos demais procedimentos estudados, as taxas  $\alpha_e$  foram pouco maiores que as taxas  $\alpha_3$ . Por sua vez, as taxas  $\alpha_e$  mostraram-se sempre menores que as taxas condicionais  $\alpha_4$ , e as diferenças entre elas, embora consideráveis, diminuíram à medida que o número de tratamentos dos cenários aumentou. Para o caso particular do procedimento de Scheffé, no entanto, as diferenças entre essas taxas para cenários com maiores números de tratamentos não foram consideravelmente grandes.

Esses resultados para os níveis de significância por comparação e por experimento se assemelham aos obtidos por Bernhardson (1975).

Para o teste de Tukey, o aumento do número de tratamentos provocou diminuição significativa nas estimativas do nível de significância por comparação  $\alpha_c$ , que sempre permaneceram abaixo do limite inferior do intervalo de confiança de 95% [0,0375; 0,0625]. Assim, infere-se que esse procedimento não controla essa taxa de erro, conforme os resultados de Boardman e Moffitt (1971), Carmer e Swanson (1971, 1973), Bernhardson (1975), Percin

e Barbosa (1988), Conagin e Gomes (2004), Conagin e Barbin (2006) e Girardi, Cargnelutti Filho e Storck (2009).

As estimativas para o nível de significância  $\alpha_e$ , no entanto, conservaram-se sempre igual ao nível nominal imposto (5%), pois nenhuma das estimativas encontradas para essa taxa de erro ficou fora do intervalo de confiança calculado. Portanto, pode-se concluir que o método de Tukey controla o nível de significância por experimento incondicional, conforme constatado por Boardman e Moffitt (1971), Carmer e Swanson (1971,1973), Bernhardson (1975), Conagin e Gomes (2004), Girardi, Cargnelutti Filho e Storck (2009) e Sousa, Lira Junior e Ferreira (2012).

Com relação ao nível de significância condicional  $\alpha_1$ , bem como ocorreu para o nível incondicional  $\alpha_c$ , foi possível notar que as estimativas obtidas para ele também sempre permaneceram abaixo do limite inferior do intervalo de confiança considerado e, conforme houve um acréscimo do número de tratamentos, esse nível decresceu ainda mais, mostrando que o procedimento também não controla essa taxa de erro.

Relativo ao nível de significância por experimento  $\alpha_3$ , com o aumento do número de tratamentos dos cenários percebeu-se uma ligeira diminuição nas estimativas para esse nível, que passaram a ficar abaixo do nível de significância nominal imposto, ou seja, o método de Tukey mostrou não controlar essa taxa de erro em cenários com grandes números de tratamentos ( $a = 7$  e  $9$ ).

Diferentemente do que ocorreu com os demais níveis até então, as estimativas da taxa de erro  $\alpha_2$  para o teste de Tukey permaneceram na maioria dos cenários simulados muito acima do limite superior do intervalo de confiança de 95%. O acréscimo no número de tratamentos dos cenários levou a uma diminuição significativa nesse nível, que na maior parte dos cenários com 9 tratamentos permaneceu dentro do intervalo de confiança calculado, ou seja, o método parece controlar essa taxa de erro quando o número de tratamentos do ensaio é alto ( $a = 9$ ).

Para o nível  $\alpha_4$ , todas as estimativas encontradas permaneceram muito acima do nível nominal declarado, e o aumento do número de tratamentos também provocou uma diminuição nessa taxa de erro, mas menos expressiva do que a obtida para  $\alpha_2$ , ou seja, ainda assim as suas estimativas permaneceram acima do limite superior do intervalo de confiança calculado.

Para o teste de Duncan, as estimativas para o nível de significância por comparação incondicional  $\alpha_c$ , na maioria dos cenários com 3 tratamentos, ficaram dentro do intervalo de confiança de 95% e, quando o número de tratamentos dos cenários aumentou, essas taxas passaram a diminuir, permanecendo abaixo do limite inferior do intervalo de confiança



calculado na grande maioria dos cenários simulados, ou seja, de um modo geral o método parece não controlar essa taxa de erro, com exceção dos cenários com 3 tratamentos, conforme constatado por Boardman e Moffitt (1971), Carmer e Swanson (1971, 1973), Bernhardson (1975), Perecin e Barbosa (1988), Conagin e Barbin (2006) e Girardi, Cargnelutti Filho e Storck (2009).

Com relação ao nível de significância por experimento incondicional  $\alpha_e$ , pode-se notar que suas estimativas se mantiveram acima do limite superior do intervalo de confiança de 95% calculado e, conforme se aumentou o número de tratamentos dos cenários, essas taxas também sofreram acréscimos, mostrando assim a falta de controle sobre essa taxa de erro, de acordo com os resultados obtidos por Boardman e Moffitt (1971), Carmer e Swanson (1971, 1973), Bernhardson (1975), Perecin e Barbosa (1988), Conagin e Barbin (2006), Girardi, Cargnelutti Filho e Storck (2009) e Sousa, Lira Junior e Ferreira (2012).

No caso do nível de significância por comparação condicional  $\alpha_1$ , as suas estimativas sempre permaneceram abaixo do limite inferior do intervalo de confiança calculado, e o aumento do número de tratamentos dos cenários resultou em uma ligeira diminuição nessas estimativas. Assim, pode-se dizer que o método não controla essa taxa de erro. Ao contrário, o teste de Duncan parece controlar o nível de significância  $\alpha_3$ , que, com exceção de apenas um único cenário, permaneceu igual ao nível de significância nominal estipulado.

Com relação aos níveis  $\alpha_2$  e  $\alpha_4$ , suas estimativas ficaram muito acima do nível de significância nominal estabelecido, mostrando que essas taxas de erro não são controladas pelo teste de Duncan. O número de tratamentos, conforme cresceu, levou a uma diminuição nas estimativas para a taxa  $\alpha_2$ . O nível de significância  $\alpha_4$ , por sua vez, permaneceu sempre igual a 1 em todos os cenários simulados, o que indica ausência total de controle sobre esse tipo de erro.

No que diz respeito ao teste DMS de Fisher, este parece controlar a taxa de erro por comparação  $\alpha_c$ , pois suas estimativas sempre estiveram dentro do intervalo de confiança de 95% calculado, a menos de um dos cenários simulados. Já as estimativas para a taxa de erro por experimento  $\alpha_e$  sempre se mantiveram acima do limite superior do intervalo de confiança estabelecido, e o aumento do número de tratamentos nesse caso levou a um acréscimo das estimativas encontradas. Esses resultados se assemelham aos de Boardman e Moffitt (1971), Carmer e Swanson (1971, 1973), Bernhardson (1975), Perecin e Barbosa (1988) e Girardi, Cargnelutti Filho e Storck (2009).

Com relação ao nível  $\alpha_1$ , o teste DMS de Fisher parece não controlar essa taxa de erro, pois as suas estimativas ficaram sempre abaixo do nível nominal imposto e, conforme houve

aumento no número de tratamentos dos ensaios, ocorreu uma ligeira diminuição nesses valores. As taxas  $\alpha_3$ , por sua vez, foram controladas pelo teste, com exceção de um único cenário simulado, e o aumento do número de tratamentos não ocasionou mudanças expressivas nos resultados.

Já as estimativas para os níveis de significância  $\alpha_2$  e  $\alpha_4$  ficaram muito acima do nível de significância nominal estipulado, mostrando que essas taxas de erro não são controladas pelo teste DMS de Fisher. O nível de significância  $\alpha_4$ , ainda, permaneceu sempre igual a 1 em todos os cenários simulados, o que sugere total falta de controle do teste sobre esse nível de significância.

O teste de SNK, por sua vez, parece não controlar o nível de significância por comparação  $\alpha_c$ , cujas estimativas sempre permaneceram abaixo do limite inferior do intervalo de 95% de confiança, a menos de um dos cenários simulados. Ainda, foi possível verificar que o aumento do número de tratamentos provocou uma diminuição nessa taxa, de acordo com o que aferiram Boardman e Moffitt (1971), Carmer e Swanson (1973), Bernhardson (1975), Percin e Barbosa (1988), Conagin e Barbin (2006) e Girardi, Cargnelutti Filho e Storck (2009).

Já as estimativas encontradas para o nível de significância  $\alpha_e$  permaneceram igual ao nível de significância nominal de 5%, a menos de um único cenário, e a alteração do número de tratamentos não levou a alterações expressivas dos resultados. O controle da taxa de erro por experimento do teste de SNK também foi notada em Boardman e Moffitt (1971), Carmer e Swanson (1973), Bernhardson (1975) e Girardi, Cargnelutti Filho e Storck (2009).

O aumento do número de tratamentos também provocou uma diminuição nas estimativas para a taxa de erro  $\alpha_1$ , que sempre permaneceram abaixo do limite inferior do intervalo de confiança de 95% estimado para o nível nominal de 5%. No caso do nível  $\alpha_3$ , a elevação do número de tratamentos dos ensaios provocou uma ligeira diminuição em suas estimativas, que passaram a ficar abaixo do limite inferior do intervalo de confiança calculado, ou seja, pode-se dizer que o teste de SNK controla essa taxa de erro quando o número de tratamentos dos ensaios é baixo ( $a = 3$  e  $5$ ).

As estimativas para a taxa de erro  $\alpha_2$  permaneceram acima do limite superior do intervalo de confiança de 95% na imensa maioria dos cenários simulados. Quando o número de tratamentos dos ensaios aumentou, entretanto, foi possível notar uma diminuição significativa dessas estimativas, que permaneceram dentro do intervalo calculado para alguns dos cenários com 9 tratamentos. No entanto, de um modo geral pode-se dizer que o teste de SNK não controla essa taxa de erro.

As estimativas para o nível  $\alpha_4$ , por sua vez, sempre se mantiveram muito acima do limite superior do intervalo de confiança calculado, o que mostra que o teste não controla esse nível de significância. Foi possível notar, ainda, uma diminuição nessa taxa quando o número de tratamentos dos cenários aumentou, mas essa mudança foi menos expressiva do que a encontrada para o nível de significância  $\alpha_2$ .

No que diz respeito ao procedimento de Scheffé para a comparação de médias de tratamentos duas a duas, as estimativas para o nível de significância por comparação  $\alpha_c$  sempre permaneceram abaixo do limite inferior do intervalo de confiança estipulado para o nível nominal de 5% e, conforme o número de tratamentos aumentou, essas estimativas decresceram ainda mais, chegando a atingir o valor zero em alguns dos cenários com 9 tratamentos. Esses resultados estão de acordo com os encontrados por Boardman e Moffitt (1971), Carmer e Swanson (1973) e Bernhardson (1975).

As estimativas para o nível de significância por experimento  $\alpha_e$ , por sua vez, também se mantiveram abaixo do limite inferior do intervalo de confiança calculado, na maioria dos cenários simulados, com exceção para alguns dos cenários com 3 tratamentos, em que essas estimativas estiveram dentro do intervalo encontrado. Da mesma forma, foi possível notar que com o aumento do número de tratamentos houve uma diminuição significativa nos valores para essa taxa de erro. Os resultados se assemelham aos de Boardman e Moffitt (1971), Carmer e Swanson (1973) e Bernhardson (1975).

Bem como ocorreu para os níveis  $\alpha_c$  e  $\alpha_e$ , o aumento do número de tratamentos levou a uma diminuição nas taxas  $\alpha_1$  e  $\alpha_3$ . As estimativas encontradas para  $\alpha_1$  sempre foram inferiores ao nível nominal estabelecido, enquanto que as estimativas de  $\alpha_3$  foram inferiores na maioria dos casos, com exceção para alguns dos cenários com 3 tratamentos, em que elas estiveram dentro do intervalo de confiança calculado.

As estimativas da taxa  $\alpha_2$  se mantiveram acima do nível de significância nominal estabelecido, a menos de alguns dos cenários com 7 tratamentos e da maioria dos cenários com 9 tratamentos, em que essas estimativas conservaram-se dentro do intervalo de confiança de 95%. Para o caso dos valores obtidos para a taxa  $\alpha_4$ , estes também se mantiveram acima do limite superior do intervalo de confiança de 95%, com exceção para alguns cenários em que o número de tratamentos foi igual a 9, em que essa taxa foi controlada pelo procedimento. Pode-se dizer que o aumento do número de tratamentos provocou uma diminuição em ambos os níveis de significância.

Os resultados obtidos para os níveis de significância condicionais de cada um dos procedimentos de comparação de médias se assemelham aos obtidos por Bernhardson (1975). Ainda, a variação do número de repetições para tratamentos e do coeficiente de variação dos ensaios mostrou não influenciar os resultados encontrados de forma que pudesse ser considerada expressiva.

A fim de resumir o comportamento de cada um dos procedimentos no que diz respeito ao controle de cada um dos níveis de significância, as Figuras de 4.1 a 4.6 descrevem o que ocorre com cada um dos níveis para cada um dos testes à medida que o número de tratamentos dos ensaios aumenta. Apenas a variação no número de tratamentos foi considerada pois este foi o único fator que levou à alteração significativa dos resultados. Para tanto, fixou-se  $r = 10$ ,  $CV = 10\%$  e  $\alpha = 5\%$ . Todos os pontos nessas figuras foram obtidos por meio do estudo de Monte Carlo. As linhas entre os pontos foram desenhadas para representar o que ocorre com os níveis com o aumento do número de tratamentos apenas para fins de apresentação. Apenas o caso de  $r = 10$  e  $CV = 10\%$  foi reportado aqui a fim de economizar espaço. No entanto, os resultados para os demais cenários simulados foram muito similares em sua aparência.

De um modo geral, é possível dizer que o teste DMS de Fisher controla o nível de significância por comparação  $\alpha_c$ , enquanto que os testes de Tukey e SNK controlam o nível de significância por experimento  $\alpha_e$ . Os testes de Duncan e DMS de Fisher, por sua vez, controlam o nível de significância por experimento condicional  $\alpha_3$ . Já o procedimento de Scheffé mostrou não controlar nenhuma das taxas de erro consideradas.

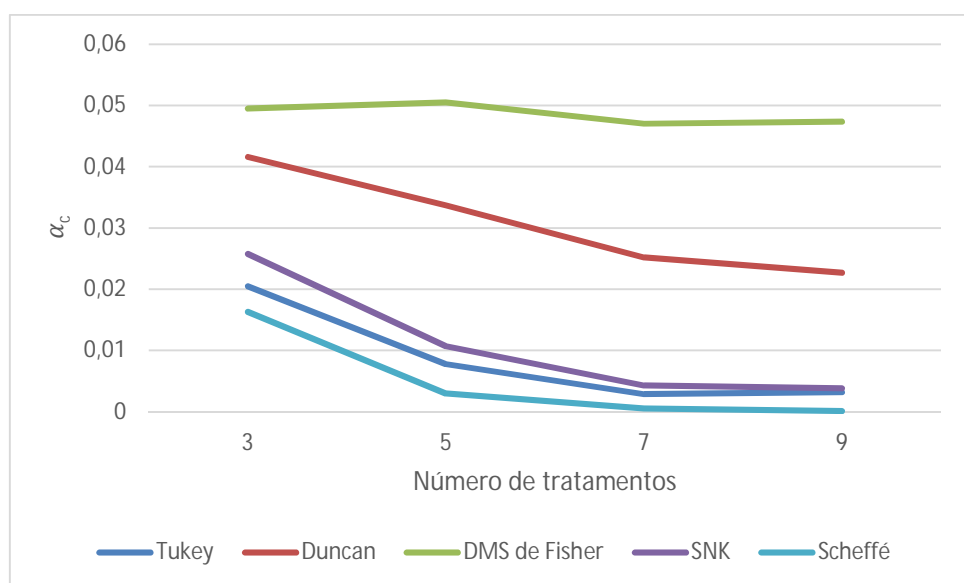


Figura 4.1 – Comportamento do nível de significância incondicional  $\alpha_c$  para  $r = 10$ ,  $CV = 10\%$  e  $\alpha = 5\%$ , com relação a variação do número de tratamentos

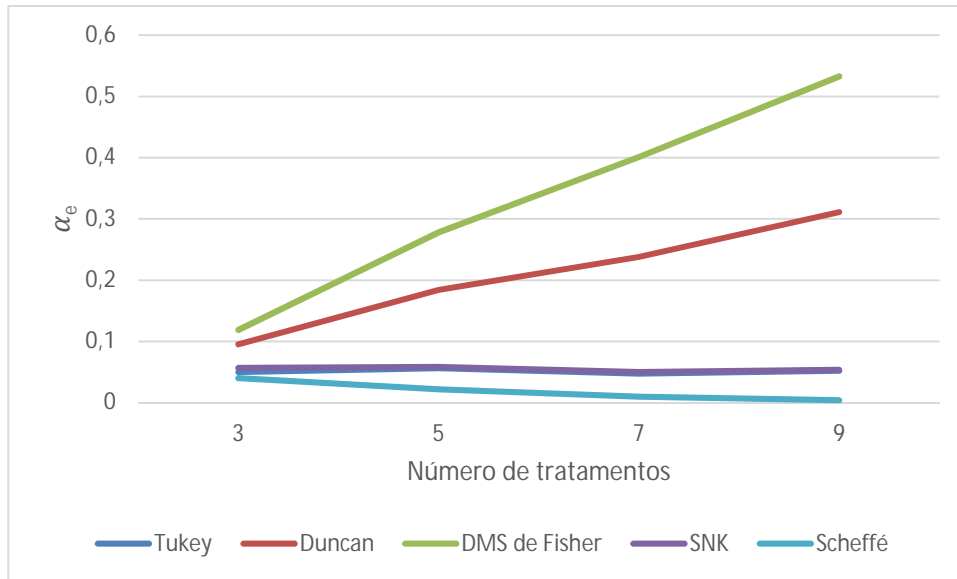


Figura 4.2 – Comportamento do nível de significância incondicional  $\alpha_e$  para  $r = 10$ ,  $CV = 10\%$  e  $\alpha = 5\%$ , com relação a variação do número de tratamentos

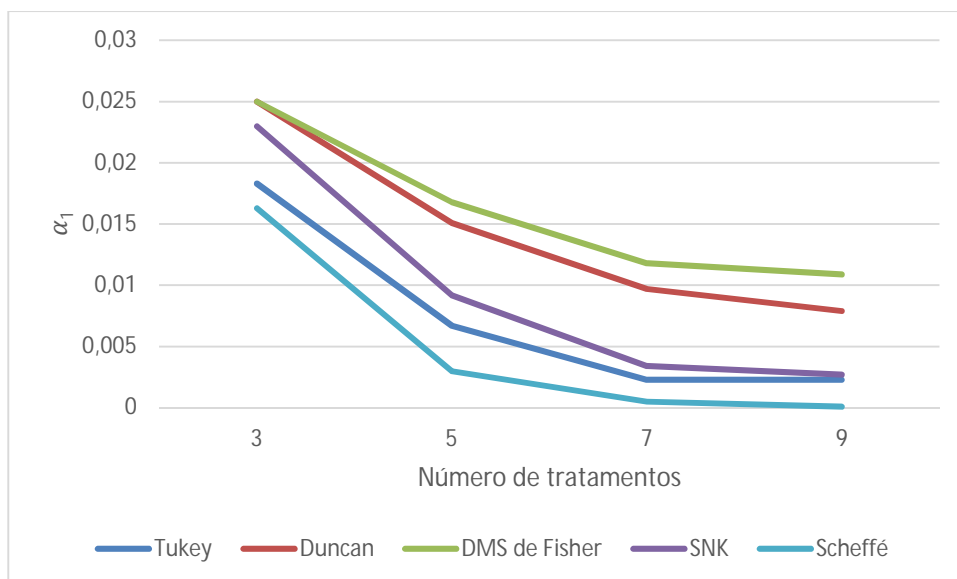


Figura 4.3 – Comportamento do nível de significância condicional  $\alpha_1$  para  $r = 10$ ,  $CV = 10\%$  e  $\alpha = 5\%$ , com relação a variação do número de tratamentos

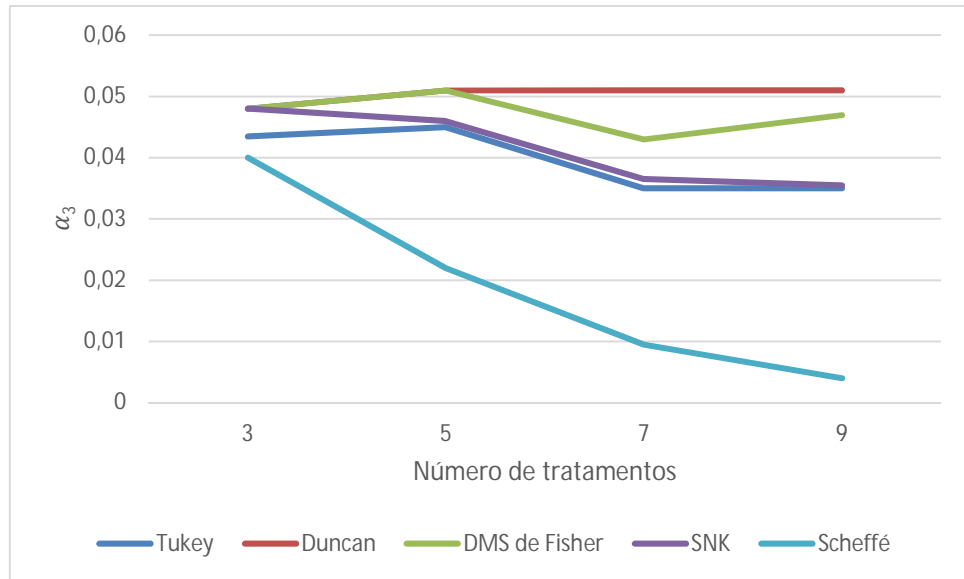


Figura 4.4 – Comportamento do nível de significância condicional  $\alpha_3$  para  $r = 10$ ,  $CV = 10\%$  e  $\alpha = 5\%$ , com relação a variação do número de tratamentos

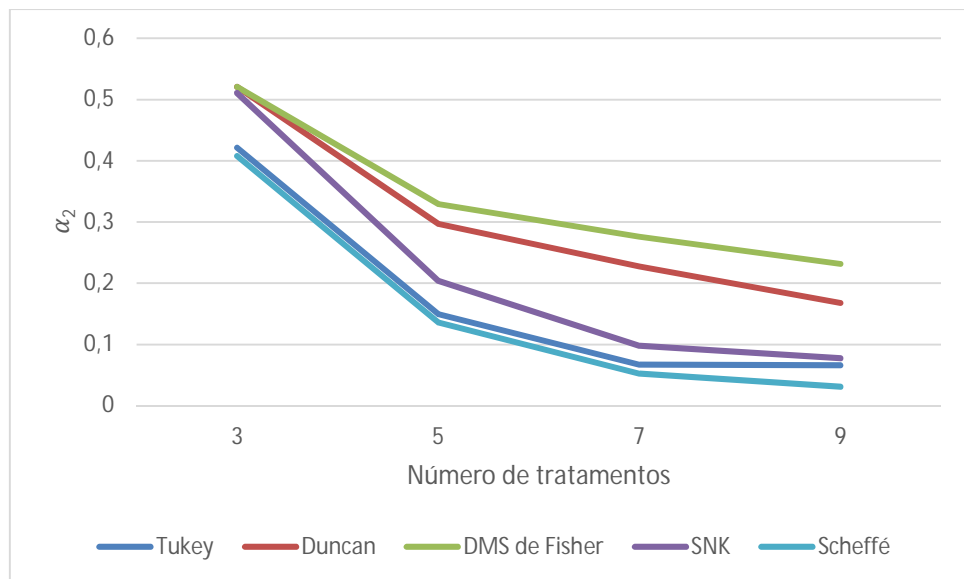


Figura 4.5 – Comportamento do nível de significância condicional  $\alpha_2$  para  $r = 10$ ,  $CV = 10\%$  e  $\alpha = 5\%$ , com relação a variação do número de tratamentos

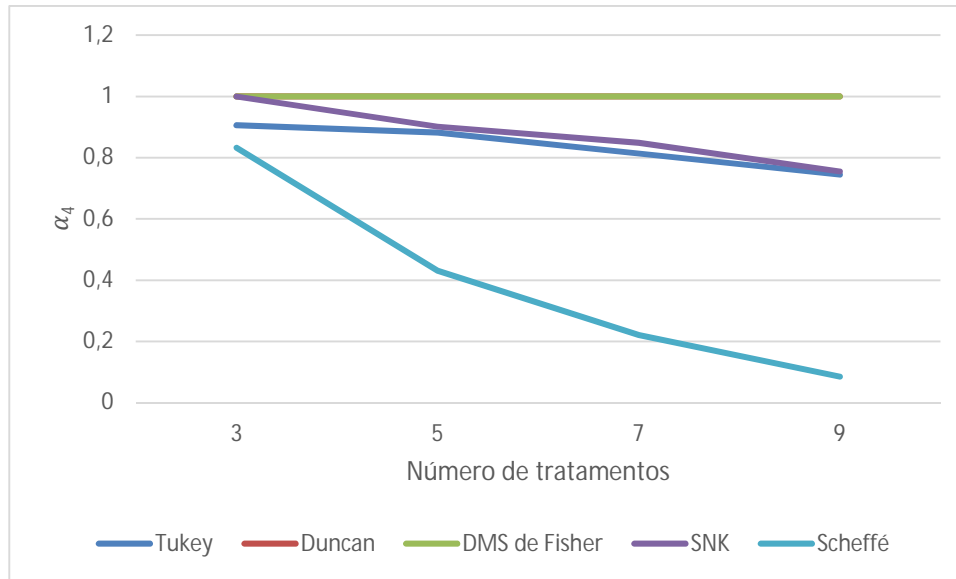


Figura 4.6 – Comportamento do nível de significância condicional  $\alpha_4$  para  $r = 10$ ,  $CV = 10\%$  e  $\alpha = 5\%$ , com relação a variação do número de tratamentos

#### 4.4 Poder

Com relação ao estudo do poder dos testes foi possível notar o seguinte comportamento dos resultados (ANEXO C1): as estimativas obtidas para o poder incondicional  $\pi$  foram sempre iguais as estimativas obtidas para o poder condicional  $\pi_1$  para o procedimento de Scheffé, enquanto houve uma tendência para as estimativas obtidas para  $\pi$  serem ligeiramente maiores que as obtidas para o poder  $\pi_1$  para cada um dos demais testes considerados no estudo, nos cenários com menor número de tratamentos e menor diferença real entre médias de tratamentos adjacentes. Nos cenários em que o número de tratamentos e a diferença real entre médias adjacentes foram altos, entretanto, as estimativas para  $\pi$  e  $\pi_1$  foram iguais entre si, para todos os procedimentos de comparação de médias avaliados.

Por sua vez, as estimativas encontradas para o poder condicional  $\pi_2$  foram maiores que as obtidas para o poder incondicional  $\pi$  nos cenários com menor número de tratamentos e menor diferença real entre médias de tratamentos adjacentes. Mais uma vez, nos cenários em que o número de tratamentos e a diferença real entre médias foram altos, as estimativas para  $\pi$  e  $\pi_2$  foram iguais entre si, para cada um dos procedimentos de comparação de médias analisados.

Foi possível notar, ainda, que à medida que houve um acréscimo da diferença real entre médias adjacentes atribuídas de acordo com o erro padrão da média, para um mesmo número de tratamentos, houve também um acréscimo do poder incondicional e condicional de cada um

dos testes de Tukey, Duncan, DMS de Fisher, SNK e Scheffé (ANEXO C1). Esse comportamento já era esperado, uma vez que quanto maior é a diferença entre duas médias, maior a probabilidade de ela ser considerada significativa por um procedimento para a comparação de médias de tratamentos (GIRARDI; CARGNELUTTI FILHO; STORCK, 2009).

Essa mesma tendência foi notada nos trabalhos de Carmer e Swanson (1971, 1973), Thomas (1974), Perecin e Barbosa (1988), Silva, Ferreira e Bearzoti (1999), Conagin e Gomes (2004), Conagin e Barbin (2006) e Girardi, Cargnelutti Filho e Storck (2009) no estudo do poder incondicional de procedimentos para a comparação de médias de tratamentos.

A Figura 4.7 descreve bem esse comportamento para o poder  $\pi$ , nos cenários em que o número de tratamentos foi igual a 5. Mais uma vez todos os pontos da figura foram obtidos por meio do estudo de simulações de Monte Carlo. As linhas entre os pontos foram desenhadas apenas para representar o que ocorre com o poder com o aumento da diferença entre médias de tratamentos adjacentes. Apenas o caso de  $\alpha = 5$  foi reportado a fim de economizar espaço. No entanto, os resultados para os demais cenários simulados foram muito similares em sua aparência. O mesmo pode-se dizer para o comportamento dos poderes  $\pi_1$  e  $\pi_2$ .

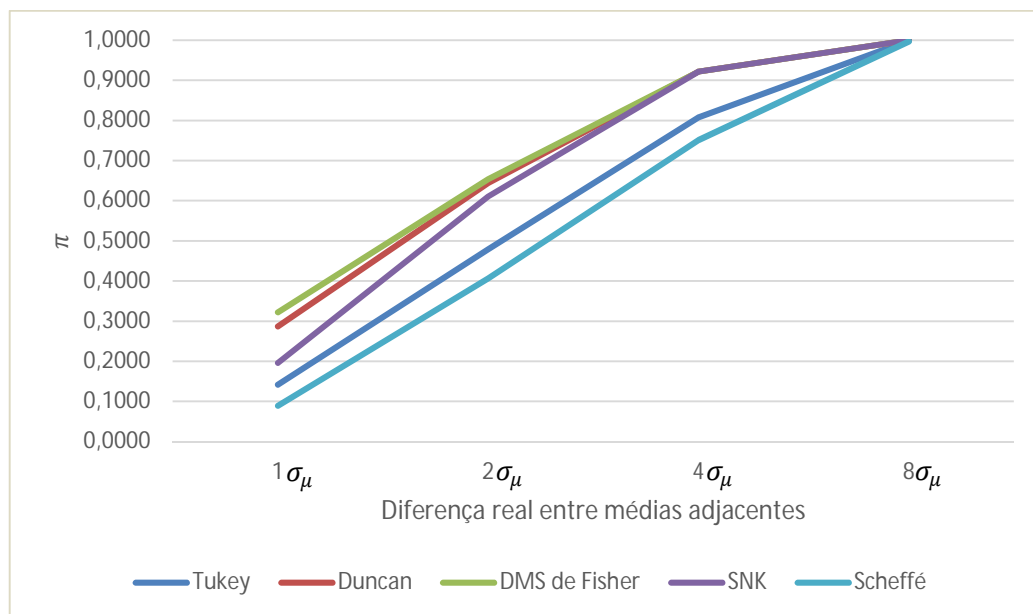


Figura 4.7 – Comportamento do poder incondicional  $\pi$  para  $\alpha = 5$  e  $\alpha = 5\%$ , com relação a variação da diferença real entre médias adjacentes

Ainda, foi possível verificar que se mantendo constante a diferença real entre médias de tratamentos adjacentes, ocorreu um acréscimo do poder de cada um dos cinco testes à medida



que houve aumento do número de tratamentos dos cenários simulados (ANEXO C1). Com efeito, esse comportamento dos testes também já era esperado, pois para obter as estimativas para o poder, tanto incondicional como condicional, foram considerados todos os contrastes de médias, duas a duas, e não apenas os contrastes com médias adjacentes.

Perecin e Barbosa (1988), que também utilizaram da configuração igualmente espaçada para atribuir valores aos efeitos de tratamentos, encontraram o mesmo comportamento para o teste de SNK, SNK modificado, Duncan,  $t$  e  $t$ -Bayesiano quando avaliados com relação ao seu poder incondicional. Para o teste de Tukey, entretanto, o aumento do número de tratamentos refletiu em diminuição do poder. Situação similar foi considerada no estudo de Girardi, Cargnelutti Filho e Storck (2009), que notaram aumento do poder incondicional dos testes  $t$ , Tukey, Duncan, Bonferroni e SNK com o aumento do número de tratamentos. Para os demais trabalhos avaliados que realizaram o estudo do poder de procedimentos de comparação de médias, entretanto, não será feita uma comparação dos resultados devido a diferença entre a metodologia utilizada neste trabalho e as descritas nos demais.

O aumento do poder dos testes com o acréscimo do número dos tratamentos dos cenários pode ser notado na Figura 4.8. Nesse caso, fica ilustrado o aumento do poder incondicional  $\pi$  dos procedimentos de Tukey, Duncan, DMS de Fisher, SNK e Scheffé com o aumento do número de tratamentos dos cenários, fixando-se a diferença real entre médias de  $2\sigma_\mu$  e  $\alpha = 5\%$ . Embora apenas essa situação tenha sido ilustrada aqui, padrão similar pode ser notado para cada um dos testes considerando outros valores de diferenças reais entre médias adjacentes, assim também como no que diz respeito ao seu desempenho com relação ao poder condicional  $\pi_1$  ou  $\pi_2$ .

De modo geral, pode-se dizer que o teste DMS de Fisher foi o que apresentou maior poder, seguido do teste Duncan. O procedimento de SNK ficou em situação intermediária, enquanto que o teste de Tukey e o de Scheffé foram os que exibiram menor poder, nessa mesma ordem.

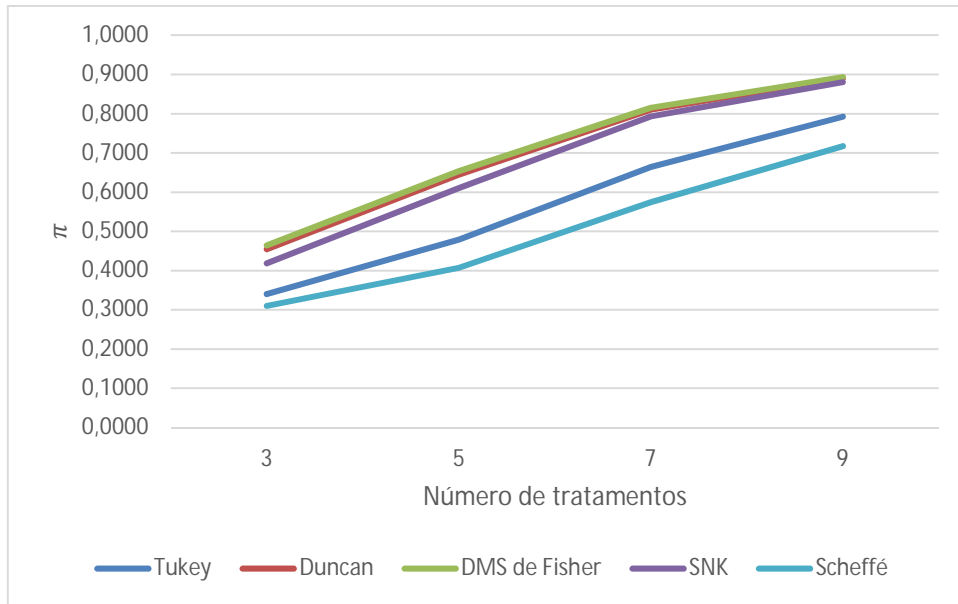


Figura 4.8 – Comportamento do poder incondicional  $\pi$  para diferença real entre médias adjacentes de  $2\sigma_\mu$  e  $\alpha = 5\%$ , com relação a variação do número de tratamentos



## 5 DISCUSSÃO

De modo geral, a presença das situações de divergência 2 e 4 foi notada em parte dos experimentos simulados para cada um dos testes de médias avaliados na pesquisa e, portanto, pode-se afirmar que é possível obter um teste  $F$  significativo na análise de variância e nenhum contraste entre duas médias de tratamentos significativo por algum dos testes de médias avaliados. Reciprocamente, é possível que o teste  $F$  aceite a hipótese de igualdade entre as médias de tratamentos mas um método de comparação detecte diferenças significativas entre as médias quando comparadas entre si duas a duas. No entanto, a presença das situações de divergência não parece estar associada ao desvio de ao menos uma das pressuposições da análise de variância.

Com exceção da situação de divergência 2 para o teste de Scheffé, o aumento do número de tratamentos levou ao aumento do número de experimentos por cenário que apresentaram as situações de divergência 2 ou 4. O coeficiente de variação, por sua vez, quando passou de 1% para 5% levou a um aumento nesse número de experimentos, que passou a diminuir à medida que o coeficiente passou a aumentar. O número de repetições para tratamentos, por sua vez, não levou a alterações expressivas nos resultados.

Com relação a situação 2 para o procedimento de Scheffé, o aumento do número de tratamentos de 3 para 7 provocou um ligeiro acréscimo no número de ensaios por cenário com essa situação de divergência que, contudo, zerou conforme o número de tratamentos dos ensaios subiu para 9. O aumento do número de repetições para tratamentos não provocou alterações expressivas nos resultados, enquanto que o acréscimo do coeficiente de variação, por sua vez, levou a uma diminuição no número de experimentos por cenário com essa situação de divergência.

No que diz respeito ao estudo dos níveis de significância por comparação dos testes, foi possível verificar que as estimativas encontradas para a taxa  $\alpha_c$  foram ligeiramente maiores que as obtidas para a taxa condicional  $\alpha_1$ , exceto para o procedimento de Scheffé, em que elas foram sempre iguais. As estimativas obtidas para a taxa condicional  $\alpha_2$ , entretanto, mostraram-se maiores que as estimativas para a taxa  $\alpha_c$ .

O mesmo comportamento pode ser notado para o nível de significância por experimento dos testes de médias. As estimativas encontradas para o nível incondicional  $\alpha_e$  foram ligeiramente maiores que as obtidas para o nível condicional  $\alpha_3$ , exceto para o procedimento de Scheffé, em que elas se mostraram iguais. As estimativas obtidas para o nível condicional  $\alpha_4$ , no entanto, revelaram-se maiores que as estimativas para a taxa  $\alpha_e$ .

Uma vez que para fazer o cálculo das estimativas dos níveis de significância condicionais  $\alpha_1$  e  $\alpha_3$  leva-se em conta todos os experimentos simulados em cada um dos cenários, e não apenas aqueles que apresentaram um resultado significativo do teste  $F$  global, é possível dizer que testes de médias que são aplicados apenas mediante um resultado preliminar significativo do teste  $F$  possuem níveis de significância por comparação e por experimento maiores do que se fossem aplicados independentemente do resultado do teste  $F$ , pois as estimativas obtidas para os níveis condicionais  $\alpha_2$  e  $\alpha_4$ , as quais foram calculadas levando em consideração apenas o número de ensaios com resultado significativo para o teste  $F$ , foram sempre maiores que as estimativas obtidas para  $\alpha_c$  e  $\alpha_e$ , respectivamente.

O aumento do número de tratamentos dos ensaios levou a alterações significativas nos resultados, acarretando na diminuição das diferenças entre cada um dos níveis de significância condicionais e o nível incondicional, tanto para o nível por comparação assim também como para o nível por experimento, o que não ocorreu quando o número de repetições para tratamentos e o coeficiente de variação dos experimentos foram alterados.

O aparecimento da situação de divergência 2 ao confrontar o resultado do teste  $F$  global na análise de variância com o resultado de cada um dos testes de comparação de médias parece estar ligado à probabilidade de ocorrência de erro tipo I, afinal, trata-se de obter um teste  $F$  não significativo e ao menos um contraste entre médias significativo pelo procedimento de comparação de médias em avaliação, o que seria similar a não aceitar a hipótese nula  $H_0: \mu_i = \mu_{i'}$ , dado que ela é verdadeira.

Ora, os testes DMS de Fisher e Duncan foram os procedimentos que mostraram maior percentual de experimentos por cenário com essa situação de divergência, nessa mesma ordem, e esse percentual aumentou à medida que o número de tratamentos dos ensaios aumentou. O teste de Scheffé, ao contrário, foi o procedimento que menos apresentou essa situação contraditória por cenário, estando os testes de Tukey e SNK em situação intermediária.

Ainda, o teste DMS de Fisher apresentou as maiores estimativas para o nível de significância  $\alpha_e$ , seguido pelo teste de Duncan, e essas estimativas também aumentaram com o acréscimo do número de tratamentos. O teste de Scheffé, ao contrário, foi o que apresentou as menores estimativas para essa taxa de erro (Figura 4.2). Assim, parece existir uma relação entre a situação de divergência 2 e o nível de significância por experimento incondicional  $\alpha_e$ , pois quanto maior a estimativa obtida para esse nível, maior a probabilidade de o teste de médias apresentar essa situação de contradição ao ter o seu resultado confrontado com o do teste  $F$ .

Com relação ao estudo do poder dos testes, as estimativas para o poder incondicional  $\pi$  foram ligeiramente maiores que as obtidas para o poder condicional  $\pi_1$  nos cenários com menor

número de tratamentos e menor diferença real entre médias de tratamentos adjacentes, exceto para os demais cenários e também para o teste de Scheffé, em que elas foram sempre iguais. Já as estimativas encontradas para o poder condicional  $\pi_2$  foram maiores que as encontradas para  $\pi$  nos cenários com menor número de tratamentos e menor diferença real entre médias de tratamentos adjacentes. Nos demais cenários, porém, as estimativas obtidas para  $\pi$  e  $\pi_2$  foram sempre iguais.

Como o cálculo do poder condicional  $\pi_2$  leva em consideração somente o número de experimentos simulados com resultado significativo do teste  $F$  global, o que não ocorre para o cálculo da estimativa de  $\pi_1$ , então é possível dizer que o poder dos testes de médias aplicados mediante um resultado preliminar significativo para o teste  $F$  é maior do que se estes fossem aplicados independentemente do resultado desse teste.

Foi possível notar, ainda, que à medida que houve um acréscimo da diferença real entre médias adjacentes atribuídas de acordo com o erro padrão da média, para um mesmo número de tratamentos, houve também um acréscimo do poder incondicional e condicional de cada um dos testes, bem como ocorreu uma aproximação das estimativas do poder de cada um dos testes de médias, e esse padrão foi notado tanto para o poder incondicional  $\pi$  assim também como para  $\pi_1$  e  $\pi_2$ .

Foi possível verificar também que se mantendo constante a diferença real entre médias de tratamentos adjacentes, ocorreu um acréscimo do poder de cada um dos cinco testes à medida que houve aumento do número de tratamentos dos cenários simulados. O aumento da diferença real entre médias levou também a uma aproximação das estimativas do poder dos testes.

O aparecimento da situação de divergência 4 ao confrontar o resultado do teste  $F$  global com o resultado de cada um dos testes de médias parece estar ligado à probabilidade de erro tipo II, afinal, trata-se de obter um teste  $F$  significativo e nenhum contraste entre médias significativo pelo procedimento de comparação de médias em avaliação, o que seria similar a aceitar a hipótese nula  $H_0: \mu_i = \mu_j$ , dado que ela é falsa.

O procedimento de Scheffé foi o que exibiu maior ocorrência da situação de divergência 4 por cenário, embora numa percentualidade ainda baixa dos ensaios. O teste DMS de Fisher, ao contrário, foi o que apresentou menor número de experimentos por cenário com essa situação de divergência, precedido pelo teste de Duncan. Os procedimentos de Tukey e SNK ficaram em situação intermediária.

Além disso, o teste DMS de Fisher foi o que apresentou maior poder, seguido do teste Duncan, enquanto que o teste de Scheffé mostrou-se o procedimento menos poderoso (ANEXO C1). Assim, de fato a situação de contradição 4 está ligada à probabilidade de ocorrência de

erro tipo II, pois o teste de Scheffé, que apresentou mais casos de ensaios por cenário com essa situação de divergência, foi o que apresentou menor poder, ou seja, maior probabilidade de erro tipo II, enquanto que os testes DMS de Fisher e Duncan, que apresentaram menos casos da contradição 4, mostraram-se mais poderosos, ou seja, com menor probabilidade de ocorrência para essa taxa de erro, estando os testes de Tukey e SNK em situação intermediária.

## 6 CONSIDERAÇÕES FINAIS

Neste trabalho, a partir de experimentos simulados no delineamento inteiramente casualizado pelo método de Monte Carlo, foi possível verificar, em parte dos experimentos simulados, a ausência de coerência entre o resultado do teste  $F$  global da análise de variância e os resultados dos procedimentos de Tukey, Duncan, DMS de Fisher, SNK e Scheffé.

De modo geral, a variação do número de tratamentos e do coeficiente de variação dos ensaios provocou mudanças expressivas na porcentagem de experimentos por cenário com resultados contraditórios, o que não ocorreu quando o número de repetições para tratamentos foi modificado. Os resultados contraditórios, entretanto, parecem não estar ligados ao desvio de ao menos uma das pressuposições da análise.

Os testes DMS de Fisher e Duncan foram os que exibiram maior ocorrência da situação de divergência 2 por cenário e, em contrapartida, foram os que exibiram menor ocorrência da situação de divergência 4. Em situação contrária, o procedimento de Scheffé apresentou a menor ocorrência da situação de divergência 2 e, de forma oposta, o maior percentual de experimentos por cenário com a situação de divergência 4. Por fim, os testes de Tukey e SNK ficaram em situação mediana com relação aos dois casos de contradição.

Seguindo a ideia de Bernhardson (1975), os níveis de significância por comparação e por experimento condicionais e incondicionais de cada um dos testes de médias foram também estimados, por meio do mesmo processo de simulação. O teste DMS de Fisher mostrou controlar o nível de significância por comparação  $\alpha_c$ , enquanto que os testes de Tukey e SNK parecem controlar o nível de significância por experimento  $\alpha_e$ . Os testes de Duncan e DMS de Fisher, por sua vez, controlam o nível de significância por experimento condicional  $\alpha_3$ . Já o procedimento de Scheffé mostrou não controlar nenhuma das taxas de erro consideradas.

Ainda, verificou-se que testes de médias que são aplicados apenas mediante um resultado preliminar significativo do teste  $F$  global possuem níveis de significância por comparação e por experimento maiores do que se fossem aplicados independentemente do resultado do teste  $F$ .

Por fim, os procedimentos de comparação de médias foram avaliados e comparados quanto ao seu poder, condicional e incondicional a um resultado significativo do teste  $F$ . Os testes DMS de Fisher e Duncan mostraram-se mais poderosos, enquanto que o procedimento de Scheffé foi o que apresentou menor poder, estando os procedimentos de SNK e de Tukey em situação intermediária. Nesta etapa também ficou claro que a aplicação condicional dos



testes pode levar a um aumento do poder dos mesmos em experimentos com baixo número de tratamentos e pequena diferença real entre médias de tratamentos adjacentes.

Além disso, o aparecimento de um maior número de experimentos por cenário com a situação de divergência 2 parece estar ligado ao fato de o teste de médias em avaliação possuir uma estimativa para o nível de significância por experimento muito acima do nível nominal declarado. Ainda, o aparecimento da situação de divergência 4 em maior número de ensaios por cenário parece estar ligado a um baixo poder do procedimento em avaliação.

Assim, como colocado por Hsu (1996), a aplicação de procedimentos para a comparação de médias de tratamentos apenas quando o teste  $F$  global da análise de variância apresenta um resultado significativo é um equívoco e deve, portanto, ser reconsiderada.

## REFERÊNCIAS

ANSCOMBE, F. J.; TUKEY, J. W. The examination and analysis of the residuals. **Technometrics**, Washington D. C., v.5, n.2, p.141-160, May, 1963.

BARBIN, D. **Planejamento e Análise Estatística de Experimentos Agronômicos**. Arapongas: Midas, 2003. 208p.

BERNHARDSON, C. S. Type I erros rates when multiple comparison procedures follow a significant F teste f ANOVA. **Biometrics**, Washington, v.31, n.1, p.229-232, Mar., 1975.

BERTOLDO, J. G.; COIMBRA, J. L. M.; GUIDOLIN, A. F.; MANTOVANI, A.; VALE, N. M. do. Problemas relacionados com o uso de testes de comparação de médias em artigos científicos. **Revista Biotemas**, Florianópolis, v.21, n.2, p.145-153, jun., 2008.

BERTOLDO, J. G.; COIMBRA, J. L. M.; GUIDOLIN, A. F.; MIQUELOTO, A.; TOALDO, D. Uso ou abuso em testes de comparações de média: conhecimento científico ou empírico? **Ciência Rural**, Santa Maria, v.38, n.4, p.1145-1148, jul., 2008.

BEZERRA NETO, F.; NUNES, G. H. S.; NEGREIROS, M. Z. de. Avaliação de procedimentos de comparações múltiplas em trabalhos publicados na revista Horticultura Brasileira de 1983 a 2000. **Horticultura Brasileira**, Brasília, v.20, n.1, p.5-9, mar., 2002.

BOARDMAN, T. J.; MOFFITT, D. R. Graphical Monte Carlo Type I error rates for multiple comparison procedures. **Biometrics**, Washington, v.27, n.3, p.738-744, Sept., 1971.

BONFERRONI, C. E. Il calcolo delle assicurazioni sui gruppi di teste. In: CARBONI, S. O. (Ed.) **Studi in Onore del Professore Salvatore Ortu Carboni**. Roma: Bardi, 1935. p.13-60.

\_\_\_\_\_. Teoria statistica delle classi e calcolo delle probabilità. **Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze**, Firenze, v.8, p.3-62, 1936.

BOX, G. E. P.; COX, D. R. An Analysis of Transformations. **Journal of the Royal Statistical Society**. London, v.26, n.2, p.211-252, Apr., 1964. Disponível em: < <http://www.ime.usp.br/~abe/lista/pdfQWaCMboK68.pdf> >. Acesso em: 21 fev. 2014.

BOX, G. E. P.; HUNTER Jr., W. G.; HUNTER, J. S. **Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building**. New York: John Wiley & Sons, 1978. 653p.

CARDELLINO, R. A.; SIEWERDT, F. Utilização correta e incorreta dos testes de comparação de médias. **Revista da Sociedade Brasileira de Zootecnia**, Viçosa, v.21, n.6, p.985-995, nov.-dez., 1992.

CARMER, S. G.; SWANSON, M. R. Detection of Differences Between Means: A Monte Carlo Study of Five Pairwise Multiple Comparison Procedures. **Agronomy Journal**, Madison, v.63, p.940-945, Nov.-Dec., 1971.

\_\_\_\_\_. An Evaluation of Ten Pairwise Multiple Comparison Procedures by Monte Carlo Methods. **Journal of the American Statistical Association**, Boston, v.68, n.341, p.66-74, Mar., 1973.

CHEW, V. Comparing treatment means: a compendium. **Hortscience**, Stanford, v.11, n.4, p.348-357, Aug., 1976.

CONAGIN, A.; BARBIN, D.; DEMÉTRIO, C. G. B. Modifications for the Tukey test procedure and evaluation of the power and efficiency of multiple comparison procedures. **Scientia Agricola**, Piracicaba, v.65, n.4, p.428-432, July-Aug., 2008.

CONAGIN, A.; GOMES, F. P. Escolha adequada dos testes estatísticos para comparações múltiplas. **Revista de Agricultura**, Piracicaba, v.79, n.3, p.288-295, dez., 2004.

DEAN, A.; VOSS, D. **Design and Analysis of Experiments**. New York: Springer, 1999. 740p.

DEMÉTRIO, C. G. B. **Transformação de dados. Efeitos sobre a análise de variância**. 1978. 113p. Dissertação (Mestrado em Experimentação e Estatística) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1978.

DRAPER, N. R.; SMITH, H. **Applied regression analysis**. 3. ed. New York: John Wiley & Sons, 1998. 706p.

DUNCAN, D. B. Multiple Range and Multiple F Tests. **Biometrics**, Raleigh, v.11. n.1, p.1-42, Mar., 1955.

DUNNETT, C. W. A multiple comparison procedure for comparing several treatments with a control. **Journal of the American Statistical Association**, Boston, v.50, n.272, p.1096-1121, Dec., 1955. Disponível em:  
< <http://www.tandfonline.com/doi/pdf/10.1080/01621459.1955.10501294>>. Acesso em: 18 dez. 2013.

EINOT, I.; GABRIEL, K. R. A study of the powers of several methods of multiple comparisons. **Journal of the American Statistical Association**, New York, v.70, n.351. p.574-583, Sep., 1975.

ENGELS, J. M.; DIEHR, P. Imputation of missing longitudinal data: a comparison of methods. **Journal of Clinical Epidemiology**, Oxford, v.56, n.10, p.968-976, Oct., 2003. Disponível em: <[http://www.jclinepi.com/article/S0895-4356\(03\)00170-7/fulltext](http://www.jclinepi.com/article/S0895-4356(03)00170-7/fulltext)>. Acesso em: 15 mar. 2014.

FISHER, R. A. **The design of experiments**. Edinburgh: Oliver & Boyd, 1935. 252p.

GIRARDI, L. H.; CARGNELUTTI FILHO, A.; STORCK, L. Erro tipo I e poder de cinco testes de comparação múltipla de médias. **Revista Brasileira de Biometria**, São Paulo, v.27, n.1, p.23-36, jan.-mar., 2009.

GOSSETT, S. W. S. Errors of routine analysis. **Biometrika**, Oxford, v.19, n.1-2, p.151-164, July, 1927. Disponível em: <<http://biomet.oxfordjournals.org/content/19/1-2/151.full.pdf>>. Acesso em: 11 fev. 2014.

HOCHBERG, Y.; TAMHANE, A. C. **Multiple comparisons procedures**. New York: John Wiley & Sons, 1987. 450p.

HSU, J. C. **Multiple Comparisons: theory and methods**. Great Britain: Chapman & Hall, 1996. 277p.

IEMMA, A. F. **Análise de Variância de Dados Desbalanceados**. In: Congresso Brasileiro de Usuários do SAS, 1., 1995, Piracicaba: ESALQ/USP, 1995. P. 111.

\_\_\_\_\_. **Dados Estatísticos Desbalanceados: Proc Sas/Glm**. Rio de Janeiro: Universidade Federal Rural, 1995. 130p.

KEULS, M. The use of the 'Studentized range' in connection with an analysis of variance. **Euphytica**, New York, v.1, n.2, p.112-122, July, 1952. Disponível em: <<http://link.springer.com/article/10.1007%2FBF01908269>>. Acesso em: 03 mar. 2014.

KOLMOGOROV, A. Sulla determinazione empirica di una legge di distribuzione. **Giornale dell'Istituto Italiano degli Attuari**, Roma, v.4, n.1, p.1-11, genn. 1933.

KUTNER, M.; NACHTSHEIM, C. J.; NETER, J.; LI, W. **Applied linear statistical models**. 5. ed. New York: McGraw-Hill/Irwin, 2004. 1396p.

LEVENE, H. Robust tests for equality of variances. In: OLKIN, I. ; GHURYE, S. G.; HOEFFDING, W.; MADOW, W. G.; MANN, H. B. (Ed.) **Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling**. Menlo Park: Stanford University Press, 1960, p.278-292.

MACHADO, A. A.; DEMÉTRIO, C. G. B.; FERREIRA, D. F.; SILVA, J. G. C. **Estatística experimental: uma abordagem fundamentada no planejamento e no uso de recursos computacionais**. Londrina: RBRAS/SEAGRO, 2005. 300p.

MONTGOMERY, D. C.; PECK, E.; VINING, G. G. **Introduction to Linear Regression Analysis**. 4. ed. New York: John Wiley & Sons, 2006. 640p.

NETER, J.; WASSERMAM, W. **Applied Linear Statistical Models: Regression, Analysis of Variance and Experimental Designs**. Homewood: Richard D. Irwin, 1974. 842p.

NEWMAN, D. The distribution of the range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. **Biometrika**, Oxford, v.31, n.1-2, p.16-31, July, 1939. Disponível em:  
<<http://biomet.oxfordjournals.org/content/31/1-2/20.full.pdf>>. Acesso em: 01 fev. 2014.

NOGUEIRA, M. C. S. **Experimentação Agrônômica I – conceitos, planejamentos e análise estatística**. Piracicaba: ESALQ, 2007. 479p.

NUNES, L. N. **Métodos de imputação de dados aplicados na área da saúde**. 2007. 120p. Tese (Doutorado em Epidemiologia) – Faculdade de Medicina, Universidade Federal do Rio Grande do Sul – UFRGS, Porto Alegre, 2007. Disponível em:  
<<http://www.lume.ufrgs.br/bitstream/handle/10183/11422/000615212.pdf?sequence=1>> Acesso em: 13 mar. 2014.

NUNES, L. N.; KLÜCK, M. M.; FACHEL, J. M. G. Uso da imputação múltipla de dados faltantes: uma simulação utilizando dados epidemiológicos. **Cadernos de Saúde Pública**, Rio de Janeiro, v.25, n.2, p.268-278, fev., 2009. Disponível em:  
<<http://www.scielo.br/pdf/csp/v25n2/05.pdf>>. Acesso em: 05 fev. 2014.

OLSHEN, R. A. The conditional level of the F-test. **Journal of the American Statistical Association**, Lancaster, v.68, n.343, p.692-698, Sept., 1973.

PARENTE, R. C. P. **Aspectos da análise de resíduos**. 1984. 139p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1984.

PERECIN, D.; BARBOSA, J. C. Uma avaliação de seis procedimentos para comparações múltiplas. **Revista de Matemática e Estatística**, São Paulo, v.6, p.95-103, mar., 1988.

PETERSEN, G. R. Use and misuse of multiple comparison procedures. **Agronomy Journal**, Madison, v.69, n.2, p.205-208, Mar.-Apr., 1977.

R Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<http://www.R-project.org/>>. Acesso em: 02 fev. 2013.

RAFTER, J. A.; ABEL, M. L.; BRASELTON, J. P. Multiple Comparison Methods for Means. **Society for Industrial and Applied Mathematics**, Philadelphia, v.44, n.2, p.259-278, May, 2002. Disponível em: <<http://www.siam.org/journals/sirev/44-2/35723.html>>. Acesso em: 10 dez. 2013.

RUBIN, D. B. **Multiple Imputation for Nonresponse in Surveys**. New York: John Wiley & Sons, 1987. 258p.

SANTOS, J. W. dos; MOREIRA, J. A. N.; BELTRÃO, N. E. de M. Avaliação do emprego dos testes de comparação de médias na revista pesquisa agropecuária brasileira (PAB) de 1980 a 1994. **Pesquisa Agropecuária Brasileira**, Brasília, v.33, n.3, p.225-230, mar., 1998.

SEBER, G. A. F. **Linear Regression Analysis**. New York: John Wiley & Sons, 1977. 465p.

SHAPIRO, S. S.; WILK, M. B. An Analysis of Variance Test for Normality (Complete Samples). **Biometrika**, Oxford, v.52, n.3-4, p.591-611, Dec., 1965. Disponível em: <<http://sci2s.ugr.es/keel/pdf/algorithm/articulo/shapiro1965.pdf>>. Acesso em: 23 nov. 2013.

SCHEFFÉ, H. A method for judging all contrasts in the analysis of variance. **Biometrika**, Oxford, v.40, n.1-2, p.87-104, Jun., 1953.

SCHEFFÉ, H. **The Analysis of Variance**. New York: John Wiley & Sons, 1959. 477p.

SILVA, E. C. da. FERREIRA, D. F. BEARZOTI, E. Avaliação do poder e taxas de erro tipo I do teste de Scott-Knott por meio do método de Monte Carlo. **Ciência e Agrotecnologia**, Lavras, v.23, n.3 p.687-696, jul.-set., 1999.

SMIRNOV, N. V. Sur la distribution de  $w^2$  (Critérium de M. v. Mises). **Comptes Rendus de l'Académie des Sciences**, Paris, v.202, p.449-452, févr., 1936. Disponível em: <<https://archive.org/details/ComptesRendusAcademieDesSciences0202>>. Acesso em: 17 fev. 2014.

\_\_\_\_\_. Sur la distribution de  $w^2$  (Critérium de M. v. Mises). **Rec. Math. [Matematicheskii Sbornik] N. S.**, Moscou, v.2, n.44, n.5, p.973-993, févr., 1937. Disponível em: <<http://www.mathnet.ru/links/0cadbafc64030a408c02a5485ab386a1/sm5636.pdf>>. Acesso em: 17 fev. 2014.

SOUZA, C. A. de; LIRA JUNIOR, M. A.; FERREIRA, R. L. C. Avaliação de testes estatísticos de comparações múltiplas de médias. **Revista Ceres**, Viçosa, v.59, n.3, p.350-354, mai.-jun., 2012. Disponível em: [http://www.scielo.br/scielo.php?pid=S0034-737X2012000300008&script=sci\\_arttext](http://www.scielo.br/scielo.php?pid=S0034-737X2012000300008&script=sci_arttext). Acesso em: 29/12/2014.

TUKEY, J. W. One Degree of Freedom for Non-Aditivity. **Biometrics**. Raleigh, v.5, n.3, p.232-242, Sept., 1949. Disponível em: <<http://www.jstor.org/discover/10.2307/3001938?uid=3737664&uid=2&uid=4&sid=21104368808277>>. Acesso em: 01 mar. 2014.

\_\_\_\_\_. The problem of multiple comparisons. In: BRAUN, H. I. (Ed.) **The collected works of John W. Tukey, Volume VIII, Multiple comparisons: 1948-1983**. New York: Chapman & Hall, 1953. p.1-300.

**BIBLIOGRAFIA CONSULTADA**

BARBIN, D. **Componentes de variância: teoria e aplicações**. 2. ed., Piracicaba: FEALQ, 1998. 120p.

BUSSAB, W.O.; MORETTIN, P.A. **Estatística Básica**. 6. ed. São Paulo: Saraiva, 2004. 526 p.

CONAGIN, A.; BARBIN, D. Bonferroni's and Sidak's modified tests. **Scientia Agricola**, Piracicaba, v.63, n.1, p.70-76, Jan.-Feb., 2006.

DURBIN, J.; WATSON, G. S. Testing for serial correlation in least squares regression II. **Biometrika**, Oxford, v.38, n.1-2, p.159-178, June, 1951. Disponível em: <<http://biomet.oxfordjournals.org/content/38/1-2/159.full.pdf>>. Acesso em: 12 jan. 2014.

GOMES, F. P. **Curso de Estatística Experimental**. 15. ed. Piracicaba: FEALQ, 2009. 451p.

MONTGOMERY, D. C. **Design and Analysis of Experiments**. 5. ed. New York: John Wiley & Sons, 2001. 684p.

THOMAS, D. A. H. Error Rate in Multiple Comparisons among Means – Results of a Simulation Exercise. **Applied Statistics**, Washington, v.23, n.3, p.284-294, 1974.

TUKEY, J. W. The philosophy of multiple comparisons. **Statistical Science**, Michigan, v.6, n.1, p.100-116, Feb., 1991.

VIEIRA, S. **Estatística Experimental**. 2. ed., São Paulo: Atlas, 1999. 185p.

\_\_\_\_\_. **Análise de Variância (Anova)**. São Paulo: Atlas, 2006. 204p.





## **APÊNDICES**



APÊNDICE A – Programa R para análise de um experimento simulado no delineamento inteiramente casualizado com 5 tratamentos, 4 repetições e coeficiente de variação de 5%. Ocorrência da situação de divergência 2 entre o resultado do teste F global e os resultados dos métodos de Tukey, Duncan, DMS de Fisher e SNK.

```
# Pacotes
library(dae)
require(dae)
library(agricolae)
require(agricolae)
library(car)
require(car)
# Conjunto de dados
CRD.dat <- data.frame(Parcelas = factor(1:20),
  Variedades = factor(rep(c("V1","V2","V3","V4","V5"), rep = 4)))
CRD.dat$resposta <- c(127, 112, 93, 104, 97, 114, 120, 111, 113, 104, 90, 109, 91, 100, 106,
  105, 133, 85, 111, 94)
CRD.dat
attach(CRD.dat)
# Média, variância e desvio padrão de cada uma das variedades
v <- tapply(resposta, Variedades, var)
data.frame(Mean = tapply(resposta, Variedades, mean), Variance = v, StandardError= sqrt(v))
rm(v)
# ANOVA
CRD.aov <- aov(resposta ~ Variedades, CRD.dat)
summary(CRD.aov)
# F_calc igual a 2,905
# F_tab ao nível de 5% igual a 3,056 - teste F global não significativo
# Diagnóstico
res <- residuals(CRD.aov) # resíduos
fit <- fitted(CRD.aov) # valores ajustados
res_standard <- rstandard(CRD.aov) # resíduos padronizados
# Gráfico para verificar independência dos erros
par(mfrow=c(1,1))
ordem <- c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20)
data.frame(ordem, res_standard)
plot(ordem, res_standard, pch=16, main="Resíduos padronizados versus ordem de coleta",
  xlab="Ordem de coleta", ylab="Resíduos padronizados")
abline(h = 0)
# Pontos distribuídos aleatoriamente em torno do zero - suposição de independência satisfeita
# Ausência de pontos discrepantes
# Gráfico para verificar suposição de homocedasticidade
plot(fit, res_standard, pch = 16, main = "Resíduos padronizados versus valores ajustados", xlab
  = "Valores ajustados", ylab = "Resíduos padronizados")
abline(h = 0)
# Teste de Levene para verificar suposição de homocedasticidade
leveneTest(resposta ~ Variedades, CRD.dat)
# Teste não significativo ao nível de 5% - suposição de homocedasticidade satisfeita
# Gráfico normal de probabilidades
```

```
qqnorm(res_standard, pch = 16, main = "Gráfico normal de probabilidades", xlab = "Quantis
teóricos da normal", ylab = "Resíduos padronizados")
qqline(res_standard)
# Teste de Shapiro-Wilk para verificar suposição de normalidade
shapiro.test(res_standard)
# Teste não significativo ao nível de 5% de probabilidade - suposição de normalidade satisfeita
# Testes para comparação de médias
# Teste de Tukey
tukey <- HSD.test(resposta, Variedades, group = TRUE, DFerror = 15,
MSerror = 110.2, alpha = 0.05)
# Teste de Duncan
duncan <- duncan.test(resposta, Variedades, group = TRUE, DFerror = 15,
MSerror = 110.2, alpha = 0.05)
# Teste DMS de Fisher
lsd <- LSD.test(resposta, Variedades, group = TRUE, DFerror = 15,
MSerror = 110.2, alpha = 0.05)
# Teste de Student-Newman-Keuls
snk <- SNK.test(resposta, Variedades, group = TRUE, DFerror = 15,
MSerror = 110.2, alpha = 0.05)
# Teste de Scheffé
scheffe <- scheffe.test(resposta, Variedades, group = TRUE, DFerror = 15,
MSerror = 110.2, alpha = 0.05)
```

APÊNDICE B – Programa R para análise de um experimento simulado no delineamento inteiramente casualizado com 5 tratamentos, 4 repetições e coeficiente de variação de 10%. Presença da situação de contradição 4 entre o resultado do teste F global e os resultados dos procedimentos de Tukey, SNK e Scheffé.

```
# Pacotes
library(dae)
require(dae)
library(agricolae)
require(agricolae)
library(car)
require(car)
# Conjunto de dados
CRD.dat <- data.frame(Parcelas = factor(1:20),
  Variedades = factor(rep(c("V1","V2","V3","V4","V5"), rep = 4)))
CRD.dat$resposta <- c(102, 92, 88, 106, 101, 108, 104, 84, 114, 93, 102, 100, 103, 108, 83, 91,
  110, 92, 100, 93)
CRD.dat
attach(CRD.dat)
# Média, variância e desvio padrão de cada uma das variedades
v <- tapply(resposta, Variedades, var)
data.frame(Mean = tapply(resposta, Variedades, mean), Variance = v, StandardError= sqrt(v))
rm(v)
# ANOVA
CRD.aov <- aov(resposta ~ Variedades, CRD.dat)
summary(CRD.aov)
# F_calc igual a 3,202
# F_tab ao nível de 5% é igual a 3,056 - teste F global significativo
# Diagnóstico
res <- residuals(CRD.aov) # resíduos
fit <- fitted(CRD.aov) # valores ajustados
res_standard <- rstandard(CRD.aov) # resíduos padronizados
# Gráfico para verificar independência dos erros
par(mfrow=c(1,1))
ordem <- c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20)
data.frame(ordem, res_standard)
plot(ordem, res_standard, pch=16, main="Resíduos padronizados versus ordem de coleta",
  xlab="Ordem de coleta", ylab="Resíduos padronizados")
abline(h = 0)
# Pontos distribuídos aleatoriamente em torno do zero - suposição de independência satisfeita
# Ausência de pontos discrepantes
# Gráfico para verificar suposição de homocedasticidade
plot(fit, res_standard, pch = 16, main = "Resíduos padronizados versus valores ajustados", xlab
  = "Valores ajustados", ylab = "Resíduos padronizados")
abline(h = 0)
# Teste de Levene para verificar suposição de homocedasticidade
leveneTest(produção ~ Variedades, CRD.dat)
# Teste não significativo ao nível de 5% - suposição de homocedasticidade satisfeita
# Gráfico normal de probabilidades
```

```
qqnorm(res_standard, pch = 16, main = "Gráfico normal de probabilidades", xlab = "Quantis
teóricos da normal", ylab = "Resíduos padronizados")
qqline(res_standard)
# Teste de Shapiro-Wilk para verificar suposição de normalidade
shapiro.test(res_standard)
# Teste não significativo ao nível de 5% de probabilidade - suposição de normalidade satisfeita
# Testes para comparação de médias
# Teste de Tukey
tukey <- HSD.test(resposta, Variedades, group = TRUE, DFerror = 15,
MSerror = 52.37, alpha = 0.05)
# Teste de Duncan
duncan <- duncan.test(resposta, Variedades, group = TRUE, DFerror = 15,
MSerror = 52.37, alpha = 0.05)
# Teste DMS de Fisher
lsd <- LSD.test(resposta, Variedades, group = TRUE, DFerror = 15,
MSerror = 52.37, alpha = 0.05)
# Teste de Student-Newman-Keuls
snk <- SNK.test(resposta, Variedades, group = TRUE, DFerror = 15,
MSerror = 52.37, alpha = 0.05)
# Teste de Scheffé
scheffe <- scheffe.test(resposta, Variedades, group = TRUE, DFerror = 15,
MSerror = 52.37, alpha = 0.05)
```

APÊNDICE C – Programa R com simulação de Monte Carlo para gerar dados de experimentos realizados de acordo com o delineamento inteiramente casualizado com 3 tratamentos, 4 repetições, coeficiente de variação de 10%, média geral arbitrada como 100 e efeitos de tratamentos  $\tau_1 = -2$ ,  $\tau_2 = 0$  e  $\tau_3 = 2$ .

```
# Semente
set.seed(5813002)
# Pacotes
library(dae)
require(dae)
library(agricolae)
require(agricolae)
library(car)
require(car)
# Processo de simulação de Monte Carlo
# Variáveis fixas
b <- 2000 # número de loopings
t <- 3 # número de tratamentos
r <- 4 # número de repetições
mu <- 100 # média geral
var.e <- 100 # variância do erro
x <- c(-2, 0, 2) # vetor dos efeitos dos tratamentos
a.dat <- matrix(x, nrow = t, ncol = r, byrow = FALSE) # matriz dos efeitos dos tratamentos
# Definindo matrizes para guardar os resultados
# F tabelado ao nível de 5% igual a 4,256
Fcalc <- matrix(0, b, 1)
Levenecalc <- matrix(0, b, 1)
Shapirocalc <- matrix(0, b, 1)
Tukeycalc <- matrix(0, b, t)
Duncancalc <- matrix(0, b, t)
LSDcalc <- matrix(0, b, t)
SNKcalc <- matrix(0, b, t)
Scheffecalc <- matrix(0, b, t)
# Simulando
for(i in 1:b){
  e.dat <- matrix(rnorm(t*r, 0, sqrt(var.e)), nrow = t, ncol = r) # matriz dos efeitos dos erros
  plan <- transform(data.frame(expand.grid(trat = 1:t, rep = 1:r), # conjunto de dados
  obs = c(round(mu + a.dat + e.dat))),
  trat = factor(trat),
  rep = factor(rep))
  respostas <- plan[,3]
  tratamentos <- plan[,1]
  anova <- aov(respostas ~ factor(tratamentos))
  mod <- summary(anova) # resultado da anova
  # Extrair o valor de F da tabela da ANOVA
  calc <- mod[[1]][1,4]
  Fcalc[i] <- calc[1]
  # teste de Levene
  testelevencalc <- leveneTest(plan[,3] ~ plan[,1], plan)
  pvalorlevencalc <- testelevencalc[1,3]
```



```

Levenecalc[i] <- pvalorlevene[1]
# teste de Shapiro-Wilk
res_standard <- rstandard(anova)
testeshapiro <- shapiro.test(res_standard)
shapiro <- as.numeric(testeshapiro)
pvalorshapiro <- shapiro[2]
Shapirocalc[i] <- pvalorshapiro[1]
# Testes para comparação de médias
# Teste de Tukey
tukey <- HSD.test(respostas, tratamentos, group = TRUE, DFerror = mod[[1]][2,1],
MSerror = mod[[1]][2,3], alpha = 0.05)
letras.tuk <- tukey$groups$M
letras.tuk
as.vector(letras.tuk)
# Teste de Duncan
duncan <- duncan.test(respostas, tratamentos, group = TRUE, DFerror = mod[[1]][2,1],
MSerror = mod[[1]][2,3], alpha = 0.05)
letras.dun <- duncan$groups$M
as.vector(letras.dun)
# Teste DMS de Fisher
lsd <- LSD.test(respostas, tratamentos, group = TRUE, DFerror = mod[[1]][2,1],
MSerror = mod[[1]][2,3], alpha = 0.05)
letras.lsd <- lsd$groups$M
as.vector(letras.lsd)
# Teste de Student-Newman-Keuls
snk <- SNK.test(respostas, tratamentos, group = TRUE, DFerror = mod[[1]][2,1],
MSerror = mod[[1]][2,3], alpha = 0.05)
letras.snk <- snk$groups$M
as.vector(letras.snk)
# Teste de Scheffé
scheffe <- scheffe.test(respostas, tratamentos, group = TRUE, DFerror = mod[[1]][2,1],
MSerror = mod[[1]][2,3], alpha = 0.05)
letras.sch <- scheffe$groups$M
as.vector(letras.sch)
for(j in 1:t){
  Tukeycalc[i,j] <- as.vector(letras.tuk[j])
  Duncancalc[i,j] <- as.vector(letras.dun[j])
  LSDcalc[i,j] <- as.vector(letras.lsd[j])
  SNKcalc[i,j] <- as.vector(letras.snk[j])
  Scheffecalc[i,j] <- as.vector(letras.sch[j])
}
}
# Matrizes com os resultados
Fcalc
Bartlettcalc
Levenecalc
Shapirocalc
Tukeycalc
Duncancalc
LSDcalc

```

SNKcalc  
Scheffecalc

APÊNDICE D – Programa R com simulação de Monte Carlo para gerar dados de experimentos realizados de acordo com o delineamento inteiramente casualizado com 3 tratamentos, 3 repetições, coeficiente de variação de 1%, média geral igual a 100 e na ausência dos efeitos de tratamentos, a fim de encontrar os níveis de significância por comparação (condicional e incondicional) e por experimento (condicional e incondicional) para cada um dos testes de Tukey, Duncan, DMS de Fisher, SNK e Scheffé.

```
# Semente
set.seed(5813002)
# Pacotes
library(dae)
require(dae)
library(agricolae)
require(agricolae)
# Processo de simulação de Monte Carlo
# Variáveis fixas
b <- 2000 # número de loopings
t <- 3 # número de tratamentos
r <- 3 # número de repetições
mu <- 100 # média geral
var.e <- 1 # variância do erro
x <- c(0, 0, 0) # vetor dos efeitos dos tratamentos
a.dat <- matrix(x, nrow = t, ncol = r, byrow = FALSE) # matriz dos efeitos dos tratamentos
# Definindo matrizes para guardar os resultados
# F tabelado ao nível de 5% igual a 5,143
Fcalc <- matrix(0, b, 1)
Tukeycalc <- matrix(0, b, t)
DuncanCalc <- matrix(0, b, t)
LSDcalc <- matrix(0, b, t)
SNKcalc <- matrix(0, b, t)
Scheffecalc <- matrix(0, b, t)
# Simulando
for(i in 1:b){
  e.dat <- matrix(rnorm(t*r, 0, sqrt(var.e)), nrow = t, ncol = r) # matriz dos efeitos dos erros
  plan <- transform(data.frame(expand.grid(trat = 1:t, rep = 1:r), # conjunto de dados
  obs = c(round(mu + a.dat + e.dat))),
  trat = factor(trat),
  rep = factor(rep))
  respostas <- plan[,3]
  tratamentos <- plan[,1]
  anova <- aov(respostas ~ factor(tratamentos))
  mod <- summary(anova) # resultado da anova
  # Extraindo o valor de F da tabela da ANOVA
  calc <- mod[[1]][1,4]
  Fcalc[i] <- calc[1]
  # Testes para comparação de médias
  # Teste de Tukey
  tukey <- HSD.test(respostas, tratamentos, group = TRUE, DFerror = mod[[1]][2,1],
  MSerror = mod[[1]][2,3], alpha = 0.05)
  letras.tuk <- tukey$groups$M
```

```

letras.tuk
as.vector(letras.tuk)
# Teste de Duncan
duncan <- duncan.test(respostas, tratamentos, group = TRUE, DFerror = mod[[1]][2,1],
MSerror = mod[[1]][2,3], alpha = 0.05)
letras.dun <- duncan$groups$M
as.vector(letras.dun)
# Teste DMS de Fisher
lsd <- LSD.test(respostas, tratamentos, group = TRUE, DFerror = mod[[1]][2,1],
MSerror = mod[[1]][2,3], alpha = 0.05)
letras.lsd <- lsd$groups$M
as.vector(letras.lsd)
# Teste de Student-Newman-Keuls
snk <- SNK.test(respostas, tratamentos, group = TRUE, DFerror = mod[[1]][2,1],
MSerror = mod[[1]][2,3], alpha = 0.05)
letras.snk <- snk$groups$M
as.vector(letras.snk)
# Teste de Scheffé
scheffe <- scheffe.test(respostas, tratamentos, group = TRUE, DFerror = mod[[1]][2,1],
MSerror = mod[[1]][2,3], alpha = 0.05)
letras.sch <- scheffe$groups$M
as.vector(letras.sch)
for(j in 1:t){
  Tukeycalc[i,j] <- as.vector(letras.tuk[j])
  Duncanalc[i,j] <- as.vector(letras.dun[j])
  LSDcalc[i,j] <- as.vector(letras.lsd[j])
  SNKcalc[i,j] <- as.vector(letras.snk[j])
  Scheffecalc[i,j] <- as.vector(letras.sch[j])
}
}
# Matrizes com os resultados
Fcalc
Tukeycalc
Duncanalc
LSDcalc
SNKcalc
Scheffecalc

```

APÊNDICE E – Programa R com simulação de Monte Carlo para gerar dados de experimentos realizados de acordo com o delineamento inteiramente casualizado com 3 tratamentos, 20 repetições, coeficiente de variação de 5%, média geral igual a 100 e diferença real entre médias de  $1\sigma_\mu$ , a fim de encontrar o poder incondicional e condicional de cada um dos testes de Tukey, Duncan, DMS de Fisher, SNK e Scheffé.

```
# Semente
set.seed(5813002)
# Pacotes
library(dae)
require(dae)
library(agricolae)
require(agricolae)
# Processo de simulação de Monte Carlo
# Variáveis fixas
b <- 2000 # número de loopings
t <- 3 # número de tratamentos
r <- 20 # número de repetições
mu <- 100 # média geral
var.e <- 25 # variância do erro
x <- c(-1.12, 0, 1.12) # vetor dos efeitos dos tratamentos
a.dat <- matrix(x, nrow = t, ncol = r, byrow = FALSE) # matriz dos efeitos dos tratamentos
# Definindo matrizes para guardar os resultados
# F tabelado ao nível de 5% igual a 3,150
Fcalc <- matrix(0, b, 1)
Tukeycalc <- matrix(0, b, t)
DuncanCalc <- matrix(0, b, t)
LSDcalc <- matrix(0, b, t)
SNKcalc <- matrix(0, b, t)
Scheffecalc <- matrix(0, b, t)
# Simulando
for(i in 1:b){
  e.dat <- matrix(rnorm(t*r, 0, sqrt(var.e)), nrow = t, ncol = r) # matriz dos efeitos dos erros
  plan <- transform(data.frame(expand.grid(trat = 1:t, rep = 1:r), # conjunto de dados
  obs = c(round(mu + a.dat + e.dat))),
  trat = factor(trat),
  rep = factor(rep))
  respostas <- plan[,3]
  tratamentos <- plan[,1]
  anova <- aov(respostas ~ factor(tratamentos))
  mod <- summary(anova) # resultado da anova
  # Extraindo o valor de F da tabela da ANOVA
  calc <- mod[[1]][1,4]
  Fcalc[i] <- calc[1]
  # Testes para comparação de médias
  # Teste de Tukey
  tukey <- HSD.test(respostas, tratamentos, group = TRUE, DFerror = mod[[1]][2,1],
  MSerror = mod[[1]][2,3], alpha = 0.05)
  letras.tuk <- tukey$groups$M
  letras.tuk
}
```

```

as.vector(letras.tuk)
# Teste de Duncan
duncan <- duncan.test(respostas, tratamentos, group = TRUE, DFerror = mod[[1]][2,1],
MSerror = mod[[1]][2,3], alpha = 0.05)
letras.dun <- duncan$groups$M
as.vector(letras.dun)
# Teste DMS de Fisher
lsd <- LSD.test(respostas, tratamentos, group = TRUE, DFerror = mod[[1]][2,1],
MSerror = mod[[1]][2,3], alpha = 0.05)
letras.lsd <- lsd$groups$M
as.vector(letras.lsd)
# Teste de Student-Newman-Keuls
snk <- SNK.test(respostas, tratamentos, group = TRUE, DFerror = mod[[1]][2,1],
MSerror = mod[[1]][2,3], alpha = 0.05)
letras.snk <- snk$groups$M
as.vector(letras.snk)
# Teste de Scheffé
scheffe <- scheffe.test(respostas, tratamentos, group = TRUE, DFerror = mod[[1]][2,1],
MSerror = mod[[1]][2,3], alpha = 0.05)
letras.sch <- scheffe$groups$M
as.vector(letras.sch)
for(j in 1:t){
  Tukeycalc[i,j] <- as.vector(letras.tuk[j])
  Duncanalc[i,j] <- as.vector(letras.dun[j])
  LSDcalc[i,j] <- as.vector(letras.lsd[j])
  SNKcalc[i,j] <- as.vector(letras.snk[j])
  Scheffecalc[i,j] <- as.vector(letras.sch[j])
}
}
# Matrizes com os resultados
Fcalc
Tukeycalc
Duncanalc
LSDcalc
SNKcalc
Scheffecalc

```



**ANEXOS**





ANEXO A – Proporção de experimentos, por cenário, em cada uma das quatro situações possíveis quando os testes de comparação de médias possuem os seus resultados confrontados com o resultado do teste F global, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 1%, com  $\tau_1 = -2, \tau_2 = 0, \tau_3 = 2$  para os cenários com 3 tratamentos,  $\tau_1 = -2, \tau_2 = -1, \tau_3 = 0, \tau_4 = 1, \tau_5 = 2$  para os cenários com 5 tratamentos,  $\tau_1 = -3, \tau_2 = -2, \tau_3 = -1, \tau_4 = 0, \tau_5 = 1, \tau_6 = 2, \tau_7 = 3$  para os cenários com 7 tratamentos e  $\tau_1 = -4, \tau_2 = -3, \tau_3 = -2, \tau_4 = -1, \tau_5 = 0, \tau_6 = 1, \tau_7 = 2, \tau_8 = 3, \tau_9 = 4$  para os cenários com 9 tratamentos

TRAT	REP	Tukey				Duncan				DMS de Fisher				SNK				Scheffé			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
3	3	0,0705	0,0065	0,9175	0,0055	0,0225	0,0545	0,9230	0,0000	0,0170	0,0600	0,9230	0,0000	0,0675	0,0095	0,9225	0,0005	0,0760	0,0010	0,9070	0,0160
3	4	0,0085	0,0035	0,9875	0,0005	0,0015	0,0105	0,9880	0,0000	0,0005	0,0115	0,9880	0,0000	0,0080	0,0040	0,9880	0,0000	0,0110	0,0010	0,9845	0,0035
3	10	0,0000	0,0005	0,9995	0,0000	0,0000	0,0005	0,9995	0,0000	0,0000	0,0005	0,9995	0,0000	0,0000	0,0005	0,9995	0,0000	0,0000	0,0005	0,9995	0,0000
3	20	0,0000	0,0255	0,9745	0,0000	0,0000	0,0255	0,9745	0,0000	0,0000	0,0255	0,9745	0,0000	0,0000	0,0255	0,9745	0,0000	0,0000	0,0255	0,9745	0,0000
5	3	0,0645	0,2020	0,7105	0,0230	0,0115	0,2550	0,7335	0,0000	0,0070	0,2595	0,7335	0,0000	0,0625	0,2040	0,7220	0,0115	0,0700	0,1965	0,6305	0,1030
5	4	0,0055	0,0005	0,9890	0,0050	0,0005	0,0055	0,9940	0,0000	0,0000	0,0060	0,9940	0,0000	0,0055	0,0005	0,9910	0,0030	0,0060	0,0000	0,9675	0,0265
5	10	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000
5	20	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000
7	3	0,0000	0,0000	0,9995	0,0005	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	0,9995	0,0005	0,0000	0,0000	0,9945	0,0055
7	4	0,0000	0,2570	0,7430	0,0000	0,0000	0,2570	0,7430	0,0000	0,0000	0,2570	0,7430	0,0000	0,0000	0,2570	0,7430	0,0000	0,0000	0,2570	0,7425	0,0005
7	10	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000
7	20	0,0000	0,0385	0,9615	0,0000	0,0000	0,0385	0,9615	0,0000	0,0000	0,0385	0,9615	0,0000	0,0000	0,0385	0,9615	0,0000	0,0000	0,0385	0,9615	0,0000
9	3	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000
9	4	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000
9	10	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000
9	20	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000

ANEXO B – Proporção de experimentos, por cenário, em cada uma das quatro situações possíveis quando os testes de comparação de médias possuem os seus resultados confrontados com o resultado do teste F global, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 1%, com  $\tau_1 = -0,5, \tau_2 = 0, \tau_3 = 0,5$  para os cenários com 3 tratamentos,  $\tau_1 = -1, \tau_2 = -0,5, \tau_3 = 0, \tau_4 = 0,5, \tau_5 = 1$  para os cenários com 5 tratamentos,  $\tau_1 = -1,5, \tau_2 = -1, \tau_3 = -0,5, \tau_4 = 0, \tau_5 = 0,5, \tau_6 = 1, \tau_7 = 1,5$  para os cenários com 7 tratamentos e  $\tau_1 = -2, \tau_2 = -1,5, \tau_3 = -1, \tau_4 = -0,5, \tau_5 = 0, \tau_6 = 0,5, \tau_7 = 1, \tau_8 = 1,5, \tau_9 = 2$  para os cenários com 9 tratamentos

TRAT	REP	Tukey				Duncan				DMS de Fisher				SNK				Scheffé			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
3	3	0,8725	0,0010	0,1080	0,0185	0,7790	0,0945	0,1265	0,0000	0,7535	0,1200	0,1265	0,0000	0,8605	0,0130	0,1260	0,0005	0,8735	0,0000	0,1075	0,0190
3	4	0,9235	0,0000	0,0515	0,0250	0,8990	0,0245	0,0765	0,0000	0,8845	0,0390	0,0765	0,0000	0,9225	0,0010	0,0620	0,0145	0,9235	0,0000	0,0435	0,0330
3	10	0,5530	0,0185	0,4115	0,0170	0,4410	0,1305	0,4285	0,0000	0,4060	0,1655	0,4285	0,0000	0,5420	0,0295	0,4270	0,0015	0,5715	0,0000	0,3825	0,0460
3	20	0,2150	0,0140	0,7600	0,0110	0,1405	0,0885	0,7710	0,0000	0,1225	0,1065	0,7710	0,0000	0,2050	0,0240	0,7705	0,0005	0,2290	0,0000	0,7340	0,0370
5	3	0,6265	0,0385	0,2915	0,0435	0,3410	0,3240	0,3350	0,0000	0,2695	0,3955	0,3350	0,0000	0,6185	0,0465	0,3085	0,0265	0,6425	0,0225	0,2010	0,1340
5	4	0,4425	0,0470	0,4615	0,0490	0,2115	0,2780	0,5105	0,0000	0,1510	0,3385	0,5105	0,0000	0,4365	0,0530	0,4805	0,0300	0,4700	0,0195	0,3420	0,1685
5	10	0,0250	0,1115	0,8505	0,0130	0,0060	0,1305	0,8635	0,0000	0,0030	0,1335	0,8635	0,0000	0,0245	0,1120	0,8545	0,0090	0,0280	0,1085	0,8060	0,0575
5	20	0,0005	0,0000	0,9995	0,0000	0,0005	0,0000	0,9995	0,0000	0,0005	0,0000	0,9995	0,0000	0,0005	0,0000	0,9995	0,0000	0,0005	0,0000	0,9990	0,0005
7	3	0,2070	0,0280	0,6945	0,0705	0,0475	0,1875	0,7650	0,0000	0,0220	0,2130	0,7650	0,0000	0,2045	0,0305	0,7140	0,0510	0,2350	0,0000	0,4630	0,3020
7	4	0,0665	0,0510	0,8385	0,0440	0,0115	0,1060	0,8825	0,0000	0,0055	0,1120	0,8825	0,0000	0,0655	0,0520	0,8505	0,0320	0,0715	0,0460	0,6470	0,2355
7	10	0,0000	0,2340	0,7660	0,0000	0,0000	0,2340	0,7660	0,0000	0,0000	0,2340	0,7660	0,0000	0,0000	0,2340	0,7660	0,0000	0,0000	0,2340	0,7655	0,0005
7	20	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000
9	3	0,0155	0,0030	0,9440	0,0375	0,0000	0,0185	0,9790	0,0025	0,0000	0,0185	0,9810	0,0005	0,0155	0,0030	0,9385	0,0430	0,0185	0,0000	0,7240	0,2575
9	4	0,0000	0,0000	0,9930	0,0070	0,0000	0,0000	0,9985	0,0015	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	0,9885	0,0115	0,0000	0,0000	0,9380	0,0620
9	10	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000
9	20	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000

ANEXO C – Proporção de experimentos, por cenário, em cada uma das quatro situações possíveis quando os testes de comparação de médias possuem os seus resultados confrontados com o resultado do teste F global, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 5%, com  $\tau_1 = -2, \tau_2 = 0, \tau_3 = 2$  para os cenários com 3 tratamentos,  $\tau_1 = -2, \tau_2 = -1, \tau_3 = 0, \tau_4 = 1, \tau_5 = 2$  para os cenários com 5 tratamentos,  $\tau_1 = -3, \tau_2 = -2, \tau_3 = -1, \tau_4 = 0, \tau_5 = 1, \tau_6 = 2, \tau_7 = 3$  para os cenários com 7 tratamentos e  $\tau_1 = -4, \tau_2 = -3, \tau_3 = -2, \tau_4 = -1, \tau_5 = 0, \tau_6 = 1, \tau_7 = 2, \tau_8 = 3, \tau_9 = 4$  para os cenários com 9 tratamentos

TRAT	REP	Tukey				Duncan				DMS de Fisher				SNK				Scheffé			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
3	3	0,8965	0,0035	0,0950	0,0050	0,8300	0,0700	0,1000	0,0000	0,8125	0,0875	0,1000	0,0000	0,8825	0,0175	0,1000	0,0000	0,9000	0,0000	0,0875	0,0125
3	4	0,8630	0,0095	0,1200	0,0075	0,7830	0,0895	0,1275	0,0000	0,7640	0,1085	0,1275	0,0000	0,8575	0,0150	0,1275	0,0000	0,8725	0,0000	0,1045	0,0230
3	10	0,6805	0,0160	0,2920	0,0115	0,5750	0,1215	0,3035	0,0000	0,5330	0,1635	0,3035	0,0000	0,6680	0,0285	0,3025	0,0010	0,6965	0,0000	0,2655	0,0380
3	20	0,3850	0,0205	0,5740	0,0205	0,2770	0,1285	0,5945	0,0000	0,2490	0,1565	0,5945	0,0000	0,3725	0,0330	0,5920	0,0025	0,4055	0,0000	0,5505	0,0440
5	3	0,8960	0,0160	0,0765	0,0115	0,7075	0,2045	0,0880	0,0000	0,6405	0,2715	0,0880	0,0000	0,8875	0,0245	0,0805	0,0075	0,9075	0,0045	0,0480	0,0400
5	4	0,8575	0,0205	0,1055	0,0165	0,6700	0,2080	0,1220	0,0000	0,5875	0,2905	0,1220	0,0000	0,8540	0,0240	0,1095	0,0125	0,8775	0,0005	0,0595	0,0625
5	10	0,6895	0,0285	0,2430	0,0390	0,4560	0,2620	0,2820	0,0000	0,3405	0,3775	0,2820	0,0000	0,6860	0,0320	0,2520	0,0300	0,7180	0,0000	0,1530	0,1290
5	20	0,3920	0,0325	0,5205	0,0550	0,2010	0,2235	0,5755	0,0000	0,1285	0,2960	0,5755	0,0000	0,3900	0,0345	0,5335	0,0420	0,4210	0,0035	0,3810	0,1945
7	3	0,8015	0,0255	0,1375	0,0355	0,4860	0,3410	0,1730	0,0000	0,3625	0,4645	0,1730	0,0000	0,7965	0,0305	0,1430	0,0300	0,8270	0,0000	0,0625	0,1105
7	4	0,7525	0,0235	0,1760	0,0480	0,4045	0,3715	0,2240	0,0000	0,2645	0,5115	0,2240	0,0000	0,7515	0,0245	0,1805	0,0435	0,7755	0,0005	0,0670	0,1570
7	10	0,3170	0,0250	0,5755	0,0825	0,1065	0,2355	0,6580	0,0000	0,0475	0,2945	0,6580	0,0000	0,3150	0,0270	0,5810	0,0770	0,3415	0,0005	0,2910	0,3670
7	20	0,0355	0,0120	0,9150	0,0375	0,0065	0,0410	0,9525	0,0000	0,0020	0,0455	0,9525	0,0000	0,0355	0,0120	0,9215	0,0310	0,0410	0,0065	0,7345	0,2180
9	3	0,6880	0,0285	0,2165	0,0670	0,2840	0,4325	0,2785	0,0050	0,1600	0,5565	0,2800	0,0035	0,6880	0,0285	0,2100	0,0735	0,7165	0,0000	0,0685	0,2150
9	4	0,5300	0,0350	0,3460	0,0890	0,1860	0,3790	0,4310	0,0040	0,0910	0,4740	0,4325	0,0025	0,5300	0,0350	0,3395	0,0955	0,5650	0,0000	0,1095	0,3255
9	10	0,0540	0,0090	0,8690	0,0680	0,0080	0,0550	0,9290	0,0080	0,0030	0,0600	0,9335	0,0035	0,0540	0,0090	0,8555	0,0815	0,0630	0,0000	0,5215	0,4155
9	20	0,0000	0,0010	0,9990	0,0000	0,0000	0,0010	0,9990	0,0000	0,0000	0,0010	0,9990	0,0000	0,0000	0,0010	0,9990	0,0000	0,0010	0,0000	0,9590	0,0400

ANEXO D – Proporção de experimentos, por cenário, em cada uma das quatro situações possíveis quando os testes de comparação de médias possuem os seus resultados confrontados com o resultado do teste F global, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 5%, com  $\tau_1 = -0,5, \tau_2 = 0, \tau_3 = 0,5$  para os cenários com 3 tratamentos,  $\tau_1 = -1, \tau_2 = -0,5, \tau_3 = 0, \tau_4 = 0,5, \tau_5 = 1$  para os cenários com 5 tratamentos,  $\tau_1 = -1,5, \tau_2 = -1, \tau_3 = -0,5, \tau_4 = 0, \tau_5 = 0,5, \tau_6 = 1, \tau_7 = 1,5$  para os cenários com 7 tratamentos e  $\tau_1 = -2, \tau_2 = -1,5, \tau_3 = -1, \tau_4 = -0,5, \tau_5 = 0, \tau_6 = 0,5, \tau_7 = 1, \tau_8 = 1,5, \tau_9 = 2$  para os cenários com 9 tratamentos

TRAT	REP	Tukey				Duncan				DMS de Fisher				SNK				Scheffé			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
3	3	0,8965	0,0035	0,0950	0,0050	0,8300	0,0700	0,1000	0,0000	0,8125	0,0875	0,1000	0,0000	0,8825	0,0175	0,1000	0,0000	0,9000	0,0000	0,0875	0,0125
3	4	0,9435	0,0025	0,0500	0,0040	0,8965	0,0495	0,0540	0,0000	0,8815	0,0645	0,0540	0,0000	0,9385	0,0075	0,0535	0,0005	0,9460	0,0000	0,0425	0,0115
3	10	0,9305	0,0075	0,0575	0,0045	0,8790	0,0590	0,0620	0,0000	0,8565	0,0815	0,0620	0,0000	0,9265	0,0115	0,0620	0,0000	0,9380	0,0000	0,0485	0,0135
3	20	0,9155	0,0070	0,0690	0,0085	0,8695	0,0530	0,0775	0,0000	0,8455	0,0770	0,0775	0,0000	0,9150	0,0075	0,0765	0,0010	0,9225	0,0000	0,0610	0,0165
5	3	0,9330	0,0100	0,0480	0,0090	0,7825	0,1605	0,0570	0,0000	0,7245	0,2185	0,0570	0,0000	0,9300	0,0130	0,0505	0,0065	0,9420	0,0010	0,0305	0,0265
5	4	0,9240	0,0100	0,0530	0,0130	0,7700	0,1640	0,0660	0,0000	0,6950	0,2390	0,0660	0,0000	0,9210	0,0130	0,0570	0,0090	0,9710	0,0290	0,0000	0,0000
5	10	0,8960	0,0385	0,0605	0,0050	0,7195	0,2150	0,0655	0,0000	0,6100	0,3245	0,0655	0,0000	0,8895	0,0450	0,0650	0,0005	0,9320	0,0025	0,0440	0,0215
5	20	0,8150	0,0180	0,1445	0,0225	0,6130	0,2200	0,1670	0,0000	0,5000	0,3330	0,1670	0,0000	0,8125	0,0205	0,1485	0,0185	0,8330	0,0000	0,0785	0,0885
7	3	0,9025	0,0175	0,0655	0,0145	0,6590	0,2610	0,0800	0,0000	0,5345	0,3855	0,0800	0,0000	0,9015	0,0185	0,0685	0,0115	0,9200	0,0000	0,0210	0,0590
7	4	0,9060	0,0165	0,0610	0,0165	0,6460	0,2765	0,0775	0,0000	0,5055	0,4170	0,0775	0,0000	0,9030	0,0195	0,0615	0,0160	0,9220	0,0005	0,0230	0,0545
7	10	0,8030	0,0200	0,1340	0,0430	0,5075	0,3155	0,1770	0,0000	0,3455	0,4775	0,1770	0,0000	0,8030	0,0200	0,1370	0,0400	0,8230	0,0000	0,0450	0,1320
7	20	0,6040	0,0385	0,2945	0,0630	0,2990	0,3435	0,3575	0,0000	0,1815	0,4610	0,3575	0,0000	0,6005	0,0420	0,2990	0,0585	0,6425	0,0000	0,1015	0,2560
9	3	0,8900	0,0205	0,0665	0,0230	0,5505	0,3600	0,0855	0,0040	0,3895	0,5210	0,0865	0,0030	0,8895	0,0210	0,0625	0,0270	0,9105	0,0000	0,0180	0,0715
9	4	0,8525	0,0265	0,0915	0,0295	0,5085	0,3705	0,1190	0,0020	0,3355	0,5435	0,1205	0,0005	0,8525	0,0265	0,0875	0,0335	0,8790	0,0000	0,0240	0,0970
9	10	0,6445	0,0335	0,2390	0,0830	0,2930	0,3850	0,3125	0,0095	0,1380	0,5400	0,3130	0,0090	0,6450	0,0330	0,2325	0,0895	0,6780	0,0000	0,0535	0,2685
9	20	0,2985	0,0310	0,5500	0,1205	0,0815	0,2480	0,6575	0,0130	0,0320	0,2975	0,6650	0,0055	0,2965	0,0330	0,5400	0,1305	0,3295	0,0000	0,1850	0,4855

ANEXO E – Proporção de experimentos, por cenário, em cada uma das quatro situações possíveis quando os testes de comparação de médias possuem os seus resultados confrontados com o resultado do teste F global, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 10%, com  $\tau_1 = -2, \tau_2 = 0, \tau_3 = 2$  para os cenários com 3 tratamentos,  $\tau_1 = -2, \tau_2 = -1, \tau_3 = 0, \tau_4 = 1, \tau_5 = 2$  para os cenários com 5 tratamentos,  $\tau_1 = -3, \tau_2 = -2, \tau_3 = -1, \tau_4 = 0, \tau_5 = 1, \tau_6 = 2, \tau_7 = 3$  para os cenários com 7 tratamentos e  $\tau_1 = -4, \tau_2 = -3, \tau_3 = -2, \tau_4 = -1, \tau_5 = 0, \tau_6 = 1, \tau_7 = 2, \tau_8 = 3, \tau_9 = 4$  para os cenários com 9 tratamentos

TRAT	REP	Tukey				Duncan				DMS de Fisher				SNK				Scheffé			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
3	3	0,9320	0,0030	0,0620	0,0030	0,8820	0,0530	0,0650	0,0000	0,8720	0,0630	0,0650	0,0000	0,9240	0,0110	0,0645	0,0005	0,9350	0,0000	0,0550	0,0100
3	4	0,9295	0,0045	0,0635	0,0025	0,8755	0,0585	0,0660	0,0000	0,8605	0,0735	0,0660	0,0000	0,9260	0,0080	0,0660	0,0000	0,9340	0,0000	0,0520	0,0140
3	10	0,8830	0,0095	0,1000	0,0075	0,8245	0,0680	0,1075	0,0000	0,7845	0,1080	0,1075	0,0000	0,8765	0,0160	0,1065	0,0010	0,8925	0,0000	0,0865	0,0210
3	20	0,8160	0,0075	0,1660	0,0105	0,7305	0,0930	0,1765	0,0000	0,6900	0,1335	0,1765	0,0000	0,8070	0,0165	0,1765	0,0000	0,8235	0,0000	0,1500	0,0265
5	3	0,9345	0,0110	0,0455	0,0090	0,7845	0,1610	0,0545	0,0000	0,7230	0,2225	0,0545	0,0000	0,9280	0,0175	0,0480	0,0065	0,9440	0,0015	0,0295	0,0250
5	4	0,9230	0,0130	0,0500	0,0140	0,7735	0,1625	0,0640	0,0000	0,6965	0,2395	0,0640	0,0000	0,9200	0,0160	0,0565	0,0075	0,9360	0,0000	0,0280	0,0360
5	10	0,8795	0,0105	0,0890	0,0210	0,7185	0,1715	0,1100	0,0000	0,6135	0,2765	0,1100	0,0000	0,8770	0,0130	0,0975	0,0125	0,8900	0,0000	0,0465	0,0635
5	20	0,8165	0,0165	0,1425	0,0245	0,6150	0,2180	0,1670	0,0000	0,5030	0,3300	0,1670	0,0000	0,8145	0,0185	0,1490	0,0180	0,8330	0,0000	0,0750	0,0920
7	3	0,9000	0,0190	0,0675	0,0135	0,6605	0,2585	0,0810	0,0000	0,5265	0,3925	0,0810	0,0000	0,8980	0,0210	0,0685	0,0125	0,9185	0,0005	0,0205	0,0605
7	4	0,9220	0,0240	0,0490	0,0050	0,6450	0,3010	0,0540	0,0000	0,5020	0,4440	0,0540	0,0000	0,9185	0,0275	0,0495	0,0045	0,9455	0,0005	0,0235	0,0305
7	10	0,8045	0,0230	0,1370	0,0355	0,5060	0,3215	0,1725	0,0000	0,3445	0,4830	0,1725	0,0000	0,8045	0,0230	0,1395	0,0330	0,8275	0,0000	0,0470	0,1255
7	20	0,6080	0,0395	0,2875	0,0650	0,3000	0,3475	0,3520	0,0005	0,1745	0,4730	0,3525	0,0000	0,6050	0,0425	0,2935	0,0590	0,6475	0,0000	0,1020	0,2505
9	3	0,8890	0,0215	0,0680	0,0215	0,5510	0,3595	0,0855	0,0040	0,3890	0,5215	0,0865	0,0030	0,8900	0,0205	0,0640	0,0255	0,9105	0,0000	0,0185	0,0710
9	4	0,8550	0,0255	0,0930	0,0265	0,5070	0,3735	0,1175	0,0020	0,3420	0,5385	0,1195	0,0000	0,8550	0,0255	0,0890	0,0305	0,8805	0,0000	0,0245	0,0950
9	10	0,6425	0,0350	0,2380	0,0845	0,2860	0,3915	0,3120	0,0105	0,1370	0,5405	0,3140	0,0085	0,6425	0,0350	0,2305	0,0920	0,6775	0,0000	0,0525	0,2700
9	20	0,2990	0,0315	0,5530	0,1165	0,0800	0,2505	0,6575	0,0120	0,0300	0,3005	0,6640	0,0055	0,2975	0,0330	0,5435	0,1260	0,3305	0,0000	0,1855	0,4840

ANEXO F – Proporção de experimentos, por cenário, em cada uma das quatro situações possíveis quando os testes de comparação de médias possuem os seus resultados confrontados com o resultado do teste F global, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 10%, com  $\tau_1 = -0,5, \tau_2 = 0, \tau_3 = 0,5$  para os cenários com 3 tratamentos,  $\tau_1 = -1, \tau_2 = -0,5, \tau_3 = 0, \tau_4 = 0,5, \tau_5 = 1$  para os cenários com 5 tratamentos,  $\tau_1 = -1,5, \tau_2 = -1, \tau_3 = -0,5, \tau_4 = 0, \tau_5 = 0,5, \tau_6 = 1, \tau_7 = 1,5$  para os cenários com 7 tratamentos e  $\tau_1 = -2, \tau_2 = -1,5, \tau_3 = -1, \tau_4 = -0,5, \tau_5 = 0, \tau_6 = 0,5, \tau_7 = 1, \tau_8 = 1,5, \tau_9 = 2$  para os cenários com 9 tratamentos

TRAT	REP	Tukey				Duncan				DMS de Fisher				SNK				Scheffé			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
3	3	0,9475	0,0145	0,0340	0,0040	0,9025	0,0595	0,0380	0,0000	0,8915	0,0705	0,0380	0,0000	0,9435	0,0185	0,0380	0,0000	0,9515	0,0105	0,0300	0,0080
3	4	0,9420	0,0050	0,0465	0,0065	0,9000	0,0470	0,0530	0,0000	0,8840	0,0630	0,0530	0,0000	0,9395	0,0075	0,0525	0,0005	0,9470	0,0000	0,0400	0,0130
3	10	0,9410	0,0085	0,0450	0,0055	0,8960	0,0535	0,0505	0,0000	0,8720	0,0775	0,0505	0,0000	0,9375	0,0120	0,0505	0,0000	0,9495	0,0000	0,0395	0,0110
3	20	0,9385	0,0050	0,0515	0,0050	0,9010	0,0425	0,0565	0,0000	0,8795	0,0640	0,0565	0,0000	0,9365	0,0070	0,0565	0,0000	0,9435	0,0000	0,0445	0,0120
5	3	0,9430	0,0065	0,0435	0,0070	0,8060	0,1435	0,0505	0,0000	0,7400	0,2095	0,0505	0,0000	0,9390	0,0105	0,0450	0,0055	0,9490	0,0005	0,0260	0,0245
5	4	0,9375	0,0105	0,0415	0,0105	0,7970	0,1510	0,0520	0,0000	0,7255	0,2225	0,0520	0,0000	0,9345	0,0135	0,0455	0,0065	0,9480	0,0000	0,0245	0,0275
5	10	0,9225	0,0070	0,0555	0,0150	0,7880	0,1415	0,0705	0,0000	0,7035	0,2260	0,0705	0,0000	0,9175	0,0120	0,0575	0,0130	0,9295	0,0000	0,0305	0,0400
5	20	0,9140	0,0135	0,0555	0,0170	0,7565	0,1710	0,0725	0,0000	0,6525	0,2750	0,0725	0,0000	0,9115	0,0160	0,0590	0,0135	0,9275	0,0000	0,0320	0,0405
7	3	0,9260	0,0155	0,0480	0,0105	0,7020	0,2395	0,0585	0,0000	0,5885	0,3530	0,0585	0,0000	0,9235	0,0180	0,0500	0,0085	0,9410	0,0005	0,0155	0,0430
7	4	0,9430	0,0200	0,0325	0,0045	0,7220	0,2410	0,0370	0,0000	0,5795	0,3835	0,0370	0,0000	0,9400	0,0230	0,0330	0,0040	0,9630	0,0000	0,0125	0,0245
7	10	0,9125	0,0165	0,0550	0,0160	0,6965	0,2325	0,0710	0,0000	0,5300	0,3990	0,0710	0,0000	0,9090	0,0200	0,0555	0,0155	0,9290	0,0000	0,0120	0,0590
7	20	0,8825	0,0235	0,0720	0,0220	0,6015	0,3045	0,0940	0,0000	0,4345	0,4715	0,0940	0,0000	0,8795	0,0265	0,0750	0,0190	0,9060	0,0000	0,0205	0,0735
9	3	0,9335	0,0125	0,0385	0,0155	0,6570	0,2890	0,0520	0,0020	0,4785	0,4675	0,0535	0,0005	0,9325	0,0135	0,0370	0,0170	0,9460	0,0000	0,0095	0,0445
9	4	0,9155	0,0150	0,0525	0,0170	0,6370	0,2935	0,0675	0,0020	0,4450	0,4855	0,0685	0,0010	0,9155	0,0150	0,0515	0,0180	0,9305	0,0000	0,0095	0,0600
9	10	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000
9	20	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000

ANEXO G – Proporção de experimentos, por cenário, em cada uma das quatro situações possíveis quando os testes de comparação de médias possuem os seus resultados confrontados com o resultado do teste F global, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 20%, com  $\tau_1 = -2, \tau_2 = 0, \tau_3 = 2$  para os cenários com 3 tratamentos,  $\tau_1 = -2, \tau_2 = -1, \tau_3 = 0, \tau_4 = 1, \tau_5 = 2$  para os cenários com 5 tratamentos,  $\tau_1 = -3, \tau_2 = -2, \tau_3 = -1, \tau_4 = 0, \tau_5 = 1, \tau_6 = 2, \tau_7 = 3$  para os cenários com 7 tratamentos e  $\tau_1 = -4, \tau_2 = -3, \tau_3 = -2, \tau_4 = -1, \tau_5 = 0, \tau_6 = 1, \tau_7 = 2, \tau_8 = 3, \tau_9 = 4$  para os cenários com 9 tratamentos

TRAT	REP	Tukey				Duncan				DMS de Fisher				SNK				Scheffé			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
3	3	0,9435	0,0035	0,0480	0,0050	0,8990	0,0480	0,0530	0,0000	0,8845	0,0625	0,0530	0,0000	0,9370	0,0100	0,0530	0,0000	0,9470	0,0000	0,0435	0,0095
3	4	0,9430	0,0035	0,0475	0,0060	0,8945	0,0520	0,0535	0,0000	0,8820	0,0645	0,0535	0,0000	0,9385	0,0080	0,0535	0,0000	0,9465	0,0000	0,0425	0,0110
3	10	0,9305	0,0070	0,0575	0,0050	0,8775	0,0600	0,0625	0,0000	0,8580	0,0795	0,0625	0,0000	0,9275	0,0100	0,0615	0,0010	0,9375	0,0000	0,0500	0,0125
3	20	0,9190	0,0050	0,0710	0,0050	0,8635	0,0605	0,0760	0,0000	0,8420	0,0820	0,0760	0,0000	0,9160	0,0080	0,0755	0,0005	0,9240	0,0000	0,0630	0,0130
5	3	0,9450	0,0055	0,0430	0,0065	0,8095	0,1410	0,0495	0,0000	0,7430	0,2075	0,0495	0,0000	0,9415	0,0090	0,0440	0,0055	0,9505	0,0000	0,0255	0,0240
5	4	0,9390	0,0095	0,0400	0,0115	0,7955	0,1530	0,0515	0,0000	0,7240	0,2245	0,0515	0,0000	0,9355	0,0130	0,0440	0,0075	0,9485	0,0000	0,0245	0,0270
5	10	0,9220	0,0085	0,0550	0,0145	0,7890	0,1415	0,0695	0,0000	0,7080	0,2225	0,0695	0,0000	0,9165	0,0140	0,0580	0,0115	0,9305	0,0000	0,0305	0,0390
5	20	0,9150	0,0150	0,0550	0,0150	0,7505	0,1795	0,0700	0,0000	0,6530	0,2770	0,0700	0,0000	0,9115	0,0185	0,0595	0,0105	0,9300	0,0000	0,0310	0,0390
7	3	0,9270	0,0160	0,0455	0,0115	0,7010	0,2420	0,0570	0,0000	0,5890	0,3540	0,0570	0,0000	0,9240	0,0190	0,0475	0,0095	0,9430	0,0000	0,0165	0,0405
7	4	0,9420	0,0205	0,0330	0,0045	0,7210	0,2415	0,0375	0,0000	0,5795	0,3830	0,0375	0,0000	0,9395	0,0230	0,0335	0,0040	0,9625	0,0000	0,0125	0,0250
7	10	0,9110	0,0200	0,0540	0,0150	0,6950	0,2360	0,0690	0,0000	0,5325	0,3985	0,0690	0,0000	0,9085	0,0225	0,0540	0,0150	0,9310	0,0000	0,0105	0,0585
7	20	0,9015	0,0660	0,0305	0,0020	0,6075	0,3600	0,0325	0,0000	0,4325	0,5350	0,0325	0,0000	0,8965	0,0710	0,0315	0,0010	0,9635	0,0040	0,0165	0,0160
9	3	0,9350	0,0125	0,0405	0,0120	0,6620	0,2855	0,0495	0,0030	0,4795	0,4680	0,0510	0,0015	0,9345	0,0130	0,0390	0,0135	0,9475	0,0000	0,0095	0,0430
9	4	0,9165	0,0140	0,0525	0,0170	0,6390	0,2915	0,0675	0,0020	0,4515	0,4790	0,0685	0,0010	0,9165	0,0140	0,0520	0,0175	0,9305	0,0000	0,0100	0,0595
9	10	0,8860	0,0210	0,0720	0,0210	0,5515	0,3555	0,0890	0,0040	0,3525	0,5545	0,0890	0,0040	0,8860	0,0210	0,0700	0,0230	0,9070	0,0000	0,0105	0,0825
9	20	0,7910	0,0390	0,1215	0,0485	0,4335	0,3965	0,1645	0,0055	0,2240	0,6060	0,1660	0,0040	0,7905	0,0395	0,1175	0,0525	0,8300	0,0000	0,0145	0,1555



ANEXO H – Proporção de experimentos, por cenário, em cada uma das quatro situações possíveis quando os testes de comparação de médias possuem os seus resultados confrontados com o resultado do teste F global, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 20%, com  $\tau_1 = -0,5, \tau_2 = 0, \tau_3 = 0,5$  para os cenários com 3 tratamentos,  $\tau_1 = -1, \tau_2 = -0,5, \tau_3 = 0, \tau_4 = 0,5, \tau_5 = 1$  para os cenários com 5 tratamentos,  $\tau_1 = -1,5, \tau_2 = -1, \tau_3 = -0,5, \tau_4 = 0, \tau_5 = 0,5, \tau_6 = 1, \tau_7 = 1,5$  para os cenários com 7 tratamentos e  $\tau_1 = -2, \tau_2 = -1,5, \tau_3 = -1, \tau_4 = -0,5, \tau_5 = 0, \tau_6 = 0,5, \tau_7 = 1, \tau_8 = 1,5, \tau_9 = 2$  para os cenários com 9 tratamentos

TRAT	REP	Tukey				Duncan				DMS de Fisher				SNK				Scheffé			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
3	3	0,9435	0,0035	0,0480	0,0050	0,8990	0,0480	0,0530	0,0000	0,8845	0,0625	0,0530	0,0000	0,9370	0,0100	0,0530	0,0000	0,9470	0,0000	0,0435	0,0095
3	4	0,9430	0,0050	0,0465	0,0055	0,9025	0,0455	0,0520	0,0000	0,8845	0,0635	0,0520	0,0000	0,9395	0,0085	0,0520	0,0000	0,9480	0,0000	0,0380	0,0140
3	10	0,9465	0,0060	0,0440	0,0035	0,8985	0,0540	0,0475	0,0000	0,8780	0,0745	0,0475	0,0000	0,9430	0,0095	0,0475	0,0000	0,9525	0,0000	0,0395	0,0080
3	20	0,9450	0,0030	0,0440	0,0080	0,9080	0,0400	0,0520	0,0000	0,8900	0,0580	0,0520	0,0000	0,9395	0,0085	0,0515	0,0005	0,9480	0,0000	0,0395	0,0125
5	3	0,9475	0,0050	0,0400	0,0075	0,8225	0,1300	0,0475	0,0000	0,7480	0,2045	0,0475	0,0000	0,9445	0,0080	0,0445	0,0030	0,9525	0,0000	0,0240	0,0235
5	4	0,9415	0,0070	0,0410	0,0105	0,8015	0,1470	0,0515	0,0000	0,7325	0,2160	0,0515	0,0000	0,9380	0,0105	0,0440	0,0075	0,9485	0,0000	0,0250	0,0265
5	10	0,9350	0,0095	0,0490	0,0065	0,8095	0,1350	0,0555	0,0000	0,7255	0,2190	0,0555	0,0000	0,9325	0,0120	0,0495	0,0060	0,9445	0,0000	0,0260	0,0295
5	20	0,9350	0,0105	0,0435	0,0110	0,7910	0,1545	0,0545	0,0000	0,6970	0,2485	0,0545	0,0000	0,9335	0,0120	0,0450	0,0095	0,9455	0,0000	0,0200	0,0345
7	3	0,9385	0,0110	0,0410	0,0095	0,7170	0,2325	0,0505	0,0000	0,6055	0,3440	0,0505	0,0000	0,9360	0,0135	0,0425	0,0080	0,9495	0,0000	0,0160	0,0345
7	4	0,9440	0,0090	0,0370	0,0100	0,7395	0,2135	0,0470	0,0000	0,5995	0,3535	0,0470	0,0000	0,9430	0,0100	0,0370	0,0100	0,9530	0,0000	0,0135	0,0335
7	10	0,9360	0,0150	0,0415	0,0075	0,7480	0,2030	0,0490	0,0000	0,5760	0,3750	0,0490	0,0000	0,9360	0,0150	0,0425	0,0065	0,9510	0,0000	0,0090	0,0400
7	20	0,9330	0,0130	0,0400	0,0140	0,7115	0,2345	0,0540	0,0000	0,5385	0,4075	0,0540	0,0000	0,9315	0,0145	0,0405	0,0135	0,9460	0,0000	0,0090	0,0450
9	3	0,9415	0,0135	0,0355	0,0095	0,6725	0,2825	0,0440	0,0010	0,5030	0,4520	0,0440	0,0010	0,9405	0,0145	0,0340	0,0110	0,9550	0,0000	0,0085	0,0365
9	4	0,9330	0,0110	0,0420	0,0140	0,6700	0,2740	0,0550	0,0010	0,4795	0,4645	0,0545	0,0015	0,9330	0,0110	0,0415	0,0145	0,9440	0,0000	0,0080	0,0480
9	10	0,9225	0,0190	0,0405	0,0180	0,6535	0,2880	0,0570	0,0015	0,4350	0,5065	0,0570	0,0015	0,9235	0,0180	0,0400	0,0185	0,9415	0,0000	0,0060	0,0525
9	20	0,9025	0,0240	0,0515	0,0220	0,6120	0,3145	0,0700	0,0035	0,3750	0,5515	0,0710	0,0025	0,9035	0,0230	0,0490	0,0245	0,9265	0,0000	0,0015	0,0720

ANEXO I – Proporção de experimentos, por cenário, dentre os que exibiram a situação de divergência 2 ou 4, que apresentaram resultado não significativo (NS) e significativo (S) para os testes de Levene e Shapiro-Wilk, respectivamente, ao nível de significância de 5%, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 1%, com  $\tau_1 = -2, \tau_2 = 0, \tau_3 = 2$  para os cenários com 3 tratamentos,  $\tau_1 = -2, \tau_2 = -1, \tau_3 = 0, \tau_4 = 1, \tau_5 = 2$  para os cenários com 5 tratamentos,  $\tau_1 = -3, \tau_2 = -2, \tau_3 = -1, \tau_4 = 0, \tau_5 = 1, \tau_6 = 2, \tau_7 = 3$  para os cenários com 7 tratamentos e  $\tau_1 = -4, \tau_2 = -3, \tau_3 = -2, \tau_4 = -1, \tau_5 = 0, \tau_6 = 1, \tau_7 = 2, \tau_8 = 3, \tau_9 = 4$  para os cenários com 9 tratamentos

TRAT	REP	Tukey				Duncan				DMS de Fisher				SNK				Scheffé			
		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk	
		NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S
3	3	1,0000	0,0000	0,9167	0,0833	1,0000	0,0000	0,9450	0,0550	1,0000	0,0000	0,9417	0,0583	1,0000	0,0000	0,8500	0,1500	1,0000	0,0000	0,9412	0,0588
3	4	0,8750	0,1250	0,8750	0,1250	0,9048	0,0952	0,8571	0,1429	0,8696	0,1304	0,8696	0,1304	0,8750	0,1250	0,8750	0,1250	0,7778	0,2222	0,7778	0,2222
3	10	1,0000	0,0000	0,0000	1,0000	1,0000	0,0000	0,0000	1,0000	1,0000	0,0000	0,0000	1,0000	1,0000	0,0000	0,0000	1,0000	0,0000	0,0000	1,0000	
3	20	0,9412	0,0588	0,2941	0,7059	0,9412	0,0588	0,2941	0,7059	0,9412	0,0588	0,2941	0,7059	0,9412	0,0588	0,2941	0,7059	0,9412	0,0588	0,2941	0,7059
5	3	1,0000	0,0000	0,8700	0,1300	1,0000	0,0000	0,8800	0,1200	1,0000	0,0000	0,8800	0,1200	1,0000	0,0000	0,8700	0,1300	1,0000	0,0000	0,8800	0,1200
5	4	1,0000	0,0000	0,9091	0,0909	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	0,9623	0,0377	0,9057	0,0943
5	10	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
5	20	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
7	3	1,0000	0,0000	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000
7	4	0,9377	0,0623	0,8580	0,1420	0,9377	0,0623	0,8580	0,1420	0,9377	0,0623	0,8580	0,1420	0,9377	0,0623	0,8580	0,1420	0,9379	0,0621	0,8583	0,1417
7	10	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
7	20	0,9740	0,0260	0,2597	0,7403	0,9740	0,0260	0,2597	0,7403	0,9740	0,0260	0,2597	0,7403	0,9740	0,0260	0,2597	0,7403	0,9740	0,0260	0,2597	0,7403
9	3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
9	4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
9	10	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
9	20	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000



ANEXO K – Número de experimentos, por cenário, dentre os que exibiram a situação de divergência 2 ou 4, que apresentaram resultado não significativo (NS) e significativo (S) para os testes de Levene e Shapiro-Wilk, respectivamente, ao nível de significância de 5%, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 5%, com  $\tau_1 = -2, \tau_2 = 0, \tau_3 = 2$  para os cenários com 3 tratamentos,  $\tau_1 = -2, \tau_2 = -1, \tau_3 = 0, \tau_4 = 1, \tau_5 = 2$  para os cenários com 5 tratamentos,  $\tau_1 = -3, \tau_2 = -2, \tau_3 = -1, \tau_4 = 0, \tau_5 = 1, \tau_6 = 2, \tau_7 = 3$  para os cenários com 7 tratamentos e  $\tau_1 = -4, \tau_2 = -3, \tau_3 = -2, \tau_4 = -1, \tau_5 = 0, \tau_6 = 1, \tau_7 = 2, \tau_8 = 3, \tau_9 = 4$  para os cenários com 9 tratamentos

TRAT	REP	Tukey				Duncan				DMS de Fisher				SNK				Scheffé			
		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk	
		NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S
3	3	1,0000	0,0000	0,9412	0,0588	1,0000	0,0000	0,9857	0,0143	1,0000	0,0000	0,9886	0,0114	1,0000	0,0000	0,9429	0,0571	1,0000	0,0000	0,9600	0,0400
3	4	0,9118	0,0882	1,0000	0,0000	0,9330	0,0670	0,9721	0,0279	0,9447	0,0553	0,9770	0,0230	0,9000	0,1000	1,0000	0,0000	0,9348	0,0652	0,9783	0,0217
3	10	0,9818	0,0200	0,9818	0,0182	1,0000	0,0000	0,9342	0,0658	0,9817	0,0200	0,9388	0,0612	1,0000	0,0000	0,9322	0,0678	0,9342	0,0700	0,9605	0,0395
3	20	0,9512	0,0488	0,9512	0,0488	0,9611	0,0389	0,9455	0,0545	0,9649	0,0351	0,9425	0,0575	0,9718	0,0282	0,9577	0,0423	0,9545	0,0455	0,9205	0,0795
5	3	1,0000	0,0000	0,9800	0,0200	1,0000	0,0000	0,9600	0,0400	1,0000	0,0000	0,9600	0,0400	1,0000	0,0000	0,9700	0,0300	1,0000	0,0000	0,9800	0,0200
5	4	1,0000	0,0000	0,9595	0,0405	1,0000	0,0000	0,9736	0,0264	1,0000	0,0000	0,9707	0,0293	1,0000	0,0000	0,9452	0,0548	1,0000	0,0000	0,9524	0,0476
5	10	0,9630	0,0370	0,9185	0,0815	0,9637	0,0363	0,9542	0,0458	0,9682	0,0318	0,9536	0,0464	0,9597	0,0403	0,9194	0,0806	0,9690	0,0310	0,9457	0,0543
5	20	0,9486	0,0514	0,9257	0,0743	0,9575	0,0425	0,9329	0,0671	0,9611	0,0389	0,9409	0,0591	0,9542	0,0458	0,9085	0,0915	0,9596	0,0404	0,9621	0,0379
7	3	1,0000	0,0000	0,9754	0,0246	1,0000	0,0000	0,9721	0,0279	1,0000	0,0000	0,9709	0,0291	1,0000	0,0000	0,9835	0,0165	1,0000	0,0000	0,9910	0,0090
7	4	0,9231	0,0769	0,9441	0,0559	0,9233	0,0767	0,9583	0,0417	0,9335	0,0665	0,9580	0,0420	0,9265	0,0735	0,9485	0,0515	0,9619	0,0381	0,9492	0,0508
7	10	0,9721	0,0279	0,9535	0,0465	0,9639	0,0361	0,9639	0,0361	0,9677	0,0323	0,9593	0,0407	0,9760	0,0240	0,9519	0,0481	0,9769	0,0231	0,9429	0,0571
7	20	0,9596	0,0404	0,9495	0,0505	0,9878	0,0122	0,9756	0,0244	0,9890	0,0110	0,9780	0,0220	0,9535	0,0465	0,9535	0,0465	0,9599	0,0401	0,9421	0,0579
9	3	1,0000	0,0000	0,9843	0,0157	1,0000	0,0000	0,9577	0,0423	1,0000	0,0000	0,9643	0,0357	1,0000	0,0000	0,9755	0,0245	1,0000	0,0000	0,9651	0,0349
9	4	0,9677	0,0323	0,9556	0,0444	0,9256	0,0744	0,9517	0,0483	0,9265	0,0735	0,9549	0,0451	0,9617	0,0383	0,9579	0,0421	0,9493	0,0507	0,9508	0,0492
9	10	0,9610	0,0390	0,9481	0,0519	0,9365	0,0635	0,9683	0,0317	0,9370	0,0630	0,9685	0,0315	0,9613	0,0387	0,9448	0,0552	0,9735	0,0265	0,9422	0,0578
9	20	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	0,9750	0,0250	0,9500	0,0500

ANEXO L – Proporção de experimentos, por cenário, dentre os que exibiram a situação de divergência 2 ou 4, que apresentaram resultado não significativo (NS) e significativo (S) para os testes de Levene e Shapiro-Wilk, respectivamente, ao nível de significância de 5%, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 5%, com  $\tau_1 = -0,5, \tau_2 = 0, \tau_3 = 0,5$  para os cenários com 3 tratamentos,  $\tau_1 = -1, \tau_2 = -0,5, \tau_3 = 0, \tau_4 = 0,5, \tau_5 = 1$  para os cenários com 5 tratamentos,  $\tau_1 = -1,5, \tau_2 = -1, \tau_3 = -0,5, \tau_4 = 0, \tau_5 = 0,5, \tau_6 = 1, \tau_7 = 1,5$  para os cenários com 7 tratamentos e  $\tau_1 = -2, \tau_2 = -1,5, \tau_3 = -1, \tau_4 = -0,5, \tau_5 = 0, \tau_6 = 0,5, \tau_7 = 1, \tau_8 = 1,5, \tau_9 = 2$  para os cenários com 9 tratamentos

TRAT	REP	Tukey				Duncan				DMS de Fisher				SNK				Scheffé			
		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk	
		NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S
3	3	1,0000	0,0000	0,9412	0,0588	1,0000	0,0000	0,9857	0,0143	1,0000	0,0000	0,9886	0,0114	1,0000	0,0000	0,9429	0,0571	1,0000	0,0000	0,9600	0,0400
3	4	1,0000	0,0000	1,0000	0,0000	0,9798	0,0202	1,0000	0,0000	0,9767	0,0233	0,9922	0,0078	1,0000	0,0000	1,0000	0,0000	0,9130	0,0870	0,9565	0,0435
3	10	1,0000	0,0000	1,0000	0,0000	0,9831	0,0200	1,0000	0,0000	0,9816	0,0200	0,9939	0,0061	1,0000	0,0000	1,0000	0,0000	0,9259	0,0700	0,9630	0,0370
3	20	0,9677	0,0323	0,9355	0,0645	0,9434	0,0566	0,9623	0,0377	0,9545	0,0455	0,9675	0,0325	0,9412	0,0588	0,8824	0,1176	0,9394	0,0606	0,9394	0,0606
5	3	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	0,9700	0,0300	1,0000	0,0000	0,9700	0,0300	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	0,9800	0,0200
5	4	0,9783	0,0217	0,9783	0,0217	0,9665	0,0335	0,9634	0,0366	0,9644	0,0356	0,9665	0,0335	0,9773	0,0227	0,9773	0,0227	0,9483	0,0517	0,9483	0,0517
5	10	0,9770	0,0230	0,9425	0,0575	0,9674	0,0326	0,9442	0,0558	0,9661	0,0339	0,9445	0,0555	0,9780	0,0220	0,9341	0,0659	0,9583	0,0417	0,9792	0,0208
5	20	0,9877	0,0123	0,9136	0,0864	0,9568	0,0432	0,9386	0,0614	0,9610	0,0390	0,9505	0,0495	0,9872	0,0128	0,9231	0,0769	0,9548	0,0452	0,9209	0,0791
7	3	1,0000	0,0000	0,9531	0,0469	1,0000	0,0000	0,9789	0,0211	1,0000	0,0000	0,9767	0,0233	1,0000	0,0000	0,9667	0,0333	1,0000	0,0000	0,9576	0,0424
7	4	0,9697	0,0303	0,9848	0,0152	0,9476	0,0524	0,9548	0,0452	0,9329	0,0671	0,9532	0,0468	0,9718	0,0282	0,9859	0,0141	0,9364	0,0636	0,9909	0,0091
7	10	0,9683	0,0317	0,9286	0,0714	0,9699	0,0301	0,9366	0,0634	0,9738	0,0262	0,9403	0,0597	0,9667	0,0333	0,9250	0,0750	0,9583	0,0417	0,9394	0,0606
7	20	0,9507	0,0493	0,9360	0,0640	0,9607	0,0393	0,9622	0,0378	0,9577	0,0423	0,9544	0,0456	0,9502	0,0498	0,9303	0,0697	0,9492	0,0508	0,9316	0,0684
9	3	1,0000	0,0000	0,9885	0,0115	1,0000	0,0000	0,9615	0,0385	1,0000	0,0000	0,9656	0,0344	1,0000	0,0000	0,9688	0,0313	1,0000	0,0000	0,9371	0,0629
9	4	0,9821	0,0179	0,9554	0,0446	0,9383	0,0617	0,9463	0,0537	0,9357	0,0643	0,9540	0,0460	0,9750	0,0250	0,9583	0,0417	0,9691	0,0309	0,9588	0,0412
9	10	0,9785	0,0215	0,9614	0,0386	0,9747	0,0253	0,9468	0,0532	0,9718	0,0282	0,9435	0,0565	0,9755	0,0245	0,9633	0,0367	0,9851	0,0149	0,9590	0,0410
9	20	0,9637	0,0363	0,9373	0,0627	0,9579	0,0421	0,9521	0,0479	0,9587	0,0413	0,9521	0,0479	0,9664	0,0336	0,9419	0,0581	0,9670	0,0330	0,9506	0,0494

ANEXO M – Proporção de experimentos, por cenário, dentre os que exibiram a situação de divergência 2 ou 4, que apresentaram resultado não significativo (NS) e significativo (S) para os testes de Levene e Shapiro-Wilk, respectivamente, ao nível de significância de 5%, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 10%, com  $\tau_1 = -2, \tau_2 = 0, \tau_3 = 2$  para os cenários com 3 tratamentos,  $\tau_1 = -2, \tau_2 = -1, \tau_3 = 0, \tau_4 = 1, \tau_5 = 2$  para os cenários com 5 tratamentos,  $\tau_1 = -3, \tau_2 = -2, \tau_3 = -1, \tau_4 = 0, \tau_5 = 1, \tau_6 = 2, \tau_7 = 3$  para os cenários com 7 tratamentos e  $\tau_1 = -4, \tau_2 = -3, \tau_3 = -2, \tau_4 = -1, \tau_5 = 0, \tau_6 = 1, \tau_7 = 2, \tau_8 = 3, \tau_9 = 4$  para os cenários com 9 tratamentos

TRAT	REP	Tukey				Duncan				DMS de Fisher				SNK				Scheffé			
		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk	
		NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S
3	3	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	0,9811	0,0189	1,0000	0,0000	0,9841	0,0159	1,0000	0,0000	0,9565	0,0435	1,0000	0,0000	1,0000	0,0000
3	4	1,0000	0,0000	1,0000	0,0000	0,9573	0,0427	0,9829	0,0171	0,9252	0,0748	0,9864	0,0136	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	0,9643	0,0357
3	10	0,9706	0,0300	0,9706	0,0294	0,9485	0,0500	0,9265	0,0735	0,9630	0,0400	0,9491	0,0509	0,9412	0,0600	0,9412	0,0588	0,9524	0,0500	0,9286	0,0714
3	20	0,9722	0,0278	0,9444	0,0556	0,9624	0,0376	0,9624	0,0376	0,9663	0,0337	0,9625	0,0375	0,9394	0,0606	0,9394	0,0606	0,9434	0,0566	0,9623	0,0377
5	3	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	0,9700	0,0300	1,0000	0,0000	0,9700	0,0300	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	0,9600	0,0400
5	4	0,9815	0,0185	0,9444	0,0556	0,9446	0,0554	0,9723	0,0277	0,9562	0,0438	0,9687	0,0313	0,9574	0,0426	0,9149	0,0851	0,9722	0,0278	0,9583	0,0417
5	10	1,0000	0,0000	0,9841	0,0159	0,9738	0,0262	0,9563	0,0437	0,9656	0,0344	0,9584	0,0416	1,0000	0,0000	0,9804	0,0196	0,9685	0,0315	0,9606	0,0394
5	20	0,9878	0,0122	0,8902	0,1098	0,9610	0,0390	0,9518	0,0482	0,9606	0,0394	0,9576	0,0424	0,9863	0,0137	0,9041	0,0959	0,9674	0,0326	0,9076	0,0924
7	3	1,0000	0,0000	0,9692	0,0308	1,0000	0,0000	0,9807	0,0193	1,0000	0,0000	0,9796	0,0204	1,0000	0,0000	0,9701	0,0299	1,0000	0,0000	0,9754	0,0246
7	4	0,9828	0,0172	0,9828	0,0172	0,9485	0,0515	0,9618	0,0382	0,9381	0,0619	0,9550	0,0450	0,9844	0,0156	0,9844	0,0156	0,9194	0,0806	0,9677	0,0323
7	10	0,9829	0,0171	0,9573	0,0427	0,9751	0,0249	0,9393	0,0607	0,9741	0,0259	0,9431	0,0569	0,9821	0,0179	0,9554	0,0446	0,9641	0,0359	0,9562	0,0438
7	20	0,9569	0,0431	0,8852	0,1148	0,9612	0,0388	0,9310	0,0690	0,9619	0,0381	0,9419	0,0581	0,9557	0,0443	0,8719	0,1281	0,9920	0,0080	0,9800	0,0200
9	3	1,0000	0,0000	0,9884	0,0116	1,0000	0,0000	0,9629	0,0371	1,0000	0,0000	0,9676	0,0324	1,0000	0,0000	0,9783	0,0217	1,0000	0,0000	0,9577	0,0423
9	4	0,9808	0,0192	0,9712	0,0288	0,9427	0,0573	0,9521	0,0479	0,9378	0,0622	0,9508	0,0492	0,9732	0,0268	0,9732	0,0268	0,9684	0,0316	0,9474	0,0526
9	10	0,9749	0,0251	0,9665	0,0335	0,9751	0,0249	0,9440	0,0560	0,9727	0,0273	0,9426	0,0574	0,9764	0,0236	0,9606	0,0394	0,9870	0,0130	0,9556	0,0444
9	20	0,9662	0,0338	0,9459	0,0541	0,9505	0,0495	0,9562	0,0438	0,9542	0,0458	0,9559	0,0441	0,9686	0,0314	0,9434	0,0566	0,9628	0,0372	0,9535	0,0465



ANEXO O – Proporção de experimentos, por cenário, dentre os que exibiram a situação de divergência 2 ou 4, que apresentaram resultado não significativo (NS) e significativo (S) para os testes de Levene e Shapiro-Wilk, respectivamente, ao nível de significância de 5%, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 20%, com  $\tau_1 = -2, \tau_2 = 0, \tau_3 = 2$  para os cenários com 3 tratamentos,  $\tau_1 = -2, \tau_2 = -1, \tau_3 = 0, \tau_4 = 1, \tau_5 = 2$  para os cenários com 5 tratamentos,  $\tau_1 = -3, \tau_2 = -2, \tau_3 = -1, \tau_4 = 0, \tau_5 = 1, \tau_6 = 2, \tau_7 = 3$  para os cenários com 7 tratamentos e  $\tau_1 = -4, \tau_2 = -3, \tau_3 = -2, \tau_4 = -1, \tau_5 = 0, \tau_6 = 1, \tau_7 = 2, \tau_8 = 3, \tau_9 = 4$  para os cenários com 9 tratamentos

TRAT	REP	Tukey				Duncan				DMS de Fisher				SNK				Scheffé			
		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk	
		NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S
3	3	1,0000	0,0000	0,9412	0,0588	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	0,9474	0,0526
3	4	1,0000	0,0000	1,0000	0,0000	0,9712	0,0288	0,9808	0,0192	0,9612	0,0388	0,9845	0,0155	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	0,9545	0,0455
3	10	1,0000	0,0000	0,9167	0,0833	0,9750	0,0300	0,9583	0,0417	0,9686	0,0300	0,9434	0,0566	1,0000	0,0000	1,0000	0,0000	0,9600	0,0400	0,8800	0,1200
3	20	1,0000	0,0000	0,9500	0,0500	0,9504	0,0496	0,9669	0,0331	0,9634	0,0366	0,9634	0,0366	1,0000	0,0000	1,0000	0,0000	0,9231	0,0769	0,9615	0,0385
5	3	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	0,9700	0,0300	1,0000	0,0000	0,9800	0,0200	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	0,9800	0,0200
5	4	0,9048	0,0952	0,9524	0,0476	0,9575	0,0425	0,9641	0,0359	0,9555	0,0445	0,9666	0,0334	0,9268	0,0732	0,9512	0,0488	0,9444	0,0556	0,9815	0,0185
5	10	0,9565	0,0435	0,9565	0,0435	0,9859	0,0141	0,9399	0,0601	0,9775	0,0225	0,9416	0,0584	0,9608	0,0392	0,9804	0,0196	0,9487	0,0513	0,9744	0,0256
5	20	0,9333	0,0667	0,9667	0,0333	0,9833	0,0167	0,9415	0,0585	0,9801	0,0199	0,9495	0,0505	0,9483	0,0517	0,9655	0,0345	0,8846	0,1154	0,9231	0,0769
7	3	1,0000	0,0000	0,9636	0,0364	1,0000	0,0000	0,9752	0,0248	1,0000	0,0000	0,9732	0,0268	1,0000	0,0000	0,9649	0,0351	1,0000	0,0000	0,9753	0,0247
7	4	0,9400	0,0600	0,9600	0,0400	0,9441	0,0559	0,9586	0,0414	0,9413	0,0587	0,9582	0,0418	0,9444	0,0556	0,9630	0,0370	0,9000	0,1000	1,0000	0,0000
7	10	0,9429	0,0571	0,9286	0,0714	0,9725	0,0275	0,9492	0,0508	0,9787	0,0213	0,9460	0,0540	0,9467	0,0533	0,9333	0,0667	0,9658	0,0342	0,9316	0,0684
7	20	0,9632	0,0368	0,9706	0,0294	0,9458	0,0542	0,9528	0,0472	0,9523	0,0477	0,9579	0,0421	0,9653	0,0347	0,9722	0,0278	0,9750	0,0250	0,9000	0,1000
9	3	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	0,9619	0,0381	1,0000	0,0000	0,9691	0,0309	1,0000	0,0000	0,9811	0,0189	1,0000	0,0000	0,9767	0,0233
9	4	0,9677	0,0323	0,9516	0,0484	0,9574	0,0426	0,9506	0,0494	0,9604	0,0396	0,9500	0,0500	0,9841	0,0159	0,9524	0,0476	0,9748	0,0252	0,9580	0,0420
9	10	0,9643	0,0357	0,9762	0,0238	0,9791	0,0209	0,9555	0,0445	0,9767	0,0233	0,9526	0,0474	0,9659	0,0341	0,9773	0,0227	0,9818	0,0182	0,9879	0,0121
9	20	0,9714	0,0286	0,9257	0,0743	0,9714	0,0286	0,9552	0,0448	0,9689	0,0311	0,9500	0,0500	0,9728	0,0272	0,9239	0,0761	0,9550	0,0450	0,9486	0,0514



ANEXO P – Proporção de experimentos, por cenário, dentre os que exibiram a situação de divergência 2 ou 4, que apresentaram resultado não significativo (NS) e significativo (S) para os testes de Levene e Shapiro-Wilk, respectivamente, ao nível de significância de 5%, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 20%, com  $\tau_1 = -0,5, \tau_2 = 0, \tau_3 = 0,5$  para os cenários com 3 tratamentos,  $\tau_1 = -1, \tau_2 = -0,5, \tau_3 = 0, \tau_4 = 0,5, \tau_5 = 1$  para os cenários com 5 tratamentos,  $\tau_1 = -1,5, \tau_2 = -1, \tau_3 = -0,5, \tau_4 = 0, \tau_5 = 0,5, \tau_6 = 1, \tau_7 = 1,5$  para os cenários com 7 tratamentos e  $\tau_1 = -2, \tau_2 = -1,5, \tau_3 = -1, \tau_4 = -0,5, \tau_5 = 0, \tau_6 = 0,5, \tau_7 = 1, \tau_8 = 1,5, \tau_9 = 2$  para os cenários com 9 tratamentos

TRAT	REP	Tukey				Duncan				DMS de Fisher				SNK				Scheffé			
		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk		Levene		Shapiro-Wilk	
		NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S
3	3	1,0000	0,0000	0,9412	0,0588	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	0,9474	0,0526
3	4	1,0000	0,0000	1,0000	0,0000	0,9670	0,0330	0,9890	0,0110	0,9685	0,0315	0,9685	0,0315	1,0000	0,0000	1,0000	0,0000	0,9643	0,0357	0,9643	0,0357
3	10	0,9474	0,0500	1,0000	0,0000	1,0000	0,0000	0,9259	0,0741	0,9664	0,0300	0,9329	0,0671	1,0000	0,0000	0,9474	0,0526	0,8750	0,1300	1,0000	0,0000
3	20	0,9545	0,0455	1,0000	0,0000	0,9875	0,0125	0,9375	0,0625	0,9828	0,0172	0,9310	0,0690	0,9444	0,0556	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000
5	3	1,0000	0,0000	0,9600	0,0400	1,0000	0,0000	0,9700	0,0300	1,0000	0,0000	0,9800	0,0200	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	0,9800	0,0200
5	4	0,9429	0,0571	0,9714	0,0286	0,9490	0,0510	0,9660	0,0340	0,9514	0,0486	0,9653	0,0347	0,9167	0,0833	0,9722	0,0278	0,9623	0,0377	0,9623	0,0377
5	10	0,9375	0,0625	0,9688	0,0313	0,9704	0,0296	0,9519	0,0481	0,9749	0,0251	0,9452	0,0548	0,9444	0,0556	0,9722	0,0278	0,9661	0,0339	0,9322	0,0678
5	20	0,9767	0,0233	0,9535	0,0465	0,9838	0,0162	0,9385	0,0615	0,9779	0,0221	0,9437	0,0563	0,9767	0,0233	0,9535	0,0465	0,9275	0,0725	0,9275	0,0725
7	3	1,0000	0,0000	0,9756	0,0244	1,0000	0,0000	0,9742	0,0258	1,0000	0,0000	0,9738	0,0262	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	0,9710	0,0290
7	4	0,9474	0,0526	0,9737	0,0263	0,9625	0,0375	0,9532	0,0468	0,9406	0,0594	0,9576	0,0424	0,9500	0,0500	0,9750	0,0250	0,9104	0,0896	0,9851	0,0149
7	10	0,9778	0,0222	0,8444	0,1556	0,9680	0,0320	0,9433	0,0567	0,9773	0,0227	0,9453	0,0547	0,9767	0,0233	0,8372	0,1628	0,9750	0,0250	0,9500	0,0500
7	20	0,9815	0,0185	0,9444	0,0556	0,9552	0,0448	0,9701	0,0299	0,9485	0,0515	0,9571	0,0429	0,9821	0,0179	0,9464	0,0536	0,9556	0,0444	0,9111	0,0889
9	3	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	0,9700	0,0300	1,0000	0,0000	0,9713	0,0287	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	0,9863	0,0137
9	4	1,0000	0,0000	0,9200	0,0800	0,9491	0,0509	0,9527	0,0473	0,9474	0,0526	0,9539	0,0461	1,0000	0,0000	0,9216	0,0784	0,9688	0,0313	0,9375	0,0625
9	10	0,9730	0,0270	1,0000	0,0000	0,9741	0,0259	0,9637	0,0363	0,9774	0,0226	0,9547	0,0453	0,9726	0,0274	1,0000	0,0000	1,0000	0,0000	0,9905	0,0095
9	20	0,9457	0,0543	0,9457	0,0543	0,9701	0,0299	0,9544	0,0456	0,9693	0,0307	0,9531	0,0469	0,9474	0,0526	0,9474	0,0526	0,9583	0,0417	0,9444	0,0556

ANEXO Q – Taxas de erro tipo I por comparação ( $\alpha_c$ ) e por experimento ( $\alpha_e$ ) incondicionais a um resultado significativo do teste F global para diferentes testes de comparação de médias ao nível de significância nominal de 5%, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 1%

TRAT	REP	Tukey		Duncan		DMS de Fisher		SNK		Scheffé	
		$\alpha_c$	$\alpha_e$	$\alpha_c$	$\alpha_e$	$\alpha_c$	$\alpha_e$	$\alpha_c$	$\alpha_e$	$\alpha_c$	$\alpha_e$
3	3	0,0177*	0,0395	0,0430	0,0950**	0,0500	0,1120**	0,0257*	0,0500	0,0175*	0,0390
3	4	0,0193*	0,0465	0,0443	0,0980**	0,0488	0,1115**	0,0273*	0,0570	0,0133*	0,0335*
3	10	0,0192*	0,0490	0,0408	0,0965**	0,0482	0,1165**	0,0262*	0,0575	0,0153*	0,0395
3	20	0,0170*	0,0445	0,0362*	0,0850**	0,0425	0,1030**	0,0238*	0,0510	0,0128*	0,0325*
5	3	0,0079*	0,0495	0,0390	0,1800**	0,0513	0,2420**	0,0123*	0,0620	0,0033*	0,0225*
5	4	0,0074*	0,0540	0,0389	0,1990**	0,0533	0,2610**	0,0104*	0,0645**	0,0033*	0,0260*
5	10	0,0071*	0,0480	0,0333*	0,1815**	0,0496	0,2625**	0,0103*	0,0515	0,0029*	0,0230*
5	20	0,0068*	0,0540	0,0316	0,1820**	0,0495	0,2780**	0,0094*	0,0585	0,0025*	0,0215*
7	3	0,0047*	0,0500	0,0370*	0,2770**	0,0551	0,3940**	0,0080*	0,0580	0,0014*	0,0175*
7	4	0,0032*	0,0455	0,0295*	0,2560**	0,0490	0,3930**	0,0049*	0,0505	0,0008*	0,0115*
7	10	0,0034*	0,0490	0,0265*	0,2605**	0,0479	0,4130**	0,0046*	0,0500	0,0004*	0,0080*
7	20	0,0030*	0,0455	0,0255*	0,2550**	0,0479	0,4290**	0,0041*	0,0460	0,0004*	0,0055*
9	3	0,0028*	0,0460	0,0288*	0,3125**	0,0494	0,4720**	0,0044*	0,0510	0,0003*	0,0080*
9	4	0,0027*	0,0530	0,0266*	0,3255**	0,0502	0,5025**	0,0046*	0,0545	0,0002*	0,0045*
9	10	0,0017*	0,0450	0,0231*	0,3225**	0,0578	0,5335**	0,0022*	0,0460	0,0001*	0,0025*
9	20	0,0019*	0,0475	0,0233*	0,3300**	0,0494	0,5470**	0,0025*	0,0480	0,0000*	0,0015*

\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram abaixo do limite inferior do intervalo de confiança com 95% de probabilidade (0,0375) para a proporção empírica desta taxa.

\*\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram acima do limite superior do intervalo de confiança de 95% de probabilidade (0,0625) para a proporção empírica desta taxa.

ANEXO R – Taxas de erro tipo I por comparação ( $\alpha_1$ ) e por experimento ( $\alpha_3$ ) condicionais a um resultado significativo do teste F global para diferentes testes de comparação de médias ao nível de significância nominal de 5%, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 1%

TRAT	REP	Tukey		Duncan		DMS de Fisher		SNK		Scheffé	
		$\alpha_1$	$\alpha_3$	$\alpha_1$	$\alpha_3$	$\alpha_1$	$\alpha_3$	$\alpha_1$	$\alpha_3$	$\alpha_1$	$\alpha_3$
3	3	0,0175*	0,0390	0,0258*	0,0445	0,0258*	0,0445	0,0241*	0,0445	0,0175*	0,0390
3	4	0,0165*	0,0380	0,0218*	0,0395	0,0218*	0,0395	0,0218*	0,0395	0,0133*	0,0335*
3	10	0,0181*	0,0460	0,0253*	0,0515	0,0253*	0,0515	0,0238*	0,0510	0,0153*	0,0395
3	20	0,0161*	0,0420	0,0230*	0,0470	0,0230*	0,0470	0,0220*	0,0460	0,0128*	0,0325*
5	3	0,0073*	0,0445	0,0198*	0,0560	0,0210*	0,0560	0,0112*	0,0515	0,0032*	0,0225*
5	4	0,0062*	0,0425	0,0159*	0,0530	0,0188*	0,0530	0,0090*	0,0490	0,0033*	0,0260*
5	10	0,0062*	0,0390	0,0146*	0,0470	0,0162*	0,0470	0,0091*	0,0415	0,0029*	0,0230*
5	20	0,0055*	0,0420	0,0142*	0,0535	0,0166*	0,0535	0,0079*	0,0445	0,0024*	0,0215*
7	3	0,0045*	0,0480	0,0197*	0,0815**	0,0230*	0,0815**	0,0070*	0,0525	0,0014*	0,0175*
7	4	0,0026*	0,0330*	0,0097*	0,0415	0,0117*	0,0415	0,0037*	0,0345*	0,0007*	0,0115*
7	10	0,0026*	0,0340*	0,0092*	0,0430	0,0115*	0,0430	0,0038*	0,0350*	0,0003*	0,0080*
7	20	0,0023*	0,0315*	0,0085*	0,0430	0,0111*	0,0430	0,0032*	0,0315*	0,0003*	0,0055*
9	3	0,0024*	0,0350*	0,0112*	0,0500	0,0137*	0,0500	0,0038*	0,0385	0,0003*	0,0080*
9	4	0,0022*	0,0375	0,0099*	0,0510	0,0134*	0,0510	0,0036*	0,0385	0,0002*	0,0045*
9	10	0,0012*	0,0295*	0,0075*	0,0460	0,0113*	0,0460	0,0015*	0,0295*	0,0000*	0,0025*
9	20	0,0015*	0,0335*	0,0080*	0,0475	0,0110*	0,0475	0,0020*	0,0335*	0,0000*	0,0015*

\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram abaixo do limite inferior do intervalo de confiança com 95% de probabilidade (0,0375) para a proporção empírica desta taxa.

\*\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram acima do limite superior do intervalo de confiança de 95% de probabilidade (0,0625) para a proporção empírica desta taxa.

ANEXO S – Taxas de erro tipo I por comparação ( $\alpha_2$ ) e por experimento ( $\alpha_4$ ) condicionais a um resultado significativo do teste F global para diferentes testes de comparação de médias ao nível de significância nominal de 5%, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 1%

TRAT	REP	Tukey		Duncan		DMS de Fisher		SNK		Scheffé	
		$\alpha_2$	$\alpha_4$	$\alpha_2$	$\alpha_4$	$\alpha_2$	$\alpha_4$	$\alpha_2$	$\alpha_4$	$\alpha_2$	$\alpha_4$
3	3	0,4487**	0,8764**	0,5805**	1,0000**	0,5805**	1,0000**	0,5753**	1,0000**	0,4487**	0,8764**
3	4	0,4342**	0,9620**	0,5527**	1,0000**	0,5527**	1,0000**	0,5527**	1,0000**	0,3980**	0,8481**
3	10	0,3949**	0,8932**	0,4919**	1,0000**	0,4919**	1,0000**	0,4863**	0,9902**	0,3881**	0,7669**
3	20	0,3849**	0,8936**	0,4893**	1,0000**	0,4893**	1,0000**	0,4888**	0,9787**	0,3948**	0,6914**
5	3	0,1651**	0,7946**	0,3544**	1,0000**	0,3750**	1,0000**	0,2472**	0,9196**	0,1444**	0,4017**
5	4	0,1470**	0,8018**	0,3009**	1,0000**	0,3556**	1,0000**	0,1989**	0,9245**	0,1269**	0,4905**
5	10	0,1589**	0,8297**	0,3117**	1,0000**	0,3446**	1,0000**	0,2287**	0,8829**	0,1260**	0,4893**
5	20	0,1321**	0,7850**	0,2663**	1,0000**	0,3102**	1,0000**	0,1827**	0,8317**	0,1139**	0,4018**
7	3	0,0952**	0,5889**	0,2424**	1,0000**	0,2822**	1,0000**	0,1400**	0,6441**	0,0816**	0,2147**
7	4	0,0793**	0,7951**	0,2352**	1,0000**	0,2839**	1,0000**	0,1125**	0,8313**	0,0662**	0,2771**
7	10	0,0777**	0,7906**	0,2159**	1,0000**	0,2685**	1,0000**	0,1088**	0,8139**	0,0476	0,1860**
7	20	0,0755**	0,7325**	0,1993**	1,0000**	0,2591**	1,0000**	0,1035**	0,7325**	0,0649**	0,1279**
9	3	0,0702**	0,7000**	0,2244**	1,0000**	0,2750**	1,0000**	0,1039**	0,7700**	0,0399	0,1600**
9	4	0,0600	0,7352**	0,1944**	1,0000**	0,2630**	1,0000**	0,0950**	0,7549**	0,0524	0,0882**
9	10	0,0428	0,6413**	0,1630**	1,0000**	0,2469**	1,0000**	0,0541	0,6413**	0,0388	0,0543
9	20	0,0456	0,7052**	0,1687**	1,0000**	0,2318**	1,0000**	0,0613	0,7052**	0,0277	0,0315

\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram abaixo do limite inferior do intervalo de confiança com 95% de probabilidade (0,0375) para a proporção empírica desta taxa.

\*\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram acima do limite superior do intervalo de confiança de 95% de probabilidade (0,0625) para a proporção empírica desta taxa.

ANEXO T – Taxas de erro tipo I por comparação ( $\alpha_c$ ) e por experimento ( $\alpha_e$ ) incondicionais a um resultado significativo do teste F global para diferentes testes de comparação de médias ao nível de significância nominal de 5%, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 5%

TRAT	REP	Tukey		Duncan		DMS de Fisher		SNK		Scheffé	
		$\alpha_c$	$\alpha_e$	$\alpha_c$	$\alpha_e$	$\alpha_c$	$\alpha_e$	$\alpha_c$	$\alpha_e$	$\alpha_c$	$\alpha_e$
3	3	0,0201*	0,0470	0,0433	0,0925**	0,0471	0,1040**	0,0288*	0,0540	0,0163*	0,0385
3	4	0,0215*	0,0510	0,0471	0,1025**	0,0516	0,1160**	0,0298*	0,0595	0,0160*	0,0365*
3	10	0,0201*	0,0495	0,0398	0,0930**	0,0471	0,1150**	0,0255*	0,0530	0,0165*	0,0410
3	20	0,0173*	0,0450	0,0383	0,0885**	0,0455	0,1090**	0,0246*	0,0545	0,0130*	0,0345*
5	3	0,0075*	0,0470	0,0364*	0,1720**	0,0483	0,2405**	0,0123*	0,0515	0,0034*	0,0255*
5	4	0,0064*	0,0445	0,0372*	0,1970**	0,0512	0,2645**	0,0101*	0,0495	0,0029*	0,0220*
5	10	0,0074*	0,0525	0,0340*	0,1840**	0,0502	0,2725**	0,0101*	0,0565	0,0030*	0,0225*
5	20	0,0059*	0,0480	0,0328*	0,1965**	0,0495	0,2815**	0,0081*	0,0530	0,0021*	0,0195*
7	3	0,0048*	0,0520	0,0354*	0,2765**	0,0522	0,3850**	0,0077*	0,0550	0,0013*	0,0140*
7	4	0,0035*	0,0450	0,0282*	0,2535**	0,0478	0,3910**	0,0051*	0,0470	0,0008*	0,0115*
7	10	0,0031*	0,0470	0,0251*	0,2410**	0,0449	0,4030**	0,0044*	0,0485	0,0005*	0,0100*
7	20	0,0025*	0,0405	0,0251*	0,2510**	0,0462	0,4160**	0,0038*	0,0420	0,0002*	0,0055*
9	3	0,0023*	0,0415	0,0285*	0,3315**	0,0622	0,4880**	0,0033*	0,0430	0,0003*	0,0075*
9	4	0,0040*	0,0445	0,0403	0,3225**	0,0496	0,5085**	0,0058*	0,0460	0,0007*	0,0080*
9	10	0,0020*	0,0490	0,0228*	0,3160**	0,0475	0,5330**	0,0026*	0,0505	0,0001*	0,0035*
9	20	0,0035*	0,0520	0,0387	0,3260**	0,0492	0,5535**	0,0030*	0,0530	0,0000*	0,0025*

\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram abaixo do limite inferior do intervalo de confiança com 95% de probabilidade (0,0375) para a proporção empírica desta taxa.

\*\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram acima do limite superior do intervalo de confiança de 95% de probabilidade (0,0625) para a proporção empírica desta taxa.

ANEXO U – Taxas de erro tipo I por comparação ( $\alpha_1$ ) e por experimento ( $\alpha_3$ ) condicionais a um resultado significativo do teste F global para diferentes testes de comparação de médias ao nível de significância nominal de 5%, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 5%

TRAT	REP	Tukey		Duncan		DMS de Fisher		SNK		Scheffé	
		$\alpha_1$	$\alpha_3$	$\alpha_1$	$\alpha_3$	$\alpha_1$	$\alpha_3$	$\alpha_1$	$\alpha_3$	$\alpha_1$	$\alpha_3$
3	3	0,0193*	0,0445	0,0276*	0,0490	0,0276*	0,0490	0,0266*	0,0480	0,0163*	0,0385
3	4	0,0193*	0,0445	0,0276*	0,0485	0,0276*	0,0485	0,0256*	0,0485	0,0160*	0,0365*
3	10	0,0188*	0,0455	0,0243*	0,0480	0,0243*	0,0480	0,0238*	0,0475	0,0165*	0,0410
3	20	0,0158*	0,0405	0,0231*	0,0460	0,0231*	0,0460	0,0215*	0,0460	0,0130*	0,0345*
5	3	0,0068*	0,0405	0,0177*	0,0510	0,0188*	0,0510	0,0113*	0,0430	0,0034*	0,0255*
5	4	0,0056*	0,0365*	0,0158*	0,0465	0,0170*	0,0465	0,0092*	0,0400	0,0029*	0,0220*
5	10	0,0065*	0,0435	0,0153*	0,0515	0,0167*	0,0515	0,0088*	0,0460	0,0030*	0,0225*
5	20	0,0047*	0,0365*	0,0130*	0,0475	0,0144*	0,0475	0,0069*	0,0395	0,0021*	0,0195*
7	3	0,0043*	0,0410	0,0137*	0,0480	0,0154*	0,0480	0,0065*	0,0410	0,0013*	0,0140*
7	4	0,0030*	0,0345*	0,0099*	0,0425	0,0121*	0,0425	0,0042*	0,0355*	0,0008*	0,0115*
7	10	0,0025*	0,0355*	0,0095*	0,0440	0,0116*	0,0440	0,0037*	0,0365*	0,0005*	0,0100*
7	20	0,0020*	0,0285*	0,0087*	0,0410	0,0105*	0,0410	0,0029*	0,0285*	0,0002*	0,0055*
9	3	0,0021*	0,0345*	0,0103*	0,0445	0,0143*	0,0445	0,0030*	0,0355*	0,0003*	0,0075*
9	4	0,0034*	0,0350*	0,0135*	0,0510	0,0133*	0,0510	0,0049*	0,0365*	0,0007*	0,0080*
9	10	0,0016*	0,0330*	0,0080*	0,0480	0,0110*	0,0480	0,0020*	0,0330*	0,0001*	0,0035*
9	20	0,0026*	0,0360*	0,0116*	0,0465	0,0106*	0,0465	0,0024*	0,0370*	0,0000*	0,0025*

\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram abaixo do limite inferior do intervalo de confiança com 95% de probabilidade (0,0375) para a proporção empírica desta taxa.

\*\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram acima do limite superior do intervalo de confiança de 95% de probabilidade (0,0625) para a proporção empírica desta taxa.

ANEXO V – Taxas de erro tipo I por comparação ( $\alpha_2$ ) e por experimento ( $\alpha_4$ ) condicionais a um resultado significativo do teste F global para diferentes testes de comparação de médias ao nível de significância nominal de 5%, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 5%

TRAT	REP	Tukey		Duncan		DMS de Fisher		SNK		Scheffé	
		$\alpha_2$	$\alpha_4$	$\alpha_2$	$\alpha_4$	$\alpha_2$	$\alpha_4$	$\alpha_2$	$\alpha_4$	$\alpha_2$	$\alpha_4$
3	3	0,4344**	0,9081**	0,5646**	1,0000**	0,5646**	1,0000**	0,5673**	0,9795**	0,4242**	0,7857**
3	4	0,4344**	0,9175**	0,5704**	1,0000**	0,5704**	1,0000**	0,5641**	1,0000**	0,4383**	0,7525**
3	10	0,4139**	0,9479**	0,5069**	1,0000**	0,5069**	1,0000**	0,5070**	0,9895**	0,4024**	0,8541**
3	20	0,3909**	0,8804**	0,5036**	1,0000**	0,5036**	1,0000**	0,4942**	1,0000**	0,3768**	0,7500**
5	3	0,1691**	0,7941**	0,3470**	1,0000**	0,3696**	1,0000**	0,2658**	0,8431**	0,1333**	0,500**
5	4	0,1534**	0,7849**	0,3397**	1,0000**	0,3655**	1,0000**	0,2358**	0,8602**	0,1318**	0,4731**
5	10	0,1505**	0,8446**	0,2980**	1,0000**	0,3252**	1,0000**	0,1966**	0,8932**	0,1333**	0,4368**
5	20	0,1301**	0,7684**	0,2736**	1,0000**	0,3042**	1,0000**	0,1828**	0,8315**	0,1102**	0,4105**
7	3	0,1062**	0,8541**	0,2867**	1,0000**	0,3214**	1,0000**	0,1602**	0,8541**	0,0969**	0,2916**
7	4	0,0869**	0,8117**	0,2341**	1,0000**	0,2868**	1,0000**	0,1228**	0,8352**	0,0724**	0,2705**
7	10	0,0711**	0,8068**	0,2170**	1,0000**	0,2635**	1,0000**	0,1005**	0,8295**	0,0524	0,2273**
7	20	0,0710**	0,6951**	0,2125**	1,0000**	0,2566**	1,0000**	0,1035**	0,6951**	0,0519	0,1341**
9	3	0,0619	0,7752**	0,2315**	1,0000**	0,3224**	1,0000**	0,0873**	0,7977**	0,0407	0,1685**
9	4	0,0996**	0,6862**	0,2660**	1,0000**	0,2622**	1,0000**	0,1396**	0,7156**	0,0954**	0,1568**
9	10	0,0492	0,6875**	0,1684**	1,0000**	0,2309**	1,0000**	0,0606	0,6875**	0,0317	0,0729**
9	20	0,0744**	0,7741**	0,2505**	1,0000**	0,2290**	1,0000**	0,0669**	0,7956**	0,0277	0,0537

\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram abaixo do limite inferior do intervalo de confiança com 95% de probabilidade (0,0375) para a proporção empírica desta taxa.

\*\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram acima do limite superior do intervalo de confiança de 95% de probabilidade (0,0625) para a proporção empírica desta taxa.

ANEXO W – Taxas de erro tipo I por comparação ( $\alpha_c$ ) e por experimento ( $\alpha_e$ ) incondicionais a um resultado significativo do teste F global para diferentes testes de comparação de médias ao nível de significância nominal de 5%, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 10%

TRAT	REP	Tukey		Duncan		DMS de Fisher		SNK		Scheffé	
		$\alpha_c$	$\alpha_e$	$\alpha_c$	$\alpha_e$	$\alpha_c$	$\alpha_e$	$\alpha_c$	$\alpha_e$	$\alpha_c$	$\alpha_e$
3	3	0,0211*	0,0495	0,0435	0,0940**	0,0482	0,1070**	0,0291*	0,0540	0,0163*	0,0395
3	4	0,0203*	0,0475	0,0451	0,0980**	0,0505	0,1140**	0,0276*	0,0575	0,0158*	0,0375
3	10	0,0205*	0,0500	0,0416	0,0955**	0,0495	0,1190**	0,0258*	0,0570	0,0163*	0,0400
3	20	0,0178*	0,0465	0,0380	0,0865**	0,0455	0,1080**	0,0253*	0,0575	0,0126*	0,0330*
5	3	0,0075*	0,0450	0,0359*	0,1735**	0,0495	0,2460**	0,0111*	0,0515	0,0034*	0,0230*
5	4	0,0064*	0,0455	0,0365*	0,1915**	0,0506	0,2645**	0,0098*	0,0515	0,0029*	0,0235*
5	10	0,0078*	0,0560	0,0337*	0,1845**	0,0505	0,2785**	0,0107*	0,0585	0,0030*	0,0220*
5	20	0,0058*	0,0490	0,0323*	0,1970**	0,0496	0,2865**	0,0083*	0,0555	0,0022*	0,0190*
7	3	0,0050*	0,0540	0,0471	0,2750**	0,0693**	0,3830**	0,0079*	0,0565	0,0015*	0,0140*
7	4	0,0033*	0,0450	0,0280*	0,2495**	0,0485	0,3895**	0,0050*	0,0455	0,0007*	0,0105*
7	10	0,0029*	0,0475	0,0252*	0,2380**	0,04707	0,4015**	0,0043*	0,0500	0,0005*	0,0095*
7	20	0,0026*	0,0405	0,0249*	0,2535**	0,04826	0,4190**	0,0040*	0,0430	0,0003*	0,0060*
9	3	0,0022*	0,0435	0,0283*	0,3250**	0,04928	0,4900**	0,0033*	0,0435	0,0002*	0,0080*
9	4	0,0025*	0,0475	0,0264*	0,3205**	0,04953	0,5060**	0,0035*	0,0485	0,0003*	0,0080*
9	10	0,0032*	0,0520	0,0227*	0,3115**	0,04735	0,5330**	0,0038*	0,0535	0,0001*	0,0040*
9	20	0,0020*	0,0505	0,0295*	0,3230**	0,04846	0,5560**	0,0028*	0,0525	0,0000*	0,0025*

\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram abaixo do limite inferior do intervalo de confiança com 95% de probabilidade (0,0375) para a proporção empírica desta taxa.

\*\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram acima do limite superior do intervalo de confiança de 95% de probabilidade (0,0625) para a proporção empírica desta taxa.



ANEXO X – Taxas de erro tipo I por comparação ( $\alpha_1$ ) e por experimento ( $\alpha_3$ ) condicionais a um resultado significativo do teste F global para diferentes testes de comparação de médias ao nível de significância nominal de 5%, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 10%

TRAT	REP	Tukey		Duncan		DMS de Fisher		SNK		Scheffé	
		$\alpha_1$	$\alpha_3$	$\alpha_1$	$\alpha_3$	$\alpha_1$	$\alpha_3$	$\alpha_1$	$\alpha_3$	$\alpha_1$	$\alpha_3$
3	3	0,0198*	0,0455	0,0271*	0,0470	0,0271*	0,0470	0,0268*	0,0470	0,0163*	0,0395
3	4	0,0186*	0,0425	0,0286*	0,0510	0,0286*	0,0510	0,0250*	0,0500	0,0158*	0,0375
3	10	0,0183*	0,0435	0,0250*	0,0480	0,0250*	0,0480	0,0230*	0,0480	0,0163*	0,0400
3	20	0,0170*	0,0440	0,0238*	0,0490	0,0238*	0,0490	0,0218*	0,0490	0,0126*	0,0330*
5	3	0,0070*	0,0400	0,0177*	0,0510	0,0191*	0,0510	0,0104*	0,0440	0,0034*	0,0230*
5	4	0,0056*	0,0370*	0,0154*	0,0475	0,0169*	0,0475	0,0086*	0,0400	0,0029*	0,0235*
5	10	0,0067*	0,0450	0,0151*	0,0510	0,0168*	0,0510	0,0092*	0,0460	0,0030*	0,0220*
5	20	0,0047*	0,0375	0,0128*	0,0485	0,0146*	0,0485	0,0069*	0,0410	0,0022*	0,0190*
7	3	0,0044*	0,0415	0,0173*	0,0515	0,0191*	0,0515	0,0066*	0,0425	0,0015*	0,0140*
7	4	0,0027*	0,0330*	0,0105*	0,0455	0,0130*	0,0455	0,0041*	0,0330*	0,0007*	0,0105*
7	10	0,0023*	0,0350*	0,0097*	0,0430	0,0118*	0,0430	0,0034*	0,0365*	0,0005*	0,0095*
7	20	0,0021*	0,0300*	0,0086*	0,0425	0,0111*	0,0425	0,0032*	0,0310*	0,0003*	0,0060*
9	3	0,0020*	0,0360*	0,0103*	0,0460	0,0131*	0,0460	0,0030*	0,0360*	0,0002*	0,0080*
9	4	0,0022*	0,0380	0,0106*	0,0530	0,0136*	0,0530	0,0031*	0,0390	0,0003*	0,0080*
9	10	0,0023*	0,0350*	0,0079*	0,0470	0,0109*	0,0470	0,0027*	0,0355*	0,0001*	0,0040*
9	20	0,0016*	0,0370*	0,0075*	0,0490	0,0106*	0,0490	0,0023*	0,0385	0,0000*	0,0025*

\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram abaixo do limite inferior do intervalo de confiança com 95% de probabilidade (0,0375) para a proporção empírica desta taxa.

\*\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram acima do limite superior do intervalo de confiança de 95% de probabilidade (0,0625) para a proporção empírica desta taxa.

ANEXO Y – Taxas de erro tipo I por comparação ( $\alpha_2$ ) e por experimento ( $\alpha_4$ ) condicionais a um resultado significativo do teste F global para diferentes testes de comparação de médias ao nível de significância nominal de 5%, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 10%

TRAT	REP	Tukey		Duncan		DMS de Fisher		SNK		Scheffé	
		$\alpha_2$	$\alpha_4$	$\alpha_2$	$\alpha_4$	$\alpha_2$	$\alpha_4$	$\alpha_2$	$\alpha_4$	$\alpha_2$	$\alpha_4$
3	3	0,4358**	0,9680**	0,5780**	1,0000**	0,5780**	1,0000**	0,5770**	1,0000**	0,4135**	0,8404**
3	4	0,4392**	0,8333**	0,5620**	1,0000**	0,5620**	1,0000**	0,5555**	0,9803**	0,4222**	0,7352**
3	10	0,4214**	0,9062**	0,5208**	1,0000**	0,5208**	1,0000**	0,5111**	1,0000**	0,4083**	0,8333**
3	20	0,3863**	0,8979**	0,4863**	1,0000**	0,4863**	1,0000**	0,4746**	1,0000**	0,3838**	0,6734**
5	3	0,1762**	0,7843**	0,3470**	1,0000**	0,3745**	1,0000**	0,2476**	0,8627**	0,1500**	0,4509**
5	4	0,1513**	0,7789**	0,3242**	1,0000**	0,3557**	1,0000**	0,2205**	0,8421**	0,1255**	0,4947**
5	10	0,1500**	0,8823**	0,2970**	1,0000**	0,3294**	1,0000**	0,2044**	0,9019**	0,1363**	0,4313**
5	20	0,1253**	0,7731**	0,2639**	1,0000**	0,3010**	1,0000**	0,1792**	0,8453**	0,1157**	0,3917**
7	3	0,1061**	0,8058**	0,3365**	1,0000**	0,3726**	1,0000**	0,1592**	0,8252**	0,1071**	0,2718**
7	4	0,0836**	0,7252**	0,2328**	1,0000**	0,2878**	1,0000**	0,1248**	0,7252**	0,0725**	0,2307**
7	10	0,0673**	0,8139**	0,2275**	1,0000**	0,2763**	1,0000**	0,0979**	0,8488**	0,0526	0,2209**
7	20	0,0706**	0,7058**	0,2044**	1,0000**	0,2621**	1,0000**	0,1059**	0,7294**	0,0515	0,1411**
9	3	0,0555	0,7826**	0,2255**	1,0000**	0,2865**	1,0000**	0,0852**	0,7826**	0,0364	0,1739**
9	4	0,0592	0,7169**	0,2017**	1,0000**	0,2570**	1,0000**	0,0826**	0,7358**	0,0399	0,1509**
9	10	0,0662**	0,7446**	0,1681**	1,0000**	0,2319**	1,0000**	0,0777**	0,7553**	0,0312	0,0851**
9	20	0,0446	0,7551**	0,1544**	1,0000**	0,2168**	1,0000**	0,0629**	0,7857**	0,0277	0,0510

\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram abaixo do limite inferior do intervalo de confiança com 95% de probabilidade (0,0375) para a proporção empírica desta taxa.

\*\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram acima do limite superior do intervalo de confiança de 95% de probabilidade (0,0625) para a proporção empírica desta taxa.

ANEXO Z – Taxas de erro tipo I por comparação ( $\alpha_c$ ) e por experimento ( $\alpha_e$ ) incondicionais a um resultado significativo do teste F global para diferentes testes de comparação de médias ao nível de significância nominal de 5%, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 20%

TRAT	REP	Tukey		Duncan		DMS de Fisher		SNK		Scheffé	
		$\alpha_c$	$\alpha_e$	$\alpha_c$	$\alpha_e$	$\alpha_c$	$\alpha_e$	$\alpha_c$	$\alpha_e$	$\alpha_c$	$\alpha_e$
3	3	0,0205*	0,0485	0,0448	0,0950**	0,0490	0,1075**	0,0296*	0,0545	0,0161*	0,0390
3	4	0,0213*	0,0490	0,0456	0,1000**	0,0506	0,1140**	0,0278*	0,0575	0,0168*	0,0395
3	10	0,0203*	0,0490	0,0406	0,0945**	0,0480	0,1155**	0,0255*	0,0550	0,0163*	0,0400
3	20	0,0166*	0,0440	0,0380	0,0870**	0,0448	0,1065**	0,0245*	0,0555	0,0130*	0,0345*
5	3	0,0075*	0,0455	0,0353*	0,1675**	0,0493	0,2455**	0,0113*	0,0520	0,0034*	0,0235*
5	4	0,0064*	0,0450	0,0366*	0,1945**	0,0501	0,2660**	0,0097*	0,0505	0,0028*	0,0225*
5	10	0,0077*	0,0560	0,0337*	0,1810**	0,0503	0,2765**	0,0110*	0,0590	0,0031*	0,0235*
5	20	0,0058*	0,0480	0,0325*	0,1980**	0,0501	0,2855**	0,0081*	0,0525	0,0020*	0,0185*
7	3	0,0050*	0,0530	0,0352*	0,2750**	0,0522	0,3855**	0,0077*	0,0560	0,0015*	0,0155*
7	4	0,0036*	0,0455	0,0287*	0,2575**	0,0486	0,3915**	0,0049*	0,0465	0,0008*	0,0115*
7	10	0,0029*	0,0470	0,0248*	0,2400**	0,0452	0,4035**	0,0042*	0,0485	0,0005*	0,0095*
7	20	0,0025*	0,0395	0,0250*	0,2515**	0,0483	0,4185**	0,0037*	0,0415	0,0003*	0,0060*
9	3	0,0025*	0,0425	0,0280*	0,3240**	0,0491	0,4890**	0,4151**	0,0435	0,0002*	0,0080*
9	4	0,0025*	0,0485	0,0264*	0,3205**	0,0492	0,5130**	0,0035*	0,0510	0,0003*	0,0080*
9	10	0,0020*	0,0495	0,0225*	0,3110**	0,0474	0,5320**	0,0026*	0,0495	0,0001*	0,0040*
9	20	0,0021*	0,0535	0,0229*	0,3215**	0,0485	0,5575**	0,0029*	0,0550	0,0000*	0,0025*

\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram abaixo do limite inferior do intervalo de confiança com 95% de probabilidade (0,0375) para a proporção empírica desta taxa.

\*\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram acima do limite superior do intervalo de confiança de 95% de probabilidade (0,0625) para a proporção empírica desta taxa.

ANEXO A1 – Taxas de erro tipo I por comparação ( $\alpha_1$ ) e por experimento ( $\alpha_3$ ) condicionais a um resultado significativo do teste F global para diferentes testes de comparação de médias ao nível de significância nominal de 5%, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 20%

TRAT	REP	Tukey		Duncan		DMS de Fisher		SNK		Scheffé	
		$\alpha_1$	$\alpha_3$	$\alpha_1$	$\alpha_3$	$\alpha_1$	$\alpha_3$	$\alpha_1$	$\alpha_3$	$\alpha_1$	$\alpha_3$
3	3	0,0193*	0,0450	0,0278*	0,0475	0,0278*	0,0475	0,0275*	0,0475	0,0161*	0,0390
3	4	0,0201*	0,0455	0,0283*	0,0505	0,0283*	0,0505	0,0256*	0,0505	0,0168*	0,0395
3	10	0,0183*	0,0430	0,0240*	0,0465	0,0240*	0,0465	0,0228*	0,0460	0,0163*	0,0400
3	20	0,0160*	0,0420	0,0235*	0,0480	0,0235*	0,0480	0,0211*	0,0480	0,0130*	0,0345*
5	3	0,0069*	0,0395	0,0177*	0,0510	0,0189*	0,0510	0,0105*	0,0440	0,0034*	0,0235*
5	4	0,0056*	0,0370*	0,0155*	0,0480	0,0170*	0,0480	0,0087*	0,0400	0,0028*	0,0225*
5	10	0,0066*	0,0450	0,0154*	0,0515	0,0167*	0,0515	0,0094*	0,0465	0,0031*	0,0235*
5	20	0,0048*	0,0380	0,0126*	0,0490	0,0149*	0,0490	0,0068*	0,0400	0,0020*	0,0185*
7	3	0,0043*	0,0400	0,0139*	0,0485	0,0154*	0,0485	0,0066*	0,0415	0,0015*	0,0155*
7	4	0,0030*	0,0340*	0,0106*	0,0445	0,0128*	0,0445	0,0041*	0,0345*	0,0008*	0,0115*
7	10	0,0024*	0,0360*	0,0097*	0,0450	0,0118*	0,0450	0,0035*	0,0370*	0,0005*	0,0095*
7	20	0,0020*	0,0285*	0,0084*	0,0410	0,0108*	0,0410	0,0029*	0,0290*	0,0003*	0,0060*
9	3	0,0022*	0,0340*	0,0098*	0,0445	0,0125*	0,0445	0,0177*	0,0345*	0,0002*	0,0080*
9	4	0,0022*	0,0380	0,0104*	0,0520	0,0133*	0,0520	0,0031*	0,0405	0,0003*	0,0080*
9	10	0,0015*	0,0335*	0,0077*	0,0470	0,0111*	0,0470	0,0019*	0,0335*	0,0001*	0,0040*
9	20	0,0016*	0,0365*	0,0077*	0,0490	0,0105*	0,0490	0,0022*	0,0375	0,0000*	0,0025*

\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram abaixo do limite inferior do intervalo de confiança com 95% de probabilidade (0,0375) para a proporção empírica desta taxa.

\*\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram acima do limite superior do intervalo de confiança de 95% de probabilidade (0,0625) para a proporção empírica desta taxa.

ANEXO B1 – Taxas de erro tipo I por comparação ( $\alpha_2$ ) e por experimento ( $\alpha_4$ ) condicionais a um resultado significativo do teste F global para diferentes testes de comparação de médias ao nível de significância nominal de 5%, em função do número de tratamentos (TRAT) e de repetições (REP), ao coeficiente de variação de 20%

TRAT	REP	Tukey		Duncan		DMS de Fisher		SNK		Scheffé	
		$\alpha_2$	$\alpha_4$	$\alpha_2$	$\alpha_4$	$\alpha_2$	$\alpha_4$	$\alpha_2$	$\alpha_4$	$\alpha_2$	$\alpha_4$
3	3	0,4296**	0,9473**	0,5859**	1,0000**	0,5859**	1,0000**	0,5851**	1,0000**	0,4145**	0,8210**
3	4	0,4432**	0,9009**	0,5610**	1,0000**	0,5610**	1,0000**	0,5519**	1,0000**	0,4261**	0,7821**
3	10	0,4263**	0,9247**	0,5161**	1,0000**	0,5161**	1,0000**	0,5131**	0,9892**	0,4083**	0,8602**
3	20	0,3809**	0,8750**	0,4895**	1,0000**	0,4895**	1,0000**	0,4756**	1,0000**	0,3768**	0,7187**
5	3	0,1759**	0,7745**	0,3470**	1,0000**	0,3705**	1,0000**	0,2542**	0,8627**	0,1468**	0,4607**
5	4	0,1513**	0,7708**	0,3239**	1,0000**	0,3552**	1,0000**	0,2272**	0,8333**	0,1244**	0,4687**
5	10	0,1477**	0,8737**	0,3000**	1,0000**	0,3252**	1,0000**	0,2065**	0,9029**	0,1319**	0,4563**
5	20	0,1263**	0,7755**	0,2581**	1,0000**	0,3051**	1,0000**	0,1779**	0,8163**	0,1108**	0,3775**
7	3	0,1095**	0,8247**	0,2866**	1,0000**	0,3186**	1,0000**	0,1634**	0,8556**	0,0998**	0,3195**
7	4	0,0903**	0,7640**	0,2402**	1,0000**	0,2894**	1,0000**	0,1218**	0,7752**	0,0724**	0,2584**
7	10	0,0681**	0,8000**	0,2169**	1,0000**	0,2634**	1,0000**	0,0971**	0,8222**	0,0526	0,2111**
7	20	0,0718**	0,6951**	0,2049**	1,0000**	0,2642**	1,0000**	0,1035**	0,7073**	0,0515	0,1463**
9	3	0,0665**	0,7640**	0,2222**	1,0000**	0,2815**	1,0000**	0,3982**	0,7752**	0,0364	0,1797**
9	4	0,0584	0,7307**	0,2003**	1,0000**	0,2572**	1,0000**	0,0836**	0,7788**	0,0399	0,1538**
9	10	0,0472	0,7127**	0,1643**	1,0000**	0,2367**	1,0000**	0,0588	0,7127**	0,0312	0,0851**
9	20	0,0449	0,7448**	0,1584**	1,0000**	0,2157**	1,0000**	0,0624	0,7653**	0,0277	0,0510

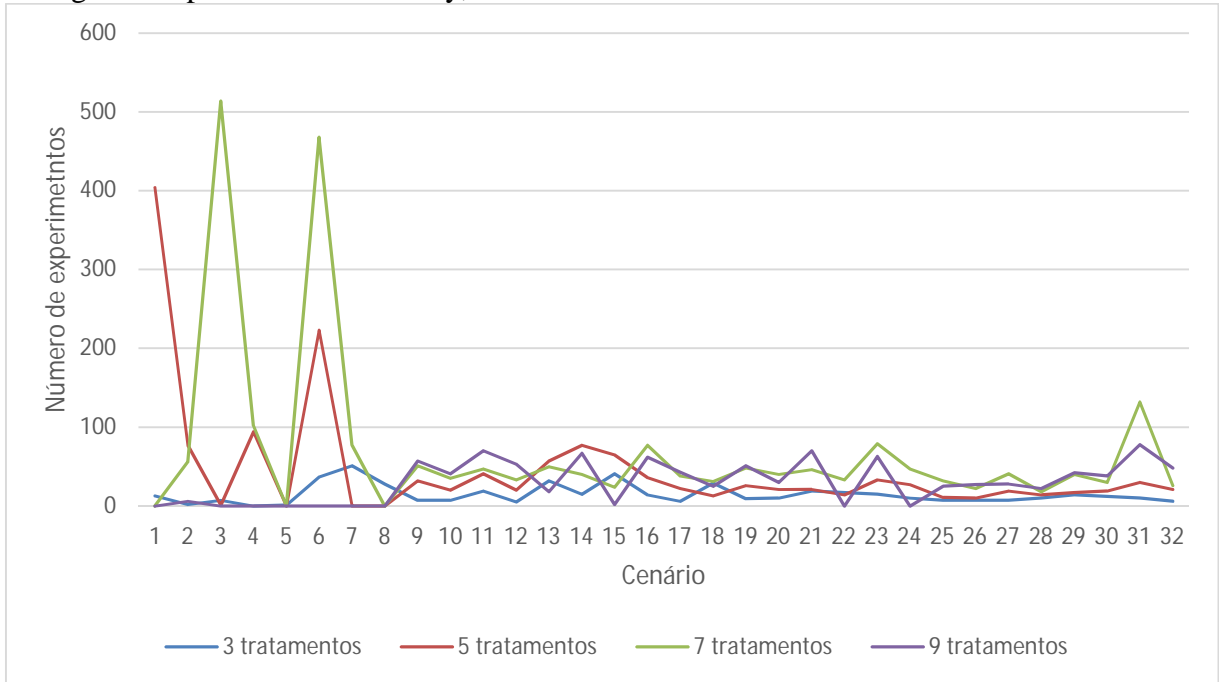
\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram abaixo do limite inferior do intervalo de confiança com 95% de probabilidade (0,0375) para a proporção empírica desta taxa.

\*\*Taxas de erro tipo I (por comparação ou por experimento) que ficaram acima do limite superior do intervalo de confiança de 95% de probabilidade (0,0625) para a proporção empírica desta taxa.

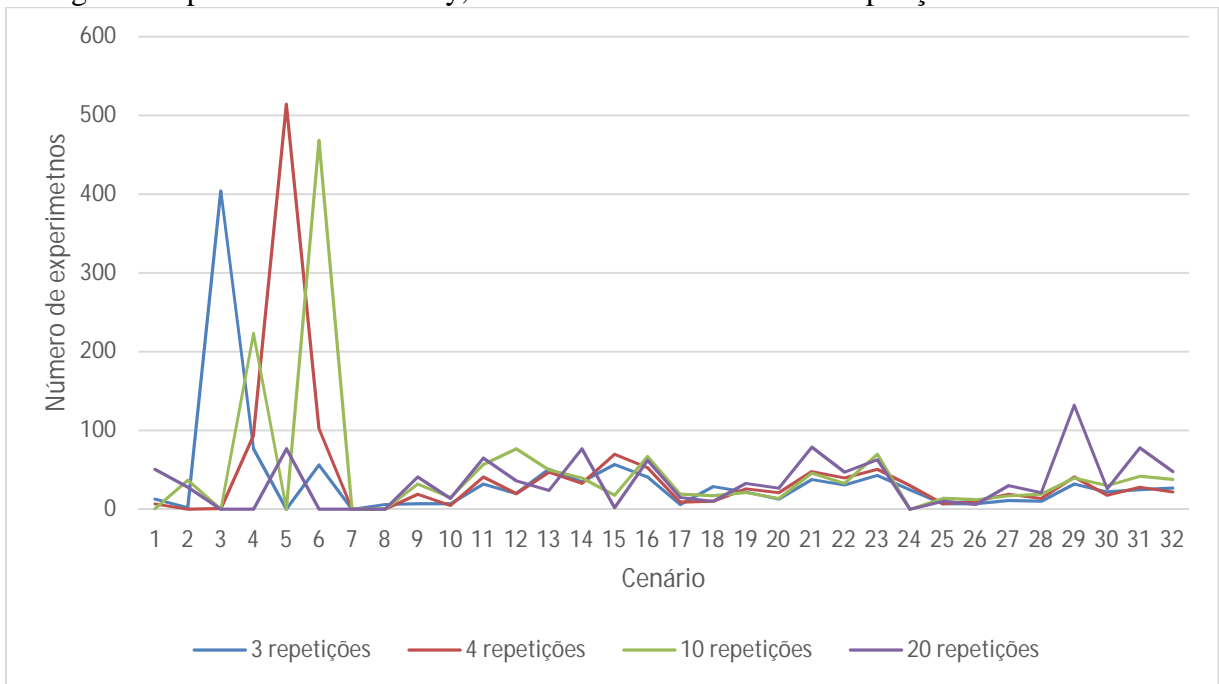
ANEXO C1 – Poder incondicional ( $\pi$ ) e condicional ( $\pi_1$  e  $\pi_2$ ) a um resultado significativo do teste F global, em percentual, para diferentes testes de comparação de médias ao nível de significância nominal de 5%, em função do número de tratamentos (TRAT) e da diferença real entre médias adjacentes (DRM), com 20 repetições para tratamentos e coeficiente de variação de 5%

TRAT	DRM	Tukey			Duncan			DMS de Fisher			SNK			Scheffé		
		$\pi$	$\pi_1$	$\pi_2$	$\pi$	$\pi_1$	$\pi_2$	$\pi$	$\pi_1$	$\pi_2$	$\pi$	$\pi_1$	$\pi_2$	$\pi$	$\pi_1$	$\pi_2$
3	$1\sigma_\mu$	0,0855	0,0812	0,4206	0,1475	0,1112	0,5306	0,1633	0,1112	0,5306	0,1117	0,1053	0,5280	0,0727	0,0727	0,4117
3	$2\sigma_\mu$	0,3408	0,3367	0,4855	0,4545	0,4213	0,5880	0,4643	0,4213	0,5880	0,4192	0,4127	0,5874	0,3107	0,3107	0,4682
3	$4\sigma_\mu$	0,7798	0,7798	0,7802	0,8673	0,8673	0,8673	0,8673	0,8673	0,8673	0,8673	0,8673	0,8673	0,7537	0,7537	0,7540
3	$8\sigma_\mu$	0,9990	0,9990	0,9990	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9987	0,9987	0,9987
5	$1\sigma_\mu$	0,1424	0,1399	0,2164	0,2872	0,2633	0,3719	0,3222	0,2839	0,4009	0,1966	0,1938	0,2960	0,0894	0,0894	0,1802
5	$2\sigma_\mu$	0,4793	0,4793	0,4800	0,6446	0,6446	0,6452	0,6542	0,6541	0,6547	0,6107	0,6107	0,6113	0,4073	0,4073	0,4085
5	$4\sigma_\mu$	0,8080	0,8080	0,8080	0,9219	0,9219	0,9219	0,9220	0,9220	0,9220	0,9218	0,9218	0,9218	0,7511	0,7511	0,7511
5	$8\sigma_\mu$	0,9992	0,9992	0,9992	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9974	0,9974	0,9974
7	$1\sigma_\mu$	0,3290	0,3290	0,3292	0,5174	0,5174	0,5176	0,5416	0,5415	0,5418	0,4243	0,4243	0,4245	0,3558	0,3557	0,3558
7	$2\sigma_\mu$	0,6645	0,6645	0,6645	0,8108	0,8108	0,8108	0,8155	0,8155	0,8155	0,7931	0,7931	0,7931	0,5749	0,5749	0,5749
7	$4\sigma_\mu$	0,8920	0,8920	0,8920	0,9637	0,9637	0,9637	0,9637	0,9637	0,9637	0,9636	0,9636	0,9636	0,8464	0,8464	0,8464
7	$8\sigma_\mu$	0,9994	0,9994	0,9994	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9962	0,9962	0,9962
9	$1\sigma_\mu$	0,5428	0,5428	0,5428	0,7051	0,7051	0,7051	0,7223	0,7223	0,7223	0,6531	0,6531	0,6531	0,4292	0,4292	0,4292
9	$2\sigma_\mu$	0,7924	0,7924	0,7924	0,8910	0,8910	0,8910	0,8940	0,8940	0,8940	0,8806	0,8806	0,8806	0,7178	0,7178	0,7178
9	$4\sigma_\mu$	0,9302	0,9302	0,9302	0,9784	0,9784	0,9784	0,9784	0,9784	0,9784	0,9783	0,9783	0,9783	0,8961	0,8961	0,8961
9	$8\sigma_\mu$	0,9992	0,9992	0,9992	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9947	0,9947	0,9947

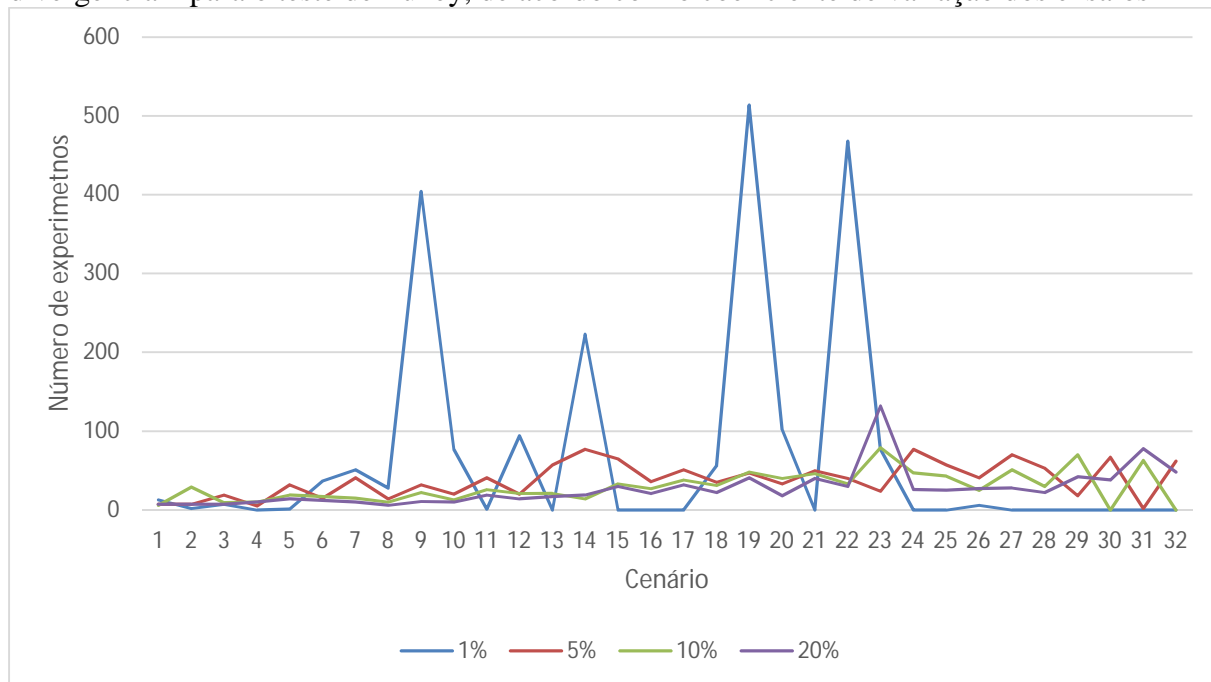
ANEXO D1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 2 para o teste de Tukey, de acordo com o número de tratamentos dos ensaios



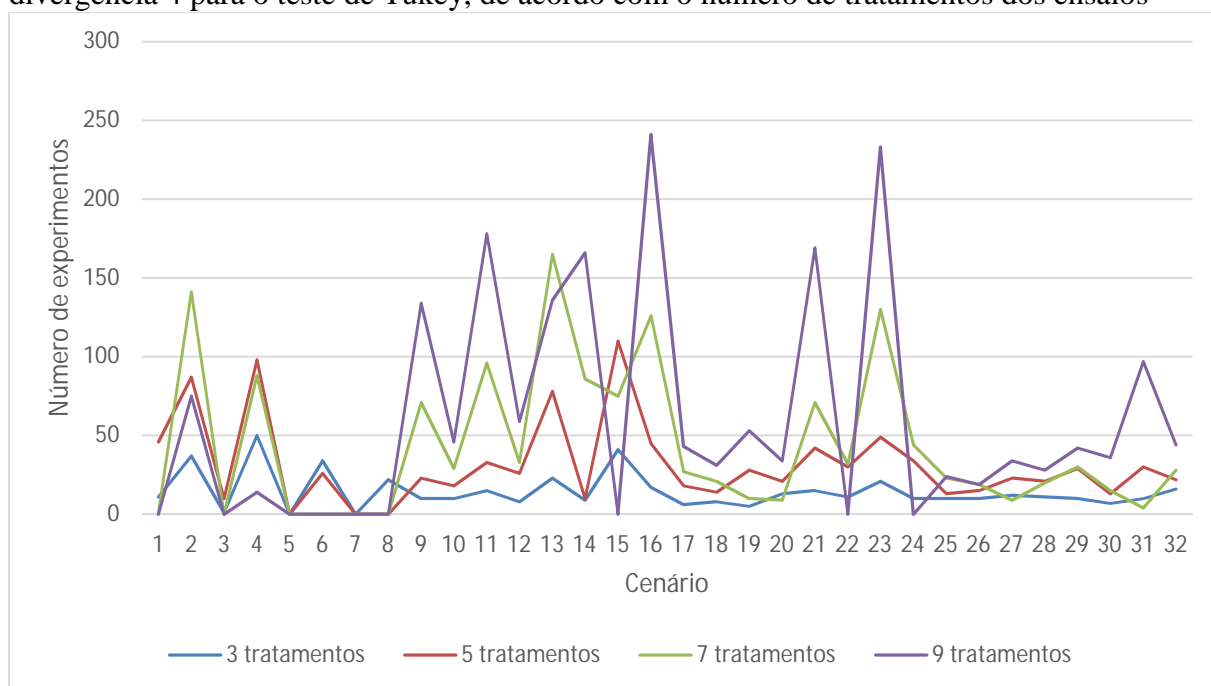
ANEXO E1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 2 para o teste de Tukey, de acordo com o número de repetições dos tratamentos



ANEXO F1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 2 para o teste de Tukey, de acordo com o coeficiente de variação dos ensaios

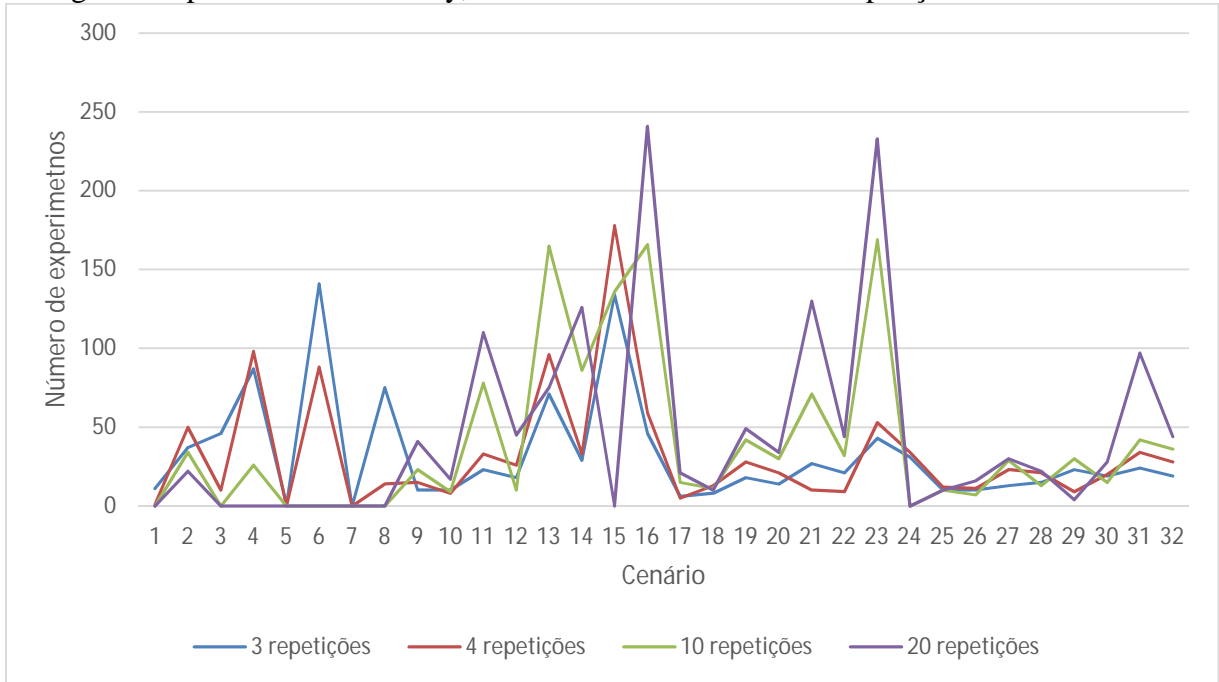


ANEXO G1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 4 para o teste de Tukey, de acordo com o número de tratamentos dos ensaios

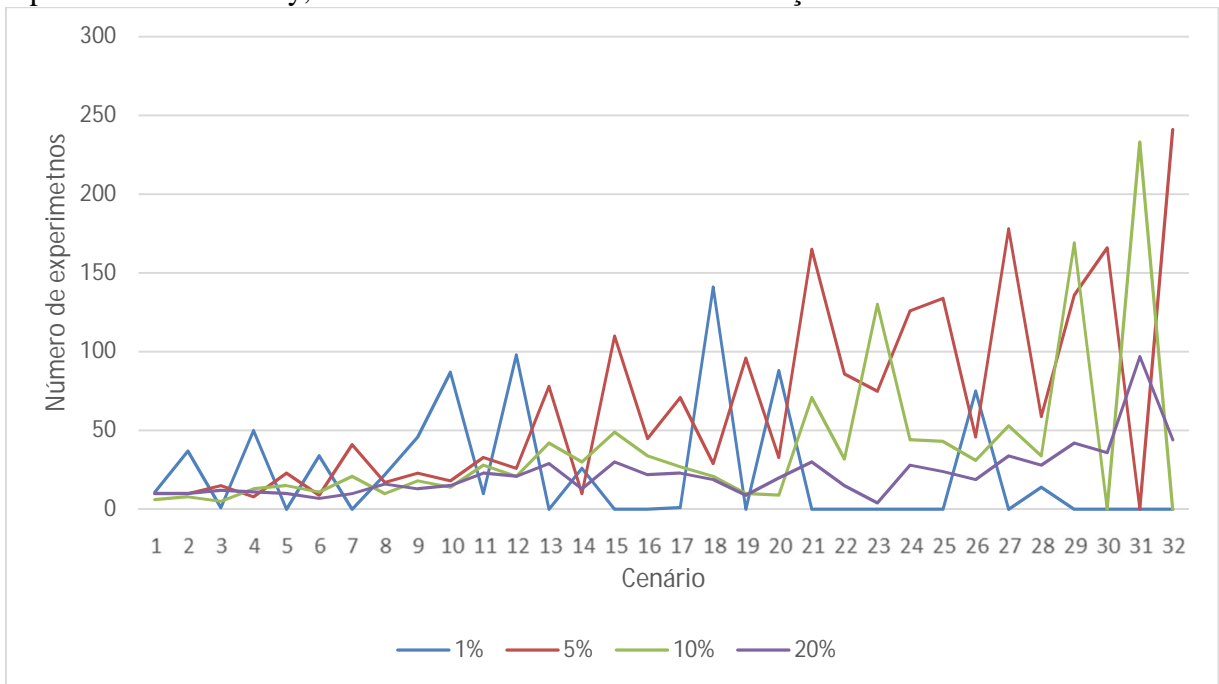




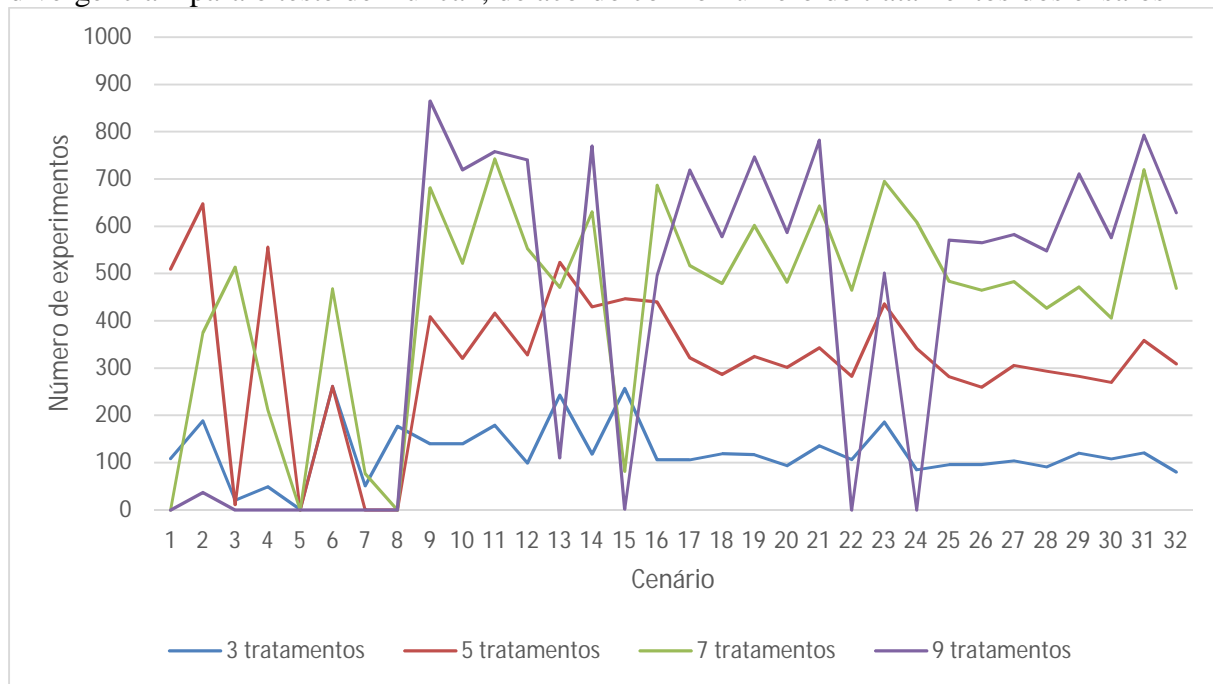
ANEXO H1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 4 para o teste de Tukey, de acordo com o número de repetições dos tratamentos



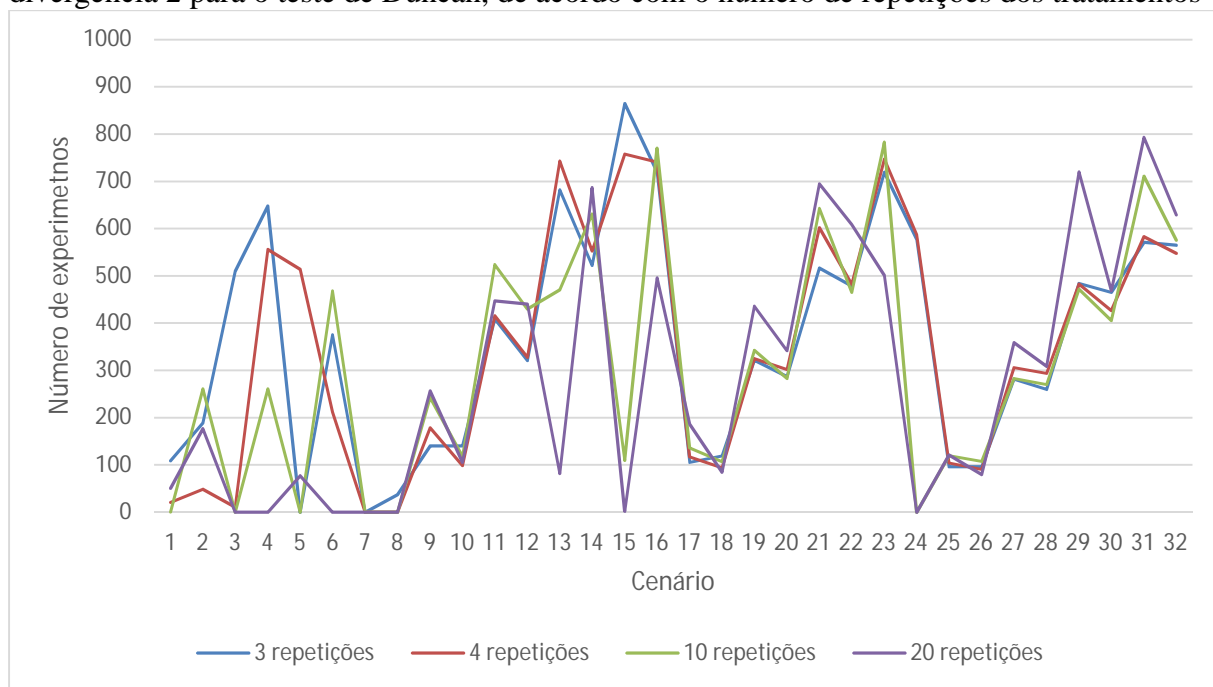
ANEXO I1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 4 para o teste de Tukey, de acordo com o coeficiente de variação dos ensaios



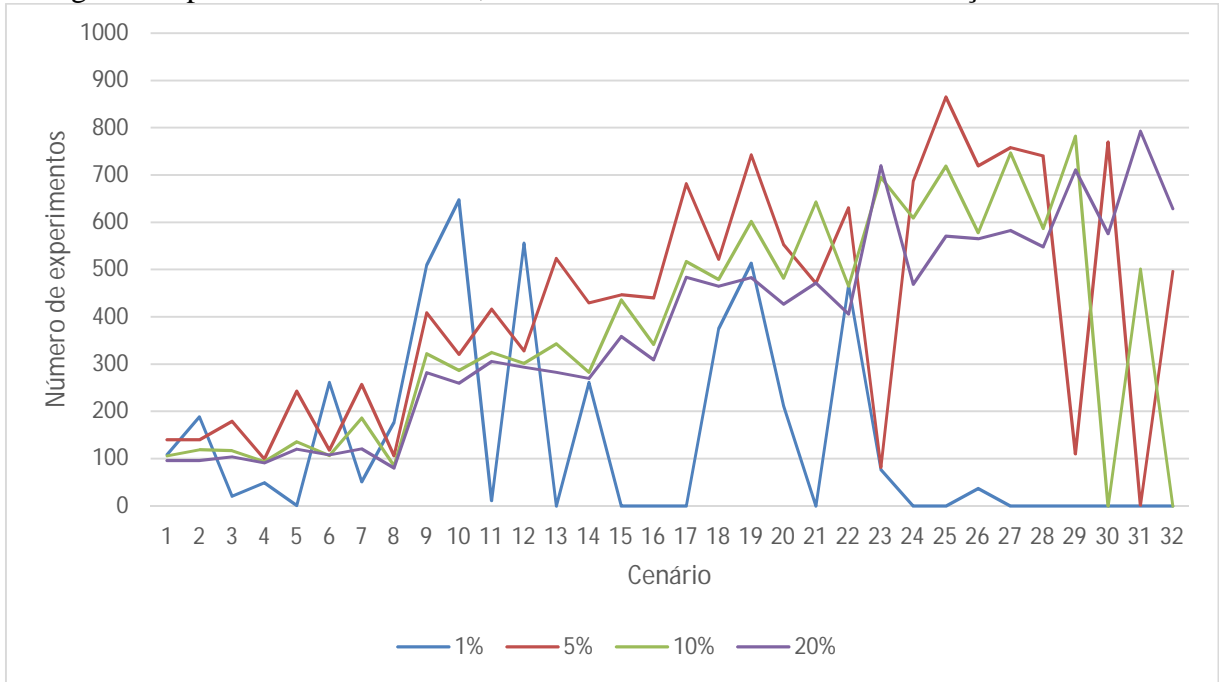
ANEXO J1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 2 para o teste de Duncan, de acordo com o número de tratamentos dos ensaios



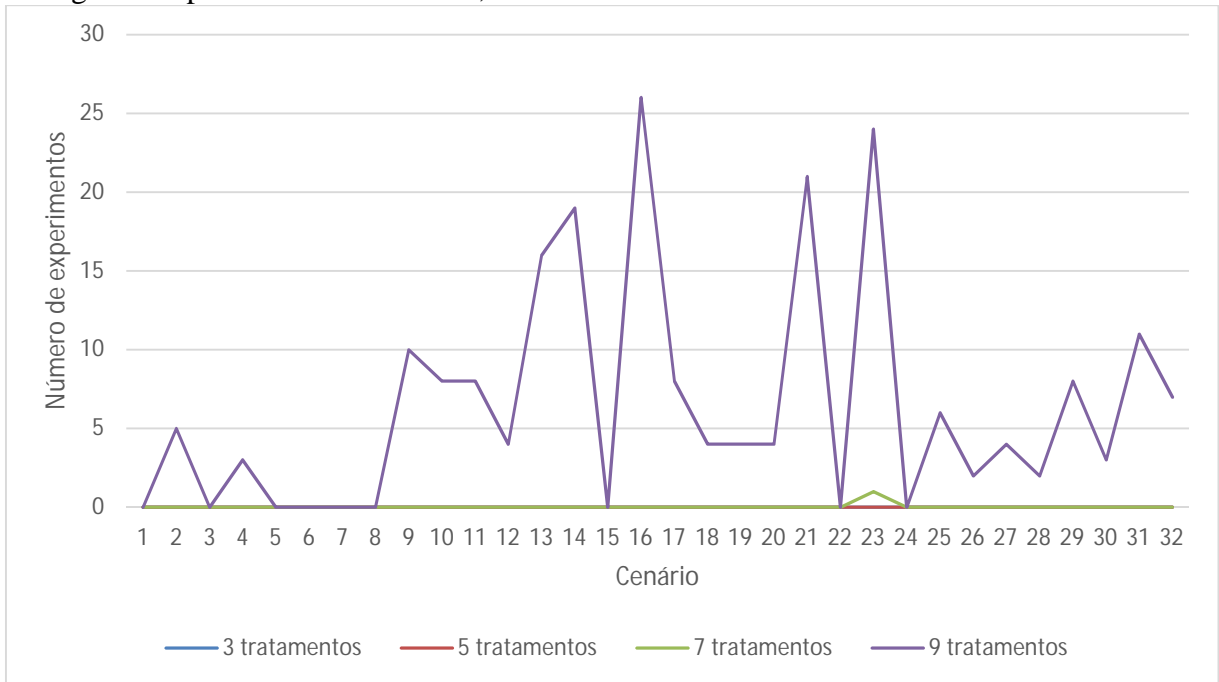
ANEXO K1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 2 para o teste de Duncan, de acordo com o número de repetições dos tratamentos



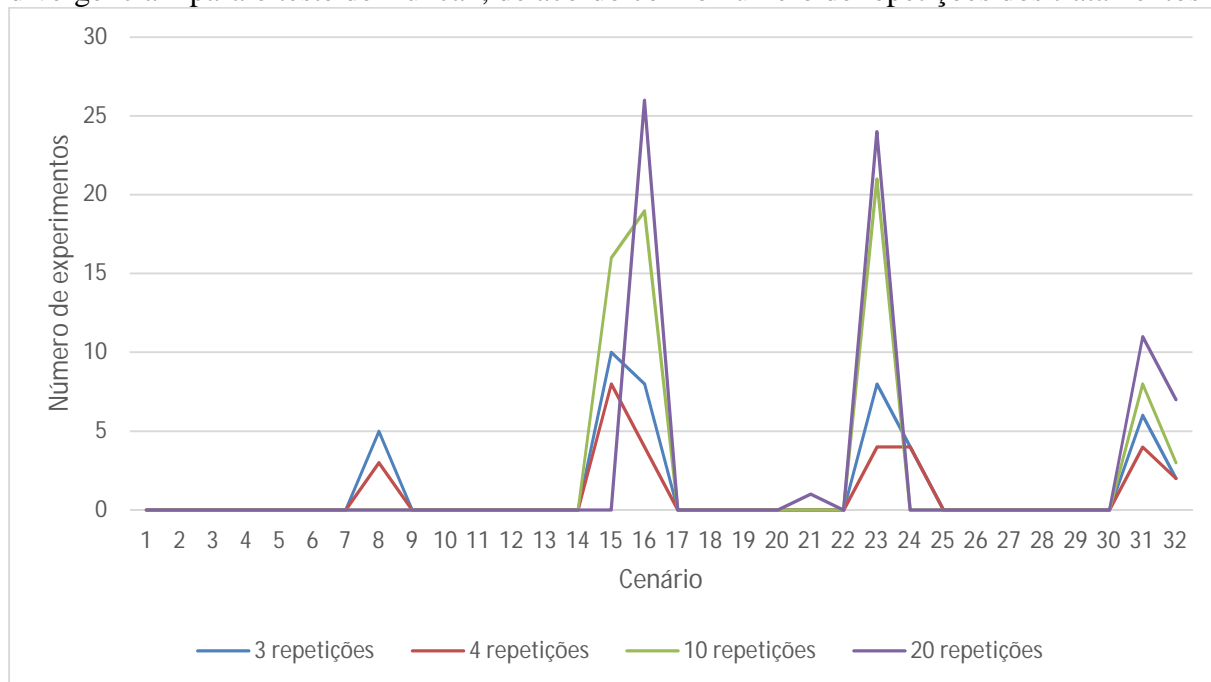
ANEXO L1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 2 para o teste de Duncan, de acordo com o coeficiente de variação dos ensaios



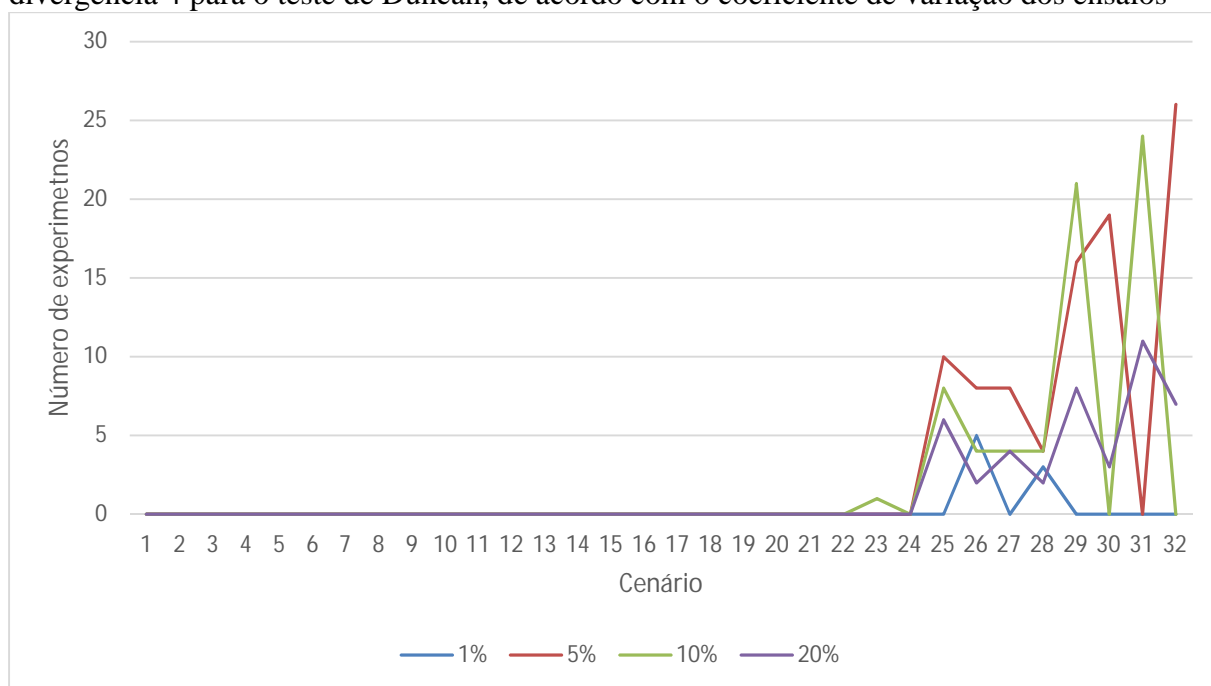
ANEXO M1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 4 para o teste de Duncan, de acordo com o número de tratamentos dos ensaios



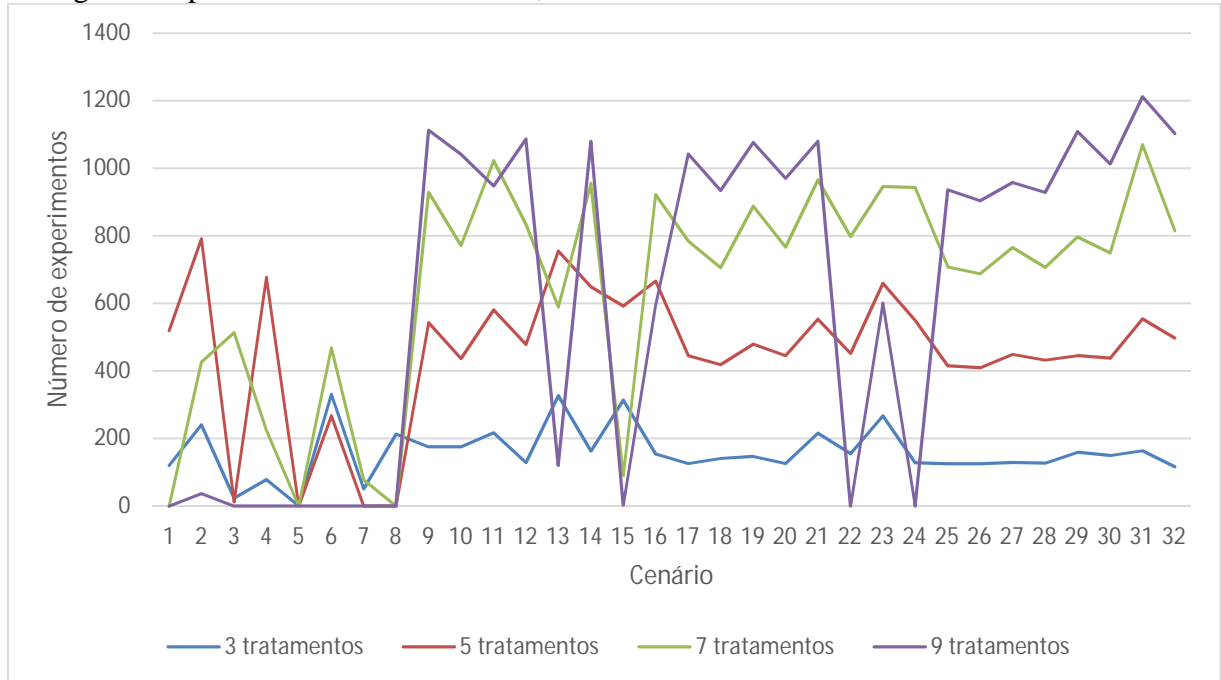
ANEXO N1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 4 para o teste de Duncan, de acordo com o número de repetições dos tratamentos



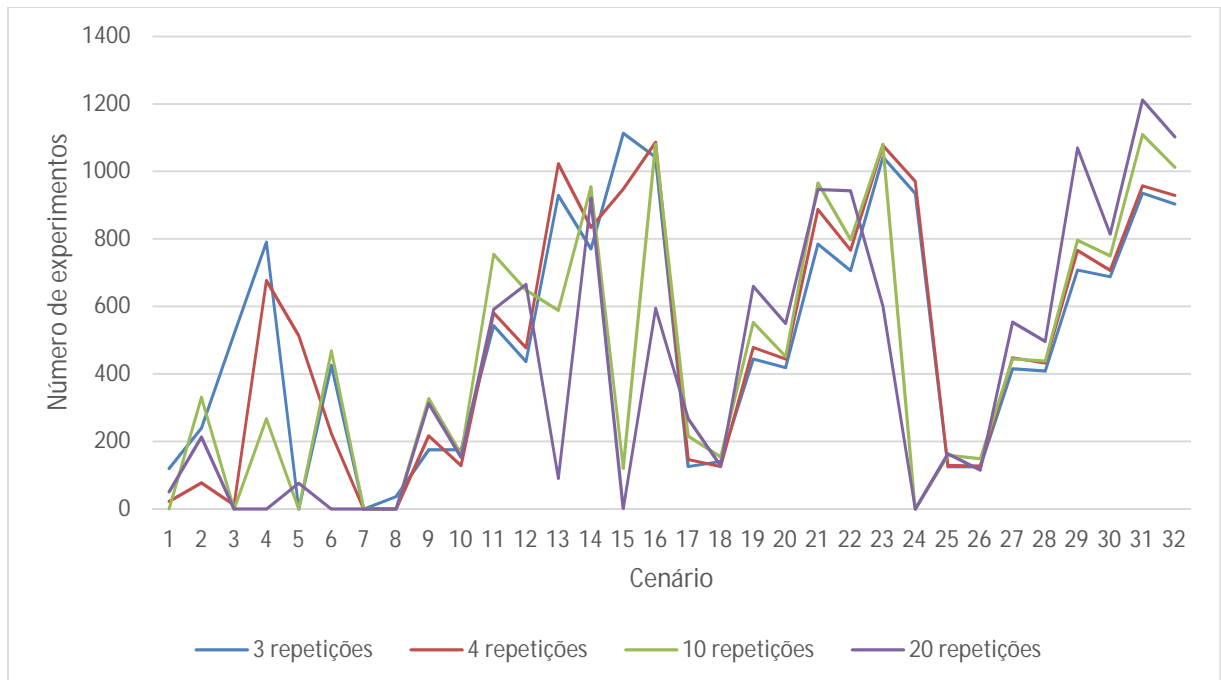
ANEXO O1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 4 para o teste de Duncan, de acordo com o coeficiente de variação dos ensaios



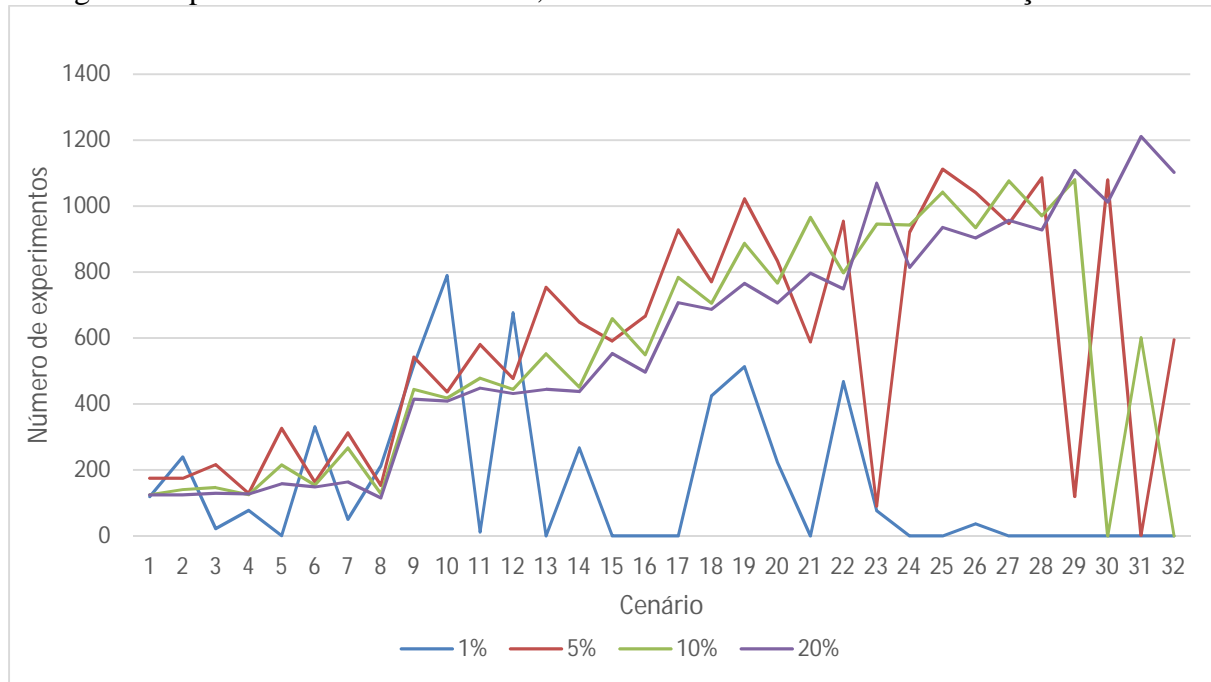
ANEXO P1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 2 para o teste DMS de Fisher, de acordo com o número de tratamentos dos ensaios



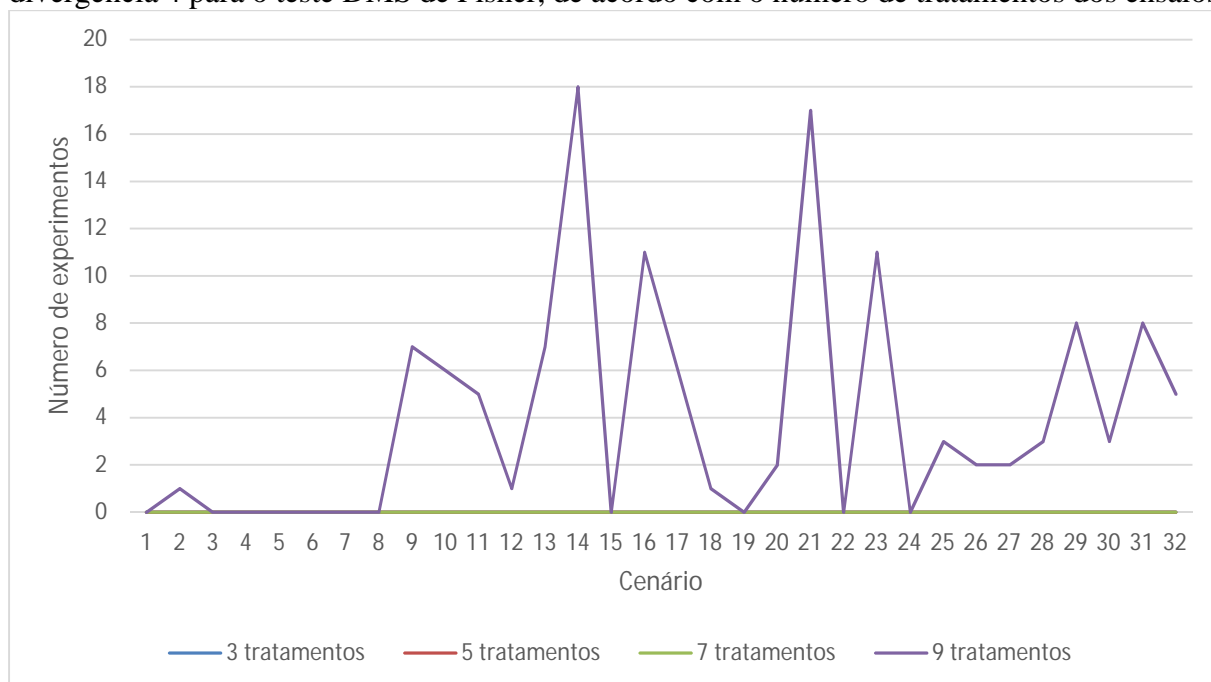
ANEXO Q1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 2 para o teste DMS de Fisher, de acordo com o número de repetições dos tratamentos



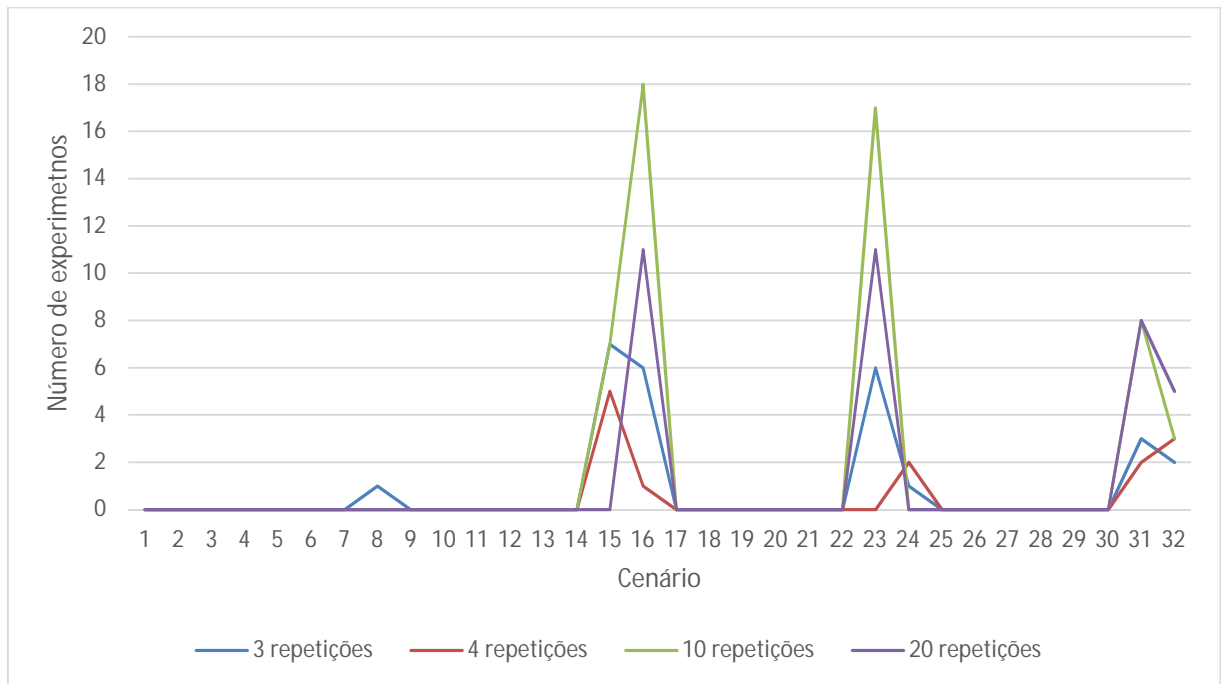
ANEXO R1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 2 para o teste DMS de Fisher, de acordo com o coeficiente de variação dos ensaios



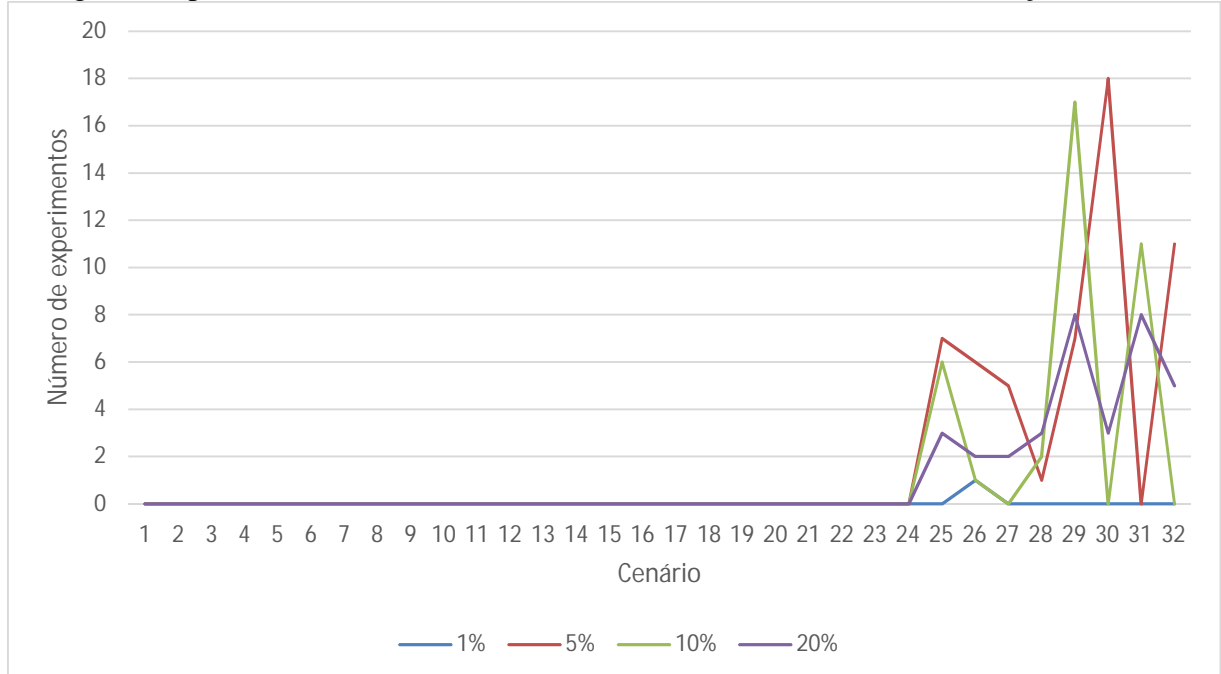
ANEXO S1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 4 para o teste DMS de Fisher, de acordo com o número de tratamentos dos ensaios



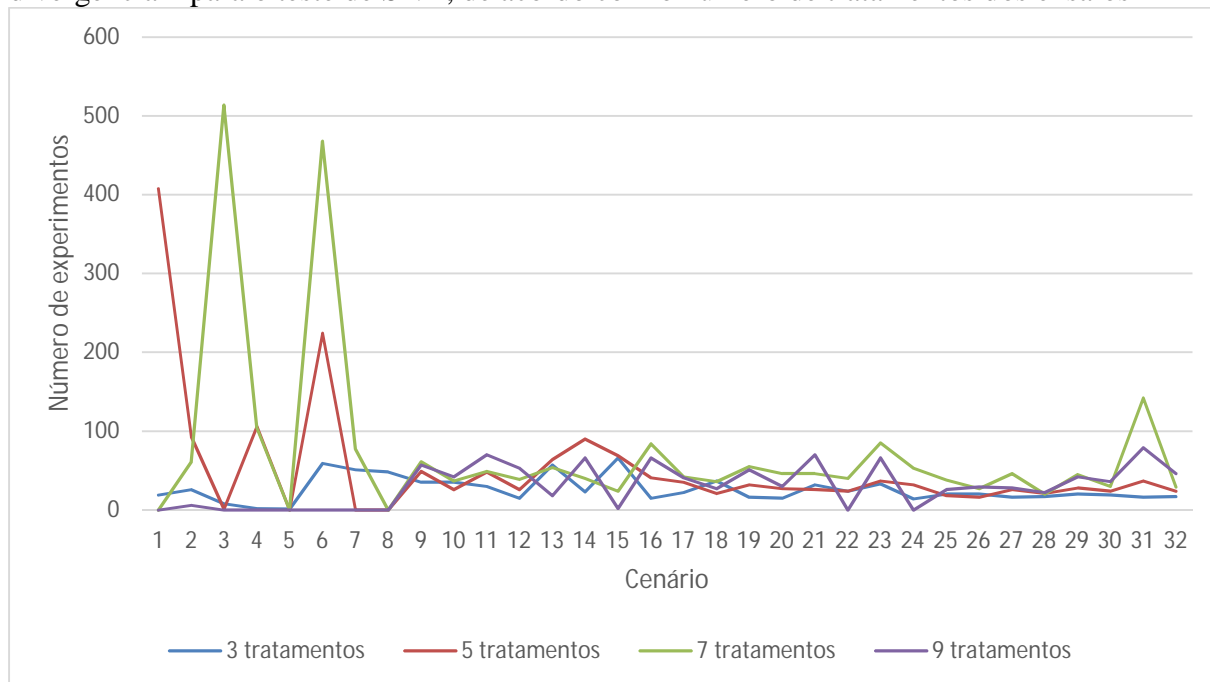
ANEXO T1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 4 para o teste DMS de Fisher, de acordo com o número de repetições dos tratamentos



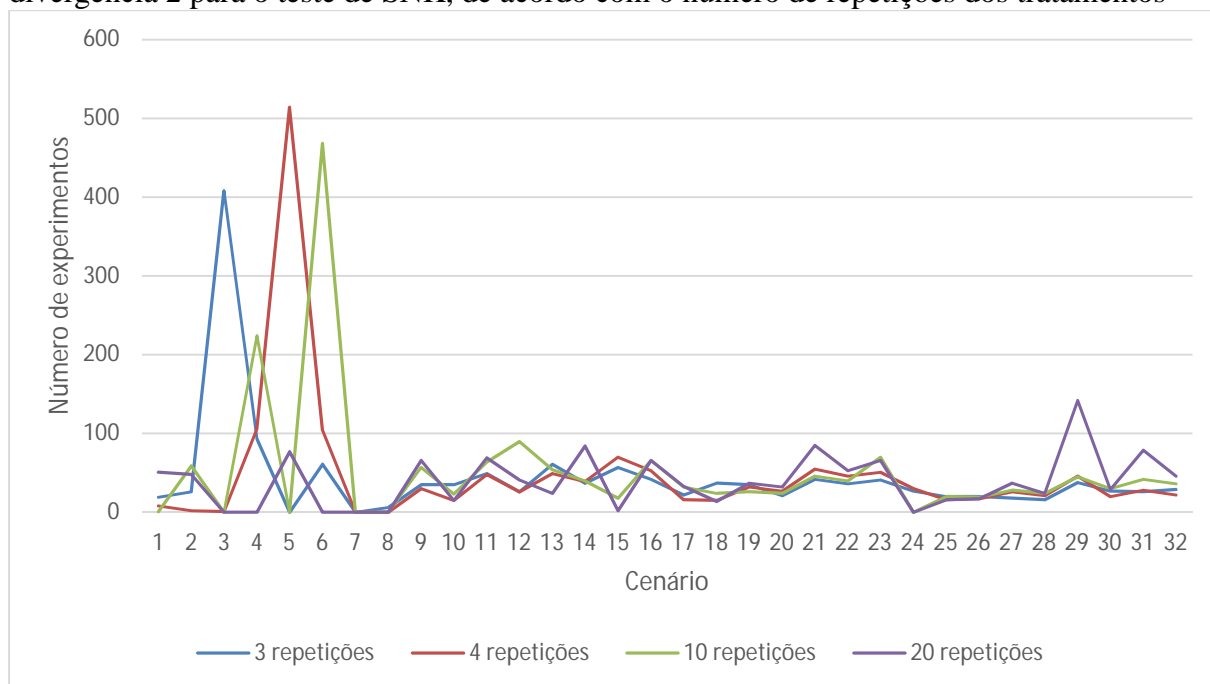
ANEXO U1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 4 para o teste DMS de Fisher, de acordo com o coeficiente de variação dos ensaio.



ANEXO V1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 2 para o teste de SNK, de acordo com o número de tratamentos dos ensaios

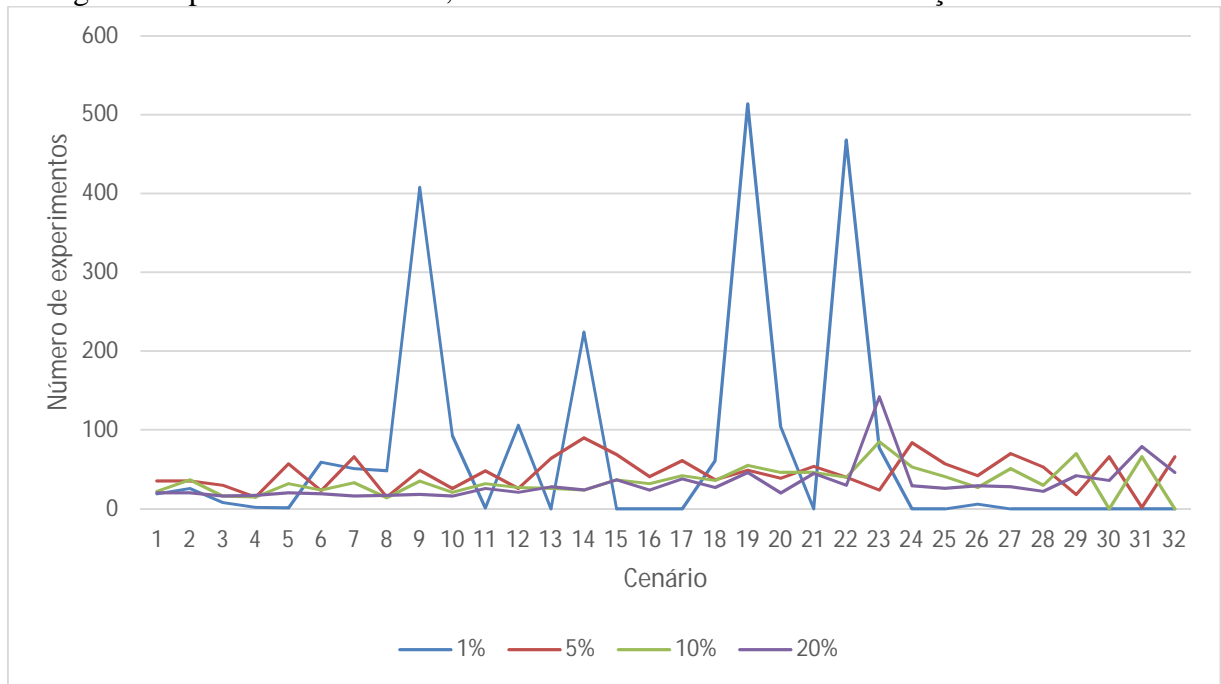


ANEXO W1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 2 para o teste de SNK, de acordo com o número de repetições dos tratamentos

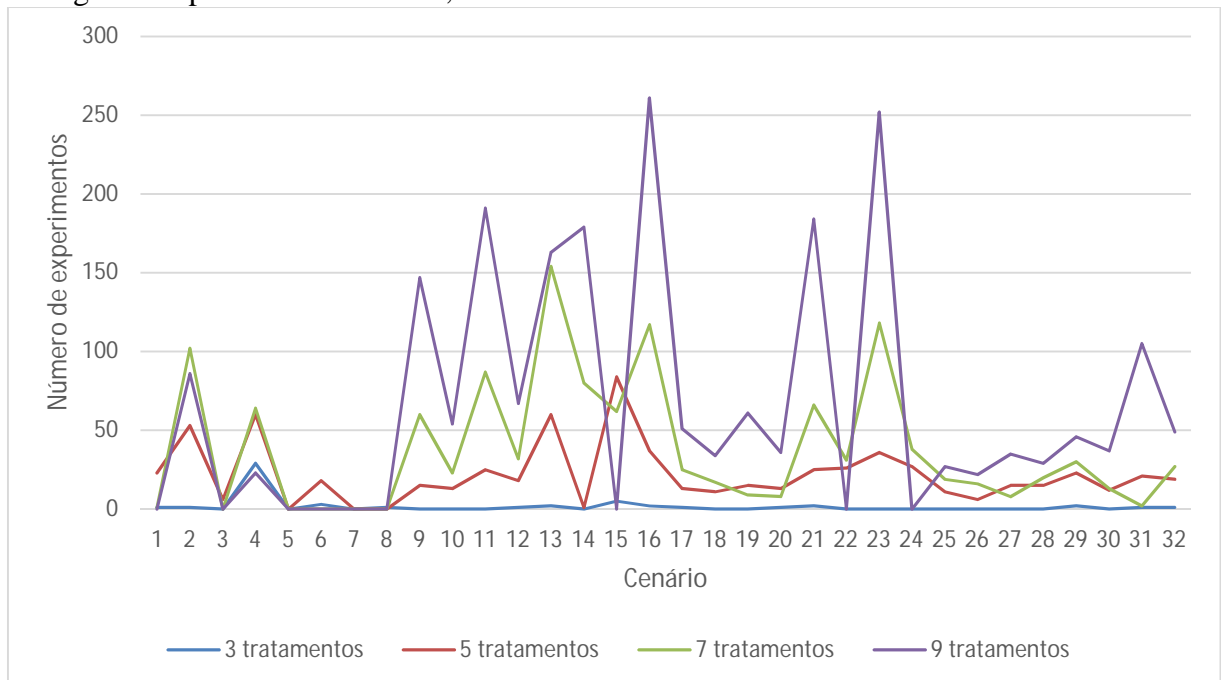




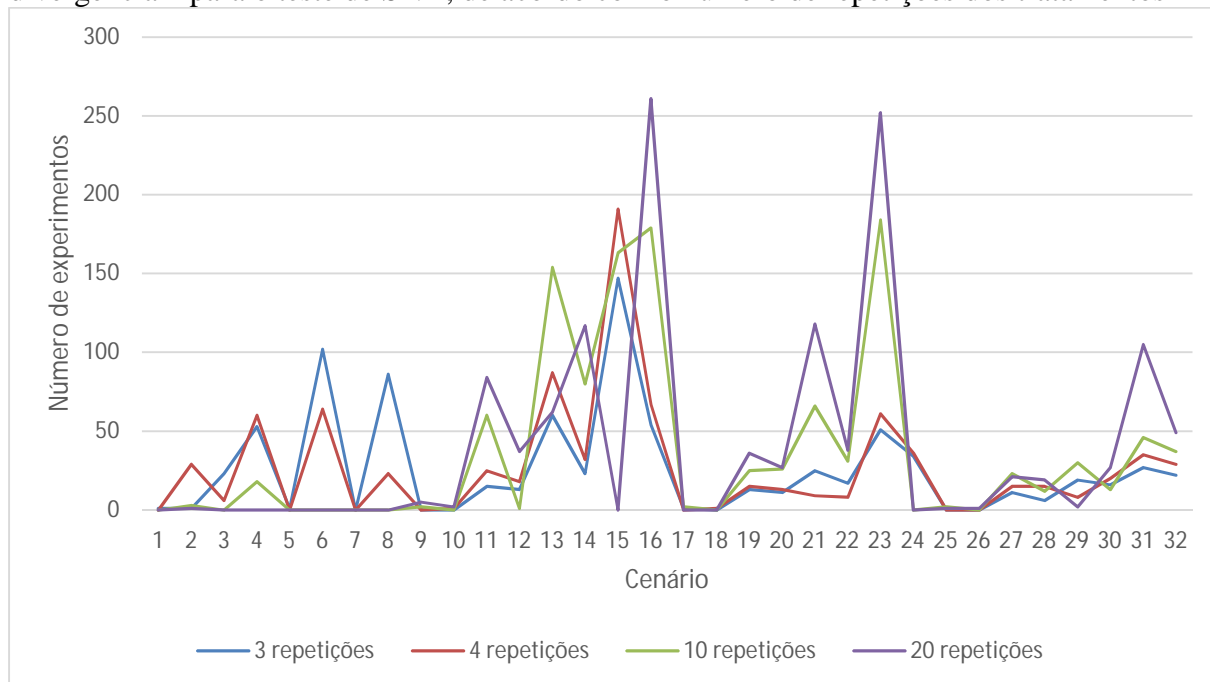
ANEXO X1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 2 para o teste de SNK, de acordo com o coeficiente de variação dos ensaios



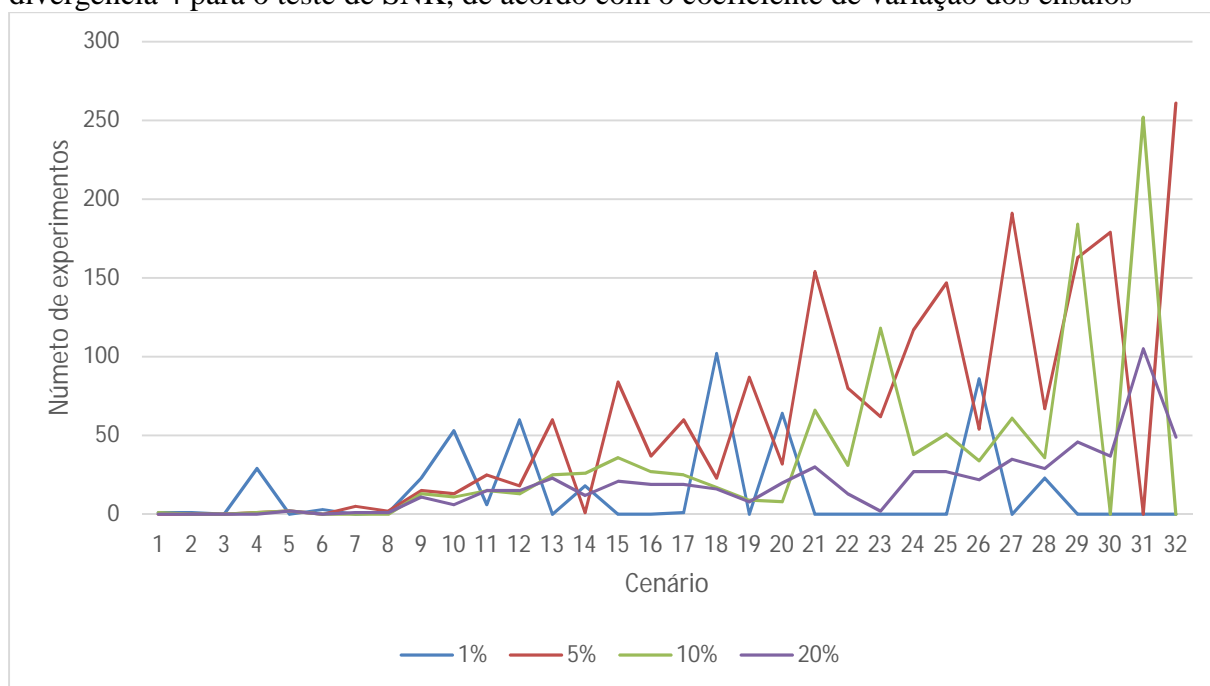
ANEXO Y1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 4 para o teste de SNK, de acordo com o número de tratamentos dos ensaios



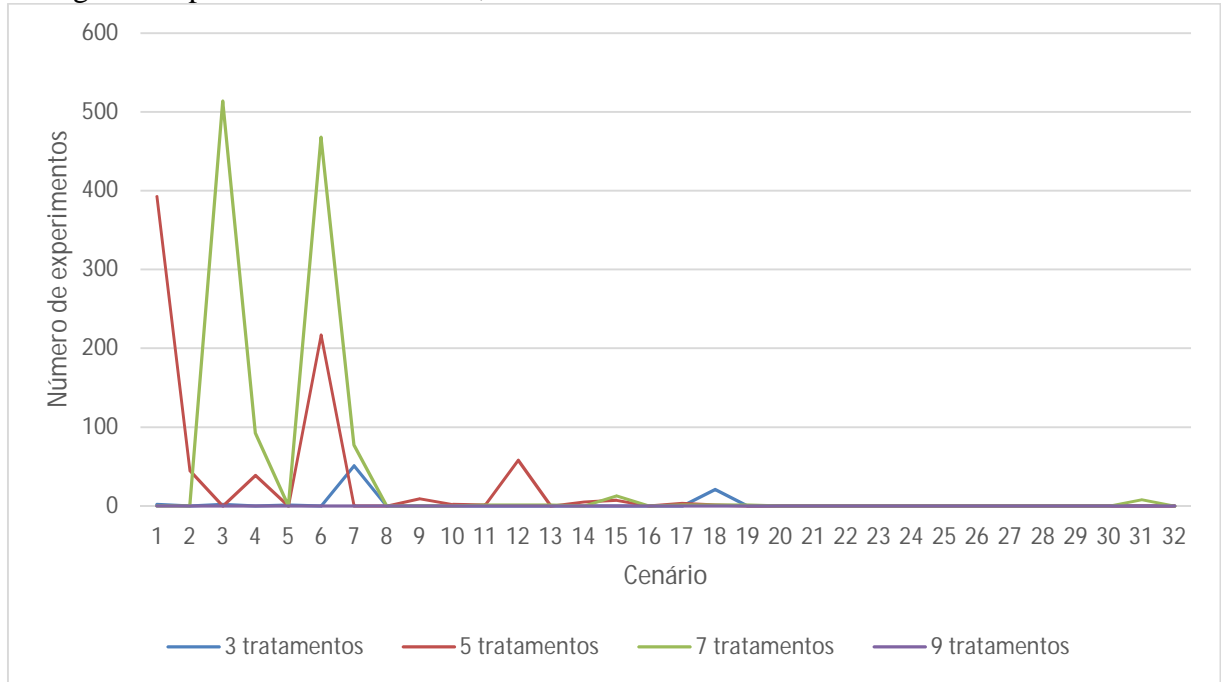
ANEXO Z1 – Número, por cenário, de experimentos que apresentaram a situação de divergência 4 para o teste de SNK, de acordo com o número de repetições dos tratamentos



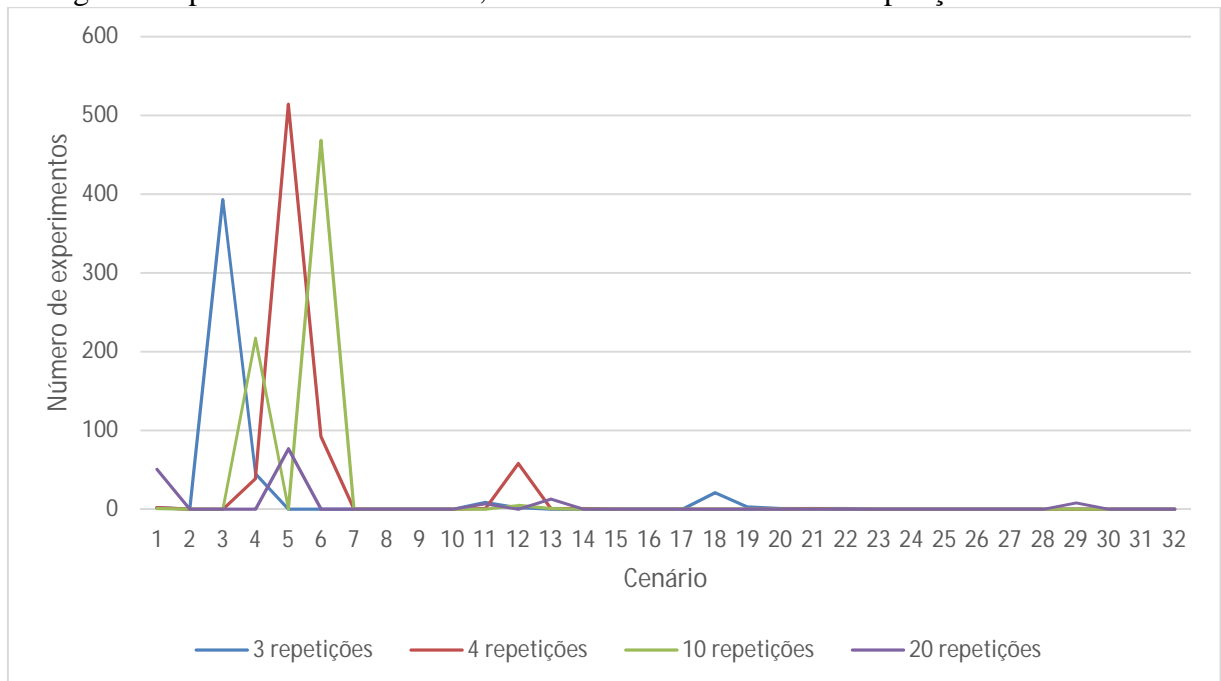
ANEXO A2 – Número, por cenário, de experimentos que apresentaram a situação de divergência 4 para o teste de SNK, de acordo com o coeficiente de variação dos ensaios



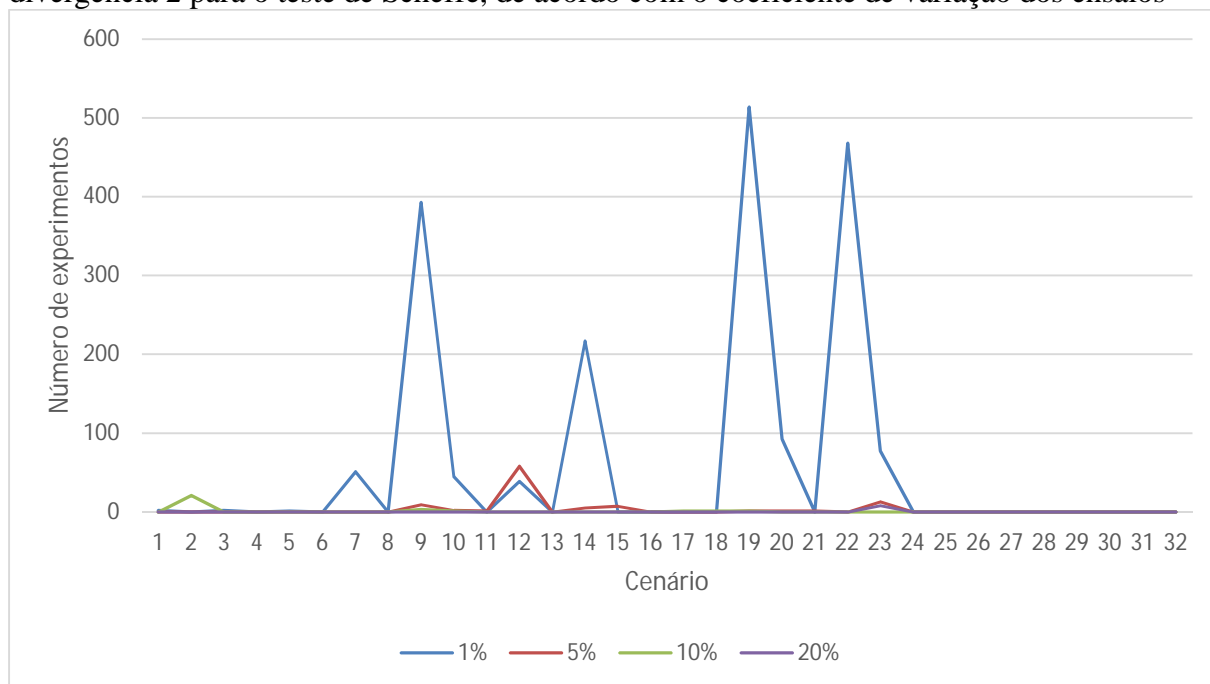
ANEXO B2 – Número, por cenário, de experimentos que apresentaram a situação de divergência 2 para o teste de Scheffé, de acordo com o número de tratamentos dos ensaios



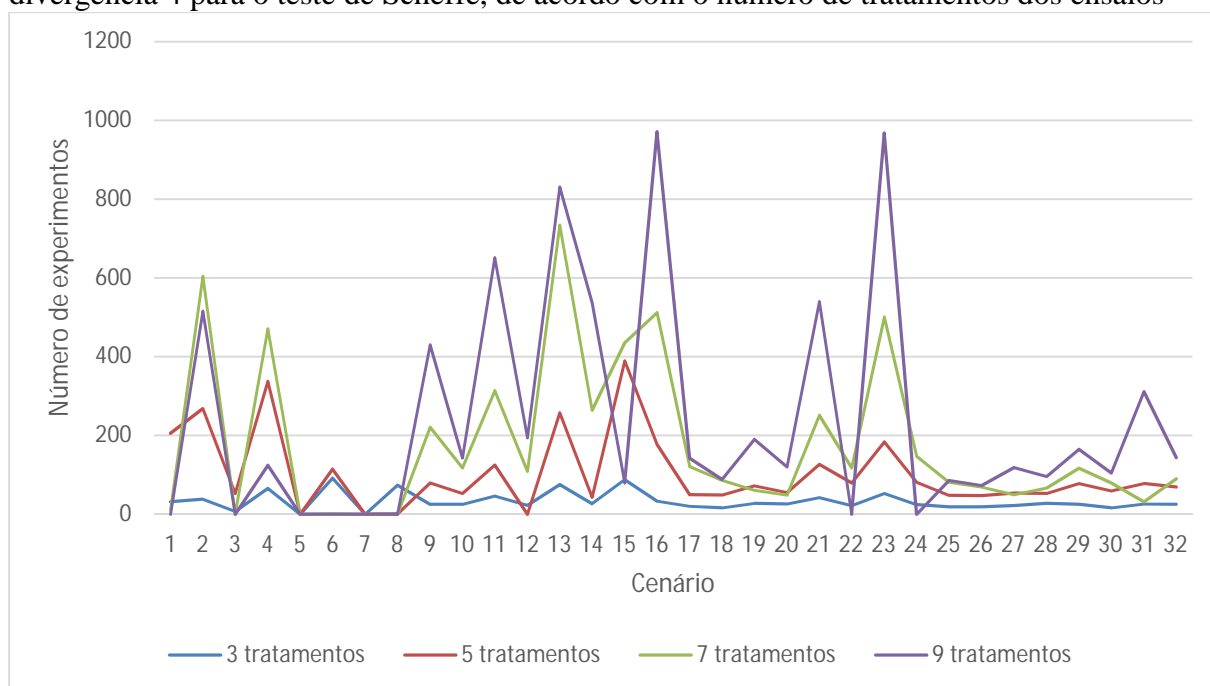
ANEXO C2 – Número, por cenário, de experimentos que apresentaram a situação de divergência 2 para o teste de Scheffé, de acordo com o número de repetições dos tratamentos



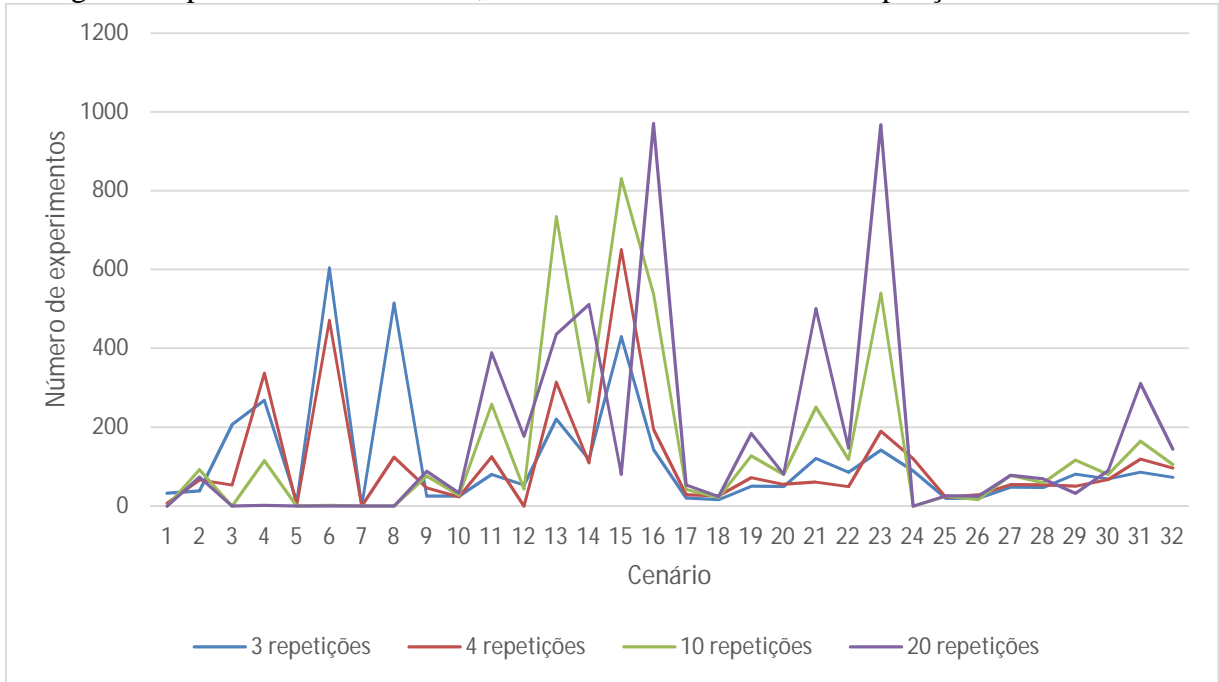
ANEXO D2 – Número, por cenário, de experimentos que apresentaram a situação de divergência 2 para o teste de Scheffé, de acordo com o coeficiente de variação dos ensaios



ANEXO E2 – Número, por cenário, de experimentos que apresentaram a situação de divergência 4 para o teste de Scheffé, de acordo com o número de tratamentos dos ensaios



ANEXO F2 – Número, por cenário, de experimentos que apresentaram a situação de divergência 4 para o teste de Scheffé, de acordo com o número de repetições dos tratamentos



ANEXO G2 – Número, por cenário, de experimentos que apresentaram a situação de divergência 4 para o teste de Scheffé, de acordo com o coeficiente de variação dos ensaios

