UNIVERSIDADE FEDERAL DE VIÇOSA
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
DEPARTAMENTO DE ESTATÍSTICA

EST 220 – ESTATÍSTICA EXPERIMENTAL

Viçosa – Minas Gerais
2010 / II

UNIVERSIDADE FEDERAL DE VIÇOSA
Departamento de Estatística
EST 220 – Estatística Experimental – 2010 / II

1. CONTEÚDO

Capítulo 1 – Testes de hipóteses
Capítulo 2 – Contrastes
Capítulo 3 – Introdução à Experimentação
Capítulo 4 – Delineamento Inteiramente Casualizado
Capítulo 5 – Procedimentos para Comparações Múltiplas
Capítulo 6 – Delineamento em Blocos Casualizados
Capítulo 7 – Delineamento em Quadrado Latino
Capítulo 8 – Experimentos Fatoriais
Capítulo 9 – Experimentos em Parcelas Subdivididas
Capítulo 10 – Regressão e Correlação

2. AVALIAÇÃO

Prova	Data	Horário	Local
1	03/09 (Sex)	18:20 h	A CONFIRMAR
2	15/10 (Sex)	20:30 h	A CONFIRMAR
3	26/11 (Sex)	20:30 h	A CONFIRMAR

O sistema de avaliação constará de três provas com pesos iguais, cujas datas foram sugeridas ao Registro Escolar. A nota final será a média das provas.

Será aplicada uma quarta prova escrita (29/11 – Seg – 12:00 h) que abordará todo o assunto do semestre, somente para o estudante que perder pelo menos uma das três provas por qualquer motivo.

Levar documento com foto para fins de fiscalização durante as provas.

Levar tabelas dos testes de hipóteses, formulário e calculadora para as provas, pois são de uso individual.

O coordenador da disciplina marcará um único período de revisão para cada uma das provas que deverá ser respeitado, dado que não serão abertas exceções para revisões de provas fora do período estabelecido.

As revisões de provas serão realizadas com o monitor durante o seu horário numa sala do Departamento de Estatística no prédio do CCE, mesmo que a monitoria regular esteja marcada para outro local.

A data da prova final será marcada pelo Registro Escolar.

3. MONITORIA

O horário e local da monitoria serão divulgados na terceira semana de aula.

Serão agendados horários extras durante a semana de cada prova, sendo o horário e local, divulgados no quadro de avisos do Departamento de Estatística no prédio do CCE.

4. BIBLIOGRAFIA

BARBETTA, P.A.; REIS, M.M. e BORNIA, A.C. Estatística para cursos de engenharia e informática. Editora Atlas, São Paulo, 2004. 410 p.

BANZATTO, D.A. e KRONKA, S.N. Experimentação agrícola. FUNESP, Jaboticabal, 1989. 249 p.

COSTA NETO, P.L.O. Estatística. Editora Edgard Blücher, São Paulo, 1977, 264 p.

GOMES, F.P. Curso de estatística experimental. 12ª edição, Livraria Nobel S.A, São Paulo, 1987. 467 p.

HINES, W.W.; MONTGOMERY, D.C.; GOLDSMAN, D.M. e BORROR, C.M. Probabilidade e estatística na engenharia. 4ª edição, LTC Editora, Rio de Janeiro, 2006. 588 p.

HOFFMANN, R. e VIEIRA, S. Análise de regressão: uma introdução à econometria. 2ª edição, Editora Hucitec, São Paulo, 1983. 379 p.

MONTGOMERY, D.C. e RUNGER, G.C. Estatística aplicada e probabilidade para engenheiros. 4ª edição, LTC Editora, Rio de Janeiro, 2009. 490 p.

MOORE, D.S. e McCABE, G.P. Introdução à prática da estatística. 3ª edição, LTC Editora, Rio de Janeiro, 2002. 536 p.

RIBEIRO JÚNIOR, J.I. Análises estatísticas no Excel – guia prático. Editora UFV, Viçosa, 2004. 249 p.

RIBEIRO JÚNIOR, J.I. e MELO, A.L.P. Guia prático para utilização do SAEG. Folha Artes Gráficas Ltda, Viçosa, 2008. 288 p.

VIEIRA, S. e HOFFMANN, R. Estatística experimental. Editora Atlas, São Paulo, 1989, 179 p.

5. PROFESSORES

Antonio Policarpo Souza Carneiro – CCE 313B – Ramal 1786
José Ivo Ribeiro Júnior – CCE 306B – Ramal 1783 (Coordenador)
Nerilson Terra Santos – CCE 312B – Ramal 1784
Sebastião Martins Filho – CCE 316B – Ramal 1773

6. HORÁRIOS DAS TURMAS

Horário	Seg	Ter	Qua	Qui	Sex
8		T1 - PVB310 Nerilson		T4 - PVB209 Sebastião	T3 - PVB209 Nerilson
10	T4 - PVB209 Sebastião	T3 - PVB209 Nerilson		T1 - PVB310 Nerilson	
14	T6 - PVB304 Policarpo		T5 - PVB105 José Ivo	T2 - PVB209 Sebastião	
16	T2 - PVB209 Sebastião		T6 - PVB304 Policarpo		T5 - PVB105 José Ivo

7. PLANEJAMENTO

Aula	Semana	Assunto
1 2	02 a 06/08	Apresentação da disciplina Testes de hipóteses: conceitos
3 4	09 a 13/08	Teste t e intervalo de confiança para uma média Teste F para duas variâncias, teste t para duas médias independentes
5 6	16 a 20/08	Intervalo de confiança para duas médias independentes, teste t para duas médias dependentes e intervalo de confiança para duas médias independentes Teste t e intervalo de confiança para duas médias dependentes
7 8	23 a 27/08	Contrastes: conceitos Métodos para obtenção de contrastes ortogonais
9 10	30 a 03/09	Princípios básicos da experimentação Tira dúvidas
		Prova 1 – 03/09 – Sex – 18:20 h
11	08 a 10/09	Delineamento inteiramente casualizado (DIC)
12 13	13 a 17/09	Análise de variância e pressuposições Delineamento em blocos casualizados (DBC)
14 15	20 a 24/09	Delineamento em quadrado latino (DQL) Testes de Tukey e Duncan
16	27 a 29/09	Testes t e de Scheffé
17 18	04 a 08/10	Experimento fatorial (EF) Interação AxB não significativa de EF
19	11 a 15/10	Interação AxB significativa de EF
20		Tira dúvidas
		Prova 2 – 15/10 – Sex – 20:30 h
21 22	25 a 29/10	Experimento em parcelas subdivididas (EPS) Interação AxB não significativa de EPS
23	03 a 05/11	Interação AxB significativa de EPS
24 25	08 a 12/11	Regressão linear de 1º grau Regressão linear de 2º grau
26	17 a 19/11	Regressão linear com delineamento experimental
27 28	22 a 26/11	Análise de correlação Tira dúvidas
		Prova 3 – 26/11 – Sex – 20:30 h
29 30		Prova 4 – 29/11 – Seg – 12:00 h Software estatístico – 30/11 – Ter – 12:00 h

Índice

Capítulo 1 - Testes de Hipóteses 1

Capítulo 2 - Contrastes 22

Capítulo 3 – Introduçao à Experimentação 30

Capítulo 4 - Delineamento Inteiramente Casualizado 37

Capítulo 5 – Procedimentos para Comparações Múltiplas 45

Capítulo 6 - Delineamento em Blocos Casualizados 53

Capítulo 7 - Delineamento em Quadrado Latino 65

Capítulo 8 - Experimentos Fatoriais 71

Capítulo 9 - Experimentos em Parcelas Subdivididas 95

Capítulo 10 - Regressão 111

Capítulo 11 – Respostas dos Exercícios 125

Anexo 1 - Formulário e Tabelas 151

Anexo 2 – Fórmula Geral para o Cálculo de Soma de Quadrados 167

Anexo 3 – Introdução ao Uso do Programa SAS 169

Anexo 4 – p-valor 190

Anexo 5 – Exemplo Extra ANOVA 191

1. Testes de Hipóteses

1.1. Introdução

Os testes de hipóteses fazem parte de um conjunto de procedimentos inferenciais usados em estatística. O uso de tais procedimentos permite ao pesquisador fazer inferências a respeito de uma população a partir de uma ou mais amostras representativas da população da qual as amostras foram retiradas.

No dia a dia usamos de inferência para tomarmos certas decisões. Por exemplo, quando vamos a feira para comprar abacaxi e um feirante nos oferece um pedaço de abacaxi. Qual o nosso procedimento? Se aquele pedaço de abacaxi for doce, concluímos que todo o lote de abacaxi vendido por aquele feirante é doce. Por outro lado, se o pedaço for azedo, inferimos que todo o lote é azedo. É lógico que podemos tomar decisões erradas devido à amostragem. Por exemplo, corremos o risco de levar abacaxi azedo para casa, mesmo que a nossa prova tenha sido doce. Isto pode acontecer porque o lote de abacaxi pode não ser completamente uniforme no teor de açúcar, ou porque experimentamos um abacaxi doce no meio de um lote composto por abacaxis azedos.

Este é um exemplo prático que ilustra o princípio básico do teste de hipóteses. Porém, em ciência é necessário que todos os procedimentos sejam padronizados e bem especificados. O objetivo deste capítulo é fornecer os conceitos teóricos fundamentais para um correto uso dos testes de hipóteses. Neste capítulo, serão abordados alguns dos testes de hipóteses mais comuns para comparar no máximo parâmetros de duas populações. Outros testes de hipóteses aplicáveis para comparações de parâmetros envolvendo mais de duas populações serão apresentados no Capítulo 5.

1.2. Conceitos fundamentais em testes de hipóteses

1.2.1 Parâmetro

Parâmetro é uma medida usada para caracterizar uma população. Assim sendo para se obter o valor de um parâmetro é necessário coletar a informação a respeito de uma ou mais variáveis em todos os indivíduos dessa população, ou seja, realizar um censo da mesma. É possível caracterizar uma população por meio de duas medidas principais: posição e dispersão.

As medidas de posição são também conhecidas como medidas de tendência central, pois elas indicam em que posição, a distribuição dos valores de uma população tendem a se concentrar. Alguns exemplos de medidas de posição são a média aritmética ($m = \mu = E(X)$), a mediana (Md) e a moda (Mo).

As medidas de dispersão indicam quanto os valores de uma população estão dispersos em torno de sua média. Como exemplo de medidas de dispersão temos a variância ($\sigma^2 = V(X)$) e o desvio-padrão (σ).

1.2.2 Estimador

Na grande maioria das situações, não é possível realizar o censo de uma população, porque ou a população é muito grande ou é de tamanho infinito. Para contornar este problema, o pesquisador pode retirar uma amostra da população e a partir desta amostra caracterizar a população de onde a amostra foi retirada sem nenhum viés.

Para alcançar este objetivo deve-se usar fórmulas estatísticas, conhecidas como estimadores, que apresentem características estatísticas desejáveis, tais como nãotendenciosidade, variância mínima, fornecer estimativas que se aproximem do valor paramétrico à medida que o tamanho da amostra aumenta, e etc..

Exemplos de estimadores são a média aritmética amostral, $\hat{m}$, que é usada para estimar a média populacional; e a variância amostral, s^2, que é usada para estimar a variância populacional. Outras simbologias comuns para a média amostral são $\hat{\mu}$ e $\overline{X}$, e para a variância amostral são $\hat{\sigma}^2$ e $\hat{V}(X)$.

Observe que algumas vezes a simbologia usada para representar os parâmetros e seus respectivos estimadores é muito parecida. Por exemplo, podemos representar a média populacional por m e seu estimador por $\hat{m}$, ou seja, a diferença entre o parâmetro e o seu estimador é o chapéu que existe no símbolo usado para representar o estimador. Isto parece ser uma diferença mínima, mas do ponto de vista estatístico, a diferença conceitual entre parâmetro e estimador é enorme.

O parâmetro é sempre um valor constante, pois para a obtenção do mesmo são usados todos os elementos da população. Por outro lado, o estimador representa uma variável aleatória, pois os seus valores mudam de amostra para amostra. Isto acontece porque os elementos que pertencem a uma amostra geralmente não são os mesmos em outras amostras. Conseqüentemente, é possível estabelecer uma distribuição de probabilidades para os valores de um estimador. Para o parâmetro, isto não é possível, pois se assume que ele tem um valor constante. Por isto recomenda-se muito cuidado para usar corretamente a simbologia para o parâmetro e paro o estimador.

Conforme mencionado anteriormente, os estimadores podem assumir valores diferentes em amostras diferentes. Estes diferentes valores que um estimador assume são também conhecidos como estimativas.

1.2.3 Hipóteses em um teste estatístico

Para realizar um teste de hipóteses e divulgar as conclusões é necessário seguir um procedimento aceito pela comunidade científica. Neste procedimento, o pesquisador deve deixar claro qual a hipótese que ele deseja testar. Para isto ele precisa escrever em termos estatísticos a sua hipóteses cientifica. A hipótese científica do pesquisador, nada mais é o que o levou a realizar a sua investigação.

Por exemplo, suponha que um tecnólogo em laticineos deseja verificar se os sabores de sorvete morango e chocolate apresentam um mesmo valor para o teor médio de glicose. Em termos estatísticos esta hipótese é expressa por

$$m_{morango} = m_{chocolate}$$

Em que:

$m_{morango}$: média do teor de glicose do sorvete sabor morango; e

$m_{chocolate}$: média do teor de glicose do sorvete sabor chocolate.

O pesquisador deseja testar esta hipótese porque ele desconfia que o teor médio de glicose não seja o mesmo para os dois sabores de sorvete. Então ele tem que ter uma alternativa para esta hipótese inicial. Nesta alternativa, ele lança a sua desconfiança a respeito do que pode acontecer. Se ele desconfiar que o sabor de morango tem um teor médio de glicose maior do que o de chocolate, então a hipótese alternativa é expressa por

$$m_{morango} > m_{chocolate}$$

Por outro lado, se ele desconfiar que o sabor de chocolate tem um teor de glicose maior do que o de morango, então a hipótese alternativa é expressa por

$$m_{morango} < m_{chocolate}$$

Uma outra alternativa seria a situação em que ele não tem nenhuma desconfiança de qual sabor teria um teor médio de glicose maior do que o outro. Neste caso, a hipótese alternativa é expressa por

$$m_{morango} \neq m_{chocolate}$$

Neste ponto fica claro que para realizar um teste de hipóteses é necessário que o pesquisador lance duas hipóteses. A primeira que contém um sinal de igualdade é conhecida como hipótese de nulidade, comumente denotada por Ho. É dado este nome, pois ela representa uma nulidade de diferença entre médias. Já a outra hipótese que contém um sinal de desigualdade, é conhecida como hipótese alternativa, comumente designada por Ha ou H1. Como o próprio nome diz, ela é uma alternativa a hipótese de nulidade. Na verdade, quando um pesquisador realiza um experimento, a hipótese de nulidade é construída com o expresso propósito de ser rejeitada. Isto faz sentido porque, quem teria o trabalho de realizar um experimento se achasse que duas médias são iguais? Qualquer um se daria ao trabalho de instalar um experimento, apenas se desconfiar que existe diferença significativa entre as médias de duas populações. No entanto, num teste de hipóteses, até que se prove o contrário, a Ho é considerada como a hipótese verdadeira.

Para o exemplo dado, supondo que o pesquisador não desconfie a princípio qual sabor que apresenta maior teor médio de glicose, o par de hipóteses a ser lançado é expresso por

$$H_0: \ m_{morango} = m_{chocolate}$$

$$H_a: \ m_{morango} \neq m_{chocolate}$$

Observe que apesar de ser possível existir três possibilidades para Ha, apenas uma possibilidade foi lançada. Outro ponto importante é que as hipóteses foram lançadas em termos dos parâmetros e não em termos dos seus estimadores. Não faz sentido lançar as hipóteses usando os estimadores, pois os mesmos não possuem um valor fixo, ou seja, apresentam valores diferentes para amostras diferentes, enquanto que o parâmetro possui um valor fixo.

1.2.4 Decisão em um teste de hipóteses

Para decidirmos se devemos ou não devemos rejeitar a hipótese de nulidade, baseamos na comparação do valor especificado para o parâmetro com aquele estimado a partir de uma amostra da população. Raramente, o valor estimado será idêntico àquele especificado para o parâmetro.

Conforme mencionado anteriormente, um estimador pode assumir valores diferentes para amostras diferentes, sendo que existem intervalos de valores mais prováveis de ocorrer do que outros. Portanto pode-se construir uma distribuição de probabilidades para os valores de um estimador.

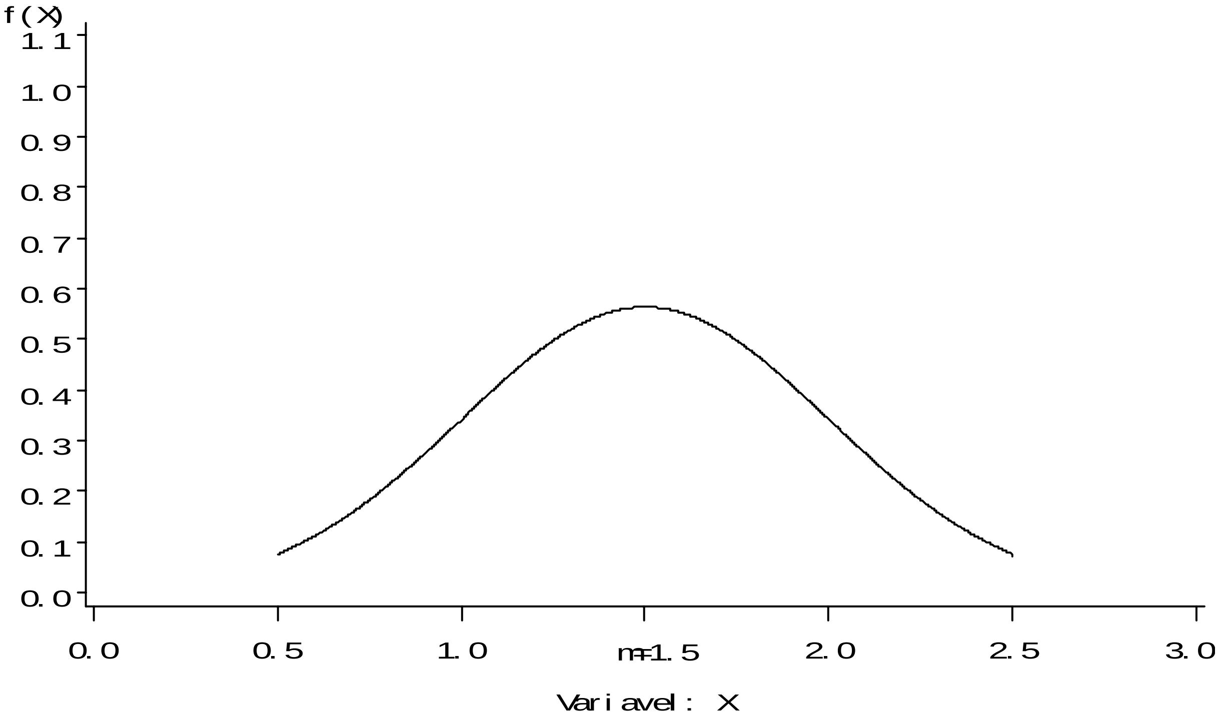
O valor fornecido pelos estimadores poderá diferir, do ponto de vista matemático, do valor esperado para o parâmetro. Esta diferença matemática nem sempre representa que a hipótese de nulidade deve ser rejeitada, pois como o estimador é uma variável aleatória, é esperado que ele possa assumir valores dentro de um intervalo. O que um teste de hipóteses geralmente faz é comparar duas fontes de variação. A primeira fonte de variação diz respeito a variação entre o valor paramétrico e uma estimativa. A segunda fonte de variação diz respeito a variação existente na população.

Se as duas fontes de variação apresentarem valores semelhantes então o valor do parâmetro não difere do valor especificado na hipótese de nulidade. Neste caso, a variação observada entre o valor paramétrico e sua estimativa é uma variação própria dos dados. Conclui-se portanto que a hipótese H_0 não deve ser rejeitada.

Por outro lado, se as duas fontes de variação apresentarem valores bem diferentes, conclui-se que a variação entre o valor especificado para o parâmetro e o de sua estimativa não é própria dos dados. Neste caso a variação entre o valor paramétrico e a estimativa é significativa, o que leva a rejeitar-se a hipótese de nulidade.

Para então decidirmos entre rejeitar ou não-rejeitar a hipótese de nulidade devemos estabelecer o que é uma “pequena” e uma “grande” variação. Para isto, precisamos conhecer a distribuição de probabilidades do estimador usado para estimar o parâmetro. Vamos ilustrar esta situação com o seguinte exemplo.

Suponha que um pesquisador desconfie que a estatura média de adolescentes na faixa etária de 13 a 15 anos é menor do que aquela informada por um órgão oficial como sendo igual a 1,5 metros. Este pesquisador sabe de fontes seguras que a estatura é uma variável aleatória que segue uma distribuição normal com variância igual a 0,25 metros^2. Se a informação do órgão oficial for verdadeira, ou seja a média de estatura igual a 1,50 metros, poderíamos descrever a distribuição de valores da variável estatura, digamos X, como $X \sim N(1{,}5\,;\ 0{,}25)$ e representar esta distribuição por meio do gráfico



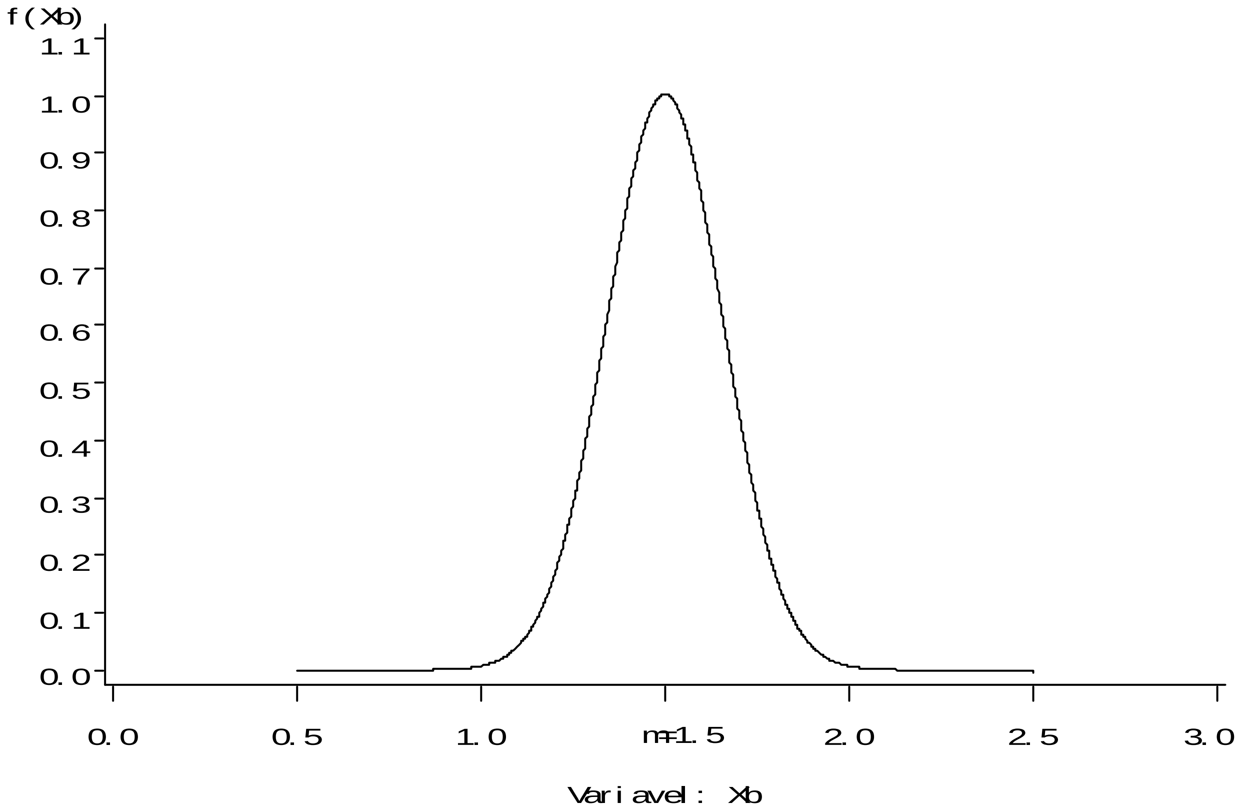
A função densidade de probabilidade de uma variável aleatória contínua que tem distribuição normal, no caso, f(X) é dada por:

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

Para verificar se a informação do órgão oficial é correta, o pesquisador tem duas opções: medir a estatura da população de todos os adolescentes, ou então tomar uma amostra de adolescentes e medir a estatura dos mesmos e usar um teste de hipóteses. Na primeira opção nenhum teste de hipóteses seria necessário, pois o pesquisador teria condições de conhecer o verdadeiro valor da média de estatura, ou seja, ele conheceria o parâmetro média daquela população de adolecentes. Na segunda opção, o pesquisador teria que usar uma média da amostra para tomar a sua decisão.

É evidente que a segunda opção é operacionalmente mais fácil, pois o custo e o tempo gasto são muito menores. Para realizar a segunda opção, o pesquisador deve escolher um tamanho de amostra adequado, por exemplo, suponha que para este exemplo o tamanho amostral ideal seja igual a 10 indivíduos. Da população de adolecentes é possível retirar um grande número de diferentes amostras de tamanho 10.

Cada amostra fornece um valor para a média amostral. Pode ser demonstrado que a média de todas as médias amostrais é igual à média da variável original, a variância é igual à variância original dividido pelo tamanho da amostra e que a variável aleatória $\hat{m}$ também segue distribuição normal, ou seja, $\hat{m} \sim N(1{,}5\,;\quad 0{,}025)$. O gráfico da distribuição das médias amostrais seria



em que Xb = $\hat{m}$ e f(Xb) = f($\hat{m}$).

Como pode ser notado, a distribuição das médias amostrais para a variável estatura, representadas no gráfico por Xb, é mais concentrada em torno da média do que a variável original X. Isto acontece porque a variância das médias amostrais é menor do que a variância da variável original estatura.

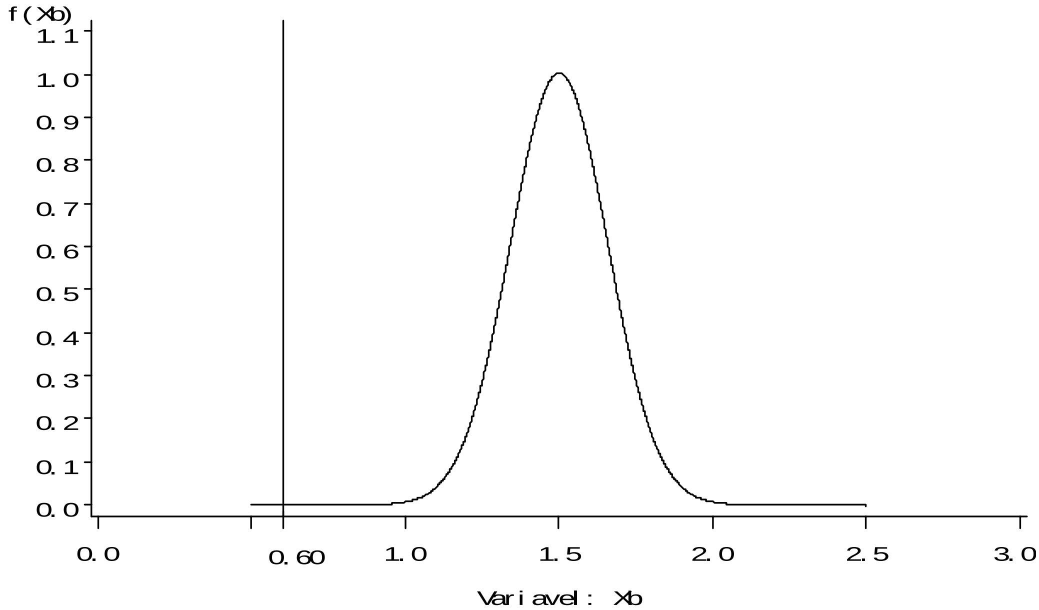
Deve ficar entendido que é possível retirar um número muito grande de amostras de mesmo tamanho de uma população, principalmente se a população for muito grande. No entanto, numa pesquisa geralmente toma-se decisão usando-se apenas uma única amostra. As hipóteses estatísticas para esta situação seriam:

$$H_O : m_{altura} = 1{,}5 \text{ metros}$$

$$H_a : m_{altura} < 1{,}5 \text{ metros}$$

Para se entender a lógica dos testes de hipóteses, vamos supor diferentes resultados possíveis para a média amostral obtida a partir de uma amostra de 10 estudantes. Suponha inicialmente que o pesquisador, obtenha uma média amostral, digamos $\hat{m}$, igual a 1,49 metros. Neste caso, a variação entre o valor observado igual a 1,49 e o valor suposto igual a 1,50 é muito *pequena*. Poder-se-ia atribuir esta variação ao acaso, ou seja, esta variação é uma variação própria de uma população que apresente média igual a 1,5 metros. Em termos probabilísticos poderíamos dizer que existe uma grande probabilidade de numa população com média igual a 1,50 metros existir grupos de 10 indivíduos que apresentem uma média de estatura igual ou inferior a 1,49 metros. Justificativa semelhante poderia ser atribuída a médias amostrais que tivessem valores *próximos* ao valor suposto, tais como: 1,48; 1,47; 1,42; etc.

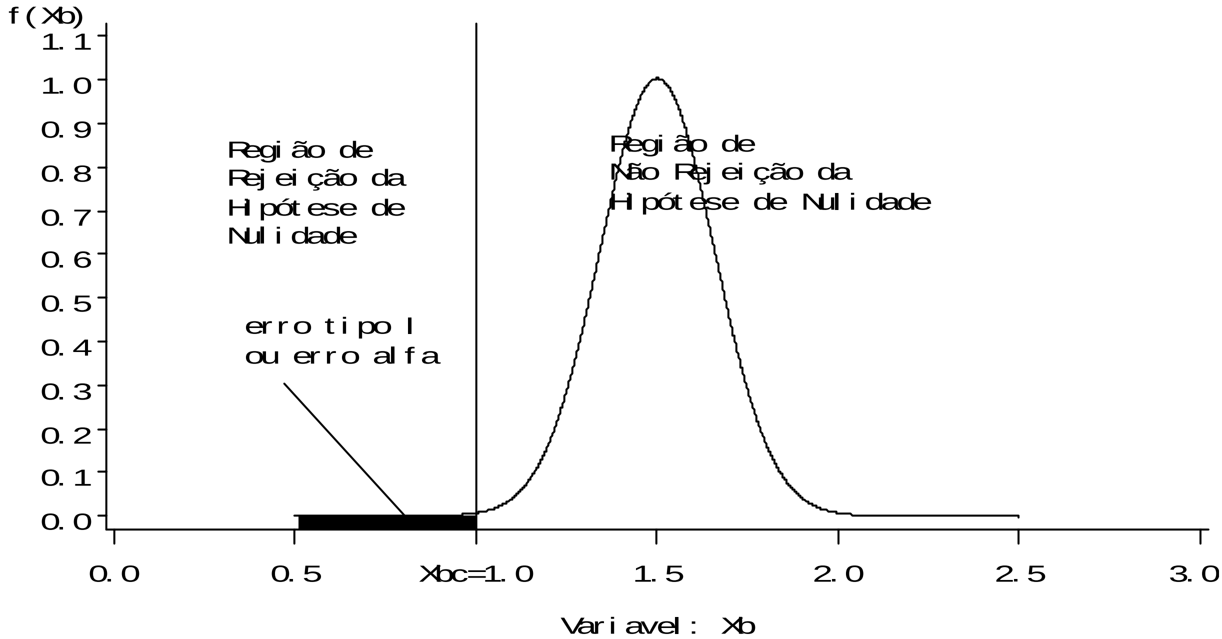
Por outro lado, se a média amostral apresentar um valor muito distante do valor suposto, como por exemplo, 0,60 metros, o pesquisador tem a tendência de rejeitar a hipótese de nulidade, isto porque há um forte indício de que a amostra foi retirada de uma população que apresenta uma média menor do que a suposta de 1,5 metros. Em termos probabilísticos poderia se dizer que a probabilidade de encontrar um grupo de indivíduos com média igual ou inferior a 0,60 metros é muito pequena, em uma população que apresenta uma média igual a 1,5 metros. Veja na figura a seguir



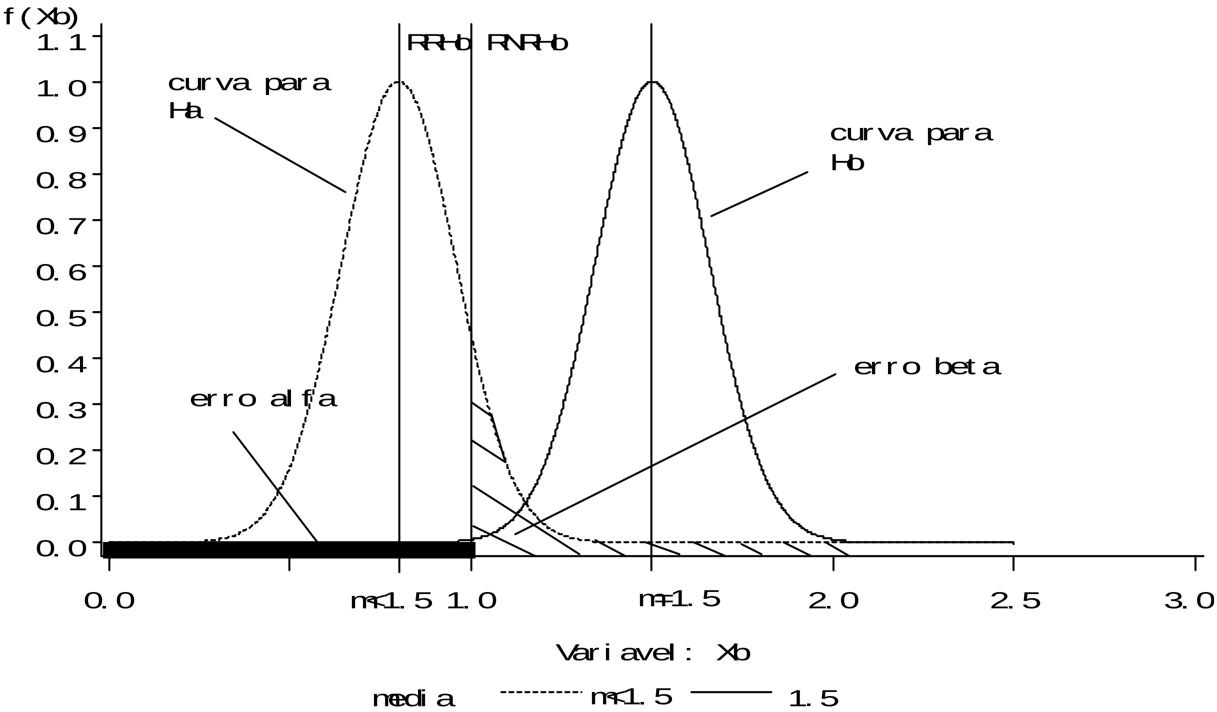
A função densidade de probabilidade da média amostral de uma variável aleatória que tem distribuição normal, no caso, f(Xb), é dada por:

$$f(Xb) = \frac{1}{\sigma\sqrt{\frac{2\pi}{n}}} e^{-\frac{1}{2}\left(\frac{x-m}{\frac{\sigma}{\sqrt{n}}}\right)^2}$$

A área sob a curva abaixo do valor 0,60 m, indica a probabilidade de se encontrar um valor igual ou inferior a 0,60 metros em uma população com média igual a 1,5 metros. Como pode ser notado, esta probabilidade é pequena em relação à área total do gráfico. Com base neste raciocínio é que o pesquisador estabelece um valor crítico que o ajuda a decidir sobre rejeitar ou não-rejeitar a hipótese de nulidade. Este valor crítico pode a princípio ser estabelecido de duas maneiras. A primeira delas seria a situação em que o pesquisador de posse de seu conhecimento prévio no assunto estabeleceria um valor crítico antes de coletar a amostra. Este valor crítico seria um valor para a média amostral tal que acima dele o pesquisador não-rejeitaria a hipótese de nulidade e abaixo dele rejeitaria a hipótese de nulidade. Digamos que neste caso o valor crítico adotado fosse igual a 1,0 metro. O valor para a média igual a 1,0 metro determinaria duas regiões na distribuição das médias amostrais, conforme é apresentado na figura a seguir. Estas duas regiões são denominadas como Região de Não-Rejeição da Hipótese de Nulidade **(RNRHo)** e Região de Rejeição da Hipótese de Nulidade **(RRHo)** . Como os respectivos nomes indicam, se o valor da média amostral estiver contido na **RNRHo**, o pesquisador não deve rejeitar a hipótese de nulidade. Caso contrário, se o valor da média amostral estiver contido na **RRHo**, o pesquisador deve rejeitar a hipótese de nulidade e considerar a hipótese alternativa como sendo a hipótese verdadeira.

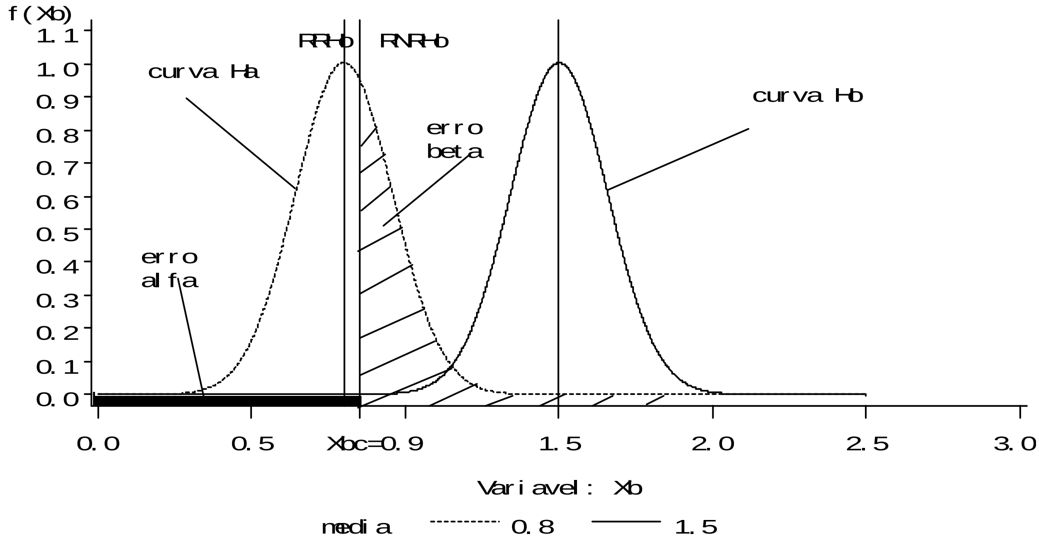


Deve-se observar que ao adotar o critério acima, o pesquisador sempre estará sujeito a cometer um de dois erros possíveis. Um destes erros, conhecido como erro tipo I ou erro alfa ($\propto$), se refere à probabilidade de rejeitar uma hipótese verdadeira, no caso a hipótese de nulidade. Na figura citada anteriormente, o critério adotado pelo pesquisador foi que se a média amostral assumisse um valor menor que 1,0 metro, então rejeitar-se-ia a hipótese de nulidade. É exatamente a adoção deste critério que pode levar o pesquisador a cometer um erro em sua tomada de decisão, pois como se pode observar na figura, em uma população que realmente apresenta média igual a 1,5 metros, existe uma pequena percentagem de indivíduos que podem apresentar uma altura média inferior a 1,0 metro. No entanto, o pesquisador acaba assumindo que devido ao fato daquela chance ser muito pequena, ele decide que se uma amostra de elementos apresentar média menor que 1,0 metro, ela pertence a uma população com média inferior à especificada de 1,5 metros, conforme é mostrado na figura a seguir.



Nesta figura, pode-se observar duas curvas: a da esquerda quando se assume que a população tem uma média inferior a especificada, isto é a curva para a hipótese alternativa (Ha) com m < 1,5 metros; e a curva da direita para a situação em que a população apresenta média igual à especificada, ou seja, curva para a hipótese de nulidade (Ho) com média m = 1,5 metros. Quando o pesquisador toma a decisão de rejeitar a Ho, ele na verdade acaba por concluir que a população de onde foi retirada a amostra pertence aquela população com média m < 1,5 metros. Observe, valores nesta região podem levar a duas conclusões que a rigor ambas estariam “corretas”, mas a probabilidade de encontrar indivíduos com média inferior ou igual ao valor crítico, no caso 1,0 metro, é bem maior numa população com m < 1,5 metros do que numa população com média m = 1,5 metros. É esta diferença nas probabilidades que leva o pesquisador a rejeitar Ho ao invés de não rejeitá-la.

Conforme mencionado anteriormente, a área sob a curva da hipótese Ho que leva a sua rejeição se refere à probabilidade de se rejeitar Ho quando Ho é verdadeira. Isto foi definido anteriormente como erro alfa. Um raciocínio lógico que se tem é tentar fazer este erro ser o menor possível. No entanto, em todo teste de hipóteses existe também um outro erro, conhecido como erro tipo II ou erro beta (β), o qual aumenta o seu valor à medida que se diminui o erro alfa. Este erro se refere à probabilidade não-rejeitar a hipótese Ho quando Ho é falsa (ver figura anterior). No exemplo que estamos trabalhando, este erro beta será tanto maior, quanto menor for o valor crítico. Se por exemplo, fizermos que o valor crítico para a média amostral seja igual a 0,9 m, então a nova proporção entre os erros alfa e beta seria conforme figura a seguir.



Nós acabamos de ver a maneira empírica de realizar um teste de hipótese, a qual se baseia no fato do pesquisador estabelecer o valor crítico de rejeição da hipótese Ho com base em seu prévio conhecimento do problema. Este procedimento, embora seu forte apelo prático, traz a desvantagem de não poder estabelecer a princípio qual seria a probabilidade de se cometer o erro tipo I, ou seja, a que nível de significância que o teste de hipóteses será realizado. É de consenso que se publique, que nos trabalhos científicos, a que nível de significância um teste de hipóteses foi realizado. Desta forma, é possível comparar os resultados e conclusões de diferentes trabalhos de pesquisa, pois existe uma tendência que, para determinada área do conhecimento, o nível de significância esteja dentro de uma faixa de valores aceito pela maioria dos pesquisadores. A determinação do nível de significância quando se usa o método empírico é possível, embora computacionalmente não seja uma tarefa fácil, pois envolve a integração de funções complexas tais como exponenciais, gama, beta, e etc. . Devido a todas estas razões, o método não-empírico é o mais usado.

O procedimento para um teste de hipóteses usando o método não-empírico é similar ao método empírico. A diferença está basicamente que no método não-empírico, o valor crítico é conhecido a partir do nível de significância estabelecido e o uso de tabelas estatísticas. Existe uma tabela estatística apropriada para cada tipo de teste de hipóteses. Estas tabelas fornecem valores críticos que delimitam regiões de rejeição e de não-rejeição de Ho. O valor obtido de uma ou mais amostras retirada da(s) população(ões) é então usado para calcular o valor de uma estatística que tem distribuição de probabilidades idêntica àquela usada para identificar o valor tabelado. A comparação dos valores calculado e tabelado permite ao pesquisador decidir entre rejeitar ou não-rejeitar Ho.

Os próximos itens deste capítulo irão tratar sobre alguns testes de hipóteses que usam este método não-empírico.

1.3. Alguns testes de hipóteses

1.3.1 Teste t de Student - Teste para pequenas amostras

A aplicação do teste t é indicada quando o tamanho amostral é igual ou inferior a 30 elementos. Para amostras com tamanho superior a 30, recomenda-se o teste Z. O uso do teste t pressupõe que a característica em análise é normalmente distribuída com variância populacional desconhecida.

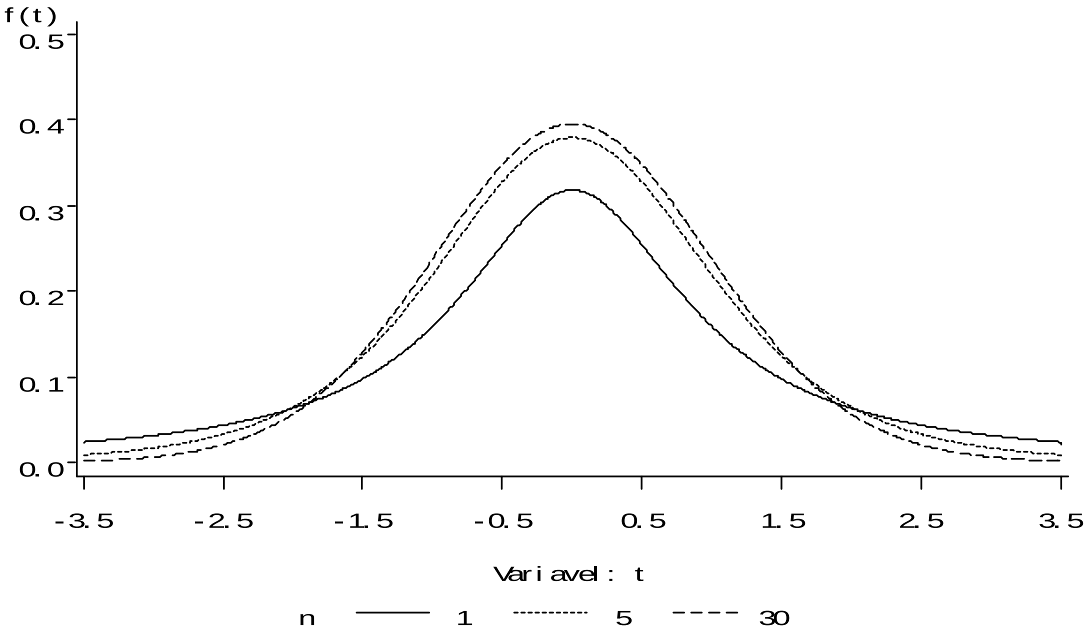
O teste t tem três aplicações principais: teste para uma média populacional, teste para duas médias populacionais e teste para mais que duas médias populacionais. As duas primeiras aplicações vão ser apresentadas neste capítulo. A terceira aplicação será apresentada no Capítulo 5.

1.3.1.1 Teste de hipóteses para uma média populacional

Este teste é usado para verificar se a média de uma característica de uma população assume um valor especificado, digamos m_0. Para aplicação deste teste devemos selecionar uma amostra aleatória de tamanho n da população. Digamos que os elementos amostrais sejam; $X_1, X_2, ..., X_n$. Com base nestes elementos amostrais, calculamos a sua média, $\hat{m}$, e seu desvio padrão, s. Estas estatísticas são então utilizadas para calcular o valor de t usando a expressão

$$t = \frac{\hat{m} - m_0}{\frac{s}{\sqrt{n}}}$$

Esta estatística t, tem distribuição t de Student com n-1 graus de liberdade, ou seja, é uma distribuição de probabilidades que depende do número de graus de liberdade associado. A figura a seguir, ilustra a distribuição t para três valores diferentes no número de graus de liberdade.



As hipóteses num teste t, para uma média populacional, são do seguinte tipo

H_0: $m = m_0$ versus
H_a: $m > m_0$ ou
H_a: $m < m_0$ ou
H_a: $m \neq m_0$

Para decidirmos entre Rejeitar ou Não-Rejeitar H_O, comparamos o valor de t com o valor tabelado de t obtido por $t_{tab} = t_\alpha(n-1)$. A tabela apresentada no final deste livro é uma tabela elaborada para testes bilaterais. Neste caso, para encontrarmos o valor tabelado basta entrar com o valor de α e o respectivo número de graus de liberdade. Por outro lado, se desejarmos realizar um teste unilateral e usarmos uma tabela bilateral, devemos entrar na tabela com 2α como nível de significância. Este procedimento garante que realizaremos o teste ao nível de significância α como desejado para testes unilaterais.

Depois de obtido o valor calculado e o valor tabelado de t, usamos a seguinte regra decisória:

- se $|t| \geq t_{tab}$ então Rejeita-se H_o
- se $|t| < t_{tab}$ então Não-Rejeita-se H_O.

Exercícios

1.1. Em indivíduos sadios, o consumo renal de oxigênio distribui-se normalmente em torno de 12 cm^3/min. Deseja-se investigar, com base em cinco indivíduos portadores de certa moléstia, se esta tem influência no consumo renal médio de oxigênio. Os consumos medidos para os cincos pacientes foram:

14,4 12,9 15,0 13,7 13,5

Qual a conclusão ao nível de 1% de significância?

1.2. Uma amostra de seis elementos, extraída de uma população normal, forneceu

$$\sum_{i=1}^{6} X_i = 84{,}0 \quad \text{e} \quad \sum_{i=1}^{6} (X_i - \hat{m})^2 = 55{,}0$$

Deseja-se saber se a média da população pode ser considerada como superior a 11. Qual a conclusão, ao nível de 5% de significância?

1.3.1.2 Teste de hipóteses para duas médias populacionais

O objetivo deste teste é verificar se duas populações, digamos população 1 e população 2 apresentam um mesmo valor médio para uma determinada característica, isto é deseja-se verificar se $m_1 = m_2$. Com esta finalidade é necessário obter uma amostra de cada população. Estas duas amostras podem ser relacionadas ou não, ou seja, podem ser dependentes ou independentes uma da outra. Esta distinção no relacionamento das duas amostras gera dois testes distintos.

1.3.1.2.1 Teste de hipóteses para o caso de duas amostras independentes

Duas amostras são ditas serem independentes quando **não existe** nada que as relacione. Nesta situação, os valores amostrais foram obtidos em conjuntos amostrais distintos, ou seja, os elementos amostrais que originaram os valores de uma amostra são distintos dos elementos amostrais que originaram a segunda amostra.

Conforme mencionado anteriormente, para comparar as médias das duas populações, toma-se uma amostra de cada população. Suponha que as amostras geradas sejam X_{11}, X_{12},... , X_{1n} e X_{21}, X_{22}, ... , X_{2m}, onde o tamanho das amostras podem ser diferentes, ou seja, n pode ser diferente de m. Para cada amostra, então calcula-se a sua média e variância. Um estimador comum para a variância é obtido tomando-se uma média ponderada das estimativas de variância obtidas para as duas amostras. O tamanho da amostra é utilizado como um peso para o cálculo desta variância média ponderada. A obtenção de um estimador comum para a variância pressupõe que a variância das duas populações sejam idênticas, ou seja $\sigma_1^2 = \sigma_2^2$. A fórmula do estimador comum é:

$$s_c^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

em que s_1^2e s_2^2são as variâncias amostrais das populações 1 e 2, respectivamente. A fórmula geral para o cálculo da variância amostral é dada por

$$s^2 = \frac{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}{n-1}$$

Uma vez obtidas estas estimativas, calcula-se o valor da estatística t dada por:

$$t = \frac{(\hat{m}_1 - \hat{m}_2) - (m_1 - m_2)}{\sqrt{s_c^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Esta estatística tem distribuição t de Student com $(n_1 + n_2 - 2)$ graus de liberdade. A comparação do valor calculado de t com o valor tabelado dado por $t_{tab} = t_\alpha(n_1 + n_2 - 2)$, é usada para testar a hipótese de nulidade

H_0: $m_1 = m_2$ versus
H_a: $m_1 > m_2$ ou
H_a: $m_1 < m_2$ ou
H_a: $m_1 \neq m_2$

A regra de decisão é idêntica ao caso anterior, ou seja:

- se **|** t **|** $\geq t_{tab}$ $\rightarrow$ Rejeita-se H_o
- se **|** t **|** $\geq t_{tab}$ $\rightarrow$ Não-Rejeita-se H_O.

Exercício

1.3. Os dados que seguem referem-se a cinco determinações da resistência de dois tipos de concreto. Ao nível de 5% de significância, há evidência de que o concreto 1 seja mais resistente que o concreto 2?

Concreto 1	54	55	58	51	57
Concreto 2	50	54	56	52	53

1.3.1.2.2 Teste de hipóteses para o caso de duas amostras dependentes

Duas amostras de elementos são ditas serem dependentes quando **existe** algo que as relacione. Por exemplo, se os valores de duas amostras foram obtidos de um mesmo conjunto de elementos amostrais, podemos dizer que as duas amostras de valores são dependentes uma vez que foram tomados de um conjunto de elementos amostrais comum.

O objetivo neste caso é verificar se houve alteração na média de uma população quando a mesma é avaliada sob duas condições diferentes. Cada condição representa uma população distinta, embora se suponha que os elementos populacionais sejam os mesmos nas duas condições. Para verificar se houve alteração na média, avalia-se uma característica de interesse do pesquisador num conjunto de elementos amostrais tomados ao acaso na população quando a mesma esteja sob a condição 1. Digamos que a avaliação da característica resulte nos seguintes valores amostrais X_{11}, X_{12},... , X_{1n}. Depois de feita esta avaliação, os elementos amostrais que originaram a primeira amostra, sejam submetidos à condição 2. Os mesmos elementos amostrais são novamente avaliados para a mesma característica na nova condição 2. Digamos que esta nova avaliação resulte nos seguintes valores amostrais X_{21}, X_{22}, ... , X_{2n}. Se a condição 2 não tiver nenhum efeito, espera-se que em média os valores observados nas duas condições sejam iguais.

Em termos de desvios, se a alteração das condições não resultasse em nenhum efeito significativo, poderíamos dizer que a diferença entre os valores observados na primeira condição e na segunda condição seria em média igual a zero. Portanto para verificar se houve alteração na média de uma população avaliada em duas condições diferentes, pode-se testar a hipótese de que o desvio médio ser estatisticamente igual a zero. Portanto, a partir de duas amostras obtém-se uma outra baseada nos desvios, conforme é mostrado a seguir.

Elemento amostral i	1	2	...	n
Amostra 1	X_{11}	X_{11}	...	X_{1n}
Amostra 2	X_{21}	X_{22}	...	X_{2n}
$d_i=X_{1i}-X_{2i}$	d_1	d_2	...	d_n

Apresentado desta forma, o teste t para duas amostras dependentes reduz-se teste t para uma média populacional, visto anteriormente. No presente caso, deseja-se testar se a média dos desvios é igual por exemplo a um valor m_0. Escrevendo em termos de hipóteses estatísticas teríamos

H_0: $m = m_0$ versus
H_a: $m > m_0$ ou
H_a: $m < m_0$ ou
H_a: $m \neq m_0$

Para decidir entre Rejeitar ou Não-Rejeitar a hipótese de nulidade, deve-se calcular o valor da estatística t dada por

$$t = \frac{\hat{m} - m_0}{\sqrt{\frac{s^2}{n}}}$$

em que

$$\hat{m} = \frac{\sum_{i=1}^{n} d_i}{n}$$

$$s^2 = \frac{\sum_{i=1}^{n} d_i^2 - \frac{\left(\sum_{i=1}^{n} d_i\right)^2}{n}}{n-1}$$

Sob H_o, esta estatística t tem distribuição t de Student com n-1 graus de liberdade. A comparação deste valor calculado com o valor de t_{tab} dado por $t_{tab} = t_{\alpha}(n-1)$.

Depois de obtido os valores calculado e tabelado de t, usamos a seguinte regra decisória:

- se $|t| \geq t_{tab}$ então Rejeita-se H_o
- se $|t| < t_{tab}$ então Não-Rejeita-se H_O.

Exercícios

1.4. Com o objetivo de avaliar se determinado produto químico é eficiente para repelir insetos domésticos, foi realizada uma contagem do número de insetos, antes e após a aplicação deste produto químico, em 7 residências. O número de insetos observado em cada residência foi

Residênca	1	2	3	4	5	6	7
Antes da aplicação	8	6	7	8	9	6	7
Após a aplicação	4	0	3	5	3	4	2

Por meio destes dados e ao nível de 5% de probabilidade, é possível concluir, em termos médios, que o produto utilizado é eficiente para repelir insetos?

1.5. Com a finalidade de testar se determinado método de secagem rápida consegue reduzir significativamente a quantidade média de água de grãos de cereais, uma porção de cada um dos seguintes tipos de cereais: Milho, Cevada, Trigo, Arroz e Sorgo, foi exposta ao referido método de secagem. Os resultados obtidos, para o peso da porção (em g) amostrada por cereal, com a realização do experimento foram:

	Milho	Cevada	Trigo	Arroz	Sorgo
Sem a secagem	30	34	41	25	36
Com a secagem	21	28	33	21	31

É possível concluir ao nível de 5% de significância que o método de secagem proposto, é eficiente para secar os grãos?

1.3.2 Teste F para Comparação de Variâncias de Duas Populações

Este teste é indicado para verificar se duas populações, digamos 1 e 2, apresentam igual valor para o parâmetro variância. Em termos de hipóteses estatísticas teríamos:

H_0: $\sigma_1^2 = \sigma_2^2$ versus

H_a: $\sigma_1^2 > \sigma_2^2$ ou

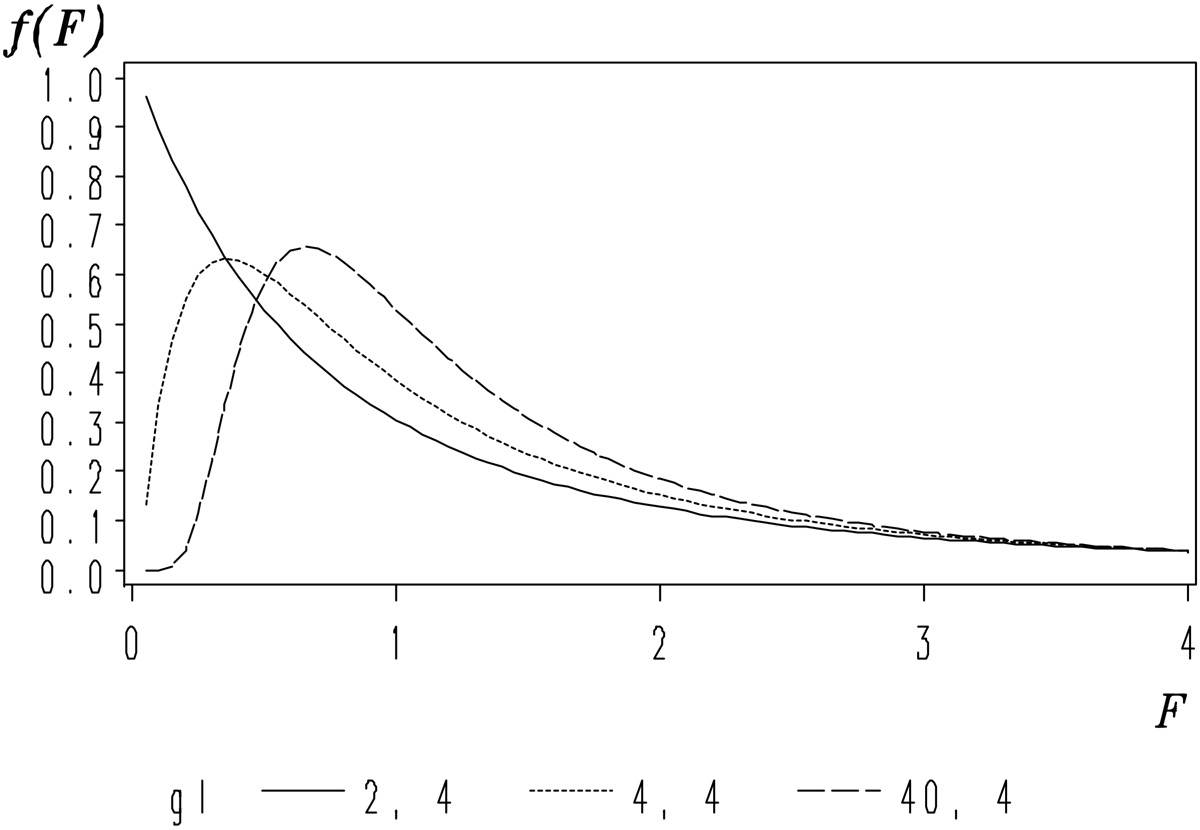
H_a: $\sigma_1^2 < \sigma_2^2$ ou

H_a: $\sigma_1^2 \neq \sigma_2^2$

A estatística F usada para decidir entre Rejeitar ou Não-Rejeitar Ho é dada pelo quociente entre as duas estimativas de variância, ou seja:

$$F = \frac{s_1^2}{s_2^2}$$

Sob a hipótese de nulidade, este quociente tem distribuição F, de Fisher-Snedecor, com n_1 e n_2 graus de liberdade, ou seja a distribuição de probabilidades da estatística F depende dos números de graus de liberdade n_1 e n_2. Um gráfico para a distribuição F, para três diferentes pares de graus de liberdade é ilustrado na figura a seguir.



A conclusão do teste é feita mediante a comparação do valor de F com o valor de $F_{tab}=F_{\alpha} = (n_1, n_2)$.

Se $F \geq F_{tab} \Rightarrow$ Rejeita-se H_0 ao nível α de probabilidade. Caso contrário Não-Rejeita-se H_O

Exercícios

1.6. Com o intuito de controlar a homogeneidade da produção de certas partes ao longo do tempo, amostras semanais são retiradas da produção corrente. Uma primeira amostra, de dez elementos, forneceu média 284,55 e desvio padrão 0,320, ao passo que, numa segunda amostra, forneceu, nas mesmas unidades, os seguintes valores:

284,6 283,9 284,8 285,2 284,3 283,7 284,0

Ao nível de 5% de significância, podemos concluir que a semana 2 apresentou maior variabilidade que a semana 1?

1.7. A qualidade de rebites é tanto melhor quanto maior sua homogeneidade. Seis rebites de duas marcas foram ensaiados ao cisalhamento, tendo-se obtido as seguintes cargas de ruptura:

Rebite	1	2	3	4	5	6
Marca A	34,9	35,5	38,8	39,2	33,7	37,6
Marca B	38,5	39,0	40,7	42,9	37,8	41,4

Estes resultados ratificam a afirmação do produtor da marca B, de que seus rebites são melhores? Use o nível de 5% de significância.

1.4. Exercícios Suplementares

1.8. Uma fábrica de cerâmica produz um tipo de peça usando o processo A de fabricação. Com o objetivo de melhorar a média de resistência das peças, quando submetidas a determinado grau de temperatura, o processo B foi introduzido. Com os dados amostrais abaixo, relativos à temperatura de rompimento das peças, testar a hipótese H_o e concluir para $\alpha = 5\%$.

PROCESSO A	90,3	93,4	96,8	91,4	92,6	102,5	103,4
PROCESSO B	101,4	98,5	104,6	95,8	96,2	94,6	99,5

1.9. Um material isolante foi utilizado com a finalidade de reduzir a temperatura média interna em ambientes similares. Para testar a hipótese H_o, 10 ambientes foram selecionados ao acaso e expostos a uma determinada fonte de radiação de calor. Testar a hipótese H_o e concluir para $\alpha = 5\%$. Os dados obtidos (em ºC) são fornecidos abaixo.

AMBIENTE	1	2	3	4	5	6	7	8	9	10
s/isolante	30,5	35,3	33,2	40,8	42,3	41,5	36,3	43,2	34,6	38,5
c/isolante	28,2	35,1	33,2	35,6	40,2	37,4	34,2	42,1	30,5	38,4

1.10. Dois processos que têm por objetivo o controle da temperatura média interna em ambientes foram colocados em competição. Para testar a H_o, 20 ambientes foram convenientemente preparados. Testar e concluir para $\alpha = 5\%$, considerando os dados abaixo.

PROCESSO	Temperatura °C									
s/isolamento	30,5	35,3	33,2	40,8	42,3	41,5	36,3	43,2	34,6	38,5
c/isolamento	28,2	35,1	33,2	35,6	40,2	37,4	34,2	42,1	30,5	38,4

1.11. Um produto foi desenvolvido com o objetivo de reduzir a média da temperatura do funcionamento de motores. Para testar o produto, foram selecionados ao acaso 8 motores e após 10 minutos de funcionamento, em cada condição, foram obtidos os dados (em ° C) do quadro abaixo. Testar a hipótese H_o e concluir, para $\alpha = 5\%$.

MOTOR	1	2	3	4	5	6	7	8
SEM PRODUTO	80,5	99,6	83,4	100,2	81,5	84,6	85,0	105,8
COM PRODUTO	75,8	98,8	77,6	99,9	74,2	80,5	83,6	105,8

1.12. Um experimentador deseja testar o efeito de certo fertilizante na média de produção de milho. Para realizar o experimento tinha-se 12 unidades experimentais de áreas iguais, onde 7 receberam o fertilizante e as outras não; sendo as outras condições mantidas iguais. As produções em kg/unidade experimental foram as seguintes:

Com Fertilizante	25	35	45	30	20	25	30
Sem Fertilizante	35	25	20	15	30		

De posse dos dados acima, pode o experimentador concluir que houve aumento da média de produção de milho por causa do fertilizante, com nível de significância igual a 5%.

1.13. Desejando comparar os efeitos de dois analgésicos A e B, em termos do tempo médio de ação sobre pacientes com certa doença (bastante prolongada), ambos foram aplicados a 14 doentes, em dias diferentes, sendo que 7 pacientes receberam primeiro o A, e outros 7 primeiro o B. A situação foi controlada de forma a não haver interferência do efeito de um sobre o outro. Os resultados (em minutos) foram:

Paciente	1	2	3	4	5	6	7	8	9	10	11	12	13	14
X_A	362	345	356	370	360	365	345	363	358	332	335	370	335	362
X_B	320	330	315	325	323	328	318	322	320	310	308	332	307	325

Testar a hipótese de diferença nula entre as médias populacionais, ao nível de significância igual a 1%.

1.14. Os dados abaixo se referem aos pesos, em gramas, de ratos machos da raça Wistar com 15 dias de idade, segundo a condição normal e submetidos à extirpação do timo (timectomização) aos 4 dias de idade. Verificar se a timectomização piora o ganho médio de peso destes animais, usando $\alpha = 5\%$.

Condição Normal	40,3	40,0	39,6	35,2	32,0
Timectomizado	18,6	20,3	23,6	22,2	20,9

1.15. Em determinada propriedade rural, foi avaliado o efeito de Suprimento Mineral (SM) na engorda de suínos. Para tanto, tomou-se 14 suínos similares em peso. Cada animal recebeu um dos SM. Os resultados obtidos, após certo período de tempo, foram os seguintes:

	Pesos (Kg)						
SM 1	38	33	36	35	36	32	30
SM 2	40	30	31	37	38	32	37

É possível afirmar ao nível de 1% de probabilidade que o SM 1 promove menor média de ganho de peso que o SM 2?

1.16. Determinada fábrica, interessada em ampliar o seu quadro de pessoal com indivíduos do sexo que apresentam menor variabilidade no tempo gasto para realizar a montagem de determinado equipamento eletrônico, realizou uma pesquisa. Os dados (em minutos) obtidos são fornecidos abaixo. Ao nível de 1% de probabilidade, pode-se concluir que indivíduos do sexo masculinos deveriam ser contratados porque apresentaram menor variabilidade no tempo gasto?

Masculino	4	8	3	9	7	5
Feminino	1	5	2	14	3	11

1.17. Um fazendeiro, visando otimizar os recursos de sua propriedade e aumentar a média de produção de leite, realizou uma pesquisa para verificar se o fornecimento da cama de galinha da sua granja poderia substituir, em parte, o fornecimento de ração ao seu gado. Para tanto, seguindo as recomendações de um zootecnista, selecionou um plantel de 10 animais e obteve os seguintes dados, em kg de leite por dia:

Ração com cama	45	47	49	48	46
Ração sem cama	38	37	35	39	37

De acordo com os resultados obtidos e ao nível de 5% de probabilidade, você recomendaria o uso de cama de galinha para substituir parte a ração?

1.18. Por meio dos dados amostrais fornecidos abaixo, é possível concluir que a média salarial de determinada empresa é inferior a R$ 950,00? (use o nível de 1% de significância)

Média	945
N^o de indivíduos avaliados	15
Variância	25

1.19. Dentre um rebanho de vacas reprodutoras, foram selecionadas ao acaso 10 animais. Dos animais selecionados, foram anotadas as produções médias diárias (kg/dia) durante o período de amamentação das crias 1 e 2. Pode-se afirmar que durante a amamentação da 2^a cria ocorre maior produção de leite? Use α = 5%

Cria	Produção de cada animal (Kg de leite/dia)									
1	15,6	16,3	19,5	14,5	16,2	20,2	14,6	13,1	16,2	17,1
2	18,3	16,3	17,2	19,8	18,5	19,1	18,3	16,5	19,5	19,8

1.20. Dois novos tipos de embalagens (A e B) foram testados para armazenar extrato de tomate. Uma boa embalagem mantém o pH do extrato de tomate em 7,2 até três meses após a sua armazenagem. Para comparar estes dois tipos de embalagens, 10 embalagens de cada um dos dois tipos testados, receberam a mesma quantidade de extrato de tomate e foram avaliados quanto ao seu pH três meses após a sua armazenagem. Os resultados das avaliações são apresentados a seguir

Embalagem A	6,8	7,0	7,1	7,0	7,1	7,3	7,4	7,5	7,4	7,4
Embalagem B	7,2	7,3	7,4	7,3	7,4	7,6	7,7	7,8	7,6	7,7

Admitindo-se que a variabilidade do pH em extratos armazenados nas embalagens A e B é a mesma, pede-se:

a. Pode-se concluir que existe diferença significativa entre as duas embalagens com relação a média do pH do extrato de tomate três meses após a sua armazenagem? Use o nível de 5% de probabilidade.

b. Baseado nos seus cálculos do item a, qual embalagem deveria ser recomendada? Justifique.

1.21. Em humanos é relativamente comum o hipotiriodismo, a qual é uma deficiência da glândula tireóide para produzir certos hormônios. Uma indústria farmacêutica, visando testar um novo tipo de droga, realizou uma pesquisa com 6 indivíduos portadores desta doença. Com tal finalidade, fez a avaliação da dosagem do hormônio H nos indivíduos portadores da doença antes e depois de serem medicados com a nova droga. Os resultados desta pesquisa são fornecidos a seguir

Indivíduo	1	2	3	4	5	6
Antes	100	110	98	105	108	105
Depois	140	135	125	145	135	140

Pode-se concluir que a nova droga é capaz de aumentar a dosagem média do hormônio H ao nível de 5% de significância?

1.22. Um fabricante de componentes eletrônicos elaborou um novo tipo de microprocessador. No entanto, é desejável que este novo microprocessador tenha velocidade média de processamento superior a 2,5 GHz. Para testar o novo microprocessador, o fabricante retirou ao acaso, uma amostra de 6 unidades, da qual obteve as seguintes informações:

Processador	1	2	3	4	5	6
Velocidade (GHz)	3,0	2,0	3,7	4,1	1,9	3,8

Média da velocidade de processamento dos 6 processadores amostrados = 3,08 GHz

Desvio padrão da velocidade de processamento dos 6 processadores amostrados = 0,95 GHz

Com base nas informações fornecidas, pergunta-se:

1.22.1 As hipóteses estatísticas para este problema são

a. $Ho: m = 2{,}5\text{ GHz}, \quad Ha: m < 2{,}5\text{ GHz}$

b. $Ho: m = 2{,}5\text{ GHz}, \quad Ha: m \neq 2{,}5\text{ GHz}$

c. $Ho: m = 2{,}5\text{ GHz}, \quad Ha: m > 2{,}5\text{ GHz}$

d. $Ho: \hat{m} = 2{,}5\text{ GHz}, \quad Ha: \hat{m} < 2{,}5\text{ GHz}$

e. $Ho: \hat{m} = 2{,}5\text{ GHz}, \quad Ha: \hat{m} \neq 2{,}5\text{ GHz}$

f. $Ho: m > 2{,}5\text{ GHz}, \quad Ha: m = 2{,}5\text{ GHz}$

g. nenhuma das anteriores

1.22.2 O valor da estatística t calculada para este problema, ao nível de 5% de probabilidade, leva a conclusão de que o novo microprocessador possui velocidade média de processamento

a. superior a 2,5 GHz

b. inferior a 2,5 GHz

c. igual a 2,5 GHz

d. nenhuma das anteriores

1.22.3 O valor da velocidade média amostral a partir do qual a hipótese H_0 é rejeitada é igual a

a. $\hat{m} = 1{,}72\text{ GHz}$

b. $\hat{m} = 1{,}50\text{ GHz}$

c. $\hat{m} = 3{,}28$ GHz
d. $\hat{m} = 3{,}50$ GHz
e. nenhuma das anteriores

1.23. Selecionaram-se aleatoriamente oito comprimidos diferentes de cada um de dois remédios antigripais concorrentes, Dozenol (D) e Niteze (N). Fez-se um teste do conteúdo de acetaminofena em cada um deles, obtendo-se os seguintes resultados (em mg):

Dozenol	472	487	506	512	489	503	511	501
Niteze	562	512	523	528	554	513	516	510

Ao nível de 5% de significância, teste a afirmação de que a quantidade média de acetaminofena é a mesma nas duas marcas.

1.24. Uma máquina foi regulada para fabricar placas de 5 mm de espessura, em média. Iniciada a produção, foi colhida uma amostra de tamanho 10, que forneceu as seguintes medidas de espessura, em mm:

5,1 4,8 5,0 4,7 4,8 5,0 4,5 4,9 4,8 5,2

Ao nível α = 0,01, pode-se aceitar a hipótese de que a regulagem da máquina foi satisfatória?

1.25. Um banho de óleo é aquecido aos poucos e sua temperatura medida de meia em meia-hora por dois termômetros. Tendo-se obtido os valores abaixo, há diferença entre as indicações dos dois termômetros, a α = 5%?

Termômetro 1:	38,2	44,5	53,0	59,0	66,4	71,3
Termômetro 2:	37,5	44,2	51,6	58,0	66,8	72,4

1.26. Um aparelho é utilizado para testar a durabilidade de lâmpadas, o qual consta de oito soquetes ligados em paralelo e de um reostato ligado em série com um gerador. Oito lâmpadas da marca A e oito lâmpadas da marca B foram ensaiadas nesse aparelho, sob as mesmas condições, fornecendo as seguintes durações, em horas:

Marca A:	35	26	40	35	31	49	38	24
Marca B:	23	28	31	35	36	30	27	26

Podemos concordar com a afirmação do fabricante da marca A, de que suas lâmpadas têm maior média de durabilidade que as da marca B (α = 1%).

1.27. Dois produtos A e B, foram avaliados quanto ao gosto, de acordo com as notas fornecidas por 10 indivíduos. Admitindo-se os valores 1 (péssimo), 2 (ruim), 3 (regular), 4 (bom) e 5 (ótimo) e um nível de significância de 5%, qual o melhor produto em termos da média da nota recebida?

Indivíduo:	1	2	3	4	5	6	7	8	9	10
Produto A:	3	3	4	4	4	4	4	5	5	5
Produto B:	3	3	3	3	2	2	2	1	1	1

1.28. Dois candidatos a um emprego, A e B, foram submetidos a um conjunto de oito questões, sendo anotados os minutos que cada um gastou na solução. Podemos, ao nível de 5% de significância, concluir que B seja mais rápido que A?

Questão	1	2	3	4	5	6	7	8
Indivíduo A	11	8	15	2	7	18	9	10
Indivíduo B	5	7	13	6	4	10	3	2

1.29. Numa competição de mercado de lâmpadas fluorescentes, duas marcas alegam para si o título de, em média, apresentar mais economia de energia. Para sanar esta dúvida, uma associação de consumidores resolve fazer uma bateria de testes com lâmpadas das duas marcas. O resultado do consumo em watts/hora desta bateria de testes é fornecido a seguir:

Marca	Consumo (watts/hora)				
A	69	72	73	72	70
B	89	92	93	92	90

Com base em um teste de hipótese, qual marca de lâmpada a associação de consumidores deveria recomendar? Utilize o nível de 5% de significância.

1.30. Suponha que um pesquisador da área de saúde deseja mostrar que os indivíduos portadores de febre amarela apresentam um teor de glicose inferior à média de 120 mg dos indivíduos não portadores. Para tanto, coletou uma amostra de sangue em sete indivíduos portadores de febre amarela e para cada um deles fez a avaliação do teor de glicose, em mg. Os resultados obtidos foram:

Indivíduo	1	2	3	4	5	6	7
Teor de glicose	119	122	120	110	112	115	116

Com base em um teste de hipóteses apropriado, qual deveria ser a conclusão do pesquisador? Utilize o nível de 5% de significância.

2. Contrastes

2.1. Introdução

O estudo de contrastes é muito importante na Estatística Experimental, principalmente quando o experimento em análise é composto por mais do que dois tratamentos. Com o uso de contrastes é possível ao pesquisador estabelecer comparações, entre tratamentos ou grupos de tratamentos, que sejam de interesse.

Este capítulo visa dar fundamentos para estabelecer grupos de contrastes, obter a estimativa para cada contraste estabelecido, bem com estimar a variabilidade associada a cada um destes contrastes. Todos os conhecimentos adquiridos neste capítulo serão utilizados no Capítulo 5 para se realizar testes de hipóteses para o grupo de contrastes estabelecidos.

2.2. Definições

Contraste

Considere a seguinte função linear de médias populacionais de tratamentos

$$C = a_1 m_1 + a_2 m_2 + ... + a_I m_I$$

C será um contraste entre médias se satisfizer a seguinte condição: $\sum_{i=1}^{I} a_i = 0$

Estimador do Contraste

Na prática, geralmente não se conhece os valores das médias populacionais m_i, mas suas estimativas. Daí, em Estatística Experimental, não se trabalhar com o contraste C mas com o seu estimador $\hat{C}$, que também é uma função linear de médias obtidas por meio de experimentos ou amostras. Assim tem-se que o estimador para o contraste de médias é dado por:

$$\hat{C} = a_1 \hat{m}_1 + a_2 \hat{m}_2 + ... + a_I \hat{m}_I$$

Exercício

2.1 Num experimento de consórcio na cultura do abacaxi, com 5 repetições, as médias de produção de frutos de abacaxi (em t/ha), foram as seguintes:

Tratamentos	$\hat{m}_i$
1 - Abacaxi (0,90 x 0,30m) monocultivo	53,5
2 - Abacaxi (0,80 x 0,30 m) monocultivo	56,5
3 - Abacaxi (0,80 x 0,30 m) + amendoim	62,0
4 - Abacaxi (0,80 x 0,30 m) + feijão	60,4

Pede-se obter as estimativas dos seguintes contrastes:

$$C_1 = m_1 + m_2 - m_3 - m_4$$

$$C_2 = m_1 - m_2$$

$$C_3 = m_3 - m_4$$

2.3. Medidas de dispersão associadas a contrastes

Considere o estimador do contraste C, dado por:

$$\hat{C} = a_1\hat{m}_1 + a_2\hat{m}_2 + ... + a_I\hat{m}_I$$

A variância do estimador do contraste é dada por:

$$V(\hat{C}) = V(a_1\hat{m}_1 + a_2\hat{m}_2 + ... + a_I\hat{m}_I)$$

Admitindo independência entre as médias

$$V(\hat{C}) = V(a_1\hat{m}_1) + V(a_2\hat{m}_2) + ... + V(a_I\hat{m}_I)$$

$$V(\hat{C}) = a_1^2 V(\hat{m}_1) + a_2^2 V(\hat{m}_2) + ... + a_I^2 V(\hat{m}_I)$$

Sabe-se que: $V(\hat{m}_i) = \frac{\sigma_i^2}{r_i}$, assim

$$V(\hat{C}) = a_1^2\frac{\sigma_1^2}{r_1} + a_2^2\frac{\sigma_2^2}{r_2} + ... + a_I^2\frac{\sigma_I^2}{r_I}$$

Admitindo-se homogeneidade de variâncias, ou seja, $\sigma_1^2 = \sigma_2^2 = ... = \sigma_n^2 = \sigma^2$, então

$$V(\hat{C}) = \left(\frac{a_1^2}{r_1} + \frac{a_2^2}{r_2} + ... + \frac{a_I^2}{r_I}\right)\sigma^2 = \sigma^2\sum_{i=1}^{I}\frac{a_i^2}{r_i}$$

Na prática, geralmente, não se conhece a variância σ^2, mas sua estimativa a qual obtida por meio de dados experimentais. Esta estimativa é denominada como estimador comum (s_c^2). Então o que normalmente se obtém é o valor do estimador da variância do estimador do contraste, a qual é obtida por

$$\hat{V}(\hat{C}) = s_c^2\sum_{i=1}^{I}\frac{a_i^2}{r_i}$$

Exercício

2.2 Por meio dos dados e dos contrastes fornecidos abaixo, obter as estimativas dos contrastes e as estimativas das variâncias das estimativas dos contrastes.

$$\hat{m}_1 = 11{,}2 \qquad \hat{m}_2 = 10{,}5 \qquad \hat{m}_3 = 10{,}0 \qquad \hat{m}_4 = 21{,}0$$

$$r_1 = r_2 = 6 \quad r_3 = 4 \quad r_4 = 5 \qquad s_c^2 = 0{,}45$$

$C_1 = m_1 + m_2 - m_3 - m_4$

$C_2 = m_1 - m_2$

$C_3 = m_3 - m_4$

2.4. Contrastes Ortogonais

Em algumas situações desejamos testar um grupo de contrastes relacionados com o experimento em estudo. Alguns tipos de testes indicados para este objetivo, necessitam que os contrastes, que compõem o grupo a ser testado, sejam ortogonais entre si.

A ortogonalidade entre os contrastes indica independência linear na comparação estabelecida por um contraste com a comparação estabelecida pelos outros contrastes.

Sejam os estimadores dos contrastes de C_1 e C_2 dados, respectivamente, por:

$$\hat{C}_1 = a_1\hat{m}_1 + a_2\hat{m}_2 + ... + a_I\hat{m}_I$$

$$\hat{C}_2 = b_1\hat{m}_1 + b_2\hat{m}_2 + ... + b_I\hat{m}_I$$

A covariância entre $\hat{C}_1$ e $\hat{C}_2$, supondo independência entre tratamentos, é obtida por

$$Cov\left(\hat{C}_1,\hat{C}_2\right) = a_1b_1V\left(\hat{m}_1\right) + a_2b_2V\left(\hat{m}_2\right) + ... + a_Ib_IV\left(\hat{m}_I\right)$$

A variância da média amostral é dada por: $V\left(\hat{m}_i\right) = \dfrac{\sigma_i^2}{r_i}$, para i = 1, 2, ..., I. Logo,

$$Cov\left(\hat{C}_1,\hat{C}_2\right) = a_1b_1\frac{\sigma_1^2}{r_1} + a_2b_2\frac{\sigma_2^2}{r_2} + ... + a_Ib_I\frac{\sigma_I^2}{r_I}$$

Admitindo que exista homogeneidade de variâncias entre os tratamentos, ou seja: $\sigma_1^2 = \sigma_2^2 = ... = \sigma_I^2 = \sigma^2$, então.

$$Cov\left(\hat{C}_1,\hat{C}_2\right) = \left(\frac{a_1b_1}{r_1} + \frac{a_2b_2}{r_2} + ... + \frac{a_Ib_I}{r_I}\right)\sigma^2 = \sigma^2\sum_{i=1}^{I}\frac{a_ib_i}{r_i}$$

Sabe-se que, se duas variáveis aleatórias são independentes, a covariância entre elas é igual a zero. Assim, se $\hat{C}_1$ e $\hat{C}_2$ são independentes, a covariância entre eles é igual a zero, isto é:

$$Cov\left(\hat{C}_1,\hat{C}_2\right) = 0$$

Para que a covariância seja nula, é necessário, portanto que:

$$\sum_{i=1}^{I}\frac{a_ib_i}{r_i} = 0.$$

Esta é a condição de ortogonalidade entre dois contrastes para um experimento com número diferente de repetições para os tratamentos. Para um experimento com o mesmo número de repetições, satisfazendo as mesmas pressuposições (médias independentes e homogeneidade de variâncias), a condição de ortogonalidade se resume a:

$$\sum_{i=1}^{I}a_ib_i = 0$$

Para um experimento com I tratamentos, podem ser formados vários grupos de contrastes ortogonais, no entanto cada grupo deverá conter no máximo (I-1) contrastes ortogonais, o que corresponde ao número de graus de liberdade para tratamentos.

Dentro de um grupo de contrastes ortogonais, todos os contrastes tomados dois a dois, serão também ortogonais.

Exercícios

2.3. Verificar se os contrastes do Exercício 2.1 formam um grupo de contrastes ortogonais.

2.4. Verificar se os contrastes do Exercício 2.2 formam um grupo de contrastes ortogonais.

2.5. Métodos para obtenção de grupos de contrastes mutuamente ortogonais

Obtenção por Meio de Sistema de Equações Lineares

Neste método, deve-se estabelecer, a princípio, um contraste que seja de interesse e, a partir deste é que os demais são obtidos. Por meio da imposição da condição de ortogonalidade e da condição para ser um contraste, obtém-se equações lineares, cujas incógnitas são os coeficientes das médias que compõem o contraste. Como o número de incógnitas é superior ao número de equações existentes, será sempre necessário atribuir valores a algumas incógnitas. É desejável que os valores a serem atribuídos, permitam que os coeficientes sejam números inteiros.

Exercício

2.5. Foi instalado para avaliar a produção de 4 híbridos cujas características são apresentadas na tabela a seguir.

Hibrido	1	2	3	4
Porte	Alto	Alto	Alto	Baixo
Inicio do Florescimento	Precoce	Tardio	Tardio	Precoce
Índice de acamamento	Médio	Alto	Baixo	Médio
r_i	3	3	3	3

Suponha que ao estabelecer as comparações dos híbridos com relação a produção, seja levado em consideração

- o porte;
- o início do florescimento;
- o índice de acamamento.

Obtenha um grupo de contrastes ortogonais que permita testar as comparações segundo os critérios citados.

Obtenção por Meio de Regras Práticas

Por meio desta metodologia, é possível estabelecer facilmente um grupo de contrastes ortogonais. A metodologia pode ser resumida nos seguintes passos (BANZATTO e KRONKA, 1989):

Divide-se o conjunto das médias de todos os tratamentos do experimento em dois grupos. O primeiro contraste é obtido pela comparação das médias de um grupo contra as médias do outro grupo. Para isso atribui-se sinais positivos para membros de um grupo e negativos para membros do outro grupo.

Dentro de cada grupo formado no passo anterior, que possui mais que uma média, aplica-se o passo 1, subdividindo-os em subgrupos. Repete-se este passo até que se forme subgrupos com apenas uma média. Ao final, deveremos ter formado (I-1) comparações.

Para se obter os coeficientes que multiplicam cada média que compõem os contrastes estabelecidos, deve-se, para cada contraste:

Verificar o número de parcelas experimentais envolvidas no 1º grupo, digamos g_1, e o número de parcelas experimentais envolvidas no 2º grupo, digamos g_2. Calcula-se o mínimo múltiplo comum (m.m.c.) entre g_1 e g_2.

Dividir o m.m.c. por g_1. O resultado será o coeficiente de cada média do 1º grupo.

Dividir o m.m.c. por g_2. O resultado será o coeficiente de cada média do 2º grupo.

Multiplicar os coeficientes obtidos pelo número de repetições da respectiva média. Se possível, simplificar os coeficientes obtidos por uma constante. No caso em que o número de repetições é igual para todos os tratamentos, este passo pode ser eliminado.

Exercício

2.6. Num experimento inteiramente casualizado, com 4 repetições, foram comparados os efeitos de 5 tratamentos em relação ao crescimento de mudas de *Pinus oocarpa*, 60 dias após a semeadura. Os tratamentos utilizados e os resultados obtidos foram (BANZATTO e KRONKA, 1989):

Tratamentos	Totais
1 – Solo de cerrado (SC)	21,0
2 – Solo de cerrado + esterco (SC+E)	27,1
3 – Solo de cerrado + esterco + NPK (SC+E+NPK)	26,6
4 – Solo de cerrado + vermiculita (SC+V)	22,1
5 – Solo de cerrado + vermiculita + NPK (SC+V+NPK)	25,6

Obtenha um grupo de contrastes ortogonais entre as médias.

2.7. Suponha agora para o exemplo 1 que os tratamentos 1 e 4 tenham 3 repetições e os tratamentos 2, 3 e 5 tenham 4 repetições. Obtenha um grupo de contrastes ortogonais entre médias.

2.6. Exercícios Suplementares

2.8. Dados

Tratamentos	$\hat{m}_i$	r_i
1	25,0	5
2	18,7	5
3	30,4	5
4	27,5	6

e os contrastes

$$C_1 = m_1 - m_2$$
$$C_2 = m_1 + m_2 - 2m_3$$
$$C_3 = m_1 + m_2 + m_3 - 3m_4$$

Admitindo-se que os estimadores das médias sejam independentes e que $s_c^2 = 0{,}45$, pede-se

a) $\hat{C}_1$, $\hat{C}_2$ e $\hat{C}_3$

b) $\hat{V}(\hat{C}_1)$, $\hat{V}(\hat{C}_2)$, e $\hat{V}(\hat{C}_3)$

c) as estimativas das covariâncias entre os estimadores dos contrastes, e por meio das mesmas, dizer quais são os contrastes ortogonais entre si.

2.9. Supondo independência entre médias, homogeneidade de variâncias entre tratamentos e admitindo que m_1, m_2 e m_3 têm, respectivamente, 5, 3 e 6 repetições, verificar se os contrastes dados abaixo são ortogonais.

$$C_1 = m_1 - m_2$$
$$C_2 = m_1 + m_2 - 2m_3$$

.

2.10. Considere um experimento com 4 tratamentos e as seguintes informações:

$$s_c^2 = 4{,}10$$
$$r_1 = r_2 = r_3 = 4; \quad r_4 = 3$$
$$C_1 = m_1 + m_2 + m_3 - 3m_4$$
$$C_2 = m_1 - 2m_2 + m_3$$

Pede-se:

a) Forme um grupo de contrastes ortogonais, a partir dos contrastes C_1 e C_2, por meio do método do sistema de equações lineares.

b) Obtenha $\hat{V}(\hat{C}_1)$

c) Obtenha $V(C_1)$

2.11. Num experimento com 4 tratamentos e 5 repetições, são dados os seguintes contrastes ortogonais:

$C_1 = m_2 - m_4$

$C_2 = -2m_1 + m_2 + m_4$

Determinar um contraste C_3 que seja ortogonal a C_1 e C_2.

2.12. Com os dados abaixo, obter o contraste C_3 ortogonal aos contrastes C_1 e C_2.

$$C_1 = m_1 - m_2 \qquad r_1 = r_3 = 4$$
$$C_2 = 4m_1 + 5m_2 - 9m_4 \qquad r_2 = r_4 = 5$$

2.13. Dado o contraste $C_1 = 2m_1 - m_2 - m_3$, referente a um experimento com 3 tratamentos ($r_1 = r_2 = r_3 = 5$), obter um contraste ortogonal C_2 em relação a C_1.

2.14. Dado o contraste $C_1 = 2m_1 - m_2 - m_3$, referente a um experimento com 3 tratamentos ($r_1 = r_2 = 4$ e $r_3 = 5$), obter um contraste ortogonal C_2 em relação a C_1

2.15. Dado o contraste $C_1 = 9m_1 - 4m_2 - 5m_3$, referente a um experimento com 3 tratamentos ($r_1 = r_2 = 4$ e $r_3 = 5$), obter um contraste ortogonal C_2 em relação a C_1.

2.16. Dados os contrastes $C_1 = m_2 - m_4$ e $Y_2 = -2m_1 + m_2 + m_4$, referente a um experimento com 4 tratamentos ($r_1 = r_2 = r_3 = r_4 = 5$), obter um contraste ortogonal C_3 em relação a C_1 e C_2.

2.17. Dados os contrastes $C_1 = m_1 + m_2 + m_3 - 3m_4$ e $C_2 = m_1 - 2m_2 + m_3$, referente a um experimento com 4 tratamentos ($r_1 = r_2 = r_3 = 4$ e $r_4 = 3$), obter um contraste ortogonal C_3 em relação a C_1 e C_2.

2.18. Dados os contrastes $C_1 = m_1 - 4m_2 + m_3 + 2m_4$ e $C_2 = m_1 - m_3$, referente a um experimento com 4 tratamentos ($r_1 = r_3 = 6$, $r_2 = 4$ e $r_4 = 5$), obter um contraste ortogonal C_3 em relação a C_1 e C_2.

2.19. Para verificar o efeito de três tipos de adoçantes no teor de glicose no sangue, foi realizada uma pesquisa em que se ministrou cada um destes tipos de adoçantes a um determinado grupo de cobaias, por certo período de tempo. Ao final deste período, o teor médio de glicose ($\hat{m}_i$) no sangue foi avaliado para cada grupo, obtendo-se os seguintes resultados:

Adoçante	Nº de Cobaias	$\hat{m}_i$	s^2
1-Químico	8	115	30
2- Químico	10	90	30
3- Natural	5	75	30

A partir dos dados fornecidos acima, pede-se:

2.19.1 Desejando-se testar o teor médio de glicose do conjunto de cobaias que recebeu adoçante químico contra o grupo que recebeu adoçante natural, qual seria o contraste apropriado? Qual o valor da estimativa deste contraste?

2.19.2 Suponha que seja de interesse testar a seguinte comparação: $C_1 = m_2 - m_3$, no entanto, desejamos testar outros contrastes que sejam ortogonais a C_1. Obtenha o (s) outro (s) contraste (s) ortogonal (is) necessário (s) para completar o grupo de contrastes ortogonais a C_1.

2.20. Num experimento, 4 novos tipos de herbicida foram comparados para verificar se são eficazes para combater ervas daninhas e assim manter a produção de milho em níveis elevados. Um resumo do experimento é dado a seguir

Herbicida	Média de produção (kg/ha)	Repetições
1 – Biológico	46	4
2 – Químico à base de nitrogênio e enxofre	31	4
3 – Químico à base de nitrogênio e fósforo	32	4
4 – Químico à base de inativadores enzimáticos	25	4

Suponha que seja de interesse testar o seguinte contraste entre as médias de tratamentos $C_1 = 3m_1 - m_2 - m_3 - m_4$. Suponha ainda que todos os tratamentos possuam uma mesma variância e que sua estimativa é igual a 35 $(kg/ha)^2$. Pergunta-se:

a) Qual a comparação que está sendo feita pelo contraste C_1? Qual a estimativa para este contraste?

b) Por meio da estimativa obtida para o contraste C_1 pode-se AFIRMAR que exista um grupo melhor de herbicidas do que outro? Justifique a sua resposta.

c) Qual a estimativa da variância para a estimativa do contraste C_1?

d) Forme um grupo de contrastes ortogonais a partir do contraste C_1. Descreva qual comparação que está sendo feita por cada contraste que você obteve. Baseando-se nos dados amostrais fornecidos, obtenha também a estimativa para cada um dos contrastes.

2.21. Considere um experimento, onde foi avaliada a variável produção (kg/parcela) de quatro tratamentos (adubações), denominados como: T_1 = Sulfato de Amônio, T_2 = Sulfato de Amônio + Enxofre, T_3 = Nitrocálcio e T_4 = Nitrocálcio + Enxofre. Os resultados obtidos foram:

Tratamentos	$\hat{m}_i$	r_i
1 – Sulfato de Amônio	24,0	4
2 – Sulfato de Amônio + Enxofre	28,0	5
3 – Nitrocálcio	27,0	4
4 – Nitrocálcio + Enxofre	25,0	5

$s_c^2 = 0{,}75$

a) Estabelecer as seguintes comparações de interesse (as comparações solicitadas, não são necessariamente ortogonais):

i) Sulfato de Amônio versus Nitrocálcio na ausência de Enxofre
ii) Sulfato de Amônio versus Sulfato de Amônio + Enxofre
iii) Nitrocálcio versus Nitrocálcio + Enxofre

b) Sendo dados, com base em outros critérios, os seguintes contrastes:

$C_1 = m_1 - m_2$
$C_2 = 4m_1 + 5m_2 + 4m_3 - 13m_4$
Pede-se:
i) Obter a estimativa do contraste C_2.
ii) Obter a estimativa da variância da estimativa do contraste C_2.
iii) Obter a variância do contraste C.
iv) Os contrastes C_1 e C_2 são ortogonais? Justifique a sua resposta.

3. Introdução à Experimentação

3.1. Introdução

A experimentação tem por objetivo o estudo dos experimentos, isto é, seu planejamento, execução, análise dos dados obtidos e interpretação dos resultados.

3.2. Alguns Conceitos Básicos

a. Tratamento ou fator: é o método, elemento ou material cujo efeito desejamos medir ou comparar em um experimento. Exemplos: a) variedades de milho; b) níveis de proteína na ração e c) diferentes temperaturas de pasteurização do leite.

b. Unidade experimental: é a unidade que vai receber o tratamento e fornecer os dados que deverão refletir o seu efeito. Exemplos: a) uma fileira de plantas com 3 metros de comprimento no campo; b) um leitão e c) um litro de leite.

c. Delineamento experimental: é a maneira como os tratamentos são designados às unidades experimentais. Exemplos: Delineamento Inteiramente Casualizado (Capítulo 4), Delineamento em Blocos Casualizados (Capítulo 6) e Delineamento em Quadrado Latino (Capítulo 7).

d. Esquema: quando em um mesmo experimento são avaliados dois ou mais fatores os níveis dos fatores podem ser combinados de maneiras diferentes. O esquema é justamente a maneira utilizada pelo pesquisador ao combinar os níveis dos fatores para se obter os tratamentos. Exemplos: Esquema Fatorial (Capítulo 8) e Esquema em Parcelas subdivididas (Capítulo 9).

e. Variável resposta: é a variável mensurada usada para avaliar o efeito de tratamentos.

f. Erro experimental: é o efeito de fatores que atuam de forma aleatória e que não são passíveis de controle pelo experimentador.

A pesquisa científica está constantemente se utilizando de experimentos para provar suas hipóteses. É claro que o procedimento para realizar um experimento varia de acordo com a área para a qual está se fazendo uma pesquisa. Porém, todo experimento deve seguir alguns princípios básicos, para que as conclusões sejam válidas.

3.3. Princípios Básicos da Experimentação

São três os princípios básicos da experimentação: repetição, casualização e controle local.

Princípio da Repetição

A repetição consiste em aplicar o mesmo tratamento a várias unidades experimentais, ou seja, consiste na reprodução do experimento básico. Não existe uma regra dizendo qual deve ser o número mínimo de repetições. Isto depende do conhecimento do pesquisador sobre o assunto e do conjunto de condições em que será realizado o experimento. Como regra prática, sugere-se que os experimentos tenham pelo menos 20 unidades experimentais e 10 graus de liberdade para o resíduo. Quanto maior é o número de repetições, espera-se que seja maior a precisão do experimento.

Em termos estatísticos, o uso do princípio da repetição tem por finalidade obter uma estimativa do erro experimental.

Princípio da Casualização

O princípio da casualização consiste em distribuir ao acaso os tratamentos às unidades experimentais. Este princípio tem por finalidade propiciar, a todos os tratamentos, a mesma chance de serem designados a qualquer uma das unidades experimentais, visando evitar que algum dos tratamentos seja sistematicamente favorecido ou desfavorecido por fatores fora de controle do pesquisador. Sendo assim com o uso do princípio da casualização, as variações que contribuem para o erro experimental são convertidas em variáveis aleatórias.

Do ponto de vista estatístico, com o uso do princípio da casualização em um experimento:

a. obtém-se uma estimativa válida do erro experimental;

b. fica garantido o uso de testes de significância, pois os erros experimentais atuam de forma independente nas diversas unidades experimentais.

Todo experimento deve conter no mínimo os princípios básicos da repetição e da casualização.

Princípio do Controle na Casualização

O uso do princípio do controle na casualização só é recomendado quando as unidades experimentais não são ou não estão sob condições homogêneas devido a influência de um ou mais fatores. Para utilizar este princípio, é necessário inicialmente dividir as unidades experimentais em blocos de unidades de tal forma que dentro de cada bloco haja homogeneidade e um número de unidades igual ao número de tratamentos do experimento. A distribuição dos tratamentos as unidades é feita então dentro de cada bloco. Daí o nome do princípio *controle* na casualização.

A finalidade, do uso do princípio do controle na casualização, é reduzir o efeito do erro experimental através do controle da variação existente entre as unidades experimentais. Espera-se que com o controle na casualização a estimativa obtida para o erro experimental seja menor.

3.4. Fontes de variação de um experimento

Em um experimento podem ocorrer as seguintes fontes de variação:

Premeditada

É aquela introduzida pelo pesquisador com a finalidade de fazer comparações. Por exemplo: tratamentos.

Sistemática

Variações não intencionais, mas de natureza conhecida. Variação inerente ao material experimental. Podem ser controladas pelo pesquisador. Por exemplo: heterogeneidade do solo, tamanho de semente, etc.

Aleatória

São variações de origem desconhecida, não podendo ser controladas. Constituem o erro experimental. São devidas a duas fontes: variações no material experimental e falta de uniformidade nas condições experimentais.

3.5. Exercícios

3.1. Um experimento deve conter no mínimo o(s) seguinte(s) princípio(s) básico(s) da experimentação:

a) repetição
b) casualização
c) controle local
d) repetição e controle local
e) repetição e casualização
f) casualização e controle local
g) nenhuma das respostas anteriores

3.2. A repetição tem a função de:

a) fornecer uma estimativa do erro experimental
b) validar a estimativa do erro experimental
c) controlar a heterogeneidade das unidades experimentais
d) nenhuma das anteriores

3.3. A casualização tem a função de:

a) fornecer uma estimativa do erro experimental
b) validar a estimativa do erro experimental
c) controlar a heterogeneidade das unidades experimentais
d) nenhuma das anteriores

3.4. Um extensionista, desejando comparar 10 rações para ganho de peso em animais, procedeu da seguinte forma:

- tomou 10 animais de uma propriedade rural. Estes 10 animais visivelmente não eram homogêneos entre si, porque foram oriundos de diferentes cruzamentos raciais e apresentavam idades diferentes.
- as rações que o extensionista julgou ser as melhores foram designadas aos melhores animais, e as rações que o extensionista julgou ser as piores foram designadas aos piores animais, de tal forma que cada animal recebeu uma única ração.
- ao final de sua pesquisa, o extensionista recomendou a ração que proporcionou maior ganho de peso nos animais.

Baseado nestas informações, pergunta-se:

3.4.1 Quantos e quais foram os tratamentos em teste nesta pesquisa? Justifique sua resposta.
3.4.2 Qual foi a constituição de cada unidade experimental nesta pesquisa? Justifique sua resposta.
3.4.3 Qual(is) foi(ram) o(s) princípio(s) básico(s) da experimentação utilizados nesta pesquisa? Justifique a sua resposta.
3.4.4 É possível estimar o erro experimental nesta pesquisa? Justifique sua resposta.
3.4.5 A conclusão dada pelo extensionista ao final da pesquisa, é estatisticamente aceitável? Justifique a sua resposta.

3.5. Um bioquímico desejando verificar qual entre 5 enzimas (identificadas como E1, E2, E3, E4 e E5) produz maiores fragmentos de DNA de células epiteliais de cobaias, realizou o seguinte ensaio:

- selecionou um conjunto de 15 cobaias (sistematicamente identificadas como 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 e 15) que eram supostamente homogêneas para as características essenciais;
- de cada uma das 15 cobaias, tomou uma amostra de tecido epitelial de cada um dos seguintes membros: superior, mediano e inferior. Procedeu posteriormente a uma mistura das amostras coletadas dos três membros, denominada de amostra composta;
- cada amostra composta foi convenientemente tratada para a extração do DNA. A amostra obtida contendo apenas o DNA foi denominada amostra genômica. As amostras genômicas foram identificadas de acordo com o número da cobaia que a originou, ou seja, a amostra genômica identificada como C1, conteve DNA extraído da cobaia 1; a amostra genômica identificada como C2, conteve DNA extraído da cobaia 2; e assim por diante. Ao final obteve-se as amostras genômicas C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C12, C13, C14 e C15;
- cada uma das amostras genômicas foi tratada com um tipo de enzima. A distribuição das enzimas às amostras foi feita da seguinte forma sistemática: E1 foi destinada às amostras genômicas C1, C2 e C3; E2 foi destinada às amostras genômicas C4, C5 e C6; E3 foi destinada às amostras genômicas C7, C8 e C9; E4 foi destinada às amostras genômicas C10, C11 e C12; e E5 foi destinada às amostras genômicas C13, C14 e C15;
- uma amostra de 1 ml de cada substrato químico dos fragmentos de DNA foi colocado para correr em um gel. O tempo, em minutos, gasto por cada uma das 15 amostras para percorrer a distância de 25 cm foi registrado para comparar o efeito das enzimas E1, E2, E3, E4 e E5.

Com base nas informações fornecidas deste ensaio e das explicações fornecidas em sala de aula, pergunta-se:

3.5.1 Quais foram os tratamentos em teste neste experimento? Justifique a sua resposta.
3.5.2 Neste experimento os tratamentos surgiram de uma forma aleatória, premeditada ou sistemática? Justifique a sua resposta.
3.5.3 Qual foi a unidade experimental nesta pesquisa? Justifique a sua resposta.
3.5.4 O princípio da repetição foi utilizado nesta pesquisa? Justifique a sua resposta. Em caso afirmativo, explique porque diferentes observações obtidas para um mesmo tratamento não são iguais. Em caso negativo, faça uma análise crítica quanto à necessidade do uso de repetições num experimento.
3.5.5 O princípio da casualização foi utilizado nesta pesquisa? Justifique a sua resposta.
3.5.6 O princípio do controle local foi utilizado nesta pesquisa? Justifique a sua resposta. Em termos gerais, quando que o princípio do controle local deve ser utilizado em um experimento?
3.5.7 É possível estimar o erro experimental nesta pesquisa? Justifique a sua resposta. Em caso afirmativo, a estimativa do erro experimental é válida? Justifique a sua resposta. Em caso negativo, indique o que deveria ser feito de diferente neste ensaio para ser possível estimar o erro experimental. Justifique a sua resposta.
3.5.8 Neste ensaio, qual foi a variável resposta utilizada para comparar os efeitos de tratamentos? Justifique a sua resposta.

3.6. Um pesquisador desejava comparar os efeitos que 8 tipos de óleo têm sobre o teor de gordura total em preparos de maionese. Com esta finalidade, esse pesquisador procedeu da seguinte forma:

- para a avaliação do teor de gordura total, o pesquisador tinha à sua disposição 8 bioquímicos. Devido à falta de experiência dos bioquímicos, o pesquisador temia que a medição dos mesmos pudesse interferir na comparação dos tipos de óleo. Visando controlar esta fonte de variação, o pesquisador decidiu que cada um dos 8 bioquímicos deveria fazer a medição do teor de gordura dos preparos de maionese produzidos utilizando os 8 tipos de óleo;
- baseado em experimentos anteriores, o pesquisador sabia que, apesar do controle de qualidade, havia variação entre os lotes de substrato de preparos de maionese. O substrato de preparo da maionese é o composto que tem todos os ingredientes do preparo da maionese, exceto o óleo. Como um lote de substrato não seria suficiente para testar os 8 tipos de óleo em todas as repetições desejadas, o pesquisador decidiu que prepararia 8 lotes de substrato e dividiria cada lote em 8 partes iguais. Cada uma das 64 partes, assim obtidas, seria denominada de amostra básica;
- foi então realizada uma distribuição ao acaso dos 8 tipos de óleo às amostras básicas, tendo as seguintes restrições na casualização:

1ª) cada tipo de óleo deveria ser aplicado em uma única amostra básica de cada um dos 8 lotes de substrato.

2ª) os 8 tipos de preparo de maionese obtidos misturando cada uma das amostras básicas com cada um dos 8 tipos de óleo, deveriam ser avaliadas por cada um dos 8 bioquímicos;

No local que foi conduzido o experimento, o pesquisador constatou que, após certo tempo do experimento ter sido instalado, houve uma pequena contaminação por fungo em algumas unidades experimentais. O pesquisador, usando do seu conhecimento técnico na área, julgou que a contaminação não comprometeria os resultados obtidos no experimento.

Baseando-se nestas informações, responda com objetividade e clareza, as seguintes perguntas:

3.6.1 Quais foram os tratamentos em teste? Justifique a sua resposta.
3.6.2 Como você classificaria a fonte de variação contaminação por fungo, observada nesse experimento? Justifique a sua resposta.
3.6.3 Qual foi a unidade experimental utilizada nesta pesquisa? Justifique a sua resposta.
3.6.4 O princípio da repetição foi utilizado nesta pesquisa? Se sua resposta for afirmativa, responda qual foi o número de repetições utilizado. Se a sua resposta for negativa, responda se o procedimento do pesquisador está correto.
3.6.5 O princípio da casualização foi utilizado nesta pesquisa? Justifique a sua resposta.
3.6.6 O princípio do controle local foi utilizado nesta pesquisa? Se a sua resposta for afirmativa, explique como este princípio foi utilizado. Se a sua resposta for negativa, explique por que não houve a necessidade da utilização deste princípio.
3.6.7 Qual foi a característica utilizada pelo pesquisador para avaliar o efeito de tratamentos neste experimento. Justifique a sua resposta.

3.7. Um fabricante de móveis realizou um experimento para verificar qual dentre cinco marcas de verniz proporciona maior brilho. Com esta finalidade, procedeu da seguinte forma:

- Em sua fábrica identificou amostras de madeira que estariam disponíveis para a realização deste experimento. Verificou que possuía cinco tábuas de Jatobá, cinco tábuas de Cerejeira, cinco tábuas de Mogno, cinco tábuas de Goiabão e cinco tábuas de Castanheira. Constatou também que as cinco tábuas de cada tipo de madeira eram homogêneas para as características essenciais e que havia uma grande variedade de cores entre os cinco tipos de madeira (Jatobá, Cerejeira, Mogno, Goiabão e Castanheira). Sabe-se que a cor da madeira pode influenciar muito o brilho da mesma quando envernizada;
- Resolveu então distribuir ao acaso as cinco marcas de verniz às tábuas de madeira, de tal forma que cada tipo de madeira fosse testada com todas as marcas de verniz;
- O brilho foi medido por meio de um aparelho que mede a refletância da luz branca projetado sobre a tábua de madeira envernizada;

Baseado nas informações deste experimento, pergunta-se:

3.7.1. Qual foi a unidade experimental utilizada neste experimento? Justifique a sua resposta.
3.7.2. Quais foram os tratamentos comparados neste experimento? Justifique a sua resposta.
3.7.3. Quais foram os princípios básicos da experimentação utilizados neste experimento? Justifique a sua resposta.
3.7.4. É possível estimar o erro experimental neste experimento? Justifique a sua resposta. Se a resposta for afirmativa, a estimativa do erro é válida? Justifique. Se a resposta foi negativa, explique o que deveria ser feito para obter uma estimativa válida para o erro experimental.
3.7.5. O que faz surgir o erro num experimento? É possível eliminar totalmente o efeito do erro experimental em um experimento? Justifique a sua resposta.
3.7.6. O procedimento adotado pelo pesquisador de distribuir as marcas de verniz ao acaso dentro de cada tipo de madeira foi realmente necessário? Justifique a sua resposta.

3.8. Um pesquisador de uma indústria de alimentos desejava verificar se seis sabores de sorvete apresentavam o mesmo o teor de glicose. O pesquisador, baseado em experimentos anteriores, sabia que duas outras fontes de variação indesejáveis poderiam influenciar o valor mensurado do teor de glicose: o tipo de recipiente utilizado para armazenagem do sorvete e o equipamento utilizado para mensuração do teor de glicose. Para controlar estas duas fontes de variação o pesquisador decidiu que cada sabor deveria ser avaliado em cada um dos seis equipamentos disponíveis; e armazenado em cada um dos seis tipos de recipientes disponíveis. Com esta finalidade, o pesquisador planejou o experimento da seguinte maneira:

- preparar 6 lotes de 100 ml de cada sabor. O total de lotes a serem preparados seria de 36 lotes;
- os lotes de sorvetes deveriam ser distribuídos ao acaso aos recipientes, com a restrição de que cada tipo de recipiente recebesse todos os 6 sabores uma única vez;

- os lotes de sorvetes seriam designados ao acaso aos equipamentos para a análise do teor de glicose, com a restrição de que cada equipamento avaliasse cada um dos seis sabores uma única vez.

Baseando-se nestas informações, pergunta-se:

3.8.1. Quais foram os tratamentos em teste neste experimento? Justifique a sua resposta.
3.8.2. O princípio da repetição foi utilizado neste experimento? Justifique a sua resposta.
3.8.3. O princípio do controle local foi utilizado neste experimento? Justifique a sua resposta. Se a resposta for afirmativa, quantas vezes o mesmo foi utilizado? Se a resposta for negativa, discuta sobre a necessidade do mesmo ser utilizado neste experimento.

4. Delineamento Inteiramente Casualizado

4.1. Introdução

No Delineamento Inteiramente Casualizado (DIC) a distribuição dos tratamentos às unidades experimentais é feita inteiramente ao acaso. Os outros delineamentos experimentais, por exemplo: blocos casualizados e quadrado latino, se originam do DIC pelo uso de restrição na casualização. O DIC utiliza apenas os princípios básicos da repetição e da casualização.

Como não faz restrições na casualização, o uso do DIC pressupõe que as unidades experimentais estão sob condições homogêneas. Estas condições homogêneas geralmente são obtidas em locais com ambientes controlados tais como laboratórios, estufas e casas de vegetação.

4.2. Quadro de tabulação dos dados

A título de exemplo, considere um experimento instalado no DIC com I tratamentos e J repetições. A coleta de dados da pesquisa pode ser resumida, num quadro do tipo a seguir:

	Tratamentos			
Repetições	1	2	...	I
1	Y_{11}	Y_{21}	...	Y_{I1}
2	Y_{12}	Y_{22}	...	Y_{I2}
...	...	...	...	...
J	Y_{1J}	Y_{2J}	...	Y_{IJ}
Totais	T_1	T_2	...	T_I

Deste quadro pode-se retirar algumas informações de interesse:

- nº de unidades experimentais: N = I x J
- Total geral: $G = \sum_{i=1,j=1}^{I,J} Y_{ij} = \sum_{i=1}^{I} T_i = Y_{\bullet\bullet}$
- Total para o tratamento i: $T_i = \sum_{j=1}^{J} Y_{ij} = Y_{i\bullet}$
- Média para o tratamento i: $\hat{m}_i = \frac{T_i}{J}$
- Média geral do experimento: $\hat{m} = \frac{G}{IJ}$.

4.3. Modelo estatístico

Existe um modelo estatístico específico para cada tipo de delineamento. O modelo estatístico identifica quais são as fontes de variação dos valores de uma variável resposta em estudo.

Para os dados oriundos de um experimento instalado segundo o DIC, o seguinte modelo estatístico deve ser utilizado nas análises estatísticas:

$$Y_{ij} = m + t_i + e_{ij}$$

em que,

Y_{ij} é o valor observado para a variável resposta obtido para o i-ésimo tratamento em sua j-ésima repetição;

m média de todos os valores possíveis da variável resposta;

t_i é o efeito do tratamento i no valor observado Y_{ij};

$$t_i = m_i - m$$

e_{ij} é o erro experimental associado ao valor observado Y_{ij};

$$e_{ij} = Y_{ij} - m_i$$

O erro experimental ocorre em todos os experimentos, porque não é possível controlar o efeito de fontes de variações que ocorrem de forma aleatória e desconhecida. Este erro é o responsável pela variação observada entre as observações obtidas nas repetições para cada tratamento.

4.4. Análise de Variância

É uma técnica de análise estatística que permite decompor a variação total, ou seja, a variação existente entre todas as observações, na variação devido à diferença entre os efeitos dos tratamentos e na variação devido ao acaso, que também é denominada de erro experimental ou resíduo.

No entanto, para que esta técnica seja empregada é necessário que sejam satisfeitas as seguintes pressuposições:

1ª) os efeitos do modelo estatístico devem ser aditivos;

2ª) os erros experimentais devem ser normalmente distribuídos, independentes, com média zero e com variância comum.

Partindo do modelo estatístico, pode-se decompor a variação entre os valores observados nas diferentes causas de variabilidade, como demonstrado a seguir:

Considere o modelo estatístico para um experimento instalado segundo o DIC:

$$Y_{ij} = m + t_i + e_{ij}$$

fazendo $t_i = m_i - m$ e $e_{ij} = Y_{ij} - m_i$, tem-se:

$$Y_{ij} - m = (m_i - m) + (Y_{ij} - m_i),$$

substituindo m, m_i e e_{ij} por seus estimadores tem-se:

$$Y_{ij} - \hat{m} = (\hat{m}_i - \hat{m}) + (Y_{ij} - \hat{m}_i),$$

elevando ambos os membros ao quadrado

$$(Y_{ij} - \hat{m})^2 = [(\hat{m}_i - \hat{m}) + (Y_{ij} - \hat{m}_i)]^2,$$

aplicando somatório

$$\sum_{i=1,j=1}^{I,J}(Y_{ij} - \hat{m})^2 = \sum_{i=1,j=1}^{I,J}[(\hat{m}_i - \hat{m}) + (Y_{ij} - \hat{m}_i)]^2 ,$$

$$\sum_{i=1,j=1}^{I,J}(Y_{ij} - \hat{m})^2 = \sum_{i=1,j=1}^{I,J}(\hat{m}_i - \hat{m})^2 + \sum_{i=1,j=1}^{I,J}(Y_{ij} - \hat{m}_i)^2 + \sum_{i=1,j=1}^{I,J}\text{duplos produtos}$$

pode-se verificar que: $\sum_{i=1,j=1}^{I,J}\text{duplos produtos} = 0$.

Escrevendo de uma forma mais simplificada a igualdade anterior temos:

SQTotal = SQTrat + SQRes

Por meio das fórmulas obtidas anteriormente, pode-se obter os valores para as respectivas somas de quadrados. No entanto, essas fórmulas demandam muitos cálculos.

Fórmulas de mais fácil aplicação podem ser obtidas, conforme é mostrado a seguir. Inicialmente trabalharemos com a fórmula da SQTotal.

Tem-se que:

$$\text{SQTotal} = \sum_{i=1,j=1}^{I,J} \left(Y_{ij} - \hat{m}\right)^2$$

desenvolvendo o quadrado perfeito,

$$\sum_{i=1,j=1}^{I,J} \left(Y_{ij} - \hat{m}\right)^2 = \sum_{i=1,j=1}^{I,J} \left(Y_{ij}^2 - 2\hat{m}Y_{ij} + \hat{m}^2\right)$$

aplicando-se as propriedades de somatório, temos:

$$\sum_{i=1,j=1}^{I,J} \left(Y_{ij} - \hat{m}\right)^2 = \sum_{i=1,j=1}^{I,J} Y_{ij}^2 - 2\hat{m}\sum_{i=1,j=1}^{I,J} Y_{ij} + \sum_{i=1,j=1}^{I,J} \hat{m}^2$$

$$\sum_{i=1,j=1}^{I,J} \left(Y_{ij} - \hat{m}\right)^2 = \sum_{i=1,j=1}^{I,J} Y_{ij}^2 - 2\hat{m}\sum_{i=1,j=1}^{I,J} Y_{ij} + IJ\hat{m}^2$$

A média geral pode ser escrita como: $\hat{m} = \dfrac{\sum_{i=1,j=1}^{I,J} Y_{ij}}{IJ}$, assim

$$\sum_{i=1,j=1}^{I,J} \left(Y_{ij} - \hat{m}\right)^2 = \sum_{i=1,j=1}^{I,J} Y_{ij}^2 - 2\frac{\sum_{i=1,j=1}^{I,J} Y_{ij}}{IJ}\sum_{i=1,j=1}^{I,J} Y_{ij} + IJ\left(\frac{\sum_{i=1,j=1}^{I,J} Y_{ij}}{IJ}\right)^2$$

simplificando tem-se,

$$\sum_{i=1,j=1}^{I,J} \left(Y_{ij} - \hat{m}\right)^2 = \sum_{i=1,j=1}^{I,J} Y_{ij}^2 - 2\frac{\left(\sum_{i=1,j=1}^{I,J} Y_{ij}\right)^2}{IJ} + \frac{\left(\sum_{i=1,j=1}^{I,J} Y_{ij}\right)^2}{IJ}$$

finalmente temos:

$$\text{SQTotal} = \sum_{i=1,j=1}^{I,J} \left(Y_{ij} - \hat{m}\right)^2 = \sum_{i=1,j=1}^{I,J} Y_{ij}^2 - \frac{\left(\sum_{i=1,j=1}^{I,J} Y_{ij}\right)^2}{IJ}$$

que é a fórmula mais prática para se calcular a SQTotal.

Para a SQTratamentos tem-se:

$$\text{SQTrat} = \sum_{i=1,j=1}^{I,J} \left(\hat{m}_i - \hat{m}\right)^2$$

desenvolvendo o quadrado perfeito,

$$\sum_{i=1,j=1}^{I,J} \left(\hat{m}_i - \hat{m}\right)^2 = \sum_{i=1,j=1}^{I,J} \left(\hat{m}_i^2 - 2\hat{m}\cdot\hat{m}_i + \hat{m}^2\right)$$

aplicando-se as propriedades de somatório, temos:

$$\sum_{i=1,j=1}^{I,J} \left(\hat{m}_i - \hat{m}\right)^2 = \sum_{i=1,j=1}^{I,J} \hat{m}_i^2 - 2\hat{m}\sum_{i=1,j=1}^{I,J} \hat{m}_i + \sum_{i=1,j=1}^{I,J} \hat{m}^2$$

$$\sum_{i=1,j=1}^{I,J} \left(\hat{m}_i - \hat{m}\right)^2 = J\sum_{i=1}^{I} \hat{m}_i^2 - 2\hat{m}J\sum_{i=1}^{I} \hat{m}_i + IJ\hat{m}^2$$

A média geral e a média para tratamentos podem ser escritas respectivamente como:

$$\hat{m} = \frac{\sum_{i=1,j=1}^{I,J} Y_{ij}}{IJ} \quad \text{e} \quad \hat{m}_i = \frac{T_i}{J}$$

substituindo na expressão anterior, tem-se:

$$\sum_{i=1,j=1}^{I,J} (\hat{m}_i - \hat{m})^2 = J\sum_{i=1}^{I} \frac{T_i^2}{J^2} - 2\frac{\sum_{i=1,j=1}^{I,J} Y_{ij}}{IJ} J\sum_{i=1}^{I} \frac{T_i}{J} + IJ\left(\frac{\sum_{i=1,j=1}^{I,J} Y_{ij}}{IJ}\right)^2$$

sabe-se que $T_i = \sum_{j=1}^{J} Y_{ij}$, então

$$\sum_{i=1,j=1}^{I,J} (\hat{m}_i - \hat{m})^2 = J\sum_{i=1}^{I} \frac{T_i^2}{J^2} - 2\frac{\sum_{i=1,j=1}^{I,J} Y_{ij}}{IJ} J\frac{\sum_{i=1,j=1}^{I,J} Y_{ij}}{J} + IJ\left(\frac{\sum_{i=1,j=1}^{I,J} Y_{ij}}{IJ}\right)^2$$

simplificando, tem-se.

$$\sum_{i=1,j=1}^{I,J} (\hat{m}_i - \hat{m})^2 = \sum_{i=1}^{I} \frac{T_i^2}{J} - 2\frac{\left(\sum_{i=1,j=1}^{I,J} Y_{ij}\right)^2}{IJ} + \frac{\left(\sum_{i=1,j=1}^{I,J} Y_{ij}\right)^2}{IJ}$$

finalmente tem-se:

$$SQTrat = \sum_{i=1,j=1}^{I,J} (\hat{m}_i - \hat{m})^2 = \sum_{i=1}^{I} \frac{T_i^2}{J} - \frac{\left(\sum_{i=1,j=1}^{I,J} Y_{ij}\right)^2}{IJ}$$

A fórmula anterior é utilizada quando o número de repetições é igual para todos os tratamentos. No caso em que o número de repetições varia de acordo com o tratamento a fórmula apropriada é

$$SQTrat = \sum_{i=1}^{I} \frac{T_i^2}{r_i} - \frac{\left(\sum_{i=1,j=1}^{I,r_i} Y_{ij}\right)^2}{N}$$

em que,

N é o número de unidades experimentais = $\sum_{i=1}^{I} r_i$

r_i é número de unidades experimentais do tratamento i.

A Soma de Quadrados do Resíduo (SQRes) é obtida por diferença,

SQRes = SQTotal - SQTrat

O quadro da análise de variância, geralmente denotada por ANOVA (**AN**alysis **O**f **VA**riance) para a análise de um experimento instalado segundo o DIC, com igual número de repetições para todos os tratamentos é do seguinte tipo:

FV	GL	SQ	QM	F	$F_{tab;\ \alpha}$
Tratamentos	(I-1)	SQTrat	$\frac{SQTrat}{I-1}$	$\frac{QMTrat}{QMRes}$	[(I-1); I(J-1)]
Resíduo	I(J-1)	SQRes	$\frac{SQRes}{I(J-1)}$		
Total	IJ - 1	SQTotal			

A partir das SQTrat e SQRes, obtém-se os respectivos quadrados médios, por meio do quociente entre a soma de quadrados com o respectivo número de graus de liberdade.

Para se concluir se existe diferença entre tratamentos, calcula-se o valor de F, que é obtido pelo quociente do QMTrat com o QMRes. Este valor de F calculado deve ser comparado com o valor de F tabelado, o qual é obtido na tabela de distribuição da variável aleatória F, de acordo com o nível de significância do teste, graus de liberdade para tratamentos e graus de liberdade para resíduo.

As hipóteses para o teste F da análise de variância para tratamentos são as seguintes:

$H_0 : m_1 = m_2 = ... = m_I = m$, o que equivale a dizer que todos os possíveis contrastes entre as médias dos tratamentos, são estatisticamente nulos, ao nível de probabilidade que foi executado o teste.

H_a : não H_0, o que equivale a dizer que existe pelo menos um contraste entre as médias dos tratamentos, estatisticamente diferentes de zero, ao nível de probabilidade que foi realizado o teste.

A regra decisória para o teste F é a seguinte:

- se o valor do F calculado for maior ou igual ao valor do F tabelado, então rejeita-se H_0 e conclui-se que os tratamentos tem efeito diferenciado ao nível de significância em que foi realizado o teste;
- se o valor de F calculado for menor que o valor do F tabelado, então não rejeita-se H_0 e conclui-se que os tratamentos têm efeitos iguais ao nível de significância em que foi realizado o teste.

4.5. Coeficiente de Variação

O coeficiente de variação é calculado da seguinte maneira:

$$CV = \frac{\sqrt{QMRes}}{\hat{m}} \cdot 100$$

O CV é utilizado para avaliação da precisão de experimentos. Quanto menor o CV mais preciso tende a ser o experimento. A título de classificação geral pode-se utilizar a seguinte tabela

C.V.	Avaliação	Precisão
< 10%	Baixo	Alta
10 a 20%	Médio	Média
20 a 30%	Alto	Baixa
>30%	Muito Alto	Muito Baixa

Porém o valor do CV não tem nada de absoluto, pois existe uma variabilidade inerente a cada área de pesquisa. Por exemplo, experimentos realizados em locais com ambiente controlado geralmente são mais precisos e podem apresentar CV menores que 5%.

4.6. Vantagens e Desvantagens do delineamento inteiramente casualizado

Vantagens

a) não existem exigências quanto ao número de tratamentos e repetições;

b) é o delineamento experimental que apresenta o maior valor para o número de graus de liberdade associado ao resíduo.

Desvantagens

a) não é fácil conseguir e manter total homogeneidade das condições durante a toda a realização do experimento;

b) todas as variações exceto a devida a tratamentos, são consideradas como sendo variações que ocorrem ao acaso. Isto pode acarretar em uma estimativa muito alta para o erro experimental.

4.7. Exercícios

4.1. Para comparar a produtividade de quatro variedades de milho, um agrônomo tomou vinte parcelas similares e distribuiu, inteiramente ao acaso, cada uma das 4 variedades em 5 parcelas experimentais. A partir dos dados experimentais fornecidos abaixo, é possível concluir que existe diferença significativa entre as variedades com relação a produtividade, utilizando o nível de significância de 5%?

	Variedades			
	A	B	C	D
	25	31	22	33
	26	25	26	29
	20	28	28	31
	23	27	25	34
	21	24	29	28
Totais	115	135	130	155
Médias	23	27	26	31

4.2. Um treinador de corrida rústica, objetivando melhorar o desempenho de seus atletas, testou três novas técnicas de preparação. Para tanto trabalhou com um grupo de 15 atletas completamente homogêneos para as características essenciais. A designação das técnicas de preparação aos atletas foi feita totalmente ao acaso e de tal forma que o número de atletas avaliados em cada uma das técnicas fosse o mesmo. Os resultados obtidos, após um determinado período de tempo de aprendizado da técnica pelos atletas, foram os seguintes (minutos / 25 Km):

	Técnicas de Preparação		
Repetições	1	2	3
1	130	125	135
2	129	131	129
3	128	130	131
4	126	129	128
5	130	127	130
Totais	643	642	653

De acordo com os resultados obtidos, pede-se.

a) Quais foram os Princípios Básicos da Experimentação utilizados pelo pesquisador neste experimento?

b) Qual foi a unidade experimental nesta pesquisa?

c) É possível concluir que existe diferença entre as técnicas de preparação com relação ao tempo médio gasto para percorrer a distância de 25 km? $(\alpha = 1\%)$

d) Qual seria a técnica a ser recomendada?

4.3. Com o objetivo de diminuir o consumo dos motores à gasolina, uma determinada indústria petroquímica testou 4 novas formulações de gasolina, as quais se diferenciavam pelo tipo de aditivo que era acrescentado à mesma durante o seu processo de fabricação. Para efetuar o teste, a indústria petroquímica utilizou carros completamente homogêneos para todas as características. A designação das formulações aos carros foi feita inteiramente ao acaso. Após os testes de rodagem, os resultados obtidos foram (km/l):

Aditivo a base de	Ácido Forte	Ácido Fraco	Base Forte	Base Fraca
Médias	14,81	6,56	10,06	10,09
Nº de carros	10	10	10	10

SQResíduo=6,0264

Com base nos resultados acima, pede-se:

a) Existe diferença entre os 4 tipos de formulações? $(\alpha = 5\%)$

b) Estabeleça um contraste entre o grupo à base de formulação ácida contra o grupo à base de formulação básica. Obtenha a estimativa para este contraste.

c) Estabeleça um contraste para comparar aditivos de formulação ácida. Obtenha a estimativa para este contraste.

d) Estabeleça um contraste para comparar aditivos de formulação básica. Obtenha a estimativa para este contraste.

4.4. Com o objetivo de verificar se a parótida tem influência na taxa de glicose no sangue, em ratos, um experimento no DIC foi realizado. Vinte e quatro ratos machos da raça W foram escolhidos aleatoriamente e separados em três grupos. Os dados referentes as taxas de glicose, em miligramas por 100 ml de sangue, segundo o grupo, em ratos machos com 60 dias de idade são dados abaixo:

Parotidectomizado	96,0	95,0	100,0	108,0	120,0	110,5	97,0	92,5
Pseudoparotidectomizado	90,0	93,0	89,0	88,0	87,0	92,5	87,5	85,0
Normal	86,0	85,0	105,0	105,0	90,0	100,0	95,0	95,0

Usando $\alpha = 5\%$, testar a hipótese de que as médias relativas aos três grupos são iguais, e concluir.

4.5.O resultado das vendas efetuadas por 3 vendedores de uma indústria de pesticidas durante certo período é dado a seguir. Ao nível de 5% de probabilidade e considerando os vendedores como tratamentos de um D.I.C., verifique se há diferença de eficiência entre os vendedores.

	Vendedores		
	A	B	C
	29	27	30
	27	27	30
	31	30	31
	29	28	27
	32		29
	30		
Totais	178	112	147

4.6. Baseado nas informações fornecidas abaixo e supondo que os tratamentos que possuem as maiores médias são os desejados, pergunta-se:

Qual(is) tratamento(s) deve(m) ser recomendado(s)? Justifique a sua resposta. Use o nível de 1% de significância.

FV	GL	SQ	QM	F
Tratamentos	2	14,80	7,40	
Resíduo				
Total	14	78,40		

Médias de tratamentos:

$$\hat{m}_1 = 128{,}6 \qquad \hat{m}_2 = 128{,}4 \qquad \hat{m}_3 = 130{,}6$$

4.7. Os seguintes dados referem-se a ganhos de peso, em kg, de animais durante um período experimental.

	Repetições				
Rações	1	2	3	4	Totais
A	7,1	8,9	6,0	7,0	29,0
B	6,2	8,8	4,9	6,1	26,0
C	6,0	5,0	9,1	3,9	24,0
D	11,1	10,8	10,2	11,9	44,0
E	7,0	11,3	10,0	11,7	40,0
					163,0

Tais dados são descritos segundo o modelo estatístico: $Y_{ij} = m + t_i + e_{ij}$. Baseando nas informações fornecidas, pede-se:

4.7.1. Proceda a análise de variância dos dados (use α = 5%)

4.7.2. De acordo com o resultado do teste F, pode-se concluir que existe efeito significativo de rações com relação ao ganho de peso médio proporcionado pelas mesmas?

4.7.3. Proponha um contraste que compare as rações B e C juntas contra as rações D e E. Obtenha a estimativa para este contraste.

4.7.4. Calcule o coeficiente de variação e interprete-o.

5. Procedimentos para Comparações Múltiplas

5.1. Introdução

O fator ou fatores em avaliação em um experimento podem ser classificados como qualitativo ou quantitativo. Um fator quantitativo é aquele onde cada nível é descrito por uma quantidade numérica em uma escala. Como exemplos tem-se temperatura, umidade, concentração de um princípio ativo, níveis de insumo, pH, etc ... Para estudar o efeito deste tipo de fator recomenda-se realizar uma análise de regressão, assunto que será abordado no Capítulo 10.

Por outro lado, um fator qualitativo é aquele onde os níveis diferem por algum atributo qualitativo. Como exemplos têm-se variedades, tipos de defensivos, métodos de conduzir uma determinada tarefa, etc. Para estudar o efeito deste tipo de fator, deve-se proceder à análise de variância dos dados e, se for conveniente, proceder às comparações entre as médias dos níveis do fator usando algum dos procedimentos para comparações múltiplas descritos neste capítulo.

A análise de variância, conforme visto no capítulo anterior, serve para verificar se existe alguma diferença significativa entre as médias dos níveis de um fator a um determinado nível de significância. Se o teste F para a fonte de variação que representa o fator em estudo for não-significativa, ou seja, a hipótese de nulidade (H_o: $m_1 = m_2 = ... = m_I$) não for rejeitada, todos os possíveis contrastes entre médias de tratamentos são estatísticamente nulos. Neste caso, não é necessário a aplicação de nenhum procedimento de comparações múltiplas.

Por outro lado se o teste F for significativo, ou seja a hipótese de nulidade for rejeitada, implica que existe pelo menos um contraste entre médias estatisticamente diferente de zero. Os procedimentos de comparações múltiplas a serem vistos neste capítulo, visam identificar qual(is) é(são) esse(s) contraste(s), para podermos por conseqüência identificarmos qual(is) é(são) o(s) nível(is) do fator em estudo que apresentou(ram) maior(es) média(s).

Dentre os diversos testes existentes na literatura, serão vistos os quatro testes mais comumente utilizados. Estes testes podem ser divididos em duas categorias principais de acordo com os tipos de contrastes que podem ser testados:

1ª) Procedimentos para testar todos os possíveis contrastes entre duas médias dos níveis do fator em estudo

a) Teste de Tukey
b) Teste de Duncan

2ª) Prodedimentos para testar todos os possíveis contrastes entre médias dos níveis do fator em estudo

a) Teste t de Student
b) Teste de Scheffé

Todos os procedimentos se baseiam no cálculo de uma diferença mínima significativa (dms). A dms representa o menor valor que a estimativa de um contraste deve apresentar para que se possa considerá-lo como significativo. Por exemplo, para um contraste entre duas médias, a dms representa qual é o menor valor que tem que ser detectado entre as suas estimativas para que se possa concluir que os dois tratamentos produzam efeitos significativamente diferentes.

A princípio um determinado contraste, por exemplo, entre duas médias poderia ser testado por cada um dos procedimentos aqui apresentados. A conclusão a respeito da significância do contraste pode variar de um procedimento para outro, pois o valor da dms varia de um teste para outro, pois cada um se baseia numa distribuição de probabilidades específica. Devido a esta possibilidade na diferença de conclusões a respeito da significância do contraste, nós podemos dizer que um teste é mais conservador (ou rigoroso) que o outros. Na estatística dizemos que um teste é mais conservador que o outro quando a dms dele é maior, pois ele tende a “conservar” a hipótese de igualdade entre médias como verdadeira. Isto porque quanto maior a dms mais difícil se torna rejeitar a hipótese de nulidade.

Este maior ou menor conservadorismo de um teste pode ajudar o pesquisador a escolher um procedimento de comparação múltipla. Se por exemplo, por experiência própria o pesquisador sabe que as diferenças entre os efeitos dos níveis do fator em teste são *pequenas* e ele deseja detectar estas *pequena*s diferenças, então ele deve usar um procedimento menos conservador, ou seja, que apresenta uma menor dms. Se por outro lado, ele quer concluir que os níveis do fator têm efeitos diferentes somente quando a diferença nos seus efeitos for realmente *grande*, então ele deve usar um teste mais conservador, ou seja, com maior dms.

Vamos ver a partir de agora cada procedimento com mais detalhe. Considere para tanto, que estamos interessados em comparar as médias dos I níveis de um fator qualitativo, as quais foram obtidas a partir da realização de um experimento no delineamento inteiramente casualizado com J repetições, para o qual o teste F para fator foi significativo; e que o número de graus de liberdade para o fator em estudo foi igual a n_1 e para o resíduo foi igual a n_2, ou seja,

FV	GL	SQ	QM	F
Fator	I-1	SQFator	QMTrat	significativo
Resíduo	I(J-1)	SQRes	QMRes	
Total	IJ - 1	SQTotal		

5.2. Alguns Procedimentos Para Comparações Múltiplas

Dentre vários procedimentos existentes para comparações múltiplas, serão apresentados quatro: teste de Tukey, teste de Duncan, teste t de Student e teste de Scheffé.

Teste de Tukey

O teste de Tukey, pode ser utilizado para comparar a totalidade dos contrastes entre duas médias, ou seja, para os I(I–1)/2 contrastes do tipo $C=m_i - m_u$; para $1 \leq i < u \leq I$, em que I é o número de níveis do fator em estudo. Este teste baseia-se na diferença mínima significativa (d.m.s.) representada por Δ e dada por:

$$\Delta = q\sqrt{\frac{1}{2}\hat{V}\left(\hat{C}\right)}$$

em que,

$q = q_\alpha(I, n_2)$ é o valor tabelado da amplitude total estudentizada, que é obtido em função do nível α de significância do teste, número de níveis do fator em estudo (I) e número de graus de liberdade do resíduo (n_2) da análise de variância.

$$\hat{V}(\hat{C}) = QMRes\left(\frac{1}{r_i} + \frac{1}{r_u}\right)$$

No caso em que todos os tratamentos apresentaram o mesmo número de repetições, ou seja, $r_i = r_u = K$, o valor de Δ é simplificado com a seguinte expressão

$$\Delta = q\sqrt{\frac{QMRes}{K}}$$

Para a realização do teste Tukey,a um nível de significância α, é necessário:

1. enunciar as hipóteses: H_0: C = 0 vs H_a: C$\neq$ 0, em que C = $m_i - m_u$, para $i \neq u$;
2. obtenção das estimativas dos contrastes, $\hat{C} = \hat{m}_i - \hat{m}_u$, com base nos valores amostrais;
3. cálculo do Δ;
4. concluir a respeito da significância dos I(I–1)/2 contrastes em teste, usando a seguinte relação: se $|\hat{C}| \geq \Delta$, rejeita-se H_0; caso contrário, não se rejeita H_0. Neste caso, indicar as médias iguais, seguidas por uma mesma letra.

Considerações:

1. O teste de Tukey é válido para a totalidade dos contrastes de duas médias.
2. O teste de Tukey exige, em princípio, balanceamento. Mas, no caso dos tratamentos apresentarem números de repetições diferentes, o resultado obtido por este teste é apenas uma aproximação.
3. O teste de Tukey é exato para testar a maior diferença, nos demais casos é conservador.

Teste de Duncan

Tal como o teste de Tukey, o teste de Duncan é um procedimento seqüencial, válido para a totalidade dos contrastes de duas médias do tipo C = $m_i - m_u$. O teste de Duncan necessita a prévia ordenação das médias, dos níveis do fator em estudo. Este teste baseia-se na amplitude total mínima significativa (D_i) dada por:

$$D_i = z_i\sqrt{\frac{1}{2}\hat{V}(\hat{C})}$$

em que,

$z_i = z_\alpha(n,n_2)$ é o valor tabelado da amplitude total estudentizada, que é obtido em função do nível α de probabilidade, número de médias ordenadas abrangidas pelo contraste entre os níveis do fator em estudo (i) e número de g.l. do resíduo da ANOVA (n_2). Como se trata de um processo seqüencial, n_1 varia seu valor durante a aplicação do teste;

$$\hat{V}(\hat{C}) = QMRes\left(\frac{1}{r_i} + \frac{1}{r_u}\right)$$

No caso em que todos os tratamentos apresentaram o mesmo número de repetições, ou seja, $r_i = r_u = K$, o valor de D_i é simplificado com a seguinte expressão

$$D_i = z_i\sqrt{\frac{QMRes}{K}}$$

Para a realização do teste Duncan a um nível de significância α é necessário:

1. enunciar as hipóteses: H_0: C = 0 vs H_a: C$\neq$ 0, em que $C = m_i - m_u$, para $i \neq u$;
2. ordenar as médias do fator em estudo em ordem crescente ou decrescente;
3. obter o valor da estimativa do contraste entre a maior e a menor média, com base nos valores amostrais;
4. calcular o valor de D_i, com base no número de médias ordenadas abrangidas pelo contraste. Neste primeiro passo i= I;
5. concluir a respeito da significância do contraste em teste, usando o seguinte critério:
   a) Se o valor de D_i for maior do que o módulo da estimativa do contraste, não rejeita-se H_0 e as médias são ligadas por um traço, indicando que não há diferença entre elas;
   b) Caso contrário, reduzir de uma unidade o valor de n_1. Calcula-se o novo valor de D_i e, para todos os pares de médias que não estejam ligadas por um mesmo traço e que envolvem n_1 médias, repetir o procedimento que consta no item 3 e nos seguintes;
6. Proceder ao item 3 e seguintes até que i = 2.

Este teste tem como inconveniente, além de ser um teste trabalhoso, o fato das médias ordenadas não serem independentes e o valor de z_i em conseqüência, não ser exato.

Considerações:

1. O teste Duncan é um procedimento seqüencial válido para a totalidade dos contrastes de duas médias.
2. Tal como o teste de Tukey, o teste de Duncan exige, em princípio, balanceamento. Mas, no caso de serem diferentes os números de repetições este teste pode ainda ser usado, mas então é apenas aproximado.
3. Quando a maior média não diferir significativamente da menor, não se admitirá diferença significativa, entre as médias intermediárias.

Teste t de Student

O teste t pode ser utilizado para testar contrastes envolvendo duas ou mais médias. Porém este teste exige que:

1. as comparações a serem realizadas sejam escolhidas a *priori*, ou seja, antes de serem examinados os dados;
2. podem-se testar no máximo, tantos contrastes quantos são os graus de liberdade para tratamentos, e estes contrastes devem ser ortogonais.

A ortogonalidade entre os contrastes indica independência linear na comparação estabelecida por um contraste com a comparação estabelecida pelos outros contrastes. Entre I médias de um fator, podem ser obtidos I – 1 contrastes ortogonais.

Consideremos um contraste de médias, entre os níveis de um fator, em sua forma geral:

$$C = a_1 m_1 + a_2 m_2 + \ldots + a_I m_I$$

do qual obtemos a estimativa por meio do estimador

$$\hat{C} = a_1 \hat{m}_1 + a_2 \hat{m}_2 + \ldots + a_I \hat{m}_I \ ,$$

que pode ser testada pelo teste t, calculando-se a estatística t, dada por.

$$t = \frac{\hat{C} - C}{\sqrt{\hat{V}(\hat{C})}} = \frac{\hat{C} - C}{\sqrt{QMRes \sum_{i=1}^{I} \frac{a_i^2}{r_i}}}$$

que tem distribuição t de Student com n_2 graus de liberdade, sendo n_2 o número de graus de liberdade do resíduo e QMResíduo o quadrado médio residual da análise de variância.

Caso o número de repetições seja o mesmo para todos os tratamentos, ou seja $r_1=r_2=...=r_I=K$, então a fórmula para a aplicação do teste t é

$$t = \frac{\hat{C} - C}{\sqrt{\frac{QMRes}{K} \sum_{i=1}^{I} a_i^2}}$$

Quando aplicamos o teste t a um contraste, C, geralmente o interesse é testar as hipóteses: H_0: C = 0 vs H_a: C $\neq$ 0.

O valor tabelado de t é obtido por $t_{tab}=t_{\alpha}(n_2)$.

A regra de decisão, neste caso, é a seguinte:

Se $|t| \geq t_{tab} \Rightarrow$ rejeita-se H_0 .
Caso contrário não se rejeita H_0 .

Considerações:

1. O nível de significância α é válido para um único contraste, e não para uma série deles;
2. O nível de significância α é válido somente se o contraste for estabelecido a *priori* e não sugerido pelos dados, pois, pode ficar caracterizado uma estatística de ordem ao querer comparar a maior com a menor média, o que acarretaria certa dependência entre as médias.

Teste de Scheffé

Este teste pode ser aplicado para testar todo e qualquer contraste entre médias, mesmo quando sugerido pelos dados. É freqüentemente utilizado para testar contrastes que envolvam grupos de médias. É um teste mais conservador que o teste t, porém não exige que os contrastes a serem testados sejam ortogonais e nem que estes contrastes sejam estabelecidos antes de se examinar os dados.

Se o valor de F obtido não for significativo, nenhum contraste poderá ser significativo pelo teste de Scheffé, e sua utilização não se justifica.

A estatística do teste, denotada por S, é calculada por:

$$S = \sqrt{(I-1)F_{tab}\hat{V}(\hat{C})}$$

em que,

I = é o número de níveis do fator em estudo;

$F_{tab} = F_{\alpha}(I\text{-}1;n_2)$ é o valor tabelado de F, obtido em função do nível α de probabilidade, número de graus de liberdade do fator em estudo, ou seja I-1, e número de graus de liberdade do resíduo, ou seja n_2;

$$\hat{V}(\hat{C}) = QMRes \sum_{i=1}^{I} \frac{a_i^2}{r_i}$$

Caso o número de repetições seja o mesmo para todos os tratamentos, ou seja, $r_1=r_2=...=r_I=K$, então a fórmula para a aplicação do teste Schheffé é

$$S = \sqrt{(I-1)F_{tab}\frac{QMRes}{K}\sum_{i=1}^{I} a_i^2}$$

Deve-se então, calcular a estimativa do contraste C, ou seja,

$$\hat{C} = a_1\hat{m}_1 + a_2\hat{m}_2 + ... + a_I\hat{m}_I$$

Se verificarmos que $|\hat{C}| \geq S$, dizemos que o contraste é significativamente diferente de zero ao nível α de probabilidade, indicando que os grupos de médias confrontados no contraste diferem entre si a esse nível de probabilidade.

Considerações:

1. O teste de Scheffé é válido para a totalidade dos contrastes.
2. Para testar um único contraste, ou para testar um número pequeno deles, o teste de Scheffé é bastante rigoroso.

5.3. Vantagens e Desvantagens dos Procedimentos Para Comparações Múltiplas

O teste t não é recomendado para testar todas as possíveis comparações entre médias de um experimento, pois este teste aponta *pequenas* diferenças como significativas. O procedimento de Duncan também é sensitivo, no sentido de declarar *pequenas* diferenças como significativas. Para estes dois testes, Duncan e t, o nível de significância conjunto para um grande número de comparações é elevado. Quando são utilizados para esta finalidade, estes testes podem apontar como significativos contrastes, quando na verdade estes contrastes são não-significativos. Neste acaso o erro tipo I tende a ocorrer mais frequentemente do que o estabelecido pelo nível de significância do teste.

O teste de Tukey é bastante rigoroso no sentido de apontar diferenças significativas. Este teste é útil quando se deseja informações preliminares a respeito das diferenças entre os efeitos dos níveis de um fator. O procedimento de Scheffé é ainda mais rigoroso que o Tukey para comparar pares de médias.

Para a comparação de um número grande de médias, não há um procedimento ideal. Testes como Tukey ou Scheffé, tornam-se extremamente rigorosos, pois o nível de significância conjunto para a maioria dos contrastes é muito menor do que o estabelecido. O inverso ocorre com o teste t e Duncan.

5.4. Exercícios

5.1. Aplique os testes Tukey e Duncan, aos exemplos dados ao final da apostila do Capítulo de Delineamento Inteiramente Casualizado.

5.2. Para os dados fornecidos a seguir, conclua pelo teste Duncan e Tukey $(\alpha = 5\%)$.

$\hat{m}_1 = 370 \quad \hat{m}_2 = 338 \quad \hat{m}_3 = 380 \quad \hat{m}_4 = 320 \quad \hat{m}_5 = 325 \quad \hat{m}_6 = 367$

$D_6 = 31 \quad D_5 = 30{,}2 \quad D_4 = 28{,}7 \quad D_3 = 26 \quad D_2 = 24{,}6 \quad \Delta = 33$

5.3. Aplicar o teste de Duncan às comparações múltiplas obtidas com as médias dos tratamentos instalados em um experimento segundo o Delineamento Inteiramente Casualizado (DIC).Concluir para $\alpha = 5\%$ de probabilidade.

$T_1 = 452{,}16$ $T_2 = 481{,}80$ $T_3 = 442{,}56$ $T_4 = 469{,}52$ $T_5 = 439{,}48$

$T_6 = 461{,}6$ SQTratamento $= 331{,}8677$ SQTotal $= 783{,}4964$ $r = 4$

5.4. Um experimento para avaliar a influência de 4 tipos de aleitamento no ganho de peso de leitões foi conduzido utilizando-se o delineamento inteiramente casualizado com 4 repetições. Foram obtidos os seguintes resultados parciais:

Tratamentos	1	2	3	4
Totais	37,2	44,8	31,6	32,8

FV	GL	SQ	QM	F
Tratamento		26,76		
Resíduo				
Total		33,82		

Complete o quadro da ANOVA e, considerando-se α = 1%, responda qual(is) o(s) melhor(es) tipo(s) de aleitamento. (Use o teste de Tukey, se necessário)

5.5. Com o objetivo de verificar se existe diferença, no tempo médio gasto para ir de 0-100 km/h, entre 5 marcas de carro de mesma categoria, 4 carros de cada marca foram escolhidos inteiramente ao acaso da linha de produção de cada marca e avaliados em uma pista de provas apropriada. Os resultados obtidos, em segundos, foram:

Marcas				
1	2	3	4	5
12	12	8	12	13
11	10	7	12	14
11	10	8	10	15
13	11	6	11	13

Usando o nível de 5% de probabilidade

a. Existe de diferença significativa entre as marcas de carro quanto ao tempo médio gasto para ir de 0-100 km/h?

b. Qual(is) é(são) a(s) marca(s) mais lenta(s) para ir de 0-100 km/h, pelo teste de Duncan?

c. Qual(is) é(são) a(s) marca(s) mais rápida(s) para ir de 0-100 km/h, pelo teste de Tukey?

d. Suponha que em termos de custo final ao consumidor pode-se classificar os carros produzidos pela marca 1 como de custo alto, os produzidos pelas marcas 2 e 3 de custo médio e aqueles produzidos pelas marcas 4 e 5 como de custo alto. Suponha também que este experimento tinha como objetivos verificar se existe diferença no tempo médio para ir de 0-100 km/h entre: 1) os carros de custo alto e os demais carros; 2) entre os carros de custo médio e os de custo alto; 3) os carros de custo médio; e 4) os carros de custo baixo. Utilize os testes de Scheffé e de t para verificar se estas comparações são significativas.

5.6. Quatro padarias da cidade de São Paulo, foram fiscalizadas para verificar a quantidade de bromato de potássio existente nos pães franceses que elas produzem. Com esta finalidade foi tomada uma amostra de pães, inteiramente ao acaso, de cada padaria e para cada um deles foi avaliado o teor de bromato de potássio (mg de bromato de potássio/1kg de pão). O resumo da avaliação é fornecido a seguir:

Padaria	1	2	3	4
Teor médio	10	11	8	9
Núm. de pães avaliados	7	8	7	8

SQResíduo = 52

Usando o nível de 5% de probabilidade
a. Pode-se concluir que existe diferença significativa no teor médio de bromato de potássio no pão entre as padarias avaliadas?
b. Suponha que as padarias 1 e 2 suprem a classe social A, a padaria 3 a classe B e a 4 a classe C. Verifique, por meio de um contraste, pelo teste de Scheffé e pelo teste t, se existe diferença no teor médio de bromato de potássio entre as padarias que suprem as classes A e C.

5.7. Com os dados fornecidos a seguir oriundos de um experimento instalado no DIC com 4 repetições, para o qual o teste F da ANOVA para tratamentos foi significativo, aplicar o teste de Duncan e o teste de Tukey para se concluir qual(is) tratamento(s) apresentou(aram) maior(es) média(s) ao nível de 5% de probabilidade.

SQResíduo = 905,6790

T_1 = 813,44 T_2 = 729,52 T_3 = 786,32 T_4 = 661,52 T_5 = 755,44 T_6 = 612,50

6. Delineamento em Blocos Casualizados

6.1. Introdução

O principal objetivo do planejamento e execução de um experimento é apontar diferenças significativas entre os efeitos os níveis de um fator em avaliação. Inicialmente isto é realizado mediante o teste F para o fator. Se o teste F for não-significativo, concluímos que os efeitos são estatisticamente iguais e nada mais precisa ser feito. Por outro lado, se o teste F for significativo, concluímos que existe diferença significativa nos efeitos dos niveis do fator. O passo seguinte seria o uso de um procedimento de comparações múltiplas para identificar quais níveis dos fatores proporcionam efeitos significativamente diferentes entre si do ponto de vista estatístico.

Tal como o teste F, todos os procedimentos de comparação múltipla tem como base para o cálculo do valor da diferença mínima significativa a estimativa da variabilidade associada ao efeito do erro experimental, a qual é conhecida como Quadrado Médio do Resíduo (QMRes). Sendo assim fica fácil entender que, para o pesquisador conseguir atingir o seu objetivo, apontar diferenças significativas entre os efeitos de níveis do fator, ele deve planejar e executar o seu experimento de tal forma que a influência do erro experimental seja a menor possível.

O delineamento inteiramente casualizado pressupõe para ser utilizado que, as unidades experimentais sejam e estejam durante todo o experimento em condições ambientais completamente homogêneas. Caso o pesquisador perceba que algum fator perturbe a homogeneidade das unidades experimentais ou nas condições ambientais que as mesmas vão estar sujeitas durante o experimento, é necessário que o pesquisador controle o efeito deste fator pertubador. Entenda-se aqui fator pertubador como uma fonte de variação indesejável entre as unidades experimentais ou nas condições ambientais. Um exemplo seria a situação em que um pesquisador deseja comparar o efeito de analgésicos em cobaias. No entanto as cobaias não são de mesma idade. Se o pesquisador achar que a idade da cobaia pode influenciar na avaliação dos analgésicos, ele deve controlar o efeito do fator pertubador idade.

O controle do efeito do fator pertubador é feito pela formação de grupos, ou seja, blocos de unidades experimentais homogêneas e fazendo com que todos os níveis do fator em estudo sejam avaliados em cada nível do fator pertubador, ou seja, em cada bloco de unidades homogêneas. No delineamento em blocos casualizados (DBC), a distribuição ao acaso dos níveis do fator em estudo às unidades experimentais, sofre a restrição de ser feita dentro de cada bloco. Portanto o DBC faz uso dos três princípios básicos da experimentação: repetição, casualização e controle na casualização. Vale lembrar que no delineamento inteiramente casualizado (DIC), não existe nenhuma restrição na casualização, uma vez que os níveis do fator em estudo são distribuídos **inteiramente** ao acaso em relação a **todas** unidades experimentais.

Em experimentos instalados segundo o DBC, espera-se que as condições experimentais de um bloco sejam diferentes das condições experimentais do outro bloco e que haja homogeneidade das condições experimentais dentro de cada bloco.

Se um pesquisador instala o seu experimento segundo o DBC, o efeito do fator perturbador é controlado sendo portanto possível quantificar o seu efeito e eliminar tal efeito na análise estatística dos dados experimentais. Caso o pesquisador não controle o efeito do fator perturbador por meio da formação de blocos de unidades experimentais homogêneas e controle na casualização, o efeito do fator pertubador é absorvido pelo erro experimental. Tal absorção tende a provocar um aumento no valor do QMRes, o que pode acarretar em não identificar nenhuma diferença nos efeitos dos tratamentos, quando de fato uma ou mais diferenças possam existir.

No entanto, a instalação de um experimento no DBC quando o mesmo não é necessário, pode implicar na perda de eficiência do experimento, pois quando se instala um experimento no DBC com J blocos, quando na verdade o DIC seria suficiente, são *perdidos* (J-1) graus de liberdade para o resíduo. No DBC o nº de graus de liberdade para o resíduo é menor. Conseqüente o F tabelado é maior. Portanto maior deverá ser a diferença entre os efeitos dos níveis do fator para que tais diferenças atinjam significância estatística.

6.2. Quadro de tabulação dos dados

A título de exemplo, considere um experimento instalado no DBC com I tratamentos e J repetições (blocos). A coleta de dados da pesquisa pode ser resumida, num quadro do tipo a seguir:

	Tratamentos				
Blocos	1	2	...	I	Totais
1	Y_{11}	Y_{21}	...	Y_{I1}	B_1
2	Y_{12}	Y_{22}	...	Y_{I2}	B_2
...	...	...	...	...	...
J	Y_{1J}	Y_{2J}	...	Y_{IJ}	B_J
Totais	T_1	T_2	...	T_I	G

Deste quadro pode-se retirar algumas informações de interesse:

- nº de unidades experimentais: N = I x J;
- Total geral: $G = \sum_{i=1,j=1}^{I,J} Y_{ij} = \sum_{i=1}^{I} T_i = \sum_{j=1}^{J} B_j = Y_{\bullet\bullet}$;
- Total para o tratamento i: $T_i = \sum_{j=1}^{J} Y_{ij} = Y_{i\bullet}$;
- Total para o bloco j: $B_j = \sum_{i=1}^{I} Y_{ij} = Y_{\bullet j}$;
- média para o tratamento i: $\hat{m}_i = \dfrac{T_i}{J}$;
- média para o bloco j: $\hat{m}_j = \dfrac{B_j}{I}$;
- média geral do experimento: $\hat{m} = \dfrac{G}{IJ}$.

6.3. Modelo Estatístico

Para o DBC o modelo estatístico é:

$$Y_{ij} = m + t_i + b_j + e_{ij}$$

em que,

Y_{ij} é o valor observado para a variável em estudo referente ao tratamento i no bloco j,

m média de todas as unidades experimentais para a variável em estudo.
t_i é o efeito do particular tratamento i no valor observado Y_{ij}:

$$t_i = m_i - m$$

b_j é o efeito do bloco j no valor observado Y_{ij}:

$$b_j = m_j - m$$

e_{ij} é o erro associado a observação Y_{ij}:

$$e_{ij} = Y_{ij} + m - m_i - m_j$$

6.4. Análise de Variância

Para realizar a análise dos dados obtidos de um experimento instalado segundo o DBC, deve-se decompor a variação total que existe entre todas as observações nas partes que a compõe. Neste tipo de delineamento, a decomposição é feita da seguinte forma:

SQTotal = SQTratamentos + SQBlocos + SQResíduo

conforme é demonstrado a seguir.

Considere o modelo estatístico para um experimento instalado segundo o DBC:

$$Y_{ij} = m + t_i + b_j + e_{ij}$$

fazendo $t_i = m_i - m$ e $b_j = m_j - m$, tem-se:

$$Y_{ij} - m = (m_i - m) + (m_j - m) + e_{ij},$$

substituindo m m_i, m_j e e_{ij} por seus estimadores tem-se:

$$Y_{ij} - \hat{m} = (\hat{m}_i - \hat{m}) + (\hat{m}_j - \hat{m}) + \hat{e}_{ij},$$

elevando ambos os membros ao quadrado

$$(Y_{ij} - \hat{m})^2 = \left[(\hat{m}_i - \hat{m}) + (\hat{m}_j - \hat{m}) + \hat{e}_{ij}\right]^2,$$

aplicando somatório

$$\sum_{i=1,j=1}^{I,J}(Y_{ij} - \hat{m})^2 = \sum_{i=1,j=1}^{I,J}\left[(\hat{m}_i - \hat{m}) + (\hat{m}_j - \hat{m}) + \hat{e}_{ij}\right]^2,$$

$$\sum_{i=1,j=1}^{I,J}(Y_{ij} - \hat{m})^2 = \sum_{i=1,j=1}^{I,J}(\hat{m}_i - \hat{m})^2 + \sum_{i=1,j=1}^{I,J}(\hat{m}_j - \hat{m})^2 + \sum_{i=1,j=1}^{I,J}\hat{e}_{ij}^{\,2} + \sum_{i=1,j=1}^{I,J}\text{duplos produtos}$$

Ou seja:

$$\text{SQTotal} = \text{SQTratamentos} + \text{SQBlocos} + \text{SQResíduo} + \sum_{i=1,j=1}^{I,J}\text{duplos produtos}$$

pode-se verificar que: $\sum_{i=1,j=1}^{I,J}\text{duplos produtos} = 0$.

Por meio das fórmulas obtidas no desenvolvimento anterior, pode-se obter os valores para as respectivas somas de quadrados. No entanto, essas fórmulas são muito trabalhosas para se obter tais valores.

São fornecidas a seguir, fórmulas mais práticas para se obter as somas de quadrados.

$$SQTotal = \sum_{i=1,j=1}^{I,J} Y_{ij}^{\ 2} - \frac{\left(\sum_{i=1,j=1}^{I,J} Y_{ij}\right)^2}{IJ}$$

$$SQTratamentos = \sum_{i=1}^{I} \frac{T_i^2}{J} - \frac{\left(\sum_{i=1,j=1}^{I,J} Y_{ij}\right)^2}{IJ}$$

$$SQBlocos = \sum_{j=1}^{J} \frac{B_j^2}{I} - \frac{\left(\sum_{i=1,j=1}^{I,J} Y_{ij}\right)^2}{IJ}$$

SQResíduo = SQTotal - SQTratamentos - SQBlocos

Estas fórmulas práticas são deduzidas a partir das somas de quadrados, obtidas no desenvolvimento anterior, mediante o desenvolvimento do quadrado do binômio, aplicação dos somatórios a todos os termos e substituição de cada uma das médias pelo quociente do total pelo nº de observações que origina cada total. As deduções são semelhantes àquelas apresentadas no capítulo de Delineamento Inteiramente Casualizado.

O quadro da ANOVA para a análise de um experimento instalado segundo o DBC é do seguinte tipo:

FV	GL	SQ	QM	F
Blocos	(J-1)	SQBlocos	-	-
Tratamentos	(I-1)	SQTratamentos	$\frac{SQTrat}{I-1}$	$\frac{QMTrat}{QMRes}$
Resíduo	(I-1)(J-1)	SQResíduo	$\frac{SQRes}{(I-1)(J-1)}$	-
Total	IJ - 1	SQTotal	-	-

Geralmente, o que interessa na análise de um experimento, é avaliar se existe diferença entre os tratamentos, o que pode ser verificado por meio do teste F para tratamentos.

As hipóteses para o teste F da análise de variância para tratamentos são as seguintes:

$H_0 : m_1 = m_2 = ... = m_I = m$, o que equivale a dizer que todos os possíveis contrastes entre médias de tratamentos, são estatisticamente nulos, ao nível de probabilidade que foi executado o teste.

H_a : não H_0, o que equivale a dizer que existe pelo menos um contraste entre médias, estatisticamente diferente de zero, ao nível de probabilidade que foi realizado o teste.

O teste F para blocos, ou seja, comparação entre blocos, geralmente é desnecessária, pois ao instalar o experimento no DBC, o pesquisador utilizou os blocos para controlar uma causa de variação conhecida.

Nos casos em que a variação entre blocos é duvidosa, o pesquisador pode realizar o teste F para blocos, para servir como orientação para a instalação de futuros experimentos.

6.5. Exercícios

6.1. Os dados abaixo, se referem a um experimento instalado segundo o DBC, em que os tratamentos, 5 produtos comerciais para suprir deficiência de micronutriente em caprinos, foram fornecidos aos animais os quais foram separados em 3 grupos segundo a idade. Os resultados obtidos, expressos em ppm de micronutriente/ml de sangue, foram os seguintes:

	Produtos comerciais					
Bloco	1	2	3	4	5	Totais
1	83	86	103	116	132	520
2	63	69	79	81	98	390
3	55	61	79	79	91	365
Totais	201	216	261	276	321	1275

Pede-se proceder a ANOVA e aplicar o teste Tukey e Duncan, usando o nível de 5% de probabilidade.

6.2. Com a finalidade de aumentar a produção de lã de suas ovelhas, por meio de uma alimentação mais apropriada um criador separou 28 ovelhas de sua criação. Como as ovelhas eram de idades diferentes, dividiu-as em 7 grupos, sendo que dentro de cada um destes grupos havia 4 ovelhas de mesma idade e homogeneidade para as demais características. Dentro de cada grupo foi realizado um sorteio para distribuir ao acaso, os 4 Tipos de Alimentação (TA) às ovelhas do grupo. O experimento se iniciou logo após as ovelhas terem sido submetidas a uma tosquia e se encerrou quando já era o momento de se realizar uma nova tosquia da qual foram obtidos os seguintes resultados, expressos em unidade de medida de lã por animal:

	grupos							
TA	1	2	3	4	5	6	7	Totais
1	30	32	33	34	29	30	33	221
2	29	31	34	31	33	33	29	220
3	43	47	46	47	48	44	47	322
4	23	25	21	19	20	21	22	151
Totais	125	135	134	131	130	128	131	914

Com base nas informações anteriores, pede-se ($\alpha = 1\%$):

a) Qual o tipo de delineamento experimental que o criador utilizou? Justifique sua resposta.

b) Existe diferença entre os tipos de alimentação fornecidos às ovelhas com relação a produção de lã?

c) Com base no teste Tukey, qual(is) seria(m) o(s) tipo(s) de alimentação a ser(em) recomendada(s) às ovelhas?

6.3. Um experimento no DBC com 4 repetições forneceu os dados abaixo:

Tratamento	Blocos				Total
	1	2	3	4	
1	142,36	144,78	145,19	138,88	571,21
2	139,28	137,77	144,44	130,61	552,10
3	140,73	134,06	136,07	144,11	554,97
4	150,88	135,83	136,97	136,36	560,04
5	153,49	165,02	151,75	150,22	620,48
Total	726,74	717,46	714,42	700,18	2858,80

Para o nível de 5% de significância, pede-se:
a) ANOVA
b) Teste Tukey
c) Teste Duncan
d) Aplicar o teste Scheffé ao contraste $C = m_1 + m_2 - 2m_5$
e) Aplicar o teste t aos contrastes

$$C_1 = m_1 + m_2 - 2m_4$$
$$C_2 = m_2 + m_3 - m_1 - m_4$$
$$C_3 = m_1 - m_2$$

6.4. O resumo da Análise de Variância de um experimento instalado segundo o Delineamento em Blocos Casualizados, para verificar se existe diferença entre 5 tipos de Levedura na produção de cerveja, é fornecido a seguir:

FV	GL	QM	F
Blocos	3	---	---
Tratamentos			
Resíduo		4,895	
Total			

Totais de Tratamentos:

$$T_1 = 12{,}0 \quad T_2 = 25{,}2 \quad T_3 = 22{,}0 \quad T_4 = 24{,}0 \quad T_5 = 45{,}6$$

Ao nível de 5% de probabilidade, pede-se:

a)Existe diferença entre os 5 tipos de Levedura, na produção de cerveja?

b)Pelo teste Tukey, qual(is) o(s) tipo(s) de Levedura que apresentou(aram) maior produção?

c)Pelo teste Duncan, qual(is) o(s) tipo(s) de Levedura que apresentou(aram) menor produção?

6.5. Um Engenheiro-Agrícola, com o objetivo de verificar qual tipo de pneu que proporciona menor consumo de combustível, para trabalhar em terrenos encharcados, testou 4 diferentes tipos de pneus. Como a área que dispunha para realizar o experimento era heterogênea com relação à declividade, ele subdividiu a área total em 3 sub-áreas de tal forma que dentro de cada uma delas existia uniformidade com relação à declividade. Após isto, dentro de cada sub-área realizou um sorteio ao acaso, dos tipos de pneus às unidades experimentais. Com a realização da pesquisa, obteve-se os seguintes resultados de consumo expressos em litros/hora trabalhada.

	Pneu			
Sub-áreas	Tipo 1	Tipo 2	Tipo 3	Tipo 4
1	30	32	33	35
2	29	30	31	33
3	25	26	30	31

Por meio das informações fornecidas acima, pede-se (use o nível de 5% de significância, quando necessário).

a)Quais foram os Princípios Básicos da Experimentação utilizados neste experimento? Justifique sua resposta.

b)Qual foi o tipo de delineamento experimental utilizado pelo Engenheiro-Agrícola? Justifique sua resposta.

c)Em termos do consumo, conclua com relação aos tipos de pneus, por meio de uma análise de variância.

d)Qual tipo de pneu que proporciona o pior consumo? Use o teste Duncan, se necessário.

6.6. Suponha que alguém solicite sua ajuda, na aplicação de testes de médias aos dados de um experimento, instalado segundo o DBC com 4 repetições, para o qual o F da Análise de Variância para tratamentos foi significativo. Para tanto você recebe as seguintes informações:

Tratamentos	1	2	3
Totais	400	440	360

SQResíduo=360 $\alpha = 5\%$

$C_1 = 3m_1 - 2m_2 - 2m_3$ $C_2 = m_1 - 2m_2 + m_3$ $C_3 = m_1 - m_2$

a)Obtenha a $V(C_2)$

b)Admita que ele deseja aplicar o teste de Scheffé em C_1 e C_2. Proceda a aplicação do teste Scheffé de maneira adequada conforme visto em sala de aula.

c)Admita que ele deseja aplicar o teste t em C_2 e C_3. Proceda a aplicação do teste t de maneira adequada conforme visto em sala de aula.

d)Obtenha um grupo de contrastes ortogonais a partir apenas de C_3, usando o método do sistema de equações lineares.

6.7. Um pesquisador foi encarregado de verificar se havia diferença de durabilidade entre 4 tipos de microaspersores presentes no mercado, produzidos por duas fábricas diferentes, conforme quadro abaixo:

Tratamentos	Microaspersor	Fabricado por
1	Tipo A	Água Boa S.A.
2	Tipo B	Água Boa S.A.
3	Tipo C	Água Boa S.A.
4	Tipo Único	Água Ardente Ltda.

Desconsiderando como o experimento foi conduzido, bem como o tipo de informação usado na avaliação, considere os seguintes dados, após uma análise parcial dos mesmos:

F.V.	G.L.	Q.M.	F
Tratamentos	3	1760,00	35,2
Bloco	4	----	
Resíduo	12	50,00	
Total	19		

Médias dos Trat$^{\underline{os}}$
$\hat{m}_1 = 36$
$\hat{m}_2 = 40$
$\hat{m}_3 = 60$
$\hat{m}_4 = 40$

Com base nas informações acima pede-se: (use α=5%)

a)Cada tratamento foi repetido quantas vezes? Justifique sua resposta.

b)Que hipótese estaríamos testando pela ANOVA? Qual a sua conclusão no presente caso?

c)Para responder qual é o melhor microaspersor, o que deveríamos fazer? Apenas comente rapidamente.

d)Faça um teste (à sua escolha) para saber se há diferença entre os resultados médios apresentados pelos microaspersores da fábrica Água Boa S.A. com o apresentado pelo microaspersor da fábrica Água Ardente Ltda.

6.8. Em um experimento com 5 variedades de batatinhas (A, B, C, D e E), em blocos casualizados, as produções, em toneladas por hectare, foram:

	Variedades				
Blocos	A	B	C	D	E
1	9	21	22	15	12
2	13	27	29	11	18
3	11	26	24	10	18
4	9	25	25	12	17

Para o nível de significância igual a 5%, pede-se:

a) O quadro da ANOVA

b) Aplicar o teste de Duncan

c) Teste t para o contraste : $C = m_A + m_B - 2m_D$

6.9. Obtenha o quadro da Análise de Variância, proceda ao teste de média se necessário e conclua para $\alpha = 1\%$, considerando os dados do delineamento em blocos casualizados (DBC), fornecidos a seguir:

$$T_1 = 130{,}6 \quad T_2 = 183{,}4 \quad T_3 = 152{,}6 \quad T_4 = 185{,}6 \quad T_5 = 143{,}2$$

$$\sum_{j=1}^{4} B_j^2 = 159.306{,}92 \quad \sum_{i,j} Y_{ij}^2 = 32.889{,}70$$

6.10. Um melhorista de plantas instalou um experimento visando selecionar as melhores progênies para dar continuidade ao seu programa de melhoramento. Na instalação do experimento, ele verificou que a área a ser utilizada não era completamente homogênea. Então dividiu a área em 4 sub-áreas de tal forma que cada uma fosse completamente homogênea e pudesse conter todas as progênies em teste. Após esta divisão, as progênies foram distribuídas ao acaso dentro de cada sub-área. Na época da colheita ele avaliou a produção de grãos por planta (kg/planta), cujos resultados foram:

	Sub-áreas				
Progênie	1	2	3	4	Totais
1	2,7	2,8	2,9	3,3	11,7
2	2,7	2,5	2,8	2,4	10,4
3	2,6	3,2	3,0	3,5	12,3
4	2,6	3,1	2,8	2,5	11,0
5	2,7	2,8	2,8	2,5	10,8
Totais	13,3	14,4	14,3	14,2	56,2

Com base nestas informações, pede: (utilize α = 5% quando necessário)
6.10.1. Qual delineamento experimental foi utilizado? Justifique a sua resposta.
6.10.2. Verificar se existe diferença entre as progênies com relação à produção. Faça a ANOVA e aplique o teste de Duncan, se necessário, concluindo corretamente. **(Dado: SQTotal = 1,5780)**

6.11. Uma nutricionista formulou dois novos tipos de dieta (A e B) para diminuir o peso de pessoas obesas. Desejando verificar qual tipo de dieta proporciona maior perda de peso, resolveu fazer um teste com os seus pacientes. Com esta finalidade, solicitou que aqueles que estivessem interessados em participar deste teste se apresentassem como voluntários. Um grupo de 8 indivíduos apresentou-se para trabalhar com o nutricionista. No entanto, a nutricionista verificou que naquele grupo de indivíduos havia diferentes faixas de idade, sendo 2 indivíduos pertencentes à faixa infantil, 2 à faixa adolescente, 2 à faixa adulta e 2 à faixa idosa. Com receio de que a diferença de idade dos indivíduos pudesse diminuir a precisão do seu experimento, a nutricionista resolveu que cada um dos dois tipos de dieta fosse testado em cada uma das faixas de idade. Para tanto, dividiu o grupo de 8 indivíduos em subgrupos de tal forma que cada subgrupo incluísse indivíduos de mesma faixa de idade. Após isso, fez a distribuição dos tipos de dieta ao acaso dentro de cada subgrupo. As perdas de peso (em Kg) obtidas por cada um dos oito indivíduos são fornecidas a seguir:

	Faixa de Idade				
Dieta	Infantil	Adolescente	Adulta	Idosa	Totais
A	3	7	14	8	32
B	7	13	22	14	56
Totais	10	20	36	22	88

Dado: SQResíduo = 4,00

Com base nas informações fornecidas pede-se:
6.11.1.Qual foi a unidade experimental utilizada?
a) cada faixa de idade
b) cada dieta
c) cada indivíduo
d) todos os indivíduos
e) os dois tipos de dieta
f) nenhuma das alternativas anteriores

6.11.2. Qual(is) foi(ram) o(s) princípio(s) básico(s) da experimentação utilizado(s)?
a) repetição e casualização
b) repetição e controle local
c) casualização e controle local

d) controle local
e) repetição
f) casualização
g) controle local
h) repetição, casualização e controle local
i) nenhuma das alternativas anteriores

6.11.3 Qual foi o delineamento experimental utilizado?
a) Delineamento em Quadrado Latino
b) Delineamento Inteiramente Casualizado
c) Delineamento em Blocos Casualizados
d) Delineamento em Lático
e) nenhuma das alternativas anteriores

6.11.4. De acordo com o teste F da análise de variância para a fonte de variação dieta, pode-se concluir ao nível de 5% de probabilidade que:
a) não existe diferença entre os tipos de dieta
b) o valor de F é menor que um e não é possível concluir
c) a dieta B possui a maior média
d) nenhuma das alternativas anteriores

6.11.5 Qual o tipo de dieta deveria ser recomendado? Use o teste Duncan e o nível de 5% de probabilidade, se necessário.
a) qualquer uma das dietas
b) todas as dietas
c) nenhuma das dietas
d) a dieta B
e) a dieta A
f) nenhuma das alternativas anteriores

6.12. Um experimento instalado segundo o Delineamento em Blocos Casualizados produziu os seguintes resultados:

Tratamentos	1	2	3	4
Totais	221	220	322	151

Blocos	1	2	3	4	5	6	7
Totais	125	135	134	131	130	128	131

Dados: SQTotal = 2214,43
SQTratamentos = 2125,29

Com base nas informações fornecidas, pede-se (use o nível de significância de 1% quando necessário):
6.12.1. Conclua a respeito dos efeitos de tratamentos, com base na análise de variância.
6.12.2. Se são desejados tratamentos que propiciam **menores** médias, qual(is) tratamento(s) deve(m) ser recomendado(s)? Utilize o teste de Tukey, se necessário. Caso contrário, justifique a sua resposta.

6.13. Quatro pesquisadores realizaram um experimento com 4 tratamentos (A, B, C e D) e 5 repetições segundo um delineamento em blocos casualizados (DBC), obtendo-se as seguintes médias de tratamentos:

Tratamentos	A	B	C	D
Médias	27,8	26,0	23,8	31,4

Dados: QMRes = 1,82; GLRes = 12; F calculado = 28,34 e α = 5%.

Os procedimentos adotados por cada um dos quatro pesquisadores foram os seguintes:

- O pesquisador 1 estabeleceu, a priori, três contrastes ortogonais com o objetivo de aplicar, separadamente a cada um deles, o teste de Scheffé. Os contrastes foram os seguintes:

  $Y_1 = 3m_A - m_B - m_C - m_D$, $Y_2 = m_B + m_C - 2m_D$ e $Y_3 = m_B - m_C$

- O pesquisador 2 estabeleceu, a priori, os mesmos três contrastes ortogonais estabelecidos pelo pesquisador 1, porém com o objetivo de aplicar, separadamente a cada um deles, o teste t

- O pesquisador 3 estabeleceu seis contrastes entre duas médias ($Y_3 = m_B - m_C$ $Y_4 = m_A - m_B$, $Y_5 = m_A - m_C$, $Y_6 = m_A - m_D$, $Y_7 = m_B - m_D$ e $Y_8 = m_C - m_D$), com o objetivo de aplicar, separadamente a cada um deles, o teste de Tukey

- O pesquisador 4, após observar os dados, estabeleceu o seguinte contraste: $Y_9 = m_A + m_B - m_C - m_D$.

No entanto, os pesquisadores 1, 2 e 3, utilizando um mesmo valor para o nível de significância, apresentaram divergências com relação aos resultados das análises estatísticas para a característica estudada, conforme mostrado a seguir:

- O pesquisador 2 obteve uma conclusão diferente da encontrada pelo pesquisador 1, em relação ao contraste Y_3
- O pesquisador 3 obteve uma conclusão diferente da encontrada pelo pesquisador 2 em relação ao contraste Y_3

Pede-se:

6.13.1 Com base na diferença das conclusões encontradas pelos pesquisadores 1 e 2 em função da utilização de testes diferentes, marque a alternativa correta e **justifique a sua resposta**.

a) O procedimento adotado pelo pesquisador 1 está correto e o procedimento adotado pelo pesquisador 2 está errado
b) O procedimento adotado pelo pesquisador 2 está correto e o procedimento adotado pelo pesquisador 1 está errado
c) Ambos os procedimentos adotados pelos pesquisadores 1 e 2 estão corretos
d) Ambos os procedimentos adotados pelos pesquisadores 1 e 2 estão errados.

6.13.2. Com base na diferença das conclusões encontradas pelos pesquisadores 2 e 3 em função da utilização de testes diferentes, marque a alternativa correta e **justifique a sua resposta**.

a) O procedimento adotado pelo pesquisador 2 está correto e o procedimento adotado pelo pesquisador 3 está errado
b) O procedimento adotado pelo pesquisador 3 está correto e o procedimento adotado pelo pesquisador 2 está errado
c) Ambos os procedimentos adotados pelos pesquisadores 2 e 3 estão corretos

d) Ambos os procedimentos adotados pelos pesquisadores 2 e 3 estão errados

6.13.3. O procedimento adotado pelo pesquisador 4 foi correto? Se a sua resposta for afirmativa, aplique o teste de **Scheffé** ao contraste Y_9. Se a sua resposta for negativa, justifique a sua resposta.

6.14. Considere um experimento no delineamento em blocos casualizados, com 4 tratamentos e 3 repetições, para o qual o teste F para a fonte de variação tratamentos foi significativo ao nível de 5% de probabilidade. Baseando-se nestas informações, pede-se:

6.14.1 Qual é a fórmula geral dos contrastes a serem testados pelo teste de Tukey? Qual é o número máximo de contrastes a serem testados pelo teste de Tukey?

6.14.2. Suponha que para este experimento, a diferença mínima significativa de Tukey foi igual a 10, ou seja $\Delta = 10$. Usando-se este Δ, o seguinte resultado foi obtido para as comparações de médias de tratamentos

$\hat{m}_1 = 100$ a
$\hat{m}_4 = 92$ ab
$\hat{m}_3 = 88$ bc
$\hat{m}_2 = 79$ c

Pode-se observar que as médias 1 e 4, médias 4 e 3 e médias 3 e 2 são estatisticamente iguais. Deste modo, pode-se concluir que as médias 1 e 2 são também estatisticamente iguais? Justifique a sua resposta.

6.14.3. Caso dois outros pesquisadores realizassem o mesmo experimento e obtivessem, respectivamente $\Delta_1 = 5$ e $\Delta_2 = 20$, pelo teste de Tukey, qual dos dois pesquisadores obteve maior precisão experimental? Justifique a sua resposta.

6.14.4. Aplique o teste de Tukey às médias de tratamentos com base no $\Delta_2 = 20$, considerando que o pesquisador 2 obteve as mesmas médias listadas no item b, com teste F significativo.

6.14.5. Se o interesse fosse testar os quatro contrastes:

$$Y_1 = m_1 - m_2$$
$$Y_2 = m_1 + m_2 - 2m_3$$
$$Y_3 = m_1 + m_2 - 2m_4$$
$$Y_4 = m_1 + m_2 - m_3 - m_4$$

Qual(is) o(s) teste(s) visto(s) em sala de aula, que poderia(m) ser aplicado(s) a todos estes contrastes? Justifique a sua resposta.

7. Delineamento em Quadrado Latino

7.1. Introdução

No Delineamento em Quadrado Latino (DQL), além dos princípios da repetição e da casualização, é utilizado também duas vezes o princípio do controle na casualização para controlar o efeito de dois fatores perturbadores que causam variabilidade entre as unidades experimentais. Para controlar esta variabilidade, é necessário dividir as unidades experimentais em blocos homogêneos de unidades experimentais em relação a cada fator perturbador. O número de blocos para cada fator perturbador deve ser igual ao número de tratamentos. Por exemplo, se no experimento estão sendo avaliados I tratamentos, deve ser formado para cada fator perturbador I blocos e cada um destes blocos deve conter I unidades experimentais. Ao final são necessários I^2 unidades experimentais. Cada uma destas I^2 unidades experimentais é classificada segundo cada um dos dois fatores perturbadores.

Uma vez formados os blocos, distribui-se os tratamentos ao acaso com a restrição que cada tratamento seja designado uma única vez em cada um dos blocos dos dois fatores perturbadores.

Geralmente, na configuração de um experimento instalado segundo o DQL, os níveis de um fator perturbador são identificados por linhas em uma tabela de dupla entrada e os níveis do outro fator perturbador são identificados por colunas na tabela.

Alguns exemplos ilustrativos

Exemplo 1 - Num laboratório devem ser comparados 5 métodos de análise (A, B, C, D e E), programados em 5 dias úteis e, em cada dia, é feita uma análise a cada hora, num período de 5 horas. O quadrado latino assegura que todos os métodos sejam processados, uma vez em cada período e em cada dia. O croqui abaixo ilustra a configuração a ser adotada.

	Dia				
Período	1	2	3	4	5
1	A	E	C	D	B
2	C	B	E	A	D
3	D	C	A	B	E
4	E	D	B	C	A
5	B	A	D	E	C

Note que os níveis de uma fonte formam as linhas e os níveis da outra fonte formam as colunas

Exemplo 2 - Num experimento com suínos pretende-se testar 4 tipos de ração (A,B,C,D), em 4 raças e 4 idades de animais. Sendo interesse fundamental o comportamento dos 4 tipos de ração, toma-se a raça e a idade como blocos, ou seja:

	Raça			
Idade	R_1	R_2	R_3	R_4
I_1	A	B	D	C
I_2	B	C	A	D
I_3	D	A	C	B
I_4	C	D	B	A

Exemplo 3 - Um experimento de competição de 6 variedades de cana-de-açúcar em que a área experimental apresenta gradiente de fertilidade do solo em duas direções. O quadrado latino possibilita a formação de blocos nas duas direções, ou seja, procedemos a um duplo controle local. O croqui seguinte ilustra a distribuição das variedades (A, B, C, D, E, F) nas parcelas.

	Colunas					
Linhas	1	2	3	4	5	6
1	F	B	C	E	D	A
2	B	D	E	A	F	C
3	D	F	A	C	B	E
4	A	C	D	F	E	B
5	C	E	F	B	A	D
6	E	A	B	D	C	F

7.2. Características do DQL

a) O número total de unidades experimentais necessárias para um experimento nesse delineamento é igual a I^2, sendo I o número de tratamentos;
b) Cada tratamento é representado uma única vez e ao acaso em cada linha e em cada coluna;
c) O número de tratamentos é igual ao número de repetições;
d) Este delineamento é aconselhável quando o número de tratamentos oscila entre 3 e 10. Mas, para 3 e 4 tratamentos, somente quando se puder repetir o experimento em vários quadrados latinos.

7.3. Casualização no delineamento em quadrado latino

Consideremos 5 tratamentos: A, B, C, D, E.

1º) Faz-se a distribuição sistemática dos tratamentos dentro das linhas, de maneira que cada coluna contenha também todos os tratamentos;

	Colunas				
Linhas	1	2	3	4	5
1	A	B	C	D	E
2	E	A	B	C	D
3	D	E	A	B	C
4	C	D	E	A	B
5	B	C	D	E	A

2º) Em seguida distribui-se ao acaso as linhas entre si, e depois as colunas, podendo-se obter um quadrado final semelhante ao apresentado abaixo.

→ Casualizando as linhas (2, 4, 5, 1, 3)

E	A	B	C	D
C	D	E	A	B
B	C	D	E	A
A	B	C	D	E
D	E	A	B	C

→ Casualizando as colunas (3, 5, 1, 4, 2)

B	D	E	C	A
E	B	C	A	D
D	A	B	E	C
C	E	A	D	B
A	C	D	B	E

⇒ Quadrado final

7.4. Modelo estatístico

O delineamento em quadrado latino apresenta o seguinte modelo estatístico:

$$Y_{ij(k)} = m + l_i + c_j + t_k + e_{ij(k)} \; ,$$

em que,

$Y_{ij(k)}$ é o valor observado para a variável em estudo referente ao k-ésimo tratamento, na i-ésima linha e na j-ésima coluna;

m é média de todas as unidades experimentais para a variável em estudo;

l_i é o efeito da linha i;

c_j é o efeito da coluna j;

t_k é o efeito do tratamento k;

$e_{ij(k)}$ é o erro experimental.

Admitindo-se I tratamentos, conseqüentemente I linhas e I colunas, o esquema da análise de variância fica:

FV	GL	Exemplo 1	Exemplo 2	Exemplo 3
Linhas	I-1	4	3	5
Colunas	I-1	4	3	5
Tratamentos	I-1	4	3	5
Resíduo	(I-1)(I-2)	12	6	20
Total	I^2-1	24	15	35

Considerando

L_i = Total da linha i;
C_j = Total da coluna j;
T_k = Total do tratamento k;
G = total geral;

as somas de quadrados são dadas por:

$$SQTotal = \sum_{i,j} Y_{ij}^2 - C, \quad \text{onde} \quad C = \frac{G^2}{I \cdot I} = \frac{G^2}{I^2}$$

$$SQLinhas = \frac{1}{I}\sum_{i=1}^{I} L_i^2 - C$$

$$SQColunas = \frac{1}{I}\sum_{j=1}^{J} C_j^2 - C$$

$$SQTratamentos = \frac{1}{I}\sum_{k=1}^{K} T_i^2 - C$$

$$SQ\,Re\,siduo = SQTotal - SQL - SQC - SQT \;.$$

7.5. Exercícios

7.1. Num experimento de competição de variedades de cana forrageira foram usadas 5 variedades: A=CO290; B=CO294; C=CO297; D=CO299 e E=CO295, dispostas em um quadrado latino 5x5. O controle feito através de blocos horizontais e verticais teve por objetivo eliminar influências devidas a diferenças de fertilidade em duas direções. As produções, em kg/parcela, foram as seguintes:

	Colunas					
Linhas	1	2	3	4	5	Totais
1	432(D)	518(A)	458(B)	583(C)	331(E)	2322
2	724(C)	478(E)	524(A)	550(B)	400(D)	2676
3	489(E)	384(B)	556(C)	297(D)	420(A)	2146
4	494(B)	500(D)	313(E)	486(A)	501(C)	2294
5	515(A)	660(C)	438(D)	394(E)	318(B)	2325
Totais	2654	2540	2289	2310	1970	11763

Considerando $\alpha = 5\%$, pede-se:

a. Análise de Variância

b. Qual a variedade a ser recomendada? Utilize teste de Tukey, se necessário.

7.2. Em um experimento no delineamento em quadrado latino com 5 tratamentos, são dados:

$\hat{m}_1 = 50{,}0;\quad \hat{m}_2 = 60{,}0;\quad \hat{m}_3 = 47{,}5;\quad \hat{m}_4 = 40{,}0;\quad \hat{m}_5 = 52{,}5 \qquad SQ\,Resíduo = 388{,}80$

a. Verificar se existe efeito significativo de tratamentos, pelo teste F, e concluir para $\alpha = 5\%$.

b. Qual o tratamento deve ser recomendado nos seguintes casos:

b.1. Se estivéssemos avaliando a produção de uma certa cultura (em kg/ha)?

b.2. Se estivéssemos avaliando a perda de grãos, durante a colheita, de uma certa cultura (em g/parcela)?

Obs.: Utilize $\alpha = 5\%$ e o Teste de Duncan (se necessário)

7.3. Aplicar o teste de Tukey para comparar as médias de tratamentos, relativos ao Quadrado Latino 5x5, dados:

$T_1 = 3024{,}0;\quad T_2 = 2549{,}0;\quad T_3 = 2349{,}0;\quad T_4 = 1970{,}0;\quad T_5 = 1734{,}0$

$SQ\,Re\,siduo = 34116{,}0 \qquad \alpha = 5\%$

7.4. O objetivo de um experimento foi estudar o efeito da época de castração no desenvolvimento e produção de suínos. Dispunha-se para esse estudo, de 5 matrizes da mesma raça, que foram submetidas à mesma alimentação e manejo durante o período de gestação. Os tratamentos foram: (A) Castração aos 56 dias de idade; (B) Castração aos 7 dias de idade; (C) Castração aos 36 dias de idade; (D) Inteiros; (E) Castração aos 21 dias de idade. Foi utilizado o delineamento em quadrado latino buscando controlar a variação entre leitegadas (linhas) e a variação no peso inicial dos leitões (colunas), sendo a parcela experimental constituída de um leitão. Os ganhos de pesos, em kg, após o período experimental (28 semanas), estão apresentados no quadro abaixo:

Leitegadas	Faixas de Peso Inicial 1	2	3	4	5	Totais
1	93,0(A)	115,4(C)	116,9(E)	110,2(D)	110,4(B)	545,9
2	110,6(C)	96,5(E)	108,9(B)	97,6(A)	112,0(D)	525,6
3	102,1(B)	108,6(D)	77,9(A)	102,0(E)	111,7(C)	502,3
4	115,4(D)	94,9(A)	114,0(C)	100,2(B)	118,5(E)	543,0
5	117,6(E)	114,1(B)	118,7(D)	108,8(C)	80,2(A)	539,4
Totais	538,7	529,5	536,4	518,8	532,8	2656,2

Considerando $\alpha = 5\%$, pede-se:

a. Faça a análise de variância. DADO: $\text{SQTotal} = 2998{,}4824$
b. Formule um contraste que permita avaliar o efeito médio da prática de castração.
c. Teste o contraste obtido no item anterior. Utilize os teste de Scheffé e t.

7.5. Um experimento foi conduzido numa região do Pantanal com o objetivo de selecionar forrageiras que garantissem uma maior produção de matéria seca. Foi utilizado o delineamento em quadrado latino, buscando controlar diferenças de fertilidade em duas direções, sendo avaliadas 7 forrageiras (A, B, C, D, E, F, G). Foram obtidos os seguintes resultados parciais com a realização do experimento:

Tratamentos	A	B	C	D	E	F	G
Totais	30,8	25,2	19,6	14,0	13,3	9,8	8,4

Linhas	1	2	3	4	5	6	7
Totais	18,9	19,9	14,5	18,1	15,6	17,4	16,7

SQTotal=72,36 SQColunas=1,27

Verificar se existe efeito significativo de forrageiras, pelo teste F, e concluir para α=1%.

7.6. Um pesquisador instalou um experimento para comparar 5 tipos de bacilos (A, B, C, D, e E) usados para produção de iogurte. No momento da instalação do experimento, o pesquisador verificou que o material experimental disponível (25 unidades de 1 litro de leite) não era completamente homogêneo entre si, pois apresentavam variação quanto ao teor de gordura e grau de acidez. Para controlar estas duas fontes de variação, o pesquisador distribuiu os bacilos ao acaso às amostras de leite de tal forma que cada bacilo pudesse ser testado em todas as condições de teor de gordura e grau de acidez. O quadro dado a seguir ilustra a distribuição dos bacilos às amostras de leite bem como o volume (em ml) de iogurte produzido:

	Grau de Acidez					
Teor de Gordura	1	2	3	4	5	Totais
1	450 A	620 E	680 C	620 D	780 B	3150
2	750 C	990 B	750 E	660 A	830 D	3980
3	750 D	910 C	690 A	990 B	760 E	4100
4	650 E	890 D	835 B	850 C	875 A	4100
5	750 B	720 A	850 D	770 E	890 C	3980
Totais	3350	4130	3805	3890	4135	19310

$T_A = 3395$ $T_B = 4345$ $T_C = 4080$ $T_D = 3940$ $T_E = 3550$

Com base nas informações fornecidas, pergunta-se:

7.6.1. Qual foi a unidade experimental utilizada?
7.6.2. Quais foram os tratamentos em teste?
7.6.3. Quantas vezes o princípio do controle local foi utilizado neste experimento?
7.6.4. Qual foi o Delineamento experimental utilizado nesta pesquisa?
7.6.5. Usando os dados experimentais fornecidos anteriormente e o teste F para testar a fonte de variação bacilos, pode-se concluir que ao nível de 5% de probabilidade que

a) existe pelo menos um contraste entre médias de bacilos estatisticamente diferente de zero
b) todos os possíveis contrastes entre médias de bacilos são estatisticamente nulos
c) o bacilo A é o melhor
d) o bacilo B é o melhor
e) o bacilo C é o melhor
f) nenhuma das alternativas anteriores

7.6.6. O teste de Tukey indica que o(s) bacilo(s) que proporciona(m) maior(es) média(s) de produção de iogurte é (são) (use o nível de 5% de significância) foi(ram)

a) o bacilo A
b) o bacilo B
c) o bacilo C
d) o bacilo D
e) o bacilo E
f) os bacilos A, B e C
g) os bacilos B, C e D
h) os bacilos C, D e E
i) os bacilos A, D e E
j) nenhuma das alternativas anteriores

8. Experimentos Fatoriais

8.1. Introdução

Experimentos fatoriais são aqueles em que se estudam simultaneamente dois ou mais fatores, cada um deles com dois ou mais níveis. O fatorial é um tipo de esquema, ou seja, uma das maneiras de organizar os tratamentos e não um tipo de delineamento, que representa a maneira pela qual os tratamentos são distribuídos às unidades experimentais. Na verdade, os experimentos fatoriais são montados segundo um tipo de delineamento experimental, como por exemplo: o DIC e o DBC.

Nos experimentos fatoriais, os tratamentos são obtidos pelas combinações dos níveis dos fatores. Num experimento fatorial completo, cada nível de um fator combina com todos os níveis dos outros fatores. A principal aplicação de experimentos fatoriais é quando se quer saber sobre o efeito de diversos fatores que influenciam na variável em estudo e o relacionamento entre eles.

A simbologia comumente utilizada, para experimentos fatoriais é indicar o produto dos níveis dos fatores em teste. Por exemplo: Experimento Fatorial 2x4x6. O produto 2x4x6 informa que no experimento foram testados simultaneamente 3 fatores. O primeiro possui 2 níveis, o segundo 4 níveis e o terceiro 6 níveis. Quando o número de níveis é igual para todos os fatores, pode-se utilizar a seguinte simbologia: n^F, em que F é o número de fatores n é o número de níveis de cada fator. Por exemplo: Experimento Fatorial 4^3. A potência 4^3 informa que o experimento tem 3 fatores com 4 níveis cada um.

8.2. Tipos de efeitos avaliados em um experimento fatorial

Nos experimentos fatoriais, podem ser estudados os seguintes efeitos:

- Efeito Principal: é o efeito de cada fator, independente do efeito dos outros fatores;

- Efeito de Interação: é o efeito simultâneo dos fatores sobre a variável em estudo. Dizemos que ocorre interação entre os fatores quando os efeitos dos níveis de um fator são modificados pelos níveis do outro fator.

O efeito da interação pode ser mais facilmente compreendido por meio de gráficos. Para ilustrar o efeito da interação, considere um experimento fatorial 3x2, em que os fatores em testes são Variedade (V) e Espaçamento (E). Os tratamentos para este experimento são os seguintes:

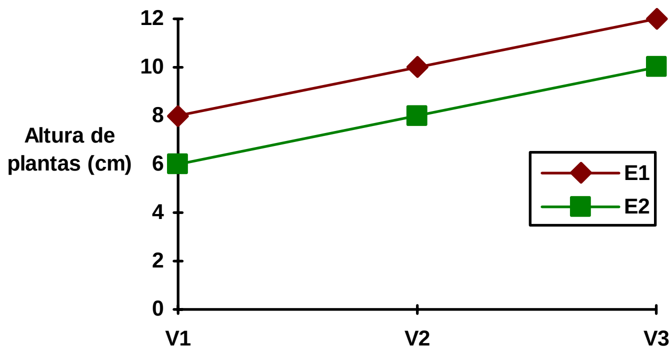
V1E1	V2E1	V3E1
V1E2	V2E2	V3E2

Suponha os seguintes resultados fictícios, para a variável altura de plantas (cm), deste experimento, nas seguintes situações:

1) Não há interação

Espaçamentos	Variedades		
	V1	V2	V3
E1	8	10	12
E2	6	8	10

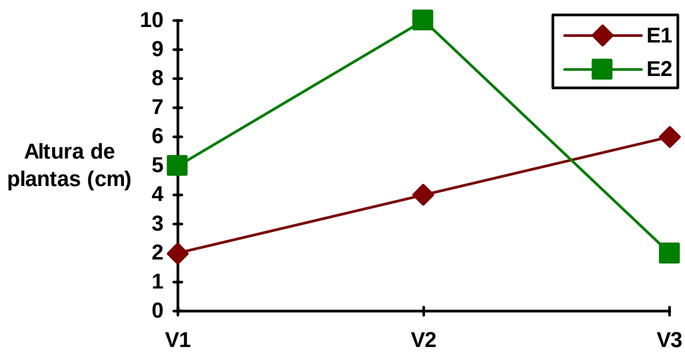
Quando não há interação as diferenças entre os resultados dos níveis de um fator são estatisticamente iguais para todos os níveis do outro fator.



2) Há interação

Espaçamentos	Variedades		
	V1	V2	V3
E1	2	4	6
E2	5	10	2

Quando há interação as diferenças entre os níveis de um fator dependem dos níveis do outro fator.



8.3. Quadro de tabulação de dados

Uma maneira de tabular os dados de um experimento fatorial, com dois fatores A e B, com I e níveis, respectivamente, instalados segundo o DIC, com K repetições, é fornecida a seguir:

	A1				A2				...	AI			
Repetição	B1	B2	...	BJ	B1	B2	...	BJ	...	B1	B2	...	BJ
1	Y_{111}	Y_{121}	...	Y_{1J1}	Y_{211}	Y_{221}	...	Y_{2J1}	...	Y_{I11}	Y_{I21}	...	Y_{IJ1}
2	Y_{112}	Y_{122}	...	Y_{1J2}	Y_{212}	Y_{222}	...	Y_{2J2}	...	Y_{I12}	Y_{I22}	...	Y_{IJ2}
...	...	...	...	...	...	...	...	...	...	...	...	...	...
K	Y_{11K}	Y_{12K}	...	Y_{1JK}	Y_{21K}	Y_{22K}	...	Y_{2JK}	...	Y_{I1K}	Y_{I2K}	...	Y_{IJK}
Total	$Y_{11\bullet}$	$Y_{12\bullet}$	...	$Y_{1J\bullet}$	$Y_{21\bullet}$	$Y_{22\bullet}$	...	$Y_{2J\bullet}$	...	$Y_{I1\bullet}$	$Y_{I2\bullet}$	...	$Y_{IJ\bullet}$

Deste quadro, pode-se tirar algumas informações que posteriormente serão úteis na análise de variância:

- Total do ij-ésimo tratamento: $(AB)_{ij} = \sum_{k=1}^{K} Y_{ijk} = Y_{ij\bullet}$
- Total do i-ésimo nível do fator A: $A_i = \sum_{j=1,k=1}^{J,K} Y_{ijk} = Y_{i\bullet\bullet}$
- Total do j-ésimo nível do fator B: $B_j = \sum_{i=1,k=1}^{I,K} Y_{ijk} = Y_{\bullet j\bullet}$
- Total Geral: $G = \sum_{i=1,j=1,k=1}^{I,J,K} Y_{ijk} = \sum_{i=1}^{I} A_i = \sum_{j=1}^{J} B_j = Y_{\ldots}$
- Média do i-ésimo nível do fator A: $\hat{m}_{A_i} = \dfrac{A_i}{JK}$
- Média do j-ésimo nível do fator B: $\hat{m}_{B_i} = \dfrac{B_j}{IK}$,
- Média geral: $\hat{m} = \dfrac{G}{N}$
- Número total de parcelas: N=IJK
-

Pode-se montar um quadro auxiliar contendo os totais de tratamentos, cujos valores são obtidos pela soma de todas as repetições para o tratamento em questão. Este quadro facilita o cálculo das somas de quadrados devido aos fatores A e B, e da interação entre eles. Para a situação citada, o quadro de totais de tratamentos é do seguinte tipo:

Fator A	Fator B				Totais
	B1	B2	...	BJ	
A1	$Y_{11.}$	$Y_{12.}$	...	$Y_{1J.}$	A_1
A2	$Y_{21.}$	$Y_{22.}$	...	$Y_{2J.}$	A_2
...	...	...	...	...	...
AI	$Y_{I1.}$	$Y_{I2.}$	...	$Y_{IJ.}$	A_I
Totais	B_1	B_2	...	B_j	G

8.4. Modelo estatístico

Considere um experimento fatorial, com dois fatores: o fator A com I níveis e o fator B com J níveis, instalados segundo o DIC, com K repetições. O modelo estatístico para um experimento como este é:

$$Y_{ijk} = m + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$$

em que,

Y_{ijk} é o valor observado para a variável em estudo referente a k-ésima repetição da combinação do i-ésimo nível do fator A com o j-ésimo nível do fator B;

m é a média de todas as unidades experimentais para a variável em estudo;

α_i é o efeito do i-ésimo nível do fator A no valor observado Y_{ijk};

β_j é o efeito do j-ésimo nível do fator B no valor observado Y_{ijk};

$(\alpha\beta)_{ij}$ é o efeito da interação do i-ésimo nível do fator A com o j-ésimo nível do fator B;

e_{ijk} é o erro associado a observação Y_{ijk}.

Para um experimento fatorial instalado segundo o DBC, com K blocos, o modelo estatístico seria:

$$Y_{ijk} = m + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \omega_k + e_{ijk}$$

em que,

ω_k é o efeito do k-ésimo bloco na observação Y_{ijk}.

8.5. Análise de Variância

A análise de variância de um experimento fatorial é feita desdobrando-se a soma de quadrados de tratamentos nas partes devido aos efeitos principais de cada fator e na parte devido à interação entre os fatores.

O quadro a seguir apresenta como seria a análise de um experimento fatorial, com 2 fatores A e B, com I e J níveis, respectivamente, e K repetições, instalado segundo o DIC.

FV	GL	SQ	QM	F	$F_{tab;\alpha}$
A	(I-1)	SQA	-	-	-
B	(J-1)	SQB	-	-	-
AxB	(I-1)(J-1)	SQAxB	$\frac{SQAxB}{(I-1)(J-1)}$	$\frac{QMAxB}{QMRes}$	[(I-1)(J-1); n_2]
(Trat)	(IJ-1)	(SQTrat)	-	-	-
Resíduo	n_2=IJ(K-1)	SQRes	$\frac{SQRes}{IJ(K-1)}$	-	-
Total	IJK – 1	SQTotal	-	-	

As fórmulas para a obtenção das somas de quadrados são as seguintes:

$$SQTotal = \sum_{i=1,j=1,k=1}^{I,J,K} Y_{ijk}^2 - C$$

$$C = \frac{\left(\sum_{i=1,j=1,k=1}^{I,J,K} Y_{ijk}\right)^2}{IJK}$$

$$SQTrat = \sum_{i=1,j=1}^{I,J} \frac{Y_{ij.}^2}{K} - C$$

$$SQA = \sum_{i=1}^{I} \frac{A_i^2}{JK} - C$$

$$SQB = \sum_{j=1}^{J} \frac{B_j^2}{IK} - C$$

SQAxB = SQTrat – SQA – SQB

SQResíduo = SQTotal – SQTrat

O quadro abaixo apresenta como seria a análise de um experimento fatorial, com 2 fatores A e B, com I e J níveis, respectivamente, e K repetições (ou blocos), instalado segundo o DBC.

FV	GL	SQ	QM	F	$F_{tab,\alpha}$
A	(I-1)	SQA	-	-	-
B	(J-1)	SQB	-	-	-
AxB	(I-1)(J-1)	SQAxB	$\frac{SQAxB}{(I-1)(J-1)}$	$\frac{QMAxB}{QMRes}$	[(I-1)(J-1); n_2]
(Trat)	(IJ-1)	(SQTrat)	-	-	-
Blocos	K-1	SQBlocos	-	-	-
Resíduo	n_2=(IJ-1)(K-1)	SQRes	$\frac{SQRes}{(IJ-1)(K-1)}$	-	-
Total	IJK - 1	SQTotal	-	-	-

Nesta situação,

$$SQBlocos = \sum_{k=1}^{K} \frac{W_k^2}{IJ} - C$$

em que,

→ Total do k-ésimo bloco: $W_k = \sum_{i=1,j=1}^{I,J} Y_{ijk} = Y_{\bullet\bullet k}$

Conforme apresentado nas duas tabelas anteriores, na análise dos dados oriundos de um experimento fatorial, para os dois tipos de delineamentos, deve-se inicialmente proceder ao teste F para a interação entre os fatores. As hipóteses para o teste F da interação são:

H_0 : Os fatores A e B atuam independentemente sobre a variável resposta em estudo.

H_a : Os fatores A e B não atuam independentemente sobre a variável resposta em estudo.

O resultado deste teste F para a interação indica como as comparações dos níveis de um fator devem ser realizadas. Temos dois resultados possíveis para o teste F da interação os quais serão apresentados a seguir.

8.5.1 Interação não-significativa

Este caso ocorre quando a hipótese H_0 para a interação entre os fatores não é rejeitada. Este resultado implica que os efeitos dos fatores atuam de forma independente.

Portanto recomenda-se que as comparações dos níveis de um fator sejam feitas de forma geral em relação ao outro fator, ou seja, independente dos níveis outro fator. O passo seguinte na análise estatística dos dados experimentais é proceder ao teste F para cada fator como ilustrado na tabela apresentada a seguir para o caso do DBC.

FV	GL	SQ	QM	F	$F_{tab,\ \alpha}$
A	(I-1)	SQA	$\frac{SQA}{(I-1)}$	$\frac{QMA}{QMRes}$	[(I-1);n2]
B	(J-1)	SQB	$\frac{SQB}{(J-1)}$	$\frac{QMB}{QMRes}$	[(J-1);n2]
AxB	(I-1)(J-1)	SQAxB	$\frac{SQAxB}{(I-1)(J-1)}$	não-siginificativo	-
(Trat)	(IJ-1)	(SQTrat)	-	-	-
Blocos	K-1	SQBlocos	-	-	-
Resíduo	n_2=(IJ-1)(K-1)	SQRes	$\frac{SQRes}{(IJ-1)(K-1)}$	-	-
Total	IJK - 1	SQTotal	-	-	-

As hipóteses para realizar o teste F para os efeitos principais são

Fator A

$H_0: m_{A1} = m_{A2} = ... = m_{AI}$ ou seja, todos os possíveis contrastes entre as médias dos níveis do fator A, são estatisticamente nulos, ao nível de probabilidade em que foi executado o teste.

H_a :não H_0 ou seja, existe pelo menos um contraste entre as médias dos níveis do fator A, que é estatisticamente diferente de zero, ao nível de probabilidade em que foi executado o teste.

Fator B

$H_0: m_{B1} = m_{B2} = ... = m_{BJ}$ ou seja, todos os possíveis contrastes entre as médias dos níveis do fator B, são estatisticamente nulos, ao nível de probabilidade em que foi executado o teste.

H_a :não H_0 ou seja, existe pelo menos um contraste entre as médias dos níveis do fator B, que é estatisticamente diferente de zero, ao nível de probabilidade em que foi executado o teste.

Se os fatores A e B forem qualitativos, e o teste F para A e/ou B, for não significativo, a aplicação do teste de médias é desnecessária. Se o teste F for significativo, para A e/ou B, aplica-se um teste de médias para comparar os níveis do fator. As estimativas das médias dos níveis dos fatores são obtidas por

Fator A $\rightarrow \hat{m}_{Ai} = \dfrac{A_i}{JK}$

Fator B $\rightarrow \hat{m}_{Bj} = \dfrac{B_j}{IK}$

Para realizar o teste de Tukey para comparar as medias dos níveis dos fatores em teste temos que usar

	Δ	q_α
A	$q\sqrt{\dfrac{QMRes}{JK}}$	$(I;n_2)$
B	$q\sqrt{\dfrac{QMRes}{IK}}$	$(J;n_2)$

Para o teste de Duncan temos que usar

	D_i	z_α
A	$z_i\sqrt{\dfrac{QMRes}{JK}}$	$(n_A;n_2)$
B	$z_i\sqrt{\dfrac{QMRes}{IK}}$	$(n_B;n_2)$

Em que n_A e n_B são os números de médias ordenadas abrangidas pelo contraste sendo testados.

As hipóteses para os testes Tukey e Duncan para comparar as médias dos níveis dos fatores são

Fator A $\rightarrow H_0 : m_{Ai} = m_{Au}$ versus $H_a : m_{Ai} \neq m_{Au}$ para $i \neq u = 1, 2, 3, \ldots, I$
Fator B $\rightarrow H_0 : m_{Bj} = m_{Bu}$ versus $H_a : m_{Bj} \neq m_{Bu}$ para $j \neq u = 1, 2, 3, \ldots, J$

Para a aplicação do teste t temos que usar

	t	t_{tab}
A	$\dfrac{\hat{C}_A - C_A}{\sqrt{\dfrac{QMRes}{JK}\sum_{i=1}^{I} a_i^2}}$	$t_\alpha\ (n_2)$
B	$\dfrac{\hat{C}_B - C_B}{\sqrt{\dfrac{QMRes}{IK}\sum_{j=1}^{J} b_j^2}}$	$t_\alpha\ (n_2)$

Em que

$C_A = a_1m_{A1} + a_2m_{A2} + \ldots + a_Im_{AI}$ e

$C_B = b_1m_{B1} + b_2m_{B2} + \ldots + b_Im_{BJ}$

Para a aplicação do teste Scheffé para testar os contrastes Y_A e Y_B temos que usar

	S	F_{tab}
A	$S=\sqrt{(I-1)F_{tab}\frac{QMRes}{JK}\sum_{i=1}^{I}a_i^2}$	F_α [(I -1); n_2]
B	$S=\sqrt{(J-1)F_{tab}\frac{QMRes}{IK}\sum_{j=1}^{J}b_j^2}$	F_α [(J -1); n_2]

As hipóteses para os testes de Scheffé e t para testar os contrastes são

Fator A $\rightarrow H_0 : C_A = 0$ versus $H_a : C_A \neq 0$
Fator B $\rightarrow H_0 : C_B = 0$ versus $H_a : C_B \neq 0$

8.5.2 Interação significativa

Este caso ocorre quando a hipótese H_0 para a interação entre os fatores é rejeitada. Este resultado implica que os efeitos dos fatores atuam de forma dependente. Neste caso as comparações entre os níveis de um fator levam em consideração o nível do outro fator, pois o resultado significativo para a interação indica que o efeito de um fator depende do nível do outro fator.

Portanto, não é recomendado realizar o teste F para cada fator isoladamente tal como foi apresentado para o caso da interação não-significativa. O procedimento recomendado é realizar o desdobramento do efeito da interação.

Para realizar este desdobramento deve-se fazer uma nova análise de variância em que os níveis de um fator são comparados dentro de cada nível do outro fator, tal como apresentado nas tabelas a seguir.

Desdobramento para comparar os níveis de A dentro de cada nível de B, ou seja, estudar A/B

FV	GL	SQ	QM	F	$F_{tab,\ \alpha}$
A/B1	(I-1)	SQA/B1	$\frac{SQA/B1}{(I-1)}$	$\frac{QMA/B1}{QMRes}$	[(I-1);n_2]
A/B2	(I-1)	SQA/B2	$\frac{SQA/B2}{(I-1)}$	$\frac{QMA/B2}{QMRes}$	[(I-1);n_2]
...	...	...	...	...	...
A/BJ	(I-1)	SQA/BJ	$\frac{SQA/BJ}{(I-1)}$	$\frac{QMA/BJ}{QMRes}$	[(I-1);n_2]
Resíduo	n_2		QMRes	-	
Total	IJK - 1	SQTotal	-	-	-

As hipóteses para testar as fontes de variação da tabela acima, para j=1, 2, 3, ..., J, são

$H_0 : m_{A1/Bj} = m_{A2/Bj} = ... = m_{AI/Bj}$
H_a :não H_0

Desdobramento para comparar os níveis de B dentro de cada nível de A, ou seja estudar B/A

FV	GL	SQ	QM	F	$F_{tab,\,\alpha}$
B/A1	(J-1)	SQB/A1	$\frac{SQB/A1}{(J-1)}$	$\frac{QMB/A1}{QMRes}$	$[(J-1);n_2]$
B/A2	(J-1)	SQB/A2	$\frac{SQB/A2}{(J-1)}$	$\frac{QMB/A2}{QMRes}$	$[(J-1);n_2]$
...	...	...	...	...	...
B/AI	(J-1)	SQB/AI	$\frac{SQB/AI}{(J-1)}$	$\frac{QMB/AI}{QMRes}$	$[(J-1);n_2]$
Resíduo	n_2		QMRes	-	
Total	IJK - 1	SQTotal	-	-	-

As hipóteses para testar as fontes de variação da tabela acima, para i=1, 2, 3, ..., I, são

$H_0 : m_{B1/Ai} = m_{B2/Ai} = ... = m_{BJ/Ai}$
H_a :não H_0

Em que as SQA/Bj e SQB/Ai podem ser obtidas usando a fórmula geral para a soma de quadrados dada por

$$SQ = \sum_{i=1}^{k} \frac{X_i^2}{r_i} - \frac{\left(\sum_{i=1}^{k} X_i\right)^2}{\sum_{i=1}^{k} r_i}$$

Se os fatores forem qualitativos, procede-se ao teste F para cada fonte de variação do desdobramento. Nas fontes de variação em que o teste F foi significativo e o fator tem mais de dois níveis, recomenda-se a aplicação de um teste de médias. As estimativas das médias dos níveis dos fatores são obtidas por

Fator A $\rightarrow \hat{m}_{Ai} = \frac{A_i}{K}$

Fator B $\rightarrow \hat{m}_{Bj} = \frac{B_j}{K}$

Para realizar o teste de Tukey para comparar as médias dos níveis dos fatores em teste temos que usar

	Δ	q_α
A	$q\sqrt{\frac{QMRes}{K}}$	$(I;n_2)$
B	$q\sqrt{\frac{QMRes}{K}}$	$(J;n_2)$

Para o teste de Duncan temos que usar

	D_i	z_α
A	$z_i\sqrt{\frac{QMRes}{K}}$	$(n_A;n_2)$
B	$z_i\sqrt{\frac{QMRes}{K}}$	$(n_B;n_2)$

Em que n_A e n_B são os números de médias ordenadas abrangidas pelo contraste sendo testados.

As hipóteses para os testes Tukey e Duncan para comparar as médias dos níveis dos fatores são

Fator A $\rightarrow$ $H_0 : m_{Ai/Bj} = m_{Au/Bj}$ versus $H_a : m_{Ai/Bj} \neq m_{Au/Bj}$ para $i \neq u = 1, 2, 3, \ldots, I$ e $j = 1, 2, \ldots, J$

Fator B $\rightarrow$ $H_0 : m_{Bj/Ai} = m_{Bu/Ai}$ versus $H_a : m_{Bj/Ai} \neq m_{Bu/Ai}$ para $j \neq u = 1, 2, 3, \ldots, J$ e $i = 1, 2, \ldots, I$

Para a aplicação do teste t temos que usar

	t	t_{tab}
A	$\dfrac{\hat{C}_A - C_A}{\sqrt{\frac{QMRes}{K}\sum_{i=1}^{I} a_i^2}}$	$t_\alpha\ (n_2)$
B	$\dfrac{\hat{C}_B - C_B}{\sqrt{\frac{QMRes}{K}\sum_{j=1}^{JI} b_i^2}}$	$t_\alpha\ (n_2)$

Em que

$C_A = a_1m_{A1/Bj} + a_2m_{A2/Bj} + \ldots + a_Im_{AI/Bj}$ para $j = 1, 2, \ldots, J$ e

$C_B = b_1m_{B1/Ai} + b_2m_{B2/Ai} + \ldots + b_Jm_{BJ/Ai}$ para $i = 1, 2, \ldots, I$

Para a aplicação do teste Scheffé para testar os contrastes C_A e C_B temos que usar

	S	F_{tab}
A	$S = \sqrt{(I-1)F_{tab}\frac{QMRes}{K}\sum_{i=1}^{I} a_i^2}$	$F_\alpha\ [(I-1);\ n_2]$
B	$S = \sqrt{(J-1)F_{tab}\frac{QMRes}{K}\sum_{j=1}^{J} b_j^2}$	$F_\alpha\ [(J-1);\ n_2]$

As hipóteses para os testes de Scheffé e t para testar os contrastes são

Fator A $\rightarrow$ $H_0 : C_A = 0$ versus $H_a : C_A \neq 0$

Fator B $\rightarrow$ $H_0 : C_B = 0$ versus $H_a : C_B \neq 0$

8.6. Vantagens e desvantagens de um experimento fatorial

8.6.1 Vantagens

a. Permite o estudo dos efeitos principais e o efeito da interação entre os fatores.
b. O nº de graus de liberdade associado ao resíduo é alto quando comparado com os experimentos simples dos mesmos fatores, o que contribui para diminuir a variância residual, aumentando a precisão do experimento.

8.6.2 Desvantagem

b. Requer maior número de unidades experimentais em relação aos experimentos simples.

8.7. Exercícios

8.1. Seja um experimento fatorial instalado no DIC, com dois fatores: Irrigação (A) e Calagem (B), cada um deles com dois níveis: presença $(A_1 \text{ e } B_1)$ e ausência $(A_0 \text{ e } B_0)$. Os dados obtidos (kg de planta/parcela) para cada tratamento são fornecidos abaixo. Pede-se realizar a ANOVA e obter as conclusões sobre os fatores. Use $\alpha = 5\%$.

A0B0	A0B1	A1B0	A1B1
25	35	41	60
32	28	35	67
27	33	38	59

8.2. Em um experimento fatorial no DIC em que foram combinadas duas doses de N e duas doses de fósforo, com 5 repetições, são dados:

	P0					P1				
N0	10,5	11,0	9,8	11,2	9,9	11,2	11,0	10,4	13,1	10,6
N1	11,5	12,4	10,2	12,7	10,4	14,0	14,1	13,8	13,5	14,2

Considerando o nível de significância de 5%, concluir sobre os efeitos dos fatores.

8.3. Foi realizada uma pesquisa para testar dois tipos de ambiente (com luz artificial e sem luz artificial no período da noite) e dois tipos de ração (com cálcio e sem cálcio). Para tanto foram utilizadas 24 poedeiras similares, escolhidas aleatoriamente. Ao final da avaliação foram obtidos os seguintes resultados (ovos/poedeira):

	Ambiente à noite											
Ração	com luz artificial						sem luz artificial					
com cálcio	50	52	48	54	52	50	49	52	50	48	46	45
sem cálcio	42	44	46	43	44	45	40	40	38	39	41	43

Ao nível de 1% de probabilidade e admitindo que se trata de um experimento instalado segundo o DIC, pede-se:

a) Pode-se afirmar que o tipo de Ração e o tipo de Ambiente atuam independentemente na produção de ovos?

b) Qual seria o tipo de Ração recomendada? (Use o teste Tukey se necessário).

c) Qual seria o tipo de Ambiente recomendado? (Use o teste Tukey se necessário).

8.4. Para um experimento montado no DBC e que se pretendia verificar o efeito dos fatores tipo de vasilhame e tipo de fonte nutritiva no crescimento de colônias bacterianas em laboratório, foram obtidos os seguintes resultados, após o término da realização do experimento:

Totais de Tratamento para o nº de colônias bacterianas

Vasilhame (V)	Fonte nutritiva (F) a base de			Total
	N	P	K	
Tubo de Ensaio	25	30	10	65
Placa de Petri	20	15	40	75
Total	45	45	50	140

Resumo da ANOVA

FV	GL	SQ	QM	F
V				
F				
VxF				
Blocos	4		1,72	
Resíduo			2,05	
Total				

Com base nos resultados fornecidos acima, pede-se, ao nível de 5% de probabilidade:

a) Os fatores fonte nutritiva e vasilhame atuam independentemente no nº de colônias bacterianas? Justifique sua resposta.

b) Qual a melhor fonte nutritiva para o vasilhame placa de petri? (Use o teste Tukey, se necessário).

c) Qual o melhor vasilhame para a fonte nutritiva a base de K? (Use o teste Tukey, se necessário).

8.5. Um experimento, instalado segundo o Delineamento Inteiramente Casualizado com 5 repetições, com o objetivo de verificar o efeito de 2 cultivares de Eucalipto e de 2 espaçamentos na produção de carvão, forneceram os seguintes resultados:

Totais de Tratamentos

Espaçamentos	Variedades V1	Variedades V2	Totais
E1	30	35	65
E2	38	39	77
Totais	68	74	142

SQResíduo = 17,00

Usando o nível de significância de 5% e aplicando o <u>teste Tukey</u> quando necessário, pede-se:

a) Os fatores variedades e espaçamentos atuam independentemente na produção de carvão?

b) Qual foi a variedade que forneceu a <u>menor</u> produção?

c) Qual foi o espaçamento que forneceu a <u>maior</u> produção?

8.6. Abaixo são fornecidos o Quadro da Análise de Variância e o Quadro de Interação para um experimento fatorial instalado segundo o Delineamento em Blocos Casualizados, com 4 repetições, que foi realizado por um zootecnista para comparar 3 raças de suínos e 2 tipos de rações com relação ao teor de gordura na carcaça.

Totais de Tratamentos

Raça	Ração 1	Ração 2	Totais
1	45	40	85
2	38	45	83
3	39	48	87
Totais	122	133	255

FV	GL	SQ	QM	F
Ração		5,0400	-	
Raça		1,0000		
Interação				
(Tratamentos)		(20,3750)		
Blocos		-		
Resíduo		15,0000		
Total				

Ao nível de 5% de probabilidade, pede-se:

a)Os fatores Raça e Ração atuam independentemente no teor de gordura dos suínos?

b)Proceda a análise do fator Ração, da maneira adequada, conforme o resultado obtido para o teste F da Análise de Variância para a Interação Raça*Ração.

8.7. Suponha que você esteja participando de uma seleção para um emprego numa empresa de pesquisa. Dentre as várias áreas em avaliação, consta a área de Estatística, que objetiva avaliar seus conhecimentos na área, não simplesmente pedindo-lhe para fazer "contas" (o que eles acham ser de menor importância), mas sim com respeito à estratégia de análise, interpretação, discussão e tomada de decisão.

São feitas as seguintes perguntas:

a) Como você faria um "leigo" entender o que vem a ser INTERAÇÃO ENTRE DOIS FATORES A e B. Para explicar você pode usar exemplos, gráficos, tabelas, etc, à sua escolha.

b) Qual a estratégia de análise a ser efetuada (ou os passos da análise subseqüente) nos seguintes casos de um fatorial com dois fatores A e B:

b.1) INTERAÇÃO NÃO SIGNIFICATIVA;

b.2) INTERAÇÃO SIGNIFICATIVA.

8.8. Para se avaliar o comportamento de 4 espécies de fungos (A, B, C e D) com relação ao crescimento em meio mínimo (m.m.) com (c/) ou sem (s/) a fonte nutritiva extrato de levedura, foi realizado um experimento fatorial 4x2 no D.B.C. com 5 repetições. Após a coleta e tabulação dos dados (numa unidade de medida qualquer) foi montado o seguinte quadro de interação de totais de tratamentos:

Meio	Fungo A	Fungo B	Fungo C	Fungo D	Totais
m.m.c/	52	60	60	90	262
m.m s/	50	56	40	40	186
Totais	102	116	100	130	448

A análise de variância dos dados no computador forneceu o seguinte quadro (incompleto) da ANOVA:

F.V.	G.L.	QM
Fator A	1	144,40
Fator B	3	19,40
Int. AxB		49,20
(Trat)		----
Blocos		----
Resíduo		10,00
Total		

Com base nos resultados fornecidos acima, pede-se: (obs.: use α=1%)

a) Cada valor interno no quadro de interação acima veio de quantas observações? Justifique.

b) Complete a coluna de G.L. do quadro acima, explicando como obteve cada um deles.

c) A que se refere o Fator A do quadro da ANOVA acima? E o Fator B? Justifique suas respostas.

d) Os fatores em estudo atuam independentemente na variável em análise (crescimento)? Justifique sua resposta.

e) Qual meio de cultura (meio mínimo com extrato de levedura ou meio mínimo sem extrato de levedura) você usaria para propiciar um maior crescimento do fungo B? Justifique sua resposta.

f) Compare por meio de um contraste, a média do grupo de fungos A e B com a média do grupo de fungos C e D pelo teste t e Scheffé quando o meio de cultura com extrato de levedura foi utilizado. Assuma que as pressuposições dos testes t e Scheffé foram satisfeitas.

8.9. Em um experimento no esquema fatorial, com dois fatores qualitativos A e B, em que se deseja estudar os efeitos dos dois fatores, qual procedimento deve-se adotar quando:

a) A interação for não-significativa.

b) A interação for significativa.

8.10. Do fatorial 4x3, no DIC com 3 repetições, são dados:

	A1			A2			A3			A4		
B1	12	14	16	15	17	18	20	21	23	23	24	26
B2	18	17	20	22	23	23	25	26	28	29	30	32
B3	22	21	20	30	31	32	29	32	32	34	35	37

Para o nível de significância de 5%, pede-se:

a) ANOVA

b) Teste de Tukey

c) Testar o contraste $C = m_{B1/A1} + m_{B2/A1} - 2m_{B3/A1}$ pelo teste de Scheffé.

8.11. De um experimento no DBC, no esquema fatorial, foram obtidos os seguintes resultados:

Totais de Tratamentos

	A1	A2	A3
B1	9068	8841	9278
B2	9932	9960	9779
B3	10709	9560	10023

$\sum Y_{ijk}^2 = 283282054$

Bloco	1	2	3
Total	28218	29641	29291

Pede-se $(\alpha = 5\%)$:

a) ANOVA

b) Teste de Duncan

8.12. Dizer o que você entende e como interpreta uma interação entre dois fatores A e B significativa, para um determinado α.

8.13. Analisar os dados do fatorial 2x3, ambos qualitativos, resumidos nos quadros de interações e ANOVA, e aplicar o teste de Duncan, se for o caso, para $\alpha = 1\%$.

Totais de Tratamentos

	B1	B2	B3	Total
A1	20,3	21,4	20,9	62,6
A2	21,4	22,3	35,6	79,3
Total	41,7	43,7	56,5	141,9

$\sum Y_{ijk}^2 = 814{,}56$.

FV	GL	SQ
Bloco	5	200,3146
A		7,7469
B		10,7467
AxB		10,4289
Resíduo		
Total		

8.14. Com os dados do quadro de interação do fatorial 2x6, no delineamento em Blocos Casualizados com 2 repetições, e considerando $\alpha = 5\%$ com os fatores A e B atuando dependentemente, pede-se :

a) testar e concluir a respeito do fator A dentro do nível B4

b) Fazer o estudo do fator B dentro dos níveis de A procedendo a análise de variância e o teste de Tukey se necessário

	B1	B2	B3	B4	B5	B6	Total
A1	46,8	48,2	47,3	49,0	48,5	46,9	286,7
A2	47,2	60,8	69,3	71,6	61,5	46,8	357,2
Total	94,0	109,0	116,6	120,6	110,0	93,7	643,9

SQResíduo = 120,8325

8.15. Em um experimento fatorial 4x2 no delineamento em Blocos Casualizados com 3 repetições, são dados:

	A1	A2	A3	A4	Total
B1	20,5	26,2	31,0	36,5	114,2
B2	15,6	21,4	26,4	29,8	93,2
Total	36,1	47,6	57,4	66,3	207,4

SQTotal = 159,98

Blocos	I	II	III
Totais	60,4	74,6	72,4

Sabendo-se que os fatores A e B atuam independentemente e adotando-se $\alpha = 5\%$, pede-se:

a) aplicar o teste de Scheffé ao contraste $C = m_{A1} + 2m_{A2} + 3m_{A3} - 6m_{A4}$

b) concluir a respeito do fator B

8.16. Em um experimento fatorial em que foram combinados 4 níveis do fator A com 2 níveis do fator B, no delineamento em Blocos Casualizados com 5 repetições, são dados:

Níveis de A	A1	A2	A3	A4
Totais	198	184	162	154

SQResíduo = 223,9680

Admitindo que os fatores atuam independentemente, aplicar o teste Tukey aos níveis do fator A e concluir para $\alpha = 5\%$.

8.17. Em um experimento fatorial, no DIC foram combinados 2 níveis do fator A com 3 níveis do fator B (ambos qualitativos), com 3 repetições. Os valores obtidos para cada repetição nos tratamentos avaliados, são dados abaixo.

	B1	B2	B3
A1	12 14 16	15 17 18	12 11 13
A2	14 13 16	11 12 11	12 12 13

Pede-se:

a) Verificar se os dois fatores atuam independentemente. Use $\alpha = 5\%$.

b) Faça um estudo completo acerca dos níveis do fator A. Concluir para $\alpha = 5\%$.

8.18. Num experimento com suínos foram comparadas três rações (A, B, C) e dois níveis de proteína (1-Alto, 2-Médio), utilizando um delineamento inteiramente casualizado num esquema fatorial com 5 repetições. Ao final do experimento, obteve-se o seguinte quadro de interação para os totais de tratamentos:

Proteína	Rações			Totais
	A	B	C	
1	498	428	477	1403
2	469	350	406	1225
Totais	967	778	883	2628

Ao nível de 5% de probabilidade, pede-se:

a. Complete o quadro da ANOVA e verifique se os fatores rações e níveis de proteína atuam independentemente.

FV	GL	SQ	QM	F
Ração				
Proteína				
Interação	2	140,4667		0,34
(Tratamentos)				
Resíduo		4957,20		
Total				

b. Qual seria a ração a ser recomendada? (Use o teste de Duncan se necessário)

c. Qual seria o nível de proteína a ser recomendado? (Use o teste de Duncan se necessário).

8.19. Um pesquisador instalou um experimento para avaliar o efeito que o horário de colheita e o tipo de colheitadeira têm na perda de grãos. Para isto foram escolhidos três horários de colheita (H1, H2 e H3) e dois tipos de colheitadeira (T1 e T2). O pesquisador definiu como unidade experimental uma área de 10×20 metros. Como as unidades experimentais eram homogêneas, o pesquisador distribuiu inteiramente ao acaso os tratamentos, ou seja, as combinações dos níveis dos fatores, H1T1, H1T2, H2T1, H2T2, H3T1 e H3T2, às unidades experimentais conforme ilustrado na Figura 1.

Figura 1 – Distribuição dos tratamentos às unidades experimentais e respectivas perdas (em gramas) observadas durante a colheita

10 m

20 m

H2T1 (43)	H1T2 (54)	H3T2 (71)	H3T2 (74)	H3T1 (56)	H2T2 (65)
H1T1 (39)	H2T2 (67)	H3T2 (73)	H2T1 (48)	H1T1 (49)	H3T1 (59)
H2T1 (41)	H3T1 (52)	H1T1 (35)	H1T2 (56)	H2T2 (62)	H1T2 (58)
H1T2 (61)	H2T1 (47)	H3T1 (58)	H2T2 (59)	H1T1 (40)	H2T2 (64)
H1T2 (59)	H3T1 (57)	H2T1 (38)	H3T2 (77)	H1T1 (45)	H3T2 (75)

Valores observados tabulados

	H1	H2	H3	Total
T1	35 40 45 49 39	43 41 47 38 48	52 57 58 56 59	
T2	54 58 56 61 59	67 59 62 65 64	71 73 74 77 75	
				$1682^{(30)}$

Totais de Tratamentos

	H1	H2	H3	Totais
T1	$208^{(5)}$	217	282	$707^{(15)}$
T2	288	317	370	975
Totais	$496^{(10)}$	534	652	$1682^{(30)}$

Com base nas informações fornecidas, pede-se:

8.19.1. Os fatores, horário de colheita e tipo de colheitadeira, atuam independentemente na perda de grãos?

8.19.2. Em qual(is) horário(s) de colheita ocorreu maior(es) média(s) de perda de grãos? Use o teste de Tukey e de Duncan.

8.19.3. Testar o contraste $C = 2m_{H1} – m_{H2} – m_{H3}$ pelos testes de Scheffé e t. Suponha que todas as pressuposições para a realização de tais testes sejam satisfeitas.

8.19.4. Qual tipo de colheitadeira ocorreu maior média de perda de grãos? Use o teste de Tukey e de Duncan se necessário.

8.19.5. Testar o contraste $C = m_{T1} – m_{T2}$ pelos testes de Scheffé e t. Suponha que todas as pressuposições para a realização de tais testes sejam satisfeitas.

8.20. Um Engenheiro de Produção, objetivando aumentar a eficiência de uma linha de produção, instalou um experimento fatorial segundo o delineamento inteiramente casualizado com 5 repetições. Neste experimento foram comparados dois tipos de controle de qualidade (A1 e A2). Cada um dos dois tipos de controle de qualidade foi testado usando dois processos de fabricação (B1 e B2). O tempo gasto, em minutos, para completar o processo de fabricação foi medido. O quadro de totais de tratamentos é fornecido a seguir:

Totais de Tratamentos

	Fator A		
Fator B	A1	A2	Totais
B1	92	113	205
B2	112	90	202
Totais	204	203	407

Resumo da ANOVA

FV	GL	SQ	QM	F
A		0,05		
B				
A*B				
Tratamentos		92,95		
Resíduo				
Total		122,55		

Com base nestas informações, pede-se (use o nível de 5% de significância):
8.20.1. Os fatores controle de qualidade e processo de fabricação atuam independentemente sobre o tempo gasto para fabricação? Justifique a sua resposta.

8.20.2. Qual processo de fabricação é mais rápido quando o controle de qualidade A1 é utilizado? Utilize o teste de Tukey, se necessário. Justifique a sua resposta.

8.21. Uma fábrica de automóveis realizou um experimento fatorial segundo o delineamento inteiramente casualizado com seis repetições, para verificar o efeito de dois fatores sobre o consumo de combustível. O primeiro fator se refere ao método de aceleração: eletrônica (A1) ou via cabo mecânico (A2). O outro fator se refere ao porte do motor: pequeno (B1), médio (B2) ou grande (B3). Os níveis destes dois fatores foram combinados, obtendo-se um total de seis tratamentos. Foram montados 36 carros e o consumo destes carros, expresso em km/l, foram medidos. Os totais observados para cada tratamento foram

Totais de Tratamentos

	Fator A		
Fator B	A1	A2	Totais
B1	73	69	142
B2	85	79	164
B3	58	52	110
Totais	216	200	416

FV	GL	SQ	QM	F
A		7,11		
B		122,99		
A*B				
Tratamentos				
Resíduo		48,67		
Total		178,89		

Baseado nestas informações e usando o nível de 1% de signficância, pede-se:
8.21.1. Os fatores método de aceleração e porte do motor atuam independentemente sobre o consumo de combustível dos carros? Justifique a sua resposta.
8.21.2. Qual método de aceleração proporciona maior consumo? Utilize o teste de Duncan se necesário. Justifque a sua resposta.

8.22. Foram obtidos os seguintes resultados parciais com a realização de um experimento com dois fatores A e B, instalado segundo o DBC com 3 blocos:

Resumo (incompleto) da ANOVA

FV	GL	SQ	QM	F
A				
B				
AxB		10315,33		
(Trat)				
Blocos				
Resíduo		180,00		
Total				

Totais de Tratamentos

	B1	B2	Totais
A1	114	85	199
A2	209	58	267
A3	330	405	735
A4	114	299	413
Totais	767	847	1614

Usando o nível de 5% de significância, pede-se:
8.22.1 Os fatores A e B atuam independentemente?
8.22.2 Proceda ao estudo do fator B dentro do nível A2 e conclua (use o teste de Tukey se necessário).
8.22.3. Supondo que o teste F da análise de variância para o estudo de A dentro de B2 foi significativo, proceda ao teste de Tukey para comparar os níveis de A dentro de B2.

8.23. Em um experimento fatorial instalado segundo o delineamento inteiramente casualizado, foram testados três tipos de suplementos minerais (Fator A) e dois tipos de suplementos vegetais (Fator B) no confinamento de bovinos. Os ganhos de peso obtidos pelos animais em teste foram:

	Repetições				
Tratamentos	1	2	3	4	Totais
A1B1	35,2	36,0	35,0	35,4	141,6
A1B2	32,8	34,6	36,7	35,2	139,3
A2B1	34,7	36,3	35,1	36,4	142,5
A2B2	28,6	31,1	29,0	28,6	117,3
A3B1	33,8	29,4	28,8	29,2	121,2
A3B2	30,8	31,4	32,8	31,3	126,3

	Fator B		
Fator A	1	2	Totais
1	141,6	139,3	280,9
2	142,5	117,3	259,8
3	121,2	126,3	247,5
Totais	405,3	382,9	788,2

FV	GL	SQ	QM	F
A		71,34		
B				
AxB				
(Trat)		154,33		
Resíduo		33,59		
Total		188,22		

Com base nas informações fornecidas, pede-se:

8.23.1 O valor do F calculado para testar o efeito da interação entre os fatores A e B.

8.23.2. O valor do F calculado para comparar os níveis de B dentro do nível A2

8.23.3. O valor do F calculado para comparar os níveis de A dentro do nível B2

8.24. Em um experimento instalado segundo o delineamento inteiramente casualizado, com 4 repetições, foram estudados os fatores A e B, com 3 e 2 níveis respectivamente. Deste experimento, são fornecidas as seguintes informações:

FV	GL	SQ	QM	F
A		92,86		
B		19,08		
AxB				
(Trat)		(175,70)		
Resíduo				
Total		198,70		

Ttotais de Tratamentos

Fator B	Fator A			Totais
	A1	A2	A3	
B1	102,6	103,5	80,2	286,3
B2	101,3	78,3	85,3	264,9
Totais	203,9	181,8	165,5	551,2

Com base nas informações fornecidas, pede-se (use o nível de 1% de significância quando necessário)

8.24.1. Os fatores A e B atuam independentemente?

8.24.2. Existe diferença entre os níveis de A dentro do nível B_1?

8.24.3. Qual o nível de B apresenta maior média dentro do nível A_2? Use o teste de Tukey, se necessário.

8.25 Foi realizado um experimento, instalado segundo o delineamento inteiramente casualizado no esquema fatorial, para avaliar o efeito do fator Recipiente e do fator Espécie na altura da muda aos 80 dias de idade. Os níveis do fator recipiente avaliados foram saco plástico pequeno, saco plástico grande e saco laminado, daqui por diante identificados como R1, R2 e R3, respectivamente. Os níveis do fator espécie avaliados foram *Eucalyptus citriodora* e *Eucalyptus grandis* daqui por diante identificados como, E1 e E2, respectivamente de eucalipto. Os valores observados, em *cm*, foram

Tratamentos	Repetições				Totais
	1	2	3	4	
1 – R1E1	26,2	26,0	25,0	25,4	102,6
2 – R1E2	24,8	24,6	26,7	25,2	101,3
3 – R2E1	25,7	26,3	25,1	26,4	103,5
4 – R2E2	19,6	21,1	19,0	18,6	78,3
5 – R3E1	22,8	19,4	18,8	19,2	80,2
6 – R3E2	19,8	21,4	22,8	21,3	85,3

Usando α=1%

a) Os fatores, Recipiente e Espécie, atuam independentemente na altura das mudas?

b) Levando em consideração o teste F para a interação entre os fatores, indique qual(is) nível(is) de Recipiente proporcionou(aram) maior média de altura das mudas? Use o teste de Tukey quando necessário.

c) Idem para Espécie. Use o teste de Tukey quando necessário.

d) Utilize os testes de Scheffé e t para testar os contrastes apropriados, os quais foram elaborados pelo pesquisador durante o planejamento deste experimento

Interação	Fator	
	Recipiente	Espécie
Significativa	$C_1 = m_{R1/E1} + m_{R2/E1} - 2m_{R3/E1}$	$C_2 = m_{E1/R1} - m_{E2/R1}$
Não-signficativa	$C_3 = m_{R1} + m_{R2} - 2m_{R3}$	$C_4 = m_{E1} - m_{E2}$

Informação adicional: Quadro de Totais de Tratamentos

Espécies	Recipientes			Totais
	R1	R2	R3	
E1	102,6 $^{(4)}$	103,5	80,2	286,3 $^{(12)}$
E2	101,3	78,3	85,3	264,9 $^{(12)}$
Totais	203,9 $^{(8)}$	181,8	165,5	551,2 $^{(24)}$

Observação: Este exercício foi adaptado de BANZATTO e KRONKA (1989)

9. Experimentos em Parcelas Subdivididas

9.1. Introdução

Tal como no caso de fatorial, o termo parcelas subdivididas não se refere a um tipo de delineamento e sim ao esquema do experimento, ou seja, a maneira pela qual os tratamentos são organizados. Nos experimentos em parcelas subdivididas, em geral, estuda-se simultaneamente dois tipos de fatores os quais são geralmente denominados de fatores primários e fatores secundários.

Em um experimento em parcelas subdivididas, as unidades experimentais são agrupadas em parcelas as quais devem conter um número de unidades experimentais (subparcelas) igual ao número de níveis do fator secundário. Na instalação os níveis do fator primário são distribuídos às parcelas segundo um tipo de delineamento experimental (DIC, DBC, etc...). Posteriormente os níveis do fator secundário são distribuídos ao acaso as subparcerlas de cada parcela.

Como a variação residual entre subparcelas é esperada ser menor do que entre parcelas, deve-se escolher como fator secundário, o fator que se espera apresentar menor diferenças, ou para o qual deseja-se maior precisão.

Às vezes o pesquisador pode optar entre um experimento com parcelas subdivididas e um experimento fatorial. Para a escolha do esquema em parcelas subdivididas, o pesquisador pode se basear nos seguintes critérios (VIEIRA, 1989):

1 - a parcela é uma unidade "física" (um vaso, um animal, uma pessoa) que pode receber vários níveis de um fator secundário;

2 - o fator principal exige "grandes parcelas" - como é o caso da irrigação e de processos industriais;

3 - o pesquisador quer comparar níveis de um fator secundário com maior precisão.

9.2. Modelo estatístico

O modelo estatístico, para um experimento em parcelas subdivididas, varia de acordo com o tipo de delineamento utilizado. Assim, para um experimento instalado segundo o DIC, em que o fator A é o fator primário e o fator B é o fator secundário, o modelo estatístico é:

$$Y_{ijk} = m + \alpha_i + \delta_{ik} + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$$

em que,

Y_{ijk} é o valor observado para a variável em estudo referente a k-ésima repetição da combinação do i-ésimo nível do fator A com o j-ésimo nível do fator B;

m é a média de todas as unidades experimentais para a variável em estudo;

α_i é o efeito do i-ésimo nível do fator A no valor observado Y_{ijk};

β_j é o efeito do j-ésimo nível do fator B no valor observado Y_{ijk};

$(\alpha\beta)_{ij}$ é o efeito da interação do i-ésimo nível do fator A com o j-ésimo nível do fator B;

δ_{ik} é o efeito residual das parcelas, caracterizado como componente do erro (a);

e_{ijk} é o efeito residual das subparcelas, caracterizado como componente do erro (b).

Para um experimento em parcelas subdivididas instalado segundo o DBC, com K blocos, o modelo estatístico seria:

$$Y_{ijk} = m + \alpha_i + \delta_{ik} + \beta_j + (\alpha\beta)_{ij} + \omega_k + e_{ijk}$$

em que,

ω_k é o efeito do k-ésimo bloco na observação Y_{ijk}.

9.3. Quadro de tabulação de dados

O quadro de tabulação de dados de um experimento em parcelas subdivididas é similar ao usado para tabular os dados de um experimento em fatorial. O quadro a seguir, ilustra a tabulação de dados de um experimento em parcelas subdivididas, no qual o fator primário é representado pelo fator A com I níveis, e o fator secundário representado pelo fator B com J níveis:

	A1				A2					AI			
Repetição	B1	B2	...	BJ	B1	B2	...	BJ	...	B1	B2	...	BJ
1	Y_{111}	Y_{121}	...	Y_{1J1}	Y_{211}	Y_{221}	...	Y_{2J1}	...	Y_{I11}	Y_{I21}	...	Y_{IJ1}
2	Y_{112}	Y_{122}	...	Y_{1J2}	Y_{212}	Y_{222}	...	Y_{2J2}	...	Y_{I12}	Y_{I22}	...	Y_{IJ2}
...	...	...	...	...	...	...	...	...	...	...	...	...	...
K	Y_{11K}	Y_{12K}	...	Y_{1JK}	Y_{21K}	Y_{22K}	...	Y_{2JK}	...	Y_{I1K}	Y_{I2K}	...	Y_{IJK}
Total	$Y_{11\bullet}$	$Y_{12\bullet}$	...	$Y_{1J\bullet}$	$Y_{21\bullet}$	$Y_{22\bullet}$	...	$Y_{2J\bullet}$	...	$Y_{I1\bullet}$	$Y_{I2\bullet}$	...	$Y_{IJ\bullet}$

Deste quadro, pode-se tirar algumas informações que posteriormente serão úteis na análise de variância:

- Total do ij-ésimo tratamento: $(AB)_{ij} = \sum_{k=1}^{K} Y_{ijk} = Y_{ij\bullet}$
- Total do i-ésimo nível do fator A: $A_i = \sum_{j=1,k=1}^{J,K} Y_{ijk} = Y_{i\bullet\bullet}$
- Total do j-ésimo nível do fator B: $B_j = \sum_{i=1,k=1}^{I,K} Y_{ijk} = Y_{\bullet j\bullet}$
- Total Geral: $G = \sum_{i=1,j=1,k=1}^{I,J,K} Y_{ijk} = \sum_{i=1}^{I} A_i = \sum_{j=1}^{J} B_j = Y_{\ldots}$
- Total de Parcelas: $P_z = \sum_{j=1}^{J} Y_{ijk}$
- Total de Blocos: $W_k = \sum_{i=1,j=1}^{I,J} Y_{ijk}$

- Média do i-ésimo nível do fator A: $\hat{m}_{A_i} = \dfrac{A_i}{JK}$
- Média do j-ésimo nível do fator B: $\hat{m}_{B_i} = \dfrac{B_j}{IK}$,
- Média geral: $\hat{m} = \dfrac{G}{N}$
- Número de parcelas Z = IK
- Número total de subparcelas: N_T=IJK

Para experimentos em parcelas subdivididas, pode-se montar dois quadros auxiliares. O primeiro deles é idêntico ao visto para experimentos fatoriais que é o quadro de totais de tratamentos, cujos valores são obtidos pela soma de todas as repetições para o tratamento em questão. Para a situação citada, o quadro de totais de tratamentos é do seguinte tipo:

Fator A	Fator B				Totais
	B1	B2	...	BJ	
A1	$Y_{11.}$	$Y_{12.}$	...	$Y_{1J.}$	A_1
A2	$Y_{21.}$	$Y_{22.}$	...	$Y_{2J.}$	A_2
...	...	...	...	...	...
AI	$Y_{I1.}$	$Y_{I2.}$	...	$Y_{IJ.}$	A_I
Totais	B_1	B_2	...	B_j	G

O segundo quadro se refere ao quadro de totais de parcelas. Este quadro facilita o cálculo das somas de quadrados de parcelas. Para a situação acima, o quadro de totais de parcelas é do seguinte tipo:

Fator A	Parcela				Totais de A
	1	2	...	Z	
A1	$Y_{1.1}$	$Y_{1.2}$	...	$Y_{1.Z}$	A_1
A2	$Y_{2.1}$	$Y_{2.2}$	...	$Y_{2.Z}$	A_2
...	...	...	...	...	...
AI	$Y_{I.1}$	$Y_{I.2}$	...	$Y_{I..}$	A_I
Totais de Parcelas	P_1	P_2	...	P_Z	G

9.4. Análise de variância

A análise de variância de um experimento em parcelas subdivididas é feita desdobrando os efeitos das parcelas e das subparcelas nas partes que as compõem. Para cada um destes desdobramentos, existe um resíduo, o qual é utilizado para testar o efeito das fontes de variação pertinentes.

O quadro a seguir apresenta como seria a análise de um experimento instalado segundo o DBC com K repetições no esquema em parcelas subdivididas, em que o fator A com I níveis foi designado às parcelas e o fator B com J níveis foi designado às subparcelas

FV	GL	SQ	QM	F	$F_{tab;\ \alpha}$
Blocos	(K-1)	SQBlocos	-	-	
A	(I-1)	SQA	$\frac{SQA}{(I-1)}$	-	
Resíduo(a)	(I-1)(K-1)	SQRes(a)	$\frac{SQResíduo(a)}{(I-1)(K-1)}$		
Parcelas	IK-1	SQParcelas	-	-	
B	(J-1)	SQB	$\frac{SQB}{(J-1)}$	-	
AxB	(I-1)(J-1)	SQAxB	$\frac{SQAxB}{(I-1)(J-1)}$	$\frac{QMAxB}{QMRes(b)}$	$[(I-1)(J-1);\ n_2]$
Resíduo(b)	n_2 = I(J-1)(K-1)	SQRes(b)	$\frac{SQResíduo(b)}{I(J-1)(K-1)}$	-	
Total	IJ K- 1	SQTotal	-	-	

em que:

$$SQTotal = \sum_{i=1,j=1,k=1}^{I,J,K} Y_{ijk}^2 - C \qquad C = \frac{\left(\sum_{i=1,j=1,k=1}^{I,J,K} Y_{ijk}\right)^2}{IJK} \qquad SQBlocos = \sum_{K=1}^{K} \frac{W_K^2}{IJ} - C$$

$$SQParcelas = \frac{\sum_{z=1}^{Z} P_z^{\ 2}}{J} - C \qquad SQTrat = \sum_{i=1,j=1}^{I,J} \frac{Y_{ij.}^2}{K} - C \qquad SQA = \sum_{i=1}^{I} \frac{A_I^2}{JK} - C$$

$$SQB = \sum_{j=1}^{J} \frac{B_J^2}{IK} - C$$

SQAxB = SQTrat – SQA – SQB

SQRes(a) = SQParcelas - SQBlocos - SQA

SQRes(b) = SQTotal - SQParcelas - SQB - SQAxB

Tal como no esquema fatorial, na análise dos dados oriundos de um experimento em parcelas subdivididas deve-se inicialmente proceder ao teste F para a interação entre os fatores. As hipóteses para o teste F da interação são:

H_0 : Os fatores A e B atuam independentemente sobre a variável resposta em estudo.

H_a : Os fatores A e B não atuam independentemente sobre a variável resposta em estudo.

O resultado deste teste F para a interação indica como as comparações dos níveis de um fator devem ser realizadas. Temos dois resultados possíveis para o teste F da interação os quais serão apresentados a seguir.

9.4.1 Interação não-significativa

Este caso ocorre quando a hipótese H_0 para a interação entre os fatores não é rejeitada. Este resultado implica que os efeitos dos fatores atuam de forma independente.

Portanto recomenda-se que as comparações dos níveis de um fator sejam feitas de forma geral em relação ao outro fator, ou seja, independente dos níveis outro fator. O passo seguinte na análise estatística dos dados experimentais é proceder ao teste F para cada fator como ilustrado na tabela apresentada a seguir para o caso do DBC.

FV	GL	SQ	QM	F	$F_{tab;\,\alpha}$
Blocos	(K-1)	SQBlocos	-	-	
A	(I-1)	SQA	$\frac{SQA}{(I-1)}$	$\frac{QMA}{QMRes(a)}$	[(I-1); n_2]
Resíduo(a)	n_2 = (I-1)(K-1)	SQRes(a)	$\frac{SQRes(a)}{(I-1)(K-1)}$	-	
Parcelas	IK-1	SQParcelas	-	-	
B	(J-1)	SQB	$\frac{SQB}{(J-1)}$	$\frac{QMB}{QMRes(b)}$	[(J-1); n_3]
AxB	(I-1)(J-1)	SQAxB	$\frac{SQAxB}{(I-1)(J-1)}$	não-signficativo	-
Resíduo(b)	n_3 = I(J-1)(K-1)	SQRes(b)	$\frac{SQRes(b)}{I(J-1)(K-1)}$	-	
Total	IJ K- 1	SQTotal	-	-	

As hipóteses para realizar o teste F para os efeitos principais são

Fator A

$H_0: m_{A1} = m_{A2} = ... = m_{AI}$ ou seja, todos os possíveis contrastes entre as médias dos níveis do fator A, são estatisticamente nulos, ao nível de probabilidade em que foi executado o teste.

$H_a:$ não H_0 ou seja, existe pelo menos um contraste entre as médias dos níveis do fator A, que é estatisticamente diferente de zero, ao nível de probabilidade em que foi executado o teste.

Fator B

$H_0: m_{B1} = m_{B2} = ... = m_{BJ}$ ou seja, todos os possíveis contrastes entre as médias dos níveis do fator B, são estatisticamente nulos, ao nível de probabilidade em que foi executado o teste.

$H_a:$ não H_0 ou seja, existe pelo menos um contraste entre as médias dos níveis do fator B, que é estatisticamente diferente de zero, ao nível de probabilidade em que foi executado o teste.

Se os fatores A e B forem qualitativos, e o teste F para A e/ou B, for não significativo, a aplicação do teste de médias é desnecessária. Se o teste F for significativo, para A e/ou B, aplica-se um teste de médias para comparar os níveis do fator. As estimativas das médias dos níveis dos fatores são obtidas por

Fator A $\rightarrow$ $\hat{m}_{Ai} = \frac{A_i}{JK}$

Fator B $\rightarrow$ $\hat{m}_{Bj} = \frac{B_j}{IK}$

Para realizar o teste de Tukey para comparar as medias dos níveis dos fatores em teste temos que usar

	Δ	q_α
A	$q\sqrt{\frac{QMRes(a)}{JK}}$	$(I;n_2)$
B	$q\sqrt{\frac{QMRes(b)}{IK}}$	$(J;n_3)$

Para o teste de Duncan temos que usar

	D_i	z_α
A	$z\sqrt{\frac{QMRes(a)}{JK}}$	$(n_A;n_2)$
B	$z\sqrt{\frac{QMRes(b)}{IK}}$	$(n_B;n_3)$

Em que n_A e n_B são os números de médias ordenadas abrangidas pelo contraste sendo testados.

As hipóteses para os testes Tukey e Duncan para comparar as médias dos níveis dos fatores são

Fator A $\rightarrow$ $H_0 : m_{Ai} = m_{Au}$ versus $H_a : m_{Ai} \neq m_{Au}$ para $i \neq u = 1, 2, 3, \ldots , I$

Fator B $\rightarrow$ $H_0 : m_{Bj} = m_{Bu}$ versus $H_a : m_{Bj} \neq m_{Bu}$ para $j \neq u = 1, 2, 3, \ldots , J$

Para a aplicação do teste t temos que usar

	t	t_{tab}
A	$\frac{\hat{C}_A - C_A}{\sqrt{\frac{QMRes(a)}{JK}\sum_{i=1}^{I} a_i^2}}$	$t_\alpha\ (n_2)$
B	$\frac{\hat{C}_B - C_B}{\sqrt{\frac{QMRes(b)}{IK}\sum_{j=1}^{J} b_i^2}}$	$t_\alpha\ (n_3)$

Em que

$C_A = a_1m_{A1} + a_2m_{A2} + \ldots + a_Im_{AI}$ e

$C_B = b_1m_{B1} + b_2m_{B2} + \ldots + b_Jm_{BJ}$

Para a aplicação do teste Scheffé para testar os contrastes C_A e C_B temos que usar

	S	F_{tab}
A	$S=\sqrt{(I-1)F_{tab}\dfrac{QMRes(a)}{JK}\sum_{i=1}^{I}a_i^2}$	F_α [(I -1); n_2]
B	$S=\sqrt{(J-1)F_{tab}\dfrac{QMRes(b)}{IK}\sum_{j=1}^{J}b_i^2}$	F_α [(J -1); n_3]

As hipóteses para os testes de Scheffé e t para testar os contrastes são

Fator A $\rightarrow$ $H_0 : C_A = 0$ versus $H_a : C_A \neq 0$
Fator B $\rightarrow$ $H_0 : C_B = 0$ versus $H_a : C_B \neq 0$

9.4.2 Interação significativa

Este caso ocorre quando a hipótese H_0 para a interação entre os fatores é rejeitada. Este resultado implica que os efeitos dos fatores atuam de forma dependente. Neste caso as comparações entre os níveis de um fator levem em consideração o nível do outro fator, pois o resultado significativo para a interação indica que o efeito de um fator depende do nível do outro fator.

Portanto, não é recomendado realizar o teste F para cada fator isoladamente tal como foi apresentado para o caso da interação não-significativa. O procedimento recomendado é realizar o desdobramento do efeito da interação.

Para realizar este desdobramento deve-se fazer uma nova análise de variância em que os níveis de um fator são comparados dentro de cada nível do outro fator, tal como apresentado nas tabelas a seguir.

Para comparar os níveis de um fator principal em cada nível do fator secundário, é necessário fazer uma combinação das duas estimativas obtidas para o erro experimental bem como do número de graus de liberdade associado as mesmas. Esta combinação é denominada de resíduo combinado (ResComb). A estimativa do quadrado médio deste resíduo combinado é obtida por

$$QMResComb = \frac{QMRes(a) + (J-1)QMRes(b)}{J}$$

O número de graus de liberdade associado a esta estimativa é obtido pela fórmula dos graus de liberdade de Satterhwaitte (n*) dada por

$$n^* = \frac{[QMRes(a) + (J-1)QMRes(b)]^2}{\dfrac{[QMRes(a)]^2}{g.l.Res(a)} + \dfrac{[(J-1)QMRes(b)]^2}{g.l.Res(b)}}$$

Desdobramento para comparar os níveis de A dentro de cada nível de B, ou seja estudar A/B

FV	GL	SQ	QM	F	$F_{tab,\ \alpha}$
A/B1	(I-1)	SQA/B1	$\frac{SQA/B1}{(I-1)}$	$\frac{QMA/B1}{QMResComb}$	[(I-1);n*]
A/B2	(I-1)	SQA/B2	$\frac{SQA/B2}{(I-1)}$	$\frac{QMA/B2}{QMResComb}$	[(I-1);n*]
...	...	...	...	...	...
A/BJ	(I-1)	SQA/BJ	$\frac{SQA/BJ}{(I-1)}$	$\frac{QMA/BJ}{QMResComb}$	[(I-1);n*]
ResComb	n*		QMResComb	-	
Total	IJK - 1	SQTotal	-	-	-

As hipóteses para testar as fontes de variação da tabela acima, para j=1, 2, 3, ..., J, são

$H_0 : m_{A1/Bj} = m_{A2/Bj} = ... = m_{AI/Bj}$
H_a :não H_0

Desdobramento para comparar os níveis de B dentro de cada nível de A, ou seja estudar B/A

FV	GL	SQ	QM	F	$F_{tab,\ \alpha}$
B/A1	(J-1)	SQB/A1	$\frac{SQB/A1}{(J-1)}$	$\frac{QMB/A1}{QMRes(b)}$	[(J-1);n_3]
B/A2	(J-1)	SQB/A2	$\frac{SQB/A2}{(J-1)}$	$\frac{QMB/A2}{QMRes(b)}$	[(J-1);n_3]
...	...	...	...	...	...
B/AI	(J-1)	SQB/AI	$\frac{SQB/AI}{(J-1)}$	$\frac{QMB/AI}{QMRes(b)}$	[(J-1);n_3]
Res(b)	n_3		QMRes(b)	-	
Total	IJK - 1	SQTotal	-	-	-

As hipóteses para testar as fontes de variação da tabela acima, para i=1, 2, 3, ..., I, são

$H_0 : m_{B1/Ai} = m_{B2/Ai} = ... = m_{BJ/Ai}$
H_a :não H_0

Em que as SQA/Bj e SQB/Ai podem ser obtidas usando a fórmula geral para a soma de quadrados dada por

$$SQ = \sum_{i=1}^{k} \frac{X_i^2}{r_i} - \frac{\left(\sum_{i=1}^{k} X_i\right)^2}{\sum_{i=1}^{k} r_i}$$

Se os fatores forem qualitativos, procede-se ao teste F para cada fonte de variação do desdobramento. Nas fontes de variação em que o teste F foi significativo e o fator tem mais de dois níveis, recomenda-se a aplicação de um teste de médias. As estimativas das médias dos níveis dos fatores são obtidas por

Fator A $\rightarrow \hat{m}_{Ai} = \frac{A_i}{K}$

Fator B $\rightarrow \hat{m}_{Bj} = \frac{B_j}{K}$

Para realizar o teste de Tukey para comparar as médias dos níveis dos fatores em teste temos que usar

	Δ	q_α
A	$q\sqrt{\frac{QMResComb}{K}}$	$(I;n^*)$
B	$q\sqrt{\frac{QMRes(b)}{K}}$	$(J;n_3)$

Para o teste de Duncan temos que usar

	D_i	z_α
A	$z_i\sqrt{\frac{QMResComb}{K}}$	$(n_A;n^*)$
B	$z_i\sqrt{\frac{QMRes(b)}{K}}$	$(n_B;n_3)$

Em que n_A e n_B são os números de médias ordenadas abrangidas pelo contraste sendo testados.

As hipóteses para os testes Tukey e Duncan para comparar as médias dos níveis dos fatores são

Fator A:

$H_0 : m_{Ai/Bj} = m_{Au/Bj}$ vs $H_a : m_{Ai/Bj} \neq m_{Au/Bj}$ para $i \neq u = 1, 2, 3, \ldots, I$ e $j = 1, 2, \ldots, J$

Fator B:

$H_0 : m_{Bj/Ai} = m_{Bu/Ai}$ vs $H_a : m_{Bj/Ai} \neq m_{Bu/Ai}$ para $j \neq u = 1, 2, 3, \ldots, J$ e $i = 1, 2, \ldots, I$

Para a aplicação do teste t temos que usar

	t	t_{tab}
A	$\frac{\hat{C}_A - C_A}{\sqrt{\frac{QMResComb}{K}\sum_{i=1}^{I} a_i^2}}$	$t_\alpha\ (n^*)$
B	$\frac{\hat{C}_B - C_B}{\sqrt{\frac{QMRes(b)}{K}\sum_{j=1}^{J} b_j^2}}$	$t_\alpha\ (n_3)$

Em que

$C_A = a_1 m_{A1/Bj} + a_2 m_{A2/Bj} + \ldots + a_I m_{AI/Bj}$ para j = 1, 2, ..., J e
$C_B = b_1 m_{B1/Ai} + b_2 m_{B2/Ai} + \ldots + b_j m_{BJ/Ai}$ para i = 1, 2, ... , I

Para a aplicação do teste Scheffé para testar os contrastes C_A e C_B temos que usar

	S	F_{tab}
A	$S = \sqrt{(I-1)F_{tab}\dfrac{QMResComb}{K}\sum_{i=1}^{I} a_i^2}$	F_α [(I -1); n^*]
B	$S = \sqrt{(J-1)F_{tab}\dfrac{QMRes(b)}{K}\sum_{j=1}^{J} b_j^2}$	F_α [(J -1); n_3]

As hipóteses para os testes de Scheffé e t para testar os contrastes são

Fator A $\rightarrow H_0 : C_A = 0$ versus $H_a : C_A \neq 0$
Fator B $\rightarrow H_0 : C_B = 0$ versus $H_a : C_B \neq 0$

9.5. Vantagens e desvantagens

Em comparação com experimentos fatoriais, experimentos em parcelas subdivididas são mais fáceis de instalar. No entanto, existe duas estimativas de variância residual: uma associada às parcelas e outra associada às subparcelas. Este desdobramento da variância residual faz com que o número de graus de liberdade associado a cada um dos resíduos seja menor do o associado ao resíduo se o experimento tivesse sido instalado segundo o esquema fatorial.

Conseqüentemente, há uma tendência de se obter maior valor para a estimativa do erro experimental. Portanto, em experimentos com parcelas subdivididas, todos os efeitos são avaliados com menor precisão que nos experimentos fatoriais correspondentes. Por isso, sempre que possível, é preferível utilizar experimentos fatoriais em lugar dos experimentos em parcelas subdivididas.

9.6. Exercícios

9.1. Considere um experimento instalado segundo o DBC e no esquema em parcelas subdivididas no qual são comparadas 4 variedades de aveia e 4 tratamentos de sementes (3 produtos químicos + testemunha não tratada) quanto aos efeitos de produção. Na instalação do experimento, as 4 variedades foram distribuídas ao acaso nas parcelas de cada um dos 4 blocos do experimento e os tratamentos de sementes foram distribuídos ao acaso nas 4 subparcelas de cada parcela (BANZATTO & KRONKA, 1989). Com base nos resultados fornecidos a seguir, pede-se, usando o nível de 5% de probabilidade, proceder a análise de variância e aplicar o teste Tukey, quando necessário:

Variedades	Sementes	Blocos				Totais
		1	2	3	4	Trat
A1 Vicland 1	B1 Testemunha	42,9	41,6	28,9	30,8	144,2
	B2 Ceresan M	53,8	58,5	43,9	46,3	202,5
	B3 Panogen	49,5	53,8	40,7	39,4	183,4
	B4 Agrox	44,4	41,8	28,3	34,7	149,2
Totais de Parcelas		190,6	195,7	141,8	151,2	679,3
A2 Vicland 2	B1 Testemunha	53,3	69,6	45,4	35,1	203,4
	B2 Ceresan M	57,6	69,6	42,4	51,9	221,5
	B3 Panogen	59,8	65,8	41,4	45,4	212,4
	B4 Agrox	64,1	57,4	44,1	51,6	217,2
Totais de Parcelas		234,8	262,4	173,3	184,0	854,5
A3 Clinton	B1 Testemunha	62,3	58,5	44,6	50,3	215,7
	B2 Ceresan M	63,4	50,4	45,0	46,7	205,5
	B3 Panogen	64,5	46,1	62,6	50,3	223,5
	B4 Agrox	63,6	56,1	52,7	51,8	224,2
Totais de Parcelas		253,8	211,1	204,9	199,1	868,9
A4 Branch	B1 Testemunha	75,4	65,6	54,0	52,7	247,7
	B2 Ceresan M	70,3	67,3	57,6	58,5	253,7
	B3 Panogen	68,8	65,3	45,6	51,0	230,7
	B4 Agrox	71,6	69,4	56,6	47,4	245,0
Totais de Parcelas		286,1	267,6	213,8	209,6	977,1
Totais de Blocos		965,3	936,8	733,8	743,9	3379,8

Totais de Parcelas

(4)	BLOCO 1	BLOCO 2	BLOCO 3	BLOCO 4	Totais (16)
A1	190,6	195,7	141,8	151,2	679,3
A2	234,8	262,4	173,3	184,0	854,5
A3	253,8	211,1	204,9	199,1	868,9
A4	286,1	267,6	213,8	209,6	977,1
Totais (16)	965,3	936,8	733,8	743,9	3379,8 (64)

Totais de Tratamentos

(4)	B1	B2	B3	B4	Totais (16)
A1	144,2	202,5	183,4	149,2	679,3
A2	203,4	221,5	212,4	217,2	854,5
A3	215,7	205,5	223,5	224,2	868,9
A4	247,7	253,7	230,7	245,0	977,1
Totais (16)	811,0	883,2	850,0	835,6	3379,8 (64)

9.2. Para se estudar o brix de mangas de acordo com a variedade e a posição dos frutos em relação aos pontos cardeais, um pesquisador procedeu a coleta de 4 frutos, cada um deles de um ponto cardeal, em cada um dos 3 exemplares de cada uma das 5 variedades em teste. Com base nos resultados (brix) fornecidos a seguir (GOMES, 1987), pede-se usando o nível de 5% de probabilidade, proceder a análise de variância e o teste Duncan quando necessário.

Variedades	B1 Norte	B2 Sul	B3 Leste	B4 Oeste	Totais Parc	Totais
A1 Carlota	18,0	17,1	17,6	17,6	70,3	
	17,5	18,8	18,1	17,2	71,6	
	17,8	16,9	17,6	16,5	68,8	
Totais Trat	53,3	52,8	53,3	51,3		210,7
A2 Extrema	16,3	15,9	16,5	18,3	67,0	
	16,6	14,3	16,3	17,5	64,7	
	15,0	14,0	15,9	15,2	60,1	
Totais Trat	47,9	44,2	48,7	51,0		191,8
A3 Oliveira	16,0	16,2	17,9	16,1	66,2	
	19,5	14,9	15,0	15,3	64,7	
	16,3	16,4	16,0	16,4	65,1	
Totais Trat	51,8	47,5	48,9	47,8		196,0
A4 Bourbon	16,6	15,2	14,2	15,5	61,5	
	15,9	13,2	18,0	17,3	64,4	
	17,5	15,8	16,7	18,4	68,4	
Totais Trat	50,0	44,2	48,9	51,2		194,3
A5 Imperial	18,9	18,6	15,3	17,0	69,8	
	18,5	13,7	18,2	18,3	68,7	
	21,5	16,4	18,3	16,6	72,8	
Totais Trat	58,9	48,7	51,8	51,9		211,3
Totais	261,9	237,4	251,6	253,2	1004,1	1004,1

Totais de Parcelas				
	REP 1	REP 2	REP 3	Totais
A1	70,3	71,6	68,8	210,7
A2	67,0	64,7	60,1	191,8
A3	66,2	64,7	65,1	196,0
A4	61,5	64,4	68,4	194,3
A5	69,8	68,7	72,8	211,3
				1004,1

Totais de Tratamentos					
	B1	B2	B3	B4	Totais
A1	53,3	52,8	53,3	51,3	210,7
A2	47,9	44,2	48,7	51,0	191,8
A3	51,8	47,5	48,9	47,8	196,0
A4	50,0	44,2	48,9	51,2	194,3
A5	58,9	48,7	51,8	51,9	211,3
Totais	261,9	237,4	251,6	253,2	1004,1

9.3. Um pesquisador, com o objetivo de verificar o efeito da dose de adubação fosfatada e o seu tipo de aplicação na cultura do milho, instalou um experimento no qual cada uma as doses de adubação fosfatada constituíram as parcelas as quais foram distribuídas segundo o DBC e o tipo de aplicação as subparcelas. Com base nos resultados fornecidos abaixo, referentes a produção de milho (kg/ha), pede-se ao nível de 5% de probabilidade, proceder a análise de variância e ao teste Tukey quando necessário (FERREIRA, 1991).

		Blocos				
Doses	Tipos de Aplicação	I	II	III	IV	Totais de tratamentos
	cova	3778	3618	2164	3996	13556
0	sulco	3467	4284	3773	3280	14804
	lanço	3422	3760	2747	2853	12782
	Totais de Parcelas	10667	11662	8684	10129	
	cova	3302	2671	2782	2502	11257
40	sulco	3653	2653	3529	2258	12093
	lanço	3711	3284	2556	3284	12835
	Totais de parcelas	10666	8608	8867	8044	
	cova	2938	2813	2560	3049	11360
80	sulco	3800	4356	3560	4013	15729
	lanço	2702	3520	3382	3524	13128
	Totais de parcelas	9440	10689	9502	10586	
	cova	3013	3787	3142	3604	13546
120	sulco	3338	3369	2507	4200	13414
	lanço	3156	4369	2831	4222	14578
	Totais de parcelas	9507	11525	8480	12026	
	Totais de blocos	40280	42484	35493	40785	159082

9.4. Suponha que para um experimento instalado segundo o delineamento inteiramente casualizado e no esquema de parcelas subdivididas com 3 repetições, foram obtidos os seguintes resultados:

FV	GL	SQ	QM	F
Fator A		29,55		
Resíduo(a)		15,71		
(Parcelas)		(45,26)		
Fator B		20,60		
Interação A*B		20,12		
Resíduo(b)		51,60		
Total		137,58		

Totais de Tratamentos

	B1	B2	B3	B4	Totais
A1	53,3	52,8	53,3	51,3	210,7
A2	47,9	44,2	48,7	51,0	191,8
A3	51,8	47,5	48,9	47,8	196,0
A4	50,0	44,2	48,9	51,2	194,3
A5	58,9	48,7	51,8	51,9	211,3
Totais	261,9	237,4	251,6	253,2	1004,1

Usando o nível de 5% de significância quando necessário, pede-se:
9.4.1. Os fatores A e B atuam independentemente? Justifique sua resposta.
9.4.2. Existe diferença entre os níveis de B pelo teste F da análise de variância?
9.4.3. Se o objetivo é obter menores médias, qual(is) o(s) nível(is) de B que devem ser recomendados? (Use o teste de Duncan, se necessário).

9.5. Considere os resultados obtidos de um experimento instalado segundo o delineamento inteiramente casualizado no esquema de parcelas subdivididas e o nível de 5% de significância quando necessário:

Totais de Parcelas

	Repetições			
	1	2	3	Totais
A1	50,3	51,6	48,8	150,7
A2	27,0	24,7	20,1	71,8
A3	26,2	24,7	25,1	76,0
A4	21,5	24,4	28,4	74,3
A5	69,8	68,7	72,8	211,3

Totais de Tratamentos

	B1	B2	B3	B4	Totais
A1	38,3	37,8	38,3	36,3	150,7
A2	17,9	14,2	18,7	21,0	71,8
A3	21,8	17,5	18,9	17,8	76,0
A4	20,0	14,2	18,9	21,2	74,3
A5	58,9	48,7	51,8	51,9	211,3
Totais	156,9	132,4	146,6	148,2	584,1

Análise de Variância

FV	GL	SQ	QM	F
A		1297,95		
Res(a)				
Parcelas				
B		20,59		
Interação A*B				
Res(b)				
Total		1405,97		

Com base nestas informações, pede-se:
9.5.1. O valor do F calculado para testar a interação entre os fatores A e B.
9.5.2. O valor do F calculado para o fator A.
9.5.3. O(s) nível(is) de A que apresentou(aram) a(s) maior(es) média(s) usando o teste de Tukey.
9.5.4. O valor do F calculado para o fator B.
9.5.5. O(s) nível(is) de B que apresentou(aram) a(s) maior(es) média(s) usando o teste de Tukey.

9.6. Considere os resultados obtidos de um experimento instalado segundo o delineamento inteiramente casualizado, com 3 repetições, no esquema de parcelas subdivididas e o nível de 5% de significância, quando necessário:

Totais de Tratamentos

	B1	B2	B3	B4	Totais
A1	53,3	52,8	53,3	51,3	210,7
A2	47,9	44,2	48,7	51,0	191,8
A3	51,8	47,5	48,9	47,8	196,0
A4	50,0	44,2	48,9	51,2	194,3
A5	58,9	48,7	51,8	51,9	211,3
Totais	261,9	237,4	251,2	253,2	1004,1

Análise de Variância

FV	GL	SQ	QM	F
A		29,55		
Res(a)				
(Parcelas)		(45,26)		
B		20,60		
AxB				
Res(b)				
Total		137,58		

SQTratamentos = 70,26

Com base nas informações fornecidas, pede-se:

9.6.1. Os fatores A e B atuam independentemente? Justifique a sua resposta.

9.6.2. Baseado no resultado do teste F para a interação, proceda ao estudo do fator B, indicando qual(is) nível(is) de B que apresenta(m) maior(es) média(s). Use o teste de Tukey, se necessário.

9.7. Num artigo científico foram apresentados os resultados abaixo referente a um experimento em parcelas subdivididas instalado segundo o delineamento em blocos casualizados, com 5 repetições, em que o fator A foi distribuído às parcelas e o fator B foi distribuído às subparcelas:

Quadro de **MÉDIAS** de Tratamentos

	A1	A2	A3	
B1	23,80	14,00	13,20	17,00 a
B2	21,60	11,60	13,60	15,60 b
	22,70 A	12,8 B	13,40 B	

As médias seguidas por uma mesma letra maiúscula na linha, ou por uma mesma letra minúscula na coluna, não diferem entre si ao nível de 5% de probabilidade, pelo teste de Tukey e pelo teste F, respectivamente.

Dados: SQRes(b) = 26,60

No entanto, o autor não menciona no seu artigo um teste para a interação entre os fatores A e B. Com base nas informações acima, pede-se usando $\alpha = 5\%$:

9.7.1. Aplique o teste F para a interação entre os fatores A e B.

9.7.2. Baseado no resultado do teste F obtido no item anterior, os procedimentos adotado para comparar os níveis de A e os níveis de B estão corretos? Justifique a sua resposta. Não é necessário conferir os cálculos do autor, apenas discuta se o procedimento adotado é coerente com o resultado do teste F para a interação.

9.8. Abaixo, são mostrados os dados de um experimento em blocos ao acaso com parcelas subdivididas, onde o fator A com três níveis foi casualizado nas parcelas e o fator B com dois níveis foi casualizado nas subparcelas.

		Blocos			
Fator A	Fator B	1	2	3	4
A1	B1	58	77	38	52
A1	B2	44	59	30	34
A2	B1	85	90	73	77
A2	B2	59	68	45	55
A3	B1	66	93	67	64
A3	B2	54	75	53	48

Efetue o teste F para a interação AxB e proceda às comparações dos níveis dos fatores A e B pelo teste de Tukey, se necessário, de acordo com o resultado de significância para a interação. Utilize α = 5%.

9.9. Considere um experimento em parcelas subdivididas no delineamento inteiramente casualizado com 4 repetições, onde o fator A foi casualizado nas parcelas e fator B casualizado nas subparcelas, sendo dados:

Totais de Tratamentos

	B1	B2	B3	
A1	20,4	19,7	32,3	72,4
A2	11,3	10,6	18,0	39,9
	31,7	30,3	50,3	112,3

SQParcelas = 55,9836 e SQTotal = 121,4907.

Efetue o teste F para a interação AxB e proceda às comparações dos níveis dos fatores A e B pelo teste de Duncan, se necessário, de acordo com o resultado de significância para a interação. Utilize α = 5%.

10. Regressão

10.1. Introdução

Um fator em estudo num experimento pode ser classificado como qualitativo ou quantitativo. Um fator qualitativo é aquele onde os seus níveis diferem por algum atributo qualitativo. Como exemplos têm-se variedades, tipos de defensivos, métodos de conduzir uma determinada tarefa, etc. Por outro lado, um fator quantitativo é aquele onde os níveis se diferem com relação a quantidade do fator. Como exemplos têm-se temperatura, umidade, concentração de um princípio ativo, níveis de insumo, pH, etc.

Quando o fator é qualitativo, deve-se proceder à análise de variância dos dados e às comparações entre médias dos níveis do fator usando algum dos procedimentos para comparações múltiplas, quando o F for significativo. Para o caso de um fator quantitativo, deve-se estudar o efeito do fator quantitativo pó r meio de uma relação funcional entre o mesmo e a variável resposta. A técnica indicada neste caso é a análise de regressão.

A análise de regressão consiste na realização de uma análise estatística com o objetivo de verificar se a relação funcional estabelecida entre um fator quantitativo e uma variável resposta é significativa. Em outras palavras, consiste na obtenção de uma equação que tenta explicar a variação significativa de uma variável resposta em função da variação dos níveis de um ou mais fatores quantitativos.

10.2. Escolha do modelo para equacionar o fenômeno em estudo

Para tentar estabelecer uma equação que representa o fenômeno em estudo, pode-se plotar um diagrama de dispersão para verificar como se comportam os valores da variável resposta (Y) em função da variação dos níveis do fator quantitativo (X).

O comportamento de Y em relação a X, pode se apresentar de diversas maneiras: linear, quadrático, cúbico, exponencial, logarítmico, etc... . Para se estabelecer o modelo para explicar o fenômeno, deve-se verificar qual tipo de curva e equação de um modelo matemático que mais se aproxime dos pontos plotados no diagrama de dispersão.

Contudo, pode-se verificar que os pontos do diagrama de dispersão, não vão se ajustar perfeitamente à curva do modelo matemático proposto. Haverá na maioria dos pontos, uma distância entre os pontos do diagrama e aqueles obtidos quando a curva do modelo proposto é traçada. Isto acontece, devido ao fato do fenômeno que está em estudo, não ser um fenômeno matemático e sim um fenômeno que está sujeito a influências de inúmeros fatores. Assim, o objetivo da regressão é obter um modelo matemático que melhor se ajuste aos valores observados de Y em função da variação dos níveis da variável X.

O modelo matemático que irá ser ajustado deve satisfazer as seguintes condições:

- Modelo selecionado deve ser coerente para representar em termos práticos, o fenômeno em estudo;
- Modelo deve conter apenas as variáveis que são relevantes para explicar o fenômeno.

10.3. Método para obter a equação estimada

Como foi dito anteriormente, os pontos do diagrama de dispersão ficam um pouco distantes da curva do modelo matemático escolhido. Um dos métodos que se pode utilizar para obter a relação funcional, se baseia na obtenção de uma equação estimada de tal forma que as distâncias entre os pontos do diagrama e os pontos da curva do modelo matemático, no todo, sejam as menores possíveis. Este método é denominado de Método dos Mínimos Quadrados (MMQ). Em resumo por este método a soma de quadrados das distâncias entre os pontos do diagrama e os respectivos pontos na curva da equação estimada é minimizada, obtendo-se, desta forma, uma relação funcional entre X e Y, para o modelo escolhido, com um mínimo de erro possível.

10.3.1. Modelo linear de 1º grau

O modelo estatístico para esta situação seria:

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

em que

Y_i é o valor observado para a variável dependente Y no i-ésimo nível da variável independente X;

β_0 é a constante de regressão. Representa o intercepto da reta com o eixo dos Y;

β_1 é o coeficiente de regressão. Representa a variação de Y em função da variação de uma unidade da variável X;

X_i é o i-ésimo nível da variável independente X $(i = 1, 2, \ldots, n)$; e

e_i é o erro que está associado à distância entre o valor observado Y_i e o correspondente ponto na curva, do modelo proposto, para o mesmo nível i de X.

Para se obter a equação estimada, vamos utilizar o MMQ, visando a minimização dos erros. Assim, tem-se que:

$$e_i = Y_i - \beta_0 - \beta_1 X_i$$

elevando ambos os membros da equação ao quadrado,

$$e_i^2 = \left[Y_i - \beta_0 - \beta_1 X_i\right]^2$$

aplicando o somatório,

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}\left[Y_i - \beta_0 - \beta_1 X_i\right]^2 \qquad (1)$$

Por meio da obtenção de estimadores de β_0 e β_1, que minimizem o valor obtido na expressão anterior, é possível alcançar a minimização da soma de quadrados dos erros.

Sabemos do Cálculo que para se encontrar o mínimo de uma equação deve-se derivar a equação em relação à variável de interesse, e igualar a derivada resultante ao valor zero. Portanto, derivando a expressão (1) em relação a β_0 e β_1 e igualando-as a zero, obtém-se:

$$\begin{cases} \dfrac{\partial \sum_{i=1}^{n} e_i^2}{\partial \beta_0} = 0 \quad \Rightarrow \quad 2\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)(-1) = 0 \qquad \Rightarrow \quad \sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0 \\ \dfrac{\partial \sum_{i=1}^{n} e_i^2}{\partial \beta_1} = 0 \quad \Rightarrow \quad 2\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)(-X_i) = 0 \quad \Rightarrow \quad \sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)(X_i) = 0 \end{cases}$$

$$\begin{cases} \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n}\hat{\beta}_0 - \sum_{i=1}^{n}\hat{\beta}_1 X_i = 0 \qquad \Rightarrow \quad \sum_{i=1}^{n} Y_i - n\hat{\beta}_0 - \hat{\beta}_1\sum_{i=1}^{n} X_i = 0 \\ \sum_{i=1}^{n} Y_i X_i - \sum_{i=1}^{n}\hat{\beta}_0 X_i - \sum_{i=1}^{n}\hat{\beta}_1 X_i^2 = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} Y_i X_i - \hat{\beta}_0\sum_{i=1}^{n} X_i - \hat{\beta}_1\sum_{i=1}^{n} X_i^2 = 0 \end{cases}$$

$$\begin{cases} \sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} Y_i X_i = \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 \end{cases}$$

Este é o sistema de equações normais, que permite a obtenção de estimativas de β_0 e β_1, que minimizam a soma de quadrados dos erros.

Uma vez obtidas estas estimativas, podemos escrever a equação estimada:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

10.3.2. Modelo linear de 2º grau

O modelo estatístico para esta situação seria:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + e_i$$

em que,

Y_i é o valor observado para a variável dependente Y no i-ésimo nível da variável independente X;

β_0 é a constante de regressão;

β_1 é o coeficiente de regressão;

X_i é o i-ésimo nível da variável independente X $(i = 1, 2, \ldots, n)$;

β_2 é o coeficiente de regressão;

X_i^2 é o i-ésimo nível da variável independente X, elevado ao quadrado;

e_i é o erro que está associado à distância entre o valor observado Y_i e o correspondente ponto na curva para o mesmo nível i de X.

Utilizando o MMQ, no modelo de 2º grau, chegar-se-á ao seguinte sistema de equações normais, para se obter as estimativas de β_0, β_1 e β_2:

$$\begin{cases} \sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i + \hat{\beta}_2 \sum_{i=1}^{n} X_i^2 \\ \sum_{i=1}^{n} Y_i X_i = \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 + \hat{\beta}_2 \sum_{i=1}^{n} X_i^3 \\ \sum_{i=1}^{n} Y_i X_i^2 = \hat{\beta}_0 \sum_{i=1}^{n} X_i^2 + \hat{\beta}_1 \sum_{i=1}^{n} X_i^3 + \hat{\beta}_2 \sum_{i=1}^{n} X_i^4 \end{cases}$$

Uma vez obtidas estas estimativas, podemos escrever a equação estimada:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2$$

10.4. Análise de variância da regressão

A equação estimada obtida, apenas estabelece uma relação funcional, entre a variável dependente e a variável independente, para representar o fenômeno em estudo. Portanto a simples obtenção da equação estimada não responde ao pesquisador se a variação da variável independente influencia significativamente na variação da variável dependente.

Para se responder a esta pergunta, é necessário realizar um teste estatístico para as estimativas dos coeficientes da equação de regressão estimada. Um teste que pode ser realizado para verificar tal fato é o teste F da análise de variância. Portanto, é necessário realizar uma análise de variância dos dados observados, em função do modelo proposto.

Contudo, a estratégia da análise de variância depende se houve ou não repetições no experimento.

10.4.1. Apenas um único valor observado para cada nível da variável independente

Nesta situação não existe repetição. A única estimativa da variância residual é aquela dada pela falta de ajuste dos valores observados ao modelo ajustado. O quadro para a análise de variância para a regressão para esta situação é do seguinte tipo:

FV	GL	SQ	QM	F	$F_{tab;\,\alpha}$
Regressão	p	SQReg	$\frac{SQReg}{p}$	$\frac{QMReg}{QMInd}$	(p;n-1-p)
Independente da Regressão	n-1-p	SQInd	$\frac{SQInd}{n-1-p}$	-	
Total	n-1	SQTotal			

em que,

p = nº de coeficientes de regressão (não inclui o β_0)

n = nº de observações.

As fórmulas para a obtenção das somas de quadrados total e da soma de quadrados do independente da regressão são as mesmas, tanto para o modelo linear de 1º grau quanto para o de 2º grau, as quais são dadas a seguir:

$$SQTotal = \sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$$

SQInd = SQTotal - SQRegressão

Já a soma de quadrados para a regressão varia de acordo com o modelo em teste.

1º grau	2º grau
$SQRegressão = \hat{\beta}_0 \sum_{i=1}^{n} Y_i + \hat{\beta}_1 \sum_{i=1}^{n} Y_i X_i - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$	$SQRegressão = \hat{\beta}_0 \sum_{i=1}^{n} Y_i + \hat{\beta}_1 \sum_{i=1}^{n} Y_i X_i + \hat{\beta}_2 \sum_{i=1}^{n} Y_i X_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$

As hipóteses estatísticas para o teste F são as seguintes:

$H_0 : \beta_1 = \beta_2 = ... = \beta_p = 0$, o que significa dizer que as p variáveis independentes não exercem influência na variável dependente, segundo o modelo proposto.

$H_a : \beta_i \neq 0$, para pelo menos um i, o que significa dizer que pelo menos uma das p variáveis independentes exerce influência na variável dependente, segundo o modelo proposto.

O valor de F da análise de variância, deve ser comparado, com o valor de F tabelado (F_{tab}), o qual se obtém na tabela da distribuição F de acordo com o nível de significância do teste, e o número de graus de liberdade para a regressão e independente da regressão, ou seja:

$$F_{tab} = F_{\alpha}(p; n-1-p).$$

A regra decisória para o teste F é:

- Se $F \geq F_{tab} \Rightarrow$ Rejeita-se H_0 ao nível de significância que foi realizado o teste. Pode-se inferir que a variável independente influência significativamente a variável dependente Y.
- Se $F < F_{tab} \Rightarrow$ Não rejeita-se H_0 ao nível de significância que foi realizado o teste. Pode-se inferir que a variável independente **não** influência significativamente a variável dependente Y.

10.4.2. Mais de um valor observado para cada nível da variável independente

Nesta situação, existe mais de um valor observado para cada nível da variável independente. Assim é possível obter uma estimativa da variância residual tal como aquela obtida em modelos de delineamento, o que não é possível quando se tem uma única observação para cada nível da variável independente.

Normalmente o que se faz numa situação como esta é inicialmente proceder a uma análise de variância usual considerando o efeito do fator quantitativo como se fosse a fonte de variação tratamentos numa análise de variância usual. Isto é realizado para que se quantifique a variância residual. Posteriormente, o efeito de tratamentos é desdobrado nos efeitos associado a um ajuste de um modelo de regressão e também a falta de ajuste deste modelo. A escolha do modelo de regressão a ser ajustado é aquele que mais se aproxima dos pontos médios observados para cada nível da variável independente. O quadro abaixo resume o que acabou de ser descrito, para uma situação geral em que se está testando I níveis da variável independente em um experimento instalado segundo o delineamento inteiramente casualizado com K repetições. Pressupõe-se também que se está testando um modelo de regressão com p coeficientes de regressão. O total de observações neste experimento é igual a N=IK.

FV	GL	SQ	QM	F	$F_{tab;\,\alpha}$
Regressão	p	SQReg	$\frac{SQReg}{p}$	$\frac{QMReg}{QMRes}$	[p; I(K – 1)]
Falta de Ajustamento	I –1 – p	SQFalta	$\frac{SQFalta}{I-1}$	$\frac{QMFalta}{QMRes}$	[I – 1 – p; I(K – 1)]
(Tratamentos)	I – 1	SQTrat	-	-	
Resíduo	I(K – 1)	SQRes	$\frac{SQRes}{I(J-1)}$		
Total	IK – 1	SQTotal	-	-	

O teste F para a falta de ajustamento é realizado para verificar se o modelo adotado está se ajustando bem aos dados. Se o teste F para a falta de ajustamento for significativo, indica que o modelo ajustado não é apropriado e um novo modelo que se ajuste melhor aos dados deve ser testado. Se por outro lado, a falta de ajustamento for não-significativa indica que o modelo adotado se ajusta bem aos dados. Conseqüentemente faz sentido analisar o teste F para a fonte de variação regressão para saber se a variável independente tem influência significativa sobre a variável dependente.

No caso de falta de ajustamento significativa não faz sentido realizar o teste para a regressão, pois o modelo de regressão não se ajustou significativamente aos dados.

As hipóteses para a falta de ajustamento são:

H_0: a falta de ajustamento não é significativa

H_a: a falta de ajustamento é significativa

O valor tabelado de F para a falta de ajustamento é encontrado usando

$F_{tab} = F_{\alpha}(p; I-1-p)$

A regra decisória para o teste F para a falta de ajustamento é:

Se $F \geq F_{tab} \Rightarrow$ Rejeita-se H_0 ao nível de significância que foi realizado o teste. O modelo adotado não se ajusta bem aos dados. Um novo modelo deve ser testado.

Se $F < F_{tab} \Rightarrow$ Não rejeita-se H_0 ao nível de significância que foi realizado o teste. O modelo adotado se ajusta bem aos dados. Não há necessidade de se testar um novo modelo. Procede-se ao teste F para regressão.

O teste F para a regressão é idêntico ao caso anterior, ou seja, com apenas uma observação para cada nível da variável independente.

10.5. Coeficiente de determinação (R^2)

O coeficiente de determinação fornece uma informação auxiliar ao resultado da análise de variância da regressão, para verificar se o modelo proposto é adequado ou não para descrever o fenômeno.

Para o caso em que se tem uma única observação para cada nível da variável independente , o R^2 é obtido por :

$$R^2 = \frac{SQReg}{SQTotal}$$

Já para o caso em que se tem mais de um valor observado para cada nível da variável independente, o valor de R^2 é obtido por:

$$R^2 = \frac{SQReg}{SQTrat}$$

O valor de R^2 varia no intervalo de 0 a 1. Valores próximos de 1 indicam que o modelo proposto é adequado para descrever o fenômeno.

10.6. Exercícios

10.1. Verificar, utilizando os dados amostrais fornecidos abaixo, se a temperatura tem influência significativa sobre o comprimento de uma barra de aço. Utilize o modelo linear de 1º grau e o nível de 5% de significância.

Temperatura (ºC)	10	15	20	25	30
Comprimento (mm)	1003	1005	1010	1011	1014

10.2. Para verificar se existe uma relação linear de 1º grau entre Umidade Relativa (UR) do ar da secagem de sementes e a germinação das mesmas, um pesquisador realizou um teste com 4 diferentes valores para a % de UR do ar que atravessava as sementes armazenadas, obtendo-se os seguintes valores amostrais:

UR (%)	20	30	40	50
Germinação (%)	94	96	95	97

Ao nível de 5% de probabilidade, qual seria a conclusão do pesquisador? Qual seria a equação estimada?

10.3. Para o seguinte conjunto de valores de X (variável independente) e Y (variável dependente), faça a análise de regressão segundo o modelo linear de 1º grau e obtenha a equação de regressão estimada. Use o nível de significância de 5%.

X	2	4	6	8	10	12	14	16	18	20
Y	10,3	18,2	25,1	35,6	43,0	50,0	59,1	67,8	75,2	85,0

10.4. De acordo com os dados fornecidos abaixo para a variável X (dose do micronutriente Zn em ppm) e a variável Y (matéria seca em g/planta), verifique, usando o nível de 5% de probabilidade e o modelo linear de 2º grau, se a relação entre as variáveis X (independente) e Y (dependente) é significativa.

X	1,0	2,5	4,0	5,5	7,0	8,5
Y	20,3	31,3	34,6	35,1	30,2	19,7

10.5. O modelo linear abaixo foi proposto para explicar a relação entre a quantidade de ração fornecida e produção de leite por cabras:

$$Y_i = a + bX_i + e_i$$

Pede-se por meio dos dados abaixo, verificar se a ração influencia significativamente a produção de leite $(\alpha = 5\%)$:

Níveis de Ração (g)	50	75	100	125	150
Produção de leite (l/dia)	1,2	1,7	2,0	2,1	2,5

10.6. Para o modelo ajustado e dados fornecidos abaixo:

$$\hat{Y} = 140{,}7835 + 0{,}2737X - 0{,}000783X^2$$

SQIndependente da Regressão = 68,1691

$\sum_{i=1}^{7} Y_i = 1094{,}800$ $\qquad \sum_{i=1}^{7} Y_i^2 = 171712{,}384$

$\sum_{i=1}^{7} Y_i X_i = 166942{,}500$ $\qquad \sum_{i=1}^{7} Y_i X_i^2 = 35986875{,}000$

Proceder a análise de variância da regressão e concluir $(\alpha = 5\%)$

10.7. Para se avaliar o efeito de diferentes dosagens de um micronutriente no desenvolvimento de duas espécies vegetais, foi realizado em experimento fatorial 4x2 no D.B.C. com 5 repetições. Após a coleta e tabulação dos dados (em produção de matéria verde por determinada unidade de área) foi montado o seguinte quadro de totais de tratamentos:

	Dose 1	Dose 2	Dose 3	Dose 4	
Espécie 1	60	52	60	90	262
Espécie 2	56	50	40	40	186
	116	102	100	130	448

A análise de variância dos dados no computador forneceu o seguinte quadro (incompleto) da ANOVA:

F.V.	G.L.	S.Q.	Q.M.
Fator A	1		
Fator B	3	58,2	
Int. AxB		----	49,20
(Trat.)		----	
Blocos		----	
Resíduo		----	10,00
Total		----	

Com base nos dados apresentados acima, pede-se: (obs.: use α = 5%):
a) Obtenha a soma de quadrados para o fator A. Apresente os cálculos.
b) Os fatores em estudo atuam independentemente na variável em análise? JUSTIFIQUE.
c) Qual espécie deveria ser usada de modo a termos uma maior produção de massa verde, quando for usada a dose 3 do micronutriente? JUSTIFIQUE.
d) Como deveríamos continuar a análise caso fosse de nosso interesse determinar a melhor dose do micronutriente? Descreva a estratégia de análise de maneira resumida, apresentando a seqüência dos procedimentos a serem realizados juntamente com algumas discussões, mas sem precisar fazer nenhum tipo de cálculo

10.8. Suponha que um colega seu tenha usado um programa de computador para realizar a análise de regressão de um experimento no DIC com 4 repetições, no qual foi avaliado o efeito de 5 níveis de adubo na produção de soja. O orientador desse seu colega pediu que ele testasse três modelos. Como seu colega "matou" todas as aulas de estatística, ele foi pedir sua ajuda para a escolha do melhor modelo a partir dos dados abaixo, referentes à análise de cada modelo. Baseado no quadro fornecido abaixo, pede-se escolher o melhor modelo. Explique, para cada modelo, a razão dele ter sido selecionado ou eliminado. Use α = 5%.

MODELO 1

F.V.	G.L.	S.Q.	Q.M.
Regressão	1	36	36
Falta de Ajust.	3	60	20
(Tratamento)	(4)	96	
Resíduo	15	75	5
Total	19	171	

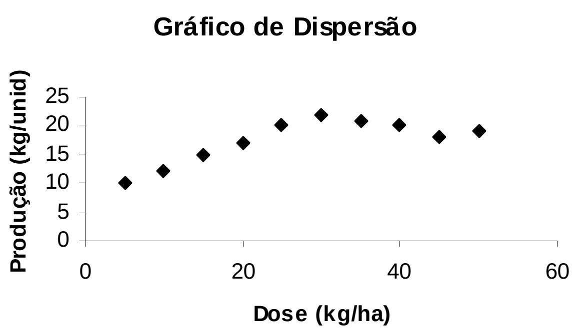
MODELO 2

F.V.	G.L.	S.Q.	Q.M.
Regressão	2	66	33
Falta de Ajust.	2	30	15
(Tratamento)	(4)	96	
Resíduo	15	75	5
Total	19		

MODELO 3

F.V.	G.L.	S.Q.	Q.M.
Regressão	3	76	25,3
Falta de Ajust.	1	20	20
(Tratamento)	(4)	96	
Resíduo	15	75	5
Total	19		

O gráfico de dispersão dos valores médios de produção em função das doses de adubo obtido pelo seu colega foi



Baseado nas informações fornecidas acima, pede-se escolher o melhor modelo. Explique, para cada modelo, a razão dele ter sido selecionado ou eliminado. Use α = 5%.

10.9. Com o objetivo de estudar o efeito da temperatura no ganho de peso de determinada espécie de animal de pequeno porte, foi realizado um estudo em que alguns animais foram submetidos a diferentes temperaturas no local em que eram confinados. Com base nos dados de ganho de peso, obtidos depois de determinado período, ajustou-se a seguinte equação de regressão:

$$\hat{Y} = -6{,}89 + 0{,}93X - 0{,}02X^2$$

Considerando que a análise de variância da regressão resultou em F significativo para regressão, pede-se:

a) Qual seria o ganho de peso (em quilos) esperado, se fosse mantida constante, no local de confinamento do animal em questão, a temperatura de 23 ºC?

b) Qual seria a temperatura a ser usada para que fosse obtido o máximo de ganho de peso?

10.10. Suponha que tenha sido realizada uma pesquisa a respeito da influência do tempo de estudo na nota da prova de determinada disciplina. Os dados obtidos com respeito a cinco alunos aleatoriamente entrevistados são dados abaixo:

X_i = Tempo de estudo (em horas)	2	3	4	5	6
Y = Nota obtida (em 10)	3	5	6	8	9

$$\sum X_i = 20 \quad \sum X_i^2 = 90 \quad \sum Y_i = 31 \quad \sum X_i Y_i = 139 \quad \sum Y_i^2 = 215$$

Pede-se:

a) Ajuste um modelo de regressão linear de 1º grau para tentar explicar a variação na nota do aluno em função do tempo de estudo.

OBS.: Indique a resolução, inclusive apresentando o sistema de equações normais.

b) Poderíamos dizer que o tempo de estudo influencia significativamente a nota obtida? (use α = 5%).

10.11. Com os dados relativos à equação de regressão $\hat{Y}_i = \hat{a}_0 - 10{,}38X_i + 1{,}08X_i^2$, obter a ANOVA da regressão e concluir para $\alpha = 5\%$.

DADOS:

$$\sum_{i=1}^{20} Y_i = 120{,}43 \quad \sum_{i=1}^{20} X_i Y_i = 340{,}87 \quad \sum_{i=1}^{20} X_i^2 Y_i = 4238{,}684$$

$$\sum_{i=1}^{20} Y_i^2 = 18375{,}38 \quad \sum_{i=1}^{20} X_i = 256{,}5 \quad \sum_{i=1}^{20} X_i^2 = 346{,}48$$

10.12. Obter a equação de regressão para o modelo $Y = a_0 + a_1X + a_2X^2 + e$ e concluir para $\alpha = 1\%$.

X	-4	-3	-2	-1	1	2	3	4
Y	1,2	10,1	13,2	14,3	14,1	12,7	8,5	0,3

10.13. Fazer a análise de variância da regressão, concluindo para $\alpha = 1\%$, dados :

$$\hat{Z} = -10,40 + 15,46W_i, i = 1,2,3,...,15$$

$$\sum Z_i^2 = 69,80 \sum Z_i = 2,3 \sum W_i = 1,46 \sum W_i Z_i = 1,92$$

10.14. Suponha que um biólogo realizou um experimento no DIC com 3 repetições, para comparar o efeito de 5 dosagens (X_i, em mg) de uma droga farmacêutica desenvolvida para aumentar o tempo de sono (Y_i, em horas). A análise dos dados oriundos deste experimento produziu as seguintes informações:

X_i	1			2			3			4			5		
Y_i	3	4	8	5	9	13	8	10	12	9	13	17	12	11	16

Usando o nível de 5% de significância, pede-se:

10.14.1. Proceda ao teste para a falta de ajustamento e conclua se o modelo de regressão linear de 1º grau é apropriado para descrever o tempo de sono em função da dosagem de sonífero.

10.14.2. O valor estimado para β_1 é estatisticamente diferente de zero? Justifique a sua resposta.

10.14.3. De acordo com a equação de regressão estimada, qual seria o tempo de sono dos ratos se uma dosagem de 17 mg fosse usada?

10.15. Foi realizada uma pesquisa para estudar o efeito de determinado medicamento usado no controle de peso de cavalos de corrida. Seis doses do medicamento foram ministradas a seis animais. A perda de peso obtida para estes animais, bem como a dose do medicamento ministrada a cada um deles é fornecida na tabela a seguir:

Dose (mg)	20	25	30	35	40	45
Perda de Peso (kg)	1,0	4,5	6,0	7,5	5,8	4,3

Suponha que o pesquisador decida usar o seguinte modelo linear de segundo grau:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i$$

$$\sum_{i=1}^{n} Y_i = 29,1 \qquad \sum_{i=1}^{n} Y_i X_i = 1000,50 \qquad \sum_{i=1}^{n} Y_i X_i^2 = 35787,50$$

$$\sum_{i=1}^{n} X_i = 195 \qquad \sum_{i=1}^{n} X_i^2 = 6775 \qquad \sum_{i=1}^{n} X_i^3 = 248625 \qquad \sum_{i=1}^{n} X_i^4 = 9521875$$

Com base nas informações fornecidas, pede-se:

10.15.1 A estimativa do intercepto (ou seja, constante da regressão)

10.15.2. As estimativas dos coeficientes de regressão, $\hat{\beta}_1$ e $\hat{\beta}_2$

10.15.3 A dose que proporciona o máximo de perda de peso

10.15.4. O valor do F da análise de variância da regressão calculado para testar se existe efeito do medicamento sobre a perda de peso, segundo o modelo proposto.

10.16. Um experimento foi instalado segundo o delineamento inteiramente casualizado para verificar se existe efeito significativo do fator quantitativo X sobre uma variável dependente Y. Suponha que foram utilizadas 2 repetições e que são fornecidas as seguintes informações:

Modelo adotado: $Y_i = \beta_o + \beta_1 X_i + \varepsilon_i$

FV	GL	SQ	QM	F
Regressão		76,05		
Falta de Ajustamento				
(Tratamentos)	(4)	101,60		
Resíduo				
Total		126,10		

Pede-se:

10.16.1. O valor do F calculado para a regressão
10.16.2. O valor do F calculado para a falta de ajustamento

10.17. Suponha que em uma pesquisa, 10 dosagens de uma droga foram ministradas a um grupo de 10 indivíduos, para verificar se o efeito da mesma era capaz de reduzir o peso em seres humanos. Cada dosagem foi testada em um único indivíduo. O modelo linear de 2º grau ajustado, a SQResíduo, as dosagens testadas e as respectivas perdas de peso observadas e alguns somatórios relacionados, foram:

$\hat{Y}_i = -1{,}15000 + 2{,}66174X_i - 0{,}09564X_i^2$ SQRegressão=179,87

Dosagem (mg)	2	4	6	8	10	12	14	16	18	20
Perda de peso (kg)	5	8	10	13	15	17	20	18	15	13

$\sum_{i=1}^{10} Y_i = 134$ $\sum_{i=1}^{10} Y_i^2 = 1990$ $\sum_{i=1}^{10} Y_iX_i = 1658$ $\sum_{i=1}^{10} Y_iX_i^2 = 23876$

$\sum_{i=1}^{10} X_i = 110$ $\sum_{i=1}^{10} X_i^2 = 1540$ $\sum_{i=1}^{10} X_i^3 = 24200$ $\sum_{i=1}^{10} X_i^4 = 405328$

Com base nas informações fornecidas acima e, usando o nível de 5% de significância quando necessário, pede-se:

10.17.1. É possível concluir que o uso da droga resulta em uma perda de peso significativa?
10.17.2. Qual a dosagem da droga que proporciona maior perda de peso?
10.17.3. Qual seria a perda de peso esperada se a dosagem de 35 mg fosse utilizada?

10.18. Suponha que um pesquisador, tendo como objetivo desenvolver uma bebida Láctea com sabor natural de laranja e temendo que o uso do suco natural resultasse em elevada acidez, resolveu testar 10 dosagens de suco natural (10, 15, 20, 25, 30, 35, 40, 45, 50 e 55 ml) com relação ao ph da bebida Láctea. Para tanto preparou um lote da fórmula básica da bebida Láctea. A fórmula básica é aquela que contém todos os ingredientes da bebida Láctea, exceto o suco de laranja. Como o lote era completamente homogêneo, dividiu o lote em 30 amostras. Procedeu-se então a distribuição inteiramente ao acaso das dosagens de suco de laranja às amostras. Ao final, cada dosagem foi designada a 3 amostras. Após a mistura do suco de laranja às amostras, o pH da bebida Láctea foi medido. Um gráfico de dispersão da dosagem versus pH, mostrou que o modelo linear de 1^o grau era indicado para estudar o fenômeno. Com base nos dados, as seguintes informações foram obtidas:

Quadro da ANOVA da Regressão

FV	GL	SQ	QM	F	$F_{5\%}$
Regressão		23,7692			
Falta de Ajustamento					
(Tratamentos)		(24,3720)			
Resíduo					
Total		25,1520			

Equação da regressão ajustada: $\hat{Y}_i = 7{,}63 - 0.08X_i$

Com base nas informações fornecidas acima e usando o nível de 5% de significância, pede-se:

OBSERVAÇÃO: UTILIZAR QUATRO DECIMAIS NOS CÁLCULOS

10.18.1. O modelo ajustado é adequado para descrever o fenômeno?
10.18.2. A dosagem do suco de laranja tem efeito significativo na acidez da bebida Láctea?
10.18.3. Quanto se espera que varie o pH da bebida Láctea em função da variação de 1 ml de suco de laranja?

10.19. Um padeiro resolveu testar 10 diferentes dosagens de um determinado tipo de fermento para verificar se o mesmo influenciava o peso final dos pães. Os resultados obtidos foram:

$\hat{Y}_i = 1{,}93 + 6{,}36X_i - 0{,}45X_i^2$ SQTotal = 310,40

$\sum_{i=1}^{10} Y_i = 194$ $\sum_{i=1}^{10} Y_i^2 = 4074$ $\sum_{i=1}^{10} Y_iX_i = 1179$ $\sum_{i=1}^{10} Y_iX_i^2 = 8461$

$\sum_{i=1}^{10} X_i = 55$ $\sum_{i=1}^{10} X_i^2 = 385$ $\sum_{i=1}^{10} X_i^3 = 3025$ $\sum_{i=1}^{10} X_i^4 = 25333$

Com base nas informações fornecidas acima e, usando o nível de 5% de significância quando necessário, pede-se:

10.19.1. É possível concluir que as dosagens do fermento influenciaram no peso final dos pães?
10.19.2. De acordo com a equação de regressão ajustada, qual é a dosagem estimada que proporciona o maior peso final de pães?

10.20. Uma droga desenvolvida para o controle do nível de açúcar (Y) foi testada em as doses 20, 30, 40, 50, 60, 70 e 80 mg (X). Os resultados apresentados abaixo foram publicados em uma revista científica:

$\hat{Y}_i = 286{,}32 + 0{,}83X_i$ SQTotal = 1933,71 SQRegressão=1905,75

Com base nestas informações, pede-se:

10.20.1. A droga tem influência significativa sobre o teor de glicose?

10.20.2. Qual é a estimativa do teor de glicose no sangue quando se usa a dose de 90 mg?

11. Respostas dos Exercícios

Pedimos aos estudantes que reportem erros nas respostas para o professor de sua turma, ou então para o professor Nerilson Terra Santos (nsantos@ufv.br). Favor reportar apenas erros nas respostas que você tiver certeza, por exemplo, a sua resposta e a de seus colegas para um determinado exercício não confere com o que está nesta seção.

A sua colaboração é muito importante. Obrigado.

Capítulo 1

1.1.	t = 5,24	$t_{1\%}$ (4) = 4,60	
1.2.	t = 2,21	$t_{5\%}$ (5) = 2,02	
1.3.	t = 1,26	$t_{5\%}$ (8) = 1,86	$s_c^2 = 6{,}25$
1.4.	t = -7,55	$t_{5\%}$ (6) = 1,94	
1.5.	t = -6,90	$t_{5\%}$ (4) = 2,13	
1.6.	F = 2,82	$F_{5\%}$ (6,9) = 3,37	
1.7.	F = 1,32	$F_{5\%}$ (5,5) = 5,05	
1.8.	t = 1,19	$t_{(5\%)}$(12) = 1,78	$s_c^2 = 20{,}29$
1.9.	t = -3,65	$t_{(5\%)}$(9) = 1,83	$s_d^2 = 3{,}41$
1.10.	t = 1,11	$t_{5\%}$ (18) = 2,10	$s_c^2 = 18{,}25$
1.11.	t = 3,10	$t_{5\%}$ (7) = 1,90	$s_d^2 = 7{,}73$
1.12.	t = 1,06	$t_{5\%}$ (10) = 1,81	$s_c^2 = 65{,}00$
1.13.	t = 14,61	$t_{1\%}$ (13) = 3,01	$s_d^2 = 75{,}46$
1.14.	t = 8,82	$t_{10\%}$ (8) = 1,86	$s_c^2 = 8{,}54$
1.15.	t = -0,39	$t_{1\%}$ (12) = 2,68	$s_c^2 = 11{,}45$
1.16.	F= 5,00	$F_{(1\%)}$ (5,5) = 10,97	
1.17.	t = 9,34	$t_{5\%}$ (8) = 1,86	$s_c^2 = 2{,}57$
1.18.	t = -3,87	$t_{1\%}$ (14) = 2,62	$s^2 = 25$
1.19.	t = -2,66	$t_{5\%}$ (9) = 1,83	$s_d^2 = 5{,}62$
1.20.	t = -3,07	$t_{5\%}$ (18) = 2,10	$s_c^2 = 0{,}0478$
1.21.	t = 11,54	$t_{5\%}$ (5) = 2,02	$s_d^2 = 47{,}06$
1.22.	1.22.1. (c)	1.22.2 (c)	1.22.3 (c)
1.23.	t = -3,45	$t_{1\%}$ (14) = 2,14	$s_c^2 = 295{,}81$
1.24.	t = -1,89	$t_{5\%}$ (9) = 3,25	$s^2 = 0{,}04$
1.25.	t = 0,84	$t_{5\%}$ (5) = 2,57	$s_d^2 = 0{,}86$
1.26.	t = 1,62	$t_{1\%}$ (14) = 2,62	$s_c^2 = 41{,}82$
1.27.	t = 4,05	$t_{5\%}$ (9) = 1,83	$s_d^2 = 2{,}44$
1.28.	t = 1,73	$t_{5\%}$ (14) = 1,76	$s_c^2 = 18{,}82$
1.29.	t = -19,53	$t_{5\%}$ (8) = 1,86	$s_c^2 = 2{,}7$. Marca A.
1.30.	t = -2,25	$t_{5\%}$ (8) = 1,94	$s^2 = 18{,}90$

Capítulo 2

2.1.

$\hat{C}_1 = -12{,}4$

$\hat{C}_2 = -3{,}0$

$\hat{C}_3 = 1{,}6$

2.2.

$\hat{C}_1 = -9{,}3$

$\hat{C}_2 = 0{,}7$

$\hat{C}_3 = -11{,}0$

$\hat{V}(\hat{C}_1) = 0{,}3525$

$\hat{V}(\hat{C}_2) = 0{,}15$

$\hat{V}(\hat{C}_3) = 0{,}2025$

2.3. São ortogonais

2.4. Não são ortogonais

2.5.

$C_1 = m_1 + m_2 + m_3 - 3m_4$

$C_2 = 2m_1 - m_2 - m_3$

$C_3 = m_2 - m_3$

2.6. Um **dos possíveis** grupos de contrastes ortogonais que podem ser formados é

$C_1 = m_1 + m_2 + m_3 + m_4 - 4m_5$

$C_2 = m_1 + m_2 + m_3 - 3m_4$

$C_3 = m_1 + m_2 - 2m_3$

$C_4 = m_1 - m_2$

2.7. Um **dos possíveis** grupos de contrastes ortogonais que podem ser formados é

$C_1 = 3m_1 + 4m_2 + 4m_3 + 3m_4 - 14m_5$

$C_2 = 3m_1 + 4m_2 + 4m_3 - 11m_4$

$C_3 = 3m_1 + 4m_2 - 7m_3$

$C_4 = m_1 - m_2$

2.8.

a) $\hat{C}_1 = 6{,}3$ $\hat{C}_2 = -17{,}1$ $\hat{C}_3 = -8{,}4$

b) $V(\hat{C}_1) = 0{,}18$ $V(\hat{C}_2) = 0{,}54$ $V(\hat{C}_3) = 0{,}9450$

c) os contrastes são ortogonais

2.9. Não são ortogonais

2.10.

a) $C_3 = m_1 - m_3$

b) 15,375

c) 0

2.11. É ortogonal. $C=m_1 + m_2 - 3m_3 + m_4$

2.12. É ortogonal. $C=4m_1 + 5m_2 - 14m_3 + 5m_4$

2.13. $C_2 = m_2 - m_3$

2.14. $C_2 = m_1 + 14m_2 - 15m_3$

2.15. $C_2 = m_2 - m_3$

2.16. $C_3 = m_1 + m_2 - 3m_3 + m_4$.

2.17. $C_3 = m_1 - m_3$.

2.18. $C_3 = 3m_1 - m_2 + 3m_3 - 5m_4$

2.19.

2.19.1 $C = m_1 + m_2 - 2m_3$, $\hat{C} = 55$;

2.19.2 $C = 3m_1 - 2m_2 - m_3$

2.20.

2.20.1. Compara o grupo de herbicida biológico com o grupo de químicos. $\hat{C} = 50$

2.20.2. Não. É necessário aplicar um teste de hipóteses para verificar se a estimativa encontrada é estatisticamente igual a zero.

2.20.3.105

2.20.4. $C_1 = 3m_1 - m_2 - m_3 - m_4$;

$C_2 = m_2 - m_3$ grupo químico nitrogênio com enxofre versus grupo químico nitrogênio com fósforo. $\hat{C}_2 = -1$

$C_3 = m_2 + m_3 - 2m_4$ grupo químico com nitrogênio versus grupo químico com inativadores de enzimas. $\hat{C}_3 = 13$

2.21.

2.21.1

a) $C_1 = m_1 - m_3$

b) $C_2 = m_1 - m_2$

c) $C_3 = m_3 - m_4$

2.21.2

a) 19

b) 35,10

c) 0

d) são ortogonais.

Capítulo 3

3.1. e

3.2. a

3.3. b

3.4.

a) Dez rações. Esta foi a fonte de variação introduzida pelo pesquisador.
b) Cada animal. Cada animal recebeu um dos tratamentos.
c) Nenhum.
d) Não. Pois o experimento não teve repetição.
e) Não, pois não foram usados os princípios básicos da experimentação.

3.5.

3.5.1 As 5 enzimas. A comparação do efeito destas 5 enzimas, foi o que motivou o pesquisador a instalar este ensaio.

3.5.2 Premeditada. O pesquisador sabia a princípio quais enzimas desejava comparar.

3.5.3 Cada amostra genômica. Esta foi a unidade que recebeu um tipo de tratamento.

3.5.4 Sim. Pois cada tratamento (enzima) foi designado a três unidades experimentais (amostra genômica). Os efeitos de ambiente que não são passíveis de controle, fazem com que as observações de um mesmo tratamento não sejam iguais.

3.5.5 Não, pois os tratamentos foram designados de uma forma sistemática às unidades experimentais.

3.5.6 Não, pois não houve nenhum controle na casualização. O princípio do controle local deve ser utilizado quando não existe uniformidade das condições experimentais.

3.5.7 Sim. Pois o princípio da repetição foi utilizado. A estimativa do erro experimental não é válida pois o princípio da casualização não foi utilizado.

3.5.8 Tempo gasto, pelo substrato químico contendo fragementos de DNA, para percorrer uma distâncias de 25 cm no gel.

3.6.

3.6.1 Os oito tipo de óleo, pois esta foi a fonte de variação introduzida pelo pesquisador com o propósito de comparação de seus efeitos.

3.6.2 Erro experimental (tipo aleatória), pois esta fonte de variação surgiu devido ao efeito de ambiente e não foi controlada pelo pesquisador.

3.6.3 Cada amostra básica, pois cada amostra básica recebeu um dos oito tipos de óleo.

3.6.4 Sim. Foram utilizados oito repetições.

3.6.5 Sim. Pois os tipos de óleo (tratamentos) foram distruibuídos ao acaso às amostras básicas (unidades experimentais).

3.6.6 Sim. Houve duas restrições na casualização de tal forma que cada bioquímico avaliou os oito tipos de óleo e cada lote recebeu os oito tipos de óleo.

3.6.7 O teor de gordura total, pois esta foi a característica avaliada para comparar o efeito dos tipos de óleo.

3.7.

3.7.1. Cada tábua de madeira, pois cada uma delas recebeu um dos 5 tratamentos em teste.

3.7.2. As 5 marcas de verniz, pois o pesquisador tinha por objetivo comparar os efeitos das 5 marcas de verniz com relação ao brilho proporcionado pelas marcas.

3.7.3.

Repetição: cada marca de verniz foi aplicada a 5 tábuas (unidade experimentais);

Casualização: a distribuição das marcas de verniz às tábuas foi feita ao acaso;

Controle local: a casualização sofreu a restrição de que cada tipo de madeira fosse testada com todas as marcas de verniz.

3.7.4. Sim, pois foram usadas repetições. A estimativa é válida pois foi usado o princípio da casualização.

3.7.5. Variações não controladas de ambiente. Não, pois geralmente não se conhece a origem destas variações não controladas.

3.7.6. Sim, pois sabia-se que a diferença de cor entre os diversos tipos de madeira poderia afetar a avaliação do verniz.

3.8.

3.8.1. Os seis sabores de sorvete. Esta foi a fonte de variação introduzida pelo pesquisador no experimento.

3.8.2. Sim, pois cada sabor apareceu seis vezes no experimento.

3.8.3. Sim, pois houve controle na casualização. Dois controles foram utilizados na casualização.

Capítulo 4

4.1. Fcal = 7,79 $\quad$ $Ftab_{5\%}$ (3,16) = 3,24 $\quad$ Rejeita-se Ho

4.2. a) Casualização e repetição

b) Cada atleta

c) Fcal = 1,39 $\quad$ $Ftab_{1\%}$ (2,12) = 6,93 $\quad$ Não Rejeita Ho

d) Qualquer técnica

4.3.

a) Sim. Fcal = 685,06. $Ftab_{5\%}$ (3,36) = 2,87 CV% = 3,94 Rejeita H_0

b) $Y = m_1 + m_2 - m_3 - m_4$ $\quad$ $\hat{Y} = 1,22$

c) $Y = m_1 - m_2$ $\quad$ $\hat{Y} = 8,25$

d) $Y = m_3 - m_4$ $\quad$ $\hat{Y} = -0,03$

4.4.

Fcal = 6,73 $Ftab_{5\%}$ (2,21) = 3,47 Rejeita H_0 As médias relativas aos 3 grupos diferem entre si.

4.5.

Fcal = 1,42 $\quad$ $Ftab_{5\%}$ (2,12) = 3,89 $\quad$ Não Rejeita-se Ho

4.6.

Fcal = 1,40 $\quad$ $Ftab_{1\%}$(2,12) = 6,93 $\quad$ Não Rejeita-se Ho

4.7.

4.7.1. Fcal = 6,94 $\quad$ $Ftab_{5\%}$(4,15) = 3,06 $\quad$ Rejeita-se Ho.

4.7.2. Existe efeito significativo das rações com relação ao ganho de peso médio proporcionado pelas mesmas.

4.7.3. $C = m_B + m_c - m_D - m_E \quad \hat{C} = -8{,}5$

4.7.4. CV = 20,68%. Valor alto indicando baixa precisão experimental.

Capítulo 5

5.1.

A numeração se refere aos exercícios do capítulo 4

4.1.

	Tukey	Duncan
	$\Delta = 4{,}79$	$D_4 = 3{,}82$ $D_3 = 3{,}71$ $D_2 = 3{,}54$
$\hat{m}_d = 31$	a	a
$\hat{m}_b = 27$	a b	b
$\hat{m}_c = 26$	b	b c
$\hat{m}_a = 23$	b	c

4.2. F não-significativo. Não é necessário aplicar teste de médias.

4.3.

	Tukey	Duncan
	$\Delta = 0{,}49$	$D_4 = 0{,}40$ $D_3 = 0{,}39$ $D_2 = 0{,}37$
$\hat{m}_1 = 14{,}81$	a	a
$\hat{m}_4 = 10{,}09$	b	b
$\hat{m}_3 = 10{,}06$	b	b
$\hat{m}_2 = 6{,}56$	c	c

4.4.

	Tukey	Duncan
	$\Delta = 9{,}23$	$D_3 = 7{,}99$ $D_2 = 7{,}61$
$\hat{m}_1 = 102{,}37$	a	a
$\hat{m}_3 = 95{,}12$	a b	a b
$\hat{m}_2 = 89{,}00$	b	b

4.5. F não-significativo. Não é necessário aplicar teste de médias.

4.6. F não-significativo. Não é necessário aplicar teste de médias.

4.7.

		Tukey	Duncan
		$\Delta = 3{,}68$	$D_5 = 2{,}78$ $D_4 = 2{,}73$ $D_3 = 2{,}66$ $D_2 = 2{,}53$
$\hat{m}_d$	= 11,00	a	a
$\hat{m}_e$	= 10,00	a b	a
$\hat{m}_a$	= 7,25	b c	b
$\hat{m}_b$	= 6,50	b c	b
$\hat{m}_c$	= 6,00	c	b

5.2.

		Tukey	Duncan
		$\Delta = 33$	$D_6 = 31$ $D_5 = 30{,}2$ $D_4 = 28{,}7$ $D_3 = 26$
$\hat{m}_3$	= 380	a	a
$\hat{m}_1$	= 370	a b	a
$\hat{m}_6$	= 367	a b	a
$\hat{m}_2$	= 338	b c	b
$\hat{m}_5$	= 325	c	b
$\hat{m}_4$	= 320	c	b

5.3.

Fcal = 2,64 $Ftab_{5\%}$ (5,18) = 2,77 Não existe diferença significativa entre os tratamentos. Portanto, o teste de Duncan não é necessário.

5.4.

Fcal = 15,23 $Ftab_{1\%}$ (3,12) = 5,95 Rejeita-se Ho. De acordo com o teste de Tukey (Δ = 2,11), os tipos de aleitamentos 1 e 2 proporcionaram as maiores médias de ganho de peso ($\hat{m}_1 = 9{,}30$ e $\hat{m}_2 = 11{,}20$).

5.5.

a. Fcal = 24,24 $Ftab_{5\%}$ (4,15) = 3,06 Rejeita-se Ho.

b. De acordo com o teste de Duncan (D_5 = 1,58, D_4 = 1,55, D_3 = 1,51, D_2 = 1,44), a marca E foi a mais lenta ($\hat{m}_E = 13{,}75$).

c. De acordo com o teste de Tukey (Δ = 2,10), as marcas E e A foram as mais lentas ($\hat{m}_E = 13{,}75$ e $\hat{m}_A = 11{,}75$).

d.

$C_1 = 4m_A - m_B - m_C - m_D - m_E$ $\hat{C}_1 = +4{,}0$ $t_c=1{,}86$ $t_{5\%(15)}=2{,}13$ NRH_0; S=7,48 NRH_0

$C_2 = m_B + m_C - m_D - m_E$ $\hat{C}_2 = -7{,}0$ $t_c=-7{,}63$ $t_{5\%(15)}=2{,}13$ RH_0; S=3,34 RH_0

$C_3 = m_B - m_C$ $\hat{C}_3 = +3{,}5$ $t_c=5{,}17$ $t_{5\%(15)}=2{,}13$ RH_0; S=2,36 RH_0

$C_4 = m_D - m_E$ $\hat{C}_4 = -2{,}5$ $t_c=-3{,}69$ $t_{5\%(15)}=2{,}13$ RH_0; S=2,36 RH_0

5.6.

a. Fcal = 6,24 $Ftab_{5\%}$ (3,26) = 2,98 Rejeita-se Ho.

b. $\hat{C}$ = 3. t = 2,42. $t_{tab} = t_{0,05}(26) = 2,06$. Rejeita-se Ho.
S = 3,70 Não rejeita-se Ho.

5.7.

	Tukey	Duncan
	Δ=15,96	D_6=11,77 D_5=11,60 D_4=11,38 D_3=11,07 D_2=10,53
$\hat{m}_1$ = 203,36	a	a
$\hat{m}_3$ = 196,58	a b	a b
$\hat{m}_5$ = 188,86	a b	b c
$\hat{m}_2$ = 182,38	b	c
$\hat{m}_4$ = 165,38	c	d
$\hat{m}_6$ = 153,20	c	e

Capítulo 6

6.1.

Fcal = 33,58 $\qquad$ $Ftab_{5\%}$ (4,8) = 3,84

	Tukey	Duncan
	Δ = 13,54	D_5= 9,75 D_4 = 9,6 D_3 = 9,39 D_2 = 9
$\hat{m}_5$ = 107	a	a
$\hat{m}_4$ = 92	b	b
$\hat{m}_3$ = 87	b	b
$\hat{m}_2$ = 72	c	c
$\hat{m}_1$ = 67	c	c

6.2

a) DBC, pois a divisão foi realizada de modo que houvesse homogeneidade entre as unidades experimentais dentro de cada grupo, ficando a heterogeneidade existente entre os grupos passível de ser quantificada e, assim, isolando os reais efeitos de tratamentos.
b) Sim. Fcal = 178 $\qquad$ $Ftab_{1\%}$ (3;18) = 5,09 $\qquad$ Rho
c) Tukey
Δ = 3,83

$\hat{m}_3$ = 46	a
$\hat{m}_1$ = 31,6	b
$\hat{m}_2$ = 31,4	b
$\hat{m}_4$ = 21,6	c

6.3.

a) Fcal = 5,87 $\qquad$ $Ftab_{5\%}$ (4,12) = 3,26

b) Tukey $\qquad$ c) Duncan
Δ = 13,12 $\qquad$ D_5= 9,77 D_4 = 9,68 D_3 = 9,39 D_2 = 8,96

$\hat{m}_5$ = 155,12 a a
$\hat{m}_1$ = 142,8 a b b
$\hat{m}_4$ = 140,01 b b
$\hat{m}_3$ = 138,74 b b
$\hat{m}_2$ = 138,02 b b

d) $\hat{Y}$ = -29,42 R H_0 S = 25,73

e) Não se aplica o teste t, pois os contrastes não são ortogonais.

6.4.
a) Fcal = 7,67 $Ftab_{5\%}$ (4,12) = 3,26 R.: Sim.
b)
Tukey Δ = 4,99

$\hat{m}_5$ = 11,4 a
$\hat{m}_2$ = 6,3 b
$\hat{m}_4$ = 6,0 b
$\hat{m}_3$ = 5,5 b
$\hat{m}_1$ = 3,0 b
R.: Levedura tipo 5

Duncan
D_5 = 3,72 D_4= 3,68 D_3= 3,57 D_2= 3,41
$\hat{m}_5$ = 11,4 a
$\hat{m}_2$ = 6,3 b
$\hat{m}_4$ = 6,0 b
$\hat{m}_3$ = 5,5 b
$\hat{m}_1$ = 3,0 b
R.: Leveduras tipo 1,2,3e 4.

6.5.
a) Repetição: cada tipo de pneu foi submetido a três repetições. Casualização; os tipos de pneus foram submetidos a sorteios dentro das respectivas sub-áreas. Controle Local: a área total foi submetido a várias subdivisões, visando proporcionar maior precisão ao experimento.
b) DBC: as sub-áreas formadas atuam como blocos no experimento.
c) Fcal = 20,9 $Ftab_{5\%}$ (3,6) = 4,76
d) Duncan
D_4 = 1,75 D_3= 1,72 D_2 = 1,67
$\hat{m}_4$ = 33 a
$\hat{m}_3$ = 31,3 b
$\hat{m}_2$ = 29,3 c
$\hat{m}_1$ = 28 c

6.6.

a) V(c) = 0
b) C_1 Não é contraste, logo não aplica teste de média
C_2 = -30 S = 30,4 N RH_0
c) Não se aplica o teste t, pois os contrastes não são ortogonais.
d) $C = m_1 + m_2 – 2m_3$

6.7.
a) 5 vezes, pois o número de graus de liberdade para blocos é igual a 4.
b) Testa-se se há diferença significativa entre a durabilidade dos 4 microaspersores. Sendo F tabelado a 5% igual a 3,49 e enunciando as hipóteses:
Ho: $m_1 = m_2 = m_3 = m_4$ vs Ha: Não Ho, verifica-se que o Fcal é significativo, portanto existe pelo menos um contraste entre as médias de durabilidade dos microaspersores estatisticamente diferente de zero.
c) Deveríamos aplicar um teste de médias; Tukey ou Duncan, por exemplo.
d) Ex.: Scheffé C = 36 + 40 + 60 – 3x40 = 16 S = 35,45 NRH_0

6.8.
a) Fcal = 37,7 $Ftab_{5\%}$ (4,12) = 3,26
b) Duncan D_5= 3,77 D_4 = 3,74 D_3 = 3,62 D_2 = 3,45

$\hat{m}_C$ = 25	a	
$\hat{m}_B$ = 24,75	a	
$\hat{m}_E$ = 16,25	b	
$\hat{m}_D$ = 12		c
$\hat{m}_A$ = 10,2		c

c) tcal = 3,99 $ttab_{5\%}$ (12) = 2,18 RHo

6.9.
Fcal = 4,20 $Ftab_{1\%}$ (4,12) = 5,41 Não Rejeita-se Ho

6.10.
6.10.1. Delineamento em Blocos Casualizados. Porque o melhorista ao instalar o experimento subdividiu a área total (heterogênea) em sub-áreas (homogêneas) entre si. Pode-se dizer também que o pesquisador utillizou os três princípios básicos da experimentação (repetição, casualização e controle local).
6.10.2. Fcal = 2,02 $Ftab_{5\%}$ (4,12) = 3,26 Não Rejeita-se Ho, consequentemente, não é necessário proceder ao teste de Duncan.

6.11.
6.11.1. (c)
6.11.2. (h)
6.11.3. (c)
6.11.4. (c)
6.11.5. (d)

6.12.

6.12.1. Fcal = 177,99 $Ftab_{1\%}(3,18) = 5,09$ Rejeita-se Ho.

6.12.2. Δ = 3,84 O tratamento 4 apresentou a menor média (21,57) portanto é o desejado.

6.13.

6.13.1. (c). Porque o teste de Scheffé pode ser aplicado a qualquer contraste sem nenhuma restrição e o teste t a contrastes estabelecidos "a priori" e ortogonais. Porém, os mesmos apresentam sensibilidades diferentes em detectar diferenças significativas.

6.13.2. (c). Porque o teste t pode ser aplicado para avaliar contrastes estabelecidos "a priori" e ortogonais e o teste de Tukey a todos os possíveis contrastes que envolvem duas médias. Porém os mesmos apresentam sensibilidades diferentes em detectar diferenças significativas.

6.13.3. Sim. S = 3,9. $\hat{Y} = -1{,}4$. Não se rejeita Ho.

6.14.

6.14.1. $Y = m_i - m_j$ para $i \neq j = 1, 2, 3$ e 4.

Número máximo de contrastes = 6.

6.14.2. Não, pois a diferença entre $\hat{m}_1$ e $\hat{m}_2$ é igual a 21 a qual é superior ao valor do Δ.

6.14.3. O que obteve $\Delta_1 = 5$, pois o Δ é função do QMResíduo, o qual é um "indicativo" da precisão experimental.

6.14.4.

$\hat{m}_1 = 100$ a

$\hat{m}_4 = 92$ a b

$\hat{m}_3 = 88$ a b

$\hat{m}_2 = 79$ b

6.14.5. O teste recomendado é o teste de Scheffé, pois existem contrastes envolvendo duas ou mais médias e os contrastes não formarem um grupo de contrastes ortogonais.

Capítulo 7

7.1.

a) Fcal = 12,09 $Ftab_{5\%}$ (4,12) = 3,26

b) Tukey Δ = 107,54

$\hat{m}_C = 604{,}8$ a

$\hat{m}_A = 492{,}6$ b

$\hat{m}_B = 440{,}8$ b

$\hat{m}_D = 413{,}4$ b

$\hat{m}_E = 401$ b

A Variedade Co 297 deve ser recomendada.

7.2.

a) Fcal = 8,198 $\quad$ Ftab$_{5\%}$ (4,12) = 3,26

b.1) Duncan D_5= 8,55 D_4 = 8,47 D_3 = 8,22 D_2 = 7,84

$\hat{m}_2$ = 60 a

$\hat{m}_5$ = 52,5 a b

$\hat{m}_1$ = 50 b

$\hat{m}_3$ = 47,5 b c

$\hat{m}_4$ = 40 c

Conclusão: os tratamentos 2 e 5 devem ser recomendados

b.2) Conclusão: os tatamentos 3 e 4 devem ser recomendados

7.3.

Fcal = 17,84 $\quad$ Ftab$_{5\%}$ (4,12) = 3,26

Tukey Δ = 107,54

$\hat{m}_1$ = 604,8 a

$\hat{m}_2$ = 509,8 a b

$\hat{m}_3$ = 469,8 b c

$\hat{m}_4$ = 394 c d

$\hat{m}_5$ = 346,8 d

7.4.

a) Fcal = 9,01 $\quad$ Ftab$_{5\%}$ (4,12) = 3,26

b) C = 4md – ma –mb – mc – me

c) Scheffé

$\hat{Y}$ = 33,66 $\quad$ S = 54,05

Teste t $\quad$ tcal = 2,249 $\quad$ ttab$_{5\%}$ (12) = 2,18

7.5.

Fcal = 31,73 Ftab$_{1\%}$ (6,30) =3,47 Rejeita-se Ho. Existe efeito significativo de forrageiras com relação a produção de matéira seca.

7.6.

7.6.1. cada litro de leite

7.6.2. os tipos de bacilos

7.6.3. 2 vezes

7.6.4. Delineamento em Quadrado Latino

7.6.5. (a)

7.6.6. (g)

Capítulo 8

8.1. Ftab = 5,32 $\quad$ FcalcAxB = 21,81 $\quad$ RHo

8.2. Ftab5% (1,16) = 4,49 FcalcAxB = 4,95 $\quad$ RHo

8.3.

a) Sim $\quad$ Ftab(1;20) = 8,10 $\quad$ FcalcAxB = 0,51 $\quad$ NRHo

b) FcalcA = 86,05 $\quad$ RHo

$\hat{m}_{R1}$ = 49,67 a
$\hat{m}_{R2}$ = 42,08 b
c) FcalcB = 15,80 Rho
$\hat{m}_{A1}$ = 47,50 a
$\hat{m}_{A2}$ = 44,25 b

8.4.
a) Não. VxF: F_{tab} = 3,49 F_{calc} = 27,23 $\rightarrow$ RHo
b) F/V1: F_{calc} = 17,07 F_{tab} (2;20) = 3,49
Teste Tukey Δ = 2,29
$\hat{m}_{F2/V1}$ = 8 a
$\hat{m}_{F1/V1}$ = 4 b
$\hat{m}_{F3/V1}$ = 3 b

c) V2/F3: $\hat{m}$ = 8 F_{tab} (1;20) = 4,35 $\rightarrow$ RHo

8.5.
a) Sim. VxE: F_{calc} = 0,75 F_{tab} = 4,49
b) Ambas fornecem a mesma produção. V: F_{calc} = 1,70 F_{tab} = 4,49
c) O espaçamento 2 ($\hat{m}_2$ = 7,7). E: F_{calc} = 6,79 F_{tab} = 4,49

8.6.
a) Não FcalcRaçãoxRaça = 7,17 FtabRaçãoxRaça = 3,68
b) F Ração/Raça1 = 3,13
F Ração/Raça2 = 6,13
F Ração/Raça3 = 10,13
Ftab(1;15) = 4,54

8.7. Questão teórica

8.8
a) 5 valores - Cada valor corresponde a um total de tratamento, repetidos 5 vezes.
b) 3, 7, 4, 28, 39
c) Meio de Cultura e tipo de Fungo, respectivamente.
d) Não. Fcalc = 4,92 Ftab = 4,57
e) Qualquer um Fcalc = 0,16 Ftab = 7,64

8.9.
a) Estudar os fatores isoladamente.
b) Efetuar o desdobramento dos fatores

8.10.
FA*B = 3,62*
FA/B1 = 30,06 FA/B2 = 36,82* FA/B3 = 51,75*

FB/A1 = 18,72* FB/A2 = 72,64* FB/A3 = 32,778 FB/A4 = 42,56
$\Delta_{B/A1}$ = 2,98 $\Delta_{A/B1}$ = 3,29

8.11.
FA = 0,81ns FB = 4,81* FA*B = 0,64ns
Teste Duncan para fator B D_3 = 262,26 D_2 = 249,77

8.12. Questão teórica -> ver teoria

8.13.
FA = 7,45ns FB = 5,16ns FA*B = 5,01ns

8.14.
a) FA/B4 = 11,62*
b) FB/A1 = 0,04ns FB/A2 = 5,10* Δ = 11,29

8.15.
a) $\hat{C}$ = -15,71ns S = 15,87
b) Fcal = 6,09* Ftab5% (1,14) = 4,60

8.16. Tukey Δ = 3,55
Médias dos niveis de A
$\hat{m}_{A1}$ = 19,8 a
$\hat{m}_{A2}$ = 18,4 ab
$\hat{m}_{A3}$ = 16,2 b
$\hat{m}_{A4}$ = 15,4 b

8.17.
a) FA*B = 9,33*
b) FA/B1 = 0,095ns
FA/B2 = 24,8*
FA/B3 = 0,097ns

8.18.
Interação: Ftab5%(2,24) = 3,40. Não Rejeita-se Ho a 5% de probabilidade.

Ração: Fcal = 4,34 Ftab5%(2,24) = 3,40. Rejeita-se Ho a 5% de probabilidade. D_3 = 13,93. D_2 = 13,25. A ração A deve ser recomendada.

Proteína: Fcal = 5,11 Ftab5%(1,24) = 4,26. Rejeita-se Ho a 5% de probabilidade. O nível alto de proteína dever ser recomendado.

8.19.

8.19.1

FV	GL	SQ	QM	F	$F_{tab;\ 5\%}$
T	2 – 1 = 1	2394,13			
H	3 – 1 = 2	1323,47			
T*H	1 × 2 = 2	20,27	10,13	0,80	(2; 24) = 3,40
(Tratamentos)	(2×3) – 1 = 5	3737,87	-		
Resíduo	29 – 5 = 24	304,00	12,67		
Total	(2 × 3 × 5) – 1 = 29	4041,87			

Conclusão: 0,80 < 3,40 → Não RH_0 a 5% de probabilidade. Logo, os fatores, horário de colheita e tipo de colheitadeira, **atuam** independentemente na perda de grãos.

8.19.2.

FV	GL	SQ	QM	F	$F_{tab;\ 5\%}$
H	2	1323,47	661,73	52,24	(2; 24) = 3,40
Resíduo	24	304,00	12,67		

H_0: $m_{H1} = m_{H2} = m_{H3}$
H_a: Não H_0

Conclusão: 52,24 > 3,40 → RH_0 a 5% de probabilidade. Logo existe pelo menos um contraste estatisticamente diferente de zero entre médias de **horário de colheita** com relação a perda de grãos.

Teste de Tukey e Duncan

Hipóteses

H_0: $m_{Hi} - m_{Hu} = 0$ para $i \neq u = 1, 2, 3.$
H_a: $m_{Hi} - m_{Hu} \neq 0$

DMS

$$\sqrt{\frac{QMRes}{I \times K}} = \sqrt{\frac{12,67}{2 \times 5}} = 1,13$$

Tukey: $\Delta = q_{tab} \times \sqrt{\frac{QMRes}{2 \times 5}}$ = 3,53 × 1,13 = 3,99

$q_{tab} = q_{0,05}(3;\ 24) = 3,53$

Duncan: $D_i = z_i \times \sqrt{\frac{QMRes}{2 \times 5}}$

i = 3 → z_3 = 3,07 → D_3 = 3,47
i = 2 → z_2 = 2,92 → D_2 = 3,30

Totais de Tratamentos

	H1	H2	H3	Totais
T1	208$^{(5)}$	217	282	707$^{(15)}$
T2	288	317	370	975
Totais	496$^{(10)}$	534	652	1682$^{(30)}$

Médias			Tukey	Duncan
$\hat{m}_{H3}$	$= \dfrac{652}{2 \times 5}$	= 65,2	a	a
$\hat{m}_{H2}$	=	= 53,4	b	b
$\hat{m}_{H1}$	=	= 49,6	b	c

8.19.3.

$\hat{C} = 2\hat{m}_{H1} - \hat{m}_{H2} - \hat{m}_{H3}$ = 2×65,2 – 53,4 – 49,6 = 27,40

H_0: C = 0
H_a: C ≠ 0

$$V\left(\hat{C}\right) = \frac{QMRes}{J \times K}\sum_{i=1}^{3} a_i^2 = \frac{12,67}{2 \times 5}\left(2^2 + (-1)^2 + (-1)^2\right) = 7,6$$

$$t = \frac{\hat{C}}{\sqrt{\hat{V}\left(\hat{C}\right)}} = \frac{27,40}{\sqrt{7,6}} = 9,94 \quad t_{tab} = t_{5\%}(24) = 2,06$$

Conclusão: |9,94| > 2,06 → RH_0 a 5% de probabilidade

$$S = \sqrt{(I-1) \times F_{tab} \times \hat{V}(\hat{C})} = \sqrt{(3-1) \times 3,40 \times 7,6} = 7,19$$

Conclusão: |27,40| >7,19 → RH_0 a 5% de probabilidade

8.19.4.

FV	GL	SQ	QM	F	$F_{tab;\ 5\%}$
T	2 – 1 = 1	2394,13	2394,13	189,01	(1; 24) = 4,20
Resíduo	29 – 5 = 24	304,00	12,67		

H_0: $m_{T1} = m_{T2}$
H_a: Não H_0

Conclusão: 189,01 > 3,40 → RH_0 a 5% de probabilidade. Logo existe pelo menos um contraste estatisticamente diferente de zero entre médias de **tipo de colheitadeira** com relação a perda de grãos.

Teste de Tukey e Duncan

Observação: A aplicação de tais testes é desnecessária, pois o teste F (1 GL para T) já é conclusivo. Apresentamos apenas para mostrar as diferenças entre aplicar para um fator com três níveis (H) e um fator com dois níveis (T)

Hipóteses	
H_0: $m_{Tj} - m_{Tu} = 0$	para $j \neq u = 1, 2$
H_a: $m_{Tj} - m_{Tu} \neq 0$	

DMS

$$\sqrt{\frac{QMRes}{3 \times 5}} = \sqrt{\frac{12,67}{15}} = 0,92$$

Tukey: $\Delta = q_{tab} \times \sqrt{\frac{QMRes}{3 \times 5}}$ = 2,92 × 0,92 = 2,69

$q_{tab} = q_{0,05}(2; 24) = 2,92$

Duncan: $D_i = z_i \times \sqrt{\frac{QMRes}{3 \times 5}}$

i = 2 $\rightarrow$ $z_2 = 2,92$ $\rightarrow$ $D_2 = 2,68$

Totais de Tratamentos

	H1	H2	H3	Totais
T1	208$^{(5)}$	217	282	707$^{(15)}$
T2	288	317	370	975
Totais	496$^{(10)}$	534	652	1682$^{(30)}$

Médias				Tukey	Duncan
$\hat{m}_{T2}$	=	$\frac{975}{3 \times 5}$	= 65,00	a	a
$\hat{m}_{T1}$	=		= 47,13	b	b

8.19.5.

Observação: A aplicação de tais testes é desnecessária, pois o teste F (1 GL para T) já é conclusivo. Apresentamos apenas para mostrar as diferenças entre aplicar para um fator com três níveis (H) e um fator com dois níveis (T)

$\hat{C} = \hat{m}_{T1} - \hat{m}_{T2}$ = 47,13 – 65,00 = - 17,87

H_0: C = 0
H_a: C $\neq$ 0

$$V\left(\hat{C}\right) = \frac{QMRes}{I \times K}\sum_{j=1}^{2} a_j^2 = \frac{12,67}{3 \times 5}\left(1^2 + \left(-1\right)^2\right) = 1,69$$

$$t = \frac{\hat{C}}{\sqrt{\hat{V}\left(\hat{C}\right)}} = \frac{-17,87}{\sqrt{1,69}} = -13,75 \quad t_{tab} = t_{5\%}(24) = 2,06$$

Conclusão: |-13,75| > 2,06 $\rightarrow$ RH_0 a 5% de probabilidade

$$S = \sqrt{(J-1) \times F_{tab} \times \hat{V}(\hat{C})} = \sqrt{(2-1) \times 4,20 \times 1,69} = 2,67$$

Conclusão: |-17,87| > 2,67 $\rightarrow$ RH_0 a 5% de probabilidade

8.20.

8.20.1. Interação: Fcal = 49,97 $Ftab_{5\%}(1,16)$ = 4,49. Rejeita-se Ho. Os fatores não atuam independentemente.

8.20.2. B/A1: Fcal = 21,62 $Ftab_{5\%}(1,16)$ = 4,49. Rejeita-se Ho. Quando se usa o controle de qualidade A1 processo de fabricação B1 é o mais rápido.

8.21.

8.21.1. Interação: Fcal = 0,037 $Ftab_{1\%}(2,30)$ = 5,39. Não rejeita-se Ho. Os fatores atuam independentemente.

8.21.2. Fator A: Fcal = 4,39. $Ftab_{1\%}(1,30)$ = 7,56. Não rejeita-se Ho. Os dois métodos de aceleração proporcionam em média igual consumo.

8.22.

8.22.1. Interação: Fcal = 267,58. $Ftab_{5\%}(3,14)$ = 3.34. Rejeita-se Ho. Os fatores A e B não atuam independentemente.

8.22.2. B/A_2: Fcal = 295,73. $Ftab_{5\%}(1,14)$ = 4,60. Rejeita-se Ho. Logo a média de B_1/A_2 é estatísticamente maior do que a de B_2/A_2.

8.22.3.

Δ = 8,51

$\hat{m}_{A3/B2} = 135,00$ a

$\hat{m}_{A4/B2} = 99,67$ b

$\hat{m}_{A1/B2} = 28,33$ c

$\hat{m}_{A2/B2} = 19,33$ d

8.23.OSERVAÇÃO: **VALORES APROXIMADOS**

8.23.1. 16,72

8.23.2. 42,54

8.23.3. 16,39

8.24.

8.24.1. Interação: Fcal = 25,10. $Ftab_{1\%}(2,18)$ = 6,01. Rejeita-se Ho. Os fatores A e B não atuam independentemente.

8.24.2. A/B1: Fcal = 34,03. $Ftab_{1\%}(2,18)$ = 6,01. Rejeita-se Ho. Existe pelo menos um contraste, estatisticamente diferente de zero, entre médias do fator A dentro do nível 1 de B.

8.24.3. B/A2: Fcal = 62,02. $Ftab_{1\%}(1,18)$ = 8,29. Rejeita-se Ho. O nível B1 apresenta maior média quando o nível 2 de A é considerado.

8.25.

FV	GL	SQ	QM	F	$F_{tab;\ 1\%}$
Recipientes (R)	2	92,86	-	-	
Espécie (E)	1	19,08	-	-	
Interação RxE	2	63,76	31,88	$24,91^{**}$	(2; 18) = 6,01
(Tratamentos)	(5)	(175,70)	-	-	
Resíduo	18	23,09	1,28		
Total	23	198,79			

** Significativo ao nível de 1% de probabilidade

R/E

FV	GL	SQ	QM	F	$F_{tab;\ 1\%}$
R/E_1	2	87,12	43,56	$34,03^{**}$	(2; 18) = 6,01
R/E_2	2	69,50	34,75	$27,15^{**}$	(2; 18) = 6,01
Resíduo	18	-	1,28		

** Significativo ao nível de 1% de probabilidade

R/E1

$\Delta = 2{,}66$

$\hat{m}_{R2/E1} = \dfrac{103{,}5}{4} = 25{,}875$ a

$\hat{m}_{R1/E1} = \quad = 25{,}650$ a

$\hat{m}_{R3/E1} = \quad = 20{,}050$ b

R/E2

$\Delta = 2{,}66$

$\hat{m}_{R1/E2} = \dfrac{101{,}3}{4} = 25{,}325$ a

$\hat{m}_{R3/E2} = \quad = 21{,}325$ b

$\hat{m}_{R2/E2} = \quad = 19{,}575$ b

E/R

FV	GL	SQ	QM	F	$F_{tab;\ 1\%}$
E/R_1	1	0,21	0,21	0,16	(1;18) = 8,29
E/R_2	1	79,38	79,38	$62,02^{**}$	(1;18) = 8,29
E/R_3	1	3,25	3,25	2,54	
Resíduo	18	-	1,28		

** Significativo ao nível de 1% de probabilidade

Capítulo 9

9.1.

Interação AxB significativa: Fcal AxB= 3,21 $Ftab_{5\%}$ (9,36) = 2,16

QMRes combinado = 32,41 N* = 27

Estudo A/B

A/B1	SQ= 1404	QM= 468,06	Fcal= 14,14	$Ftab_{5\%}$ (3,27) = 2,96
A/B2	SQ= 413	QM= 137,66	Fcal= 4,25	
A/B3	SQ= 325	QM= 108,26	Fcal= 3,34	
A/B4	SQ= 1293	QM= 430,86	Fcal= 13,2	
Resíduo	GL= 27	QM= 32,41		

Teste de Tukey A/B Δ = 11

Para A/B1

$\hat{m}_{A4/B1}$ = 61,9 a

$\hat{m}_{A3/B1}$ = 53,9 a b

$\hat{m}_{A2/B1}$ = 50,9 b

$\hat{m}_{A1/B1}$ = 36,1 c

Para A/ B2	Para A/B3	Para A/B4
a	a	a
a b	a b	a
b	a b	a
b	b	b

Estudo: B/A

B/A1	SQ= 583,49	QM= 194,50	Fcal= 9,58
B/A2	SQ= 45,21	QM= 15,07	Fcal= 0,74
B/A3	SQ= 56,96	QM= 18,99	Fcal= 0,94
B/A4	SQ= 71,34	QM= 23,78	Fcal= 1,17

Teste Tukey Δ = 6,06

$\hat{m}_{B2/A1}$ = a

$\hat{m}_{B3/A2}$ = a b

$\hat{m}_{B4/A3}$ = b

$\hat{m}_{B1/A4}$ = b

9.2.

Interação AxB não significativa Fcal AxB= 0,98 $Ftab_{5\%}$(12,30) = 2,09

Fcal A = 4,71 $Ftab_{5\%}$(4,10) = 3,48

Fcal B = 3,99 $Ftab_{5\%}$(3,30) = 2,92

Teste Duncan Fator A (D5= 1,24 D4 = 1,22 D3 = 1,19 D2 = 1,14)

Médias

$\hat{m}_{A5}$ = 17,61 a

$\hat{m}_{A1}$ = 17,56 a

$\hat{m}_{A3}$ = 16,33 b

$\hat{m}_{A4}$ = 16,19 b

$\hat{m}_{A2}$ = 15,98 b

Teste Duncan Fator B (D4= 1,056 D3= 1,029 D2= 0,978)
Médias
$\hat{m}_{B1}$ = 17,43 a
$\hat{m}_{B4}$ = 16,88 a
$\hat{m}_{B3}$ = 16,77 a b
$\hat{m}_{B2}$ = 15,83 b

9.3.
F para interação não significativo Fcal AxB = 2,08 $Ftab_{5\%}(6,24) = 2,51$
Fcal A = 1,17 $Ftab_{5\%}(3,90) = 3,86$
Fcal B = 3,57 $Ftab_{5\%}(2,24) = 3,40$

Fator B Teste tukey
Médias
$\hat{m}_{sulco}$ = 3502,5 a
$\hat{m}_{lanço}$ = 3332,7 ab
$\hat{m}_{cova}$ = 3107,4 b

9.4.

9.4.1. Interação: Fcal=0,97. $Ftab_{5\%}(12,30) = 2,09$. Não rejeita-se Ho. Os fatores atuam independentemente.

9.4.2. Fator B: Fcal = 4,00. $Ftab_{5\%}(3,30) = 2,92$. Rejeita-se Ho. Existe pelo menos um contraste entre médias de niveis do fator B estatisiticamente diferente de zero.

9.4.3.

$D_4 = 1,05$ $D_3 = 1,02$ $D_2 = 0,97$
$\hat{m}_{B1} = 17,46$ a
$\hat{m}_{B4} = 16,88$ a
$\hat{m}_{B3} = 16,77$ a b
$\hat{m}_{B2} = 15,82$ b

9.5. OBSERVAÇÃO: VALORES APROXIMADOS

9.5.1. 0,97
9.5.2. 206,57
9.5.3. nível 5
9.5.4. 3,99
9.5.5. niveis 1, 3 e 4

9.6.

9.6.1. Interação: Fcal = 0,97. $Ftab_{5\%}(12,30) = 2,09$. Não rejeita-se Ho. Os fatores A e B atuam independentemente.

9.6.2. Fator B: Fcal =3,99. $Ftab_{5\%}(3,30) = 2,92$. Rejeita-se Ho. Existe pelo menos um contraste entre médias de niveis de B estatisticamente diferente de zero.

Tukey (Δ=1,31)

$\hat{m}_{B1} = 17{,}46$ a
$\hat{m}_{B4} = 16{,}88$ a b
$\hat{m}_{B3} = 16{,}75$ a b
$\hat{m}_{B2} = 15{,}83$ b

9.7.

9.7.1. Interação: Fcal = 2,74. $Ftab_{5\%}(2,12)$ = 3,89. Não rejeita-se Ho. Os fatores A e B atuam independentemente.

9.7.2. Como a interação foi não significativa, o autor procedeu da forma correta, pois ele comparou os niveis de um fator independente do outro fator.

9.8.

Interação: Fcal = 10,03. $Ftab_{5\%}(2,9)$ = 4,26. Rejeita-se Ho. Os fatores não atuam independentemente.

Estudo: A/B

QMResíduo Combinado: 29,83. GL = 7.

A/B1: Fcal = 21,58. $Ftab_{5\%}(2,7)$ = 4,74. Rejeita-se Ho. Existe pelo menos um constraste estatisticamente diferente de zero, entre médias de niveis de A dentro do nível B1.

A/B2: Fcal = 10,58. $Ftab_{5\%}(2,6)$ = 5,14. Rejeita-se Ho. Existe pelo menos um constraste estatisticamente diferente de zero, entre médias de niveis de A dentro do nível B2.

Médias

- A/B1

$\hat{m}_{A2/B1}$ = 81.25 a
$\hat{m}_{A3/B1}$ = 72.50 a
$\hat{m}_{A1/B1}$ = 56.25 b

- A/B2

$\hat{m}_{A3/B2}$ = 57.50 a
$\hat{m}_{A2/B2}$ = 56.75 a
$\hat{m}_{A1/B2}$ = 41.75 b

B/A

B/A1: Fcal = 66,39. $Ftab_{5\%}(1,9)$ = 5,12. Rejeita-se Ho. Existe pelo menos um constraste estatisticamente diferente de zero, entre médias de niveis de B dentro do nível A1.

B/A2: Fcal = 189,55. $Ftab_{5\%}(1,9)$ = 5,12. Rejeita-se Ho. Existe pelo menos um constraste estatisticamente diferente de zero, entre médias de niveis de B dentro do nível A2.

B/A3: Fcal = 71,05. $Ftab_{5\%}(1,9)$ = 5,12. Rejeita-se Ho. Existe pelo menos um constraste estatisticamente diferente de zero, entre médias de niveis de B dentro do nível A2.

9.9.

Interação: Fcal = 0,42. $Ftab_{5\%}(2,12)$ = 3,89. Não rejeita-se Ho. Os fatores atuam independentemente.

Fator A: Fcal = 22,05. $Ftab_{5\%}(1,6)$ = 5,99. Rejeita-se Ho. Existe pelo menos um contraste estatisticamente diferente de zero entre médias de niveis do fator A.

Fator B: Fcal = 5,83. $Ftab_{5\%}(2,12)$ = 3,39. Rejeita-se Ho. Existe pelo menos um contraste estatisticamente diferente de zero entre médias de niveis do fator B.

Teste de Duncan

Fator A: não é necessário. Teste F já é conclusivo.

Fator B D3 = 1.87 e D2 = 1.78

Médias

$\hat{m}_{B3}$ = 6,29 a

$\hat{m}_{B1}$ = 3,96 b

$\hat{m}_{B2}$ = 3,79 b

Capítulo 10

10.1.

$\hat{\beta}_0$ = 997,4 $\hat{\beta}_1$ = 0,56 x i Fcal= 84,3. A variável independente influencia significativamente a variável dependente.

10.2.

$\hat{\beta}_0$ = 92,7 $\hat{\beta}_1$ = 0,08x i Fcal= 3,55 A variável independente não influencia significativamente a variável dependente.

10.3.

$\hat{\beta}_0$ = 1 52 $\hat{\beta}_1$ = 4,1282 F significativo. A variável independente influencia significativamente a variável dependente.

10.4.

$\hat{\beta}_0$ = 11,24 $\hat{\beta}_1$ = 10,4677 x i $\hat{\beta}_2$ = -1,1135 F cal=231,4. A variável independente influencia significativamente a variável dependente.

10.5.

$\hat{\beta}_0$ = 0,7 $\hat{\beta}_1$ = 0,012 x i F cal= 67,52. A variável independente influencia significativamente a variável dependente.

10.6.

F cal=12,24. A variável independente influencia significativamente a variável dependente.

10.7.

a) 144,4
b)Não. F cal interação foi significativo.
c) Espécie 1 i GL=1 F conclusivo.
d)Fazer uma análise por meio de regressão. Escolhendo o modelo mais adequado.

10.8.
Modelo 3
F falta ajustamento n.s
F regressão significativo F cal= 5,06
R^2=78,2%

10.9.
a) 3,92 kg
b) 23,25°C

10.10.
a)F cal= 225*
b) Sim. F significativo da regressão.

10.11.
$\hat{\beta}_0 = 120,44 \quad F_{cal} = 44,49$

10.12.
$\hat{Y}_i = 16,1 - 0,16x_i - 0.9x_i^2 \quad F_{cal} = 97,63$

10.13.
F =1,097 $^{n.s}$ Ftab = 9,07

10.14.
10.14.1. Falta de Ajustamento: Fcalc = 0,4. $Ftab_{5\%}$(3,10) = 3,71. Não rejeita-se Ho. O modelo linear de 1o grau é apropriado para descrever o tempo de sono em função da dosagem de sonífero.
10.14.2. Regressão: Fcalc = 12. $Ftab_{5\%}$(1,10) = 4,96. Rejeita-se Ho. O coeficiente β_1 é estatisticamente diferente de zero.
10.14.3. Não é recomendável fazer tal estimativa, pois a dose de 17 mg não está dentro do intervalo de dosagem testada.

10.15. OBSERVAÇÃO: **VALORES APROXIMADOS**
10.15.1. –23,76
10.15.2. 1,88 e -0,027
10.15.3. 34,81
10.15.4. 63,13

10.16.
10.16.1. 15,52
10.16.2. 1,74

10.17.

10.17.1. Fcalc = 43,68. $Ftab_{5\%}(2,7) = 4,74$. Rejeita-se Ho. A droga resulta em uma perda de peso significativa.

10.17.2. 13,91

10.17.3. Como 35 mg está fora do intervalo testado, então a equação de regressão ajustada não pode ser usada para estimar a perda de peso para esta dosagem.

10.18.

10.18.1. Falta de Ajustamento: Fcal = 1,93. $Ftab_{5\%}(8,20) = 2,45$. Não rejeita-se Ho. Logo o modelo ajustado é adequado para descrever o fenômeno.

10.18.2. Regressão: Fcalc = 609,47. $Ftab_{5\%}(1,20) = 4,35$. Rejeita-se Ho. A dosagem do suco de laranja tem efeito significativo na acidez da bebida láctea.

10.18.3. –0,08.

10.19.

10.19.1. Fcal = 122,69. $Ftab_{5\%}(2,7) = 4,74$. Rejeita-se Ho. O fermento tem influência significativa no peso final dos pães.

10.19.2. 7,07

10.20.

10.20.1. Fcal = 340,31. $Ftab_{5\%}(1,5) = 6,61$. Rejeita-se Ho. A droga tem influência sobre o nível de açúcar.

10.20.2. Não é possível obter tal estimativa, pois a dose de 90 mg está fora do intervalo testado.

EST 220 – Estatística Experimental

Anexo 1 - Formulário e Tabelas

Observações:

- As tabelas que aqui constam, foram adaptadas do livro: Curso de Estatística Experimental (12ª ed) de Frederico Pimentel Gomes, 1987.
- **Este material será usado em provas e portanto não deverá conter informações adicionais**

Nome:____ Matrícula:______

Formulário

$$\hat{m}_i = \frac{\sum_{i=1}^{n} x_i}{n} \qquad s^2 = \frac{SQ}{GL} \qquad SQ = \sum_{i=1}^{k} \frac{X_i^2}{r_i} - \frac{\left(\sum_{i=1}^{k} X_i\right)^2}{\sum_{i=1}^{k} r_i} \qquad s = \sqrt{s^2} \qquad s(\hat{m}) = \frac{s}{\sqrt{n}} \qquad s_c^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \qquad F = \frac{> s^2}{< s^2}$$

$$t = \frac{\hat{m} - m_0}{\frac{s}{\sqrt{n}}} \qquad t = \frac{(\hat{m}_1 - \hat{m}_2) - (m_1 - m_2)}{\sqrt{s_c^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad t = \frac{\hat{m}_D - m_D}{\sqrt{\frac{s_D^2}{n}}} \qquad \hat{m}_D = \frac{\sum_{i=1}^{n} d_i}{n} \qquad s_D^2 = \frac{\sum_{i=1}^{n} d_i^2 - \frac{\left(\sum_{i=1}^{n} d_i\right)^2}{n}}{n-1}$$

$$\Delta = q\sqrt{\frac{1}{2}\hat{V}(\hat{C})} \qquad D_i = z_i\sqrt{\frac{1}{2}\hat{V}(\hat{C})} \qquad \hat{V}(\hat{C}) = S_C^2 \sum_{i=1}^{I} \frac{a_i^2}{r_i} = QMRes \sum_{i=1}^{I} \frac{a_i^2}{r_i}$$

$$\Delta = q\sqrt{\frac{QMResíduo}{K}} \qquad D_i = z_i\sqrt{\frac{QMResíduo}{K}} \qquad S = \sqrt{(I-1)F_{tab}\hat{V}(\hat{C})}$$

$$t = \frac{\hat{C}}{\sqrt{\hat{V}(\hat{C})}} \qquad SQ = \sum_{i=1}^{k} \frac{X_i^2}{r_i} - \frac{\left(\sum_{i=1}^{k} X_i\right)^2}{\sum_{i=1}^{k} r_i} \qquad CV(\%) = 100\,\frac{\sqrt{QMResíduo}}{\hat{m}} \qquad \hat{m}_i = \frac{T_i}{r_i} \qquad \hat{m} = \frac{G}{N}$$

$$n^* = \frac{[QMRes(a) + (J-1)QMRes(b)]^2}{\frac{[QMRes(a)]^2}{g.l.Res(a)} + \frac{[(J-1)QMRes(b)]^2}{g.l.Res(b)}} \qquad QMResComb = \frac{QMRes(a) + (J-1)QMRes(b)}{J}$$

$$\begin{cases} \sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} Y_i X_i = \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 \end{cases}$$

$$\begin{cases} \sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i + \hat{\beta}_2 \sum_{i=1}^{n} X_i^2 \\ \sum_{i=1}^{n} Y_i X_i = \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 + \hat{\beta}_2 \sum_{i=1}^{n} X_i^3 \\ \sum_{i=1}^{n} Y_i X_i^2 = \hat{\beta}_0 \sum_{i=1}^{n} X_i^2 + \hat{\beta}_1 \sum_{i=1}^{n} X_i^3 + \hat{\beta}_2 \sum_{i=1}^{n} X_i^4 \end{cases}$$

$$SQTotal = \sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$$

$$R^2 = \frac{SQRegressão}{SQTotal}$$

$$R^2 = \frac{SQRegressão}{SQTratamentos}$$

$$SQRegressão = \hat{\beta}_0 \sum_{i=1}^{n} Y_i + \hat{\beta}_1 \sum_{i=1}^{n} Y_i X_i - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$$

$$SQRegressão = \hat{\beta}_0 \sum_{i=1}^{n} Y_i + \hat{\beta}_1 \sum_{i=1}^{n} Y_i X_i + \hat{\beta}_2 \sum_{i=1}^{n} Y_i X_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$$

Tabela 1 - Valores de **t** em níveis de **10% a 0,1%** de probabilidade (**Tabela Bilateral**)

Graus de liberdade	10%	5%	2%	1%	0,5%	0,1%
1	6,31	12,71	31,82	63,66	127,32	636,62
2	2,92	4,30	6,97	9,92	14,09	31,60
3	2,35	3,18	4,54	5,84	7,45	12,94
4	2,13	2,78	3,75	4,60	5,60	8,61
5	2,02	2,57	3,37	4,03	4,77	6,86
6	1,94	2,45	3,14	3,71	4,32	5,96
7	1,90	2,36	3,10	3,50	4,03	5,41
8	1,86	2,31	2,90	3,36	3,83	5,04
9	1,83	2,26	2,82	3,25	3,69	4,78
10	1,81	2,23	2,76	3,17	3,58	4,59
11	1,80	2,20	2,72	3,11	3,50	4,44
12	1,78	2,18	2,68	3,06	3,43	4,32
13	1,77	2,16	2,65	3,01	3,37	4,22
14	1,76	2,14	2,62	2,98	3,33	4,14
15	1,75	2,13	2,60	2,95	3,29	4,07
16	1,75	2,12	2,58	2,92	3,25	4,02
17	1,74	2,11	2,57	2,90	3,22	3,97
18	1,73	2,10	2,55	2,88	3,20	3,92
19	1,73	2,09	2,54	2,86	3,17	3,88
20	1,73	2,09	2,53	2,84	3,15	3,85
21	1,72	2,08	2,52	2,83	3,14	3,82
22	1,72	2,07	2,51	2,82	3,12	3,79
23	1,71	2,07	2,50	2,81	3,10	3,77
24	1,71	2,06	2,49	2,80	3,09	3,75
25	1,71	2,06	2,49	2,79	3,08	3,73
26	1,71	2,06	2,48	2,78	3,07	3,71
27	1,70	2,05	2,47	2,77	3,06	3,69
28	1,70	2,05	2,47	2,76	3,05	3,67
29	1,70	2,04	2,46	2,76	3,04	3,66
30	1,70	2,04	2,46	2,75	3,03	3,65
40	1,68	2,02	2,42	2,70	2,97	3,55
60	1,67	2,00	2,39	2,66	2,92	3,46
120	1,65	1,98	2,36	2,62	2,86	3,37
∞	1,65	1,96	2,33	2,58	2,81	3,29

Tabela 2 - Limites unilaterais de **F** ao nível de **1%** de probabilidade, para o caso de F > 1

	n_1																						
n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	20	24	30	40	60	120	∞
1	4052	5000	5403	5625	5764	5859	5928	5982	6022	6056	6082	6106	6125	6142	6157	6169	6209	6235	6261	6287	6313	6339	6366
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,41	99,42	99,42	99,43	99,43	99,44	99,45	99,46	99,47	99,47	99,48	99,49	99,50
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23	27,13	27,05	26,98	26,92	26,87	26,83	26,69	26,60	26,50	26,41	26,32	26,22	26,13
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,45	14,37	14,30	14,24	14,20	14,15	14,02	13,93	13,84	13,75	13,65	13,56	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,96	9,89	9,83	9,77	9,72	9,68	9,55	9,47	9,38	9,29	9,20	9,11	9,02
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72	7,66	7,60	7,56	7,52	7,40	7,31	7,23	7,14	7,06	6,97	6,88
7	12,25	9,55	8,45	7,85	8,46	7,19	6,99	6,84	6,72	6,62	6,54	6,47	6,41	6,35	6,31	6,27	6,16	6,07	5,99	5,91	5,82	5,74	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,74	5,67	5,61	5,56	5,52	5,48	5,36	5,28	5,20	5,12	5,03	4,95	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18	5,11	5,05	5,00	4,96	4,92	4,81	4,73	4,65	4,57	4,48	4,40	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,78	4,71	4,65	4,60	4,56	4,52	4,41	4,33	4,25	4,17	4,08	4,00	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,46	4,40	4,34	4,29	4,25	4,21	4,10	4,02	3,94	3,86	3,78	3,69	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,22	4,16	4,10	4,05	4,01	3,98	3,86	3,78	3,70	3,62	3,54	3,45	3,36
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96	3,90	3,85	3,82	3,78	3,66	3,59	3,51	3,43	3,34	3,25	3,17
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80	3,75	3,70	3,66	3,62	3,51	3,43	3,35	3,27	3,18	3,09	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67	3,61	3,56	3,52	3,48	3,37	3,29	3,21	3,13	3,05	2,96	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55	3,50	3,45	3,41	3,37	3,26	3,18	3,10	3,02	2,93	2,84	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,46	3,40	3,35	3,31	3,27	3,16	3,08	3,00	2,92	2,83	2,75	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,44	3,37	3,32	3,27	3,23	3,19	3,08	3,00	2,92	2,84	2,75	2,66	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30	3,24	3,19	3,15	3,12	3,00	2,92	2,84	2,76	2,67	2,58	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,30	3,23	3,18	3,13	3,09	3,05	2,94	2,86	2,78	2,69	2,61	2,52	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,24	3,17	3,12	3,07	3,03	2,99	2,88	2,80	2,72	2,64	2,55	2,46	2,36
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	3,12	3,07	3,02	2,98	2,94	2,83	2,75	2,67	2,58	2,50	2,40	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,14	3,07	3,02	2,97	2,93	2,89	2,78	2,70	2,62	2,54	2,45	2,35	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,09	3,03	2,98	2,93	2,89	2,85	2,74	2,66	2,58	2,49	2,40	2,31	2,21
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	3,05	2,99	2,94	2,89	2,85	2,81	2,70	2,62	2,54	2,45	2,36	2,27	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	3,02	2,96	2,91	2,86	2,81	2,77	2,66	2,58	2,50	2,42	2,33	2,23	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,98	2,93	2,88	2,83	2,78	2,74	2,63	2,55	2,47	2,38	2,29	2,20	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,95	2,90	2,85	2,80	2,75	2,71	2,60	2,52	2,44	2,35	2,26	2,17	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,92	2,87	2,82	2,77	2,73	2,68	2,57	2,49	2,41	2,33	2,23	2,14	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,90	2,84	2,79	2,74	2,70	2,66	2,55	2,47	2,39	2,30	2,21	2,11	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,73	2,66	2,61	2,56	2,52	2,49	2,37	2,29	2,20	2,11	2,02	1,92	1,80
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,56	2,50	2,45	2,40	2,35	2,32	2,20	2,12	2,03	1,94	1,84	1,73	1,60
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,40	2,34	2,29	2,24	2,19	2,16	2,03	1,95	1,86	1,76	1,66	1,53	1,38
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,24	2,18	2,12	2,07	2,04	1,99	1,88	1,79	1,70	1,59	1,47	1,32	1,00

n_1 = número de graus de liberdade do numerador
n_2 = número de graus de liberdade do denominador

Tabela 3 - Limites unilaterais de **F** ao nível de **5%** de probabilidade, para o caso de F > 1

	n_1																						
n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	20	24	30	40	60	120	∞
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,0	243,9	244,4	245,0	245,9	246,0	248,0	249,1	250,1	251,1	252,2	253,3	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,40	19,41	19,42	19,42	19,43	19,43	19,45	19,45	19,46	19,47	19,48	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,72	8,71	8,70	8,69	8,66	8,64	8,62	9,59	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91	5,89	5,87	5,86	5,84	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66	4,64	4,62	4,60	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98	3,96	3,94	3,92	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55	3,52	3,51	3,49	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,25	3,23	3,22	3,20	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,04	3,02	3,01	2,98	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,88	2,86	2,85	2,82	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76	2,74	2,72	2,70	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,66	2,64	2,62	2,60	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,57	2,55	2,53	2,51	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,56	2,53	2,50	2,48	2,46	2,44	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,45	2,43	2,40	2,39	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42	2,39	2,37	2,35	2,33	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,35	2,33	2,31	2,29	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,31	2,29	2,27	2,25	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,28	2,26	2,23	2,21	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,25	2,23	2,20	2,18	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25	2,22	2,20	2,18	2,15	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23	2,20	2,18	2,15	2,13	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,24	2,20	2,17	2,14	2,13	2,10	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,22	2,18	2,15	2,13	2,11	2,09	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16	2,13	2,11	2,09	2,06	2,01	1,96	1,92	1,87	1,82	1,77	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,12	2,10	2,07	2,05	1,99	1,95	1,90	1,85	1,80	1,75	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,16	2,13	2,10	2,08	2,06	2,03	1,97	1,93	1,88	1,84	1,79	1,73	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12	2,09	2,06	2,04	2,02	1,96	1,91	1,87	1,82	1,77	1,71	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,14	2,10	2,07	2,05	2,03	2,00	1,94	1,90	1,85	1,81	1,75	1,70	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,12	2,09	2,06	2,04	2,01	1,99	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00	1,97	1,95	1,92	1,90	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,89	1,86	1,84	1,81	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,86	1,83	1,80	1,77	1,75	1,73	1,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,79	1,75	1,72	1,69	1,67	1,64	1,57	1,52	1,46	1,39	1,32	1,22	1,00

n_1 = número de graus de liberdade do numerador
n_2 = número de graus de liberdade do denominador

Tabela 4 - Valores da amplitude total estudentizada (q), para uso no teste de **Tukey**, ao nível de **1%** de probabilidade

	I																	
n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	90,03	135,0	164,3	185,6	202,2	215,8	227,2	237,0	245,6	253,2	260,0	266,2	271,8	277,0	281,8	286,3	290,4	294,3
2	14,04	19,02	22,29	24,72	26,63	28,20	29,53	30,68	31,69	32,59	33,40	34,13	34,81	35,43	36,00	36,53	37,03	37,50
3	8,26	10,62	12,17	13,33	14,24	15,00	15,64	16,20	16,69	17,13	17,53	17,89	18,22	18,52	18,81	19,07	19,32	19,55
4	6,51	8,12	9,17	9,96	10,58	11,10	11,55	11,93	12,27	12,57	12,84	13,09	13,32	13,53	13,73	13,91	14,08	14,24
5	5,70	6,98	7,80	8,42	8,91	9,32	9,70	9,97	10,24	10,48	10,70	10,89	11,08	11,24	11,40	11,55	11,68	11,81
6	5,24	6,33	7,03	7,56	7,97	8,32	8,61	8,87	9,10	9,30	9,48	9,65	9,81	9,95	10,08	10,21	10,32	10,43
7	4,95	5,92	6,54	7,00	7,37	7,68	7,94	8,17	8,37	8,55	8,71	8,86	9,00	9,12	9,24	9,35	9,46	9,55
8	4,75	5,64	6,20	6,62	6,96	7,24	7,47	7,68	7,86	8,03	8,18	8,31	8,44	8,55	8,66	8,76	8,85	8,94
9	4,60	5,43	5,96	6,35	6,66	6,92	7,13	7,32	7,50	7,65	7,78	7,91	8,02	8,13	8,23	8,32	8,41	8,50
10	4,48	5,27	5,77	6,14	6,43	6,67	6,88	7,06	7,21	7,36	7,48	7,60	7,71	7,81	7,91	7,99	8,08	8,15
11	4,39	5,15	5,62	5,97	6,25	6,48	6,67	6,84	6,99	7,13	7,25	7,36	7,46	7,56	7,65	7,73	7,81	7,88
12	4,32	5,05	5,50	5,84	6,10	6,32	6,51	6,67	6,81	6,94	7,06	7,17	7,26	7,36	7,44	7,52	7,59	7,66
13	4,26	4,96	5,40	5,73	5,98	6,19	6,37	6,53	6,67	6,79	6,90	7,01	7,10	7,19	7,27	7,34	7,42	7,48
14	4,21	4,90	5,32	5,63	5,88	6,08	6,26	6,41	6,54	6,66	6,77	6,87	6,96	7,05	7,13	7,20	7,27	7,33
15	4,17	4,84	5,25	5,56	5,80	5,99	6,16	6,31	6,44	6,56	6,66	6,76	6,84	6,93	7,00	7,07	7,14	7,20
16	4,13	4,79	5,19	5,49	5,72	5,92	6,08	6,22	6,35	6,46	6,56	6,66	6,74	6,82	6,90	6,97	7,03	7,09
17	4,10	4,74	5,14	5,43	5,66	5,85	6,01	6,15	6,27	6,38	6,48	6,57	6,66	6,73	6,81	6,87	6,94	7,00
18	4,07	4,70	5,09	5,38	5,60	5,79	5,94	6,08	6,20	6,31	6,41	6,50	6,58	6,66	6,72	6,79	6,85	6,91
19	4,05	4,67	5,05	5,33	5,55	5,74	5,89	6,02	6,14	6,25	6,34	6,43	6,51	6,58	6,65	6,72	6,78	6,84
20	4,02	4,64	5,02	5,29	5,51	5,69	5,84	5,97	6,09	6,19	6,28	6,37	6,45	6,52	6,59	6,65	6,71	6,77
24	3,96	4,55	4,91	5,17	5,37	5,54	5,68	5,81	5,92	6,02	6,11	6,19	6,26	6,33	6,39	6,45	6,51	6,56
30	3,89	4,46	4,80	5,05	5,24	5,40	5,54	5,65	5,76	5,85	5,93	6,01	6,08	6,14	6,20	6,26	6,31	6,36
40	3,82	4,37	4,70	4,93	5,11	5,26	5,39	5,50	5,60	5,69	5,76	5,84	5,90	5,96	6,02	6,07	6,12	6,16
60	3,76	4,28	4,60	4,82	4,99	5,13	5,25	5,36	5,45	5,53	5,60	5,67	5,73	5,78	5,84	5,89	5,93	5,97
120	3,70	4,20	4,50	4,71	4,87	5,00	5,12	5,21	5,30	5,38	5,44	5,50	5,56	5,61	5,66	5,71	5,75	5,79
∞	3,64	4,12	4,40	4,60	4,76	4,88	4,99	5,08	5,16	5,23	5,29	5,35	5,40	5,45	5,49	5,54	5,57	5,61

I = número de níveis do fator em teste
n_2 = número de graus de liberdade do resíduo

Tabela 4 - Valores da amplitude total estudentizada (q), para uso no teste de **Tukey**, ao nível de **1%** de probabilidade (continuação)

									I								
n_2	20	22	24	26	28	30	32	34	36	38	40	50	60	70	80	90	100
1	298,0	304,7	310,8	316,3	321,3	326,0	330,3	334,3	338,0	341,5	344,8	358,9	370,1	379,4	387,3	394,1	400,1
2	37,95	38,76	39,49	40,15	40,76	41,32	41,84	42,33	42,78	43,21	43,61	45,33	46,70	47,83	48,80	49,64	50,38
3	19,77	20,17	20,53	20,86	21,16	21,44	21,70	21,95	22,17	22,39	22,59	23,45	24,13	24,71	25,19	25,62	25,99
4	14,40	14,68	14,93	15,16	15,37	15,57	15,75	15,92	16,08	16,23	16,37	16,98	17,46	17,86	18,20	18,50	18,77
5	11,93	12,16	12,36	12,54	12,71	12,87	13,02	13,15	13,28	13,40	13,52	14,00	14,39	14,72	14,99	15,23	15,45
6	10,54	10,73	10,91	11,06	11,21	11,34	11,47	11,58	11,69	11,80	11,90	12,31	12,65	12,92	13,16	13,37	13,55
7	9,65	9,82	9,97	10,11	10,24	10,36	10,47	10,58	10,67	10,77	10,85	11,23	11,52	11,77	11,99	12,17	12,34
8	9,03	9,18	9,32	9,45	9,57	9,68	9,78	9,87	9,96	10,05	10,13	10,47	10,75	10,97	11,17	11,34	11,49
9	8,58	8,72	8,85	8,97	9,08	9,18	9,27	9,36	9,44	9,52	9,59	9,91	10,17	10,38	10,57	10,73	10,87
10	8,23	8,36	8,48	8,60	8,70	8,79	8,88	8,97	9,04	9,12	9,19	9,49	9,73	9,93	10,10	10,25	10,39
11	7,95	8,08	8,20	8,30	8,40	8,49	8, 58	8,65	8,73	8,80	8,86	9,15	9,38	9,57	9,73	9,88	10,00
12	7,73	7,85	7,96	8,07	8,16	8,25	8,33	8,40	8,47	8,54	8,60	8,88	9,09	9,28	9,43	9,57	9,69
13	7,55	7,66	7,77	7,87	7,96	8,04	8,12	8,19	8,26	8,33	8,39	8,65	8,86	9,04	9,19	9,32	9,44
14	7,40	7,51	7,61	7,70	7,79	7,87	7,95	8,02	8,08	8,15	8,20	8,46	8,66	8,83	8,98	9,11	9,22
15	7,26	7,37	7,47	7,57	7,65	7,73	7,80	7,87	7,93	7,99	8,05	8,30	8,49	8,66	8,80	8,92	9,04
16	7,15	7,26	7,36	7,44	7,53	7,60	7,67	7,74	7,80	7,86	7,92	8,15	8,35	8,51	8,65	8,77	8,87
17	7,05	7,16	7,25	7,34	7,42	7,49	7,56	7,63	7,69	7,74	7,80	8,03	8,22	8,38	8,51	8,63	8,74
18	6,97	7,07	7,16	7,25	7,32	7,40	7,46	7,53	7,59	7,64	7,70	7,92	8,11	8,26	8,39	8,51	8,61
19	6,89	6,99	7,08	7,17	7,24	7,31	7,38	7,44	7,50	7,55	7,60	7,83	8,01	8,16	8,29	8,40	8,50
20	6,82	6,92	7,01	7,09	7,17	7,24	7,30	7,36	7,42	7,47	7,52	7,74	7,92	8,07	8,19	8,30	8,40
24	6,61	6,70	6,79	6,86	6,94	7,00	7,06	7,12	7,17	7,22	7,27	7,48	7,64	7,78	7,90	8,00	8,10
30	6,41	6,49	6,57	6,64	6,71	6,77	6,83	6,88	6,93	6,98	7,02	7,22	7,37	7,50	7,61	7,71	7,80
40	6,21	6,29	6,36	6,43	6,49	6,55	6,60	6,65	6,70	6,74	6,78	6,96	7,10	7,22	7,33	7,42	7,50
60	6,02	6,09	6,16	6,22	6,28	6,33	6,38	6,42	6,47	6,51	6,55	6,71	6,84	6,95	7,05	7,13	7,21
120	5,83	5,90	5,96	6,02	6,07	6,12	6,16	6,20	6,24	6,28	6,32	6,47	6,59	6,69	6,78	6,85	6,92
∞	5,64	5,71	5,77	5,82	5,87	5,91	5,95	5,99	6,03	6,06	6,09	6,23	6,34	6,43	6,51	6,58	6,64

I = número de níveis do fator em teste
n_2 = número de graus de liberdade do resíduo

Tabela 5 - Valores da amplitude total estudentizada (q), para uso no teste de **Tukey**, ao nível de **5%** de probabilidade

	I																		
n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	17,97	26,98	32,82	37,08	40,41	43,12	45,40	47,36	49,07	50,59	51,96	53,20	54,33	55,36	56,32	57,22	58,04	58,83	59,56
2	6,09	8,33	9,80	10,88	11,74	12,44	13,03	13,54	13,99	14,39	14,75	15,08	15,38	16,65	15,91	16,14	16,37	16,57	16,77
3	4,50	5,91	6,83	7,50	8,04	8,48	8,85	9,18	9,46	9,72	9,95	10,15	10,35	10,53	10,69	10,84	10,98	11,11	11,24
4	3,93	5,04	5,76	6,29	6,71	7,05	7,35	7,60	7,83	8,03	8,21	8,37	8,53	8,66	8,79	8,91	9,03	9,13	9,23
5	3,64	4,60	5,22	5,67	6,03	6,33	6,58	6,80	7,00	7,17	7,32	7,47	7,60	7,72	7,83	7,93	8,03	8,12	8,21
6	3,46	4,34	4,90	5,31	5,63	5,90	6,12	6,32	6,49	6,65	6,79	6,92	7,03	7,14	7,24	7,34	7,43	7,51	7,59
7	3,34	4,17	4,68	5,06	5,36	5,61	5,82	6,00	6,16	6,30	6,43	6,55	6,66	6,76	6,85	6,94	7,02	7,10	7,17
8	3,26	4,04	4,53	4,89	5,17	5,40	5,60	5,77	5,92	6,05	6,18	6,29	6,39	6,48	6,57	6,65	6,73	6,80	6,87
9	3,20	3,95	4,42	4,76	5,02	5,24	5,43	5,60	5,74	5,87	5,98	6,09	6,19	6,28	6,36	6,44	6,51	6,58	6,64
10	3,15	3,88	4,33	4,65	4,91	5,12	5,31	5,46	5,60	5,72	5,83	5,94	6,03	6,11	6,19	6,27	6,34	6,41	6,47
11	3,11	3,82	4,26	4,57	4,82	5,03	5,20	5,35	5,49	5,61	5,71	5,81	5,90	5,98	6,06	6,13	6,20	6,27	6,33
12	3,08	3,77	4,20	4,51	4,75	4,95	5,12	5,27	5,40	5,51	5,62	5,71	5,80	5,88	5,95	6,02	6,09	6,15	6,21
13	3,06	3,74	4,15	4,45	4,69	4,89	5,05	5,19	5,32	5,43	5,53	5,63	5,71	5,79	5,86	5,93	6,00	6,06	6,11
14	3,03	3,70	4,11	4,41	4,64	4,83	4,99	5,13	5,25	5,36	5,46	5,55	5,64	5,71	5,79	5,85	5,92	5,97	6,03
15	3,01	3,67	4,08	4,37	4,60	4,78	4,94	5,08	5,20	5,31	5,40	5,49	5,57	5,65	5,72	5,79	5,85	5,90	5,96
16	3,00	3,65	4,05	4,33	4,56	4,74	4,90	5,03	5,15	5,26	5,35	5,44	5,52	5,59	5,66	5,73	5,79	5,84	5,90
17	2,98	3,63	4,02	4,30	4,52	4,71	4,86	4,99	5,11	5,21	5,31	5,39	5,47	5,54	5,61	5,68	5,73	5,79	5,84
18	2,97	3,61	4,00	4,28	4,50	4,67	4,82	4,96	5,07	5,17	5,27	5,35	5,43	5,50	5,57	5,63	5,69	5,74	5,79
19	2,96	3,59	3,98	4,25	4,47	4,65	4,79	4,92	5,04	5,14	5,23	5,32	5,39	5,46	5,53	5,59	5,65	5,70	5,75
20	2,95	3,58	3,96	4,23	4,45	4,62	4,77	4,90	5,01	5,11	5,20	5,28	5,36	5,43	5,49	5,55	5,61	5,66	5,71
24	2,92	3,53	3,90	4,17	4,37	4,54	4,68	4,81	4,92	5,01	5,10	5,18	5,25	5,32	5,38	5,44	5,49	5,55	5,59
30	2,89	3,49	3,85	4,10	4,30	4,46	4,60	4,72	4,82	4,92	5,00	5,08	5,15	5,21	5,27	5,33	5,38	5,43	5,48
40	2,86	3,44	3,79	4,04	4,23	4,39	4,52	4,64	4,74	4,82	4,90	4,98	5,04	5,11	5,16	5,22	5,27	5,31	5,36
60	2,83	3,40	3,74	3,98	4,16	4,31	4,44	4,55	4,65	4,73	4,81	4,88	4,94	5,00	5,06	5,11	5,15	5,20	5,24
120	2,80	3,36	3,69	3,92	4,10	4,24	4,36	4,47	4,56	4,64	4,71	4,78	4,84	4,90	4,95	5,00	5,04	5,09	5,13
$\propto$	2,77	3,31	3,63	3,86	4,03	4,17	4,29	4,39	4,47	4,55	4,62	4,69	4,74	4,80	4,85	4,89	4,93	4,97	5,01

I = número de níveis do fator em teste
n_2 = número de graus de liberdade do resíduo

Tabela 5 - Valores da amplitude total estudentizada (q), para uso no teste de **Tukey**, ao nível de **5%** de probabilidade (continuação).

	I															
N_2	22	24	26	28	30	32	34	36	38	40	50	60	70	80	90	100
1	60,91	62,12	63,22	64,23	65,15	66,01	66,81	67,56	68,26	68,92	71,73	73,97	75,82	77,40	78,77	79,98
2	17,13	17,45	17,75	18,02	18,27	18,50	18,72	18,92	19,11	19,28	20,05	20,66	21,16	21,59	21,96	22,29
3	11,47	11,68	11,87	12,05	12,21	12,36	12,50	12,63	12,75	12,87	13,36	13,76	14,08	14,36	14,61	14,82
4	9,42	9,58	9,74	9,88	10,00	10,12	10,23	10,34	10,44	10,53	10,93	11,24	11,51	11,73	11,92	12,09
5	8,37	8,51	8,64	8,76	8,88	8,98	9,08	9,17	9,25	9,33	9,67	9,95	10,18	10,38	10,54	10,69
6	7,73	7,86	7,98	8,09	8,19	8,28	8,37	8,45	8,53	8,60	8,91	9,16	9,37	9,55	9,70	9,84
7	7,30	7,42	7,53	7,63	7,73	7,81	7,90	7,97	8,04	8,11	8,40	8,63	8,82	8,99	9,13	9,26
8	7,00	7,11	7,21	7,31	7,40	7,48	7,55	7,63	7,69	7,76	8,03	8,25	8,43	8,59	8,72	8,84
9	6,76	6,87	6,97	7,06	7,15	7,22	7,30	7,36	7,43	7,49	7,75	7,96	8,13	8,28	8,41	8,53
10	6,58	6,69	6,78	6,87	6,95	7,02	7,09	7,16	7,22	7,28	7,53	7,73	7,90	8,04	8,17	8,28
11	6,44	6,54	6,63	6,71	6,79	6,86	6,93	6,99	7,05	7,11	7,35	7,55	7,71	7,85	7,97	8,08
12	6,32	6,41	6,50	6,59	6,66	6,73	6,80	6,86	6,92	6,97	7,21	7,39	7,55	7,69	7,80	7,91
13	6,22	6,31	6,40	6,48	6,55	6,62	6,68	6,74	6,80	6,85	7,08	7,27	7,42	7,55	7,67	7,77
14	6,13	6,22	6,31	6,39	6,46	6,53	6,59	6,65	6,70	6,75	6,98	7,16	7,31	7,44	7,55	7,65
15	6,06	6,15	6,23	6,31	6,38	6,45	6,51	6,56	6,62	6,67	6,89	7,07	7,21	7,34	7,45	7,55
16	6,00	6,08	6,17	6,24	6,31	6,37	6,43	6,49	6,54	6,59	6,81	6,98	7,13	7,25	7,36	7,46
17	5,94	6,03	6,11	6,18	6,25	6,31	6,37	6,43	6,48	6,53	6,74	6,91	7,05	7,18	7,28	7,38
18	5,89	5,98	6,06	6,13	6,20	6,26	6,32	6,37	6,42	6,47	6,68	6,85	6,99	7,11	7,21	7,31
19	5,85	5,93	6,01	6,08	6,15	6,21	6,27	6,32	6,37	6,42	6,63	6,79	6,93	7,05	7,15	7,24
20	5,81	5,89	5,97	6,04	6,10	6,17	6,22	6,28	6,33	6,37	6,58	6,74	6,88	6,99	7,10	7,19
24	5,68	5,76	5,84	5,91	5,97	6,03	6,09	6,13	6,18	6,23	6,42	6,58	6,71	6,82	6,92	7,01
30	5,56	5,64	5,71	5,77	5,83	5,89	5,94	5,99	6,04	6,08	6,27	6,42	6,54	6,65	6,74	6,83
40	5,44	5,51	5,58	5,64	5,70	5,75	5,80	5,85	5,89	5,93	6,11	6,26	6,38	6,48	6,57	6,65
60	5,32	5,39	5,45	5,51	5,57	5,62	5,66	5,71	5,75	5,79	5,96	6,09	6,21	6,30	6,39	6,46
120	5,20	5,27	5,33	5,38	5,43	5,48	5,53	5,57	5,61	5,64	5,80	5,93	6,04	6,13	6,21	6,28
∝	5,08	5,14	5,20	5,25	5,30	5,35	5,39	5,43	5,46	5,50	5,65	5,76	5,86	5,95	6,02	6,09

I = número de níveis do fator em teste

n_2 = número de graus de liberdade do resíduo

Tabela 6 - Valores da amplitude total estudentizada (z), para uso no teste de **Duncan**, ao nível de **1%** de probabilidade

	n															
n_2	2	3	4	5	6	7	8	9	10	12	14	16	18	20	50	100
1	90,00	90,00	90,00	90,00	90,00	90,00	90,00	90,00	90,00	90,00	90,00	90,00	90,00	90,00	90,00	90,00
2	14,00	14,00	14,00	14,00	14,00	14,00	14,00	14,00	14,00	14,00	14,00	14,00	14,00	14,00	14,00	14,00
3	8,26	8,50	8,60	8,70	8,80	8,90	8,90	9,00	9,00	9,00	9,10	9,20	9,30	9,30	9,30	9,30
4	6,51	6,80	6,90	7,00	7,10	7,10	7,20	7,20	7,30	7,30	7,40	7,40	7,50	7,50	7,50	7,50
5	5,70	5,96	6,11	6,18	6,26	6,33	6,40	6,44	6,50	6,60	6,60	6,70	6,70	6,80	6,80	6,80
6	5,24	5,51	5,65	5,73	5,81	5,88	5,95	6,00	6,00	6,10	6,20	6,20	6,30	6,30	6,30	6,30
7	4,95	5,22	5,37	5,45	5,53	5,61	5,69	5,73	5,80	5,80	5,90	5,90	6,00	6,00	6,00	6,00
8	4,74	5,00	5,14	5,23	5,32	5,40	5,47	5,51	5,50	5,60	5,70	5,70	5,80	5,80	5,80	5,80
9	4,60	4,86	4,99	5,08	5,17	5,25	5,32	5,36	5,40	5,50	5,50	5,60	5,70	5,70	5,70	5,70
10	4,48	4,73	4,88	4,96	5,06	5,13	5,20	5,24	5,28	5,36	5,42	5,48	5,54	5,55	5,55	5,55
11	4,39	4,63	4,77	4,86	4,94	5,01	5,06	5,12	5,15	5,24	5,28	5,34	5,38	5,39	5,39	5,39
12	4,32	4,55	4,68	4,76	4,84	4,92	4,96	5,02	5,07	5,13	5,17	5,22	5,24	5,26	5,26	5,26
13	4,26	4,48	4,62	4,69	4,74	4,84	4,88	4,94	4,98	5,04	5,08	5,13	5,14	5,15	5,15	5,15
14	4,21	4,42	4,55	4,63	4,70	4,78	4,83	4,87	4,91	4,96	5,00	5,04	5,06	5,07	5,07	5,07
15	4,17	4,37	4,50	4,58	4,64	4,72	4,77	4,81	4,84	4,90	4,94	4,97	4,99	5,00	5,00	5,00
16	4,13	4,34	4,45	4,54	4,60	4,67	4,72	4,76	4,79	4,84	4,88	4,91	4,93	4,94	4,94	4,94
17	4,10	4,30	4,41	4,50	4,56	4,63	4,68	4,72	4,75	4,80	4,83	4,86	4,88	4,89	4,89	4,89
18	4,07	4,27	4,38	4,46	4,53	4,59	4,64	4,68	4,71	4,76	4,79	4,82	4,84	4,85	4,85	4,85
19	4,05	4,24	4,35	4,43	4,50	4,56	4,61	4,64	4,67	4,72	4,76	4,79	4,81	4,82	4,82	4,82
20	4,02	4,22	4,33	4,40	4,47	4,53	4,58	4,61	4,65	4,69	4,73	4,76	4,78	4,79	4,79	4,79
22	3,99	4,17	4,28	4,36	4,42	4,48	4,53	4,57	4,60	4,65	4,68	4,71	4,74	4,75	4,75	4,75
24	3,96	4,14	4,24	4,33	4,39	4,44	4,49	4,53	4,57	4,62	4,64	4,67	4,70	4,72	4,74	4,74
26	3,93	4,11	4,21	4,30	4,36	4,41	4,46	4,50	4,53	4,58	4,62	4,65	4,67	4,69	4,73	4,73
28	3,91	4,08	4,18	4,28	4,34	4,39	4,43	4,47	4,51	4,56	4,60	4,62	4,65	4,67	4,72	4,72
30	3,89	4,06	4,16	4,22	4,32	4,36	4,41	4,45	4,48	4,54	4,58	4,61	4,63	4,65	4,71	4,71
40	3,82	3,99	4,10	4,17	4,24	4,30	4,34	4,37	4,41	4,46	4,51	4,54	4,57	4,59	4,69	4,69
60	3,76	3,92	4,03	4,12	4,17	4,23	4,27	4,31	4,34	4,39	4,44	4,47	4,50	4,53	4,66	4,66
100	3,71	3,86	3,98	4,06	4,11	4,17	4,21	4,25	4,29	4,35	4,38	4,42	4,45	4,48	4,64	4,65
∝	3,64	3,80	3,90	3,98	4,04	4,09	4,14	4,17	4,20	4,26	4,31	4,34	4,38	4,41	4,60	4,68

n = nº de médias ordenadas abrangidas pelo contraste

n_2 = nº de graus de liberdade do resíduo

Tabela 7 - Valores da amplitude total estudentizada (z), para uso no teste de **Duncan**, ao nível de **5%** de probabilidade

	n															
n_2	2	3	4	5	6	7	8	9	10	12	14	16	18	20	50	100
1	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0
2	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09
3	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50
4	3,93	4,01	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02
5	3,64	3,74	3,79	3,83	3,83	3,83	3,83	3,83	3,83	3,83	3,83	3,83	3,83	3,83	3,83	3,83
6	3,46	3,58	3,64	3,68	3,68	3,68	3,68	3,68	3,68	3,68	3,68	3,68	3,68	3,68	3,68	3,68
7	3,35	3,47	3,54	3,58	3,60	3,61	3,61	3,61	3,61	3,61	3,61	3,61	3,61	3,61	3,61	3,61
8	3,26	3,39	3,47	3,52	3,55	3,56	3,56	3,56	3,56	3,56	3,56	3,56	3,56	3,56	3,56	3,56
9	3,20	3,34	3,41	3,47	3,50	3,52	3,52	3,52	3,52	3,52	3,52	3,52	3,52	3,52	3,52	3,52
10	3,15	3,30	3,37	3,43	3,46	3,47	3,47	3,47	3,47	3,47	3,47	3,47	3,47	3,48	3,48	3,48
11	3,11	3,27	3,35	3,39	3,43	3,44	3,45	3,46	3,46	3,46	3,46	3,46	3,47	3,48	3,48	3,48
12	3,08	3,23	3,33	3,36	3,40	3,42	3,44	3,44	3,46	3,46	3,46	3,46	3,47	3,48	3,48	3,48
13	3,06	3,21	3,30	3,35	3,38	3,41	3,42	3,44	3,45	3,45	3,46	3,46	3,47	3,47	3,47	3,47
14	3,03	3,18	3,27	3,33	3,37	3,39	3,41	3,42	3,44	3,45	3,46	3,46	3,47	3,47	3,47	3,47
15	3,01	3,16	3,25	3,31	3,36	3,38	3,40	3,42	3,43	3,44	3,45	3,46	3,47	3,47	3,47	3,47
16	3,00	3,15	3,23	3,30	3,34	3,37	3,39	3,41	3,43	3,44	3,45	3,46	3,47	3,47	3,47	3,47
17	2,98	3,13	3,22	3,28	3,33	3,36	3,38	3,40	3,42	3,44	3,45	3,46	3,47	3,47	3,47	3,47
18	2,97	3,12	3,21	3,27	3,32	3,35	3,37	3,39	3,41	3,43	3,45	3,46	3,47	3,47	3,47	3,47
19	2,96	3,11	3,19	3,26	3,31	3,35	3,37	3,39	3,41	3,43	3,44	3,46	3,47	3,47	3,47	3,47
20	2,95	3,10	3,18	3,25	3,30	3,34	3,36	3,38	3,40	3,43	3,44	3,46	3,46	3,47	3,47	3,47
22	2,93	3,08	3,17	3,24	3,29	3,32	3,35	3,37	3,39	3,42	3,44	3,45	3,46	3,47	3,47	3,47
24	2,92	3,07	3,15	3,22	3,28	3,31	3,34	3,37	3,38	3,41	3,44	3,45	3,46	3,47	3,47	3,47
26	2,91	3,06	3,14	3,21	3,27	3,30	3,34	3,36	3,38	3,41	3,43	3,45	3,46	3,47	3,47	3,47
28	2,90	3,04	3,13	3,20	3,26	3,30	3,33	3,35	3,37	3,40	3,43	3,45	3,46	3,47	3,47	3,47
30	2,89	3,04	3,12	3,20	3,25	3,29	3,32	3,35	3,37	3,40	3,43	3,44	3,46	3,47	3,47	3,47
40	2,86	3,01	3,10	3,17	3,22	3,27	3,30	3,33	3,35	3,39	3,42	3,44	3,46	3,47	3,47	3,47
60	2,83	2,98	3,08	3,14	3,20	3,24	3,28	3,31	3,33	3,37	3,40	3,43	3,45	3,47	3,48	3,48
100	2,80	2,95	3,05	3,12	3,18	3,22	3,26	3,29	3,32	3,36	3,40	3,42	3,45	3,47	3,53	3,53
∞	2,77	2,92	3,02	3,09	3,15	3,19	3,23	3,26	3,29	3,34	3,38	3,41	3,44	3,47	3,61	3,67

n = nº de médias ordenadas abrangidas pelo contraste

n_2 = nº de graus de liberdade do resíduo

Tabela 8 - Valores críticos (d_c) para o teste de Lilliefors (adaptado de Barbetta et al.2004)

n	α=5%	α=1%
4	0,381	0,734
5	0,337	0,405
6	0,319	0,364
7	0,300	0,348
8	0,285	0,331
9	0,271	0,311
10	0,258	0,294
11	0,249	0,284
12	0,242	0,275
13	0,234	0,268
14	0,227	0,261
15	0,220	0,257
16	0,213	0,250
17	0,206	0,245
18	0,200	0,239
19	0,179	0,235
20	0,190	0,231
25	0,173	0,200
30	0,161	0,187
N>30	$d_c = \frac{0,886}{\sqrt{N}}$	$d_c = \frac{1,031}{\sqrt{N}}$

Tabela 9 – Valores da função de distribuição acumulada da normal padrão, Z, tal que F(z) = P(0 ≤ Z ≤ z)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586
0.1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535
0.2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409
0.3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173
0.4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793
0.5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566	0.21904	0.22240
0.6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.25490
0.7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935	0.28230	0.28524
0.8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327
0.9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891
1.0	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214
1.1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900	0.38100	0.38298
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147
1.3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41308	0.41466	0.41621	0.41774
1.4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189
1.5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408
1.6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254	0.45352	0.45449
1.7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.46080	0.46164	0.46246	0.46327
1.8	0.46407	0.46485	0.46562	0.46638	0.46712	0.46784	0.46856	0.46926	0.46995	0.47062
1.9	0.47128	0.47193	0.47257	0.47320	0.47381	0.47441	0.47500	0.47558	0.47615	0.47670
2.0	0.47725	0.47778	0.47831	0.47882	0.47932	0.47982	0.48030	0.48077	0.48124	0.48169
2.1	0.48214	0.48257	0.48300	0.48341	0.48382	0.48422	0.48461	0.48500	0.48537	0.48574
2.2	0.48610	0.48645	0.48679	0.48713	0.48745	0.48778	0.48809	0.48840	0.48870	0.48899
2.3	0.48928	0.48956	0.48983	0.49010	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158
2.4	0.49180	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361
2.5	0.49379	0.49396	0.49413	0.49430	0.49446	0.49461	0.49477	0.49492	0.49506	0.49520
2.6	0.49534	0.49547	0.49560	0.49573	0.49585	0.49598	0.49609	0.49621	0.49632	0.49643
2.7	0.49653	0.49664	0.49674	0.49683	0.49693	0.49702	0.49711	0.49720	0.49728	0.49736
2.8	0.49744	0.49752	0.49760	0.49767	0.49774	0.49781	0.49788	0.49795	0.49801	0.49807
2.9	0.49813	0.49819	0.49825	0.49831	0.49836	0.49841	0.49846	0.49851	0.49856	0.49861
3.0	0.49865	0.49869	0.49874	0.49878	0.49882	0.49886	0.49889	0.49893	0.49896	0.49900
3.1	0.49903	0.49906	0.49910	0.49913	0.49916	0.49918	0.49921	0.49924	0.49926	0.49929
3.2	0.49931	0.49934	0.49936	0.49938	0.49940	0.49942	0.49944	0.49946	0.49948	0.49950
3.3	0.49952	0.49953	0.49955	0.49957	0.49958	0.49960	0.49961	0.49962	0.49964	0.49965
3.4	0.49966	0.49968	0.49969	0.49970	0.49971	0.49972	0.49973	0.49974	0.49975	0.49976
3.5	0.49977	0.49978	0.49978	0.49979	0.49980	0.49981	0.49981	0.49982	0.49983	0.49983
3.6	0.49984	0.49985	0.49985	0.49986	0.49986	0.49987	0.49987	0.49988	0.49988	0.49989
3.7	0.49989	0.49990	0.49990	0.49990	0.49991	0.49991	0.49992	0.49992	0.49992	0.49992
3.8	0.49993	0.49993	0.49993	0.49994	0.49994	0.49994	0.49994	0.49995	0.49995	0.49995
3.9	0.49995	0.49995	0.49996	0.49996	0.49996	0.49996	0.49996	0.49996	0.49997	0.49997
4.0	0.49997	0.49997	0.49997	0.49997	0.49997	0.49997	0.49998	0.49998	0.49998	0.49998

Tabela 10 – Valores críticos para teste de Cochran para homogeneidade de Variâncias

α = 1%														
I	K – 1													
	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	0.9999	0.9950	0.9794	0.9586	0.9373	0.9172	0.8988	0.8823	0.8674	0.8539	0.7949	0.7067	0.6062	0.5000
3	0.9933	0.9423	0.8831	0.8335	0.7933	0.7606	0.7335	0.7107	0.6912	0.6743	0.6059	0.5153	0.4230	0.3333
4	0.9676	0.8643	0.7814	0.7212	0.6761	0.6410	0.6129	0.5897	0.5702	0.5536	0.4884	0.4057	0.3251	0.2500
5	0.9279	0.7885	0.6957	0.6329	0.5875	0.5531	0.5259	0.5037	0.4854	0.4697	0.4094	0.3351	0.2644	0.2000
6	0.8828	0.7218	0.6258	0.5635	0.5195	0.4866	0.4608	0.4401	0.4229	0.4084	0.3529	0.2858	0.2229	0.1667
7	0.8376	0.6644	0.5685	0.5080	0.4659	0.4347	0.4105	0.3911	0.3751	0.3616	0.3105	0.2494	0.1929	0.1429
8	0.7945	0.6152	0.5209	0.4627	0.4226	0.3932	0.3704	0.3522	0.3373	0.3248	0.2779	0.2214	0.1700	0.1250
9	0.7544	0.5727	0.4810	0.4251	0.3870	0.3592	0.3378	0.3207	0.3067	0.2950	0.2514	0.1992	0.1521	0.1111
10	0.7175	0.5358	0.4469	0.3934	0.3572	0.3308	0.3106	0.2945	0.2813	0.2704	0.2297	0.1811	0.1376	0.1000
12	0.6528	0.4751	0.3919	0.3428	0.3099	0.2861	0.2680	0.2535	0.2419	0.2320	0.1961	0.1535	0.1157	0.0833
15	0.5747	0.4069	0.3317	0.2882	0.2593	0.2386	0.2228	0.2104	0.2002	0.1918	0.1612	0.1251	0.0934	0.0667
20	0.4799	0.3297	0.2654	0.2288	0.2048	0.1877	0.1748	0.1646	0.1567	0.1501	0.1248	0.0960	0.0709	0.0500
24	0.4247	0.2871	0.2295	0.1970	0.1759	0.1608	0.1495	0.1406	0.1338	0.1283	0.1060	0.0810	0.0595	0.0417
30	0.3632	0.2412	0.1913	0.1635	0.1454	0.1327	0.1232	0.1157	0.1100	0.1054	0.0867	0.0658	0.0480	0.0333
40	0.2940	0.1915	0.1508	0.1281	0.1135	0.1033	0.0957	0.0898	0.0853	0.0816	0.0668	0.0503	0.0363	0.0250
60	0.2151	0.1371	0.1069	0.0902	0.0796	0.0722	0.0668	0.0625	0.0594	0.0567	0.0461	0.0344	0.0245	0.0167
120	0.1225	0.0759	0.0585	0.0489	0.0429	0.0387	0.0357	0.0334	0.0316	0.0302	0.0242	00178	0.0125	0.0083
∞	0	0	0	0	0	0	0	0	0	0	0	0	0	0

α = 5%														
I	K – 1													
	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	0.9985	0.9750	0.9392	0.9057	0.8772	0.8534	0.8332	0.8159	0.8010	0.7880	0.7341	0.6602	0.5813	0.5000
3	0.9669	0.8709	0.7977	0.7457	0.7071	0.6771	0.6530	0.6333	0.6167	0.6025	0.5466	0.4748	0.4031	0.3333
4	0.9065	0.7679	0.6841	0.6287	0.5895	0.5598	0.5365	0.5175	0.5017	0.4884	0.4366	0.3720	0.3093	0.2500
5	0.8412	0.6838	0.5981	0.5441	0.5065	0.4783	0.4564	0.4387	0.4241	0.4118	0.3645	0.3066	0.2513	0.2000
6	0.7808	0.6161	0.5321	0.4803	0.4447	0.4184	0.3980	0.3817	0.3682	0.3568	0.3135	0.2612	0.2119	0.1667
7	0.7271	0.5612	0.4800	0.4307	0.3974	0.3726	0.3535	0.3384	0.3259	0.3154	0.2756	0.2278	0.1833	0.1429
8	0.6798	0.5157	0.4377	0.3910	0.3595	0.3362	0.3185	0.3043	0.2926	0.2829	0.2462	0.2022	0.1616	0.1250
9	0.6385	0.4775	0.4027	0.3584	0.3286	0.3067	0.2901	0.2768	0.2659	0.2568	0.2226	0.1820	0.1446	0.1111
10	0.6020	0.4450	0.3733	0.3311	0.3029	0.2823	0.2666	0.2541	0.2439	0.2353	0.2032	0.1655	0.1308	0.1000
12	0.5410	0.3924	0.3264	0.2880	0.2624	0.2439	0.2299	0.2187	0.2098	0.2020	0.1737	0.1403	0.1100	0.0833
15	0.4709	0.3346	0.2758	0.2419	0.2195	0.2034	0.1911	0.1815	0.1736	0.1671	0.1429	0.1144	0.0889	0.0667
20	0.3894	0.2705	0.2205	0.1921	0.1735	0.1602	0.1501	0.1422	0.1357	0.1303	0.1108	0.0879	0.0675	0.0500
24	0.3434	0.2354	0.1907	0.1656	0.1493	0.1374	0.1286	0.1216	0.1160	0.1113	0.0942	0.0743	0.0567	0.0417
30	0.2929	0.1980	0.1593	0.1377	0.1237	0.1137	0.1061	0.1002	0.0958	0.0921	0.0771	0.0604	0.0457	0.0333
40	0.2370	0.1576	0.1259	0.1082	0.0968	0.0887	0.0827	0.0780	0.0745	0.0713	0.0595	0.0462	0.0347	0.0250
60	0.1737	0.1131	0.0895	0.0765	0.0682	0.0623	0.0583	0.0552	0.0520	0.0497	0.0411	0.0316	0.0234	0.0167
120	0.0998	0.0632	0.0495	0.0419	0.0371	0.0337	0.0312	0.0292	0.0279	0.0266	0.0218	0.0165	0.0120	0.0083
∞	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Em que: I= nº de tratamentos e K: nº de repetições

Tabela 11 – Valores críticos (d_c) para o teste de Kolmogorov-Smirnov (adaptado de Barbetta et al.2004)

N	α=5%	α=1%
1	0,975	0,995
2	0,842	0,929
3	0,708	0,829
4	0,624	0,734
5	0,563	0,669
6	0,519	0,617
7	0,483	0,576
8	0,454	0,542
9	0,430	0,513
10	0,430	0,490
11	0,409	0,468
12	0,391	0,449
13	0,375	0,432
14	0,361	0,418
15	0,349	0,404
16	0,338	0,392
17	0,327	0,81
18	0,318	0,371
19	0,309	0,361
20	0,301	0,352
25	0,294	0,317
30	0,264	0,290
35	0,242	0,269
40	0,224	0,252
45	0,210	0,238
50	0,198	0,227
N>50	$d_c = \frac{1{,}36}{\sqrt{N}}$	$d_c = \frac{1{,}36}{\sqrt{N}}$

Em que N= nº de unidades experimentais (N=I*K; I: nº de tratamentos e K:nº de repetições)

Anexo 2 - Fórmula geral para o cálculo de soma de quadrados

Suponha que se deseje estimar a variação entre os níveis de uma determinada Fonte de Variação, digamos T. Para estimar esta variação, calcula-se o Quadrado Médio de T (QMT) de acordo com a fórmula geral dada por:

$$QMT = \frac{SQ}{gl}$$

Em termos gerais, a fórmula de Soma de Quadrados para T (SQT) pode ser escrita como:

$$SQT = \sum_{i=1}^{k} \frac{X_i^2}{r_i} - \frac{\left(\sum_{i=1}^{k} X_i\right)^2}{\sum_{i=1}^{k} r_i}$$

em que:
X_i = total observado para cada nível de T. Este total se refere à soma de todas as observações contidas no i-ésimo nível de T;

r_i = número de observações que foram somadas para se obter o total X_i;
k = número de níveis de T.

Em termos gerais, pode-se dizer que na fórmula de SQT, cada valor elevado ao quadrado deve ser dividido pelo número de observações que originou aquele valor.

Pode-se também visualizar dois termos na fórmula de SQT. O termo com sinal positivo, considera os totais individuais de cada i-ésimo nível de T. Já o termo com sinal negativo, considera a soma conjunta dos totais de todos os i-ésimos níveis de T. Neste segundo termo, o numerador se refere à soma de todos os totais (X_i) incluídos no primeiro termo e, o denominador, se refere à soma de todas as observações (r_i) incluídas no primeiro termo.

O denominador da fórmula de QMT, número de graus de liberdade, se refere à diferença entre o número de termos que estão sendo somados e subtraídos no cálculo da SQT. De acordo com a definição da fórmula geral, o número de graus de liberdade associado à T é igual a k – 1, pois no cálculo da SQT existem k valores sendo somados, ou seja, $\sum_{i=1}^{k} \frac{X_i^2}{r_i}$, e existe um valor sendo subtraído, ou seja, $\frac{\left(\sum_{i=1}^{k} X_i\right)^2}{\sum_{i=1}^{k} r_i}$.

Para o uso da fórmula geral, para cada fonte de variação, é necessário apenas identificar o que representa X_i e o que representa r_i. Como relatado anteriormente, X_i é o total observado para cada i-ésimo nível da FV e r_i é o número de observações que originou o respectivo total X_i. Veja na tabela a seguir, o que representam X_i e r_i para algumas possíveis fontes de variação de uma análise de variância.

Fonte de Variação	X_i	r_i
Total	cada observação	igual a um
Tratamentos	total do i-ésimo tratamento	número de observações associado ao total do i-ésimo tratamento
Blocos	total do i-ésimo bloco	número de observações associado ao total do i-ésimo bloco
Linhas	total da i-ésima linha	número de observações associado ao total da i-ésima linha
Colunas	total da i-ésima coluna	número de observações associado ao total da i-ésima coluna
Fator A	total do i-ésimo nível de A	número de observações associado ao total do i-ésimo nível de A
Fator A / B_J	total do i-ésimo nível de A dentro do j-ésimo nível de B	número de observações associado ao total do i-ésimo nível de A dentro do j-ésimo nível de B
Parcelas	total da i-ésima parcela	número de observações associado ao total da i-ésima parcela

OBSERVAÇÕES:

- Esta fórmula geral não se aplica para o cálculo das fontes de variação: resíduo, interação entre fatores e regressão;
- **Esta será a única fórmula fornecida em prova para o cálculo de soma de quadrados;**
- Cabe ao aluno praticar a aplicação desta fórmula nos exercícios existentes.

Anexo 3 - Introdução ao uso do programa SAS

1 Instalação do programa SAS em computadores conectados na rede da UFV

O programa SAS encontra-se disponível na rede da UFV. Para instalar este programa em seu computador, siga as seguintes instruções:

1 - ir no "Iniciar/Localizar/Computador" da barra de tarefas que fica na linha inferior da tela do seu computador;

2 - escrever no campo apropriado "estatisticos" (sem aspas e sem acento) e clicar com o mouse em "Localizar Agora";

3 - quando aparecer o computador estatisticos clicar duas vezes com o mouse nele vai aparecer uma nova janela com a pasta "SAS-V8", "clicar" UMA única vez nela;

4 - na parte superior da janela, onde aparece a pasta "SAS-V8", escolher "Arquivo/Mapear Unidade de Rede", vai aparecer uma nova janela onde se tem a opção de escolher a "Letra da Unidade", escolha a letra "S" (você pode marcar ou não a caixinha de "Reconectar ao iniciar", em geral não é recomendável), finalmente clique em "OK";

5 - na nova janela aparece uma nova pasta "SAS", clique nela duas vezes; ao aparecer um novo grupo de arquivos clique duas vezes na pasta "V8" e procure pelo arquivo "SAS.EXE" (o ".EXE" final pode ou não aparecer dependendo de como seu micro esta configurado, mas trata-se de um ícone cinza, uma pirâmide com a ponta para baixo e uma bolinha vermelha orbitando a pirâmide);

6 - com o mouse, aperte o botão esquerdo (sem soltar) neste ícone e arraste-o para sua tela de computador (será criado um atalho para o executável do SAS);

7 - pode-se agora fechar todas as janelas abertas neste processo;

Basta agora clicar duas vezes no ícone do "SAS" criado na sua tela principal para que o SAS seja executado.

Se o computador que você está instalando o SAS tiver o Windows 98 como sistema operacional, possivelmente você obterá uma mensagem de erro na janela LOG que significa que o Enhanced Editor não foi instalado. Para instalar este editor, siga os passos fornecidos a seguir:

1 - Feche todos os aplicativos;

2 - ir no "Iniciar/Localizar/Computador" da barra de tarefas que fica na linha inferior da tela;

3 - escrever no campo apropriado "estatisticos" (sem aspas e sem acento) e clicar com o mouse em "Localizar Agora";

4 - quando aparecer o computador estatisticos clicar duas vezes com o mouse nele; vai aparecer uma nova janela com a pasta "SAS-V8", "clicar" DUAS única vez nela;

5 - na nova janela aparece uma nova pasta "SAS", clique nela duas vezes; ao aparecer um novo grupo de arquivos "clique" duas vezes na pasta "BUNDLES";

6 – na nova janela aparece uma pasta Eeditor. Clicar duas vezes nela.

7 – na nova janela, procure pelo ícone Setup. Clicar duas vezes nele. Siga todos os passos para a instalação do Enhanced Editor.

8 – após isto, clicar duas vezes no ícone do SAS. O programa agora deve conter 3 janelas abertas: Log, Output e Editor

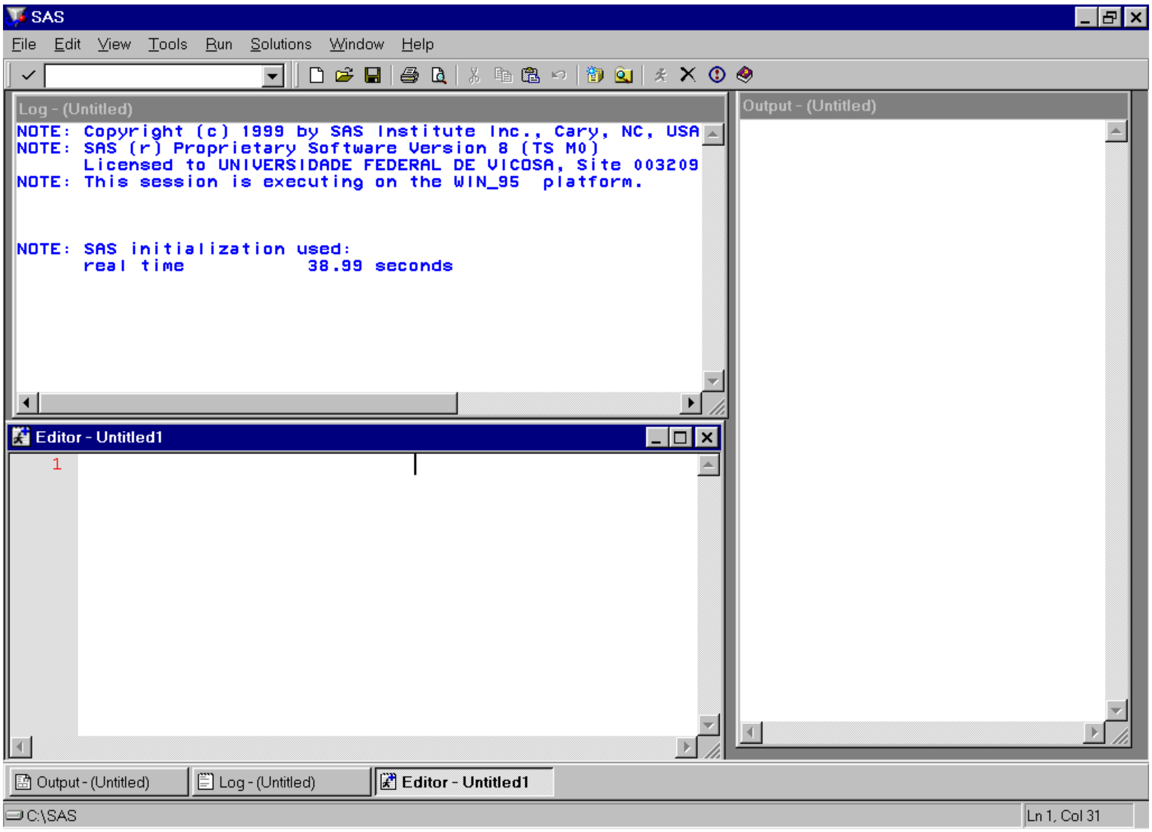
OBS: Estas instruções foram adaptadas do README que consta no computador estatisticos onde o SAS encontra-se instalado.

2 Conceitos Básicos no SAS

Esta seção tem como objetivo introduzir o usuário ao ambiente SAS. Inicialmente as "janelas" existentes no SAS são apresentadas. É por meio delas que o usuário se interage com o SAS.

2.1 Janelas existentes no programa SAS

Existem três janelas principais no ambiente SAS. São elas: editor de programas, log e output. Detalhes de cada uma delas são apresentados a seguir.



2.1.1 Janela do Editor de Programas (Editor Program window)

A janela do Editor de Programas é usada para editar programas e arquivos de dados. O usuário também pode submeter os seus programas usando esta janela.

Na versão 8.0 existem dois tipos de editores: Program Editor e Enhanced Editor. O último tem as mesmas características do primeiro com a vantagem de indicar erros de programação por meio de um jogo de cores no código do programa.

2.1.2 Janela de Mensagens (Log window)

Nesta janela são apresentadas mensagens relacionadas a execução de programas do usuário submetidos ao SAS. É aconselhável que, antes de submeter um novo programa, deletar as mensagens existentes. Aconselha-se também que, o usuário verifique as mensagens referentes ao programa que submeteu. Em geral, mensagens em azul indicam que não existem erros de programação. Mensagens em verde, indicam que "pequenos" erros de programação foram encontrados, mas que o próprio SAS "concertou" (com c mesmo, pois pode ser que o "concerto" dele não seja correto). Mensagens em vermelho, indicam que o SAS encontrou um erro de programação grave. Tais erros, podem impedir que uma programa seja executado e consequentemente nenhuma saída é obtida.

2.1.3 Janela de Saída (Output window)

Os resultados da execução de um programa são apresentados nesta janela. Vale lembrar que saídas poderão ser geradas mesmo quando existirem erros de programação. Portanto, é sempre bom olhar a janela de mensagens, antes de começar a interpretar os resultados mostrados na janela de saída.

2.2 Elaboração de programas do SAS (SAS jobs)

Um programa no SAS nada mais é do que um conjunto de comandos próprios (palavras chaves) do SAS, que são usados em uma sequência lógica, para realizar um conjunto de tarefas. Como mencionado anteriormente, os programas do SAS são normalmente editados na Janela do Editor de programas.

Um programa no SAS, em geral, consiste de dois passos (steps) distintos:

2.2.1 1º Passo: Data step

No data step, dados são lidos e convertidos em um arquivo de trabalho do SAS. Neste passo deve ser informado o nome de todas as variáveis, bem como o seu tipo (caracter ou alfanumérica), suas posições no arquivo de dados.

O conjunto de dados, se "pequeno", pode ser inserido no programa. Se ao contrário, o conjunto de dados for "grande", o mesmo deve ser salvo em arquivo à parte e o caminho para o SAS buscá-lo deve ser informado no DATA step. Normalmente, arquivos de dados gerados em outros softwares (excel, word, etc...) devem ser "importados" para o programa SAS. Consulte o help do SAS para verificar como isto deve ser realizado. O exemplo a seguir ilustra um DATA step.

Data step

```
data a1;
   input x1 $ x2 x3;
   x4 = x2 + x3;
   cards;
      a 1 14.1
      a 1 14.2
      b 2 16.3
      b 2 17.3
    ;
run;
```

Comentários:

1) a palavra chave DATA informa que o SAS deve criar um novo arquivo de trabalho, cujo nome é o que segue a palavra chave DATA. Este arquivo de trabalho, geralmente é deletado ao sair do SAS. As seguintes regras devem ser observadas ao dar um nome para um arquivo de trabalho do SAS:
- o nome deve conter de 1 a 8 caracteres;
- comece com qualquer letra ou _ (sublinhado) ;
- continue com números, letras ou _ ;
2) a palavra chave INPUT informa ao SAS os nomes das variáveis existentes no arquivo de dados. Neste exemplo, o conjunto de dados possui três variáveis cujos nomes são x1 x2 x3. O sinal $ depois do nome da primeira variável, informa ao SAS que a variável x1 é alfanumérica. A ausência deste sinal após os nomes das variáveis x2 e x3, indica que elas são apenas numéricas. Como não foi informado as colunas que cada uma delas ocupa no arquivo de dados, supõe-se que as mesmas ocupam as mesmas posições ao longo de todo o conjunto de dados e que também elas estejam separadas por pelo menos um espaço em branco;
3) uma nova variável, x4, é criada. Esta é o resultado da soma da variável x2 e x3. As regras para nomear variáveis são as mesmas dadas anteriormente para nomear SAS data sets;
4) a palavra chave CARDS informa ao SAS que, linhas a seguir são linhas que contém os valores das variáveis declaradas em INPUT;
5) as linhas a seguir são os valores das variáveis;
6) o ponto e vírgula na linha imediatamente após a última linha de dados, informa ao SAS que o conjunto de dados chegou ao fim;
7) a palavra chave RUN informa ao SAS que aquele passo, no caso DATA step, acabou e que ele pode executar esta parte da análise;
8) observe que ao final de cada linha de comandos (exceto as linhas que contém os dados, existe um ponto e vírgula. Este ponto e vírgula ao final de cada linha, informa ao SAS que aquele comando terminou. A falta de um ponto e vírgula em qualquer uma das linhas de comando, pode fazer com que o SAS não execute aquele passo. Este é um erro muito comum para principiantes do SAS. Se o usuário estiver usando o Enhanced

Editor, tais erros podem ser minimizados, pois a falta de um ponto e vírgula faz com que palavras chaves não apareçam em cores distintas;

9) a identação das linhas não é requerida, embora ajude ao usuário saber onde começa e termina cada passo do programa.

Se o conjunto de dados enconta-se em arquivo separado, por exemplo no disco *C*, na pasta *Meus Dados* e arquivo *arq1.txt*, então o caminho para o SAS ler este arquivo deve ser informado usando a palavra chave INFILE, como ilustrado a seguir:

data step	data a2 Infile ‘C:\Meus Dados\arq1.txt’ input x1 $ x2 x3 x4 = x2 + x3 run

Comentários adicionais para este exemplo:

1) a palavra chave INFILE informa ao SAS que o conjunto de dados encontra-se em um arquivo em separado do programa, cujo caminho é informado a seguir;
2) o caminho para o arquivo é escrito entre apóstrofes;

2.2.2 2º Passo: PROCedure step

Neste passo deve-se informar que tipo tarefa (análise estatística, ordenação de valores, gráficos, etc...) deseja-se realizar com um determinado arquivo de trabalho do SAS. O SAS possui uma variedade muito grande de procedimentos. O exemplo a seguir ilustra o uso do procedimento PRINT.

proc step	proc print data=a1; var x1 x2 x3; run;

O exemplo anterior é um exemplo muito simples, pois várias opções podem ser acrescentadas ao procedimento PRINT. No entanto, o objetivo agora é apenas fornecer uma idéia geral do passo PROC. Outros procedimentos e opções dos mesmos serão apresentados posteriormente.

Comentários:

1) um procedimento sempre inicia com a palavra chave PROC seguida do nome do procedimento que se deseja executar. No exemplo, o procedimento solicitado é o PRINT. Este procedimento pode ser usado para “imprimir” na tela conteúdos de um arquivo de trabalho do SAS.
2) seguindo o nome do procedimento desejado, deve-se informar qual arquivo de trabalho do SAS deve ser utilizado neste passo. Caso o usuário não indique o nome do arquivo de trabalho, o SAS utilizará o arquivo de trabalho que foi mais recentemente criado. Ao final desta linha deve existir um ponto e vírgula;

3) a palavra chave VAR é usada para informar ao SAS, quais variáveis devem ter seus valores impressos;
4) a última linha do procedimento deve conter a palavra chave RUN, seguida de ponto e vírgula. A presença desta palavra chave informa ao SAS que os comandos contidos naquele passo podem ser executados.

Em um único programa, qualquer um destes dois passos podem se repetir inúmeras vezes. Ou seja, diferentes procedimentos podem ser solicitados para um mesmo conjunto de dados, e diferentes conjunto de dados podem ser informados num mesmo programa.

2.3 Erros comuns na elaboração de um programa SAS

Como todo qualquer programa computacional, erros de programas SAS podem comprometer parcial ou totalmente sua execução. O SAS é um programa robusto, no sentido que se alguma palavra chave for escrita errada, o programa por si procede a correção e continua a execução com a versão corrigida. Neste caso, correções feitas pelo SAS são mostradas como linhas em verde na janela LOG. Erros deste tipo produzem saída. Vale lembrar, que cabe ao usuário verificar se a correção realizada pelo SAS é coerente ao desejado pelo usuário, e assim o mesmo poderá decidir se a saída obtida é satisfatória ou não.

Por outro lado, existem erros de programação que o SAS não corrige tais como:

- esquecer um ponto e vírgula no final de uma declaração;
- esquecer uma RUN statement;
- esquecer de fechar aspas.

Tais erros não produzem saídas. O SAS informa que identificou tais erros na janela LOG por meio de linhas vermelhas.

3 Análises Estatísticas

3.1 Variáveis classificatórias vs analíticas

Para realizar análises estatísticas, o SAS faz distinção entre variáveis classificatórias e analíticas.

3.1.1 Variáveis classificatórias

Em termos estatísticos, variáveis classificatórias são aquelas que poderíamos classificar como qualitativas. Por exemplo, a variável que identifica tratamentos numa análise de variância. São também conhecidas, em alguns casos, como variáveis independentes.

Estas variáveis podem ser:

- numéricas ou alfanuméricas;
- representar categorias discretas, se elas forem contínuas;
- identificar classes ou categorias nas quais os cálculos são efetuados.

3.1.2 Variáveis análiticas

Em termos estatísticos, variáveis analíticas são as variáveis que vamos usar para estudar o efeito de tratamentos, ou em outras palavras, elas são as variáveis respostas, ou valores observados em um experimento. São também conhecidas como variáveis dependentes.

Por sua vez, estas variáveis são:

- numéricas;
- apropriadas para o cálculo de médias, somas, ou outras estatísticas;
- contínuas (na maior parte dos casos).

4 Análise de dados oriundos de delineamentos experimentais

Delineamentos experimentais são utilizados para obter um maior controle do efeito do erro experiemental. A escolha de qual delineamento utilizar para instalar um experimento depende das condições do material experimental, por exemplo, uniformidade das unidades experimentais. O tipo de delineamento utilizado, define o modelo estatístico a ser usado para a análise dos dados. O programa SAS pressupõe que o usuário saiba qual delineamento foi utilizado e consequentemente o modelo estatístico a ser adotado na análise.

Basicamente o SAS tem dois procedimentos para a análise de dados de experimentos: ANOVA e GLM. O procedimento ANOVA é indicado quando os dados são balanceados e não existem valores perdidos. Por outro lado, o procedimento GLM é indicado quando os dados são desbalanceados.

A seguir são apresentados exemplos de programas para a análise de dados oriundos de experimentos instalados em diferentes tipos de delineamentos experimentais. A parte do programa que altera de delineamento para delineamento é apenas a referente a declaração MODEL, a qual está diretamente relacionada com o modelo estatístico do delineamento experimental.

4.1 Delineamento Inteiramente Casualizado (DIC)

A estrutura geral do programa abaixo pode ser utilizada para a análise de um experimento instalado segundo o DIC, com dados balanceados.

```
proc anova data=a1;
            class trt;
             model y = trt;
   means trt / duncan alpha=0.05;
run;
quit;
```

onde:

- proc ANOVA solicita que o procedimento ANOVA seja utilizado;
- data=a1 indica que o arquivo de trabalho do SAS “a1” deve ser utilizado na análise;
- A declaração CLASS informa quais fatores do modelo estatístico são classificatórias;

- A declaração MODEL informa qual modelo estatístico deve ser adotado durante a análise. No caso de um delineamento inteiramente casualizado, apenas a variável que identifica tratamentos deve ser informada, uma vez que o efeito da média geral e do erro estão presentes no modelo de todos os delineamentos;
- A declaração MEANS é opcional. Com ela é possível comparar médias de tratamentos e estabelecer o nível de significância do teste de médias. Dentre outros testes, pode ser solicitado o teste de Duncan, Tukey e Bonferroni;
- quit informa ao SAS que ele pode abandonar a execução do procedimento ANOVA.

Exemplo:
Os dados deste programa são do exercício 4.1

```
options nodate nocenter nonumber;
title 'Delineamento Inteiramente Casualizado';
data exerc_4_1;
  input varied prod @@;
  cards;
  1 25 2 31 3 22 4 33
  1 26 2 25 3 26 4 29
  1 20 2 28 3 28 4 31
  1 23 2 27 3 25 4 34
  1 21 2 24 3 29 4 28
;
run;
proc anova data=exerc_4_1;
  class varied;
  model prod = varied;
  means varied / tukey alpha=0.05;
  means varied / duncan alpha=0.05;
run;
quit;
```

A sentença “options” informa opções do formato da saída de um programa SAS. A opção “nodate” solicita ao SAS que não imprima a data da execução, a opção “nocenter” solicita que o texto da saída seja alinhado à esquerda; e a opção “nonumber” solicita que as páginas da saída não sejam numeradas.

A sentença “title” possibilita personificar as saídas do SAS. Existem 10 níveis de título de saída que podem ser definidos. O nível é identificado pelo número após “title”, ou seja “title2” se refere ao segundo nível de título.

4.2 Delineamento em Blocos Casualizados (DBC)

As únicas diferenças para o programa anterior estão nas declarações CLASS e MODEL. Ambas devem conter, adicionalmente, REP, pois, o fator repetição é uma variável classificatória e é parte do modelo estatístico do DBC. Assim uma forma geral de programa para este tipo de delineamento seria:

```
proc anova data=a1;
        class rep trt;
         model y = rep trt;
   means trt / duncan alpha=0.05;
run;
```

Exemplo

Os dados deste programa são do exercício 6.2

```
options nodate nocenter nonumber;
title 'Delineamento em Blocos Casualizados';
data exerc_6_2;
 input ta grupo prod @@;
 cards;
 1 1 30 1 2 32 1 3 33 1 4 34 1 5 29 1 6 30 1 7 33
 2 1 29 2 2 31 2 3 34 2 4 31 2 5 33 2 6 33 2 7 29
 3 1 43 3 2 47 3 3 46 3 4 47 3 5 48 3 6 44 3 7 47
 4 1 23 4 2 25 4 3 21 4 4 19 4 5 20 4 6 21 4 7 22
;
run;
proc anova data=exerc_6_2;
 class ta grupo;
 model prod = ta grupo;
 means ta / tukey alpha=0.05;
 means ta / duncan alpha=0.05;
run;
quit;
```

4.3 Delineamento em Quadrado Latino (DQL)

Em relação ao DIC, as declarações CLASS e MODEL devem ser alteradas para também conter as variáveis que identificam a linha e a coluna de cada valor observado. Uma forma geral para análise de um experimento instalado segundo o DQL seria:

```
proc anova;
    class linha coluna trt;
     model y = linha coluna trt;
     means trt / duncan;
run;
```

Exemplo

Os dados deste programa são do exercício 7.4

```
options nodate nocenter nonumber;
title 'Delineamento em Quadrado Latino';
data exerc_7_4;
 input leit faixa castracao $ ganho @@;
 cards;
   1 1 A  93.0 1 2 C 115.4 1 3 E 116.9 1 4 D 110.2 1 5 B 110.4
   2 1 C 110.6 2 2 E  96.5 2 3 B 108.9 2 4 A  97.6 2 5 D 112.0
   3 1 B 102.1 3 2 D 108.6 3 3 A  77.9 3 4 E 102.0 3 5 C 111.7
   4 1 D 115.4 4 2 A  94.9 4 3 C 114.0 4 4 B 100.2 4 5 E 118.5
   5 1 E 117.6 5 2 B 114.1 5 3 D 118.7 5 4 C 108.8 5 5 A  80.2
;
run;
proc glm data=exerc_7_4;
 class leit faixa castracao;
 model ganho = leit faixa castracao;
 means castracao / tukey alpha=0.05;
 means castracao / duncan alpha=0.05;
run;
quit;
```

4.4 Experimentos Fatoriais

Em experimentos fatoriais, existem no mínimo dois fatores sendo estudados simultaneamente num experimento. O interesse, nesta situação, é verificar se é significativo o efeito da interação e dos efeitos principais. Experimentos fatoriais podem ser instalados usando vários tipo de delineamentos.

Para contemplar análises de experimentos fatoriais, o programa proposto para o DIC deve ser modificado nas declarações CLASS e MODEL. Suponha que os dois fatores em estudos, ditos A e B, estão sendo estudados e que o experimento foi instalado segundo o DBC. Uma forma geral para um programa como este seria

```
proc glm;
  class rep a b;
  model y = rep a b a*b;
run;
```

Veja que o termo da interação foi incluído na declaração MODEL, para que seja realizado o teste F para o efeito da mesma. No caso da interação ser significativa, seria desejável proceder ao estudo de um fator dentro de cada nível do outro fator. Para atingir tal objetivo, inclua as seguintes linhas após a declaração MODEL:

```
lsmeans a*b / slice=b;
lsmeans a*b / slice=a;
```

Se a interação for não significativa, podemos estudar um fator independente do outro. Neste caso as seguintes linhas poderiam ser incluídas no programa inicial:

```
lsmeans a / tukey;
lsmeans b / tukey;
```

Exemplo

O enunciado para este exemplo foi fornecido em sala de aula, o qual foi retirado do livro BANZATTO e KRONKA (1989).

```
options nodate nocenter nonumber;
title 'Experimentos Fatoriais';
data exemplo_8_extra;
 input recipiente especie altura @@;
 cards;
   1 1 26.2 1 1 26.0 1 1 25.0 1 1 25.4
   1 2 24.8 1 2 24.6 1 2 26.7 1 2 25.2
   2 1 25.7 2 1 26.3 2 1 25.1 2 1 26.4
   2 2 19.6 2 2 21.1 2 2 19.0 2 2 18.6
   3 1 22.8 3 1 19.4 3 1 18.8 3 1 19.2
   3 2 19.8 3 2 21.4 3 2 22.8 3 2 21.3
  ;
run;
proc glm data=exemplo_8_extra;
  class recipiente especie;
  model altura = recipiente especie recipiente*especie;
  lsmeans recipiente*especie / slice=especie;
  lsmeans recipiente*especie / slice=recipiente;
run;
quit;
```

4.5 Experimentos em Parcelas Subdivididas

Tal como no caso de experimentos fatoriais, experimentos em parcelas subdivididas são usados quando se deseja estudar dois ou mais fatores simultaneamente num mesmo experimento. A diferença é que em experimentos em parcelas subdivididas um fator, dito principal, é designado segundo um tipo de delineamento as parcelas que contém várias unidades experimentais. O segundo fator, é então designado aleatoriamente às subparcelas de cada parcela.

O objetivo em parcelas subdividas também é verificar se os efeitos principais e interação entre fatores são significativos. O programa para esta situação difere no fato de ser necessário indicar o resíduo correto para testar o fator principal uma vez que o SAS assume que todos os fatores devem ser testados contra o erro(b). A seguir está uma forma geral para a análise de um experimento em parcelas subdivididas, considerando que o fator A é o fator principal e o fator B é o secundário e que o experimento foi instalado segundo o DBC:

```
proc glm;
    class rep a b;
    model y = a rep a*rep b a*b;
     test h=a e=a*rep;
 run;
 quit;
```

Exemplo

Os dados deste programa são do exercício 9.1

```
options nodate nocenter nonumber;
title 'Experimentos em Parcelas Subdivididas';
data exemplo_9_1;
 input variedade defensivo bloco producao @@;
 cards;
   1 1 1 42.9 1 1 2 41.6 1 1 3 28.9 1 1 4 30.8
   1 2 1 53.8 1 2 2 58.5 1 2 3 43.9 1 2 4 46.3
   1 3 1 49.5 1 3 2 53.8 1 3 3 40.7 1 3 4 39.4
   1 4 1 44.4 1 4 2 41.8 1 4 3 28.3 1 4 4 34.7
   2 1 1 53.3 2 1 2 69.6 2 1 3 45.4 2 1 4 35.1
   2 2 1 57.6 2 2 2 69.6 2 2 3 42.4 2 2 4 51.9
   2 3 1 59.8 2 3 2 65.8 2 3 3 41.4 2 3 4 45.4
   2 4 1 64.1 2 4 2 57.4 2 4 3 44.1 2 4 4 51.6
   3 1 1 62.3 3 1 2 58.5 3 1 3 44.6 3 1 4 50.3
   3 2 1 63.4 3 2 2 50.4 3 2 3 45.0 3 2 4 46.7
   3 3 1 64.5 3 3 2 46.1 3 3 3 62.6 3 3 4 50.3
   3 4 1 63.6 3 4 2 56.1 3 4 3 52.7 3 4 4 51.8
   4 1 1 75.4 4 1 2 65.6 4 1 3 54.0 4 1 4 52.7
   4 2 1 70.3 4 2 2 67.3 4 2 3 57.6 4 2 4 58.5
   4 3 1 68.8 4 3 2 65.3 4 3 3 45.6 4 3 4 51.0
   4 4 1 71.6 4 4 2 69.4 4 4 3 56.6 4 4 4 47.4
  ;
run;
proc glm data=exemplo_9_1;
  class variedade defensivo bloco;
  model producao = variedade defensivo bloco
            variedade*bloco variedade*defensivo;
  test h=variedade e=variedade*bloco;
run;
quit;
```

5 Regressão Linear

Regressão é geralmente usada quando deseja-se verificar se um fator quantitativo exerce influência sob uma variável dependente. No programa SAS, existem dois procedimentos distintos para tal finalidade, GLM e REG.

O procedimento REG é mais usado quando se ajusta um modelo contendo apenas fatores quantitativos como variáveis independentes. Já o procedimento GLM possibilita o ajuste de modelos de covariância, ou seja, modelos que incluem tanto fatores qualitativos como quantitativos como variáveis independentes. A seguir é fornecido explicações mais detalhadas a respeito de cada um deles.

5.1 PROC REG

O procedimento REG usa o método dos quadrados mínimos para estimar os parâmetros num modelo linear. Com este procedimento é possível. A estrutura geral de um programa usando o procedimento REG é:

```
proc reg data=a1;
    var variables;
    model y = x1 x2 ... / options;
    output out=newset keyword=name1 keyword=name2;
    test equation1, equation2;
run;
```

Comentários:

- a declaração REG solicita que o SAS utilize o procedimento REG;
- a declaração DATA informa ao SAS qual DATA set deve ser utilizado neste procedimento;
- a declaração VAR especifica todas as variáveis que serão utilizadas na análise;
- a declaração MODEL especifica que y é a variável dependente e x1, x2, ... são as variáveis independentes;
- a declaração OPTIONS solicita ao SAS que algumas saídas adicionais sejam impressas. Algumas das possibilidades são:
    + p: imprime valores preditos;
    + r: imprime resíduos;
    + xpx: imprime a matriz X'X;
    + i: imprime a inversa da matriz X'X;
    + covb: imprime a matriz de variância e covariância das estimativas dos parâmetros;
- a declaração OUTPUT statement cria um SAS DATA set com as variáveis definidas nas declarações KEYWORD. Para este procedimento as KEYWORD's podem ser PREDICT, RESIDUAL, etc ...
- TEST statement é usada para realizar teste de significância para cada uma das equações listadas. Por exemplo, TEST X1=0, X2=0; testa se os coeficientes de X1 e X2 são simultaneamente iguais a zero.

Exemplos

1) Regressão Linear Simples

Os dados deste programa são do exercício 10.5. O programa a seguir produz uma regressão linear de y em x. Usando a declaração OUTPUT, é possível salvar os valores preditos (yhat) e residuais (resid) em outro arquivo de trabalho (new).

```
options nodate nocenter nonumber;
title 'Regressão Linear Simples de 1o grau';
data exerc_10_5;
 input x y @@;
 cards;
 50 1.2 75 1.7 100 2.0 125 2.1 150 2.5
 ;
run;
proc reg data=exerc_10_5;
  model y = x;
  output out=new p=yhat r=resid;
run;
quit;
```

2) Regressão linear polinomial

Para modelos de regressão polinomial, é necessário criar as potências das variáveis logo depois da declaração INPUT.

Os dados deste programa são do exercício 10.4. O programa ilustra análise de dados segundo um modelo linear de 2º grau.

```
options nodate nocenter nonumber;
title 'Regressão Linear Simples de 2o grau';
data exerc_10_4;
 input x y @@;
 x2 = x*x;
 cards;
 1.0 20.3 2.5 31.3 4.0 34.6 5.5 35.1 7.0 30.2 8.5 19.7
 ;
run;
proc reg data=exerc_10_4;
  model y = x x2;
  output out=new p=yhat r=resid;
run;
quit;
```

6 Saídas do Programa SAS

Para identificar a qual programa pertence cada saída, basta comparar a 1ª linha de cada página destas saídas com o que está escrito na declaração **title** do programa.

Por exemplo:

Programa:

```
options nodate nocenter nonumber;
title 'Delineamento Inteiramente Casualizado';
data exerc_4_1;
...
```

1ª linha da saída do programa:

```
Delineamento Inteiramente Casualizado
```

OBSERVAÇÃO: O conteúdo das saídas aqui mostradas é apenas um resumo de uma saída normal do SAS. Aqui são apresentadas apenas os resultados mais importantes de uma saída de um programa do SAS.

Delineamento Inteiramente Casualizado

The ANOVA Procedure

```
    Class Level Information
Class         Levels    Values
var                4    1 2 3 4

Dependent Variable: y
                              Sum of
Source             DF        Squares     Mean Square    F Value     Pr > F
Model               3    163.7500000      54.5833333       7.80     0.0020
Error              16    112.0000000       7.0000000
Corrected Total    19    275.7500000

R-Square     Coeff Var      Root MSE        y Mean
0.593835      9.890659      2.645751      26.75000

Source             DF       Anova SS     Mean Square    F Value     Pr > F
var                 3    163.7500000      54.5833333       7.80     0.0020
```

Tukey's Studentized Range (HSD) Test for y

```
Alpha                                   0.05
Error Degrees of Freedom                  16
Error Mean Square                          7
Critical Value of Studentized Range  4.04609
Minimum Significant Difference        4.7874

Means with the same letter are not significantly different.
Tukey
Grouping        Mean      N    var
     A        31.000      5    4
B    A        27.000      5    2
B             26.000      5    3
B             23.000      5    1
```

Duncan's Multiple Range Test for y

```
Alpha                        0.05
Error Degrees of Freedom       16
Error Mean Square               7

Number of Means          2          3          4
Critical Range       3.547      3.720      3.828

Means with the same letter are not significantly different.

Duncan
Grouping        Mean      N    var
     A        31.000      5    4
     B        27.000      5    2
C    B        26.000      5    3
C             23.000      5    1
```

Delineamento em Blocos Casualizados

The ANOVA Procedure

Class Level Information

Class	Levels	Values
ta	4	1 2 3 4
grupo	7	1 2 3 4 5 6 7

Number of observations 28

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	2142.714286	238.079365	59.76	<.0001
Error	18	71.714286	3.984127		
Corrected Total	27	2214.428571			

R-Square	Coeff Var	Root MSE	y Mean
0.967615	6.114746	1.996028	32.64286

Source	DF	Anova SS	Mean Square	F Value	Pr > F
ta	3	2125.285714	708.428571	177.81	<.0001
grupo	6	17.428571	2.904762	0.73	0.6323

Tukey's Studentized Range (HSD) Test for y

Alpha	0.05
Error Degrees of Freedom	18
Error Mean Square	3.984127
Critical Value of Studentized Range	3.99698
Minimum Significant Difference	3.0154

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	ta
A	46.000	7	3
B	31.571	7	1
B	31.429	7	2
C	21.571	7	4

Duncan's Multiple Range Test for y

Alpha	0.05
Error Degrees of Freedom	18
Error Mean Square	3.984127

Number of Means	2	3	4
Critical Range	2.242	2.352	2.421

Means with the same letter are not significantly different.

Duncan Grouping	Mean	N	ta
A	46.000	7	3
B	31.571	7	1
B	31.429	7	2
C	21.571	7	4

Delineamento em Quadrado Latino

The GLM Procedure

Class Level Information

Class	Levels	Values
leit	5	1 2 3 4 5
faixa	5	1 2 3 4 5
castracao	5	A B C D E

Number of observations 25

Dependent Variable: ganho

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	2326.379200	193.864933	3.46	0.0204
Error	12	672.103200	56.008600		
Corrected Total	24	2998.482400			

R-Square	Coeff Var	Root MSE	ganho Mean
0.775852	7.043793	7.483889	106.2480

Source	DF	Type I SS	Mean Square	F Value	Pr > F
leit	4	257.826400	64.456600	1.15	0.3796
faixa	4	48.498400	12.124600	0.22	0.9242
castracao	4	2020.054400	505.013600	9.02	0.0013

Tukey's Studentized Range (HSD) Test for ganho

Alpha	0.05
Error Degrees of Freedom	12
Error Mean Square	56.0086
Critical Value of Studentized Range	4.50760
Minimum Significant Difference	15.086

Tukey Grouping	Mean	N	castracao
A	112.980	5	D
A	112.100	5	C
A	110.300	5	E
A	107.140	5	B
B	88.720	5	A

Duncan's Multiple Range Test for ganho

Alpha	0.05
Error Degrees of Freedom	12
Error Mean Square	56.0086

Number of Means	2	3	4	5
Critical Range	10.31	10.79	11.09	11.28

Duncan Grouping	Mean	N	castracao
A	112.980	5	D
A	112.100	5	C
A	110.300	5	E
A	107.140	5	B
B	88.720	5	A

Experimentos Fatoriais

The GLM Procedure

Class Level Information

Class	Levels	Values
recipiente	3	1 2 3
especie	2	1 2
Number of observations		24

Experimentos Fatoriais

Dependent Variable: altura

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	175.7033333	35.1406667	27.39	<.0001
Error	18	23.0900000	1.2827778		
Corrected Total	23	198.7933333			

R-Square	Coeff Var	Root MSE	altura Mean
0.883849	4.931485	1.132598	22.96667

Source	DF	Type I SS	Mean Square	F Value	Pr > F
recipiente	2	92.86083333	46.43041667	36.20	<.0001
especie	1	19.08166667	19.08166667	14.88	0.0012
recipiente*especie	2	63.76083333	31.88041667	24.85	<.0001

recipiente*especie Effect Sliced by especie for altura

especie	DF	Sum of Squares	Mean Square	F Value	Pr > F
1	2	87.121667	43.560833	33.96	<.0001
2	2	69.500000	34.750000	27.09	<.0001

recipiente*especie Effect Sliced by recipiente for altura

recipiente	DF	Sum of Squares	Mean Square	F Value	Pr > F
1	1	0.211250	0.211250	0.16	0.6897
2	1	79.380000	79.380000	61.88	<.0001
3	1	3.251250	3.251250	2.53	0.1288

Experimentos em Parcelas Subdivididas

The GLM Procedure

Class Level Information

Class	Levels	Values
variedade	4	1 2 3 4
defensivo	4	1 2 3 4
bloco	4	1 2 3 4

Number of observations 64

Dependent Variable: producao

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	27	7066.191875	261.710810	12.89	<.0001
Error	36	731.202500	20.311181		
Corrected Total	63	7797.394375			

R-Square	Coeff Var	Root MSE	producao Mean
0.906225	8.534077	4.506793	52.80938

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Variedade	3	2848.021875	949.340625	46.74	<.0001
defensivo	3	170.536875	56.845625	2.80	0.0539
bloco	3	2842.873125	947.624375	46.66	<.0001
variedade*bloco	9	618.294375	68.699375	3.38	0.0042
variedade*defensiv	9	586.465625	65.162847	3.21	0.0059

Tests of Hypotheses Using the Type III MS for variedade*bloco as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Variedade	3	2848.021875	949.340625	13.82	0.0010

Regressão Linear Simples de 1o grau

The REG Procedure

Dependent Variable: y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.90000	0.90000	67.50	0.0038
Error	3	0.04000	0.01333		
Corrected Total	4	0.94000			

Root MSE	0.11547	R-Square	0.9574
Dependent Mean	1.90000	Adj R-Sq	0.9433
Coeff Var	6.07737		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.70000	0.15492	4.52	0.0203
x	1	0.01200	0.00146	8.22	0.0038

Regressão Linear Simples de 2o grau

The REG Procedure

Dependent Variable: y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	234.81498	117.40749	259.30	0.0004
Error	3	1.35836	0.45279		
Corrected Total	5	236.17333			

Root MSE	0.67289	R-Square	0.9942
Dependent Mean	28.53333	Adj R-Sq	0.9904
Coeff Var	2.35827		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	11.24222	0.97376	11.55	0.0014
x	1	10.46770	0.47719	21.94	0.0002
x2	1	-1.11349	0.04895	-22.75	0.0002

Anexo 4 – p-valor

O p-valor representa a probabilidade estimada no experimento, de rejeitar Ho quando Ho é verdadeira, de acordo com o número de graus de liberdade (gl) e do teste de hipóteses utilizado. Portanto, quanto menor for o p-valor, mais forte será a evidência de que Ho deverá ser rejeitada. Em termos práticos, tem-se a seguinte regra de decisão em relação ao nível de significância α de referência:

- Se p-valor $\leq \alpha \Rightarrow$ rejeitar Ho;
- Se p-valor $> \alpha \Rightarrow$ não rejeitar Ho.

No Excel (inserir função estatística), os p-valores e os valores tabelados do teste t são obtidos, como seguem:

- DISTT ($|t_{cal}|$; gl; 2) = p-valor $\Rightarrow$ Ha bilateral;
- DISTT ($|t_{cal}|$; gl; 1) = p-valor $\Rightarrow$ Ha unilateral;
- INVT (α; gl) = $t_{tab} \Rightarrow$ Ha bilateral;
- INVT (2α; gl) = $t_{tab} \Rightarrow$ Ha unilateral.

No Excel (inserir função estatística), os p-valores e os valores tabelados do teste F são obtidos, como seguem:

- DISTF (F_{cal}; gl numerador; gl denominador) = p-valor $\Rightarrow$ Ha unilateral;
- INVF (α; gl numerador; gl denominador) = $F_{tab} \Rightarrow$ Ha unilateral.

Anexo 5 – Exemplo Extra ANOVA

	Tratamentos						ANOVA				
Repetições	1	2	3	4	5	....	FV	GL	SQ	QM	F
1	100	100	100	100	100		Tratamentos	4	0	0	-
2	100	100	100	100	100		Resíduo	10	0	0	
3	100	100	100	100	100		Total	14	0		
Totais	300	300	300	300	300						
Médias	100	100	100	100	100						

	Tratamentos						ANOVA				
Repetições	1	2	3	4	5	....	FV	GL	SQ	QM	F
1	90	80	70	60	50		Tratamentos	4	0	0	0
2	100	100	100	100	100		Resíduo	10	11000	1100	
3	110	120	130	140	150		Total	14	11000		
Totais	300	300	300	300	300						
Médias	100	100	100	100	100						

	Tratamentos						ANOVA				
Repetições	1	2	3	4	5	....	FV	GL	SQ	QM	F
1	98	99	100	101	102		Tratamentos	4	30	7,5	Inf
2	98	99	100	101	102		Resíduo	10	0	0	
3	98	99	100	101	102		Total	14	30		
Totais	294	297	300	303	306						
Médias	98	99	100	101	102						

	Tratamentos						ANOVA				
Repetições	1	2	3	4	5	....	FV	GL	SQ	QM	F
1	98	99	100	101	102		Tratamentos	4	30	7,5	7,50
2	99	100	101	102	103		Resíduo	10	10	1,0	
3	100	101	102	103	104		Total	14	40		
Totais	297	300	303	306	309						
Médias	100	100	100	100	100						

Pressuposições da Análise de Variância

Na análise de variância, os valores observados Y_{ik} de uma variável resposta são descritos em termos de um modelo estatístico. Uma das pressuposições para a realização da análise de variância é que o modelo estatístico seja composto pela soma de efeitos, os quais podem ser fixos ou aleatórios. Em geral, o efeito do fator em estudo é considerado fixo. Enquanto que o efeito do erro experimental é considerado aleatório.

Por exemplo, para os valores observados em um experimento instalado segundo o delineamento inteiramente casualizado (DIC) com I tratamentos e K repetições, o modelo estatístico é

$$Y_{ik} = m + t_i + e_{ik}$$

em que,

Y_{ik}: é o valor observado para a variável resposta obtido para o i-ésimo tratamento em sua k-ésima repetição;

m: é a média fixa de todos os valores possíveis da variável resposta;

t_i: é o efeito fixo do tratamento i no valor observadoY_{ik};

$$t_i = m_i - m$$

e_{ik}: é o efeito aleatório do erro ou resíduo experimental associado ao valor observado Y_{ik}, definido por

$$e_{ik} = Y_{ik} - m_i$$

As pressuposições para a validade dos resultados da análise de variância são que os erros experimentais

1. Sigam uma distribuição normal;
2. Tenham variância comum e
3. Sejam independentes

A estimativa do erro experimental, no DIC, é obtida pela diferença entre o valor observado e o respectivo valor predito $\hat{Y}_{ik}$, ou seja,

$$\hat{e}_{ik} = Y_{ik} - \hat{Y}_{ik}.$$

O valor predito é obtido por

$$\hat{Y}_{ik} = \hat{m} + \hat{t}_i$$

A estimativa do efeito do tratamento i, $\hat{t}_i$, por sua vez é obtida por

$$\hat{t}_i = \hat{m}_i - \hat{m}$$

Portanto temos que

$$\hat{Y}_{ik} = \hat{m}_i$$

Então a estimativa do resíduo experimental, $\hat{e}_{ik}$, de acordo como o modelo estatístico apresentado anteriormente é obtida por

$$\hat{e}_{ik} = Y_{ik} - \hat{m}_i.$$

Portanto, antes de interpretar os resultados da análise de variância recomenda-se verificar, por meios dos procedimentos descritos a seguir, se as estimativas dos resíduos satisfazem as pressuposições da análise de variância.

1ª Pressuposição) Normalidade da distribuição dos erros experimentais

Para verificar se os resíduos associados ao modelo estatístico utilizado aderem a uma distribuição normal, pode-se realizar o teste de hipóteses de Lilliefors. As hipóteses para este teste são:

H_0: os resíduos experimentais seguem uma distribuição normal

H_a: os resíduos experimentais não seguem uma distribuição normal.

Este teste se baseia na comparação da freqüência acumulada empírica com a freqüência acumulada teórica, as quais são obtidas para cada valor do resíduo experimental.

Após a ordenação crescente dos valores residuais, a freqüência acumulada empírica, $S(\hat{e}_{ik})$ é obtida por

$$S(\hat{e}_{ik}) = \frac{\text{nº de valores} < \hat{e}_{ik}}{n}$$

Por outro lado, para obter o valor da freqüência acumulada teórica, $F(\hat{e}_{ik})$, para cada valor $\hat{e}_{ik}$, é necessário especificar a que distribuição normal os resíduos experimentais tendem a se aderir.

Uma distribuição normal é especificada pelos parâmetros média e variância. Na realização deste teste, assume-se que os parâmetros da suposta distribuição normal dos resíduos são iguais aos valores da média e variância dos resíduos experimentais.

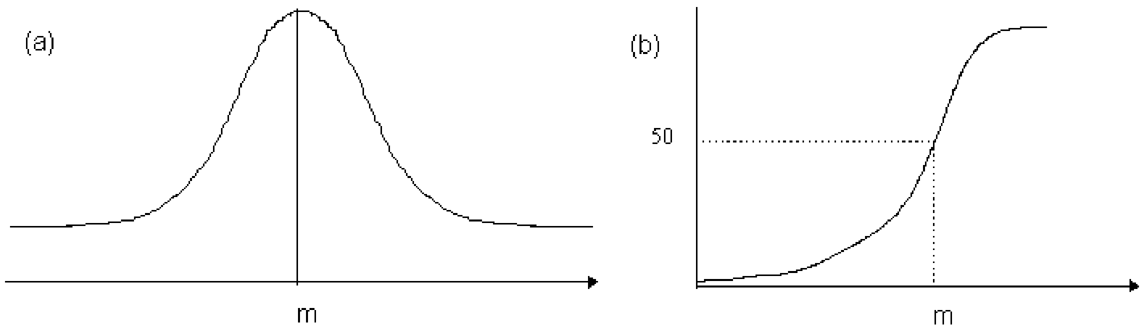
A partir da especificação dos parâmetros da distribuição normal é possível calcular a freqüência acumulada teórica. A distribuição acumulada é definida como $F(\hat{e}_{ik}) = P(\hat{E}_{ik} \leq \hat{e}_{ik})$.

Supondo que a distribuição dos resíduos experimentais tenha sido definida como $\hat{E}_{ik} \sim N(m; \sigma^2)$, então o valor de $F(\hat{e}_{ik})$ é obtido por

$$F(\hat{e}_{ik}) = \int_{-\infty}^{\hat{e}_{ik}} f(\hat{e}_{ik}) d(\hat{e}_{ik}) = \int_{-\infty}^{\hat{e}_{ik}} \frac{1}{2\pi\sqrt{\sigma^2}} e^{-\frac{1}{2}\frac{(\hat{e}_{ik}-m)^2}{\sigma^2}} d(\hat{e}_{ik})$$

Uma representação genérica para os gráficos de uma distribuição normal e respectiva distribuição acumulada teórica são apresentados na Figura 1 - (a) e (b), respectivamente.

Figura 1 – Distribuição normal (a) e respectiva distribuição acumulada (b)



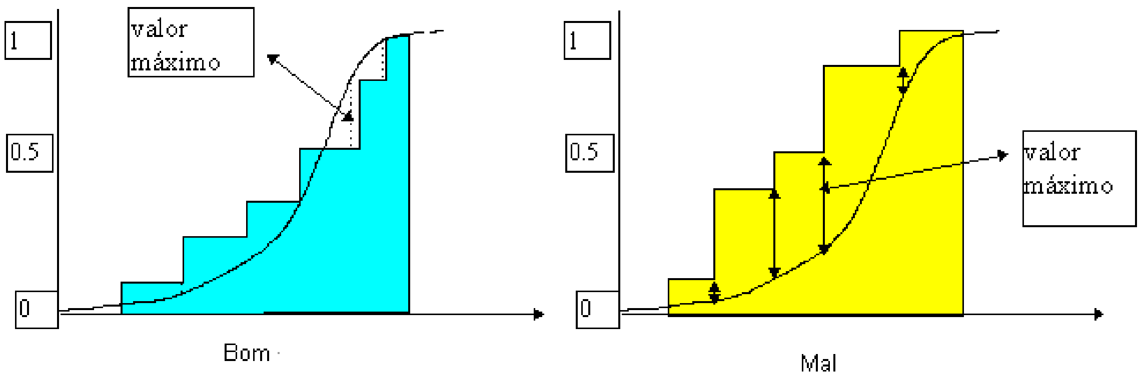
Espera-se que para cada valor $\hat{e}_{ik}$ os valores obtidos para $S(\hat{e}_{ik})$ e $F(\hat{e}_{ik})$ sejam bem similares, caso os resíduos experimentais sigam a distribuição normal especificada. É por esta razão que o teste de Lilliefors se baseia na comparação destes dois valores de distribuição acumulada.

Após a ordenação em ordem crescente (j = 1, 2, ... , n) dos resíduos experimentais são obtidos, para cada $\hat{e}_{ik}$, os módulos das diferenças entre $F(\hat{e}_{ik})_j - S(\hat{e}_{ik})_j$ e entre $F(\hat{e}_{ik})_j - S(\hat{e}_{ik})_{(j-1)}$. O teste de Lilliefors se baseia na maior diferença absoluta encontrada. Esta diferença é definida como sendo a estatística d obtida por

$$d = \max_{j}\left\{\left|F(\hat{e}_{ik})_j - S(\hat{e}_{ik})_j\right|, \left|F(\hat{e}_{ik})_j - S(\hat{e}_{ik})_{(j-1)}\right|\right\}$$

O valor da estatística d é então comparado com o valor tabelado d_{tab} de acordo com o nível de significância α e do número de resíduos experimentais na Figura 2 apresenta as situações com um bom ajustamento a uma distribuição normal e outra com um mal ajustamento. Nesta Figura 2, a curva representa a distribuição acumulada teórica, e a escada representa a distribuição acumulada empírica.

Figura 2 – Ilustrações de um bom ajuste a um mal ajuste de uma distribuição normal



Suponha os dados do Exemplo 4.1

Para comparar a produtividade de quatro variedades de milho, um agrônomo tomou vinte parcelas similares e distribuiu, inteiramente ao acaso, cada uma das 4 variedades em 5 parcelas experimentais.

	Variedades			
	1	2	3	4
	25	31	22	33
	26	25	26	29
	20	28	28	31
	23	27	25	34
	21	24	29	28
Totais	115	135	130	155
Médias	23	27	26	31

Neste caso como foi utilizado o DIC temos que o modelo estatístico é

$$Y_{ik} = m + t_i + e_{ik}$$

Portanto, segundo o exposto anteriormente, são apresentados na Tabela 1 os valores observados e respectivos valores preditos residuais.

Tabela 1 – Valores observados (Y_{ik}) e respectivos valores preditos ($\hat{Y}_{ik}$) e resíduais ($\hat{e}_{ik}$)

Variedade	Repetição	Y_{ik}	$\hat{Y}_{ik}$	$\hat{e}_{ik}$
1	1	25	23	2
1	2	26	23	3
1	3	20	23	-3
1	4	23	23	0
1	5	21	23	-2
2	1	31	27	4
2	2	25	27	-2
2	3	28	27	1
2	4	27	27	0
2	5	24	27	-3
3	1	22	26	-4
3	2	26	26	0
3	3	28	26	2
3	4	25	26	-1
3	5	29	26	3
4	1	33	31	2
4	2	29	31	-2
4	3	31	31	0
4	4	34	31	3
4	5	28	31	-3
Média	-	26,75	26,75	0
Variância	-	14,51	8,62	5,89
Desvio-padrão	-	3,81	2,94	2,43
N	-	20	20	20

A partir da Tabela 1, podemos obter as distribuições de freqüência dos valores residuais apresentadas na Tabela 2 Estas distribuições de freqüências serão denominadas daqui para frente de distribuições de freqüência empíricas.

Tabela 2 – Distribuições de freqüências empíricas dos resíduos ($\hat{e}_{ik}$) e respectivas freqüências acumuladas teóricas nas quais os resíduos aparecem em ordem (j) crescente

j	$\hat{e}_{ik}$	Freqüências Empíricas			Freqüência Teórica Acumulada		
		Simples	Relativa	Acumulada $S(\hat{e}_{ik})$	$F(\hat{e}_{ik})$	$\lvert F(\hat{e}_{ik})_j - S(\hat{e}_{ik})_{(j-1)}\rvert$	$\lvert F(\hat{e}_{ik})_j - S(\hat{e}_{ik})_{(j)}\rvert$
0				0	0		
1	-4	1	0.05	0.05	0.0497	0.0497	0.0003
2	-3	3	0.15	0.20	0.1083	0.0583	0.0917
3	-2	3	0.15	0.35	0.2050	0.0050	$\boxed{0.1450}$
4	-1	1	0.05	0.40	0.3402	0.0098	0.0598
5	0	4	0.20	0.60	0.5000	0.1000	0.1000
6	1	1	0.05	0.65	0.6598	0.0598	0.0098
7	2	3	0.15	0.80	0.7950	$\boxed{0.1450}$	0.0050
8	3	3	0.15	0.95	0.8917	0.0917	0.0583
9	4	1	0.05	1.00	0.9503	0.0003	0.0497

Na Tabela 2 também é apresentada a distribuição acumulada de freqüência teórica para os valores residuais $\hat{e}_{ik}$. A freqüência teórica acumulada foi obtido supondo que os resíduos seguem uma distribuição normal com média igual a zero e variância igual a 5,89, ou seja, $\hat{e}_{ik} \sim N(0;5,89)$. Para encontrar o valor da freqüência teórica acumulada, por exemplo, para os valores dos resíduos igual a – 4 e – 3, foram utilizados

$$\hat{e}_{ik} = -4 \rightarrow \int_{-\infty}^{-4} f(\hat{e})d(\hat{e}) = \int_{-\infty}^{-4} \frac{1}{2\pi\sqrt{5,89}} e^{-\frac{1}{2}\frac{(\hat{e}-0)^2}{5,89}} d(\hat{e}) = 0,0497$$

$$\hat{e}_{ik} = -3 \rightarrow \int_{-\infty}^{-3} f(\hat{e})d(\hat{e}) = \int_{-\infty}^{-3} \frac{1}{2\pi\sqrt{5,89}} e^{-\frac{1}{2}\frac{(\hat{e}-0)^2}{5,89}} d(\hat{e}) = 0,1083$$

Para obter estes valores sem calcular estas integrais basta converter tais valores usando a distribuição normal padrão ou seja,

$$\hat{e}_{ik} = -4 \rightarrow z = \frac{-4-0}{\sqrt{5,89}} = -1,65 \rightarrow P(Z < -1,65) = 0,0495$$

$$\hat{e}_{ik} = -3 \rightarrow z = \frac{-3-0}{\sqrt{5,89}} = -1,23 \rightarrow P(Z < -1,23) = 0,1093$$

Como pode ser notado os valores não são exatamente iguais aos apresentados na Tabela 2 Isto ocorre devido as aproximações realizadas durante o cálculo. Para gerar os valores apresentados na Tabela 2, foi utilizada a função INV.NORMP do software Excel. Para os resíduos -4 e -3 foram utilizadas INV.NORMP(-4;0;2,43) e INV.NORMP(-3;0;2,43), respectivamente.

Ao observarmos a Tabela 2 podemos verificar que a estatística d

$$d = \max_j \left\{ \left| F(\hat{e}_{ik})_j - S(\hat{e}_{ik})_j \right|, \left| F(\hat{e}_{ik})_j - S(\hat{e}_{ik})_{(j-1)} \right| \right\}$$

para os dados do Exemplo 4.1 é igual a 0,1450. O respectivo valor tabelado, obtido na Tabela 3, para α=5% e n=20 é d_{tab} =0,220.

As hipóteses para este teste são:

H_0: os resíduos experimentais seguem uma distribuição normal

H_a: os resíduos experimentais não seguem uma distribuição normal

Como 0,1450 < 0,220 não devemos rejeitar H_0. Portanto, conclui-se que os resíduos experimentais segundo o modelo estatístico adotado não diferem de uma distribuição normal. As distribuições, empírica e teórica, são apresentadas na Figura 3.

Figura 3 – Distribuições empírica e teórica obtida para o Exemplo 4.1

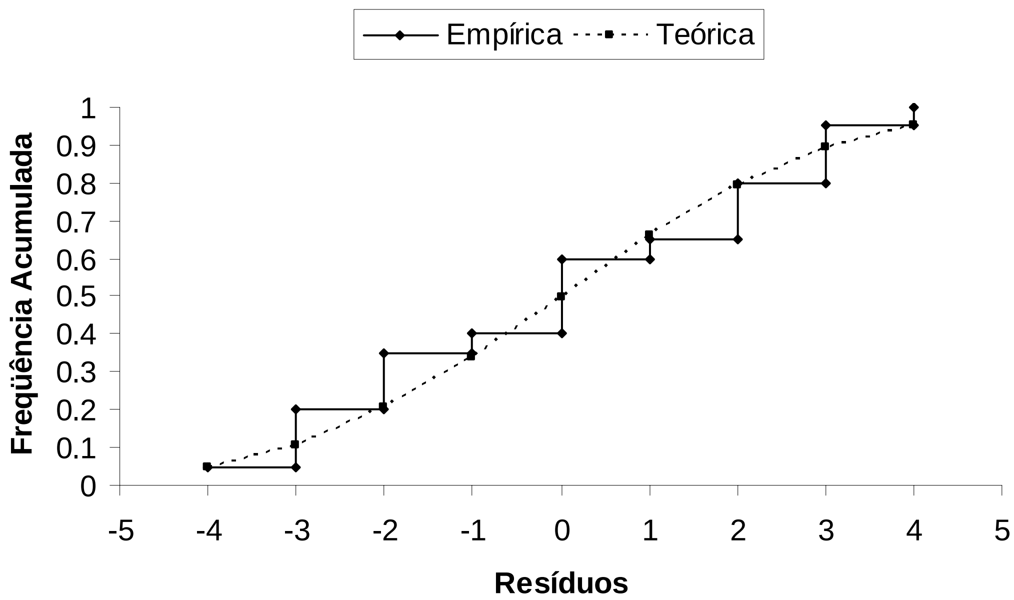


Tabela 3 – Valores críticos (d_c) para o teste de Lilliefors (adaptado de Barbetta et al.2004)

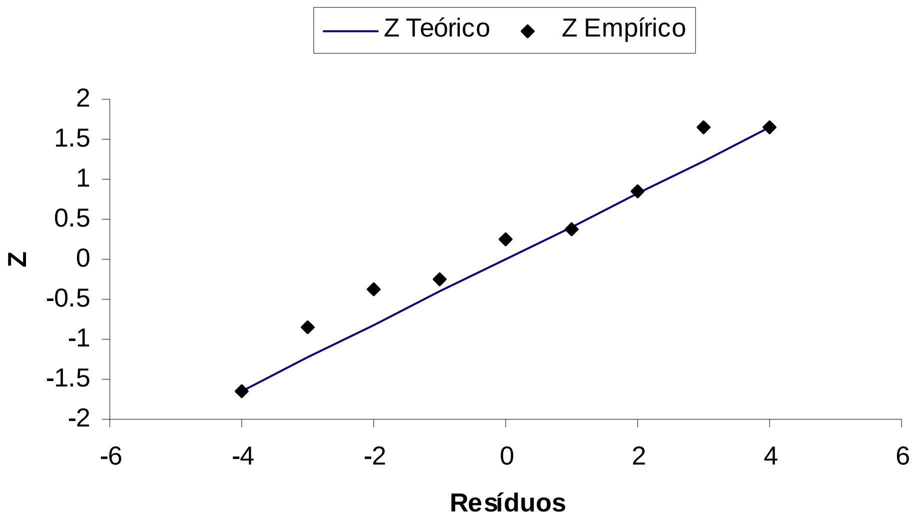
n	α=5%	α=1%
4	0,381	0,734
5	0,337	0,405
6	0,319	0,364
7	0,300	0,348
8	0,285	0,331
9	0,271	0,311
10	0,258	0,294
11	0,249	0,284
12	0,242	0,275
13	0,234	0,268
14	0,227	0,261
15	0,220	0,257
16	0,213	0,250
17	0,206	0,245
18	0,200	0,239
19	0,179	0,235
20	0,190	0,231
25	0,173	0,200
30	0,161	0,187
n>30	$d_c = \frac{0{,}886}{\sqrt{n}}$	$d_c = \frac{1{,}031}{\sqrt{n}}$

Uma avaliação visual da distribuição normal também pode ser realizada por meio do gráfico da probabilidade normal. Neste gráfico são plotados os valores da variável normal correspondente as distribuições acumulada empírica e acumulada teórica. Os valores da distribuição teórica ajustam-se perfeitamente a uma reta. Caso os resíduos apresentarem distribuição normal, os valores da distribuição empírica tenderam a se concentrar em torno da reta. Para os dados do Exemplo 4.1, os valores da variável z, correspondentes aos valores das distribuições empírica e teórica, são apresentados na Tabela 4 e o gráfico da probabilidade normal é apresentado na Figura 4.

Tabela 4 – Valores do resíduo, respectivas freqüências acumuladas teórica e empírica e valores da distribuição normal (Z)

	Freqüência Acumulada		Z	
Resíduo	Empírica	Teórica	Empírico	Teórico
-4	0.05	0.0497	-1.64485	-1.64751
-3	0.20	0.1083	-0.84162	-1.23563
-2	0.35	0.2050	-0.38532	-0.82375
-1	0.40	0.3402	-0.25335	-0.41188
0	0.60	0.5000	0.253347	-1.4E-16
1	0.65	0.6598	0.38532	0.411877
2	0.80	0.7950	0.841621	0.823754
3	0.95	0.8917	1.644854	1.235632
4	1.00	0.9503	1.644854	1.647509

Figura 4 – Gráfico de probabilidade normal para os dados do Exemplo 4.1



2ª Pressuposição) Homogeneidade das variâncias residuais

Para uma variável resposta Y, considere I tratamentos, cada um com K repetições, para os quais se deseja avaliar se a variância residual é idêntica para todos os tratamentos. As hipóteses a serem testadas são

H_0: $\sigma^2_{E1} = \sigma^2_{E2} = ... = \sigma^2_{EI} = \sigma^2_{E}$

H_a: pelo menos um tratamento apresenta variância residual diferente dos demais.

Em termos práticos estamos querendo verificar se o efeito do erro experimental afetou igualmente todos os tratamentos. Caso isto ocorra, as variâncias dentro de tratamentos tenderam a apresentar valores bem similares, sendo, portanto, viável a obtenção de um estimador comum para a variância dentro de tratamentos.

Na análise de variância, o cálculo do quadrado médio do resíduo é o estimador comum da variância dentro de tratamentos. Portanto, antes de interpretar os resultados da análise de variância faz-se necessário realizar um teste de hipóteses para a homogeneidade da variância dentro de tratamentos.

Um dos testes que podem ser utilizados é o teste de Cochran. Este teste só pode ser aplicado quando o número de graus de liberdade for o mesmo para todas as variâncias, ou seja, quando o número de repetições por tratamento for o mesmo.

A estatística do teste de Cochran é definida como

$$C_{cal} = \frac{\text{maior } s^2_{Ei}}{\sum_{i=1}^{I} s^2_{Ei}}.$$

Se $C_{cal} \geq C_{tab}$ (α, I, K–1), rejeita-se H_0. Caso contrário, se $C_{cal} < C_{tab}$, não se rejeita H_0 e conclui-se que existe homogeneidade de variâncias residuais entre os tratamentos.

Para os dados do Exemplo 4.1 as variâncias dentro de tratamento são apresentadas na Tabela 5.

Tabela 5 – Valores originais e ajustados de Y e estimativas dos efeitos do erro experimental

Trat	Rep	Y_{ik}	$\hat{Y}_{ik}$	$\hat{e}_{ik}$	s^2_{Ei}
1	1	25	23	2	
1	2	26	23	3	
1	3	20	23	-3	s^2_{E1} = 6,5
1	4	23	23	0	
1	5	21	23	-2	
2	1	31	27	4	
2	2	25	27	-2	
2	3	28	27	1	s^2_{E2} = 7,5
2	4	27	27	0	
2	5	24	27	-3	
3	1	22	26	-4	
3	2	26	26	0	
3	3	28	26	2	s^2_{E3} = 7,5
3	4	25	26	-1	
3	5	29	26	3	
4	1	33	31	2	
4	2	29	31	-2	
4	3	31	31	0	s^2_{E4} = 6,5
4	4	34	31	3	
4	5	28	31	-3	

As hipóteses testadas na pressuposição de homogeneidade de variâncias são iguais a:

H_0: $\sigma^2_{E1} = \sigma^2_{E2} = \sigma^2_{E3} = \sigma^2_{E4} = \sigma^2_{E}$;

H_a: pelo menos uma σ^2_{Ei} (i = 1, 2, 3 e 4) difere das demais.

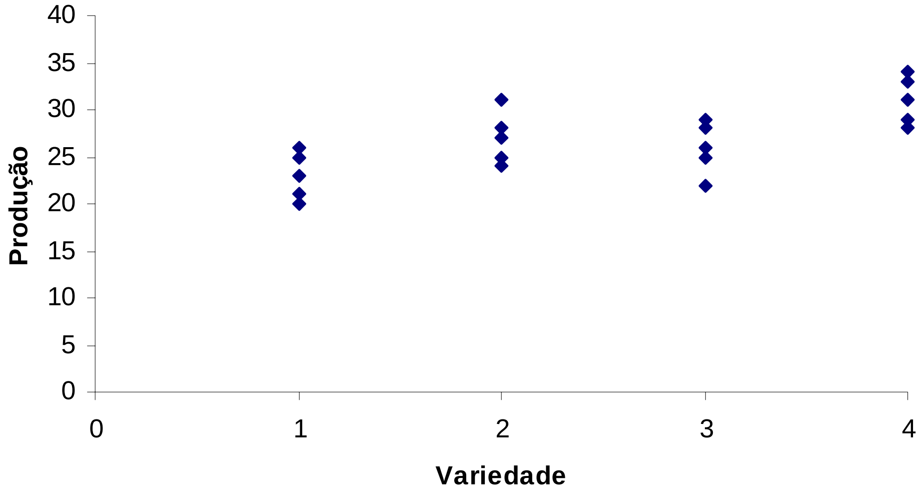
O valor da estatística de Cochran, para os dados deste exemplo, é obtido por

$$C_{cal} = \frac{7,5}{6,5 + 7,5 + 7,5 + 6,5} = 0,2679 \text{ e } C_{tab}\ (5\%, 4, 4) = 0,6287;$$

Como $C_{cal} < C_{tab}$ não rejeita-se H_0. Portanto, considera-se satisfeita a pressuposição de homogeneidade de variâncias.

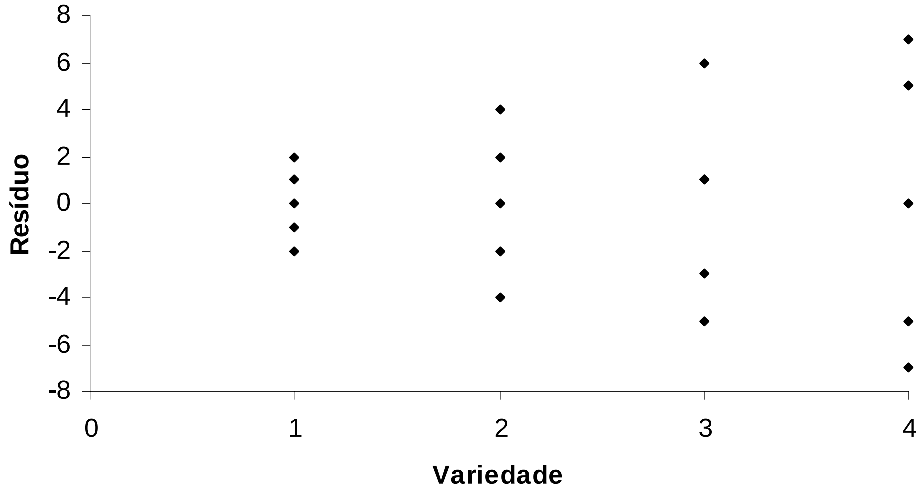
A análise gráfica da homogeneidade de variâncias pode ser feita por meio da dispersão dos valores observados para cada nível do fator em estudo. Para o exemplo em estudo este gráfico de dispersão é apresentado na Figura 5. Pode ser observado que a variabilidade da produção dentro de cada variedade, tende a ser a mesma em todas as variedades.

Figura 5 – Dispersão das produções observadas em cada variedade



Um exemplo em que visualmente poderíamos ter um indicativo de que a variância não é a mesma para todos os tratamentos é apresentado na Figura 6.

Figura 6 - Exemplo de gráfico de dispersão quando as variância dentro de tratamento não é homogênea



3ª Pressuposição) Independência dos erros

A independência dos erros da análise de variância significa que os erros não são correlacionados. Uma das situações que podem fazer com que este resultado não aconteça é aquela em que o valor do erro tende diminuir na seqüência cronológica em que os valores são observados.

Isto pode ocorrer quando, por exemplo, um laboratorista está aprendendo a usar um equipamento. No início, o erro associado a leitura é grande. À medida que são feitas novas leituras o erro tende a ser menor.

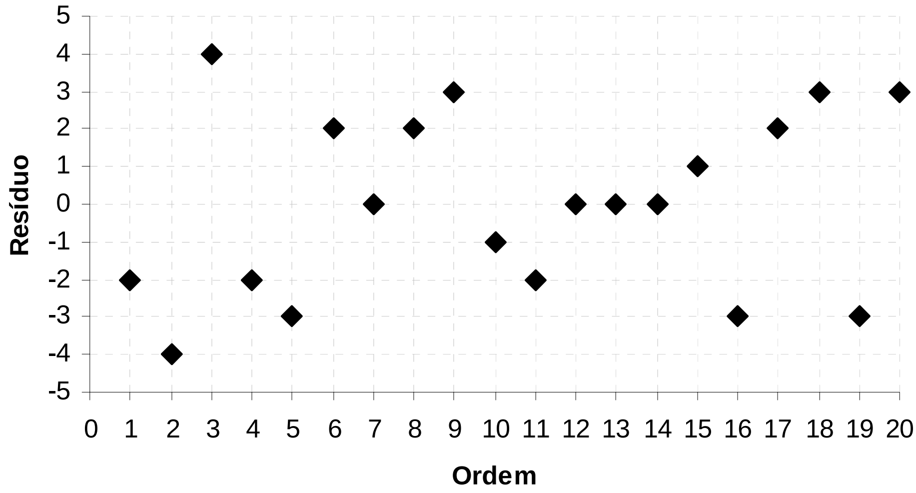
Portanto, para fazer a avaliação da independência dos erros é necessário ter informações adicionais, por exemplo ordem de coleta das observações.

A ordem de coleta das observações dos dados do Exemplo 4.1 é apresentada na Tabela 6 e o gráfico de dispersão dos resíduos versus a ordem de coleta é apresentada na Figura 7. Pode-se observar na Figura 7 que não existe nenhuma tendência nos resíduos em relação a ordem de coleta.

Tabela 6 – Valores observados com os respectivos valores preditos, residuais e ordem de coleta

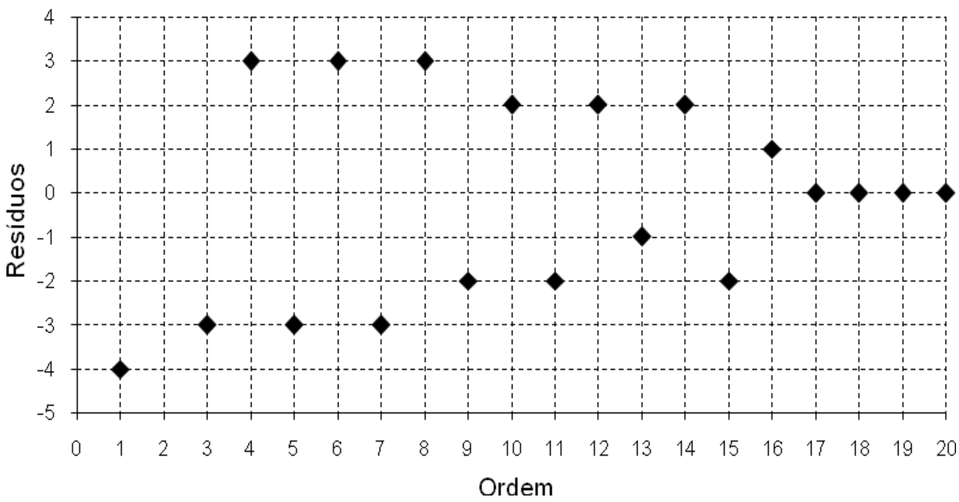
Ordem de coleta	Variedade	Repetição	Y_{ik}	$\hat{Y}_{ik}$	$\hat{e}_{ik}$
1	1	1	25	23	-2
5	1	2	26	23	-3
9	1	3	20	23	3
13	1	4	23	23	0
17	1	5	21	23	2
2	2	1	31	27	-4
6	2	2	25	27	2
10	2	3	28	27	-1
14	2	4	27	27	0
18	2	5	24	27	3
3	3	1	22	26	4
7	3	2	26	26	0
11	3	3	28	26	-2
15	3	4	25	26	1
19	3	5	29	26	-3
4	4	1	33	31	-2
8	4	2	29	31	2
12	4	3	31	31	0
16	4	4	34	31	-3
20	4	5	28	31	3

Figura 7 – Gráfico de dispersão dos resíduos versus a ordem de coleta das observações



A Figura 8 apresenta o gráfico de dispersão em que os erros não são independentes. Nesta Figura 8 pode-se observar que nas primeiras coletas, os valores residuais tendem a ser maiores do que nas últimas coletas. Uma possível explicação para isto é o aprendizado na realização do experimento.

Figura 8 – Dispersão dos resíduos em função da ordem de coleta



Capítulo 4 – Delineamento Inteiramente Casualizado

Exercícios extras

1) Considere que para o Exercício 4.7 é fornecida a ordem de coleta dos valores de ganho de peso, conforme tabela abaixo

Ração	Repetição	Ganho de Peso	Ordem
A	1	7.1	2
A	2	8.9	1
A	3	6.0	4
A	4	7.0	3
B	1	6.2	7
B	2	8.8	5
B	3	4.9	6
B	4	6.1	8
C	1	6.0	12
C	2	5.0	11
C	3	9.1	10
C	4	3.9	9
D	1	11.1	16
D	2	10.8	13
D	3	10.2	15
D	4	11.9	14
E	1	7.0	19
E	2	11.3	20
E	3	10.0	17
E	4	11.7	18

Com base nestas informações, pede-se

a) aplicar o teste de Liliefors;

b) aplicar o teste de Cochran;

c) avaliar a independência dos erros;

d) traçar o gráfico de probabilidade normal;

e) traçar o gráfico para avaliar a homogeneidade de variâncias.

2) Considere que para um experimento em que foram avaliados 5 tratamentos com 4 repetições no DIC sejam fornecidas as seguintes informações

Tratamento	Repetição	Y_{ik}	Ordem
1	1	12.1	20
1	2	2.0	2
1	3	3.0	4
1	4	1.0	1
2	1	11.2	9
2	2	11.3	10
2	3	11.5	15
2	4	11.4	12
3	1	8.1	8
3	2	8.3	11
3	3	8.4	14
3	4	8.5	16
4	1	16.1	19
4	2	10.8	5
4	3	10.2	3
4	4	11.9	7
5	1	12.0	18
5	2	11.3	12
5	3	10.0	6
5	4	11.7	17

Com base nestas informações, pede-se

a) aplicar o teste de Liliefors;

b) aplicar o teste de Cochran;

c) avaliar a independência dos erros;

d) traçar o gráfico de probabilidade normal;

e) traçar o gráfico para avaliar a homogeneidade de variâncias.

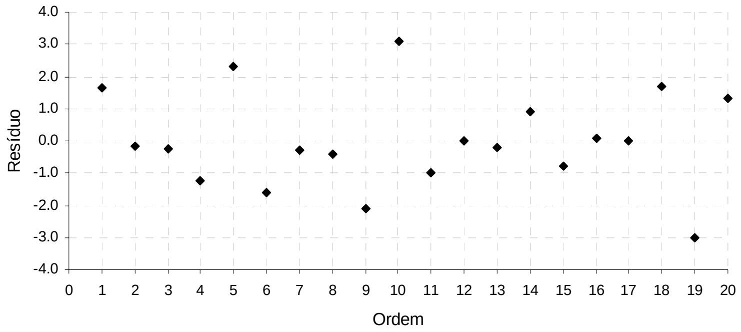
Respostas (parciais)

1)

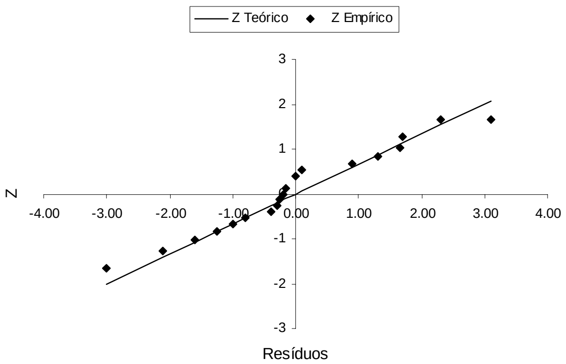
a) d = 0.17337

b) C = 0.33

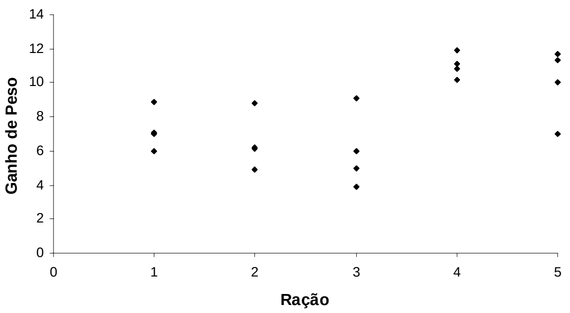
c)



d)



e)

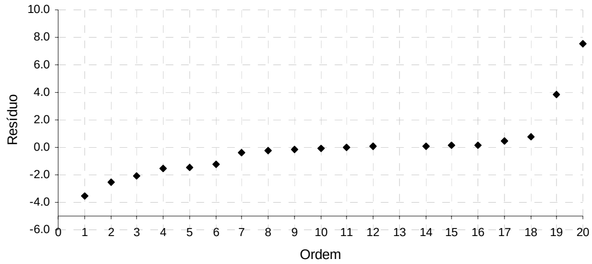


2)

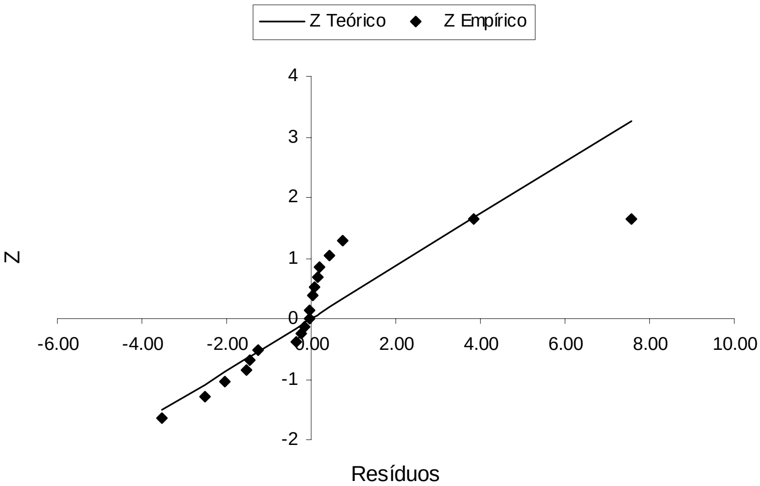
a) d = 0.27322

b) C = 0,97

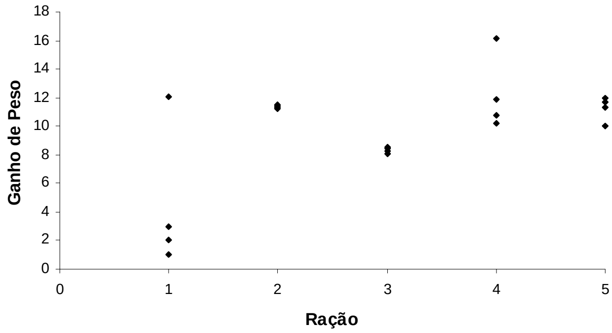
c)



d)



e)



Observação: o exercício abaixo não se refere as pressuposições da ANOVA, mas sim a uma transparência apresentada pelos professores em sala de aula.

Exercício: Considere 4 resultados possíveis (R1, R2, R3 e R4) para a realização de um experimento no DIC em que foram avaliados os efeitos de 5 tratamentos em 3 repetições. Para cada uma destas situações, pede-se:

a) Proceda a ANOVA para R1, R2, R3 e R4;
b) Para um ou mais dos resultados (R1, R2, R3 e R4) a SQ para uma ou mais FV apresentou valor zero. Explique a razão de ter sido obtido tais valores iguais a zero.

R1

	Tratamentos				
Repetições	1	2	3	4	5
1	100	100	100	100	100
2	100	100	100	100	100
3	100	100	100	100	100
Totais	300	300	300	300	300
Médias	100	100	100	100	100

R2

	Tratamentos				
Repetições	1	2	3	4	5
1	90	80	70	60	50
2	100	100	100	100	100
3	110	120	130	140	150
Totais	300	300	300	300	300
Médias	100	100	100	100	100

R3

	Tratamentos				
Repetições	1	2	3	4	5
1	98	99	100	101	102
2	98	99	100	101	102
3	98	99	100	101	102
Totais	294	297	300	303	306
Médias	98	99	100	101	102

R4

	Tratamentos				
Repetições	1	2	3	4	5
1	98	99	100	101	102
2	99	100	101	102	103
3	100	101	102	103	104
Totais	297	300	303	306	309
Médias	99	100	101	102	103