



UNIVERSIDADE ESTADUAL DA PARAÍBA  
CENTRO DE CIÊNCIAS E TECNOLOGIA  
DEPARTAMENTO DE ESTATÍSTICA

Arnete Matias Campos

# **Análise estatística de um experimento aninhado com três estágios: uma aplicação em dados de algodão**

Campina Grande - PB

Agosto de 2017

Arnete Matias Campos

**Análise estatística de um experimento aninhado com três  
estágios: uma aplicação em dados de algodão**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientador: Profa. Dra. Ana Patricia Bastos Peixoto

Campina Grande - PB

Agosto de 2017

É expressamente proibida a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano da dissertação.

C198a Campos, Arnete Matias.

Análise estatística de um experimento aninhado com três estágios [manuscrito] : uma aplicação em dados de algodão / Arnete Matias Campos. - 2017.  
35 p. : il. color.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2017.

"Orientação: Profa. Dra. Ana Patricia bastos Peixoto, Departamento de Ciências e Tecnologias".

1. Experimento hierárquico. 2. Modelo estatístico. 3. Componentes de variância. I. Título.

21. ed. CDD 519.5

Arnete Matias Campos

## **Análise estatística de um experimento aninhado com três estágios: uma aplicação em dados de algodão**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Trabalho aprovado em 08 de Agosto de 2017.

### **BANCA EXAMINADORA**



Profa. Dra. Ana Patricia Bastos Peixoto  
Universidade Estadual da Paraíba



Prof. Dr. Ricardo Alves de Oliveira  
Universidade Estadual da Paraíba



Prof. Esp. Edmundo Barbosa de Mendonça  
Universidade Federal de Campina Grande

*Aos meus pais Antônio e Ramira,  
as minhas irmãs Arlete, Maria José e Rosângela,  
dedico com muito amor e carinho.*

# Agradecimentos

A DEUS por me dar forças para enfrentar as dificuldades, por sempre me manter firme nos meus objetivos e também por me fazer persistir na realização deste sonho.

Aos meus pais Antônio e Ramira, as minhas irmãs Arlete, Maria José e Rosângela por todo carinho, amor e compreensão.

A Carlos Alberto Lucena por todo apoio e incentivo.

A orientadora, Dra. Ana Patricia por toda ajuda, atenção e empenho no desenvolvimento desse trabalho.

A Dr. Wellington dos Santos meu eterno agradecimento pelas explicações e ensinamentos que me ajudaram a entender e desenvolver os assuntos, pela amizade e por ser a pessoa maravilhosa sempre me tratando com carinho.

Aos professores, Dr. Ricardo Olinda e Ednário Barbosa por participarem da banca.

A todos os professores do departamento de estatística da UEPB em especial a Dr. Tiago Almeida de Oliveira e Dr. João Gil de Luna por me ajudarem na realização desse trabalho.

A Adrícia Mirely por me ajudar com informações e materiais e por toda a atenção me dada.

A todos os meus amigos que conviveram comigo no decorrer do curso, em especial a Luzidark Maciel e Klaini Clemente por terem me acolhido, em alguns momentos, em suas casas, meu muito obrigada!

Em fim, a todos que, de alguma forma, contribuíram para a realização deste trabalho.

*“O fato de uma opinião ser amplamente compartilhada não é nenhuma evidência de que não seja completamente absurda; de fato, tendo-se em vista a maioria da humanidade, é mais provável que uma opinião difundida seja tola do que sensata.”*

*(Bertrand Russell)*

# Resumo

Em certas ocasiões, as restrições nos impedem de cruzar todos os níveis de um fator com todos os níveis do outro fator. Assim sendo, será necessário utilizar-se de uma ferramenta conhecida como disposição hierarquizada. A análise de variância hierarquizada obtida é uma extensão da análise de variância, onde cada fator é dividido em subgrupos destes fatores. Exclusivamente, estes subgrupos são escolhidos aleatoriamente à partir de um conjunto maior de subgrupos possíveis. Para esse trabalho os dados utilizados foram obtidos da Embrapa Algodão, na safra 2014/2015, **onde foi inserido hipoteticamente o fator lote**, com o intuito de investigar se existe efeito de fornecedores, genótipos e lotes sobre o comprimento, percentagem e resistência da fibra do algodão. Os dados são balanceados de um experimento aninhado conhecido também como hierárquico. Foi utilizado o método da análise de variância para o modelo aleatório, e obtidos também as estimativas dos componentes de variância. Além disto, foram feitos estudos para as análises de resíduos. Observou-se que ao cruzar fatores como por exemplo genótipo e lote, a efetivação de tal cruzamento trás mudanças consideráveis para a proposta do experimento apresentado, já em relação a variável resistência da fibra **o fornecedor junto com genótipo e lote não tiveram efeito significativo para a variável.**

**Palavras-chave:** experimento hierárquico, fibra de algodão, componentes de variância.

# Abstract

On some occasions, the constraints prevent us from crossing all levels of one factor with all levels of the other factor. Therefore, it is necessary to use a tool knowledgeable of hierarchical disposition. A hierarchical analysis of variance obtained is an extension of the analysis of variance, where each factor is divided into subgroups these factors. Exclusively, these subgroups are randomly selected from a larger set of possible subgroups. For this work the data used were obtained from Embrapa Cotton, in the 2014/2015 harvest, **the factor lots was inserted hypothetically**, in order to investigate whether there is effect of suppliers, genotypes and lots on the length, percentage and resistance of the cotton fiber. The data is balanced from a nested experiment also known as hierarchical. The variance analysis method was used for the random model, and also obtained the estimates of the variance components. In addition, studies were carried out for residuos analysis. It was observed that when crossing factors such as genotype and batch, the accomplishment of such a crossing brings considerable changes to the proposal of the presented experiment, already in relation to the variable resistance of the fiber the supplier together with genotype and lot had not significant effects for the variable.

**Key-words:** herarchical experiment, cotton fiber, variance component.

# Lista de ilustrações

Figura 1 – Croqui de um experimento hierárquico com três estágios ( Figura adaptada de Montgomery (1984)) . . . . .	13
Figura 2 – Gráficos dos resíduos estudentizados versus valores ajustados para a variável resposta comprimento de fibra (a), percentagem de fibra (b) e resistência da fibra (c), respectivamente . . . . .	31
Figura 3 – Gráficos dos quantis normais para a variável resposta comprimento de fibra (a), percentagem de fibra (b) e resistência da fibra (c), respectivamente . . . . .	31
Figura 4 – Gráficos envelope simulado para a variável resposta comprimento de fibra (a), percentagem de fibra (b) e resistência da fibra (c), respectivamente	32

# Lista de tabelas

Tabela 1 – Parâmetros e estimadores do modelo . . . . .	18
Tabela 2 – Análise de variância para experimentos aninhados em três estágios . . .	19
Tabela 3 – Esperança do quadrado médio de um modelo matemático com três fatores e efeitos aleatórios. . . . .	20
Tabela 4 – Razão-F para o teste da hipótese nula para o modelo aleatório de três fatores aninhados . . . . .	23
Tabela 5 – Valores, em mm, da variável resposta comprimento de fibra . . . . .	26
Tabela 6 – Valores, em %, da variável resposta percentagem de fibra . . . . .	27
Tabela 7 – Valores, em gf/tex, da variável resposta resistência da fibra . . . . .	28
Tabela 8 – Análise da variância para o modelo aleatório em três estágios, com as variáveis resposta comprimento de fibra (COMP), percentagem de fibra (PF) e resistência da fibra (RF). . . . .	29
Tabela 9 – Esperança dos quadrados mínimos, estimativas da esperança dos quadra- dos e esperança dos componentes de variância ( E.C.V.) para o modelo aleatório em três estágios, com as variáveis resposta comprimento de fibra, percentagem de fibra e resistência da fibra. . . . .	30
Tabela 10 – Estatística Shapiro Wilks (W) com os respectivos valores de p das variáveis comprimento da fibra, percentagem de fibra e resistência da fibra. . . . .	32

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>11</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>12</b>
<b>2.1</b>	<b>Experimento aninhado</b>	<b>12</b>
2.1.1	Experimento aninhado com três estágios	13
2.1.1.1	Modelo Estatístico	14
2.1.1.2	Estimação dos efeitos do modelo	15
2.1.1.3	Decomposição da variabilidade total	18
2.1.1.4	Componentes da variância no modelo aleatório	19
2.1.1.4.1	Estimação dos componentes de variância	20
2.1.1.4.2	Estimativa negativa de um componente da variância	21
2.1.1.5	Razão - F	21
2.1.1.6	Diagnóstico do modelo	23
<b>3</b>	<b>MATERIAL E MÉTODOS</b>	<b>25</b>
<b>3.1</b>	<b>Resultados e discussão</b>	<b>29</b>
<b>4</b>	<b>CONCLUSÃO</b>	<b>33</b>
	<b>REFERÊNCIAS</b>	<b>34</b>

# 1 Introdução

Os primeiros vestígios históricos do algodão foram encontrados entre quatro séculos e meio há cinco mil anos atrás na América e no Paquistão respectivamente. No Brasil não temos muitos relatos sobre a história do algodão, o que se sabe é que a fibra já era articulada pelos indígenas que as transformavam em fios e tecidos. À que é mais usada é a fibra branca, coletada do algodoeiro, é natural e bastante usada na indústria têxtil. Diversos estudos vem sendo desenvolvidos ao longo dos anos, destacando-se atualmente a produção do algodão colorido.

De acordo com Santos et al. (2008) a estatística experimental é de grande importância, estando em todas as áreas do conhecimento humano como uma ferramenta para auxiliar as decisões a serem tomadas, possibilitando maior segurança e mais precisão a pesquisa. Ronald A. Fisher desenvolveu a análise de variância, a qual hoje é aplicada nos experimentos aleatórios. Técnicas como a dos componentes de variância, cujos parâmetros correspondentes às variâncias dos efeitos aleatórios de um modelo estatístico são largamente empregadas nos melhoramentos Genéticos, Animal e Vegetal, no ambiente de pesquisa em Ciências Agrárias.

Os componentes de variância têm um enorme uso na estatística, são parâmetros que correspondem as variâncias dos efeitos aleatórios de um modelo estatístico. Herbach (1959), visando à obtenção de estimativas não negativas, aplicou o princípio da máxima verossimilhança (ML) em modelos de dados balanceados; entretanto, suas estimativas eram tendenciosas. Thompson (1962) considerou a ideia de restringir a maximização da verossimilhança somente aos efeitos aleatórios do modelo, descrevendo, assim, o estimador REML. Harville (1969) projetou o estimador MIVQUE, que possui variância mínima sob normalidade. Segundo Barbin (1993), para obter os componentes de variância, destaca-se o método ANOVA para modelos balanceados. Apresenta, também, um método prático, o Método de Hicks, para obtenção dos quadrados médios nos modelos balanceados.

O presente trabalho tem como objetivo ajustar o modelo utilizado na análise de dados resultante de um planejamento experimental aninhado balanceado com três fatores aplicados a dados de algodão, dando ênfase nas estimativas de componentes de variância e realizando uma análise de diagnóstico para verificar a adequação do modelo, tal como identificar pontos discrepantes.

## 2 Fundamentação Teórica

O conteúdo dessa seção descreve os principais aspectos referentes a utilização da estatística experimental, sobretudo em experimentos aninhados com dados balanceados, utilizando-se de artigos, teses e dissertações, extraíndo os aspectos mais relevantes que permitam obter o objetivo desejado sobre o assunto.

### 2.1 Experimento aninhado

Um experimento aninhado ou hierárquico (*Nested design*) é utilizado quando se estabelece nos fatores um grau de hierarquia. Nesses experimentos comparam-se fatores (tratamentos secundários) que variam “dentro” de outros fatores (tratamentos principais). Os tratamentos secundários não se repetem para os diversos tratamentos principais. Dentro de cada tratamento principal há uma ramificação em tratamentos secundários específicos daquele tratamento principal. Como exemplo, pode-se ter certos experimentos fatoriais, em que um fator (fator B) são similares, mas não idênticos, para os diferentes níveis do outro fator (fator A). Esse arranjo para os níveis dos fatores de tratamentos é conhecido como delineamentos hierárquicos (ou aninhados). Assim sendo, diz-se que os níveis do fator B são hierárquicos (ou estão aninhados) em relação aos níveis do fator A (MONTGOMERY, 1984).

Segundo Silva e Zonta (1991), ao contrário da relação de cruzamento dos fatoriais  $A \times B$  a relação hierárquica não é recíproca, isto é, B dentro de A é diferente de A dentro de B. Por esse motivo, nos delineamentos de classificação hierárquica, não é possível avaliar interação entre os fatores, neste caso avalia-se o efeito médio dos níveis do fator B, dentro dos níveis do fator A. Em melhoramento bovino, por exemplo, analisa-se o resultado de cruzamentos de touros (fator A) com diversas vacas (fator B) através do teste de progênie. Esse tipo de experimento tem como finalidade a seleção de reprodutores para o rebanho. Embora sejam procedentes do mesmo rebanho, as vacas cruzadas com um touro, diferem daquelas cruzadas com outro touro. As ideias básicas relacionadas a esses experimentos podem ser percebidas considerando-se apenas dois estágios, porém pode-se estender para um número maior de fatores.

Esse trabalho teve como foco analisar um experimento hierárquico com efeitos aleatórios, porém, achou-se importante uma discussão o qual caracteriza um pouco do que são efeitos fixos e aleatórios. Diz-se que o modelo é de efeitos fixos quando os tratamentos são escolhidos “fixados” pelo pesquisador e que o modelo é de efeitos aleatórios quando os tratamentos são sorteados aleatoriamente de uma população de níveis de tratamento. Num modelo de efeitos aleatórios não analisamos as comparações múltiplas, devido à natureza

aleatória dos grupos.

### 2.1.1 Experimento aninhado com três estágios

Um experimento aninhado com três estágios segue o mesmo raciocínio que o delineamento fatorial hierárquico com 2 fatores, com uma diferença que tem um fator  $C$  a mais, e esse fator  $C$  está dentro de um outro fator  $B$ , que por sua vez, está dentro de um fator  $A$ , ou seja,  $A(B(C))$ . Neste caso estudam-se os efeitos médios dos níveis de  $C$  dentro de  $B$  e  $A$ , e os níveis de  $B$  dentro de  $A$  (MONTGOMERY, 1984).

Como exemplo, temos o caso em que desejamos investigar a dureza de duas diferentes formulação de liga, com três níveis do fator calor em cada liga e duas barras de metal fundido em cada fator calor, são selecionada aleatoriamente dentro de cada nível do fator calor testado, e duas medidas de dureza são medida em cada barra, (delineamento fatorial hierárquico em 3 estágios com 2 repetições), sendo assim têm-se como fator principal (fator  $A$ ) liga, como fatores secundários (fator  $B$ ) calor e como fatores terciários (fator  $C$ ) barra, como pode ser ilustrado na Figura 1.

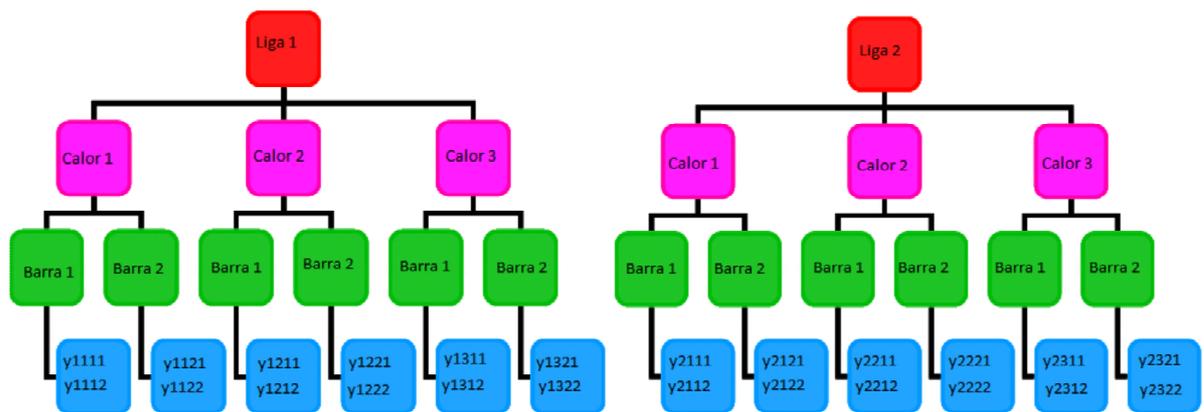


Figura 1 – Croqui de um experimento hierárquico com três estágios ( Figura adaptada de Montgomery (1984))

Percebe-se na Figura 1 que os diversos níveis de barras estão posicionados em um nível inferior à posição dos níveis do fator calor que estão posicionados ao um nível inferior a liga. Porém, se o leitor verificasse apenas a representação do croqui, poderia ter dúvida quanto ao tipo de experimento, podendo chegar a conclusão que seria um experimento fatorial. No entanto, quando um experimento é fatorial os níveis de diferentes fatores são combinados e é impossível determinar se existe interação entre os níveis de diferentes fatores. Deste modo, tem-se um experimento aninhado (hierárquico) porque as barras de cada nível do fator calor são únicas para cada níveis do fator calor que consequentemente são únicos para cada liga em particular. Ou seja, a barra 1 do calor 1 da liga 1 não tem conexão com a barra 1 de qualquer outro calor que não tem conexão com nenhuma

outra liga e assim por diante. Sendo possível observar no esquema da Figura 1 que as barras estão aninhados sob calor que estão aninhados sob liga. Caso exista dúvida se o experimento é realmente aninhado (hierárquico), tenta-se renumerar arbitrariamente, caso consiga dizemos que o fator é aninhado.

### 2.1.1.1 Modelo Estatístico

Para um experimento aninhado com três estágios, tem-se o modelo estatístico,

$$y_{ijkl} = \mu + \tau_i + \beta_{j(i)} + \gamma_{k(ij)} + \varepsilon_{(ijk)l} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, c \\ l = 1, 2, \dots, n \end{cases} \quad (2.1)$$

em que,  $y_{ijkl}$  é a  $l$ -ésima observação do  $k$ -ésimo nível do fator  $C$  dentro do  $j$ -ésimo nível do fator  $B$  (e do  $i$ -ésimo nível do fator  $A$ );  $\mu$  é a média geral;  $\tau_i$  é o efeito aleatório do  $i$ -ésimo nível do fator  $A$ ;  $\beta_{j(i)}$  é o efeito aleatório do  $j$ -ésimo nível do fator  $B$  aninhado sob o  $i$ -ésimo nível do fator  $A$ ;  $\gamma_{k(ij)}$  é o efeito aleatório do  $k$ -ésimo nível do fator  $C$  aninhado sob o  $j$ -ésimo nível do fator  $B$  (e do  $i$ -ésimo nível do fator  $A$ );  $\varepsilon_{(ijk)m}$  é o erro aleatório associado a cada observação, aninhado sob os níveis  $A$ ,  $B$  e  $C$ .

As suposições acerca do modelo nas quais  $A$ ,  $B$  e  $C$  aleatórios são que o erro tem distribuição Normal com média 0 e variância  $\sigma_\varepsilon^2$ , além disso, temos que os erros são mutuamente independentes. Com isso, obtemos,  $\varepsilon_{k(ij)} \sim N(0, \sigma_\varepsilon^2)$ . Agora, para o efeito  $\tau_i$ , assumimos que tem distribuição Normal com média zero e variância  $\sigma_\tau^2$ , e temos que os efeitos são mutuamente independentes. Assim,  $\tau_i \sim N(0, \sigma_\tau^2)$ . Temos também, para o efeito  $\beta_{j(i)}$ , assumimos que tem distribuição Normal com média zero e variância  $\sigma_{\beta(\tau)}^2$ , e temos que os efeitos são mutuamente independentes. Assim,  $\beta_{j(i)} \sim N(0, \sigma_{\beta(\tau)}^2)$ . E por fim, para o efeito  $\gamma_{k(ij)}$ , assumimos que tem distribuição Normal com média zero e variância  $\sigma_{\gamma(\tau\beta)}^2$ , e temos que os efeitos são mutuamente independentes. Assim,  $\gamma_{k(ij)} \sim N(0, \sigma_{\gamma(\tau\beta)}^2)$ .

As hipóteses há serem testadas são:

$$\begin{cases} H_0 : \sigma_\tau^2 = 0 \\ H_1 : \sigma_\tau^2 \neq 0 \end{cases}$$

$$\begin{cases} H_0 : \sigma_{\beta(\tau)}^2 = 0 \\ H_1 : \sigma_{\beta(\tau)}^2 \neq 0 \end{cases}$$

$$\begin{cases} H_0 : \sigma_{\gamma(\tau\beta)}^2 = 0 \\ H_1 : \sigma_{\gamma(\tau\beta)}^2 \neq 0 \end{cases}$$

## 2.1.1.2 Estimação dos efeitos do modelo

Para encontrar os estimadores dos parâmetros do modelo utilizaremos (2.1) o método dos mínimos quadrados, o qual consiste em encontrar os valores de  $\mu, \tau_i, \beta_{j(i)}, \gamma_{k(ij)}$  que minimizam a soma dos quadrados dos resíduos.

A partir do modelo, os resíduos poderão ser obtidos por:

$$\varepsilon_{(ijk)l} = y_{ijkl} - \mu - \tau_i - \beta_{j(i)} - \gamma_{k(ij)}$$

Segue que,

$$Z = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n \varepsilon_{(ijk)l}^2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n (y_{ijkl} - \mu - \tau_i - \beta_{j(i)} - \gamma_{k(ij)})^2$$

Para obter os valores de  $\mu, \tau_i, \beta_{j(i)}, \gamma_{k(ij)}$ , deriva-se parcialmente a função Z em relação a cada parâmetro, igualando a zero e em seguida explicitamos cada estimador, ou seja,

$$\frac{\partial Z}{\partial \mu} = 2 \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n (y_{ijkl} - \mu - \tau_i - \beta_{j(i)} - \gamma_{k(ij)})(-1) = 0 \quad (\text{I})$$

$$\frac{\partial Z}{\partial \tau_i} = 2 \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n (y_{ijkl} - \mu - \tau_i - \beta_{j(i)} - \gamma_{k(ij)})(-1) = 0 \quad (\text{II})$$

$$\frac{\partial Z}{\partial \beta_{j(i)}} = 2 \sum_{k=1}^c \sum_{l=1}^n (y_{ijkl} - \mu - \tau_i - \beta_{j(i)} - \gamma_{k(ij)})(-1) = 0 \quad (\text{III})$$

$$\frac{\partial Z}{\partial \gamma_{k(ij)}} = 2 \sum_{l=1}^n (y_{ijkl} - \mu - \tau_i - \beta_{j(i)} - \gamma_{k(ij)})(-1) = 0 \quad (\text{IV})$$

Resolvendo (I) tem-se que:

$$2 \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n (y_{ijkl} - \mu - \tau_i - \beta_{j(i)} - \gamma_{k(ij)})(-1) = 0$$

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n (y_{ijkl} - \mu - \tau_i - \beta_{j(i)} - \gamma_{k(ij)}) = 0$$

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n y_{ijkl} - abc n \hat{\mu} - bcn \sum_i \tau_i - cn \sum_i \sum_j \beta_{j(i)} - n \sum_i \sum_j \sum_k \gamma_{k(ij)} = 0$$

$$abc n \hat{\mu} + bcn \sum_i \tau_i + cn \sum_i \sum_j \beta_{j(i)} + n \sum_i \sum_j \sum_k \gamma_{k(ij)} = y \dots$$

Resolvendo (II) tem-se que:

$$2 \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n (y_{ijkl} - \mu - \tau_i - \beta_{j(i)} - \gamma_{k(ij)})(-1) = 0$$

$$a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n (y_{ijkl} - bcn \mu - bcn \tau_i - cn \sum_{j=1}^b \beta_{j(i)} - a \sum_{j=1}^b \sum_{k=1}^c \gamma_{k(ij)}) = 0$$

$$bcn\mu + bcn\tau_i + cn \sum_{j=1}^b \beta_{j(i)} + a \sum_{j=1}^b \sum_{k=1}^c \gamma_{k(ij)} = y_{i...}$$

Resolvendo (III) tem-se que:

$$\sum_{k=1}^c \sum_{l=1}^n (y_{ijkl} - \mu - \tau_i - \beta_{j(i)} - \gamma_{k(ij)}) = 0$$

$$ab \sum_{k=1}^c \sum_{l=1}^n y_{ijkl} - cn\mu - cn\tau_i - cn\beta_{j(i)} - ab \sum_{k=1}^c \gamma_{k(ij)} = 0$$

$$cn\hat{\mu} + cn\tau_i + cn\beta_{j(i)} + ab \sum_{k=1}^c \gamma_{k(ij)} = y_{ij..}$$

Resolvendo (IV) tem-se que:

$$\sum_{l=1}^n (y_{ijkl} - \mu - \tau_i - \beta_{j(i)} - \gamma_{k(ij)}) = 0$$

$$abc \sum_{l=1}^n (y_{ijkl} - n\hat{\mu} - n\tau_i - n\beta_{j(i)} - n\gamma_{k(ij)}) = 0$$

$$n\hat{\mu} + n\tau_i + n\beta_{j(i)} + n\gamma_{k(ij)} = y_{ijk.}$$

O sistema formado pelas equações I a IV é:

$$\left\{ \begin{array}{l} abc n \hat{\mu} + bcn \sum_{i=1}^a \tau_i + cn \sum_{i=1}^a \sum_{j=1}^b \beta_{j(i)} + n \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \gamma_{k(ij)} = y_{...} \\ bcn \hat{\mu} + bcn \hat{\tau}_i + cn \sum_{j=1}^b \beta_{j(i)} + a \sum_{j=1}^b \sum_{k=1}^c \gamma_{k(ij)} = y_{i...} \\ cn \hat{\mu} + cn \hat{\tau}_i + cn \hat{\beta}_{j(i)} + ab \sum_{k=1}^c \gamma_{k(ij)} = y_{ij..} \\ n \hat{\mu} + n \hat{\tau}_i + n \hat{\beta}_{j(i)} + n \hat{\gamma}_{k(ij)} = y_{ijk.} \end{array} \right. \quad (2.2)$$

O sistema de equações normais é inconsistente e para resolvê-lo é preciso impor as seguintes restrições:

$$\sum_{i=1}^a \tau_i = \sum_{i=1}^a \sum_{j=1}^b \beta_{j(i)} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \gamma_{k(ij)} = \sum_{j=1}^b \beta_{j(i)} = \sum_{j=1}^b \sum_{k=1}^c \gamma_{k(ij)} = \sum_{k=1}^c \gamma_{k(ij)} = 0$$

Assim, temos o sistema:

$$\left\{ \begin{array}{l} abc n \hat{\mu} = y_{...} \\ bcn \hat{\mu} + bcn \hat{\tau}_i = y_{i...} \\ cn \hat{\mu} + cn \hat{\tau}_i + cn \hat{\beta}_{j(i)} = y_{ij..} \\ n \hat{\mu} + n \hat{\tau}_i + n \hat{\beta}_{j(i)} + n \hat{\gamma}_{k(ij)} = y_{ijk.} \end{array} \right. \quad (2.3)$$

Resolvendo o sistema (2.3) obtém-se os estimadores de mínimos quadrados.

(1)  $abc n \hat{\mu} = y_{...}$

$$\hat{\mu} = \frac{y_{i\dots}}{abcn} = \bar{y}_{i\dots}$$

(2)

$$bcn\hat{\mu} + bcn\hat{\tau}_i = y_{i\dots}$$

$$bcn\hat{\tau}_i = y_{i\dots} - bcn\hat{\mu}$$

$$\hat{\tau}_i = \frac{y_{i\dots}}{bcn} - \hat{\mu}$$

$$\hat{\tau}_i = \bar{y}_{i\dots} - \bar{y}_{i\dots}$$

(3)

$$cn\hat{\mu} + cn\hat{\tau}_i + cn\hat{\beta}_{j(i)} = y_{ij..}$$

$$cn\hat{\beta}_{j(i)} = y_{ij..} - cn\hat{\mu} - cn\hat{\tau}_i$$

$$\hat{\beta}_{j(i)} = \frac{y_{ij..}}{KL} - \hat{\mu} - \hat{\tau}_i$$

$$\hat{\beta}_{j(i)} = \bar{y}_{ij..} - \bar{y}_{i\dots} - (\bar{y}_{i\dots} - \bar{y}_{i\dots})$$

$$\hat{\beta}_{j(i)} = \bar{y}_{ij..} - \bar{y}_{i\dots}$$

(4)

$$c\hat{\mu} + c\hat{\tau}_i + c\hat{\beta}_{j(i)} + c\hat{\gamma}_{k(ij)} = y_{ijk.}$$

$$c\hat{\gamma}_{k(ij)} = y_{ijk.} - c\hat{\mu} - c\hat{\tau}_i - c\hat{\beta}_{j(i)}$$

$$\hat{\gamma}_{k(ij)} = \frac{y_{ijk.}}{L} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_{j(i)}$$

$$\hat{\gamma}_{k(ij)} = \bar{y}_{ijk.} - \bar{y}_{i\dots} - (\bar{y}_{i\dots} - \bar{y}_{i\dots}) - (\bar{y}_{ij..} - \bar{y}_{i\dots})$$

$$\hat{\gamma}_{k(ij)} = \bar{y}_{ijk.} - \bar{y}_{ij..}$$

Logo, as estimativas de  $\hat{\mu}$ ,  $\hat{\tau}_i$ ,  $\hat{\beta}_{j(i)}$  e  $\hat{\gamma}_{k(ij)}$  são:

$$\hat{\mu} = \bar{y}_{i\dots}, \hat{\tau}_i = \bar{y}_{i\dots} - \bar{y}_{i\dots}, \hat{\beta}_{j(i)} = \bar{y}_{ij..} - \bar{y}_{i\dots}, \hat{\gamma}_{k(ij)} = \bar{y}_{ijk.} - \bar{y}_{ij..} \quad (2.4)$$

Os estimadores de  $\hat{y}_{ijkl}$  e  $\hat{\varepsilon}_{ijkl}$  são dados respectivamente por:

$$\hat{y}_{ijkl} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_{j(i)} + \hat{\gamma}_{k(ij)}$$

e

$$\hat{\varepsilon}_{ijkl} = y_{ijkl} - \hat{y}_{ijkl}$$

Substituindo os resultados encontrados em (2.4) obtém-se,

$$\hat{y}_{ijkl} = \bar{y}_{....} + (\bar{y}_{i...} - \bar{y}_{....}) + (\bar{y}_{ij..} - \bar{y}_{i...}) + (\bar{y}_{ijk.} - \bar{y}_{ij..})$$

$$\hat{y}_{ijkl} = \bar{y}_{ijk.}$$

Daí, tem-se que  $\bar{y}_{ijk.}$  é o estimador de mínimos quadrados de  $y_{ijk.}$ .

$$\hat{\varepsilon}_{ijkl} = y_{ijkl} - \bar{y}_{ijk.}$$

Tabela 1 – Parâmetros e estimadores do modelo

Parâmetro do modelo	Estimador
$\mu$	$m = \bar{y}_{....}$
$\tau_i$	$t_i = \bar{y}_{i...} - \bar{y}_{....}$
$\beta_{j(i)}$	$b_{j(i)} = \bar{y}_{ij..} - \bar{y}_{i...}$
$\gamma_{k(ij)}$	$c_{k(ij)} = \bar{y}_{ijk.} - \bar{y}_{ij..}$
$E(y_{ijkl}) = \mu + \tau_i + \beta_{j(i)} + \gamma_{k(ij)}$	$\hat{y}_{ijkl} = \bar{y}_{ijk.}$
$\varepsilon_{ijkl}$	$e_{ijkl} = y_{ijkl} - \bar{y}_{ijk.}$
$\sigma^2$	$S^2 = QM_{Res}$

Observação: O  $QM_{Res}$  é o estimador não viciado para  $\sigma^2$ , que é a variância do erro experimental  $\varepsilon_{ijkl}$ .

### 2.1.1.3 Decomposição da variabilidade total

A soma de quadrados total pode ser dividida em seus componentes aditivos, como pode-se observar abaixo

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n (y_{ijkl} - \bar{y}_{....})^2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n [(\bar{y}_{i...} - \bar{y}_{....}) + (\bar{y}_{ij..} - \bar{y}_{i...}) + (\bar{y}_{ijk.} - \bar{y}_{ij..}) + (y_{ijkl} - \bar{y}_{ijk.})]^2$$

Expandindo o lado direito temos:

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n (y_{ijkl} - \bar{y}_{....})^2 = bcn \sum_{i=1}^a (\bar{y}_{i...} - \bar{y}_{....})^2 + cn \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij..} - \bar{y}_{i...})^2 + n \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (\bar{y}_{ijk.} - \bar{y}_{ij..})^2 + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n (y_{ijkl} - \bar{y}_{ijk.})^2$$

Simbolicamente,

$$SS_T = SS_A + SS_{B(A)} + SS_{C(B(A))} + SS_E$$

Para obter a variação residual, calcula-se por diferença, ou seja,

$$SS_E = SS_T - SS_A - SS_{B(A)} - SS_{C(B(A))}$$

Resumindo esse procedimento em uma tabela de Análise de variância (Tabela 2), temos:

Tabela 2 – Análise de variância para experimentos aninhados em três estágios

Fonte de Variação	Graus de liberdade	Soma de quadrados	Quadrado médio
A	a-1	$bcn \sum_{i=1}^a (\bar{y}_{i...} - \bar{y}_{...})^2$	$QM_A$
B dentro de A	a(b-1)	$cn \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij..} - \bar{y}_{i...})^2$	$QM_{B(A)}$
C dentro de B	ab(c-1)	$n \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (\bar{y}_{ijk.} - \bar{y}_{ij..})^2$	$QM_{C(B)}$
Resíduos	abc(n-1)	$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n (y_{ijkl} - \bar{y}_{ijk.})^2$	$QM_{Res}$
Total	abcn-1	$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n (y_{ijkl} - \bar{y}_{...})^2$	

onde,

$$C = \frac{y_{...}^2}{abcn}$$

$$SQ_{total} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n (y_{ijkl}^2 - C)$$

$$SQ_A = \frac{1}{bcn} \sum_{i=1}^a (y_{i...}^2 - C)$$

$$SQ_{B(A)} = \frac{1}{cn} \sum_{i=1}^a \sum_{j=1}^b (y_{ij..}^2 - C) - SQ_A$$

$$SQ_{C(B(A))} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (y_{ijk.}^2 - C) - SQ_{B(A)} - SQ_A$$

$$SQ_{Res} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n y_{ijkl}^2 - \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c y_{ijk.}^2$$

$$QM_A = \frac{SQ_A}{(a-1)}, QM_{B(A)} = \frac{SQ_{B(A)}}{a(b-1)}, QM_{C(B(A))} = \frac{SQ_{C(B(A))}}{ab(c-1)}, QM_{Res} = \frac{SQ_{Res}}{abc(n-1)}$$

#### 2.1.1.4 Componentes da variância no modelo aleatório

De acordo com Barbin (1993), componentes de variância são variâncias associadas aos efeitos aleatórios de um modelo matemático. O efeito aleatório representa os tratamentos de uma amostra oriunda de uma determinada população, conhecido como modelo aleatório ou do tipo II. O modelo aleatório ou do tipo II para dados balanceados é mais usado na área de melhoramento, embora na prática sejam mais comum dados desbalanceados.

Para se obterem as estimativas de componentes de variância utiliza-se de vários métodos, como o dos método dos momentos, o da Máxima Verossimilhança, os Henderson (métodos I, II e III), o da Máxima Verossimilhança Restrita, os de Rao (MINQUE - norma mínima de orientação quadrática imparcial e MIVQUE - estimativa quadrática imparcial de variância mínima) e os de Bayes. Porém o mais usado é o método dos momentos que consiste em se obter a Esperança Matemática dos quadrados médios da análise da variância.

A forma analítica de encontrar as esperanças dos quadrados médios de um modelo matemático é dispendiosa e por vezes complicada. Hicks (1973) desenvolveu um algoritmo

para se encontrar a  $E(QM)$  que torna mais simples e menos dispendioso esta tarefa. Pelo método prático modificado, apresenta o seguinte algoritmo de Hicks:

1. Todos os índices referentes aos efeitos fixos recebem o algarismo 0 e os aleatórios recebem 1 (na coluna), correspondente ao efeito;
2. Todos os índices entre parênteses recebem 1 (na coluna) - hierárquico;
3. Completa-se o quadro com os índices (a,b,c etc) correspondentes, na coluna;
4. Cobrir a(s) coluna(s) correspondente(s) ao(s) índice(s) do(s) fator(es) para encontrar a  $E(QM)$ , multiplicando-se os efeitos na linha, considerando somente os fatores com o mesmo índice;

DICA: iniciar multiplicando pela última linha.

Os quadrados médios do modelo com efeitos aleatórios, considerando 3 estágios pode ser visualizada na Tabela 3.

Tabela 3 – Esperança do quadrado médio de um modelo matemático com três fatores e efeitos aleatórios.

	a	a	a	a		
Efeitos	I	J	K	L	QM	E(QM)
$A_i$	1	J	K	L	$Q_1$	$\sigma^2 + L\sigma_C^2 + KL\sigma_B^2 + JKLS\sigma_A^2$
$B_{j(i)}$	1	1	K	L	$Q_2$	$\sigma^2 + L\sigma_C^2 + KL\sigma_B^2$
$C_{k(ji)}$	1	1	1	L	$Q_3$	$\sigma^2 + L\sigma_C^2$
$e_{ijklm}$	1	1	1	1	$Q_4$	$\sigma^2$

#### 2.1.1.4.1 Estimação dos componentes de variância

De acordo com Montgomery (2001), num modelo de efeitos aleatório, usualmente é necessário estimar os componentes da variância. A variância da observação experimental  $y_{ijk}$  é chamada de variância total, denotada por  $\sigma_T^2 = \sigma_A^2 + \sigma_B^2 + \sigma_C^2 + \sigma^2$ . O procedimento para se estimar  $\sigma_A^2, \sigma_B^2, \sigma_C^2, \sigma^2$  é conhecido como método da análise de variância, o mesmo consiste em igualar os valores esperados dos quadrados médios a seus valores observados na análise da variância e resolver a equação em relação aos componentes da variância, isto é:

$$Q_1 = QM_A = \sigma^2 + L \sigma_C^2 + KL \sigma_B^2 + JKLS\sigma_A^2$$

$$Q_2 = QM_{B(A)} = \sigma^2 + L \sigma_C^2 + KL \sigma_B^2$$

$$Q_3 = QM_{C(B(A))} = \sigma^2 + L \sigma_C^2$$

$$Q_4 = QM_{Res} = \sigma^2$$

Daí, os estimadores dos componentes da variância são:

$$\begin{aligned}\hat{\sigma}^2 &= Q_4 = QM_{Res} \\ \hat{\sigma}_C^2 &= \frac{Q_3 - Q_4}{L} \\ \hat{\sigma}_B^2 &= \frac{Q_2 - Q_3}{KL} \\ \hat{\sigma}_A^2 &= \frac{Q_1 - Q_2}{JKL}\end{aligned}$$

Portanto, o estimador da variabilidade total é:

$$\hat{\sigma}_T^2 = \hat{\sigma}_A^2 + \hat{\sigma}_B^2 + \hat{\sigma}_C^2 + \hat{\sigma}^2$$

#### 2.1.1.4.2 Estimativa negativa de um componente da variância

De acordo com Costa (2012), o método da análise de variância às vezes produz uma estimativa negativa dos componentes da variância. Por definição os componentes da variância são não-negativos, no caso uma estimativa negativa seria inaceitável. Porém no caso que a estimativa é negativa, uma alternativa é aceitar e usá-la como evidência de que o verdadeiro valor do componente da variância é zero, assumindo-se que a variabilidade amostral ocasionou a uma estimativa negativa. Apesar de que isso tenha um apelo intuitivo, atrapalhará as propriedades estatísticas de outras estimativas. Outra alternativas seria reestimar o componente negativo da variância por método que sempre resulte em estimativas não-negativas ou considerar a estimativa negativa como evidência de que o modelo linear proposto é incorreto, o que requer um estudo do modelo e suas hipóteses para determinar um modelo mais apropriado.

Herbach (1959), visando à obtenção de estimativas não negativas, aplicou o princípio da máxima verossimilhança, ML, em modelos de dados balanceados; entretanto, suas estimativas eram tendenciosas. Thompson (1962) considerou a ideia de restringir a maximização da verossimilhança somente aos efeitos aleatórios do modelo, descrevendo, assim, o estimador REML. Harville (1969) projetou o estimador MIVQUE, que possui variância mínima sob normalidade. As estimativas negativas dos componentes de variâncias, obtidas pelo método ANOVA, podem ser convertidas em zero, conforme recomendado por Searle (1971) e seguido por Swallow e Monahan (1984).

#### 2.1.1.5 Razão - F

A razão F é utilizada para testar as hipótese nula em um experimento aninhado com três fatores para dados balanceados (QUINN; KEOUGH, 2002). De acordo com os resultados encontrados para as esperanças dos quadrados médios, observa-se que:

$$\text{i) Sob } H_0^{(\tau)} : \sigma_\tau^2 = 0,$$

O  $QM_A$  é um estimador não viciado para  $\sigma^2$  e que independentemente de  $H_0^{(\tau)}$  verifica-se que o  $QM_{B(A)}$  também é um estimador não viciado para  $\sigma^2$ . Logo, pode-se deduzir que uma estatística apropriada para contrastar as hipóteses  $H_0^{(\tau)} : \sigma_\tau^2 = 0$  vs  $H_1^{(\tau)} : \sigma_\tau^2 > 0$  é a que compara  $QM_A$  com  $QM_{B(A)}$ , isto é,  $F_A = \frac{QM_A}{QM_{B(A)}} \sim F_{[(a-1); a(b-1); \alpha]}$  a estatística  $F_A$  segue uma distribuição F com (a-1) graus de liberdade do numerador e a(b-1) graus de liberdade do denominador. Então, se  $H_0^{(\tau)} : \sigma_\tau^2 = 0$  for verdadeiro, tanto o numerador quanto o denominador de  $F_A$  são estimadores não viciados para  $\sigma^2$ , porém, se  $H_0^{(\tau)} : \sigma_\tau^2 = 0$  for falso, a esperança matemática do  $QM_A$  será maior que  $QM_{B(A)}$ . Deste modo, rejeita-se  $H_0^{(\tau)} : \sigma_\tau^2 = 0$ , quando o valor de  $F_A$  for maior que o valor da distribuição de F com (a-1) e [a(b-1)] graus de liberdade e nível de significância  $\alpha$ .

ii) Sob  $H_0^{(\beta)} : \sigma_\beta^2 = 0$

O  $QM_{B(A)}$  é um estimador não viciado para  $\sigma^2$  e independentemente de  $H_0^{(\beta)}$  verifica-se que o  $QM_{C(B(A))}$  também é um estimador não viciado para  $\sigma^2$ . Deste modo, a estatística apropriada para contrastar as hipóteses  $H_0^{(\beta)} : \sigma_\beta^2 = 0$  vs  $H_1^{(\beta)} : \sigma_\beta^2 > 0$  é a que compara  $QM_{B(A)}$  com  $QM_{C(B(A))}$ , isto é,  $F_{B(A)} = \frac{QM_{B(A)}}{QM_{C(B(A))}} \sim F_{[a(b-1); ab(c-1); \alpha]}$ , a estatística  $F_{B(A)}$  segue uma distribuição F com [a(b-1)] graus de liberdade do numerador e [ab(c-1)] graus de liberdade do denominador. Então, se  $H_0^{(\beta)} : \sigma_\beta^2 = 0$  for verdadeiro, tanto o numerador quanto o denominador de  $F_{B(A)}$  são estimadores não viciados para  $\sigma^2$ , porém, se  $H_0^{(\beta)} : \sigma_\beta^2 = 0$  for falso, a esperança matemática do  $QM_{B(A)}$  será maior que  $QM_{C(B(A))}$ . Deste modo, rejeita-se  $H_0^{(\beta)} : \sigma_\beta^2 = 0$ , quando o valor de  $F_{B(A)}$  for maior que o valor da distribuição de F com [a(b-1)] e [ab(c-1)] graus de liberdade e nível de significância  $\alpha$ .

iii) Sob  $H_0^{(\gamma)} : \sigma_\gamma^2 = 0$ , o  $QM_{C(B(A))}$  é um estimador não viciado para  $\sigma^2$  e independentemente de  $H_0^{(\gamma)}$  verifica-se que o  $QM_{Res}$  também é um estimador não viciado para  $\sigma^2$ . Deste modo, a estatística apropriada para contrastar as hipóteses  $H_0^{(\gamma)} : \sigma_\gamma^2 = 0$  vs  $H_1^{(\gamma)} : \sigma_\gamma^2 > 0$  é a que compara  $QM_{C(B(A))}$  com  $QM_{Res}$ , isto é,  $F_{C(B(A))} = \frac{QM_{C(B(A))}}{QM_{Res}} \sim F_{[ab(c-1); abc(n-1); \alpha]}$ , a estatística  $F_{C(B(A))}$  segue uma distribuição F com [ab(c-1)] graus de liberdade do numerador e [abc(n-1)] graus de liberdade do denominador. Então, se  $H_0^{(\gamma)} : \sigma_\gamma^2 = 0$  for verdadeiro, tanto o numerador quanto o denominador de  $F_{C(B(A))}$  são estimadores não viciados para  $\sigma^2$ , porém, se  $H_0^{(\gamma)} : \sigma_\gamma^2 = 0$  for falso, a esperança matemática do  $QM_{C(B(A))}$  será maior que  $QM_{Res}$ . Deste modo, rejeita-se  $H_0^{(\gamma)} : \sigma_\gamma^2 = 0$ , quando o valor de  $F_{C(B(A))}$  for maior que o valor da distribuição de F com [ab(c-1)] e [abc(n-1)] graus de liberdade e nível de significância  $\alpha$ . Na Tabela 4 tem-se um resumo destes testes.

Tabela 4 – Razão-F para o teste da hipótese nula para o modelo aleatório de três fatores aninhados

Fonte de variação	Razão-F (A,B e C aleatórios)
A	$\frac{QM_A}{QM_{B(A)}}$
B(A)	$\frac{QM_{B(A)}}{QM_{C(B(A))}}$
C(B(A))	$\frac{QM_{C(B(A))}}{QM_{Res}}$
Resíduo	$\sigma_\epsilon^2$

#### 2.1.1.6 Diagnóstico do modelo

A análise dos resíduos tem uma importância fundamental na variação da qualidade dos ajustes de modelos. Basicamente, essa análise fornece evidências sobre possíveis violações nas suposições do modelo, tais como a de normalidade, homocedasticidade, e quando for o caso ainda fornece indícios da falta de ajuste do modelo proposto. Além dos testes de significância e adequação já apresentados, a análise de resíduos vem completar o elenco de procedimentos que devem ser realizados após o ajuste de qualquer modelo.

Na análise de resíduos, verificam-se as pressuposições sob o modelo em estudo. Uma ou mais observações são ditas discrepantes (*outliers*) se seus resíduos são muito grandes (em valor absoluto) em relação aos demais (DRAPER; SMITH, 1998), (PRUDENTE, 2009). Já os pontos influentes são observações que, embora não apresentem resíduos grandes, podem alterar significativamente as estimativas dos parâmetros do modelo escolhido. Assim, para detectar a presença de pontos influentes ou de *outliers*, podem ser aplicadas técnicas com base nos resíduos ordinários e resíduos padronizados.

O resíduo para a  $i$ -ésima observação é obtido por meio da função  $r_i = y_i - \hat{y}_i$ , que mensura a diferença entre o valor observado e o valor ajustado, chamado de resíduo ordinário da variável resposta do modelo. Então pode-se afirmar que modelos com melhor ajuste deverão apresentar uma quantidade menor de resíduos e caso contrário modelos mal ajustados apresentarão uma quantidade maior de resíduos. De acordo com Cordeiro e Neto (2006), os resíduos ordinários não são muito informativos, pois não apresentam variância constante  $Var(r_i) = \sigma^2(1 - h_{ii})$ , pois depende dos valores de  $h_{ii}$ .

Para garantir a independência do numerador e denominador, faz-se a padronização dos resíduos, obtendo o chamado resíduo estudentizado, definido como

$$z_i^* = \frac{e_i}{\sqrt{\hat{\sigma}_i^2(1-h_{ii})}}, \quad i = 1, 2, \dots, n$$

Prudente (2009), afirmou que os resíduos projetados são melhores que os resíduos ordinários e padronizados, no que se refere a análise e diagnósticos dos modelos não lineares, pois os resíduos projetados superam os resíduos ordinários em diversos aspectos e

---

muitas das técnicas de diagnóstico utilizadas na regressão linear são, também, aplicáveis aos mesmos.

Um outra alternativa é o uso do gráfico meio-normal com envelope simulado (MORAL, 2013) é bastante útil para a verificação da qualidade do ajuste. Essa técnica, relativamente simples, para verificação da adequação do ajuste de um determinado modelo a um conjunto de observações, proposta por Atkinson (1985), consiste em se plotarem os valores absolutos ordenados de uma determinada medida de diagnóstico (diferentes tipos de resíduos, distância de Cook, medida de leverage etc.) versus as estatísticas de ordem esperadas da distribuição meio-normal.

### 3 Material e métodos

Os dados utilizados nesta aplicação foram obtidos de um experimento desenvolvido pela Empresa Brasileira de Pesquisa Agropecuária - Embrapa Algodão, localizada em Campina Grande - PB, na cidade de Apodi – RN, safra 2014/2015 (ALBUQUERQUE, 2016). Para efeito desse estudo foi inserido hipoteticamente o fator lote, cujo objetivo foi analisar se os efeitos de genótipos, fornecedores e lotes influenciam as características tecnológicas da fibra do algodão. O estudo foi conduzido por meio de um delineamento experimental inteiramente casualizado com classificação hierárquica balanceada em três estágios, com quatro fornecedores, quatro genótipos, três lotes e quatro repetições, onde foram estudadas três variáveis resposta: comprimento de fibra - COMP (mm), percentagem de fibra – PF (%) e resistência da fibra – RF (gf/tex).

O modelo adotado para este exemplo foi o modelo aleatório representado abaixo:

$$y_{ijkl} = \mu + \tau_i + \beta_{j(i)} + \gamma_{k(ij)} + \varepsilon_{(ijk)l} \begin{cases} i = 1, 2, 3, 4 \\ j = 1, 2, 3, 4 \\ k = 1, 2, 3 \\ l = 1, 2, 3, 4 \end{cases}$$

em que,  $y_{ijkl}$  é a  $l$ -ésima observação do  $k$ -ésimo nível do fator lote dentro do  $j$ -ésimo nível do fator genótipo (e do  $i$ -ésimo nível do fator fornecedor);  $\mu$  é a média geral;  $\tau_i$  é o efeito aleatório do  $i$ -ésimo nível do fator fornecedor;  $\beta_{j(i)}$  é o efeito aleatório do  $j$ -ésimo nível do fator genótipo aninhado sob o  $i$ -ésimo nível do fator fornecedor;  $\gamma_{k(ij)}$  é o efeito aleatório do  $k$ -ésimo nível do fator lote aninhado sob o  $j$ -ésimo nível do fator genótipo (e do  $i$ -ésimo nível do fator fornecedor);  $\varepsilon_{(ijk)m}$  é o erro aleatório associado a cada observação  $y_{ijkl}$ , aninhado sob os níveis fornecedor, genótipo e lote.

As hipóteses a serem testadas nesse trabalho são:

$$\begin{cases} H_0 : \text{Os efeitos hierárquico não influenciam as características da fibra;} \\ H_a : \text{Os efeitos hierárquico, pelo menos um, influenciam as características da fibra.} \end{cases}$$

Os resultados da análise foram obtidos por meio do software R (TEAM, 2017) e do software SAS (INC, 2008) versão gratuita. Os dados obtidos para análise encontram-se nas Tabelas 5, 6 e 7 a seguir.

Tabela 5 – Valores, em mm, da variável resposta comprimento de fibra

Fornecedor	Genótipo	Lotes	Repetições			
			1	2	3	4
FORN 1	Gen 1	1	27,90	28,50	27,29	30,70
		2	32,00	29,25	29,01	32,00
		3	29,00	28,70	26,35	32,98
	Gen 2	1	29,82	27,55	31,00	29,85
		2	31,03	28,55	34,26	33,48
		3	28,20	29,00	32,00	30,03
	Gen 3	1	30,88	31,23	29,99	30,97
		2	28,75	27,13	29,99	27,89
		3	28,43	32,03	34,00	29,49
	Gen 4	1	29,00	26,58	33,08	31,12
		2	27,96	29,74	33,80	31,25
		3	27,49	28,63	28,96	32,90
FORN 2	Gen 1	1	30,82	31,38	30,15	29,12
		2	28,77	29,43	30,18	30,24
		3	28,00	28,59	30,88	31,20
	Gen 2	1	31,71	32,45	28,22	29,07
		2	29,18	33,98	28,15	33,32
		3	34,25	28,62	33,24	27,65
	Gen 3	1	26,40	27,80	27,00	26,77
		2	34,45	32,98	27,76	26,35
		3	26,00	26,54	27,08	27,83
	Gen 4	1	34,56	33,65	29,48	29,45
		2	29,71	28,37	28,91	28,00
		3	29,81	26,70	27,00	27,67
FORN 3	Gen 1	1	32,30	33,14	34,12	35,00
		2	27,88	28,60	29,16	29,00
		3	27,00	34,82	25,93	26,98
	Gen 2	1	25,12	29,83	29,84	29,71
		2	28,21	28,22	27,63	34,04
		3	35,09	35,00	33,18	33,01
	Gen 3	1	32,09	32,01	28,78	28,98
		2	28,84	34,00	35,00	27,66
		3	26,46	26,00	25,42	27,00
	Gen 4	1	28,91	34,00	35,06	33,34
		2	33,08	33,18	32,20	28,30
		3	29,27	28,37	33,47	32,15
FORN 4	Gen 1	1	27,38	27,62	35,17	35,06
		2	27,13	28,19	28,17	28,78
		3	29,87	29,46	32,14	32,43
	Gen 2	1	32,98	27,00	27,87	27,66
		2	28,10	28,00	28,99	29,79
		3	30,86	34,12	29,08	29,64
	Gen 3	1	29,40	32,16	32,59	29,33
		2	29,11	28,29	33,00	28,74
		3	33,87	33,93	31,48	30,58
	Gen 4	1	28,87	28,28	30,09	28,76
		2	28,80	31,14	29,53	28,04
		3	31,71	28,27	29,65	29,17

Tabela 6 – Valores, em %, da variável resposta percentagem de fibra

Fornecedor	Genótipo	Lotes	Repetições			
			1	2	3	4
FORN 1	Gen 1	1	40,20	42,60	42,40	39,20
		2	37,70	39,65	41,87	38,25
		3	43,27	38,80	39,55	38,27
	Gen 2	1	45,20	46,97	42,33	42,00
		2	45,33	44,77	46,51	42,31
		3	45,95	46,30	43,42	43,00
	Gen 3	1	46,99	46,88	45,45	44,03
		2	43,50	46,33	45,54	45,84
		3	42,15	44,65	43,25	45,06
	Gen 4	1	46,07	43,65	43,84	44,63
		2	45,42	44,33	44,67	43,15
		3	42,96	44,23	45,86	45,73
FORN 2	Gen 1	1	43,62	46,16	46,67	45,66
		2	46,66	45,62	39,49	38,89
		3	41,84	38,59	42,19	40,18
	Gen 2	1	38,56	43,78	41,79	44,95
		2	41,20	43,50	42,70	39,59
		3	37,71	39,80	42,80	38,68
	Gen 3	1	43,43	38,87	39,68	43,23
		2	38,46	45,25	46,90	42,34
		3	45,40	44,80	46,60	42,40
	Gen 4	1	44,00	46,00	46,30	43,50
		2	47,00	46,90	45,50	44,25
		3	43,70	46,40	45,60	45,90
FORN 3	Gen 1	1	42,20	44,70	43,30	45,70
		2	46,50	43,70	43,90	44,70
		3	45,50	44,40	44,70	43,30
	Gen 2	1	43,00	44,30	45,80	45,80
		2	43,70	46,30	46,80	45,95
		3	46,75	45,75	39,57	38,96
	Gen 3	1	41,95	38,45	42,27	40,26
		2	38,45	43,87	41,87	45,03
		3	40,50	42,90	42,78	39,97
	Gen 4	1	37,98	39,63	41,98	38,48
		2	43,67	38,91	39,25	43,21
		3	45,25	44,65	46,45	42,29
FORN 4	Gen 1	1	43,87	45,86	46,16	43,34
		2	46,76	46,00	45,20	45,00
		3	43,55	44,80	43,35	47,00
	Gen 2	1	46,40	43,55	43,75	44,45
		2	45,35	44,27	44,79	43,25
		3	43,15	44,48	46,96	45,91
	Gen 3	1	43,37	46,03	46,56	45,59
		2	46,41	45,16	39,27	38,90
		3	41,76	38,65	42,10	40,17
	Gen 4	1	38,41	43,63	41,67	44,80
		2	41,88	45,42	40,06	44,67
		3	39,57	38,51	43,29	38,44

Tabela 7 – Valores, em gf/tex, da variável resposta resistência da fibra

Fornecedor	Genótipo	Lotes	Repetições			
			1	2	3	4
FORN 1	Gen 1	1	33,55	30,39	30,42	30,20
		2	32,96	32,64	27,52	36,97
		3	30,97	32,13	32,75	30,62
	Gen 2	1	26,72	31,65	27,00	32,37
		2	29,81	30,99	28,68	30,00
		3	31,85	32,57	35,00	29,75
	Gen 3	1	31,67	29,76	33,78	27,71
		2	33,79	32,49	30,73	32,94
		3	33,54	32,49	30,87	32,91
	Gen 4	1	33,54	36,12	32,02	28,26
		2	33,89	30,23	33,38	32,31
		3	30,34	33,38	34,45	30,46
FORN 2	Gen 1	1	33,01	29,94	31,98	35,26
		2	34,19	34,83	34,92	31,85
		3	31,35	32,94	34,93	31,67
	Gen 2	1	31,25	32,90	34,98	31,73
		2	30,54	30,72	35,23	28,30
		3	34,00	31,09	31,35	31,10
	Gen 3	1	33,80	33,45	28,42	37,80
		2	31,80	32,08	33,60	31,50
		3	27,62	32,35	27,89	33,12
	Gen 4	1	30,50	31,28	29,45	30,03
		2	32,76	33,50	35,09	29,88
		3	32,25	29,09	34,65	28,52
FORN 3	Gen 1	1	34,50	33,10	31,12	33,72
		2	34,31	37,09	32,01	29,21
		3	34,78	31,12	34,01	33,00
	Gen 2	1	31,01	34,02	35,33	31,12
		2	34,24	33,15	34,21	28,00
		3	33,00	30,50	32,80	36,11
	Gen 3	1	35,00	35,16	35,20	32,75
		2	32,08	33,06	35,05	32,27
		3	31,15	31,51	36,12	29,39
	Gen 4	1	31,52	29,29	29,32	29,10
		2	31,86	31,54	26,41	35,85
		3	29,86	31,01	31,64	29,51
FORN 4	Gen 1	1	25,60	31,53	25,90	31,17
		2	28,60	29,88	27,55	28,91
		3	30,74	31,46	33,98	28,55
	Gen 2	1	30,50	28,57	32,64	26,60
		2	32,49	31,38	29,61	31,82
		3	32,33	35,01	30,91	27,14
	Gen 3	1	32,78	29,11	32,08	31,01
		2	29,11	32,08	33,33	29,16
		3	32,71	31,35	32,61	26,01
	Gen 4	1	31,91	28,80	30,86	34,11
		2	33,02	33,76	33,81	30,56
		3	30,01	31,38	33,18	30,35

### 3.1 Resultados e discussão

Na Tabela 8 é apresentado os resultados da análise da variância (ANOVA), para as variáveis comprimento de fibra, percentagem de fibra e resistência da fibra. Pode-se observar através do teste F, que o fator lote dentro de genótipo dentro de fornecedor apresentou efeito significativo ao nível de 5% de probabilidade apenas nas variáveis comprimento da fibra e percentagem da fibra. O fator genótipo dentro de fornecedor teve efeito significativo nas três variáveis (comprimento da fibra, percentagem da fibra e resistência da fibra), ao nível de 5% de probabilidade. Estudando o fator fornecedor sozinho, pode-se observar através do teste F, que houve efeito significativo apenas na variável resistência da fibra ao nível de probabilidade de 5%. De acordo com Mendenhall e Sincich (1988), dado que

Tabela 8 – Análise da variância para o modelo aleatório em três estágios, com as variáveis resposta comprimento de fibra (COMP), percentagem de fibra (PF) e resistência da fibra (RF).

	C. V	G.L.	SQ	QM	F	p-valor
COMP	Forn	3	23,63	7,87	1,67	0,1754
	Gen(forn)	12	108,19	9,01	1,92	0,0369*
	<i>lotes(forn * gen)</i>	32	363,82	11,36	2,42	0,0002*
	Erro	144	677,80	4,70		
	Total	191	1173,46			
PF	Forn	3	10,36	3,45	0,84	0,47
	Gen(forn)	12	471,58	39,29	9,56	<,0001*
	<i>lotes(forn * gen)</i>	32	251,82	7,86	1,91	0,0053*
	Erro	144	591,89	4,11		
	Total	191	1325,66			
RF	Forn	3	77,12	25,70	5,01	0,0025*
	Gen(forn)	12	129,50	10,79	2,11	0,019*
	<i>lotes(forn * gen)</i>	32	124,10	3,87	0,76	0,82
	Erro	144	738,25	5,12		
	Total	191	1068,97			

o p-valor exceda  $\alpha = 0.05$ , não há evidências suficientes para indicar que os efeitos de fornecedores, genótipos e lotes, pelo menos um, influenciam as características da fibra.

Conforme a Tabela 9, podemos verificar as estimativas dos componentes de variância para as características da fibra. Para encontrar essas estimativas foram empregadas o método da ANOVA. É possível verificar a presença de estimativas negativas, e de acordo com Costa (2012), quando isso ocorre deve-se aceitar essa estimativa e usá-la como evidência que o verdadeiro valor do componente da variância é zero, ou reestimar o componente negativo da variância por um método que resulte em um valor não-negativo. Neste caso, irá ser considerado que os valores negativos das estimativas dos componentes da variância são zero. Com isso, temos as seguintes estimativas de componente de variância para cada variável da fibra.

Tabela 9 – Esperança dos quadrados mínimos, estimativas da esperança dos quadrados e esperança dos componentes de variância ( E.C.V.) para o modelo aleatório em três estágios, com as variáveis resposta comprimento de fibra, percentagem de fibra e resistência da fibra.

	C. V	E(QM)	E.C.V.
Comp	Forn	$\sigma^2 + 4\sigma_{lotes(forn*gen)}^2 + 12\sigma_{gen(forn)}^2 + 48\sigma_{(forn)}^2$	-0,023
	Gen(forn)	$\sigma^2 + 4\sigma_{lotes(forn*gen)}^2 + 12\sigma_{gen(forn)}^2$	-0,196
	$lotes(forn * gen)$	$\sigma^2 + 4\sigma_{lotes(forn*gen)}^2$	1,665
	Erro	$\sigma^2$	4,707
PF	Forn	$\sigma^2 + 4\sigma_{lotes(forn*gen)}^2 + 12\sigma_{gen(forn)}^2 + 48\sigma_{(forn)}^2$	-0,746
	Gen(forn)	$\sigma^2 + 4\sigma_{lotes(forn*gen)}^2 + 12\sigma_{gen(forn)}^2$	2,619
	$lotes(forn * gen)$	$\sigma^2 + 4\sigma_{lotes(forn*gen)}^2$	0,939
	Erro	$\sigma^2$	4,110
Res	Forn	$\sigma^2 + 4\sigma_{lotes(forn*gen)}^2 + 12\sigma_{gen(forn)}^2 + 48\sigma_{(forn)}^2$	0,310
	Gen(forn)	$\sigma^2 + 4\sigma_{lotes(forn*gen)}^2 + 12\sigma_{gen(forn)}^2$	0,576
	$lotes(forn * gen)$	$\sigma^2 + 4\sigma_{lotes(forn*gen)}^2$	-0,312
	Erro	$\sigma^2$	5,126

A estimativa da variabilidade total para o comprimento da fibra é  $\hat{\sigma}_T^2 = \hat{\sigma}_\tau^2 + \hat{\sigma}_\beta^2 + \hat{\sigma}_\gamma^2 + \hat{\sigma}^2 = 6,372$ , ou seja, estimam que 26,1% da variabilidade total é explicada pelos efeitos de Lote dentro de genótipo dentro de fornecedor e 73,9% pela variação do acaso. Para a variável percentagem da fibra do algodão, a estimativa da variabilidade total é  $\hat{\sigma}_T^2 = \hat{\sigma}_\tau^2 + \hat{\sigma}_\beta^2 + \hat{\sigma}_\gamma^2 + \hat{\sigma}^2 = 7,668$ , ou seja, estimam que 34,2% da variabilidade total é explicada pelos efeitos de genótipo dentro de fornecedor, 12,2% é explicada pelos efeitos de Lote dentro de genótipo dentro de fornecedor e 53,6% pela variação do acaso. Já para a resistência da fibra a estimativa da variabilidade total é  $\hat{\sigma}_T^2 = \hat{\sigma}_\tau^2 + \hat{\sigma}_\beta^2 + \hat{\sigma}_\gamma^2 + \hat{\sigma}^2 = 6,012$ , ou seja, estimam que 5,1% da variabilidade total é explicada pelos efeitos de fornecedor, 9,6% é explicado pelos efeitos genótipo dentro de fornecedor e 85,3% pela variação do acaso.

De acordo com os resultados apresentados, foi possível observar por meio das Figuras 2a, 2b e 2c possíveis valores discrepantes nos dados observados para as respectivas variáveis respostas comprimento da fibra, percentagem da fibra e resistência da fibra conforme os gráficos de resíduos estudentizados vs valores ajustados, pois foi verificado resíduos fora do intervalo (-2; 2), indicando que pode haver uma certa correlação entre as observações, podendo ser um possível outlier. Convém ressaltar, no entanto, que o número de observações é pequeno e assim, mesmo supondo que a distribuição dos erros seja normal, a probabilidade de aparecerem valores discrepantes é grande. Pode-se notar que, os resíduos ajustados estão dispostos de maneira aleatória em torno de zero.

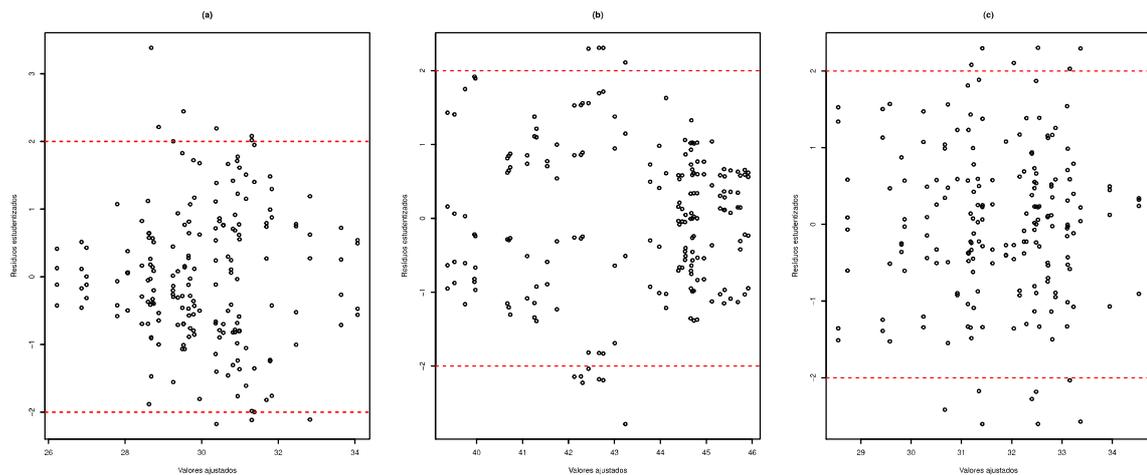


Figura 2 – Gráficos dos resíduos estudentizados versus valores ajustados para a variável resposta comprimento de fibra (a), percentagem de fibra (b) e resistência da fibra (c), respectivamente

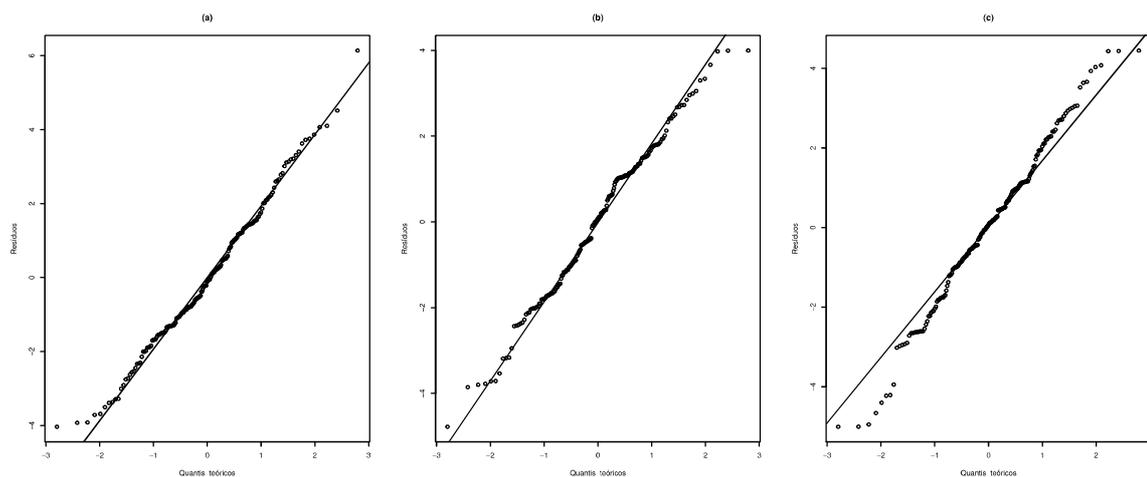


Figura 3 – Gráficos dos quantis normais para a variável resposta comprimento de fibra (a), percentagem de fibra (b) e resistência da fibra (c), respectivamente

Na Tabela 10 temos a estatística de shapiro Wilks e o valor p para as características da fibra, é possível observar que com nível de significância de 5% que as amostras das variáveis comprimento da fibra, percentagem da fibra e resistência da fibra provém de uma população normal, pois os valores de p foram maiores que o nível de significância nas três variáveis. Esse fato é confirmado pela aleatoriedade dos pontos em torno da reta na figura 3a, 3b e 3c respectivamente.

Tabela 10 – Estatística Shapiro Wilks (W) com os respectivos valores de p das variáveis comprimento da fibra, percentagem de fibra e resistência da fibra.

Variável da Fibra	(W)	valor p
Comprimento	0,99057	0,2409
Percentagem	0,99022	0,2163
Resistência	0,99083	0,2614

De acordo com os gráficos (Figura 4) é possível observar que os modelos ajustados foram adequados, os resíduos apresentam distribuição normal.

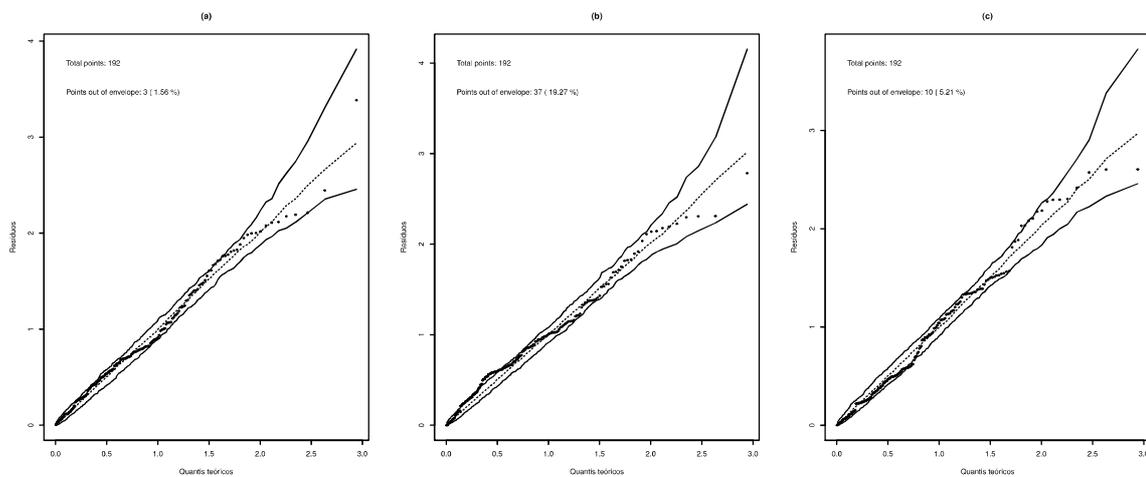


Figura 4 – Gráficos envelope simulado para a variável resposta comprimento de fibra (a), percentagem de fibra (b) e resistência da fibra (c), respectivamente

## 4 Conclusão

Pelo trabalho apresentado conseguimos identificar alguns pontos importantes sobre o experimento, em relação as variáveis resposta comprimento da fibra e percentagem da fibra, o fator fornecedor isolado por se só não influencia em nada no experimento, porem ao aninhar com outros fatores como por exemplo genótipo e lote, a efetivação de tal aninhamento já nos trás mudanças consideráveis para a proposta do experimento apresentado. Já em relação à variável resistência da fibra o fornecedor junto com genótipo e lote não tiveram efeito significativo para a variável.

Pela literatura consultada, não foi encontrado um embasamento teórico sobre os experimentos aninhados. Então, para senar parte dessa lacuna, desenvolveu-se a análise da variância para um experimento aninhado com três estágios, pelo métodos dos quadrados mínimos ordinários.

## Referências

- ALBUQUERQUE, A. M. M. d. O. Análise estatística de um experimento aninhado com uma aplicação em dados de algodão. 2016. Citado na página 25.
- ATKINSON, A. C. *Plots, transformations and regression; an introduction to graphical methods of diagnostic regression analysis*. [S.l.], 1985. Citado na página 24.
- BARBIN, D. *Componentes de Variância: teoria e aplicação*. [S.l.]: Piracicaba, 1993. Citado 2 vezes nas páginas 11 e 19.
- CORDEIRO, G.; NETO, E. *Modelos paramétricos*. [S.l.]: UFPE, 2006. Citado na página 23.
- COSTA, J. B. F. Aspectos teóricos e práticos da análise estatística de um experimento em blocos casualizados com efeitos aleatórios. 2012. Citado 2 vezes nas páginas 21 e 29.
- DRAPER, N. R.; SMITH, H. *Applied regression analysis*. [S.l.]: New York: John Wiley & Sons, 1998. Citado na página 23.
- HARVILLE, D. A. Quadratic unbiased estimation of variance components for the one-way classification. *Biometrika*, Biometrika Trust, v. 56, n. 2, p. 313–326, 1969. Citado 2 vezes nas páginas 11 e 21.
- HERBACH, L. H. Properties of model ii–type analysis of variance tests, a: Optimum nature of the f-test for model ii in the balanced case. *The Annals of Mathematical Statistics*, JSTOR, p. 939–959, 1959. Citado 2 vezes nas páginas 11 e 21.
- HICKS, C. R. *Fundamental concepts in the design of experiments*, New York Holt, Rinehart and Winston. [S.l.]: Inc, 1973. Citado na página 19.
- INC, S. I. *SAS/STAT® 9.2. User's Guide*. [S.l.]: SAS Institute Inc Cary, NC, 2008. Citado na página 25.
- MENDENHALL, W.; SINCICH, T. *Statistics for Engineering and the Sciences*. [S.l.: s.n.], 1988. Citado na página 29.
- MONTGOMERY, D. *Design and analysis of experiments*. [S.l.]: John Wiley & Sons, 1984. Citado 3 vezes nas páginas 8, 12 e 13.
- MONTGOMERY, D. C. *Design and analysis of experiments*. [S.l.]: John Wiley & Sons, 2001. Citado na página 20.
- MORAL, R. d. A. *Modelagem estatística e ecológica de relações tróficas em pragas e inimigos naturais*. Tese (Doutorado) — Universidade de São Paulo, 2013. Citado na página 24.
- PRUDENTE, A. A. Modelos não-lineares de regressão: alguns aspectos de teoria assintótica. *TEDE*, 2009. Citado na página 23.
- QUINN, G. P.; KEOUGH, M. J. *Experimental design and data analysis for biologists*. [S.l.]: Cambridge University Press, 2002. Citado na página 21.

- SANTOS, J. W. dos et al. *Estatística Experimental Aplicada*. [S.l.]: Embrapa Algodão e UFCG, 2008. Citado na página 11.
- SEARLE, S. R. Linear models. new york: J. Wiley & Sons, 1971. Citado na página 21.
- SILVA, J.; ZONTA, E. Conceitos e algoritmos úteis em estatística experimental. *Simpósio de estatística aplicada à experimentação agrônômica*, v. 4, 1991. Citado na página 12.
- SWALLOW, W. H.; MONAHAN, J. F. Monte carlo comparison of anova, mivque, reml, and ml estimators of variance components. *Technometrics*, Taylor & Francis, v. 26, n. 1, p. 47–57, 1984. Citado na página 21.
- TEAM, R. C. a language and invironment for statistical computing r, foundation for statistical computing. *Vienna, Austria (available at <http://www.R-project.org>)*, 2017. Citado na página 25.
- THOMPSON, W. A. J. The problem of negative estimates of variance components. *The Annals of Mathematical Statistics*, JSTOR, p. 273–289, 1962. Citado 2 vezes nas páginas 11 e 21.