



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

Adri cia Mirelly Marques de Oliveira Albuquerque

An lise estat stica de um experimento aninhado com uma aplica o em dados de algod o

Campina Grande - PB

Abril de 2016

Adrcia Mirelly Marques de Oliveira Albuquerque

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

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Estatística Aplicada 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 especialista em Estatística.

Orientador: Prof. Dr. Elias Dias Coelho Neto

Campina Grande - PB

Abril de 2016

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.

A345d Albuquerque, Adrícia Mirelly Marques de Oliveira
Análise estatística de um experimento aninhado com uma aplicação em dados de Algodão [manuscrito] / Adrícia Mirelly Marques de Oliveira Albuquerque. - 2016.
34 p. : il.

Digitado.

Monografia (Estatística Aplicada) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2016.

"Orientação: Prof Drº Elias Dias Coelho Neto, Departamento de Estatística".

1. Delineamento Experimental. 2. Algodão. 3. Componente de Variância. 4. Experimento hierárquico. I. Título.

21. ed. CDD 576.52

Adri cia Mirelly Marques de Oliveira Albuquerque

An lise estat stica de um experimento aninhado com uma aplica o em dados de algod o

Trabalho de Conclus o de Curso apresentado
ao Curso de Especializa o em Estat stica
Aplicada do Departamento de Estat stica do
Centro de Ci ncias e Tecnologia da Uni-
versidade Estadual da Para ba em cumprimento
 s exig ncias legais para obten o do t tulo
de especialista em Estat stica.

Trabalho aprovado em 22 de Abril de 2016.

BANCA EXAMINADORA

Elias Dias Coelho Neto

Prof. Dr. Elias Dias Coelho Neto
Universidade Estadual da Para ba

Tiago Almeida de Oliveira

Prof. Dr. Tiago Almeida de Oliveira
Universidade Estadual da Para ba

Ana Patr cia Bastos Peixoto

Prof.^a Dr.^a Ana Patr cia Bastos Peixoto
Universidade Estadual da Para ba

Dedico este trabalho aos meus pais, esposo, irmãos e filhos (Sophia e Gabriel) com carinho.

Agradecimentos

Em primeiro lugar agradeço a DEUS, que foi minha maior fortaleza, pela paciência que me deu, pelas noites de sono que estive zelando por mim e me dando força, apesar das dificuldades que tive, para não desistir.

Ao meu orientador, Dr. Elias Dias Coelho Neto pelo empenho dedicado à elaboração deste trabalho.

Ao grande Mestre José Wellington dos Santos, pessoa extraordinária, de coração sem tamanho que muito me ajudou no trabalho, muito obrigada.

Aos professores do Curso de especialização da UEPB, em especial a Ana Patrícia Peixoto e Tiago Almeida de Oliveira o meu eterno agradecimento pelas orientações, pela paciência, pela amizade, pela força e apoio, pois em muitos momentos abriram a porta de sua casa para me ajudar.

Aos meus colegas do curso da especialização, em especial a minha grande amiga-irmã-companheira-parceira de todas as horas Raquel Eloy.

Ao meu esposo, Anderson Albuquerque, pelo amor, carinho e por estar comigo nas horas boas e ruins, obrigada pela compreensão e companheirismo.

Por fim, aos meus pais Maria da Paz e Heronildes e minha querida irmã Dayanna que me deram apoio, incentivo nas horas difíceis de desânimo e cansaço e pelos vários dias que ficaram com meus filhos para que eu pudesse ir para as aulas e estudar.

"A vida é uma peça de teatro que não permite ensaios. Por isso, cante, chore, dance, ria e viva intensamente, antes que a cortina se feche e a peça termine sem aplausos". (Autor Desconhecido)

Resumo

Objetivou-se, com esta pesquisa, investigar o efeito de genótipos e fornecedores sobre as características tecnológicas da fibra de algodão. As variáveis estudadas foram: comprimento da fibra (milímetros), percentagem de fibra (%) e resistência da fibra (grama força /tex). O experimento foi conduzido pela Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA ALGODÃO, na safra 2014/2015, a qual cedeu dados de um experimento aninhado para dados balanceados em que os fatores foram: quatro fornecedores onde cada um nos fornece quatro genótipos e de cada genótipos retiramos quatro repetições. Os dados foram aplicados em dois modelos, misto e aleatório, onde foi utilizado o método da análise da variância para ambos os modelos e também o teste de Tukey apenas no modelo misto, dado o fator principal fixo. Além disso, foram realizados estudos da análise de resíduos. Assim sendo, verificou-se, em ambos os modelos, que não houve diferença significativa para as variáveis respostas comprimento de fibra e resistência da fibra. No entanto, a variável percentagem de fibra apresentou diferença significativa entre os fornecedores.

Palavras-chaves: Delineamento Experimental, Experimento hierárquico, algodão e Componente de Variância.

Abstract

This research will investigate the effect of genotypes and suppliers of the technological characteristics of cotton fibre. The variables analysed were: fibre length (mm), percentage (%) and strength (gram force /tex). The experiment was performed by the Brazilian Agricultural Research Corporation - EMBRAPA ALGODÃO, in the harvest of 2014/2015, which yielded data of an experiment nested for data balanced with the following factors: four suppliers where each one provides the four genotypes and each genotype; from each genotype, four replications were withdraw. Data were applied in two models, mixed and random, where the analysis of variance was used; the Tukey test was used only in the mixed model with the main factor fixed. Furthermore, analysis of residues was performed. As a result, it was found that there is no significant difference for the variable response fibre length and strength in both models. However, for the variable percentage of fibre a significant difference between the suppliers was found.

Key-words: Experimental Design, hierarchical experiment, cotton and variance component.

Lista de ilustrações

Figura 1 – Croqui de um experimento hierárquico	14
Figura 2 – Reescritura do Croqui de um experimento hierárquico	15
Figura 3 – Croqui aninhado de três fases com duas repetições.	20
Figura 4 – Teste de Tukey para as médias do fator fornecedor para a variável comprimento da fibra, percentagem da fibra , acima, e resistência da fibra, abaixo, respectivamente.	28
Figura 5 – Análise dos resíduos para a variável comprimento da fibra	29
Figura 6 – Análise dos resíduos para a variável percentagem da fibra	30
Figura 7 – Análise dos resíduos para a variável resistência da fibra	31

Lista de tabelas

Tabela 1 – Quadrados médios esperados nos experimentos aninhados em dois estágios	16
Tabela 2 – Análise de variância para o projeto aninhado em duas etapas	17
Tabela 3 – Análise de Variância para experimentos aninhados em três estágios . .	21
Tabela 4 – Quadrados Médios Esperados para o Experimento Aninhado em Três Estágios com A e B Fixados e C Aleatório, pelo método de Hicks	21
Tabela 5 – Razão-F para o teste da hipótese nula para o modelo ANOVA de dois fatores aninhados	22
Tabela 6 – Estimativas de componentes de variância para três fatores aninhados e com fator A aleatório, fator B dentro de A aleatório e com fator C dentro de B também aleatório. São realizadas n réplicas dentro de cada célula.	22
Tabela 7 – Valores, em milímetros, da variável resposta comprimento de fibra . . .	23
Tabela 8 – Valores, em %, da variável resposta percentagem de fibra	24
Tabela 9 – Valores, em gf/tex, da variável resposta resistência da fibra	24
Tabela 10 – Análise da variância para o modelo aleatório, com suas respectivas Somas de Quadrados (SQ), Quadrados Médios (QM), estatística F, e valor p, para as variáveis comprimento de fibra, percentagem de fibra e resistência da fibra.	26
Tabela 11 – Análise da variância para o modelo misto, dado forn fixo, com as variáveis resposta comprimento de fibra, percentagem de fibra e resistência da fibra.	27
Tabela 12 – Variáveis com suas devidas médias	27

Sumário

1	INTRODUÇÃO	11
2	FUNDAMENTAÇÃO TEÓRICA	13
2.1	Experimento aninhado	13
2.2	Modelo Estatístico	15
2.3	Verificação de diagnóstico	17
2.4	Componentes de variância	18
2.5	Experimento aninhado generalizado em m-estágios	19
2.6	Razão - F	21
3	MATERIAL E MÉTODOS	23
4	RESULTADOS OBTIDOS	26
4.1	Análise descritiva	26
4.2	Análise de resíduos	28
5	CONCLUSÃO	32
5.1	Perspectivas de trabalhos futuros	32
6	REFERÊNCIAS	33

1 Introdução

A busca pelo melhoramento da produção de alimentos é um dos problemas experimentais mais antigos que o homem tenta resolver. Busca-se na própria natureza o conhecimento necessário para a melhoria da qualidade do alimento, além do aumento da produção. Com o desenvolvimento do Cálculo por Newton e Leibniz o conhecimento científico do manejo da terra avançou e levou o homem a se aperfeiçoar, levando-o a criação de novos métodos, técnicas e procedimentos mais adequados e convenientes para a produção de alimentos. É um conhecimento que se renova e que ajuda a desenvolver procedimentos mais efetivos para solucionar problemas e responder questões relacionadas à agricultura. Esse progresso é conseguido através da permanente atividade de indagação a que se dedicam os pesquisadores.

Hoje em dia, quando um pesquisador estuda os métodos e as técnicas experimentais, ele inevitavelmente vai ler algo sobre Ronald A. Fisher. Este cientista se destacou nas décadas de 20 e 30 por ter sido o responsável pela análise de dados da Estação Experimental de Rothamsted de Londres, Inglaterra. Ele foi o primeiro cientista a usar métodos estatísticos nos delineamentos experimentais. Não contente com as técnicas que existiam até então, Fisher desenvolveu a análise de variância, que foi o primeiro método de análise de dados experimentais escrito com rigor matemático. A maioria das aplicações feitas por Fisher foram às áreas de agricultura e biologia. Hoje se sabe que essa técnica é aplicada em todas as áreas do conhecimento que trabalham com algum tipo de experimento aleatório.

De acordo com Santos et al (2008) a estatística experimental no mundo de hoje é uma necessidade real, estando em todas as áreas do conhecimento humano como uma ferramenta para auxiliar as decisões a ser tomadas, permitindo maior segurança e mais precisão a pesquisa. Segundo Spiegel (1982), o planejamento de experimentos consiste no estudo dos métodos da amostragem e dos problemas correlatos que surgem. Conforme Johnston e Pennypacker (1993a), delinear um experimento é planejar condições controle e experimentais de modo a permitir comparações significativas entre elas, verificar os efeitos da variável independente e responder ao problema de pesquisa. Por esta razão, é extremamente importante que o experimentador conheça as características, as possibilidades e os limites dos delineamentos experimentais, em geral. Realização de experimentos demandam recursos financeiros, que normalmente são escassos, como enfatizado por Milliken e Johnson (1998), que dissertaram sobre experimentos bem planejados e analisados que fornecem uma maior quantidade de informações sobre as condições investigadas. Assim, destaca-se a importância de conhecer bem os delineamentos experimentais para reduzir gastos.

No setor industrial, a teoria do planejamento experimental é uma das mais impor-

tantes por possibilitar redução de custos da produção, levando ao aumento do lucro, além de propiciar o surgimento de novos materiais e produtos. Não é raro notar a existência de uma ordem natural entre os fatores de um experimento, por exemplo, para se avaliar a qualidade do café, podem-se comparar lotes que vem de diferentes fornecedores. Nesse sentido, o modelo Experimental Aninhado, também conhecido como Experimental Hierárquico, surgiu como uma alternativa na análise de dados experimentais cujo croqui é organizado seguindo alguma hierarquia natural entre os fatores.

Nesse sentido, o objetivo deste trabalho é o de mostrar o modelo matemático utilizado na análise de dados provenientes de um planejamento experimental aninhado balanceado e aplicá-lo a dados de algodão. Os fatores que serão considerados são diferentes fornecedores e, também, diferentes genótipos.

2 Fundamentação teórica

Segundo Silva (2006), desde a pré-história o homem busca pelo conhecimento da natureza. Os recursos limitados permitiam apenas percepções dos fenômenos através dos sentidos e explicações superficiais, aderindo também ao culto de forças da natureza como forma de conhecimento e até explicações religiosas. Foi o conhecimento científico que levou o homem a se aperfeiçoar levando-o a criação de novos métodos, técnicas e procedimentos mais adequados e convenientes, é um conhecimento que se renova para desenvolver procedimentos mais efetivos para solucionar problemas e responder questões. Esse progresso é conseguido através da permanente atividade de indagação a que se dedicam os pesquisadores.

Assim surgem os princípios básicos da experimentação: repetição, casualização e controle local. A repetição se refere ao uso de mais de uma parcela por tratamento. Este princípio tem por finalidade propiciar a obtenção de uma estimativa do erro experimental, melhorar a precisão do experimento e fazer com que o teste de hipótese seja possível. Já a casualização, tem por finalidade principal propiciar, a todos os tratamentos, a mesma probabilidade de serem designados a qualquer unidade experimental (a casualização faz com que o teste de hipótese seja válido). O controle local divide um ambiente heterogêneo em sub-ambientes homogêneos e torna o delineamento experimental mais eficiente pela redução de erro experimental.

De acordo com Milliken e Johnson (1998), experimentos bem planejados e analisados fornecem uma maior quantidade de informação sobre as condições investigadas de acordo com os recursos gastos. Assim, destaca-se a importância de conhecer bem os delineamentos experimentais.

Neste trabalho, será abordado um experimento hierárquico onde diz respeito à estrutura dos fatores e seus níveis. Diz-se que um delineamento é hierárquico quando níveis de um fator B só ocorrem em determinados níveis de um fator A, ou seja, o fator A possui uma maior hierarquia e determina quais os possíveis níveis de B.

2.1 Experimento aninhado

Em um experimento aninhado comparam-se níveis de fatores principais e, também se comparam níveis de fatores secundários que variam “dentro” de cada um dos fatores 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, ou seja, os tratamentos secundários

estão aninhados ao tratamento principal, figura 1. Nessa figura, mostra-se o croqui de um experimento hierárquico em dois estágios. Para exemplificar, foram utilizados como fator principal fornecedores e lotes como fator secundário. Note que os diversos níveis de lotes estão posicionados em um nível inferior à posição dos fornecedores. No entanto, apenas pela representação do croqui, o leitor poderia ter dúvidas quanto ao tipo de experimento podendo acreditar que seria em experimento fatorial. Porém, 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. Sendo assim, tem-se um experimento aninhado (hierárquico) porque os lotes de cada fornecedor são únicos para cada fornecedor em particular. Isto é, o lote 1 do fornecedor 1 não tem conexão com o lote 1 de qualquer outro fornecedor e assim por diante.

Experimentos planejados com fatores aninhados são comumente utilizados no campo das ciências biológicas. Por exemplo, Quinn & Keough (1993) realizaram um experimento para examinar o efeito de diferentes tamanhos de invólucros sobre o crescimento *Cellana tramoserica* em rochas interdiárias, cuja variável resposta utilizada foi altura da concha; Outro exemplo, Caselle & Warner (1996) observaram as densidades populacionais de peixes residentes em recifes de corais de cinco locais diferentes das Ilhas Virgens, costa norte dos Estados Unidos da América. Eles usaram locais como o fator principal e transectos dentro de cada local como o fator aninhado; Bellgrove et al. (1997) analisaram dados da variação temporal da abundância de propágulos de algas ao longo das rochas nas linhas costeiras do oceano, eles coletaram volumes de água de uma área interdiária em diferentes datas dentro de duas estações (os dias foram um fator aninhado e fixo, pois dia não são aleatoriamente escolhidos). Algumas vezes fatores aninhados aparecem como parte de modelos mais complexos, como uma mistura de experimento fatorial com um experimento aninhado. Twombly (1996) usou um experimento misto de fatorial com aninhado para examinar os efeitos da concentração de comida para diferentes sibships (ovos provenientes de uma mesma fêmea em um dado instante de tempo) sobre o desenvolvimento de água fresca de copepodes *Mesocyclops edax*.

É possível observar no esquema da figura 1 que os lotes estão aninhados sob fornecedores. No caso de existir dúvidas se o experimento é realmente hierárquico, pode-se tentar reenumerados arbitrariamente, se sim dizemos que o fator é aninhado. Logo, ele pode ser reescrito como segue na figura 2:

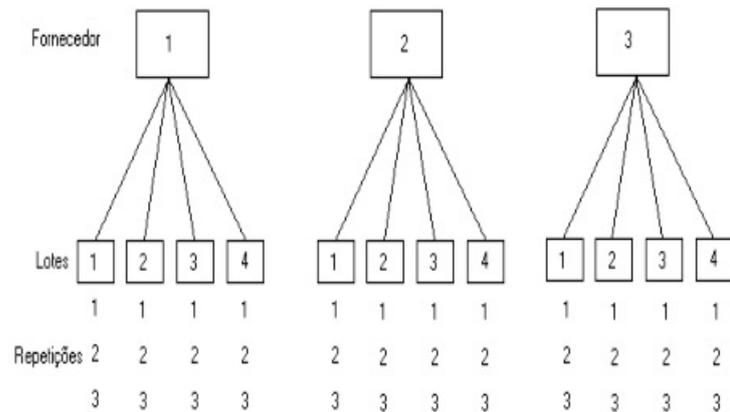


Figura 1 – Croqui de um experimento hierárquico

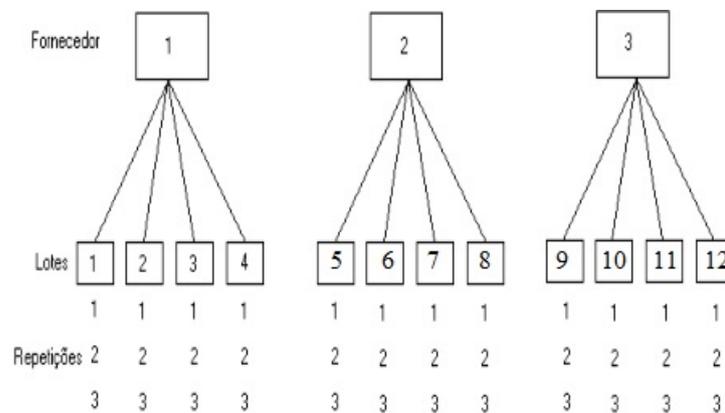


Figura 2 – Reescritura do Croqui de um experimento hierárquico

2.2 Modelo Estatístico

Para experimentos aninhados em dois estágios segue a análise estatística,

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

em que:

a é o número de níveis do fator A; b é o número de níveis do fator B inferior a cada nível de A; n é o número de repetições. O subscrito $j(i)$ indica que o j -ésimo nível do fator B é aninhado em uma posição inferior a do i -ésimo nível do fator A.

Pelas figuras 1 e 2 tem-se um planejamento experimental aninhado balanceado, pois existe um mesmo número de níveis e réplicas. É conveniente pensar que a réplica está sendo aninhada dentro da combinação de níveis de A e B; assim, o subscrito $(ij)k$ é usado para o erro. De acordo com Quinn (2002), a soma de quadrados total pode ser dividida

em seus componentes aditivos da seguinte maneira:

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

e expandindo o lado direito da equação acima, tem-se:

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

porque os três termos com produtos cruzados são iguais a zero. Esta equação indica que a soma de quadrados total (SST) pode ser particionada em soma de quadrados referente ao fator A (SSA), uma soma de quadrados referente ao fator B dentro dos níveis de A (SSB(A)), e uma soma de quadrados referente ao erro (SSE).

Algebricamente,

$$SST = SSA + SSB(A) + SSE.$$

Sendo assim para os graus de liberdade temos: $abn - 1$ graus de liberdade para SST, $a - 1$ graus de liberdade para SSA, $a(b - 1)$ graus de liberdade para SSB(A), e $ab(n - 1)$ graus de liberdade para o erro. Note que, $abn - 1 = (a - 1) + a(b - 1) + ab(n - 1)$.

Para obter-se os quadrados médios independentemente distribuídos cada um possuindo distribuição de probabilidade qui quadrado, todas elas independentes duas a duas, deve-se dividir cada soma de quadrados da equação acima pelos seus respectivos graus de liberdade, dados que os erros são normais e identicamente distribuídos $(0, \sigma^2)$. Sendo assim, a razão de qualquer dois quadrados médios possui distribuição de probabilidade F de Fisher-Snedecor.

O cálculo para determinar as estatísticas apropriadas para os efeitos dos fatores A e B, dependerá da natureza dos mesmos, ou seja, se esses fatores são fixos ou aleatórios. De acordo com Montgomery (2001), se os fatores A e B são fixos, nós assumimos que $\sum_{j=1}^b \tau_i = 0$ e $\sum_{j=1}^b \beta_{j(i)} = 0$ ($i=1,2,\dots,a$). Isto é, o efeito do tratamento A soma zero, e o efeito do tratamento B soma zero dentro de cada nível de A. Alternativamente, se A e B são aleatórios, nós assumimos que τ_i é NID($0, \sigma_\tau^2$) e $B_{j(i)}$ é NID ($0, \sigma_\beta^2$). Modelos mistos com A fixo e B aleatório também são de grande utilidade no processo produtivo.

A tabela abaixo nos fornece os quadrados médios esperados para essas situações, conforme Barbin (1993), para os três tipos de modelos que temos para um experimento aninhado.

Observando as condições da tabela 1, pode-se testar as hipóteses que irá depender do tipo de efeito considerado. Sendo assim, se considerarmos estes como efeitos fixo e aleatório respectivamente, as hipóteses testadas serão:

Se os níveis de A e B são fixados, testa-se $H_0 : \tau_i = 0$ por meio de $\frac{QMA}{QME}$ e $H_0 : \beta_{j(i)} = 0$ testa-se por meio de $\frac{QMB(A)}{QME}$.

Tabela 1 – Quadrados médios esperados nos experimentos aninhados em dois estágios

E(QM)	A Fixo B Fixo	A Fixo B Aleatório	A Aleatório B Aleatório
$E(QM_A)$	$\sigma^2 + \frac{bn \sum \tau_i^2}{a-1}$	$\sigma^2 + n\sigma_\beta^2 + \frac{bn \sum \tau_i^2}{a-1}$	$\sigma^2 + n\sigma_\beta^2 + bn\sigma_\tau^2$
$E(QM_{B(A)})$	$\sigma^2 + \frac{n \sum \sum \beta_{j(i)}^2}{a(b-1)}$	$\sigma^2 + n\sigma_\beta^2$	$\sigma^2 + n\sigma_\beta^2$
$E(QM_E)$	σ^2	σ^2	σ^2

Se A é um fator fixo e B é aleatório, então $H_0 : \tau_i = 0$ é testado por $\frac{QMA}{QMB(A)}$ e $H_0 = \sigma_\beta^2 = 0$ é testada por $\frac{QMB(A)}{QME}$.

Por fim, se ambos A e B são fatores aleatórios, nós testamos $H_0 : \sigma_\tau^2 = 0$ por $\frac{QMA}{QMB(A)}$ e $H_0 = \sigma_\beta^2 = 0$ por $\frac{QMB(A)}{QME}$.

Portanto, podemos resumir este procedimento em uma tabela de análise de variância, tabela 2.

Tabela 2 – Análise de variância para o projeto aninhado em duas etapas

Fonte de Variação	Soma de quadrados	Graus de liberdade	Quadrado médio
A	$bn \sum (\bar{y}_{i..} - \bar{y}_{...})^2$	a-1	QM_A
B dentro de A	$n \sum \sum (\bar{y}_{ij.} - \bar{y}_{j..})^2$	a(b-1)	$QM_{B(A)}$
Resíduos	$\sum \sum \sum (\bar{y}_{ijk} - \bar{y}_{ij.})^2$	ab(n-1)	QME
Total	$\sum \sum \sum (\bar{y}_{ijk} - \bar{y}_{...})^2$	abn-1	

Segundo Barbin (1993), o cálculo para as fórmulas das somas de quadrados da tabela 2 podem ser obtidas expandindo-se as quantidades na equação acima e simplificando. São elas:

$$SS_A = \frac{1}{bn \sum_{i=1}^a y_{i..}^2} - \frac{y_{...}^2}{abn}$$

$$SS_{B(A)} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b y_{ij.}^2 - \frac{1}{bn} \sum_{i=1}^a y_{i..}^2$$

$$SS_E = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}^2 - \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b y_{ij.}^2$$

$$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}^2 - \frac{y_{...}^2}{abn}$$

Podemos observar que a equação para se obter $SS_{B(A)}$ pode ser escrita como

$$SS_{B(A)} = \sum_{i=1}^a \left[\frac{1}{n} \sum_{j=1}^b y_{ij.}^2 - \frac{y_{i..}^2}{bn} \right]$$

Isto expressa a ideia de que $SS_{B(A)}$ é a soma de quadrados entre os níveis de B para cada nível de A, somados sobre todos os níveis de A.

2.3 Verificação de diagnóstico

A utilização da Análise de Variância para um conjunto de dados provenientes de algum experimento pressupõe a verificação de algumas hipóteses. As hipóteses fundamentais da análise de variância são:

- i) Os diferentes efeitos admitidos no modelo estatístico são aditivos (aditividade).
- ii) Os erros das observações não são correlacionados (independência).
- iii) Os erros têm a mesma variância (homocedasticidade).
- iv) Os erros têm distribuição Normal (normalidade)

Diante disso, devemos observar as condições para que cada um dos pressupostos sejam atendidos, como por exemplo os efeitos devem se aditivos, não havendo aditividade isso resultará na heterocedasticidade do erro que por sua vez, afeta o nível de significância para comparações entre tratamentos. Quanto a normalidade, há uma perda de eficiência na estimação dos efeitos de tratamentos e uma correspondente perda de poder dos testes. Outra ferramenta útil para a verificação da normalidade é o uso do papel normal de probabilidade onde devem ser plotados os resíduos, pois a grande vantagem da análise de resíduos é que, por simples inspeção do gráfico, pode-se concluir sobre a Normalidade.

Em relação a independência dos erros, quando eles são correlacionados, os testes de significância não são válidos. Há casos em que, devido a uma correlação positiva entre os erros, o teste de F leva a um excesso de resultados significativos. Em casos de correlação negativa, o valor da estatística F_c pode ser muito menor que um. Plotando os resíduos na ordem em que os dados foram coletados pode-se verificar facilmente a existência de correlação entre eles. Quando os resíduos se distribuem de maneira desordenada, pode-se pensar em não existência de correlação.

A falta de homogeneidade da variância é uma das mais graves quebras de suposição básica principalmente para os modelos não balanceados e os modelos de efeitos aleatórios. Através de um gráfico de resíduos versus o valor estimado ou versus a variável X (tratamentos), pode-se detectar a não homogeneidade de variância. Em geral, os resíduos devem ser não correlacionados com qualquer outra variável. Os gráficos devem apresentar a ausência de estrutura entre os resíduos e a outra variável plotada. A heterogeneidade dos erros pode ser classificada como irregular e regular. A heterogeneidade é irregular quando aparentemente não existe uma relação entre médias e variâncias enquanto que, existindo certa relação entre as médias e as variâncias dos vários tratamentos, a heterogeneidade é dita regular.

Para a verificação de diagnóstico do experimento aninhado em dois estágios a ferramenta que usamos é a análise residual. Para esse experimento, os resíduos são:

$$e_{ijk} = y_{ijk} - \hat{y}_{ijk}$$

onde os valores ajustados (previstos) são:

$$\hat{y}_{ijk} = \hat{\mu} + \hat{\tau}_j + \hat{\beta}_{j(i)}$$

e se fizermos a restrição usual sobre os parâmetros do modelo, $\sum_i \hat{\tau}_j = 0$ e $\sum_i \hat{\beta}_{j(i)} = 0$, $i=1,2,\dots,a$, então $\hat{\mu} = y_{...}$, $\hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...}$ e $\hat{\beta}_{j(i)} = \bar{y}_{ij.} - \bar{y}_{i..}$.

Assim sendo, os valores ajustados (previstos) são:

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

Desse modo, os resíduos para o experimento aninhado em dois estágios são: $e_{ijk} = y_{ijk} - \hat{y}_{ijk}$, em que $y_{ij.}$ são as médias dos efeitos.

2.4 Componentes de variância

Conforme Barbin (1993), os componentes de variância são as variâncias associadas aos parâmetros de efeitos aleatórios de um modelo estatístico.

Tem-se um efeito aleatório quando podemos fazer uma inferência sobre toda a população, dada uma amostra. Já para os efeitos fixos suas considerações são limitadas, isto é, são restritas aos tratamentos ensaiados.

No planejamento de experimentos, a escolha dos fatores e seus respectivos níveis (que farão parte de um experimento) é uma etapa fundamental, sendo basicamente um problema do pesquisador. Vejamos as definições dos modelos:

i) Modelo fixo: se todos os parâmetros do modelo, com exceção do erro, forem de efeitos fixos, o modelo é chamado fixo. Neste caso o interesse é estimar e testar as hipóteses de funções lineares dos efeitos.

ii) Modelo aleatório: se todos os parâmetros do modelo, com exceção da média, forem de efeitos aleatórios, o modelo é chamado aleatório. Neste caso o interesse é estimar componente de variância.

iii) Modelo misto: se o modelo envolver tanto parâmetros de efeitos fixos como de efeitos aleatórios, excetuando-se nestas considerações a média e o erro, o modelo é chamado misto.

Assim, para parâmetros de efeitos fixo, os resultados são válidos apenas para os tratamentos ensaiados, e para parâmetros de efeito aleatório os resultados são estendidos para toda a população de níveis da qual a amostra foi retirada.

Segundo Montgomery(2001), para o caso de efeitos aleatórios, para estimarmos as componentes de variância σ^2 , σ_β^2 e σ_τ^2 pode-se utilizar o método dos momentos via análise de variância. Sendo assim, a partir dos quadrados médios esperados obtém-se:

$$\hat{\sigma}^2 = QM_E$$

$$\hat{\sigma}_\beta^2 = \frac{QM_{B(A)} - QME}{n}$$

$$\hat{\sigma}_\tau^2 = \frac{QM_A - QM_{B(A)}}{bn}$$

Maiores detalhes do método veja Barbin (1993, pg. 97)

2.5 Experimento aninhado generalizado em m-estágios

Para o experimento aninhado generalizado em m-estágios se enquadra nos resultados mostrados até aqui. Para um melhor entendimento podemos observar a seguinte figura:

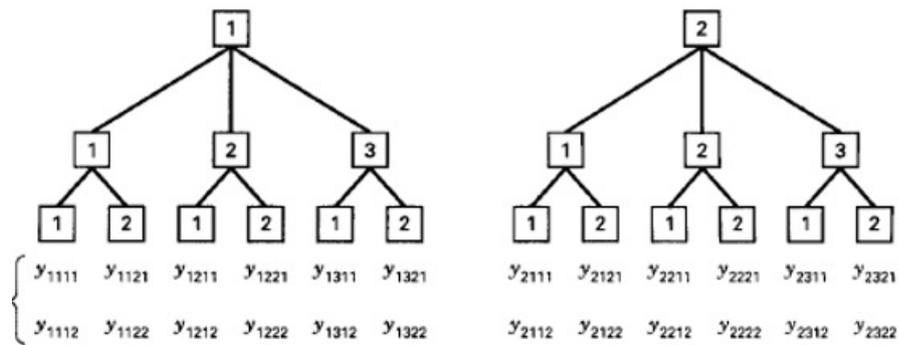


Figura 3 – Croqui aninhado de três fases com duas repetições.

Daí, observando a figura 3, tem-se por exemplo, duas formulações diferentes de um mesmo material onde retiramos três amostras, e de cada amostra são selecionados duas observações aleatoriamente para serem testadas e então são feitas duas medidas do material, por exemplo dureza. Logo, podemos ver que os níveis estão aninhados, de baixo pra cima, sempre um nível inferior ao do seu fator. Sendo assim, dizemos que este é um delineamento aninhado em três estágios, com duas réplicas. O modelo para o delineamento aninhado em três estágios é:

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

Para o exemplo acima tem-se:

τ_i é o efeito do i -ésima formulação; $\beta_{j(i)}$ é o efeito da j -ésima amostra dentro da i -ésima formulação; $\gamma_{k(ij)}$ é o efeito da k -ésima observação dentro da j -ésima amostra e i -ésima formulação $\varepsilon_{k(ijk)l}$ é o erro usual $NID(0, \sigma^2)$.

A extensão deste modelo a m fatores é direta. Note que, no suposto exemplo acima, a variabilidade total consiste de três componentes resultantes da formulação, da

amostra e do erro analítico do teste. Logo, podemos dizer que este exemplo demonstra como o delineamento aninhado é, às vezes, usado no processo de análise para identificar a maior fonte de variabilidade na produção. Assim sendo, se a componente de variância da formulação é grande, então isto implica que a variabilidade total poderia ser reduzida utilizando apenas uma formulação. Os cálculos das somas de quadrados e a análise de variância do experimento aninhado em m-estágios são similares à análises apresentadas anteriormente. Note que elas são uma simples extensão das fórmulas para o experimento aninhado em dois estágios.

Tabela 3 – Análise de Variância para experimentos aninhados em três estágios

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrados Médios
A	a-1	$bcn \sum (\bar{y}_{i...}, -\bar{y}....)^2$	MS_A
B(dentro de A)	a(b-1)	$cn \sum_i \sum_j (\bar{y}_{ij..}, -\bar{y}....)^2$	$MS_{B(A)}$
C(dentro de B)	ab(c-1)	$n \sum_i \sum_j \sum_k (\bar{y}_{ijk.}, -\bar{y}....)^2$	$MS_{C(B)}$
Erro	abc(n-1)	$\sum_i \sum_j \sum_k \sum_l (\bar{y}_{ijkl}, -\bar{y}_{ijkl})^2$	MS_E
Total	abcn-1	$\sum_i \sum_j \sum_k \sum_l (\bar{y}_{ijkl}, -\bar{y}....)^2$	

Tabela 4 – Quadrados Médios Esperados para o Experimento Aninhado em Três Estágios com A e B Fixados e C Aleatório, pelo método de Hicks

Fator	F a	F b	R c	R n	Quadrado Médio Esperado
τ_i	0	b	c	n	$\sigma^2 + n\sigma_\gamma^2 + \frac{bcn \sum \tau_i^2}{a-1}$
β_j	1	0	c	n	$\sigma^2 + n\sigma_\gamma^2 + \frac{cn \sum \beta_{j(i)}^2}{a(b-1)}$
$\gamma_{k(ij)}$	1	1	1	n	$\sigma^2 + n\sigma_\gamma^2$
$\varepsilon_{(ijk)l}$	1	1	1	1	σ^2

As propriedades estatísticas do teste para qualquer efeito ou interação podem ser obtidos da inspeção da tabela 4. Para determinar as propriedades estatísticas do teste, nós devemos encontrar os quadrados médios esperados. Por exemplo, se os fatores A e B são fixados e o fator C é aleatório, então nós podemos derivar os quadrados médios esperados.

2.6 Razão - F

Conforme Quinn (2002) a razão - F é utilizada para testar a hipótese nula em um experimento aninhado para dois fatores, onde a mesma é determinada pelo quociente de duas quiquadrados. As hipóteses testas são, de modo geral:

$$H_0(A) : \tau_1 = \tau_2 = \dots = \tau_i = 0, \text{ não há efeito em qualquer nível do fator A;}$$

$H_0(B) : \beta_{1(1)} = \beta_{2(1)} = \dots = \beta_{j(i)} = 0$, não há efeito em qualquer nível específico de cada fator B dentro de qualquer nível de A.

Sendo assim, se $H_0(A)$ é verdadeira, então todos os τ_{is} são iguais a zero e QMA e $QMB(A)$ são ambas estimativas para $\sigma_\varepsilon^2 + \sigma_\beta^2$, por isso sua razão deve ser menor ou igual a um. Já se, $H_0(B)$ é verdadeira, então, não há variância adicional devido à diferenças entre os possíveis níveis do fator B dentro de cada nível de A, $n\sigma^2 = 0$, então, $QMB(A)$ e $QM_{residual}$ são ambas estimativas de σ^2 , por isso sua razão deve ser igual a um.

Para o exemplo apresentado no trabalho as hipóteses testadas são:

H_0 : os efeitos de fornecedores e genótipos não influenciam as características tecnológicas da fibra.

H_a : os efeitos de fornecedores e genótipos, pelo menos um, influenciam as características tecnológicas da fibra.

Tabela 5 – Razão-F para o teste da hipótese nula para o modelo ANOVA de dois fatores aninhados

Fonte de Variação	Razão-F A fixo e B aleatório	Razão-F A fixo e B fixo
A	$\frac{QM_A}{QM_{B(A)}}$	$\frac{QM_A}{QM_{residual}}$
B(A)	$\frac{QM_{B(A)}}{QM_{residual}}$	$\frac{QM_{B(A)}}{QM_{residual}}$
Resíduo	σ_ε^2	$QM_{residual}$

Estas razões - F seguem uma distribuição F-Fisher-Snedecor sobre as hipóteses de homogeneidade da variância e normalidade com uma exceção. Se B é fator aleatório e o número de réplicas observadas dentro de cada nível de B é variável (desbalanceados), então a razão - F obtida de $QM_{B(A)}$ e $QM_{residual}$ não segue uma distribuição F quando σ_β^2 é maior do que zero, porque $QM_{B(A)}$ não é distribuído como um múltiplo de uma distribuição qui-quadrado (Searle et al. 1992).

Quando B é um fator aleatório, $QM_{B(A)}$ compõe o denominador da razão-F para o teste do fator A, isto é, as unidades de replicação para testar os efeitos de A são as médias de B. Isto é uma consideração importante para o poder do teste para o fator A e o delineamento de experimentos baseados sobre modelos aninhados. Quando B é fixo, o QM estimado para A não inclui uma componente para B, então a razão-F para testar A usa $QM_{residual}$ como denominador. Se A é aleatório, as razões-F são as mesmas como se A é fixado.

Tabela 6 – Estimativas de componentes de variância para três fatores aninhados e com fator A aleatório, fator B dentro de A aleatório e com fator C dentro de B também aleatório. São realizadas n réplicas dentro de cada célula.

Fonte de Variação	Quadrado Médio Esperado	Estimativa da Componente de Variância	Razão-F
A	$\sigma_\varepsilon^2 + n\sigma_\gamma^2 + nc\sigma_\beta^2 + nbc\sigma_\tau^2$	$\frac{QM_A - QM_{B(A)}}{nbc}$	$\frac{QM_A}{QM_{B(A)}}$
B(A)	$\sigma_\varepsilon^2 + n\sigma_\gamma^2 + nbc\sigma_\tau^2$	$\frac{QM_{B(A)} - QM_{C(B(A))}}{nc}$	$\frac{QM_{C(B(A))}}{QM_{B(A)}}$
C(B(A))	$\sigma_\varepsilon^2 + n\sigma_\gamma^2$	$\frac{QM_{C(B(A))} - QM_{residual}}{n}$	$\frac{QM_{C(B(A))}}{QM_{residual}}$
Resíduo	σ_ε^2	$QM_{residual}$	

3 Material e métodos

Encontram-se nesta seção as principais metodologias que serviram de base para este trabalho, no que se refere a descrição dos dados e ao uso do tipo de delineamento utilizado.

Os dados utilizados nesta aplicação são referentes a um experimento desenvolvido pela Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA ALGODÃO, localizada em Campina Grande - PB, na cidade de Apodi – RN, safra 2014/2015, e teve como objetivo analisar se os efeitos de genótipos e fornecedores influenciam as características tecnológicas da fibra do algodão. O material do presente estudo refere-se a um conjunto de dados que foram conduzidos por meio de um delineamento experimental com classificação hierárquica com fatores balanceados, onde foram estudadas três variáveis resposta: comprimento de fibra - comp (mm), percentagem de fibra – pf(%) e resistência da fibra – res (gf/tex). Para cada fornecedor (Forn), temos quatro genótipos (gen), onde cada genótipo nos dá quatro repetições. Sendo assim, temos 16 amostras dentro de cada fornecedor. Conforme as tabelas abaixo:

Tabela 7 – Valores, em milímetros, da variável resposta comprimento de fibra

Fornecedor	Genótipo	Repetições			
		1	2	3	4
Forn 1	1	28,00	28,78	27,18	30,85
	2	32,39	29,74	29,26	31,84
	3	29,10	28,75	26,45	33,83
	4	29,92	27,32	30,55	29,30
Forn 2	5	30,97	28,79	34,15	33,37
	6	28,19	29,45	29,44	28,17
	7	29,28	31,13	30,31	30,87
	8	28,54	27,69	28,89	28,12
Forn 3	9	28,63	29,59	32,01	29,93
	10	30,77	26,48	29,03	28,11
	11	29,15	28,35	32,83	29,15
	12	29,14	27,69	30,27	32,48
Forn 4	13	29,10	29,19	29,72	29,78
	14	32,48	31,70	29,79	27,68
	15	28,46	28,82	29,93	28,03
	16	30,03	28,28	29,65	29,12

Nesta seção, o procedimento estatístico utilizado foi o método dos momentos ou análise da variância. Neste estudo, foi considerado os componentes de variância no modelo com classificação hierárquica (dados balanceados). Foi realizado também o teste de Tukey

Tabela 8 – Valores, em %, da variável resposta percentagem de fibra

Fornecedor	Genótipo	Repetições			
		1	2	3	4
Forn 1	1	40,19	42,59	42,48	39,17
	2	37,60	39,61	41,78	38,16
	3	43,36	38,71	39,45	43,12
	4	38,28	45,10	46,80	42,30
Forn 2	5	45,30	44,70	46,50	42,30
	6	43,90	45,90	46,20	43,40
	7	46,90	46,80	45,40	44,00
	8	43,60	46,30	45,50	45,80
Forn 3	9	42,10	44,60	43,20	45,00
	10	46,00	43,60	43,80	44,60
	11	45,40	44,30	44,60	43,10
	12	42,90	44,20	45,80	45,70
Forn 4	13	43,60	46,10	46,60	45,60
	14	46,60	45,60	39,47	38,86
	15	41,80	38,54	42,17	40,16
	16	38,54	43,70	41,77	44,90

Tabela 9 – Valores, em gf/tex, da variável resposta resistência da fibra

Fornecedor	Genótipo	Repetições			
		1	2	3	4
Forn 1	1	33,52	30,29	30,32	30,10
	2	32,86	32,54	27,41	36,85
	3	30,86	32,01	32,64	30,51
	4	26,60	31,53	26,90	32,17
Forn 2	5	29,60	30,88	28,55	29,91
	6	31,74	32,46	34,98	29,55
	7	31,50	29,57	33,64	27,60
	8	33,49	32,38	30,61	32,82
Forn 3	9	33,33	36,01	31,91	28,14
	10	33,78	30,11	33,08	32,01
	11	30,11	33,08	34,33	30,16
	12	33,71	32,35	33,61	27,01
Forn 4	13	32,91	29,80	31,86	35,10
	14	34,01	34,75	34,80	31,65
	15	31,05	32,83	34,81	31,53
	16	30,14	30,60	35,13	28,29

no modelo misto, dado fator principal fixo. A ideia central dos modelos de componentes de variância é a dos efeitos, onde são classificados como efeito fixo ou efeito aleatório. Entretanto, sabemos que existem três modelos: fixo, aleatório e misto. Diante disso, cabe

ao pesquisador escolher o tipo de modelo que deseja estudar.

Os modelos escolhidos para serem trabalhados no exemplo foram o aleatório e o misto, onde seus modelos estatísticos estão representados abaixo:

i) Modelo aleatório

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

onde:

μ é a média geral;

τ_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 dentro do i-ésimo nível do fator fornecedor;

$\varepsilon_{(ij)k}$ é o erro casual não observável.

ii) Modelo misto

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

onde:

μ é a média geral;

τ_i é o efeito fixo do i-ésimo nível do fator fornecedor;

$\beta_{j(i)}$ é o efeito aleatório do j-ésimo nível do fator genótipo dentro do i-ésimo nível do fator fornecedor;

$\varepsilon_{(ij)k}$ é o erro casual não observável.

As análises foram realizadas através do sistema estatístico – computacional SAS versão 9.1.3, onde foram utilizados os procedimentos para análise de modelos lineares: o PROC GLM (Procedure for general linear models), permite que você especifique qualquer grau de interação em efeitos cruzados e efeitos aninhados e PROC VARCOMP (Variance Components Estimation Procedure), estima componentes de variância para um modelo linear geral.

4 Resultados obtidos

Nesta seção, serão relatados os resultados deste estudo. Abaixo, serão encontradas as tabelas da análise da variância para dos modelos aleatório e misto, os gráficos de comparação de médias (teste de Tukey) no modelo misto e os gráficos para análise dos resíduos.

4.1 Análise descritiva

Abaixo, serão encontradas as tabelas da análise da variância para dos modelos aleatório e misto, os gráficos de comparação de médias (teste de Tukey) no modelo misto e os gráficos para análise dos resíduos. A variabilidade presente em um ensaio é analisada com o auxílio da Tabela da Análise de Variância (Tabela 10). A análise da variância tem como um dos principais objetivos a obtenção da estatística F, onde a partir dela, faz-se conclusões acerca das hipóteses.

Tabela 10 – Análise da variância para o modelo aleatório, com suas respectivas Somas de Quadrados (SQ), Quadrados Médios (QM), estatística F, e valor p, para as variáveis comprimento de fibra, percentagem de fibra e resistência da fibra.

	Causas de Variação	G.L.	SQ	QM	F	Pr>F	E(QM)
Comp	Forn	3	1,06	0,35	0,08	0,09	$\sigma^2 + 4\sigma_{gen(forn)}^2 + 16\sigma_{forn}^2$
	Gen(forn)	12	50,89	4,24	1,57	0,13	$\sigma^2 + 4\sigma_{gen(forn)}^2$
	Erro	48	129,27	2,69			σ^2
	Total	63	181,22				
Pf	Forn	3	148,70	49,56	7,22	0,0050	$\sigma^2 + 4\sigma_{gen(forn)}^2 + 16\sigma_{forn}^2$
	Gen(forn)	12	82,40	6,86	1,60	0,12	$\sigma^2 + 4\sigma_{gen(forn)}^2$
	Erro	48	206,05	4,29			σ^2
	Total	63	437,16				
Res	Forn	3	21,28	7,09	1,52	0,26	$\sigma^2 + 4\sigma_{gen(forn)}^2 + 16\sigma_{forn}^2$
	Gen(forn)	12	56,05	4,67	0,85	0,59	$\sigma^2 + 4\sigma_{gen(forn)}^2$
	Erro	48	263,21	5,48			σ^2
	Total	63	340,55				

Considerando que os dados experimentais devem satisfazer algumas pressuposições, como observar o quadrado médio esperado, as hipóteses e a distribuição de probabilidade (distribuição – F), então podemos observar pelo valor do F calculado, para as variáveis comprimento de fibra e resistência da fibra, que não houve diferença significativa ao nível de 5% de significância, no entanto houve diferença significativa para a percentagem de fibra. Portanto, diante das hipóteses testadas há evidências para rejeitar-se H_0 , pois existe pelo menos um dos efeitos de fornecedores e genótipos influenciam as características

tecnológicas da fibra.

Tabela 11 – Análise da variância para o modelo misto, dado forn fixo, com as variáveis resposta comprimento de fibra, percentagem de fibra e resistência da fibra.

	Causas de Variação	G.L.	SQ	QM	F	Pr>F	E(QM)
Comp	Forn	3	1,06	0,35	0,08	0,09	$\sigma^2 + 4\sigma_{gen(forn)}^2 + 16\sigma_{forn}^2$
	Gen(forn)	12	50,89	4,24	1,57	0,13	$\sigma^2 + 4\sigma_{gen(forn)}^2$
	Erro	48	129,27	2,69			σ^2
	Total	63	181,22				
Pf	Forn	3	148,70	49,56	7,22	0,0050	$\sigma^2 + 4\sigma_{gen(forn)}^2 + 16\sigma_{forn}^2$
	Gen(forn)	12	82,40	6,86	1,60	0,12	$\sigma^2 + 4\sigma_{gen(forn)}^2$
	Erro	48	206,05	4,29			σ^2
	Total	63	437,16				
Res	Forn	3	21,28	7,09	1,52	0,26	$\sigma^2 + 4\sigma_{gen(forn)}^2 + 16\sigma_{forn}^2$
	Gen(forn)	12	56,05	4,67	0,85	0,59	$\sigma^2 + 4\sigma_{gen(forn)}^2$
	Erro	48	263,21	5,48			σ^2
	Total	63	340,55				

Analogamente, observando-se a tabela 11, para o modelo misto, dado forn fixo, que as conclusões são similares, porém foi feito o teste de Tukey, onde obtivemos os box-plots utilizado para avaliar a distribuição empírica do dados(figuras 4),que apresentam uma análise para cada um dos fornecedores. Pode-se ver que as médias amostrais foram, em alguns fornecedores, superiores aos das medianas e em outros inferiores. Tal fato, evidencia que a distribuição normal pode ser inadequada para esse modelo, e também que é possível violação da pressuposição de homogeneidade exigida na análise de variância. Podemos ainda observar, pela figura 4, que existem pontos fora dos limites superior e inferior nos box-plot, que podem ser outliers, essas observações merecem uma atenção maior e se, possível uma transformação para podermos concluir algo sobre elas.

Tabela 12 – Variáveis com suas devidas médias

Fornecedor	Comprimento de fibra (mm)	Percentagem de fibra (%)	Resistência da fibra (gf/tex)
Forn1	29,5788a	41,1687b	31,0694a
Forn2	29,8350a	45,1563a	31,2050a
Forn3	29,6006a	44,3063a	32,0456a
Forn4	29,4850a	42,7506ab	32,4538a

A tabela 12 deixa mais clara as conclusões que discutimos até aqui de que não existe diferença significativa entre as variáveis respostas comprimento da fibra e resistência da fibra, pois como elas possuem a mesma letra, significa que as médias não são significativamente diferentes. E com relação a percentagem de fibra, que tem letras diferentes, existe diferença entre forn1 e forn2 e forn1 e forn3, e em relação ao forn4 existe uma ambiguidade.

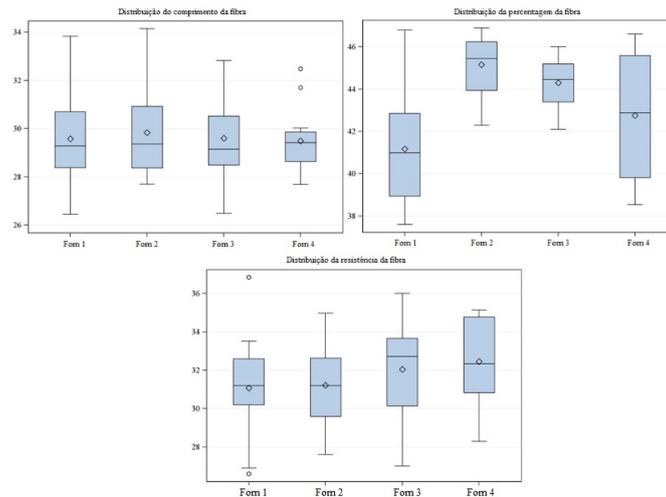


Figura 4 – Teste de Tukey para as médias do fator fornecedor para a variável comprimento da fibra, percentagem da fibra, acima, e resistência da fibra, abaixo, respectivamente.

4.2 Análise de resíduos

A fim de detectar algumas observações discrepantes nos modelos apresentados no trabalho e, também, verificar supostos desvios das pressuposições do modelo é necessário um estudo das análises de resíduos que estão apresentadas nas figuras 5, 6 e 7.

Diante dos diagnósticos de ajustes apresentados pode-se observar a não homogeneidade da variância, pois os resíduos devem ser distribuídos aleatoriamente em torno de zero e é possível perceber que existe uma correlação entre eles, não deixando dúvidas da existência de uma dependência. No entanto, o problema de falta de independência pode ser devido a falta de ajuste no modelo. Da mesma maneira, identificamos valores extremos, atípicos que são supostamente outliers que não são bem ajustadas pelo modelo e possivelmente precisam de uma transformação para podermos concluir algo sobre elas.

Como pode-se perceber as observações acima quebram os pressupostos do modelo, no entanto o gráfico da normalidade nos deixa dúvidas sugerindo que a modelo segue aproximadamente uma distribuição normal, porém a falta de homogeneidade, geralmente, vem junto a falta de normalidade

Diagnósticos de ajuste para comprimento

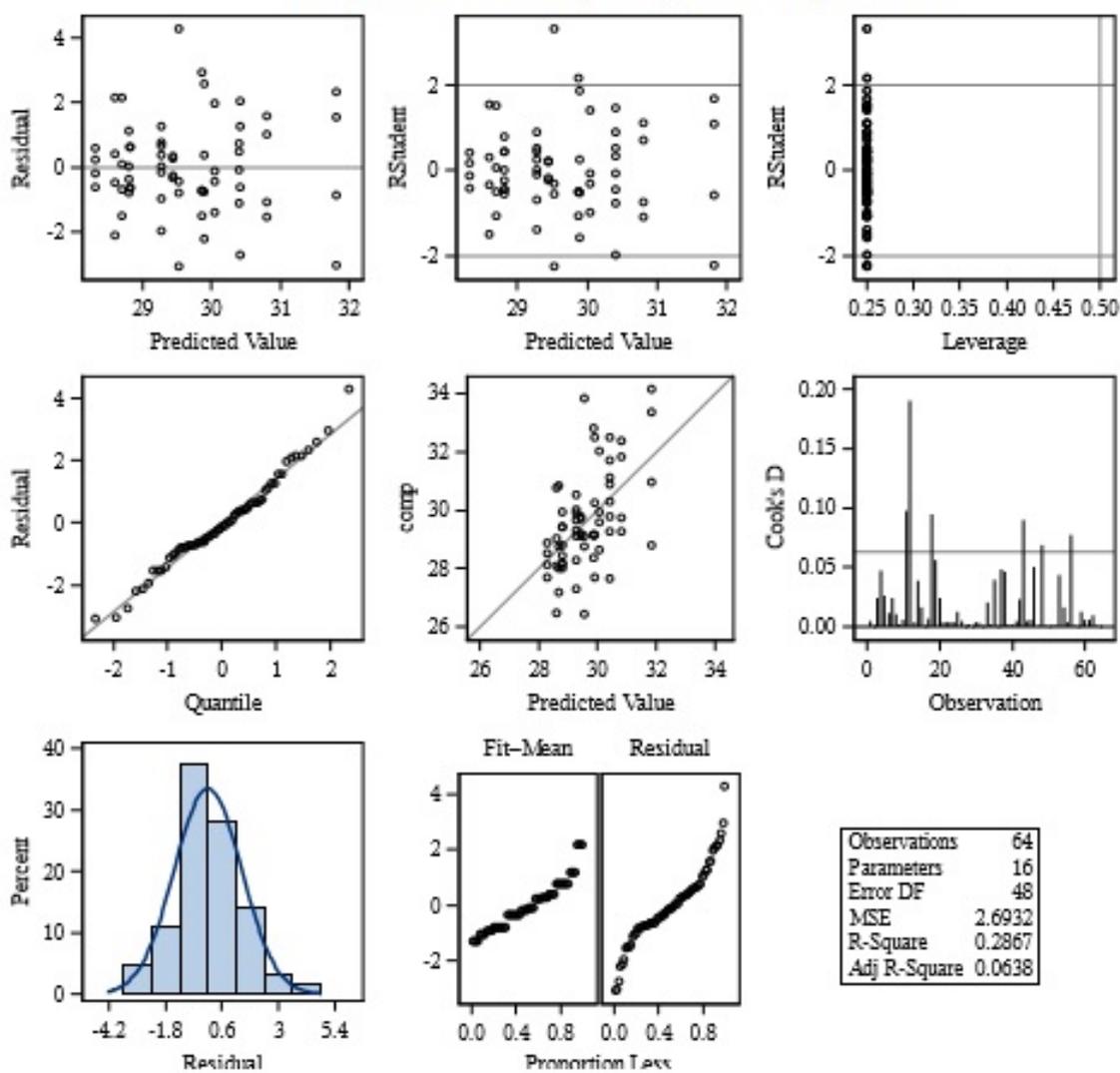


Figura 5 – Análise dos resíduos para a variável comprimento da fibra

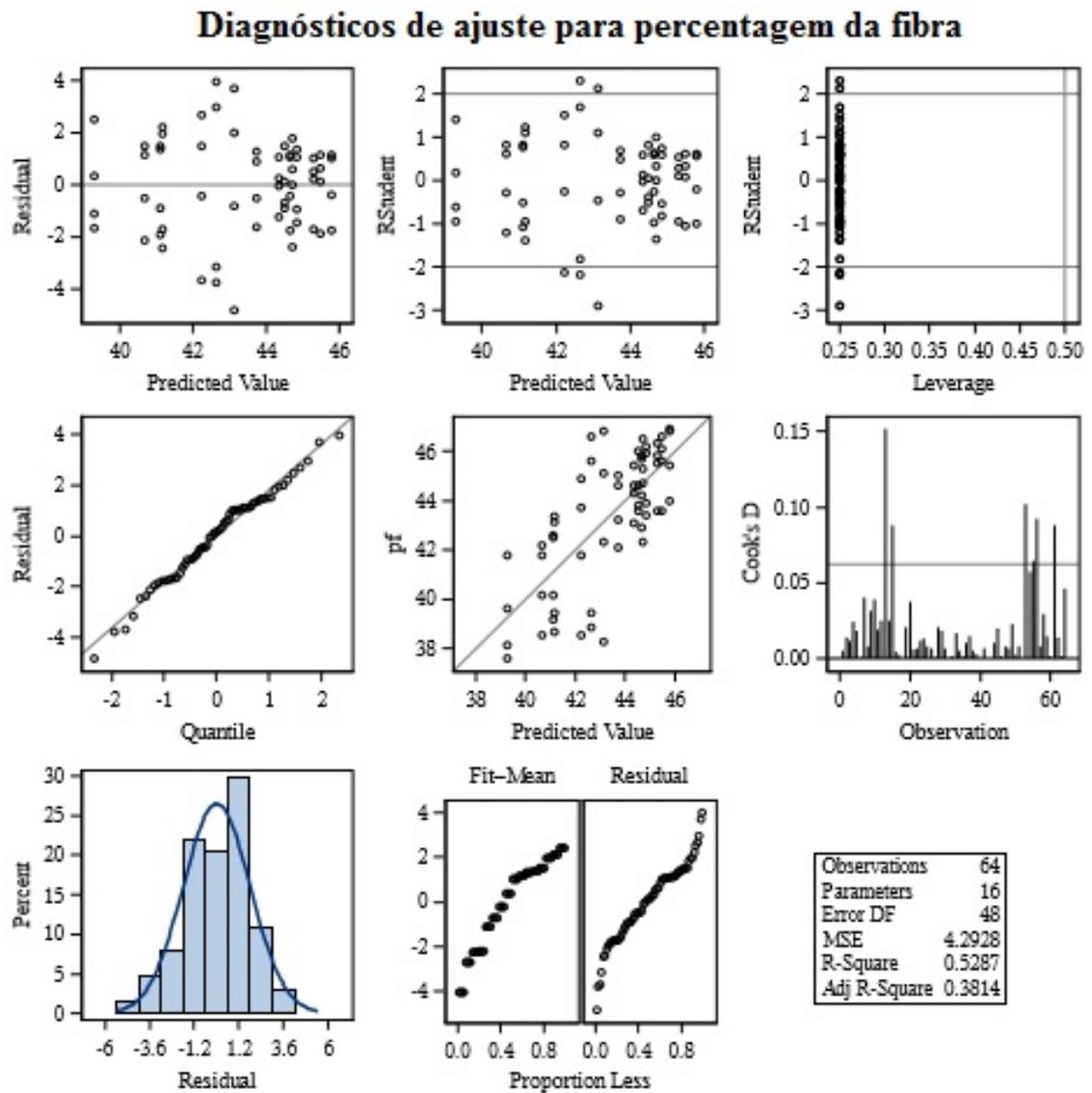


Figura 6 – Análise dos resíduos para a variável percentagem da fibra

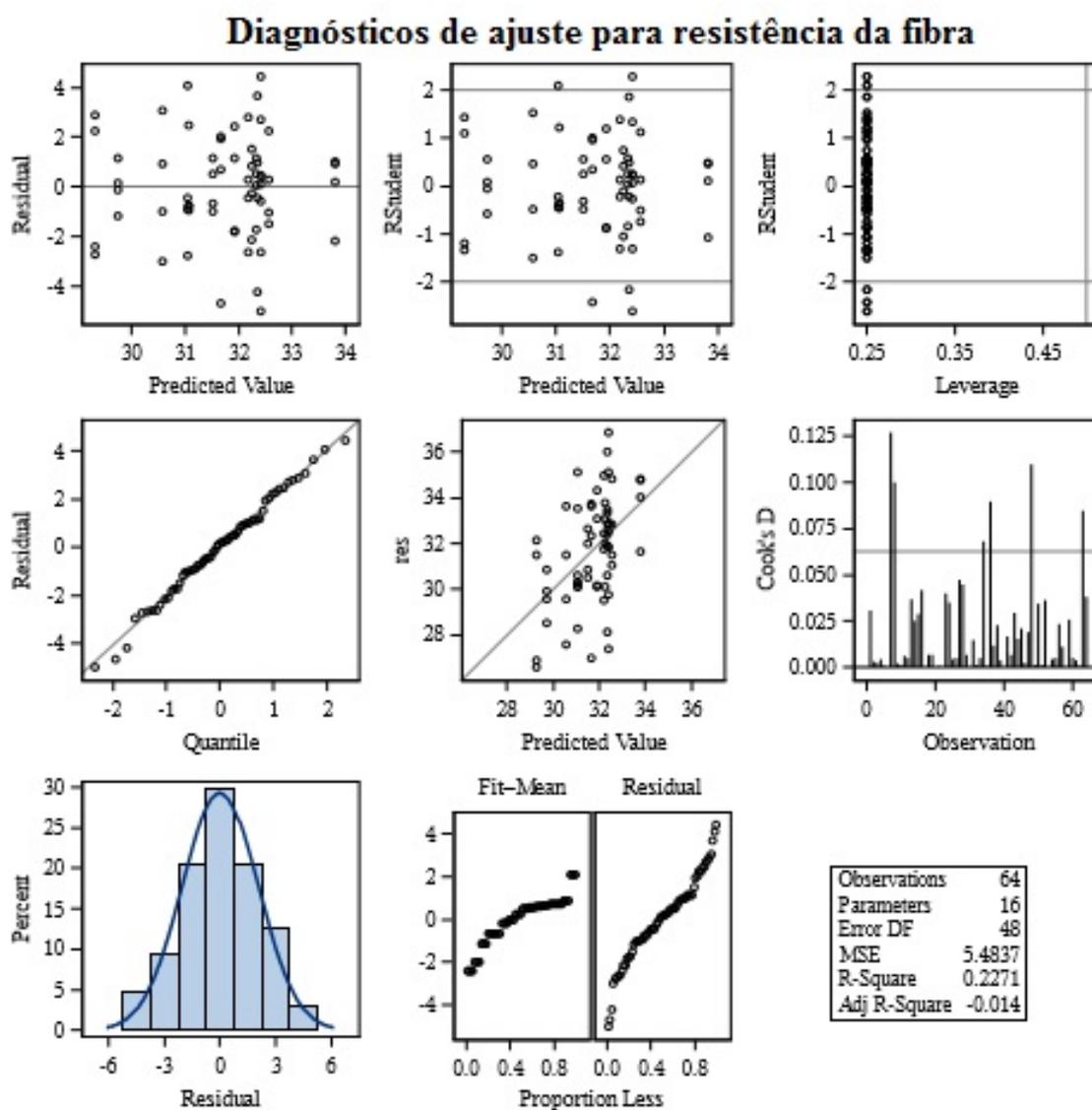


Figura 7 – Análise dos resíduos para a variável resistência da fibra

5 Conclusão

Neste trabalho foi apresentado um estudo teórico e prático do modelo experimental aninhado (dados balanceados) com uma aplicação em dados de algodão. O banco de dados foi aplicado em apenas dois modelos, misto e aleatório. Nos modelos apresentados foi aplicado o método dos momentos para ambos, o teste de Tukey apenas no modelo misto, dado seu fator principal – fornecedor fixo, e foi feita também uma análise de resíduos. Após aplicação do método e do teste foi possível concluir que rejeitaríamos H_0 , pois o estudo foi feito avaliando se os efeitos de fornecedores e genótipos influenciariam as características tecnológicas da fibra do algodão. As variáveis resposta avaliadas foram comprimento da fibra, percentagem da fibra e resistência da fibra, onde vimos que comprimento de fibra e resistência da fibra não apresentaram diferença significativa ao nível de 5% de significância, no entanto a percentagem de fibra apresentou diferença entre os fornecedores, fazendo assim a não aceitação da hipótese nula.

O estudo feito da análise dos resíduos foi de grande importância, pois só assim verificamos que os pressupostos do modelo não foram satisfeitos, vimos a não homogeneidade dos resíduos, a existência de correlação e a dependência dos resíduos, fazendo-nos perceber que o modelo apresentado precisa de ajustes e de transformações para podermos ter respostas mais claras, precisas.

5.1 Perspectivas de trabalhos futuros

Abaixo foi descrito algumas sugestões de futuros trabalhos que podem ser realizados para complementar o tema estudado:

- 1) Fazer o estudo do experimento aninhado escalonado e experimento aninhado com mais de dois estágios
- 2) Fazer um estudo do experimento aninhado para dados desbalanceados
- 3) Fazer um estudo do experimento misto (fatorial e aninhado).
- 4) Fazer um estudo sobre ambiguidade de dados.
- 5) Usar uma transformação nos dados apresentados, para poder ter-se uma melhor explicação do comportamento dos dados.

6 Referências

- BARBIN, Décio. Componentes de Variância: teoria e aplicação (2a Ed. rev. e amp.) - Piracicaba, 1993.
- BELGROVE, A., CLAYTON, M. N. & Quinn, G. P. Effects of secondarily treated sewage effluent on intertidal macroalgal recruitment processes. *Marine and Freshwater Research* 48: 137-146, 1997.
- CASELLE, J. E. & WARNER, R. R. Variability in recruitment of coral reef fishes: the importance of habitat at two spatial scales. *Ecology* 77: 2488-2504, 1996.
- JOHNSTON, J. M, & PENNYPACKER, H. S, (1993a). Strategies and tactics of behavioral research (2a Ed.). Hillsdale, NJ: Lawrence Erlbaum.
- MONTGOMERY, Douglas C. Design and analysis of experiments (5a Ed). John Wiley & Sons, 2001.
- QUINN, Gerry P.; KEOUGH, Michael J. Experimental design and data analysis for biologists. Cambridge University Press, 2002.
- SANTOS, J. W. dos; ALMEIDA, F. de A. C.; BELTRÃO, N. E. de M.; CAVALCANTI, F. B. Estatística Experimental Aplicada. 2 ed. Revisada e Ampliada. Campina Grande; Embrapa/Algodão/ Universidade Federal de Campina Grande, 2008.
- SAS/STAT® 9.1 User's Guide. Disponível em: https://support.sas.com/documentation/onlinedoc/91pdf/sasdoc_91/stat_ug_7313.pdf.
- SILVA, J. G. C. da; Estatística Experimental: Planejamento de Experimentos - Versão Preliminar. Pelotas; Universidade Federal de Pelotas/ Instituto de Física e Matemática/ Departamento de Matemática e Estatística, 2007.
- SPIEGEL, Murray R. Théorie et applications de la statistique. Paris: McGrawHill, 1982. 580p.
- TWOMBLY, S. Timing of metamorphosis in a freshwater crustacean: comparison with anuran models. *Ecology* 77: 1855-1866, 1996.