

**Universidade de São Paulo  
Escola Superior de Agricultura “Luiz de Queiroz”**

**Modelagem para probabilidade de frutificação em café *Arábica* baseado  
em ocupação de metâmeros**

**Luiza Yoko de Barros**

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

**Piracicaba  
2017**

**Luiza Yoko de Barros  
Bacharel em Estatística**

**Modelagem para probabilidade de frutificação em café *Arábica* baseado  
em ocupação de metâmeros**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientadora:

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

Dissertação apresentada para obtenção do título de Mestra  
em Ciências. Área de concentração: Estatística e Experi-  
mentação Agronômica

**Piracicaba  
2017**

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

Barros, Luiza Yoko de

Modelagem para probabilidade de frutificação em café *Arábica* baseado em ocupação de metâmeros / Luiza Yoko de Barros. -- versão revisada de acordo com a resolução CoPGr 6018 de 2011. -- Piracicaba, 2017 .

41 p.

Dissertação (Mestrado) -- USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Café *Arábica* 2. Metâmero 3. Regressão logística 4. Dados longitudinais 5. Modelos mistos I. Título.

## AGRADECIMENTOS

Meus agradecimentos vão a todos os colegas, professores, familiares e amigos que acompanharam essa trajetória, contribuindo para concretização deste trabalho. Em especial:

À minha orientadora Professora Dra. Sônia e à Dra. Miroslava, Embrapa - Campinas, pela sugestão deste tema, pela disposição e competência na condução deste trabalho;

Aos professores do departamento de Matemática e Estatística da ESALQ-USP, pelos ensinamentos;

Aos meus pais Maria Cristina Shimako Shimabukuro, Laécio Carvalho de Barros e meu irmão Otávio Akira de Barros, pelo amor, dedicação, incentivo, apoio financeiro e presença constante em todas as etapas da minha vida;

Ao meu eterno amigo Rafael pelo companheirismo;

Ao Estevão, pelo apoio nas mais diversas dúvidas envolvendo, principalmente os softwares R e Latex; à Profa. Renata, pela ajuda em partes importantes do trabalho e à minha turma da pós graduação em especial ao meu amigo Rick, que por diversas vezes me ajudou como um verdadeiro amigo;

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudos;

Ao Denys pelo apoio emocional e à Aline pela amizade, que foram cruciais nessa trajetória;

Aos funcionários do departamento de Matemática e Estatística da ESALQ-USP, pela prontidão e atenção dispensada;

Finalmente, a todos que de alguma forma contribuíram direta ou indiretamente para a realização deste trabalho.

## SUMÁRIO

RESUMO . . . . .	5
ABSTRACT . . . . .	6
LISTA DE FIGURAS . . . . .	7
LISTA DE TABELAS . . . . .	8
1 INTRODUÇÃO . . . . .	9
1.1 Objetivos . . . . .	9
1.2 Café Arábica e Robusta . . . . .	9
2 REVISÃO BIBLIOGRÁFICA . . . . .	11
2.1 Modelagem Estatística . . . . .	11
2.2 Modelos para dados longitudinais . . . . .	12
2.2.1 Algumas estruturas de covariâncias . . . . .	13
3 MATERIAL . . . . .	15
3.1 Arquitetura de planta e Metâmero . . . . .	15
3.2 Bancos de dados . . . . .	16
3.3 Análise exploratória dos dados . . . . .	17
3.3.1 Auxílio no desenvolvimento de equações para alometria e assimetria . . . . .	18
3.4 Estudo das ordens de ramificação utilizando retas de regressão . . . . .	23
4 MÉTODOS . . . . .	27
4.1 Análise exploratória dos dados para os tempos 1 e 2 . . . . .	27
4.2 Modelagem da probabilidade para cada tempo . . . . .	29
4.2.1 Modelagem da probabilidade para o tempo 1 . . . . .	30
4.2.2 Modelagem da probabilidade para o tempo 2 . . . . .	31
4.3 Modelos mistos para os dois tempos estudados . . . . .	32
5 DISCUSSÃO E CONCLUSÃO . . . . .	35
REFERÊNCIAS . . . . .	37
APÊNDICES . . . . .	39

## RESUMO

### **Modelagem para probabilidade de frutificação em café *Arábica* baseado em ocupação de metâmeros**

Este é um trabalho proveniente de vários encontros com pesquisadores da Embrapa de Campinas, cujo objetivo principal se centrou em investigar a probabilidade de frutificação em árvores de Café *Arábica*. Para isso foram analisados 20 bancos de dados para árvores de Café *Arábica* de um cultivar localizado no Instituto Agronômico do Paraná em dois momentos distintos (Junho de 2010 e em Novembro - Dezembro de 2010) e consideradas as seguintes variáveis: “OCUPAÇÃO” (quadrática ou retangular), “ESPAÇAMENTO” (6.000 plantas/ha e 10.000 plantas/ha) ambas relativas ao cultivo das plantas e “TAMANHO DO ENTRENÓ” (medição em cm de uma partição definida da planta). Na busca de um modelo representativo de tal fenômeno, foram estudadas paralelamente tópicos relativos a alometria e assimetria dessas mesmas plantas, os quais permitiram modelar determinadas associações entre algumas estruturas como largura e comprimento de folhas. Os modelos ajustados apresentaram uma grande significância para a variável “ESPAÇAMENTO” nos dois tempos estudados, enquanto que a variável “OCUPAÇÃO” foi significativa apenas no segundo tempo e variável “TAMANHO DO ENTRENÓ” não foi significativa para nenhum dos tempos. A metodologia adotada para investigar essa probabilidade se deu através dos modelos de regressão logística. Com o intuito de agregar a variável “TEMPO”, juntou-se os dois bancos de dados em diferentes tempos e, baseado na metodologia de modelos mistos, obteve-se um modelo ajustado com retas paralelas, onde apenas a variável “ESPAÇAMENTO” foi considerada significativa.

**Palavras-chave:** Café *Arábica*; Metâmero; Regressão logística; Dados longitudinais; Modelos mistos

## ABSTRACT

### Modeling for probability of fruiting in *Arabica* coffee based on occupancy metameres

This work is based on several meetings with Embrapa researchers from Campinas in an integrated study with professors from the ESALQ Department of Statistics and Agronomic Experimentation, whose main objective was to investigate the probability of fruiting in *Arabica* Coffee trees. In order to do so, it was analyzed 20 databases for Coffee trees of a cultivar located at the Agronomic Institute of Paraná at two different times (June 2010 and November - December 2010), where the following variables were considered: "OCCUPATION" (quadratic or rectangular), "SPACING" (6,000 plants / ha and 10,000 plants / ha), both related to plant cultivation and "SIZE OF ENTRENÓ" (measured in cm of a defined plant partition). In order to find a representative model of this phenomenon, topics related to allometry and asymmetry of these same plants were studied in parallel, which allowed to model certain associations between some structures such as width and length of leaves. The adjusted models presented a great significance for the variable "SPACING" in the two studied times, whereas the variable "OCCUPATION" was significant only in the second time and the variable "SIZE OF THE ENTRENÓ" was not significant for any of the times. The methodology used to investigate this probability was based on logistic regression models. In order to aggregate the variable "TIME", the two databases were combined at different times and, based on the methodology of mixed models, a model adjusted with parallel lines was obtained, where only the variable "SPACING" was considered significant.

**Keywords:** *Arabica* Coffee; Metamer; Logistic regression; Longitudinal data; Mixed models

## LISTA DE FIGURAS

1.1	Arvore de café com frutos. . . . .	10
3.1	Esquema que reúne metâmeros em uma árvore (TEAM DIGIPLANTE, 2016) . . . . .	16
3.2	Número de folhas relativo à sua orientação no espaço. Norte(N), Leste(L), Sul(S), Oeste(O). . . . .	18
3.3	Número de folhas relativo a ordem de ramificação . . . . .	19
3.4	Número de folhas pela inserção no ortotrópico (considerando todo o conjunto de plantas) . . . . .	19
3.5	Alometria de comprimento e largura de folhas (em cms) . . . . .	20
3.6	Comprimento dos eixos da 2ª ordem dependendo da sua inserção no eixo ortotrópico (1ª ordem) e orientação cardinal no espaço. . . . .	20
3.7	Número de frutos formados nos eixos da 2ª ordem dependendo da sua inserção no eixo ortotrópico (1ª ordem) e orientação cardinal no espaço. . . . .	21
3.8	Número de ramificações formados nos eixos da 2ª ordem dependendo da sua inserção no eixo ortotrópico (1ª ordem) e orientação cardinal no espaço. . . . .	21
3.9	Alometria do comprimento (em cms) entre as duas folhas do par. . . . .	22
3.10	Alometria do largura (em cms) entre as duas folhas do par. . . . .	22
3.11	Modelo de regressão nulo para o cálculo de filocrono . . . . .	23
3.12	Modelo de retas coincidentes para o cálculo de filocrono . . . . .	24
3.13	Modelo de intercepto comum para o cálculo de filocrono . . . . .	24
3.14	Modelo de retas paralelas para o cálculo de filocrono . . . . .	25
3.15	Modelo de retas concorrentes para o cálculo de filocrono . . . . .	25
3.16	Fluxograma cronológico de cada etapa realizada neste trabalho. . . . .	26
4.1	Frutificação por tipo de Ocupação para $t = 1$ . . . . .	27
4.2	Frutificação por tipo de Ocupação para $t = 2$ . . . . .	28
4.3	Frutificação por tipo de Espaçamento para $t = 1$ . . . . .	28
4.4	Frutificação por tipo de Espaçamento para $t = 2$ . . . . .	29
4.5	Tamanhos dos entrenós estudados. . . . .	29



## LISTA DE TABELAS

4.1	Proporção de metâmeros que frutificaram por plantas para t=1 . . . . .	30
4.2	Proporção de metâmeros que frutificaram por plantas para t=2 . . . . .	30
4.3	Estimativas dos Coeficientes e seus respectivos p-valores do modelo de regressão logístico apenas com a variável “ESPAÇAMENTO” relativo ao tempo t=1 . . . . .	31
4.4	Estimativas dos Coeficientes e seus respectivos p-valores do modelo de regressão logístico apenas com a variável “OCUPAÇÃO” relativo ao tempo t=1 . . . . .	31
4.5	Estimativas dos Coeficientes e seus respectivos p-valores do modelo de regressão logístico apenas com a variável “TAMANHO DO ENTRENÓ” relativo ao tempo t=1 . . . . .	31
4.6	Estimativas dos Coeficientes e seus respectivos p-valores do modelo de regressão logístico com as variáveis “ESPAÇAMENTO” e “OCUPAÇÃO” relativo ao tempo t=2 . . . . .	32
4.7	Estimativas dos Coeficientes e seus respectivos p-valores do modelo de regressão logístico apenas com a variável “ESPAÇAMENTO” relativo ao tempo t=2 . . . . .	32
4.8	Estimativas dos Coeficientes e seus respectivos p-valores do modelo de regressão logístico apenas com a variável “OCUPAÇÃO” relativo ao tempo t=2 . . . . .	32
4.9	Estimativas dos Coeficientes e seus respectivos p-valores do modelo de regressão logístico apenas com a variável “TAMANHO DO ENTRENÓ” relativo ao tempo t=2 . . . . .	32
4.10	Estimativas dos Coeficientes e seus respectivos p-valores do modelo de regressão logístico com as variáveis “ESPAÇAMENTO” e “OCUPAÇÃO” relativo ao tempo t=2 . . . . .	33
4.11	Estimativas dos Coeficientes e seus respectivos p-valores do modelo logístico com retas paralelas e inclusão do efeito aleatório referente às plantas . . . . .	33
4.12	Estimativas dos Coeficientes e seus respectivos p-valores do modelo logístico com retas concorrentes e inclusão do efeito aleatório referente às plantas . . . . .	33
4.13	Estimativas dos Coeficientes e seus respectivos p-valores do modelo logístico com retas com o mesmo intercepto e coeficientes angulares diferente e inclusão do efeito aleatório referente às plantas . . . . .	33
4.14	Estimativas dos Coeficientes e seus respectivos p-valores do modelo logístico sem o efeito de tratamento e inclusão do efeito aleatório referente às plantas . . . . .	34

## 1 INTRODUÇÃO

Este trabalho foi motivado por diversos encontros com a Dra. Miroslava da Embrapa de Campinas, os quais foram largamente estudados os processos de frutificação em árvores de cafés *Arábica*. Após repetidas medições periódicas das mesmas unidades amostrais presentes nos bancos de dados, surgiu um questionamento importante: qual seria a probabilidade de frutificação de uma determinada estrutura das plantas de café frutificarem considerando algumas características do plantio, ou até mesmo da própria planta? A metodologia adotada por nós, conjuntamente com a orientadora, foi investigar a probabilidade de frutificação por meio dos modelos de regressão logístico.

### 1.1 Objetivos

Os objetivos deste trabalho são:

1. realizar estudos sobre surgimentos de metâmeros;
2. modelar o crescimento de uma árvore do café *Arábica* baseando-se no número de metâmeros;
3. utilizar modelagem estatística para estudar o crescimento de uma árvore;
4. a partir dos dados, propor um modelo de regressão logístico para prever a probabilidade de frutificação de uma determinada estrutura das plantas de café estudadas no experimento.

### 1.2 Café Arábica e Robusta

O café é uma planta exótica que pertence à família das Rubiáceas, do gênero *Coffea*, que engloba diversas espécies (cerca de 60), das quais apenas duas são cultivadas e comercializadas: a *Coffea Arabica* e *Coffea Canephora*, conhecido como Robusta (no Brasil, Conilon).

A espécie *Coffea Arabica* foi classificada apenas em 1753. Ela é rica em aroma, muito perfumada, doce e ligeiramente ácida. Originária da Etiópia, a planta é bastante delicada e necessita de tratamentos frequentes. Os grãos, antes de torrados, têm forma ovoide alongada, de cor verde azulada e apresentam um sulco pouco pronunciado e sinuoso. O habitat ideal de cultivo situa-se entre 600 e 2000 metros de altitude e, nessa faixa quanto maior for a altitude, mais acentuadas são as qualidades organolépticas (perceptíveis aos sentidos humanos) dos grãos. Os maiores cultivadores desta espécie são os países da América do Sul e Central além de alguns países da África e Ásia. Estes grãos possuem numerosas variedades, como Bourbon, Catimorra, Mundo Novo, Caturra, Catuai, entre outros. Hoje, o Arábica representa três quartos da produção mundial de café. A taxa de cafeína deste tipo de grão é de cerca de 1,4%.

Já a espécie *Coffea Canephora* (Robusta) foi descoberta e classificada no final de 1800. Esta espécie é muito difundida na África, Ásia, Indonésia e Brasil. Ela é bem adaptada em altitudes compreendidas entre o nível do mar e 600 metros; resiste bem em condições climáticas quente e úmidas. Tem um crescimento rápido, bom rendimento e apresenta boa resistência aos parasitas. Floresce várias vezes por ano e, por isso, sua produção por planta é ligeiramente superior a do Arábica. Esse tipo contribui com cerca de um quarto da produção mundial. Antes de torrados os grãos têm forma arredondada, coloração amarelo esverdeada e contém, em média, cerca de 2,5% de cafeína. Depois de torrado, a bebida é encorpada e apresenta sabor achocolatado e persistente.

A bebida do café, tal como conhecemos, é produzida a partir dos frutos destas duas espécies, cujas combinações produzem sabores e aromas bem distintos.



**Figura 1.1.** Árvore de café com frutos.

Segundo (ANDRADE, 2013), dentre as cultivares disponíveis para o plantio no Brasil a Mundo Novo ocupa lugar de destaque, apresentando boas produções em quase todas as regiões cafeeiras do Brasil (FAZUOLI *et al.*, 2005). Essa denominação é porque as primeiras seleções que deram origem à cultivar foram realizadas em 1943 em que pesquisadores do Instituto Agrônomo de Campinas (IAC) foram informados de uma lavoura produtiva e rústica da cultivar Sumatra na Fazenda Aparecida, no município de Mundo Novo, hoje Urupês, na região araraquarense do estado paulista.

O café arábico (*Coffea Arabica*) é originário das terras altas de florestas Afromontane do Sudoeste Etiópia e Boma Planalto, região do Sudão. Considerando duas espécies diploides parentais putativos: *C. Canephora* e *C. Eugenioides*. Os ciclos de produção do café arábica são cerca de 4-5 anos na monocultura e 6-7 em sistemas tradicionais. Qualquer que seja o sistema de cultivo, a produção do cafeeiro é regulada pela poda, que tem duas funções essenciais: manter um rendimento suficiente durante toda a vida da árvore e permitir que as bagas permanecem facilmente acessível. A Figura 1.1. representa uma árvore de café.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 Modelagem Estatística

Podemos dizer que a modelagem estatística faz parte de um tema maior que é a modelagem matemática. A principal função dos modelos matemáticos é procurar descrever a realidade de algum fenômeno natural. No caso de estatística aplicada, por considerar fortemente experimentos planejadas, tem-se a resposta procurada em função de fatores previamente escolhidos pelo pesquisador além de suas interrelações (efeitos de interação ou hierarquizados). Outro aspecto importante na modelagem com experimentos é a especificação da natureza dos efeitos presentes no modelo estatístico, que pode ser fixo ou aleatório. A opção de como o efeito será considerado, em certas ocasiões, pode não ser tarefa fácil, no sentido de ter-se pouco controle. No entanto, tão importante quanto os aspectos teóricos, devemos ponderar o contexto da pesquisa, pois ao se considerar um efeito como aleatório no experimento, a inferência é ampliada para a população de referência, enquanto que para efeitos fixos, a inferência é restrita aos níveis do fator em estudo. No contexto da agricultura, (DE RESENDE, 2007) afirma que para fins de melhoramento genético, efeitos de genótipos devem ser considerados como aleatórios. O autor argumenta que a seleção durante as etapas de avaliação e, principalmente, a recomendação de cultivares aos agricultores foca o desempenho futuro das plantas. Ou seja, como esse genótipo se comportará quando for cultivado. Contrariamente, os fatores referentes aos efeitos ambientais devem ser preferencialmente considerados como fixos. Se o interesse reside no efeito genético, caso o fator de blocagem seja considerado aleatório, por exemplo, esse terá seu efeito penalizado pelo regressor. Consequentemente se adicionarão ruídos à estimativa do valor genético. Além disso, os efeitos de blocos, colheitas e ambientes dificilmente se repetirão em ocasiões futuras. Dessa forma, tais fatores devem ser considerados aleatórios desde que tenham pelo menos dez níveis. E também, quando os blocos forem incompletos, pois existe informação disponível a ser recuperada entre blocos dentro da repetição (PIEPHO *et al.*, 2004) e (SMITH *et al.*, 2005). Vários autores têm apresentado uma discussão conceitual sobre essa questão (EISENHART, 1947) e (SEARLE, 1991). Existem três tipos de modelagem com foco na seleção. A primeira, denominada tipo I ou modelo fixo, considera todos os fatores de efeito fixo, exceto o erro. O tipo II, modelo aleatório, apenas a constante populacional é de efeito fixo. Já a modelagem do tipo III é mista, coexistindo fatores de efeito fixo e aleatório, além da constante e do erro aleatório (HENDERSON, 1975). Baseando-se em (ALCARDE, 2012) e (DA COSTA, 2003) a Seção 2.2 trás uma rápida dedução dos modelos do tipo III.

Um aspecto fundamental para fins de melhoramento, independente do modelo empregado, é a busca por estimadores ou preditores que resultem em estimativas ou previsões mais próximas possíveis dos verdadeiros valores dos genótipos (*breeding values*) em teste (WHITE and HODGE, 1988). Para a seleção sob modelo fixo, temos as médias fenotípicas ajustadas dos genótipos, que se constituem nos melhores estimadores lineares não viesados dos breeding values (BLUE). Nesse caso, aplicam-se os métodos estatísticos dos mínimos quadrados, que se constitui na base para o método da análise de variância. Para a seleção sob modelo aleatório ou misto tem-se o método geral proposto por (HENDERSON, 1950), no qual a seleção dos genótipos seria pautada com base no preditor BLUP, que também se constitui numa função linear dos fenótipos, de elevada acurácia e não viciado. Mas, para que o BLUP se constitua absolutamente eficiente dentre os preditores, é preciso que as variâncias associadas aos efeitos aleatórios ou componentes de variância sejam conhecidas. Como na prática dispõe-se apenas de estimativas para essas variâncias, tais previsões são denominadas de Empirical-BLUP (LITTELL, 2002).

No método estatístico geral de seleção sob modelo misto de Henderson, a predição do efeito de genótipo, no qual centra-se, em geral, o interesse do melhorista, é efetuada uma penalização (encolhimento) da performance média fenotípica do genótipo, ajustada para os demais efeitos ambientais controlados presentes no experimento. O primeiro trabalho conhecido que utilizou tal preditor foi o de

(LUSH, 1931). Dentre elas pode-se citar o preditor de Bayes, o índice multifeitos (LUSH, 1947) o estimador de James-Stein (EFRON and MORRIS, 1977). Atualmente, o preditor de Henderson - BLUP - é o mais difundido. Para se proceder a modelagem seguem-se dois passos. Primeiro modela-se a média dos efeitos fixos e posteriormente a covariância dos efeitos aleatórios. Modelar a média significa construir as matrizes de incidência dos fatores presentes no modelo de maneira a obter mínima diferença entre o valor predito pelo modelo e o observado. Para os fatores aleatórios modela-se também a estrutura de covariância. Esse processo nada mais é que o modo como as variâncias e covariâncias irão se comportar ao predizer os efeitos da variável aleatória resposta. Ambas as modelagens são de fundamental importância na experimentação e necessitam ter bom ajustamento e parcimônia (VALLEJO *et al.*, 2011). O ajustamento do modelo é verificado pela verossimilhança, a qual informa a capacidade do modelo em interpretar a realidade, de ser verossímil. Ou seja, indica a probabilidade de adequação do modelo escolhido aos dados. Seguindo-se esse raciocínio o incremento no número de parâmetros ao modelo conduz aumento da verossimilhança. Quanto mais parâmetros, melhor o ajustamento. Porém a questão não é unilateral. O excesso de parâmetros a serem estimados compromete a capacidade preditiva do modelo. O problema agrava-se quanto menor a quantidade de dados, ou equações, disponíveis na estimação. Para contornar esse problema, alguns pesquisadores desenvolveram os chamados critérios de informação. Esses critérios penalizam a verossimilhança em função do número de parâmetros a serem estimados e, dependendo do critério, também em relação ao tamanho amostral (AKAIKE, 1974) (BURNHAM and ANDERSON, 2004) (SCHWARZ *et al.*, 1978).

A seguir será enunciado o modelo linear misto para dados longitudinais. As deduções são baseadas, principalmente, em (ALCARDE, 2012) e (DA COSTA, 2003).

## 2.2 Modelos para dados longitudinais

Em muitas situações prática, o interesse está voltado para modelar variáveis respostas a partir de medidas de uma ou mais populações ao longo de uma dimensão ordenada, por exemplo, o tempo. Essa situação é bastante frequente nos experimentos em que pretende-se estudar o comportamentos de uma variável resposta para diversas populações observadas. Há pelo menos duas situações distintas nas coletas de dados. A mais comum é aquela onde, em cada instante específico, realiza-se apenas uma medida para variável resposta para as diversas subpopulações envolvidas. A segunda estratégia é aquela em que há uma ou mais observações da variável resposta, realizadas em tempos distintos. No primeiro caso diz-se que trata-se de um planejamento transversal enquanto no segundo um planejamento longitudinal (SINGER and ANDRADE, 1986).

O modelo linear tradicional, ou fixo, tem usualmente o estimador de quadrados mínimos como estimador dos parâmetros, que pressupõe independência nos resíduos e homogeneidade nas variâncias (ALCARDE, 2012). Esse fato é um limitador para estudos longitudinais, já que em cada unidade experimental as medidas são tomadas em diferentes ocasiões (tempos) e, possivelmente, em diferentes situações. Tais conjuntos de dados são tratados a partir de modelo multivariado geral com estrutura de covariância irrestrita, que torna o estudo de difícil análise (ALCARDE, 2012). A maior parte da literatura trata esses casos por meio de uma alternativa intermediária, os chamados modelos mistos, os quais contém efeitos fixos, além da média geral, e efeitos aleatórios, além do erro. O termo “efeito fixo” em um modelo estatístico se refere ao comportamento da população por inteiro. Por outro lado, o termo “efeito aleatório” é descrito por uma distribuição de um coeficiente dentro do modelo geral da população (ALCARDE, 2012).

A formulação do modelo de dois estágios pode ser descrita como segue. Considere um modelo de regressão para cada indivíduo  $i$  e a medida no tempo  $t_{ij}$ , com  $i = 1, \dots, N$  e  $j = 1, \dots, n_i$ . Seja também

o vetor  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in})'$  tal que

$$\begin{cases} y_1 = z_1\beta_1 + \epsilon_1 \\ y_2 = z_2\beta_2 + \epsilon_2 \\ \vdots \\ y_N = z_N\beta_N + \epsilon_N \end{cases} \quad (2.1)$$

que matricialmente, o modelo (2.1) tem a forma

$$\mathbf{y}_i = \mathbf{Z}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i \quad (2.2)$$

onde  $\mathbf{Z}_i$  é uma matriz  $n_i \times q$ , chamada matriz de covariáveis,  $\boldsymbol{\beta}_i$  é um vetor de coeficientes de regressão do indivíduo  $i$  e  $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$ , onde  $\boldsymbol{\Sigma}_i$  é a matriz de variância e covariância das observações dentro do indivíduo  $i$ .

O modelo (2.1) representa o primeiro estágio do modelo misto, que trata exatamente de uma regressão clássica para o indivíduo  $i$ . O segundo estágio introduz a variabilidade entre os indivíduos e é descrito por

$$\boldsymbol{\beta}_i = \mathbf{K}_i\boldsymbol{\beta} + \mathbf{b}_i \quad (2.3)$$

em que  $\mathbf{K}_i$  é uma matriz  $q \times p$  de variáveis conhecidas,  $\boldsymbol{\beta}$  é um vetor de ordem  $p$  de parâmetros desconhecido da regressão e  $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$ . De (2.2) e (2.3) vem o modelo geral, que considera todos os indivíduos da população

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \quad (2.4)$$

onde  $\mathbf{X}_i = \mathbf{Z}_i\mathbf{K}_i$  é a matriz,  $n_i \times p$ , dos coeficientes de efeitos fixos,  $\mathbf{Z}_i$  é a matriz dos coeficientes de efeitos aleatórios e  $\mathbf{b}_i$  é o vetor de parâmetros de efeito aleatório.

### 2.2.1 Algumas estruturas de covariâncias

A depender do problema em mãos, muitas são as estruturas de covariância que se pode selecionar. A seguir elencamos algumas delas.

(i) Componentes de variância, caracterizada por variâncias iguais e observações independentes

$$\mathbf{V} = \begin{pmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix}$$

(ii) Simetria composta: se caracterizada por igualdade de variâncias e covariâncias. Apresenta covariâncias constantes entre quaisquer observações de uma mesma unidade devido a independência de erros

$$\mathbf{V} = \begin{pmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 \end{pmatrix}$$

iii) Auto-regressiva de primeira ordem - AR(1): dados de séries temporais igualmente espaçados e correlações diminuindo, ou seja, a covariância entre duas observações decresce à medida que aumenta o

intervalo de tempo entre elas e, denotando-se por  $\rho$  o parâmetro auto-regressivo, onde para um processo estacionário tem-se  $|\rho| < 1$ ,

$$V = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$

iv) Potência - Assim como a anterior, essa matriz também contém dois parâmetros e apresenta e seu uso é aconselhado quando as observações repetidas não são equidistantes.

$$V = \sigma^2 \begin{pmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} & \rho^{d_{14}} \\ \rho^{d_{21}} & 1 & \rho^{d_{23}} & \rho^{d_{24}} \\ \rho^{d_{31}} & \rho^{d_{32}} & 1 & \rho^{d_{34}} \\ \rho^{d_{41}} & \rho^{d_{42}} & \rho^{d_{43}} & 1 \end{pmatrix}$$

v) Componente de variância com heterogeneidade - Essa matriz conserva hipótese de independência nos dados, mas não ocorre homogeneidade de variâncias. O número de parâmetros dessa matriz coincide com sua ordem.

$$V = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{pmatrix}$$

vi) Banda - Essa matriz pressupõe heterogeneidade de variâncias e covariâncias. Nela considera-se uma distância pré-estabelecida em que observações que superam essa distância são consideradas independentes.

$$V = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & 0 & 0 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 & 0 \\ 0 & \rho\sigma_2\sigma_3 & \sigma_3^2 & \rho\sigma_3\sigma_4 \\ 0 & 0 & \rho\sigma_3\sigma_4 & \sigma_4^2 \end{pmatrix}$$

vii) Simétrica composta com heterogeneidade de variâncias - Essa matriz apresenta um número de parâmetros que é uma unidade a mais que sua dimensão. Os dados apresentam variâncias e covariâncias distintas, mas correlações iguais.

$$V = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_3 & \rho\sigma_1\sigma_4 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 & \rho\sigma_2\sigma_4 \\ \rho\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 & \rho\sigma_3\sigma_4 \\ \rho\sigma_1\sigma_4 & \rho\sigma_2\sigma_4 & \rho\sigma_3\sigma_4 & \sigma_4^2 \end{pmatrix}$$

### 3 MATERIAL

A seção seguinte se refere aos experimentos baseados em (RAKOCEVIC *et al.*, 2013), (MATSUNAGA *et al.*, 2016) e (MATSUNAGA *et al.*, 2015), tanto do ponto de vista dos conceitos como dos dados produzidos para nosso estudo.

#### 3.1 Arquitetura de planta e Metâmero

Arquitetura de planta é um termo aplicado para definir a organização dos componentes de uma planta dentro de um espaço e tempo. Essa área científica quantifica a morfologia no tempo e espaço e por isso tende a definir a organização dos componentes vegetais incluindo suas características geométricas (forma, tamanho e orientação dos componentes) e topológicas (conexões físicas entre os componentes). Através da modelagem estrutural de plantas consegue-se fazer uma ligação entre o mundo real e virtual, com a aplicação de conhecimentos biológicos, matemáticos e computacionais (RAKOCEVIC *et al.*, 2013).

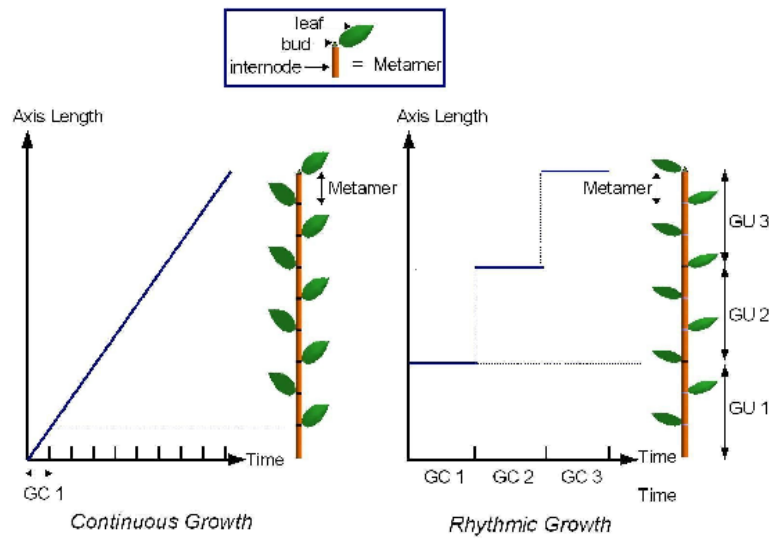
O componente básico da arquitetura vegetal chama-se metâmero. Ele é apresentado por internódio com a(s) sua(s) folha(s) e gema(s) (Figura 2).

As medições arquiteturais usadas nesse texto foram conduzidas em plantas adultas de cafeeiros arábica no Instituto Agrônomo do, Paraná, em dois experimentos. Um comparando impactos de densidades e arranjos na arquitetura e produtividade de cafeeiros e outro comparando as respostas de quatro genótipos na irrigação. Quatro tratamentos foram realizados no primeiro experimento, considerando duas densidades (6.000 e 10.000 plantas ha<sup>-1</sup>) combinadas com dois arranjos de plantios (quadrado - Q e retangular - R). As linhas tinham orientação Leste-Oeste, com a distância entre linhas em plantio retangular de 3 m. O segundo experimento seguiu dois cultivares (IAPAR 59 e Catuaí IAC 99) e duas descendências dos cafeeiros da coleção trazida de Etiópia (a coleção renovada, originada de plantas matrizes obtidas de expedição campo em 2010, em arranjo de 2,5 m na distância entre as linhas e 0,5 m em linha, orientação de linhas foi Leste-Oeste. Os quatro genótipos foram observados desde 2011 até 2013. As plantas dos cafeeiros foram codificadas a partir de grafo em “árvore multi-escalonado- MTGs (multi-scale-tree-graphs), cuja topologia foi decomposta em três escalas- (P), galhos (G) e metâmeros Na escala de metâmeros diferenciaram-se duas classes - ortotrópicos (O) e plagiotrópicos (E), permitindo a definição do dimorfismo de ramos. Para as plantas de 2010 as medições foram realizadas em alto nível de detalhamento na escala de metâmeros, enquanto para 2011-2013, de dois experimentos, as medições realizadas consideraram o detalhamento misto pela amostragem. Em amostragem, foram selecionados quatro ramos plagiotrópicos por cada camada de 40 cm sobre o sentido vertical para serem descritos completamente na escala de metâmeros. Os demais ramos foram descritos apenas pelo comprimento total, orientação cardinal e número total de frutos verdes e maduros. Neste caso, estas plantas necessitaram ser processadas pelo software AmostraCafe3D em (RAKOCEVIC *et al.*, 2013), para incluir as informações na escala de metâmeros nos ramos plagiotrópicos parcialmente codificados, obtendo o MTG da planta inteira como resultado. Para as medições de 2007 e 2008, a descrição arquitetural foi realizada na escala de metâmeros, em que o comprimento de cada entrenó, comprimento/largura/ângulo de elevação/orientação cardinal de folhas, número de frutos verdes e maduros e a posição/orientação de cada ramo foram descritos detalhadamente nos MTGs. Todos os MTGs codificados foram armazenados em um banco de dados.

Como dito antes, as plantas são codificadas em MTG. Esse processo consiste em realizar as medições de uma planta, levando em consideração as características de cada componente.

A arquitetura de plantas do gênero *Coffea* (cafeeiros) é definida pelo dimorfismo de ramos -presença de tronco principal ortotrópico e ramos laterais plagiotrópicos, que crescem quase horizontalmente. As espécies deste gênero têm o crescimento contínuo. Na espécie café arábica o tronco ortotrópico





**Figura 3.1.** Esquema que reúne metâmeros em uma árvore (TEAM DIGIPLANTE, 2016)

é composto por folhas opostas e entrenós de comprimento regular. As folhas do ortotrópico são opostas uma às outras. Os ramos laterais, de um determinado entrenó do ortotrópico estão posicionados em sentidos opostos uns em relação aos outros, contudo possuem a mesma direção, enquanto que no próximo entrenó do ortotrópico os ramos possuem direção oposta (em ângulo de  $90^\circ$ ) em relação ao galho do entrenó anterior em (MATSUNAGA *et al.*, 2015). Atualmente, os cafeeiros dos cultivares teste (IAPAR 59 e Catuaí), localizados no Instituto Agrônomo do Paraná, vêm sendo codificados com base em medições parciais. Essa estratégia consiste em completa descrição dos ortotrópicos (de todos os componentes de cada metâmero), enquanto que os ramos plagiotrópicos são descritos pela amostragem de estratos definidos por altura de copa (a cada 40 cm no eixo z). Em cada estrato escolheram-se quatro ramos orientados para quatro pontos cardeais (Norte (N), Oeste (O), Sul (S), Leste (L)) e nestes são definidos todos os metâmeros e todas as ramificações de segunda e terceira ordem. Para os demais plagiotrópicos de primeira ordem foram definidos somente a posição da sua inserção no ortotrópico, o seu comprimento total, a sua orientação e o número de frutos. A Figura 2 representa um esquema que resume metâmeros em uma árvore.

### 3.2 Bancos de dados

Os bancos de dados são formados por experimentos para investigar os vários tipos de crescimento do café do tipo arábica. Os detalhes dos experimentos o leitor pode encontrar em (RAKOCEVIC *et al.*, 2013) e MATSUNAGA *et al.* (2015). Abaixo encontra-se apenas um resumo dos experimentos.

Como já comentado antes, os experimentos foram conduzidos no Instituto Agrônomo do Paraná, Londrina ( $23^\circ 18' S$  e  $51^\circ 17' W$ , 563 m acima do nível do mar), Brasil, com árvores adultas de *Coffea arábica*, da cultivar mais popular do Paraná - IAPAR 59. As mudas foram plantadas em 1995 e podadas perto do chão em 2008. Dois períodos sucessivos de carregamento de grãos por metâmero foram observados: 1) no primeiro ano de produção (Junho de 2010) e 2) formação inicial de grãos em Novembro - Dezembro de 2010. Duas altas densidades de plantas (6.000 e 10.000 plantas por hectare por ano) combinadas com dois padrões de plantio, quadrado (Q) e retangular (R), que definem quatro tratamentos identificados como Q10, R10 (3 m x 0,33 m), Q6 (1,29 m x 1,29 m) e R6 (3 m x 0,55 m).

As plantas foram codificados seguindo a metodologia VPlants, considerando a modificação do tempo de cada metâmero por 20 plantas. A resposta binária incluída a ocorrência (1) ou não ocorrência (0) de sucessiva ocupação grãos no metâmeros que tinha pelo menos um grão em junho de 2010.

Os dados analisados aqui fazem parte de dois bancos de dados coletados em dois momentos distintos. Cada banco de dados contém informações de 20 plantas. Os dados coletados em cada planta foram:

1. Altura do Ortotrópico
2. Comprimento da folha
3. Largura da folha
4. Surgimento de chumbinho
5. Surgimento de metâmeros
6. Comprimento de entre-nó
7. Orientação do galho
8. Numero de chumbinhos.

### 3.3 Análise exploratória dos dados

conjunto de dados permitiu diferenciar, o filocrono (emissão de pares de folhas em graus dias) entre ordens de ramificação.

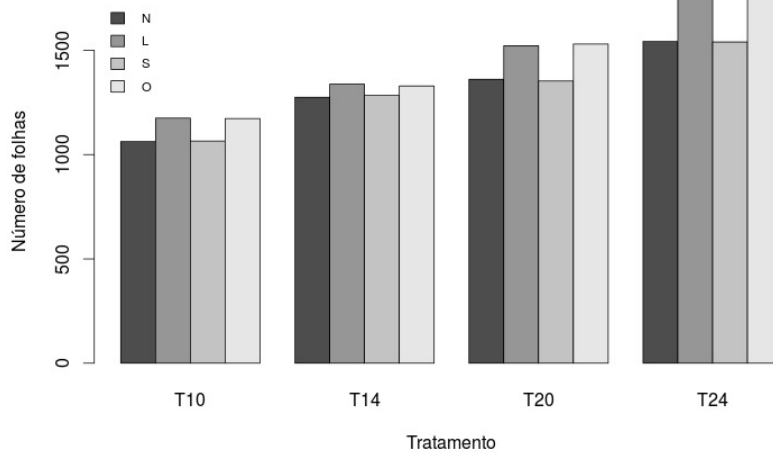
O interesse nessa seção é realizar um estudo exploratório do conjunto de dados para auxiliar no desenvolvimento de equações que envolvem alometria e assimetrias. Tal estudo pode auxiliar na questão central desse trabalho: modelos mistos para prever probabilidade no ressurgimento de metâmeros. Assim, os objetivos nesse seção os seguintes são:

- Testar assimetria de plantas, considerando pontos cardiais: norte, sul, leste e oeste;
- Testar assimetria devido a posicionamento das plantas em linhas de cultivo;
- Estimar alometria foliar.

Como dito acima, são considerados as informações coletadas e organizadas em MTGs de plantas detalhadamente codificadas (na escala de metâmeros) em Novembro e Dezembro de 2010. Os cafeeiros foram cultivados em quatro desenhos diferentes: duas altas densidades - 10.000 plantas  $h.ano^{-1}$  (T10 e T14) e 6.000 plantas  $h.ano^{-1}$  (T20 e T24) em dois arranjos - quadrado (0) e retangular (4).

As plantas geralmente têm a capacidade de crescer indefinidamente e substituir as estruturas vegetativas danificadas por novas estruturas. Alterações na forma, tamanho e estrutura são comuns durante o desenvolvimento. Esses organismos dependem fortemente de fatores extrínsecos, ou seja, fatores do ambiente em que vivem. Assim, a capacidade de modificar o crescimento de acordo com as presentes condições, pode alterar a função de um módulo ou conjunto de órgãos, e se adaptar a novas condições ambientais. Tecnicamente, a esse estudo denomina-se **Alometria** (RAKOCEVIC *et al.*, 2013).

Alometria refere-se às alterações no tamanho relativo de partes do corpo correlacionadas com as mudanças no tamanho total. Mais especificamente, durante o desenvolvimento de um organismo, o crescimento alométrico refere-se ao diferencial de crescimento em diferentes partes do corpo. As plantas geralmente têm a capacidade de crescer indefinidamente e substituir as estruturas vegetativas danificadas por novas estruturas. Alterações na forma, tamanho e estrutura são comuns durante o desenvolvimento. Essas reorganizações dependem fortemente de fatores extrínsecos, ou seja, fatores do ambiente em que vivem.



**Figura 3.2.** Número de folhas relativo à sua orientação no espaço. Norte(N), Leste(L), Sul(S), Oeste(O).

O conceito de filocrono, ou seja, o intervalo de tempo necessário para a emissão de duas folhas consecutivas (em cafeeiros dois pares de folhas consecutivas) pode ser usado para simular o aparecimento de folhas em plantas. O número de folhas, além de afetar o índice de área foliar da planta, é considerado uma excelente medida do seu tempo fisiológico. Em algumas espécies, como na batata e em gramíneas, este parâmetro fisiológico aparentemente pode mudar com as condições de cultivo, por exemplo, com a densidade de plantio. O cálculo do filocrono é uma expressão temporal-espacial e por isso pode ser calculado considerando diversas escalas. Nos cafeeiros, a ideia é considerar um conjunto de folhas de planta, de um ramo com certa ordem, ou de ramos de diferentes qualidades (ortotrópicos e plagiotrópicos). A regressão linear pode ser aplicada como um dos métodos para o cálculo do filocrono em espécies anuais de crescimento rápido. No experimento tratado aqui foram realizadas medições a cada duas semanas nos cafeeiros para a realização de cálculos utilizando a teoria de renovação. Assim, a partir de outubro de 2013 foram efetuadas 26 medições, somente no E083 (único sobrevivente na geada de 2013), com intervalo de duas a três semanas. Foram realizadas extrações destes arquivos e, pela ordem de ramificação, foi avaliado o crescimento em graus dias, com a hipótese que existem diferenças no filocrono entre as ordens de ramificação 1<sup>a</sup> versus 2<sup>a</sup>, 2<sup>a</sup> versus 3<sup>a</sup>, 3<sup>a</sup> versus 4<sup>a</sup> e 4<sup>a</sup> versus 5<sup>a</sup>.

### 3.3.1 Auxílio no desenvolvimento de equações para alometria e assimetria

Todo estudo seguir se refere apenas a um dos bancos de dados utilizados nesse texto. Resultados semelhantes foram obtidos para o outro banco. Por isso, achamos desnecessária a apresentação de seus resultados.

Nota-se que, nos quatro tratamentos, o número de folhas no eixo leste-oeste foi maior que no eixo norte-sul (Figura 3.2). A plantação de linhas encontra-se no eixo leste-oeste, indicando maior impacto negativo de passagens mecanizadas do que a competição entre folhas dentro de linhas.

A terceira ordem da ramificação é onde se concentra a maior quantidade de folhas, a segunda ordem aparece com menos quantidades de folhas, chegando a ser metade da terceira. Mas mesmo assim, com muito mais folhas que as demais ordens, independente do tratamento (Figura 3.3). Além disso, essa diferença torna-se ainda maior no tratamento 24. A Figura 3.4 mostra que, como um reflexo da metade de um cafeeiro, o número maior de folhas concentra-se na base, diminuindo conforme caminha-se em direção a ápice principal. Ao analisar a Figura 3.5, de dispersão das medidas de comprimento versus

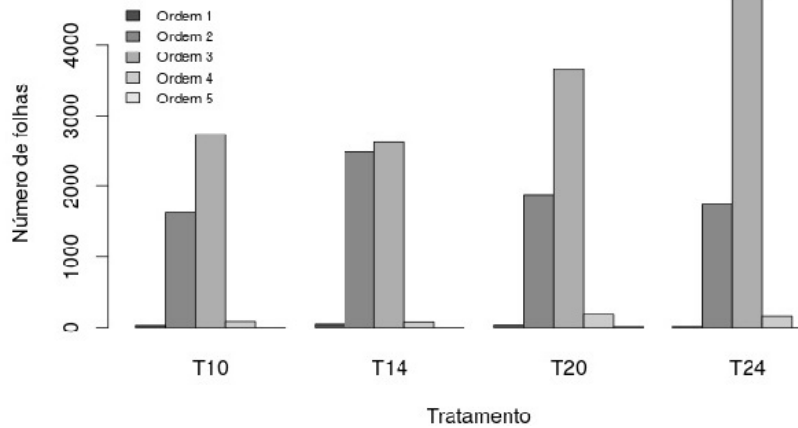


Figura 3.3. Número de folhas relativo a ordem de ramificação

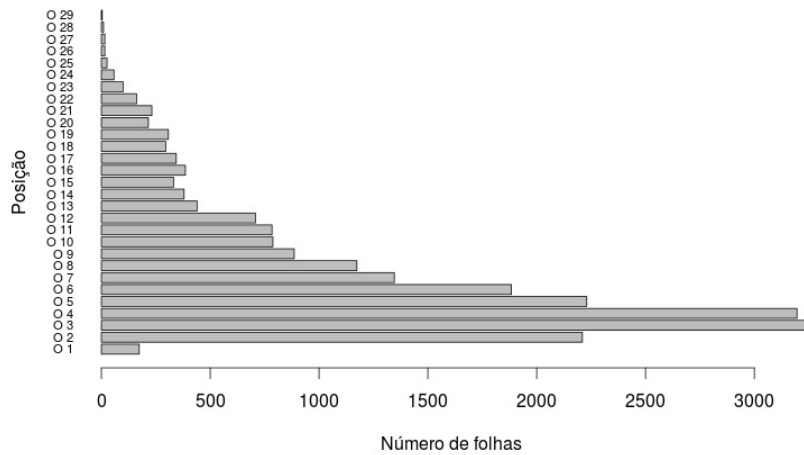


Figura 3.4. Número de folhas pela inserção no ortotrópico (considerando todo o conjunto de plantas)

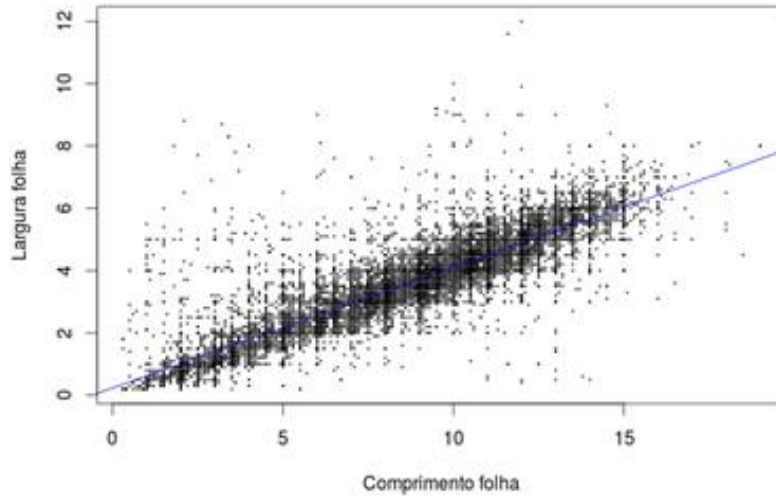
largura de uma mesma folha, nota-se que existe uma forte correlação entre estes valores. Para verificar matematicamente essa relação alométrica foi traçada uma reta de regressão, obtendo a seguinte equação:

$$largura = 0,23 + 0,39 * comprimento$$

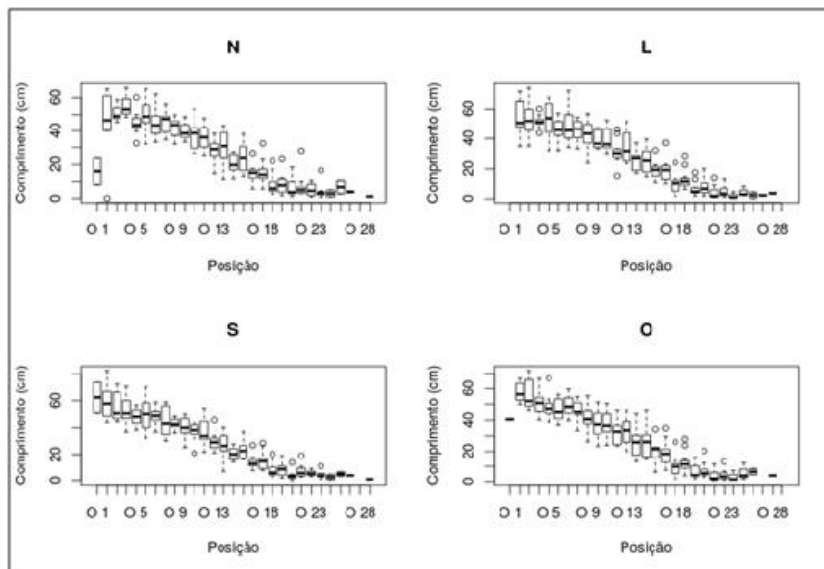
O problema é que não há indícios de homocedasticidade neste modelo, precisando ser necessário ajuste por algum outro tipo de modelo.

Os gráfico não sugerem influencia significativa da orientação cardinal. Nota-se que os ramos inferiores apresentam uma mediana baixa, muito provavelmente devido à morte (ou quebra), uma vez que eles são mais velhos (Figura 3.6). Conforme esperado, o tamanho dos ramos vão diminuindo em direção a ápice do cafeeiro, independentemente da orientação, proporcionando o formato de cone de cafeeiros

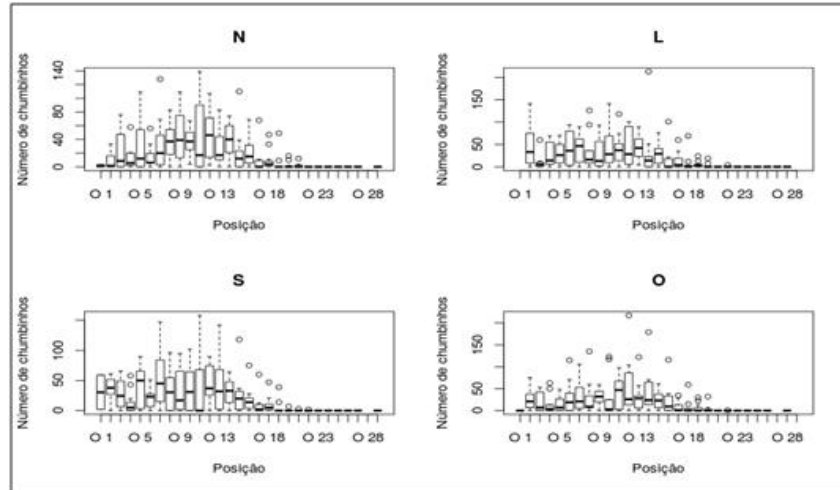
A maior concentração de chumbinhos encontra-se da meia-altura para baixo de cada árvore do cafeeiro (Figura 3.7), uma vez que em direção ao ápice este número reduz drasticamente. Novamente a orientação cardinal não parece interferir muito neste fenômeno. Outra medida de interesse é o número de ramificações. Segundo especialistas (RAKOCEVIC *et al.*, 2013), essa informação tem grande estreitamento



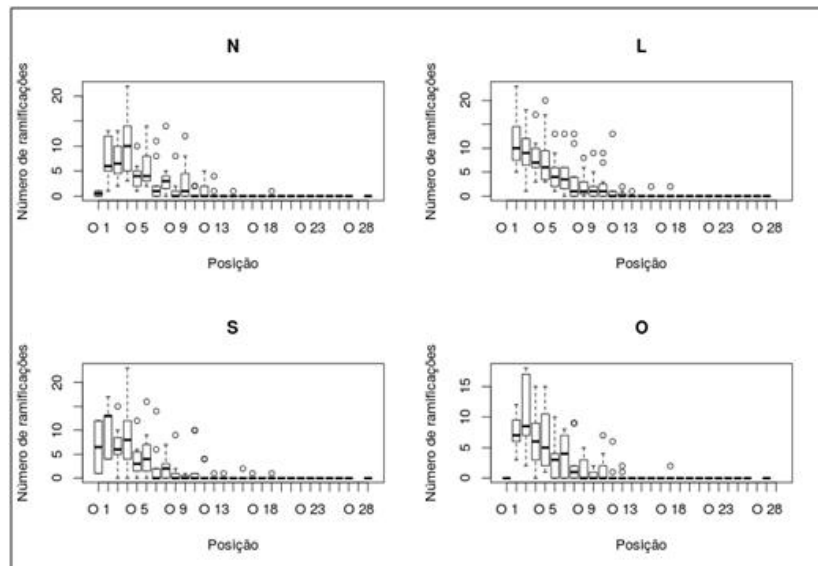
**Figura 3.5.** Alometria de comprimento e largura de folhas (em cms)



**Figura 3.6.** Comprimento dos eixos da 2ª ordem dependendo da sua inserção no eixo ortotrópico (1ª ordem) e orientação cardinal no espaço.



**Figura 3.7.** Número de frutos formados nos eixos da 2ª ordem dependendo da sua inserção no eixo ortotrópico (1ª ordem) e orientação cardinal no espaço.



**Figura 3.8.** Número de ramificações formados nos eixos da 2ª ordem dependendo da sua inserção no eixo ortotrópico (1ª ordem) e orientação cardinal no espaço.

com o surgimento de metâmeros. Como esperado, o número de ramificações é muito maior na base dos cafeeiros, não existem ramificações na metade superior (Figura 3.8).

Uma vez que as folhas se dão aos pares, foi feito um estudo alométrico entre as principais dimensões foliar. Inicialmente para comprimentos e em seguida, para as larguras de cada par de folhas. Os gráficos de dispersão estão ilustrados abaixo. Conforme pode-se ver no gráfico (Figura 3.9), existe uma forte correlação entre estas medidas. Foi traçada uma reta de regressão, obtendo a seguinte expressão:

$$\text{comprimento folha2} = 0,78 + 0,90 * \text{comprimento folha1}$$

Mais uma vez depara-se com o problema de homocedasticidade neste modelo, sendo necessário ajuste por algum outro tipo de modelo. Assim como no caso anterior, foi feito um gráfico de dispersão para analisar a relação alométrica entre a largura desses pares de folhas. Nota-se uma forte correlação entre

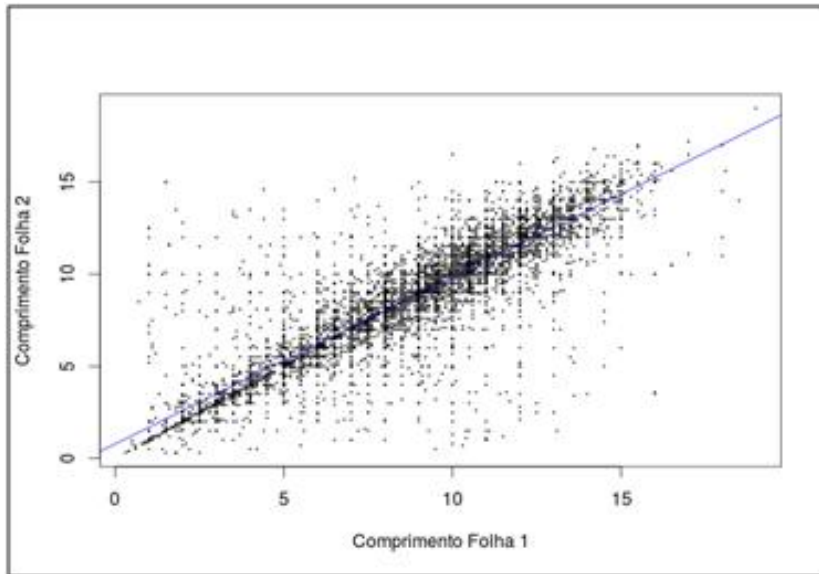


Figura 3.9. Alometria do comprimento (em cms) entre as duas folhas do par.

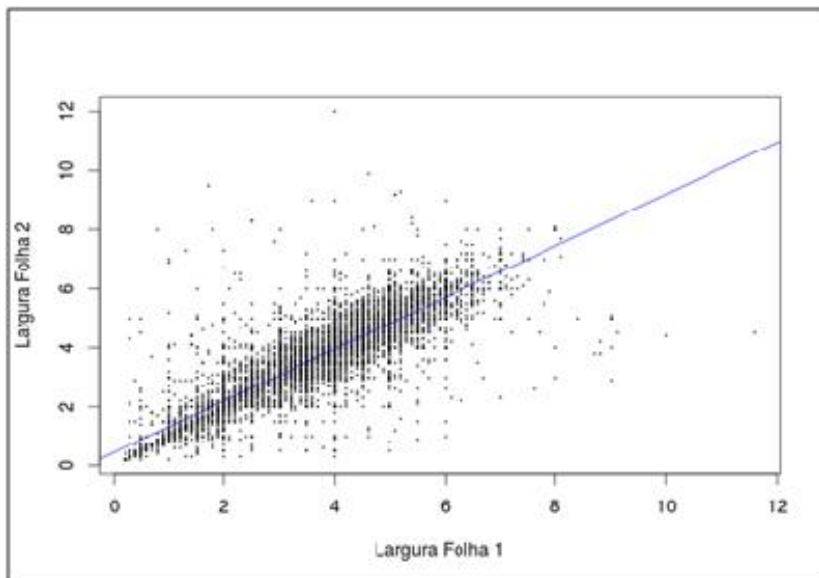
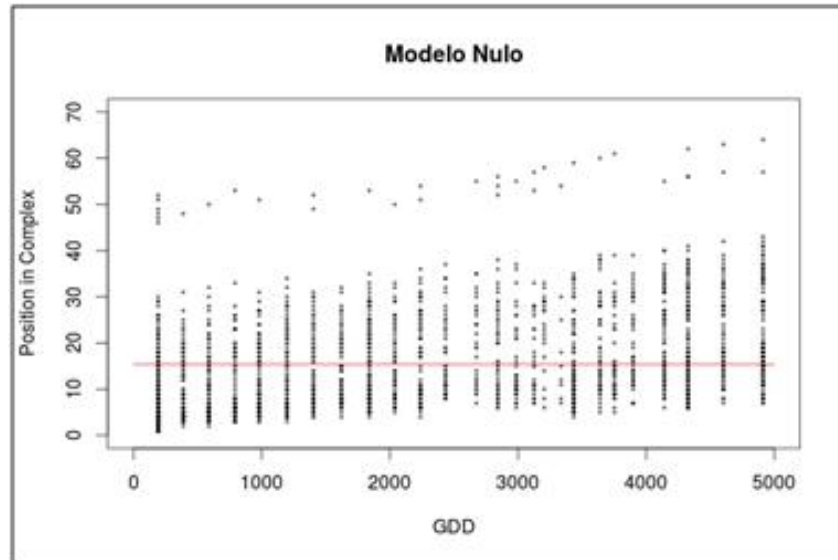


Figura 3.10. Alometria do largura (em cms) entre as duas folhas do par.



**Figura 3.11.** Modelo de regressão nulo para o cálculo de filocrono

estas medidas (Figura 3.10). Novamente foi traçada uma reta de regressão, obtendo a seguinte equação:

$$largura_{folha2} = 0,45 + 0,88 * largura_{folha1}$$

Novamente parece existir um problema de falta de homocedasticidade dos resíduos neste modelo, sendo necessário o ajuste por algum outro tipo de análise.

A seguir será apresentado estudos dos dados por meio de retas concorrentes e paralelas.

### 3.4 Estudo das ordens de ramificação utilizando retas de regressão

Na busca de melhor modelo de cálculo de filocrono pela ordem de ramificação foram testados alguns modelos, tais como modelo de regressão nulo, de retas coincidentes, de intercepto comum, de retas paralelas e de retas concorrentes. O modelo de regressão nulo (Figura 3.11), traça uma reta horizontal na média dos valores e, neste caso, é dado pela equação:  $Y = 15,37$ .

O modelo de retas coincidentes mostra que, no geral, existe uma leve inclinação na reta de regressão, indicando um crescimento baixo (Figura 3.12). A equação da reta ajustada é dada por

$$Y = 9,267 + 0,003 * X.$$

O modelo de intercepto comum mostra uma reta de regressão para cada ordem de ramificação, mas mantendo a mesma estimativa do intercepto (Figura 3.13). Pode-se verificar que aparentemente as retas são diferentes para cada uma das ordens da ramificação. As equações de cada reta são dadas por:

$$Ordem1 : Y = 9,0790 + 0,0137 * X.$$

$$Ordem2 : Y = 9,0790 + 0,0053 * X.$$

$$Ordem3 : Y = 9,0790 + 0,0016 * X.$$

$$Ordem4 : Y = 9,0790 + 0,0008 * X.$$

$$Ordem5 : Y = 9,0790 + 0,0001 * X.$$

O modelo de retas paralelas, mostra uma reta de regressão para cada ordem de ramificação, mas mantendo



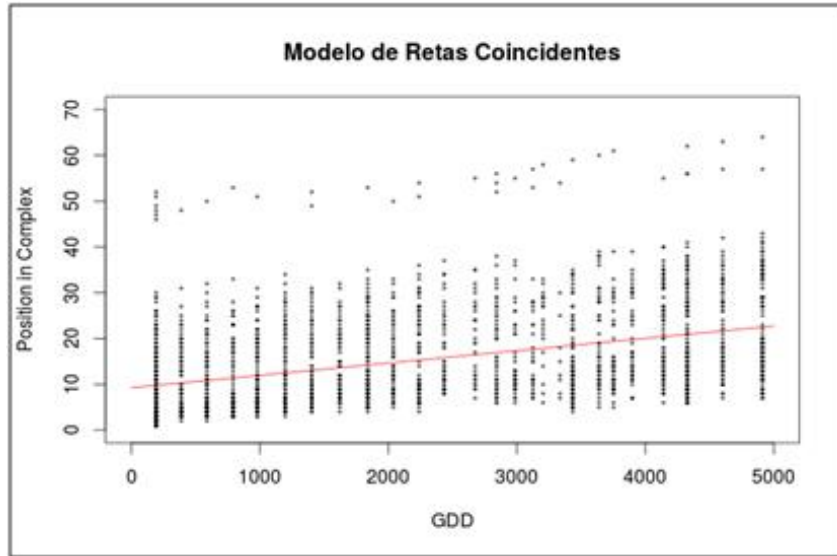


Figura 3.12. Modelo de retas coincidentes para o cálculo de filocrono

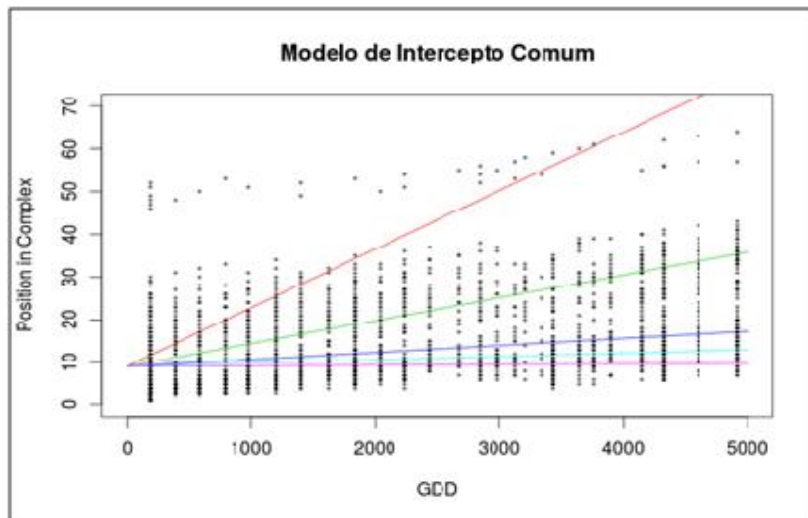


Figura 3.13. Modelo de intercepto comum para o cálculo de filocrono

a mesma inclinação (Figura 3.14). Conforme já era suspeitado, as retas são diferentes para cada uma das ordens da ramificação. As equações de cada reta são dadas por:

$$\text{Ordem1} : Y = 47,60 + 0,0026 * X.$$

$$\text{Ordem2} : Y = 17,41 + 0,0026 * X.$$

$$\text{Ordem3} : Y = 5,827 + 0,0026 * X.$$

$$\text{Ordem4} : Y = 3,136 + 0,0026 * X.$$

$$\text{Ordem5} : Y = 1,115 + 0,0026 * X.$$

Por fim, o modelo de retas concorrentes além de calcular uma reta de regressão para cada ordem de ramificação, permite variar o valor do intercepto e da inclinação (Figura 3.15). Observa-se que, aparentemente, não houve muito ganho em relação ao modelo anterior. As equações de cada reta são dadas por:

$$\text{Ordem1} : Y = 48,08 + 0,0024 * X.$$

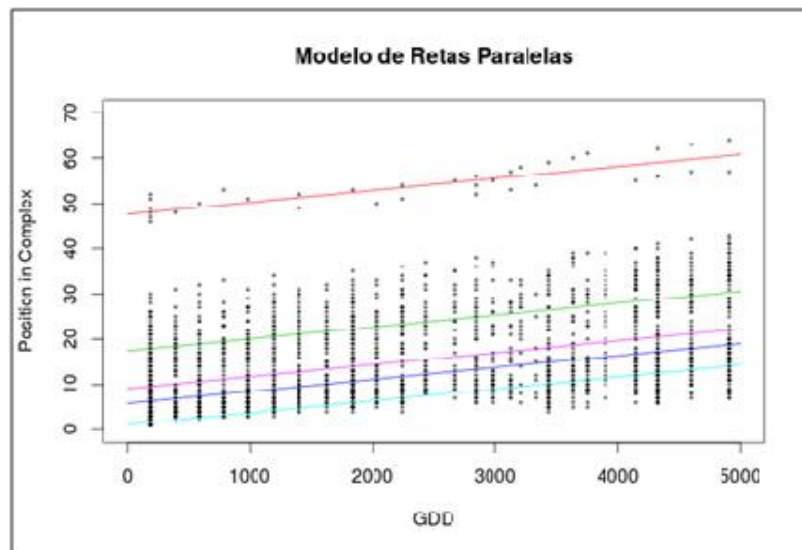


Figura 3.14. Modelo de retas paralelas para o cálculo de filocrono

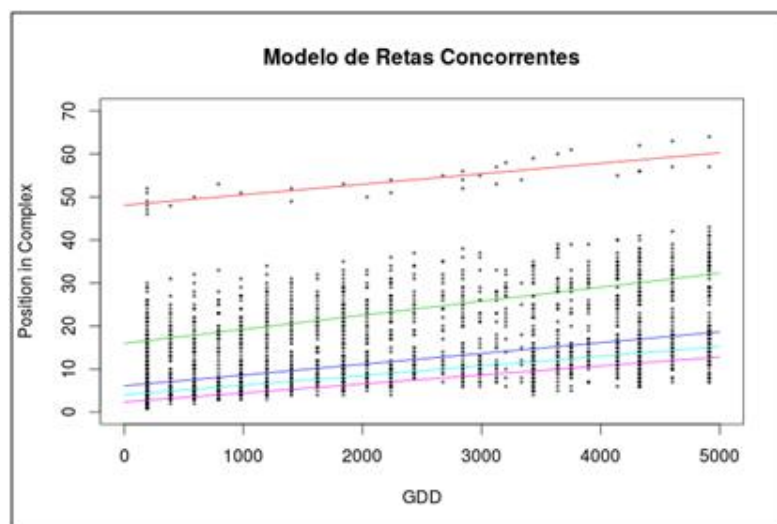


Figura 3.15. Modelo de retas concorrentes para o cálculo de filocrono

$$\text{Ordem2} : Y = 15,99 + 0,0033 * X.$$

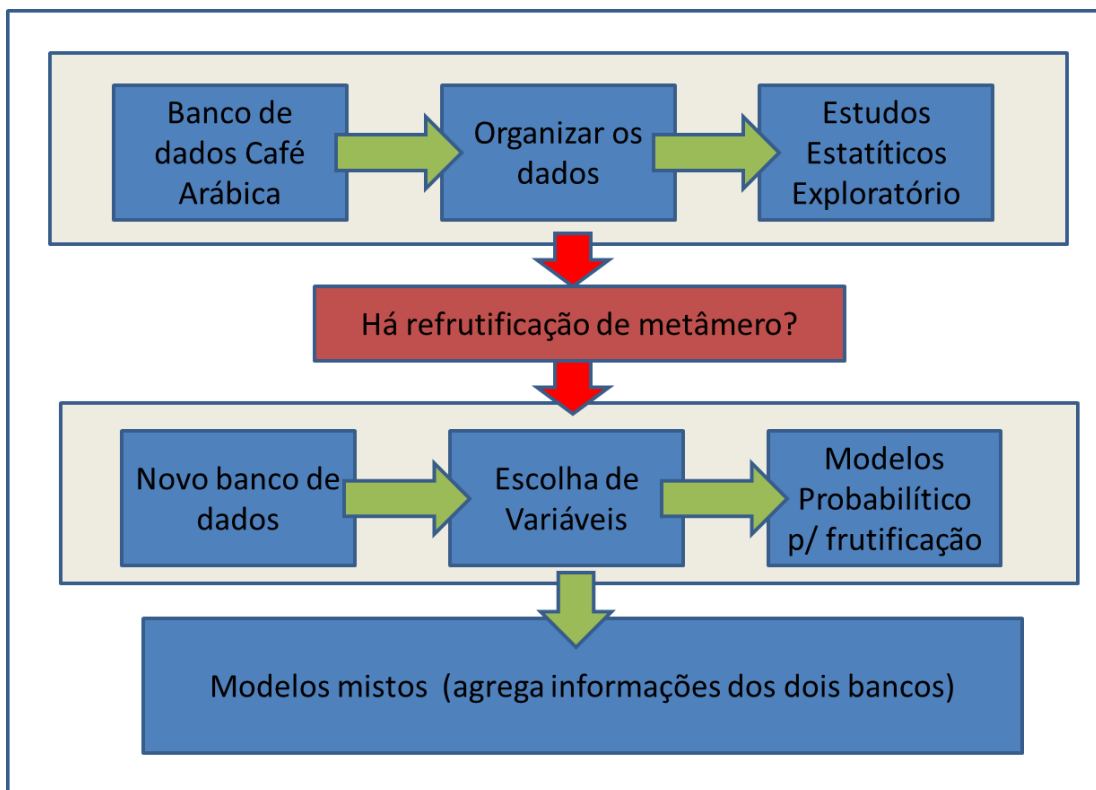
$$\text{Ordem3} : Y = 6,115 + 0,0025 * X.$$

$$\text{Ordem4} : Y = 4,025 + 0,0022 * X.$$

$$\text{Ordem5} : Y = 2,361 + 0,0021 * X.$$

O modelo de retas concorrentes aparentemente não acrescenta nenhuma informação e pelo critério da parcimônia ele seria o pior modelo por ter muitos parâmetros. O modelo de retas paralelas parece ser o mais adequado pois além de ser significativo, parece graficamente se ajustar bem aos pontos, levando em consideração as diferenças entre as ordens de ramificação, mas mantendo a inclinação fixa, indicando que o a taxa crescimento é igual em todas as ordens.

O fluxograma da Figura 3.16 ilustra, pedagogicamente e cronologicamente, cada passo seguido na realização deste estudo para um melhor entendimento das etapas e questionamentos que surgiram no decorrer da pesquisa.



**Figura 3.16.** Fluxograma cronológico de cada etapa realizada neste trabalho.

## 4 MÉTODOS

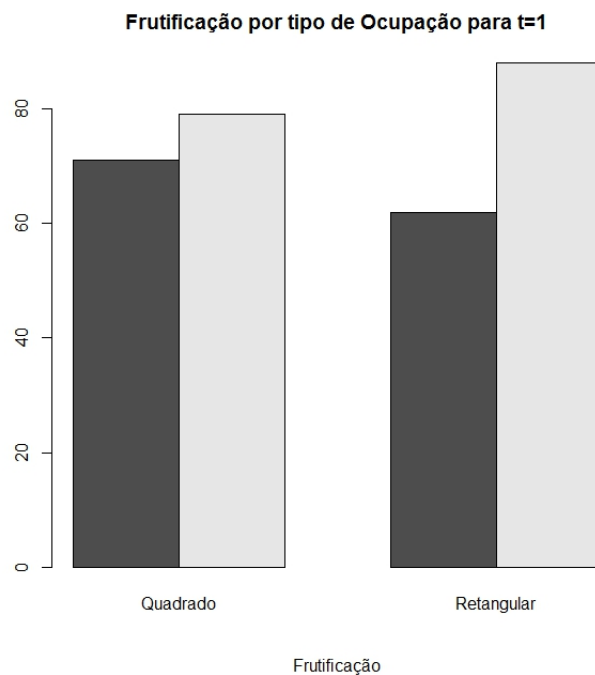
Esse capítulo tem como principal interesse a investigação da probabilidade de frutificação na árvore de café. Para isto, inicialmente serão feitos estudos da influencia de cada tratamento adotado nas plantas. A técnica utilizada para isso é através de regressão logística.

Para o presente estudo foram escolhidos aleatoriamente 15 metâmeros de cada planta em dois tempos diferentes de observação (Junho de 2010 e Novembro – Dezembro de 2010), totalizando assim em 600 unidades amostrais.

### 4.1 Análise exploratória dos dados para os tempos 1 e 2

Afim de atender alguns questionamentos botânicos específicos ligados à frutificação de plantas de café em tempos sucessivo, foram selecionas para o estudo variáveis ligadas a “OCUPAÇÃO” (quadrática ou retangular já estabelecida no plantio), “ESPAÇAMENTO” (6.000 plantas/ha ou 10.000 plantas/ha) e “TAMANHO DO ENTRENÓ” (comprimento do metâmero). Este estudo foi realizado separadamente para os dois tempos de observação.

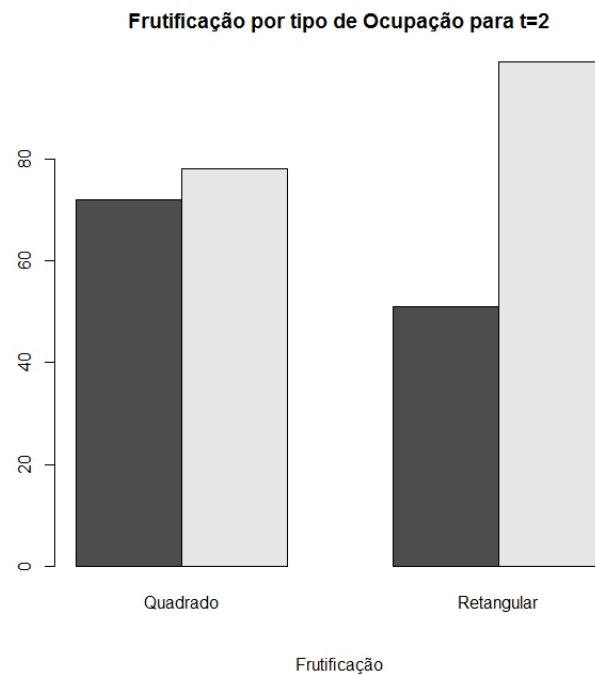
As Figura 4.1 e Figura 4.2 mostram que há maior proporção de metâmeros que frutificaram na ocupação retangular quando comparada à quadrática. Essa discrepância fica mais evidente no tempo 2, como pode ser observado na Figura 4.2.



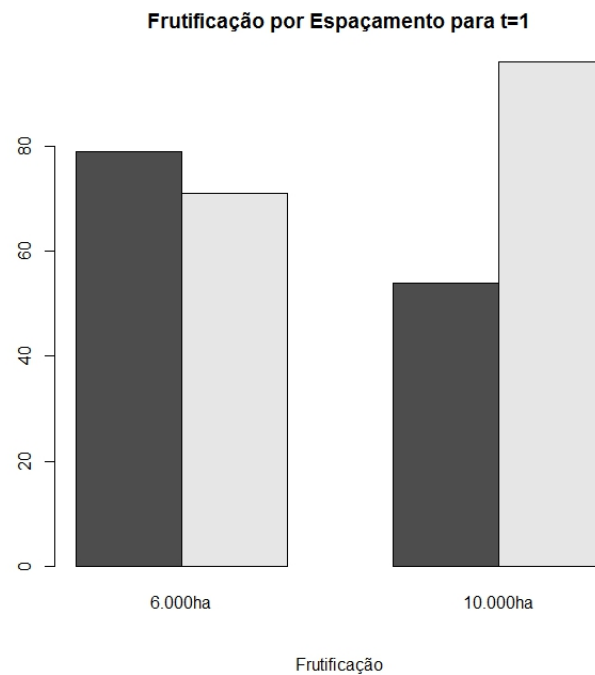
**Figura 4.1.** Frutificação por tipo de Ocupação para  $t = 1$ .

Para o tempo 1, a Figura 4.3 revela que a proporção de metâmeros que frutificaram no espaçamento de 6.000 plantas/ha é menor se comparada aos metâmeros que não frutificaram. Para o espaçamento de 10.000 plantas/ha, há uma proporção de metâmeros que frutificaram, revelando assim que há uma possibilidade do espaçamento interferir na quantidade de metâmeros frutificados.

Para o tempo 2 a Figura 4.4, revela que a proporção de metâmeros que frutificaram no espaçamento de 6.000 plantas/ha foi um pouco maior do que os que não frutificaram. Porém, para o espaçamento de 10.000 plantas/ha, a discrepância entre a quantidade de metâmeros que frutificaram em relação aos que não frutificaram manteve-se grande.



**Figura 4.2.** Frutificação por tipo de Ocupação para  $t = 2$ .

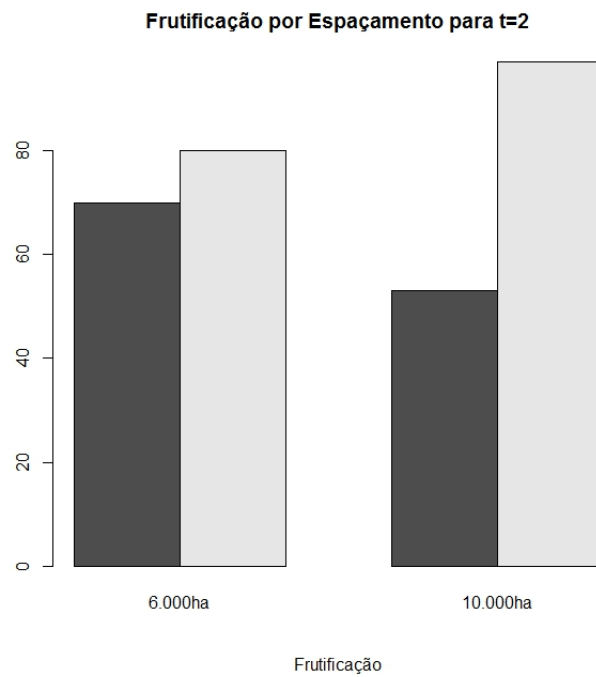


**Figura 4.3.** Frutificação por tipo de Espaçamento para  $t = 1$ .

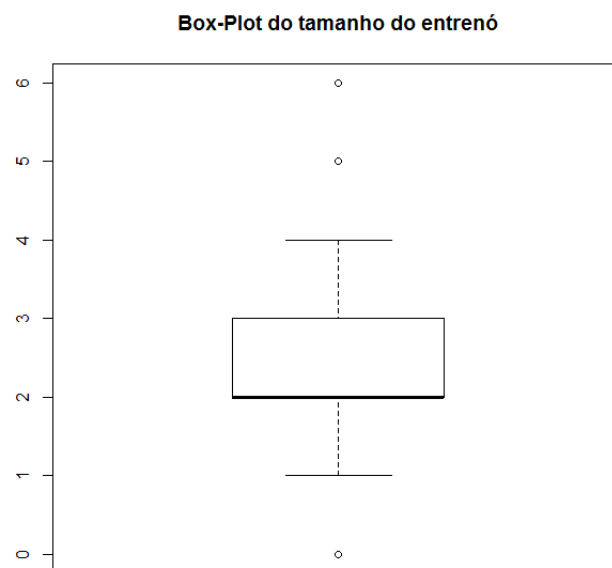
A Figura 4.5 nos traz informações sobre os tamanhos dos entrenós estudados. Esta variável não muda de acordo com o tempo estudado, uma vez que o tamanho do entrenó não se altera com o tempo. A média relacionada ao tamanho do entrenó é de 2,8cm.

As Tabelas 4.1 e 4.2 exibem a proporção de metâmeros frutificados por cada planta nos tempos  $t=1$  e  $t=2$ , respectivamente.

A seguir apresentamos a modelagem da probabilidade de frutificação em cada tempo estudado



**Figura 4.4.** Frutificação por tipo de Espaçamento para  $t = 2$ .



**Figura 4.5.** Tamanhos dos entrenós estudados.

## 4.2 Modelagem da probabilidade para cada tempo

A seguir apresentamos estudos para cada banco de dados. Primeiramente para o tempo 1 e posteriormente para o tempo 2.

**Tabela 4.1.** Proporção de metâmeros que frutificaram por plantas para t=1

Planta	Não Frutificou	Frutificou	Frequência	%
1	5	10	0,67	67%
2	7	8	0,53	53%
3	5	10	0,67	67%
4	7	8	0,53	53%
5	9	6	0,40	40%
6	5	10	0,67	67%
7	8	7	0,47	47%
8	7	8	0,53	53%
9	9	6	0,40	40%
10	6	9	0,60	60%
11	8	7	0,47	47%
12	4	11	0,73	73%
13	8	7	0,47	47%
14	7	8	0,53	53%
15	6	9	0,60	60%
16	4	11	0,73	73%
17	6	9	0,60	60%
18	8	7	0,47	47%
19	8	7	0,47	47%
20	6	9	0,60	60%

**Tabela 4.2.** Proporção de metâmeros que frutificaram por plantas para t=2

Planta	Não Frutificou	Frutificou	Frequência	%
1	4	11	0,73	73%
2	5	10	0,67	67%
3	7	8	0,53	53%
4	6	9	0,60	60%
5	7	8	0,53	53%
6	3	12	0,80	80%
7	7	8	0,53	53%
8	7	8	0,53	53%
9	9	6	0,40	40%
10	6	9	0,60	60%
11	6	9	0,60	60%
12	4	11	0,73	73%
13	9	6	0,40	40%
14	8	7	0,47	47%
15	6	9	0,60	60%
16	5	10	0,67	67%
17	4	11	0,73	73%
18	6	9	0,60	60%
19	6	9	0,60	60%
20	8	7	0,47	47%

#### 4.2.1 Modelagem da probabilidade para o tempo 1

Utilizando o pacote estatístico R, foram testados primeiramente os modelos incluindo-se, uma a uma, todas as covariáveis independentes (ocupação, espaçamento, tamanho do entrenó) com a variável resposta (frutificação). Ao nível de significância de 0,05 foram descartadas apenas as variáveis que têm pouca relação com a variável resposta. Obteve-se assim diferentes modelos para cada planta. A segunda etapa foi filtrar os modelos obtidos na primeira, considerando agora as interações entre as covariáveis por meio do backward, como método de seleção de variáveis regressoras ao nível de significância de 0,05.

Porém, nenhuma interação foi considerada significativa para nenhum dos modelos obtidos.

Para as unidades amostrais colhidas no tempo  $t=1$  obtivemos um modelo cuja única variável significativa foi a relacionada ao espaçamento do plantio. A Tabela 4.3 revela os valores das estimativas dos parâmetros e seus respectivos p-valores.

**Tabela 4.3.** Estimativas dos Coeficientes e seus respectivos p-valores do modelo de regressão logístico apenas com a variável “ESPAÇAMENTO” relativo ao tempo  $t=1$

Covariável	Categoria	$\hat{\beta}$	p-valor
(Intercept)		-0,107	0,514
ESPAÇAMENTO	6.000ha	-	-
	10.000ha	0,682	0,004

Os valores da Tabela 4.3 são relativos ao seguinte modelo de regressão:

$$\text{Logit}(\text{Pr}(\text{frutificar})) = \ln\left(\frac{\text{Pr}(\text{frutificar})}{\text{Pr}(\text{não frutificar})}\right) = -0,107 + 0,628\text{ESPAÇAMENTO}. \quad (4.1)$$

Através dessa estimativa, obteve-se a Razão de Chances (RC) relacionada à variável “ESPAÇAMENTO”. Essa estimativa, relacionada ao modelo obtido, nos revela que uma planta que está em um espaçamento de 10.000 plantas/ha tem chance de frutificar quase o dobro (1,98) do que uma planta que está em um espaçamento de 6.000 plantas/ha.

Para se avaliar a qualidade do ajuste para o modelo, realizou-se o teste de Hosmer-Lemeshow, cujo p-valor foi igual a 0,36. Este resultado levou a não rejeição da hipótese que o modelo está bem ajustado.

Os modelos logísticos envolvendo as variáveis “OCUPAÇÃO” e “TAMANHO DO ENTRENÓ” para o tempo  $t = 1$  foram rejeitados baseado no p-valor obtido tal como pode-se observar nas Tabelas 4.4 e 4.5.

**Tabela 4.4.** Estimativas dos Coeficientes e seus respectivos p-valores do modelo de regressão logístico apenas com a variável “OCUPAÇÃO” relativo ao tempo  $t=1$

	Categoria	$\hat{\beta}$	p-valor
(Intercept)		0,107	0,514
OCUPAÇÃO	Quadrática	-	-
	Retangular	0,243	0,296

**Tabela 4.5.** Estimativas dos Coeficientes e seus respectivos p-valores do modelo de regressão logístico apenas com a variável “TAMANHO DO ENTRENÓ” relativo ao tempo  $t=1$

	$\hat{\beta}$	p-valor
(Intercept)	0,166	0,561
TAMANHO DO ENTRENÓ	0,023	0,813

#### 4.2.2 Modelagem da probabilidade para o tempo 2

Para as unidades amostrais colhidas no tempo  $t=2$  obteve-se um modelo cujas variáveis significativas foram as relacionadas ao espaçamento e ocupação do plantio. A Tabela 4.2.2 apresenta os valores das estimativas dos parâmetros e seus respectivos p-valores.

Os valores da Tabela 4.2.2 são relativos ao seguinte modelo de regressão:

$$\text{Logit}(\text{Pr}(\text{frutificar})) = \ln\left(\frac{\text{Pr}(\text{frutificar})}{\text{Pr}(\text{não frutificar})}\right) = -0,159 + 0,481\text{ESPAÇAMENTO} + 0,591\text{OCUPAÇÃO}. \quad (4.2)$$



**Tabela 4.6.** Estimativas dos Coeficientes e seus respectivos p-valores do modelo de regressão logístico com as variáveis “ESPAÇAMENTO” e “OCUPAÇÃO” relativo ao tempo t=2

Covariável	Categoria	$\hat{\beta}$	p-valor
(Intercept)		-0,159	0,432
ESPAÇAMENTO	6.000ha	-	-
	10.000ha	0,481	0,044
OCUPAÇÃO	Quadrática	-	-
	Retangular	0,591	0,014

Obteve-se então as estimativas das Razões de Chances (RC) relacionadas às variáveis “ESPAÇAMENTO” e “OCUPAÇÃO”. Essas estimativas revelam que uma planta que está em um espaçamento de 10.000ha-1 tem chance 1,62 para frutificação em relação a uma planta que está em um espaçamento de 6.000ha-1. Além disso, há uma chance de 1,81 superior de uma planta que está em uma ocupação retangular frutificar quando comparado com uma que está em uma ocupação quadrática.

Para avaliar a qualidade do ajuste do modelo, realizou-se o teste de Hosmer-Lemeshow, cujo p-valor foi igual a 0,41. Este resultado leva a não rejeição da hipótese que o modelo está bem ajustado.

Como podemos observar nas Tabelas 4.7 e 4.8, as variáveis “ESPAÇAMENTO” e “OCUPAÇÃO” são significativas para explicar a probabilidade de frutificação no tempo  $t = 2$ . Entretanto, a variável “TAMANHO DO ENTRENÓ” não foi significativa como pode ser verificado na Tabela 4.9.

**Tabela 4.7.** Estimativas dos Coeficientes e seus respectivos p-valores do modelo de regressão logístico apenas com a variável “ESPAÇAMENTO” relativo ao tempo t=2

	Categoria	$\hat{\beta}$	p-valor
(Intercept)		0,134	0,415
ESPAÇAMENTO	6.000ha	-	-
	10.000ha	0,471	0,047

**Tabela 4.8.** Estimativas dos Coeficientes e seus respectivos p-valores do modelo de regressão logístico apenas com a variável “OCUPAÇÃO” relativo ao tempo t=2

	Categoria	$\hat{\beta}$	p-valor
(Intercept)		0,080	0,624
OCUPAÇÃO	Quadrática	-	-
	Retangular	0,583	0,014

**Tabela 4.9.** Estimativas dos Coeficientes e seus respectivos p-valores do modelo de regressão logístico apenas com a variável “TAMANHO DO ENTRENÓ” relativo ao tempo t=2

	$\hat{\beta}$	p-valor
(Intercept)	0,192	0,505
TAMANHO DO ENTRENÓ	0,064	0,515

Embora as variáveis “ESPAÇAMENTO” e “OCUPAÇÃO” tenham sido significativas para o modelo, a interação entre as mesmas não foi significativa como pode-se observar na Tabela 4.10.

A seguir, com o propósito de estudar efeitos aleatórios intra e inter-específico apresentados nos dois bancos de dados, foi utilizado a modelagem baseada em técnicas dos modelos mistos.

### 4.3 Modelos mistos para os dois tempos estudados

Com o objetivo de integrar a variável tempo ao estudo e considerando que possivelmente haja uma correlação entre as respostas obtidas em uma mesma unidade (planta) ao longo do tempo, optou-se

**Tabela 4.10.** Estimativas dos Coeficientes e seus respectivos p-valores do modelo de regressão logístico com as variáveis “ESPACAMENTO” e “OCUPAÇÃO” relativo ao tempo t=2

	Categoria	$\hat{\beta}$	p-valor
(Intercept)		-0.08004	0.729
ESPACAMENTO	6.000ha	-	-
	10.000ha	0.32120	0.327
OCUPACAO	Quadrática	-	-
	Retangular	0.43025	0.191
ESPACAMENTO*OCUPACAO	10.000ha*Retangular	0.34019	0.479

pela possível extensão do modelo utilizando a teoria de modelos lineares generalizados mistos. Assim, uma forma de modelar os dados longitudinais presentes no estudo seria incluindo variáveis latentes no preditor afim de incorporar a correlação entre as mesmas plantas.

Foram testados quatro modelos com a inclusão do fator aleatório associado à variável “PLANTAS”. As Tabelas 4.11, 4.12, 4.13 e 4.14 apresentam as características de cada modelo testado.

**Tabela 4.11.** Estimativas dos Coeficientes e seus respectivos p-valores do modelo logístico com retas paralelas e inclusão do efeito aleatório referente às plantas

	Categoria	Estimativa	Erro-padrão	Estatística z	p-valor
(Intercept)		-0,345	0,301	-1,147	0,252
ESPACAMENTO	6.000ha	-	-	-	-
	10.000ha	0,457	0,233	1,962	0,050
OCUPACAO	Quadrática	-	-	-	-
	Retangular	0,294	0,232	1,270	0,204
TEMPO		0,141	0,168	0,838	0,402
ESPACAMENTO*OCUPACAO	10.000ha*Retangular	0,263	0,337	0,781	0,435

**Tabela 4.12.** Estimativas dos Coeficientes e seus respectivos p-valores do modelo logístico com retas concorrentes e inclusão do efeito aleatório referente às plantas

	Categoria	Estimativa	Erro-padrão	Estatística z	p-valor
(Intercept)		-0,294	0,518	-0,568	0,570
ESPACAMENTO	6.000ha	-	-	-	-
	10.000ha	0,864	0,738	1,171	0,242
OCUPACAO	Quadrática	-	-	-	-
	Retangular	-0,109	0,733	-0,149	0,882
TEMPO		0,107	0,327	0,327	0,743
ESPACAMENTO*OCUPACAO	10.000ha*Retangular	0,035	1,061	0,033	0,973
ESPACAMENTO*TEMPO	10.000ha*Tempo	-0,271	0,466	-0,583	0,560
OCUPAÇÃO*TEMPO	Retangular*Tempo	0,270	0,464	0,581	0,561
ESPACAMENTO*TEMPO	10.000ha*Retangular*Tempo	0,152	0,674	0,226	0,821

**Tabela 4.13.** Estimativas dos Coeficientes e seus respectivos p-valores do modelo logístico com retas com o mesmo intercepto e coeficientes angulares diferente e inclusão do efeito aleatório referente às plantas

	Categoria	Estimativa	Erro-padrão	Estatística z	p-valor
(Intercept)		0,076	0,263	0,289	0,773
ESPACAMENTO*OCUPACAO*TEMPO	6.000ha*Quadrática*Tempo	-0,115	0,189	-0,609	0,542
	10.000ha*Quadrática*Tempo	0,132	0,190	0,695	0,487
	6.000ha*Retangular*Tempo	0,088	0,189	0,467	0,641
	10.000ha*Retangular*Tempo	0,515	0,200	2,575	0,010

Os valores dos AIC e BIC foram utilizados como critérios de seleção do modelo que melhor se ajusta aos dados do estudo. O modelo que apresentou menor valor de AIC (811,2) e BIC (837,6) foi o modelo de retas paralelas. Para este, nota-se que ao nível de significância de 0,05, a única variável significativa foi a “ESPACAMENTO” (ver Tabela 4.11).

**Tabela 4.14.** Estimativas dos Coeficientes e seus respectivos p-valores do modelo logístico sem o efeito de tratamento e inclusão do efeito aleatório referente às plantas

	Estimativa	Erro-padrão	Estatística z	p-valor
(Intercept)	0,092	0,262	0,350	0,726
TEMPO	0,137	0,165	0,826	0,409

## 5 DISCUSSÃO E CONCLUSÃO

A partir do estudo realizado na Seção 4.2, foi possível observar que o espaçamento e o tipo de ocupação podem interferir significativamente no processo de frutificação de plantas de café. Quanto ao espaçamento, a chance de frutificação pode chegar a praticamente, se o espaçamento for adequado. Também a forma interfere fortemente na chance de frutificação. O estudo realizado (Seção 4.2.2) indica que a forma retangular de ocupação tem chance muito maior (1,81) de frutificação que a ocupação quadrática. O estudo da Seção 4.3, que trata dos modelos mistos na tentativa de incorporar a variável tempo e a estrutura de variabilidade entre as plantas de café. Desta forma, foram considerados quatro modelos diferentes quanto a estrutura dos efeitos fixos, considerando aditivo o efeito aleatório relativo às plantas em todos eles. Através dos critérios AIC e BIC o modelo que melhor ajustou os dados foi o de retas paralelas, sendo considerada apenas a variável “ESPAÇAMENTO” significativa para a estimação da probabilidade de frutificação das plantas. Como investigações futuras, seria interessante a modelagem da probabilidade de refrutificação dos metâmeros a partir desse estudo preliminar e exploratório considerando as mais diversas variáveis presentes nesse banco de dados. Um estudo envolvendo aspectos relativos a alometria de plantas poderão ser analisadas de forma mais complexa associando os dois bancos de dados obtidos nos dois tempos de coleta.



## REFERÊNCIAS

- AKAIKE, H., 1974 A new look at the statistical model identification. *IEEE Transactions on automatic control* **19**: 716–723.
- ALCARDE, R., 2012 *Modelos lineares mistos em dados longitudinais com o uso do pacote ASReml-R*. Ph.D. thesis, Escola Superior de Agricultura Luiz de Queiroz.
- ANDRADE, V. T., 2013 *Implicações da modelagem estatística na seleção de progênies de cafeeiro*. Ph.D. thesis, Universidade Federal de Lavras.
- BURNHAM, K. P. and D. R. ANDERSON, 2004 Multimodel inference understanding aic and bic in model selection. *Sociological methods & research* **33**: 261–304.
- DA COSTA, S. C., 2003 *Modelos lineares generalizados mistos para dados longitudinais*. Ph.D. thesis, Universidade de São Paulo.
- DE RESENDE, M. D. V., 2007 *Matemática e estatística na análise de experimentos e no melhoramento genético*. Embrapa Florestas, Colombo **362**.
- EFRON, B. and C. N. MORRIS, 1977 *Stein's paradox in statistics*. WH Freeman.
- EISENHART, C., 1947 The assumptions underlying the analysis of variance. *Biometrics* **3**: 1–21.
- FAZUOLI, L. C., O. GUERREIRO FILHO, M. B. SILVAROLLA, H. P. MEDINA FILHO, and A. CARVALHO, 2005 Evaluation of the cultivars mundo novo, bourbon amarelo and bourbon vermelho of coffea arabica l., in Campinas, state of São paulo, Brazil. *Bragantia* **64**: 533–546.
- HENDERSON, C. R., 1950 Estimation of genetic parameters. In *Biometrics*, volume 6, pp. 186–187, INTERNATIONAL BIOMETRIC SOC 1441 I ST, NW, SUITE 700, WASHINGTON, DC 20005-2210.
- HENDERSON, C. R., 1975 Best linear unbiased estimation and prediction under a selection model. *Biometrics* pp. 423–447.
- LITTELL, R. C., 2002 Analysis of unbalanced mixed model data: a case study comparison of anova versus reml/gls. *Journal of Agricultural, Biological, and Environmental Statistics* **7**: 472–490.
- LUSH, J. L., 1931 The number of daughters necessary to prove a sire. *Journal of Dairy Science* **14**: 209–220.
- LUSH, J. L., 1947 Family merit and individual merit as bases for selection. part i. *The American Naturalist* **81**: 241–261.
- MATSUNAGA, F. T., W. L. PENSE, J. D. BRANCHER, M. RAKOCEVIC, and P. PRUDENTE-SP, 2015 Módulo generalizado para reconstrução e visualização 3d de cafeeiros com frutos baseado em um banco de dados de coffea arabica .
- MATSUNAGA, F. T., J. B. TOSTI, A. ANDROCIOLI-FILHO, J. D. BRANCHER, E. COSTES, and M. RAKOCEVIC, 2016 Strategies to reconstruct 3d coffea arabica l. plant structure. SpringerPlus .
- PIEPHO, H., A. BÜCHSE, and C. RICHTER, 2004 A mixed modelling approach for randomized experiments with repeated measures. *Journal of Agronomy and Crop Science* **190**: 230–247.
- RAKOCEVIC, M., J. B. TOSTI, F. T. MATSUNAGA, and C. C. P. CAFÉ, 2013 Amostracafe3d–software para inclusão de informações na escala de metâmeros em eixos plagiotrópicos parcialmente codificados .

- SCHWARZ, G. *et al.*, 1978 Estimating the dimension of a model. *The annals of statistics* **6**: 461–464.
- SEARLE, S. R., 1991 Cr henderson, the statistician; and his contributions to variance components estimation. *Journal of dairy science* **74**: 4035–4044.
- SINGER, J. M. and D. D. ANDRADE, 1986 Análise de dados longitudinais. *Simpósio Nacional de Probabilidade e Estatística* **7**.
- SMITH, A., B. R. CULLIS, and R. THOMPSON, 2005 The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *The Journal of Agricultural Science* **143**: 449–462.
- TEAM DIGIPLANTE, 2016 The equations of plant growth and development.
- VALLEJO, G., M. P. FERNÁNDEZ, P. E. LIVACIC-ROJAS, and E. TUERO-HERRERO, 2011 Selecting the best unbalanced repeated measures model. *Behavior research methods* **43**: 18–36.
- WHITE, T. and G. HODGE, 1988 Best linear prediction of breeding values in a forest tree improvement program. *Theoretical and applied genetics* **76**: 719–727.

## APÊNDICES

### COMANDOS EXECUTADOS NO SOFTWARE R

```
setwd("C:/Users/Luiza/Desktop/DADOS TESE - NOVO/Obs4/DADOS EXTRAIDOS")
```

```
DADOS_T1 <- read.csv("DADOS_T1.csv", header=T, sep=";", dec=".")
attach(DADOS_T1)
str(DADOS_T1)
```

```
DADOS_T1$PLANTA<-as.factor(DADOS_T1$PLANTA)
DADOS_T1$REPETICAO<-as.factor(DADOS_T1$REPETICAO)
DADOS_T1$ESPACAMENTO<-as.factor(DADOS_T1$ESPACAMENTO)
DADOS_T1$OCUPACAO<-as.factor(DADOS_T1$OCUPACAO)
DADOS_T1$FRUTIFICOU<-as.factor(DADOS_T1$FRUTIFICOU)
```

```
summary(DADOS_T1)
```

```
modelosimples1 <- glm(FRUTIFICOU~ESPACAMENTO,
                     family=binomial(link="logit"), data=DADOS_T1)
summary(modelosimples1)
```

```
modelosimples2 <- glm(FRUTIFICOU~OCUPACAO,
                     family=binomial(link="logit"), data=DADOS_T1)
summary(modelosimples2)
```

```
modelosimples3 <- glm(FRUTIFICOU~TAM_EN,
                     family=binomial(link="logit"), data=DADOS_T1)
summary(modelosimples3)
```

```
round(exp(modelosimples1$coef),2)
```

```
DADOS_T2 <- read.csv("DADOS_T2.csv", header=T, sep=";", dec=".")
attach(DADOS_T2)
str(DADOS_T2)
```

```
DADOS_T2$PLANTA<-as.factor(DADOS_T2$PLANTA)
DADOS_T2$REPETICAO<-as.factor(DADOS_T2$REPETICAO)
DADOS_T2$ESPACAMENTO<-as.factor(DADOS_T2$ESPACAMENTO)
DADOS_T2$OCUPACAO<-as.factor(DADOS_T2$OCUPACAO)
DADOS_T2$FRUTIFICOU<-as.factor(DADOS_T2$FRUTIFICOU)
```

```
summary(DADOS_T2)
```



```
modelosimples1.1 <- glm(FRUTIFICOU~ESPACAMENTO,
                        family=binomial(link="logit"), data=DADOS_T2)
summary(modelosimples1.1)

modelosimples2.2 <- glm(FRUTIFICOU~OCUPACAO,
                        family=binomial(link="logit"), data=DADOS_T2)
summary(modelosimples2.2)

modelosimples3.3 <- glm(FRUTIFICOU~TAM_EN,
                        family=binomial(link="logit"), data=DADOS_T2)
summary(modelosimples3.3)

modelosimples4.4 <- glm(FRUTIFICOU~ESPACAMENTO+OCUPACAO,
                        family=binomial(link="logit"), data=DADOS_T2)
summary(modelosimples4.4)

modelosimples5.5 <- glm(FRUTIFICOU~ESPACAMENTO+OCUPACAO+ESPACAMENTO*OCUPACAO,
                        family=binomial(link="logit"), data=DADOS_T2)
summary(modelosimples5.5)

round(exp(modelosimples4.4$coef),2)

DADOS <- read.csv("DADOS2.csv", header=T, sep=";", dec=".")
attach(DADOS)
str(DADOS)

DADOS$PLANTA<-as.factor(DADOS$PLANTA)
DADOS$REPETICAO<-as.factor(DADOS$REPETICAO)
DADOS$ESPACAMENTO<-as.factor(DADOS$ESPACAMENTO)
DADOS$OCUPACAO<-as.factor(DADOS$OCUPACAO)
DADOS$FRUTIFICOU<-as.factor(DADOS$FRUTIFICOU)

summary(DADOS)

require(ggplot2)
require(GGally)
require(reshape2)
require(lme4)
require(compiler)
require(parallel)
require(boot)
require(hnp)
```

```
# modelo: paralelas
MOD1 <- glmer(cbind(FRUTIFICOU==1, FRUTIFICOU==0) ~
              ESPACAMENTO*OCUPACAO + TEMPO +
              (1|PLANTA), DADOS,
              family=binomial())
hnp(MOD1)
summary(MOD1)

# modelo: concorrentes
MOD2 <- glmer(cbind(FRUTIFICOU==1, FRUTIFICOU==0) ~
              ESPACAMENTO*OCUPACAO*TEMPO +
              (1|PLANTA), DADOS,
              family=binomial())
hnp(MOD2)
summary(MOD2)
anova(MOD1, MOD2)

# modelo: mesmo intercepto e coef ang diferentes
MOD3 <- glmer(cbind(FRUTIFICOU==1, FRUTIFICOU==0) ~
              ESPACAMENTO:OCUPACAO:TEMPO +
              (1|PLANTA), DADOS,
              family=binomial())
hnp(MOD3)
summary(MOD3)
anova(MOD1, MOD3)
anova(MOD2, MOD3)
```