

**CENTRO UNIVERSITÁRIO DE ARARAQUARA**  
**MESTRADO PROFISSIONAL EM ENGENHARIA DE PRODUÇÃO**

**Marcelo Edmundo Alves Martins**

**Utilização de Métodos Bayesianos na Avaliação do Desempenho de Equipes  
de Colheita: uma Aplicação no Setor de Citricultura do Estado de São Paulo**

Dissertação de mestrado apresentada ao Programa de Mestrado Profissional em Engenharia de Produção do Centro Universitário de Araraquara – UNIARA – como parte dos requisitos para obtenção do título de Mestre em Engenharia de Produção. Área de Concentração: Gestão Estratégica e Operacional da Produção.

Orientador: Prof. Dr. Jorge Alberto Achcar

Araraquara, SP – Brasil  
2014

## **FICHA CATALOGRÁFICA**

M344u Martins, Marcelo Edmundo Alves

Utilização de Métodos Bayesianos na Avaliação do Desempenho de Equipes de Colheita: uma Aplicação no Setor de Citricultura do Estado de São Paulo/ Marcelo Edmundo Alves Martins - Araraquara: Centro Universitário de Araraquara, 2015.

62f

Dissertação - Mestrado Profissional em Engenharia de Produção  
Centro Universitário de Araraquara - UNIARA

Orientador: Prof. Dr. Jorge Alberto Achcar

1. Indicadores de desempenho. 2. Modelos de regressão linear. 3. Modelo de regressão de Poisson. 4. Análise Bayesiana. 5. Métodos de Monte Carlo em Cadeias de Markov.

CDU 62-1

## **REFERÊNCIA BIBLIOGRÁFICA**

MARTINS, M.E.A . Utilização de Métodos Bayesianos na Avaliação do Desempenho de Equipes de Colheita: uma Aplicação no Setor de Citricultura do Estado de São Paulo. 2015. 62f. Dissertação de Mestrado em Engenharia de Produção – Centro Universitário de Araraquara, Araraquara-SP.

## **ATESTADO DE AUTORIA E CESSÃO DE DIREITOS**

NOME DO AUTOR: Marcelo Edmundo Alves Martins

TÍTULO DO TRABALHO: Utilização de Métodos Bayesianos na Avaliação do Desempenho de Equipes de Colheita: uma Aplicação no Setor de Citricultura do Estado de São Paulo

TIPO DO TRABALHO/ANO: Dissertação / 2015

Conforme LEI Nº 9.610, DE 19 DE FEVEREIRO DE 1998, o autor declara ser integralmente responsável pelo conteúdo desta dissertação e concede ao Centro Universitário de Araraquara permissão para reproduzi-la, bem como emprestá-la ou ainda vender cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta dissertação pode ser reproduzida sem a sua autorização.



\_\_\_\_\_  
Marcelo Edmundo Alves Martins

Rua Voluntários da Pátria, 1309 e 1295 – Centro 14801-320 - Araraquara –  
marcelostatistics@gmail.com

**Dissertação aprovada em sua versão final pela banca examinadora:**



---

**Prof. Dr. Jorge Alberto Achear**  
Orientador(a) – UNIARA



---

**Prof. Dr. Francisco Aparecido Rodrigues**  
Universidade de São Paulo - USP



---

**Prof. Dr. Walther Azzolini Junior**  
Centro Universitário de Araraquara - UNIARA

**Araraquara, 11 de setembro de 2014**

## **AGRADECIMENTOS**

A Deus por me amparar nos momentos difíceis, me dar força interior para superar as dificuldades, mostrar os caminho nas horas incertas e me suprir em todas as minhas necessidades.

Gostaria de agradecer ao meu orientador Jorge Alberto Achcar por estar sempre presente, pela dedicação e a confiança depositada em mim em vários momentos distintos deste processo.

A meus pais, Arlindo e Rosa, meu infinito agradecimento. Sempre acreditaram em minha capacidade e me acharam O MELHOR de todos, mesmo não sendo. Isso só me fortaleceu e me fez tentar, não ser O MELHOR, mas afazer o melhor de mim.

Ao meu filho Matheus, criança alegre e de personalidade marcante, cujas brincadeiras me inspiram nos momentos de maiores dificuldades, e cujo olhar seguro e intenso me acolhe, cujas palavras firmes e sinceras demonstram todo o carinho que tem por mim.

A minhas irmãs, Viviani e Priscila, meu agradecimento especial pois, a seu modo, sempre se orgulharam de mim e confiaram em meu trabalho.

Ao meu amigo Rafael Câmara que me auxiliou em momentos importantes deste trabalho.

.

## RESUMO

O presente estudo tem como objetivo identificar, sob o enfoque de engenharia de produção, os principais fatores que contribuem para o bom desempenho das equipes de colheita na citricultura. Para isso, é considerado em um estudo de caso, a coleta de um conjunto de indicadores referentes a um grande número de equipes de colheita de uma empresa do setor de citros do interior do estado de São Paulo. Pretende-se, assim, verificar a relação existente entre variáveis do processo e os indicadores gerais de desempenho. Para a análise dos dados, consideramos modelos de regressão linear múltipla para os dados transformados e modelos de regressão de Poisson, sob um enfoque Bayesiano. Com isso, conclui-se que a quantidade de safras trabalhadas por equipe, o líder (idade e sexo), a quantidade de colhedores e a porcentagem de trabalhadores do sexo masculino (dentre outras covariáveis) têm impacto significativo sobre o indicador de desempenho (volume colhido).

***Palavras-chave:** Indicadores de desempenho ; modelos de regressão linear ; modelo de regressão de Poisson; análise Bayesiana; métodos de Monte Carlo em Cadeias de Markov.*

## ABSTRACT

This study aims to identify under a production engineering approach, the main factors that affect the performance of teams in fruit harvest. In this way, it was considered in a case study, some indicators which could be related to the performance of different teams working in the fruit harvest of a industry of the sector in the state of São Paulo. In this study, we want to examine the relationship between variables the process and some general performance indicators. For the data analysis, we considered multiple linear regression models where the response of interest was transformed to other scale to satisfy standard statistical properties and Poisson regression models. Under a Bayesian approach considered for the Poisson regression models, it was concluded that the fruit harvest volume was affected by some factors as the team leader (age and sex), the amount of lanyardsand and the percentage of male workers (among other covariable), have significant impact on the performance indicator (harvested volume).

**Keywords:** *Performance indicators; Linear regression models, Poisson regression model, Bayesian analysis, Markov chains Monte Carlo methods.*

## Lista de figuras

Figura 1.1 - Custos de produção de laranja .....	12
Figura 1.2 - Épocas de colheita .....	13
Figura 1.3 - Materiais de colheita.....	13
Figura 1.4 - Talhões de laranja .....	14
Figura 1.5 - Exemplo de distribuição de colhedores em um talhões de laranja .....	15
Figura 4.1 - Gráficos dos resíduos.....	29
Figura 5.1 - Gráficos dos valores observados e médias ajustadas versus observações.....	35
Figura 5.2 - Gráficos dos valores observados de caixas e médias ajustadas versus observações.....	36
Figura 6.1. Gráficos dos resíduos .....	41

## Lista de Quadros e Tabelas

Tabela 1.1 - Principais Frutas Produzidas no Brasil.....	10
Tabela 4.1 - Estimadores e valores-p para os coeficientes de regressão considerando a resposta $\log(\text{total caixas})$ .....	26
Tabela 4.2 - Estimadores e valores-p para os coeficientes de regressão considerando a resposta $\log(\text{produção média})$ .....	28
Tabela 4.3 - Estimadores (EMV) e valores-p para os coeficientes de regressão logística .....	31
Tabela 5.1 - Sumários a posteriori-regressão de Poisson (total caixas) .....	34
Tabela 5.2 - Sumários a posteriori-regressão de Poisson (produção média).....	35
Tabela 5.3 - Sumários a posteriori- regressão logística(produção diária acima da média) .....	37
Tabela 6.1- Estimadores e valores-p para os coeficientes de regressão considerando a resposta $\log(\text{totalcaixas})$ .....	39
Tabela 6.2 - Estimadores e valores-p para os coeficientes de regressão considerando a resposta $\log(\text{produção média})$ .....	40
Tabela 6.3 - Estimadores (EMV) e valores-p para os coeficientes de regressão logística .....	42
Tabela 6.4 - Sumários a posteriori; regressão de Poisson (total caixas) .....	43
Tabela 6.5 - Sumários a posteriori-regressão de Poisson (produção média).....	44



## Sumário

1. Introdução.....	8
1.1 Problema de Pesquisa .....	14
1.2 Questão da Pesquisa .....	14
1.3 Objetivos.....	14
1.3.1 Objetivo Geral.....	15
1.3.2 Objetivos Específicos.....	15
1.4 Justificativa.....	15
1.5 Metodologia de Pesquisa .....	15
1.6 Estrutura do trabalho .....	16
2. Revisão Bibliográfica .....	17
2.1 Algumas considerações sobre modelos de regressão linear múltipla.....	17
2.2 Algumas considerações sobre análise Bayesiana .....	19
3. Dados Coletados para o estudo.....	21
4. Análise estatística dos dados .....	22
4.1 Análise de regressão linear múltipla dos dados .....	22
4.2 Análise Variável resposta: log(total caixas) .....	23
4.3 Variável resposta: log(total caixas) .....	25
4.4 Análise de regressão logística para a resposta binária produção diária acima do piso (sim = 1 e não = 0) - uso do software MINITAB versão 16.....	27
5. Análise Bayesiana dos dados.....	29
5.1 Variável resposta: total caixas .....	31
5.2 Variável resposta: produção média.....	32
5.3 Análise Bayesiana da regressão Logística para produção acima do piso .....	34
6. Reanálise dos dados considerando as covariáveis contínuas padronizadas .....	36

6.1 Variável resposta: log(total de caixas).....	36
6.2 Variável resposta: log(produção média).....	37
6.3 Análise de regressão logística para a resposta binária produção diária acima do piso (sim = 1 e não = 0) - uso do software MINITAB versão 16.....	39
6.4 Análise Bayesiana assumindo um modelo de regressão de Poisson .....	40
6.4.1 Variável resposta: total caixas .....	40
6.4.2 Variável resposta: produção média.....	41
7. Algumas conclusões e perspectivas futuras.....	43
Referências .....	44
Apêndices .....	49
Apêndice 1 – Descritivo das variáveis a serem estudadas.....	49
Apêndice 2: Uma breve revisão da metodologia bayesiana .....	52
Apêndice 3: Artigos produzidos .....	56

## 1. Introdução

Atualmente, o Brasil tem se mantido na posição de terceiro produtor mundial de frutas, perdendo apenas para a China e a Índia, de modo que a Fruticultura é hoje um dos segmentos mais importantes da Agricultura Brasileira, respondendo por 25% do valor da produção agrícola nacional, de acordo com o Anuário da Fruticultura Brasileira, publicado em 2012.

Essa atividade envolve, no Brasil, mais de 5 milhões de pessoas, e encontra-se distribuída nas principais regiões produtoras do Sudeste, Nordeste e Sul. Em 2010, o estado de São Paulo respondeu por 32,9% da oferta nacional de frutas frescas, conforme pesquisa de Produção Agrícola Municipal (PAM) do IBGE.

Na Tabela 1.1, são apresentados os dados das quantidades das principais frutas produzidas no Brasil nos anos 2011 e 2012.

**Tabela 1.1** Principais Frutas Produzidas no Brasil

Estimativas de produção brasileira (em t)		
Frutas	2011	2012
Laranja	19.655.469	18.030.413
Banana	7.023.396	6.980.192
Abacaxi*	1.519.881	1.455.056
Coco-da-baía*	1.899.355	1.786.498
Uva	1.463.481	1.387.830
Maçã	1.364.953	1.208.658

Fonte: IBGE (Janeiro de 2012) - \* Em mil frutos. Conversão: 1 fruto = 2,5 Kg (Região Sul-Sudeste, exceto PR (1,6 Kg) e SC (1,67 Kg)), 2,1 Kg (Região Centro-Oeste) e 1,8 Kg (para as demais regiões).

O plantio da laranja no Brasil, que teve seu início no período de colonização, tem duas destinações principais: mercado *in natura* e industrialização, que se dá principalmente para o suco.

O Brasil detêm 50% da produção mundial do suco de laranja e exporta 98% do que produz. O estado de São Paulo representa 53% do total da produção.

Juntos, o estado da Flórida, nos Estados Unidos da América, e o estado de São Paulo, detêm 81% da produção mundial de suco.

Atualmente, a colheita dos frutos cítricos tem sido considerada uma das atividades mais críticas de todo o processo produtivo (Citrus BR 2012). A colheita manual baseia-se na utilização dos principais sentidos do ser humano, tais como visão, tato, etc. Esse método possui vantagens e desvantagens. Entre as vantagens, observa-se que o ser humano é completo em relação aos sentidos, visão, tato, olfato, podendo melhor empreender a colheita. Colhedores mais cuidadosos em campo podem ocasionar menor perda da fruta *in natura*.

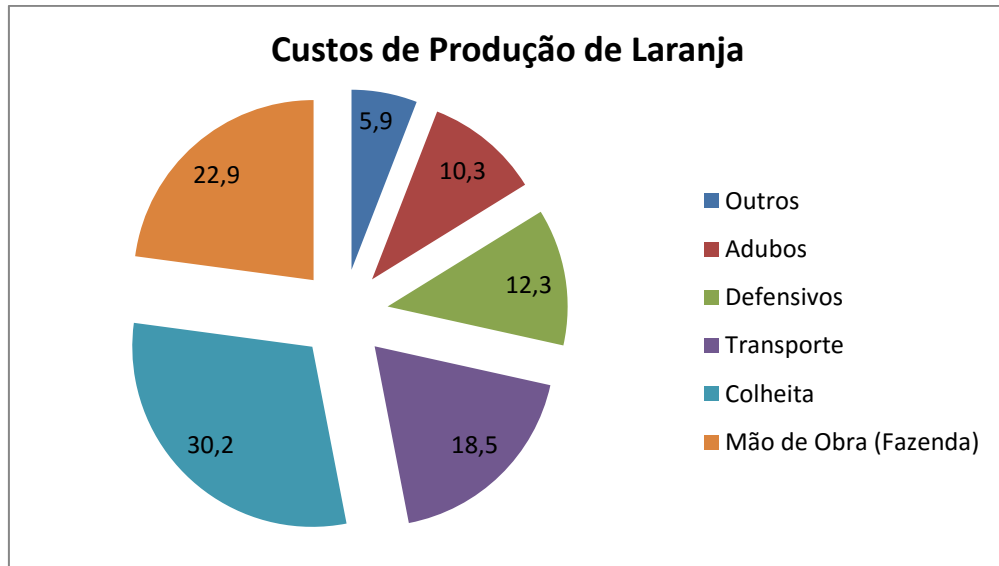
A técnica mais comumente utilizada é a da torção seguida da retirada do fruto (PETTO NETO; POMPEU JUNIOR, 1991).

A seleção e empacotamento podem ser realizados no campo, portanto com menor número de etapas. Entre as possíveis desvantagens observa-se o alto custo da mão de obra em algumas regiões; além disso, essa mão de obra muitas vezes não é treinada e pode ser desqualificada para tal operação, o que pode ocasionar problemas diversos. A sazonalidade relacionada à oferta de mão de obra pode ser também um desafio para diversas regiões (CORTEZ, 2002).

Além dos aspectos de qualidade e logística, o valor dispendido para realização da colheita têm impacto relevante em toda cadeia de produção citrícola, equivalendo a uma média de 25% a 35% do custo total de produção (POZZAN E TRIBONI, 2005).

Na figura 1.1, encontra-se os custos de produção de laranja, onde observa-se o grande percentual de custo relativo à mão de obra na colheita.

**Figura 1.1** - Custos de produção de laranja



Fonte: Citrus BR 2012

No estado de São Paulo utiliza-se de 70.000 funcionários para realização da operação de colheita. Paralelamente ao encarecimento da mão de obra, o setor produtivo tem enfrentado também a escassez de trabalhadores, sobretudo devido à competição com a construção civil, o que vem levando muitos produtores a mecanizarem seu sistema de produção no que for possível (Revista Hortifrut Brasil - CEPEA – ESALQ/USP, 2012).

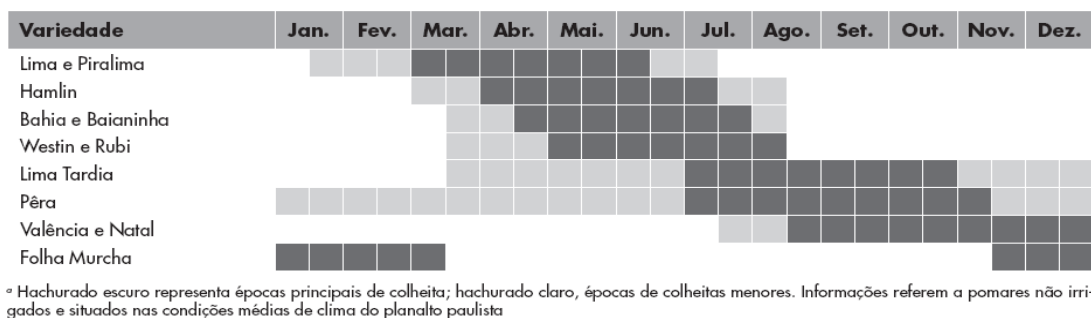
No gerenciamento da operação de colheita poucos indicadores de desempenho são utilizados para avaliação de rendimento, qualidade e atendimento. (TACHIBANA E RIGOLIN, 2002).

Com a mecanização da colheita de cana de açúcar na ordem de 68%, era esperado que parte dessa mão de obra migrasse para a colheita de citrus, fato que não ocorreu, pois grande parte migrou para o setor de construção civil (Revista Hortifrut Brasil - CEPEA – ESALQ/USP 2012).

Outro fator que impacta em relação ao custo é a sazonalidade da cultura, sendo que 95% das variedades com valor industrial são colhidas no período de maio a janeiro (Revista Hortifrut Brasil - CEPEA – ESALQ/USP, 2012).

Na Figura 1.2, observa-se as épocas de colheita para diferentes variedades de laranja.

**Figura 1.2** – Épocas de colheita



Fonte: Citrus BR

A produção de citrus normalmente é colhida manualmente, sendo que os colhedores utilizam-se de sacolinhas (capacidade de 27,2 Kg), escadas de comprimento de 3,5m ou 4,5m e *big bags* com capacidade individual de armazenamento de 20 sacolinhas ou 540 kg. (DAVIES E ALBRIGO, 1994). Ver figura 1.3

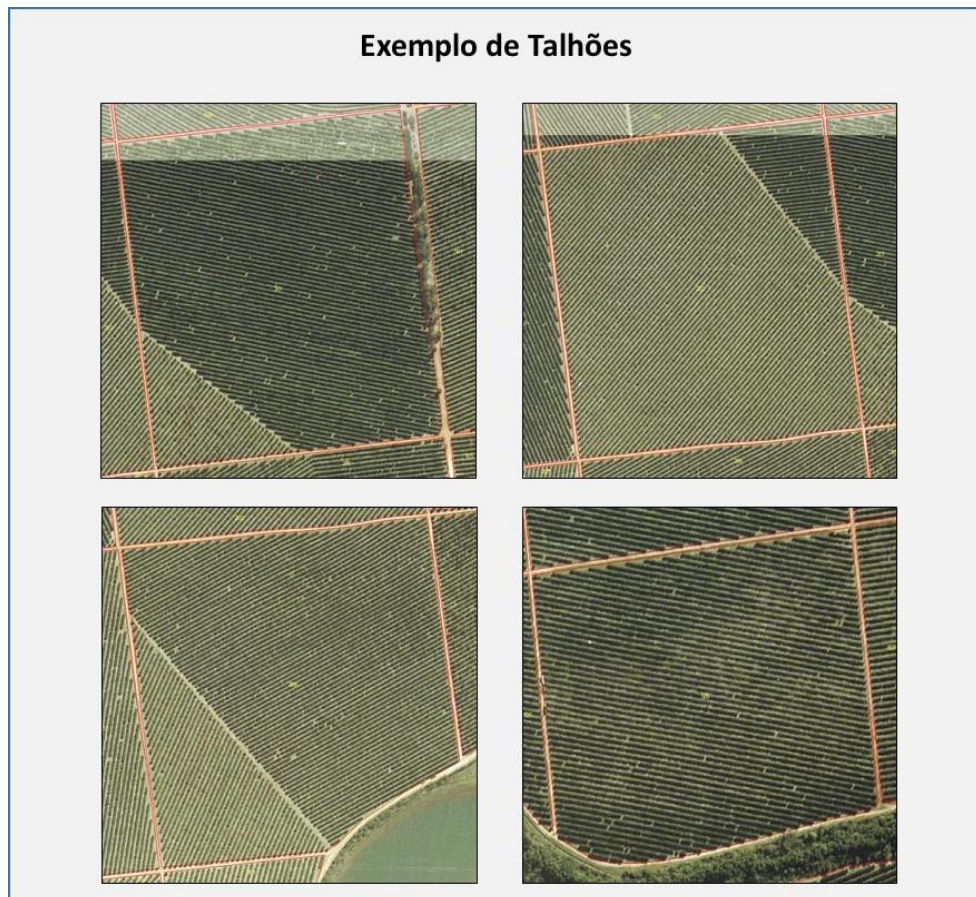
**Figura 1.3** – Materiais de Colheita



Fonte: Citrus BR

Os colhedores realizam a colheita em áreas que são denominadas talhões de laranja (figura 1.4) temos alguns exemplos destes talhões:

**Figura 1.4** Exemplo de Talhões de laranja



Os colhedores são dispostos nos talhões de colheita em regiões de colheita, que são determinadas por:

- 1) Produção das árvores de laranja;
- 2) Produtividade dos colhedores;
- 3) Posicionamento das bancas(ruas onde os caminhões de carregamento podem trafegar);
- 4) Facilidade para o gerenciamento da equipe de colheita.





dispostas em pelo menos 150 municípios paulistas, sendo Barretos, São José do Rio Preto, Araraquara, Avaré e Bauru as regiões de maior predominância.

### **1.1 Problema de Pesquisa**

Com base na revisão da literatura, a questão que se coloca é: como uma empresa contrata equipes de colheita, buscando melhorar sua produtividade, qualidade e custo?

### **1.2 Questão da Pesquisa**

O objetivo é identificar as covariáveis que efetivamente impactam no desempenho das equipes de colheita na citricultura. Metodologicamente este trabalho pode ser classificado como aplicado, de objetivo descritivo e abordagem quantitativa. Bertrand e Fransoo (2002) definem a pesquisa quantitativa em Engenharia de Produção como aquela em que se modela um problema cujas variáveis apresentam relações causais e quantitativas. Neste sentido, torna-se possível quantificar o comportamento das variáveis dependentes sob um domínio específico, permitindo ao pesquisador realizar previsões. Em geral, as pesquisas quantitativas utilizam modelagem matemática, estatística ou computacional (simulação) – especificamente, neste trabalho será adotada a modelagem estatística. Quanto às técnicas de pesquisa serão utilizadas a pesquisa bibliográfica e a observação direta intensiva, segundo a classificação de Lakatos e Marconi (2008) ou a pesquisa bibliográfica e pesquisa ação, conforme a classificação de Gil (2008).

O método proposto é replicável para outras culturas agrícolas, o que deve contribuir para respostas à várias questões de interesse.

### **1.3 Objetivos**

### **1.3.1 Objetivo Geral**

Propor uma modelagem estatística como ferramenta de apoio à contratação de equipes de colheita, com foco na redução de custo, melhoria da qualidade e aumento de produtividade.

### **1.3.2 Objetivos Específicos**

1. Descobrir quais covariáveis impactam positivamente e negativamente na produtividade de turmas de colheita;
2. Descobrir qual tipo de correlação existe entre as variáveis independentes e a variável dependente (produtividade);
3. Comparação da abordagem estatística clássica com a abordagem bayesiana na análise dos dados.

### **1.4 Justificativa**

Vários trabalhos relativos à área de citrus apontam a operação de colheita como sendo o maior custo na produção de laranja (TACHIBANA E RIGOLIN, 2002; POZZAN E TRIBONI, 2005; DAVIES E ALBRIGO, 1994).

A competição mundial que muitos setores vêm enfrentando nas últimas décadas, tem levado as empresas a melhorar continuamente seus desempenhos. Com isso, grande importância tem sido dada às descobertas de fatores que impactam na produtividade.

Outro ponto importante que justifica a realização deste trabalho está relacionado aos setores que também trabalham com mão-de-obra rural e que têm buscado a mecanização devido a falta de mão de obra, bem como a existência de problemas ambientais. Dessa forma, os custos de colheita têm tido um crescimento muito grande na última década (Cepea: custos de colheita, 2003/2012).

### **1.5 Metodologia de Pesquisa**

Os métodos a serem empregados em pesquisas científicas podem ser selecionados desde a identificação do problema, formulação das hipóteses ou delimitação do universo ou da amostra. A seleção destes aspectos dependerá de vários fatores relacionados com o estudo, tais como a natureza dos fenômenos, o objeto de pesquisa, os recursos, a abordagem do estudo (qualitativa ou quantitativa, ou uma combinação de ambas), entre outros (MARCONI E LAKATOS, 2010).

Na presente pesquisa, o método escolhido foi a da pesquisa ação com dados coletados de uma indústria do setor de citros do estado de São Paulo. Esse método tem sido, consistentemente, um dos mais poderosos em gestão de operações, particularmente no desenvolvimento de novas teorias sobre o problema. Para lidar com a crescente frequência e magnitude das mudanças na tecnologia e métodos de gestão, operações de gestão de pesquisadores têm considerado, cada vez mais, o emprego de métodos baseados em pesquisa de campo (LEWIS, 1998).

Um roteiro dessa metodologia é introduzida por Meredith (1998) que cita três pontos fortes pendentes de investigação de caso, também apresentados por Bebensat et al. (1987):

(1) O fenômeno pode ser estudado em seu ambiente natural e teoria, significado relevante gerado a partir do conhecimento adquirido através da observação prática.

(2) O método baseado no estudo de caso permite que as questões do por que, o quê e como, possam ser respondidas com uma compreensão relativamente completa da natureza e complexidade do fenômeno.

(3) O método baseado no estudo de caso se presta a investigações exploratórias precoces, nas quais as variáveis ainda são desconhecidas e o fenômeno ainda não é todo compreendido.

## **1.6 Estrutura do trabalho**

O trabalho abordará, na seção 2, uma breve revisão bibliográfica, que será dividida em sub-itens para fundamentar a Análise de Regressão Múltipla e a Análise Bayesiana; na seção 3, será apresentada a descrição dos dados coletados para a pesquisa; na seção 4 será apresentada a análise dos dados coletados sob o enfoque clássico; na seção 5 será apresentada a análise dos dados coletados sob o enfoque bayesiano; na seção 6 será apresentada a análise dos dados padronizados. Finalmente, na seção 7, serão apresentadas algumas conclusões dos resultados obtidos.

## **2. Revisão Bibliográfica**

### **2.1 Algumas considerações sobre modelos de regressão linear múltipla**

A técnica de regressão linear múltipla é uma das técnicas estatísticas mais amplamente empregada para se obter previsões de interesse (HAIR, BLACK, ROLPH E ANDERSON, 2005).

Em estatística, regressão linear é uma abordagem para modelar a relação entre uma variável  $Y$  com uma ou mais variáveis dependentes ou explicativas, denotadas por  $X$ . O caso de apenas uma variável explicativa é chamado de modelo de regressão linear simples. Com mais de uma variável explicativa, o modelo é chamado de regressão linear múltipla.

Modelos de regressão linear são amplamente utilizados em ciências biológicas, comportamentais, econômicas, sociais e engenharia para descrever possíveis relações entre as variáveis. É considerado como um dos mais importantes instrumentos utilizados nessas áreas.

Na área de engenharia de produção, podemos citar algumas das aplicações apresentada por Carvalho, Sedyama, Cecon e Alves (2004) ou por Jordan e Letti (2011).

Na análise de regressão linear, os dados são modelados utilizando-se funções de previsão linear, e os parâmetros do modelo são desconhecidos e estimados a partir dos dados. Tais modelos são chamados modelos lineares. Mas, geralmente, refere-se a regressão linear para um modelo em que a média condicional de  $Y$ , dado o valor de  $X$ , é uma função de  $X$ .

O método de regressão linear tem muitos usos práticos. A maioria das aplicações se enquadram em uma das seguintes duas grandes categorias:

- Se o objetivo é a predição ou previsão, modelos de regressão linear podem ser utilizados para ajustar um modelo preditivo para um conjunto de dados observados de valores  $Y$  e  $X$ . Após o ajuste de um modelo desse tipo, se um valor adicional de  $X$  é dado, então o modelo ajustado pode ser usado para fazer uma previsão do valor de  $y$ .
- Dadas as variáveis ou covariáveis  $X_1, \dots, X_p$ , que podem estar relacionadas com a resposta ou variável dependente  $Y$ , a análise de regressão linear, pode ser utilizada para quantificar a magnitude da relação entre  $Y$  e  $X_j$ ,  $j=1, \dots, p$ . Isso é dado por testes de hipóteses nos parâmetros de regressão.

Modelos de regressão linear são frequentemente ajustados usando-se a abordagem de mínimos quadrados.

Quando utilizamos mais de uma variável explanatória para prever o comportamento de uma variável resposta, passamos a nomeá-lo como modelo de regressão múltipla (HILL, GRIFFITHS E JUDGE, 2003).

Na análise de regressão linear múltipla verifica-se o efeito conjunto das covariáveis na resposta Y. (DRAPER E SMITH, 1981; SEBER E LEE, 2003; OU MONTGOMERY E RUNGER, 2011).

Conforme Sanders (1995), as aplicações de modelos de previsão tiveram um grande aumento nas últimas décadas, bem com a variedade e diversidade de modelos. No trabalho de Schwitzky (2001), ele comenta que, para se obter uma boa predição de comportamentos, deve-se utilizar modelos eficientes de previsão.

Um exemplo de modelo de regressão com duas variáveis independentes é dado por:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \quad (2.1)$$

para  $i=1, \dots, n$  e  $\varepsilon_i$ , é um termo relacionado ao erro (variável não-observada) suposto como uma quantidade aleatória. Este erro aleatório inclui todos os outros fatores que poderiam influenciar a variável dependente Y não incluídos no modelo de regressão.

Baseado em um modelo de regressão, um dos pontos principais a ser respondido é: “Qual o valor explicativo deste modelo?” (WEBSTER, 2006). Isso também é estimado usando técnicas de regressão.

## 2.2 Algumas considerações sobre análise Bayesiana

Em geral, as inferências para um modelo de regressão são obtidas a partir de algumas suposições sobre a estrutura do erro  $\varepsilon_i$  (2.1) considerado como uma variável aleatória com média zero, variância constante e distribuição normal  $N(0, \sigma^2)$ . Em muitas aplicações essas suposições podem não ser verificadas, o que pode invalidar as inferências obtidas, e é usual tentar transformar as respostas para satisfazer essas suposições. Um caso especial é dado pela transformação do tipo proposto por Box e Cox (1964).

Uma alternativa para analisar dados seria o uso de métodos Bayesianos, assumindo as respostas na escala original com distribuições de probabilidade diferentes da normal, sem a necessidade de uma transformação Box-Cox.

Métodos Bayesianos vêm sendo empregados em diversas aplicações em Administração, Economia e na Engenharia de Produção; ver por exemplo, Motta (1997), Bueno Neto (1997), Ramirez Pongo e Bueno Neto (1997), Droguett e Mosleh (2006), Cavalcante e Almeida (2006), Quinino e Kalatzis et al. (2006), Moura et al (2007), Ferreira et al. (2009), Barossi-Filho et al (2010) e Freitas et al. (2010).

Em estatística, inferência bayesiana é um método em que a fórmula de Bayes é usada para atualizar a estimativa de probabilidade para uma hipótese. A atualização bayesiana é especialmente importante na análise dinâmica de uma sequência de dados. E a inferência Bayesiana tem sido aplicada em uma variedade de campos, incluindo ciências biológicas, engenharia, filosofia, medicina e direito (ver por exemplo, MOALA et al , 2013).

Ela determina a probabilidade a posteriori como consequência de dois antecedentes, uma probabilidade anterior, ou probabilidade a priori; e uma "função de verossimilhança" derivada de um modelo de probabilidade para os dados observados. A inferência bayesiana estima a probabilidade a posteriori para um parâmetro de acordo com a fórmula de Bayes (BOX E TIAO,1973).

Seja  $\mathbf{X} = (X_1, X_2, X_3, \dots, X_n)$  um vetor aleatório definido no espaço de probabilidade  $(\Omega, A, P_\theta)$  onde  $\Omega$  é o espaço amostral,  $A$  é uma sigma álgebra e  $P_\theta$  é uma probabilidade. Sob a perspectiva bayesiana, a incerteza sobre o parâmetro desconhecido  $\theta$  é descrito em forma de uma distribuição de probabilidade (BOX E TIAO,1973).

Dessa maneira, associa-se uma distribuição de probabilidade para  $\theta$ , usualmente chamada de distribuição *a priori*. Posteriormente, pela análise dos dados, associamos uma distribuição de  $\theta$  condicional à amostra, também chamada de distribuição *a posteriori*.

Em geral, o valor verdadeiro de  $\theta$  é desconhecido, e o objetivo é fazer inferências sobre esse parâmetro. Para representar os diferentes graus de incerteza sobre um parâmetro  $\theta$ , diferentes modelos probabilísticos são elicitados; desta forma, cada pesquisador pode formular um modelo estatístico baseado no seu grau de conhecimento sobre o parâmetro específico. A informação sobre um parâmetro  $\theta$  é representada probabilisticamente por  $\Pi(\theta)$ , também chamada de distribuição a priori, e incorporada ao estudo através do uso do teorema de Bayes, que combina a informação prévia do pesquisador com a informação contida nos dados, resultando na distribuição a posteriori.

Segundo interpretação de Ibrahim, Chen e Sinha (2001), a análise Bayesiana é baseada em especificar um modelo probabilístico para o vetor de dados observados  $T$ , dado um vetor de parâmetros  $\theta$ , levando em consideração uma função de verossimilhança  $L(\theta | T)$ .

Assumindo, então, que  $\theta$  é aleatório, consequentemente tem-se uma distribuição a priori denotada por  $\Pi(\theta)$ . A inferência sobre  $\theta$  é baseada numa distribuição a posteriori, a qual é obtida pelo teorema de Bayes. A distribuição a posteriori de  $\theta$  é dada por,

$$\pi(\theta|T) = \frac{L(\theta|T)\pi(\theta)}{\int_{\theta} L(\theta|T)\pi(\theta)d\theta} \quad (2.2)$$

Na obtenção de sumários a posteriori, como por exemplo, a média a posteriori de  $\theta$ , precisamos resolver integrais múltiplas; muitas vezes, complicadas, o que exige o uso de métodos numéricos ou de aproximações de integrais, especialmente quando a dimensão do vetor de parâmetros é grande.

Daí surge a necessidade do uso de métodos computacionais de simulação introduzidos na literatura, como o método de Monte Carlo em Cadeias de Markov (MCMC), em especial os algoritmos de Metropolis-Hastings, e o amostrador de Gibbs (Gibbs Sampler) (GELFAND E SMITH,1990; ou CHIB E GREENBERG,1995).

Quando as distribuições condicionais a posteriori para cada parâmetro têm formas de distribuições conhecidas e são simples para gerar amostras, é mais usual utilizar-se do amostrador de Gibbs, que é baseado em um processo MCMC, o qual gera amostras das distribuições condicionais completas, que convergem para a distribuição a posteriori de interesse; caso contrário, o algoritmo de Metropolis-Hastings é utilizado, no caso onde as distribuições condicionais a posteriori não possuem formas de distribuições conhecidas e simples para geração de amostras.

### **3. Dados Coletados para o estudo**



Para o presente estudo, foi considerado um conjunto de dados relacionados à produtividade de  $n = 605$  turmas de colhedores de laranja, selecionadas de diferentes regiões do estado de São Paulo. Dentre várias respostas de interesse, consideramos neste estudo, as seguintes respostas relacionadas à produção diária de diferentes turmas de trabalhadores: total de caixas produzidas, produção média e produção acima do piso. Observando-se que as respostas total de caixas produzidas e produção média são dadas por números inteiros, o que caracteriza dados de contagem, enquanto que a resposta produção acima piso é dada por uma variável indicadora. A identificação dessas respostas é importante para a seleção de diferentes modelos estatísticos usados na identificação dos diferentes fatores significativos nas respostas e, também, para serem usados em previsão, um ponto muito importante no setor de frutas.

Dentre várias covariáveis possíveis para cada turma de trabalhadores (pontos amostrais do problema), foram selecionadas as seguintes: quantidade de safras da turma; sexo do líder da turma; idade do líder da turma; estado civil do líder da turma; escolaridade do líder da turma; região onde atua a turma; quantidade de colhedores; porcentagem de trabalhadores do sexo masculino na turma; idade média dos trabalhadores; porcentagem de trabalhadores casados na turma; média de faltas ao trabalho da turma; média diária de colheita e distância média percorrida até o local da colheita; e % trabalhadores experiência.

Observa-se que o ajuste de modelos estatísticos apropriados para os dados pode levar a grandes ganhos pelas empresas do setor de frutas, em termos de identificação dos principais fatores que controlam a variabilidade das respostas e nas previsões.

## **4. Análise estatística dos dados**

### **4.1 Análise de regressão linear múltipla dos dados**

Apresentado o problema, define-se como objetivo central deste trabalho verificar o efeito conjunto dessas covariáveis na resposta Y através de técnicas de regressão múltipla (DRAPER E SMITH, 1981; SEBER E LEE, 2003; OU MONTGOMERY E RUNGER, 2011).

Para analisar os dados de produtividade no setor de frutas cítricas, utilizou-se um modelo de regressão linear múltiplo, de acordo com a equação (4.1). Considerando as covariáveis introduzidas na capítulo 3 e considerando a resposta Y transformada para a escala logarítmica para adequação dos pressupostos de normalidade, supõe-se um modelo de regressão linear múltiplo dado por,

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \beta_7 x_{7i} + \beta_8 x_{8i} + \beta_9 x_{9i} + \beta_{10} x_{10i} + \beta_{11} x_{11i} + \beta_{12} x_{12i} + \beta_{13} x_{13i} + \beta_{14} x_{14i} + \varepsilon_i \quad (4.1)$$

onde  $i=1,2,\dots,605$ ;  $\varepsilon_i$  são erros aleatórios supostos como independentes, com uma distribuição normal com média zero, e variância constante  $\sigma^2$ ;  $x_{1i}$  denota a quantidade de safras da turma;  $x_{2i}$  denota o sexo do líder da turma;  $x_{3i}$  denota a idade do líder da turma;  $x_{4i}$  denota o estado civil do líder da turma;  $x_{5i}$  denota a escolaridade do líder da turma;  $x_{6i}$  denota a região onde atua a turma (Região definida com base no cinturão citrícola);  $x_{7i}$  denota a quantidade de colhedores;  $x_{8i}$  denota a porcentagem de trabalhadores do sexo masculino na turma;  $x_{9i}$  denota a idade média dos trabalhadores;  $x_{10i}$  denota a porcentagem de trabalhadores casados na turma;  $x_{11i}$  denota a média de faltas ao trabalho da turma;  $x_{12i}$  denota a média diária de colheita;  $x_{13i}$  denota a distância média percorrida até o local da colheita; e  $x_{14i}$  denota a porcentagem % trabalhadores com experiência. A variável resposta é dada por uma transformação logarítmica, isto é,  $y_i = \log(\text{total caixas}_i)$  ou  $y_i = \log(\text{produção média}_i)$ . Estimadores de mínimos quadrados dos coeficientes de regressão do modelo (4.1) são obtidos usando o *software* MINITAB versão 16.

## 4.2 Análise Variável resposta: $\log(\text{total caixas})$

Assumindo o modelo de regressão (4.1) temos o seguinte modelo ajustado, obtido usando o *software* MINITAB, versão 16:

$\log(\text{total.caixas}) = 10,3 + 0,00775 \text{ quant.safras} + 0,0491 \text{ sexo} + 0,00042 \text{ idade.lider} -$   
 $0,0725 \text{ est.civ.lid} - 0,0129 \text{ escol.lid} - 0,0163 \text{ região} + 0,0233 \text{ quantcatads} + 0,512 \text{ \%homens}$   
 $- 0,0206 \text{ idade.média} + 0,716 \text{ \%casados} - 0,0703 \text{ média.faltas} - 0,00559 \text{ diária-média-colh}$   
 $+ 0,000072 \text{ distância.média} - 0,0679 \text{ \%trabalhadores experiência}.$

Na Tabela 4.1, temos os sumários das inferências obtidas para esse modelo.

**Tabela 4. 1** Estimadores e valores-p para os coeficientes de regressão considerando a resposta  $\log(\text{total caixas})$

---

Predictor	Coef	SE-Coef	T	P
Constant	10,3277	0,3421	30,19	0,000
quant.safras	0,007750	0,004490	1,73	0,085
sexo	0,04908	0,05920	0,83	0,407
idade.lider	0,000417	0,002600	0,16	0,873
est.civ.lid	-0,07246	0,04119	-1,76	0,079
escol.lid	-0,01287	0,03772	-0,34	0,733
região	-0,01629	0,04487	-0,36	0,717
quant.colhedores	0,023301	0,002540	9,17	0,000
%homens	0,5116	0,1775	2,88	0,004
Idade.média	-0,020622	0,008316	-2,48	0,013
%casados	0,7158	0,2266	3,16	0,002
média.faltas	-0,07026	0,01337	-5,25	0,000
diária-média-colh	-0,0055862	0,0005593	-9,99	0,000
distância.média	0,0000716	0,0002567	0,28	0,780
%trab.experiência	-0,06794	0,09811	-0,69	0,489

S = 0,416401    R-Sq = 39,7%    R-Sq(adj) = 38,3%

---

A partir dos resultados da Tabela 4.1, verificamos que as covariáveis significativas no total de caixas diárias (valores-p menores do que 0,05) são as seguintes: quantidade colhedores; % homens; idade média; % casados; média faltas; e média diária colhida.

Considerando um nível de significância igual à 0,10, também teríamos a significância de outras duas covariáveis: quantidade de safras e estado civil do líder (valores-p menores do que 0,10).

#### **4.3 Variável resposta: log(total caixas)**

Assumindo o modelo de regressão (4.1), temos o seguinte modelo ajustado por mínimos quadrados e obtido usando o software MINITAB:

$$\log(\text{prod média}) = 4,42 + 0,00809 \text{ quant.safras} + 0,0031 \text{ sexo} - 0,00295 \text{ idade.lider} - 0,0278 \text{ est.civ.lid} - 0,0594 \text{ escol.lid} + 0,0710 \text{ região} - 0,00351 \text{ quant.colhedores} + 0,331 \text{ \%homens} - 0,00021 \text{ idade.média} + 0,046 \text{ \%casados} - 0,0324 \text{ média.faltas} - 0,00296 \text{ diária-média-colh} + 0,000379 \text{ distância.média} + 0,0573 \text{ \%trabalhadores experiência}$$

Na Tabela 4.2, temos os sumários das inferências obtidas para esse modelo. A partir dos resultados da Tabela 4.2, verificamos que as covariáveis significativas na produção média (valores-p menores do que 0,05) são as seguintes: quantidade safras; idade líder; escolaridade do líder; região; quantidade colhedores; % homens; média faltas; diária média colhida; e distância média percorrida.

Para verificação da validade dos modelos temos, na Figura 4.1, os gráficos dos resíduos para os dois casos.

**Tabela 4.2** – Estimadores e valores-p para os coeficientes de regressão considerando a resposta  $\log(\text{produção média})$

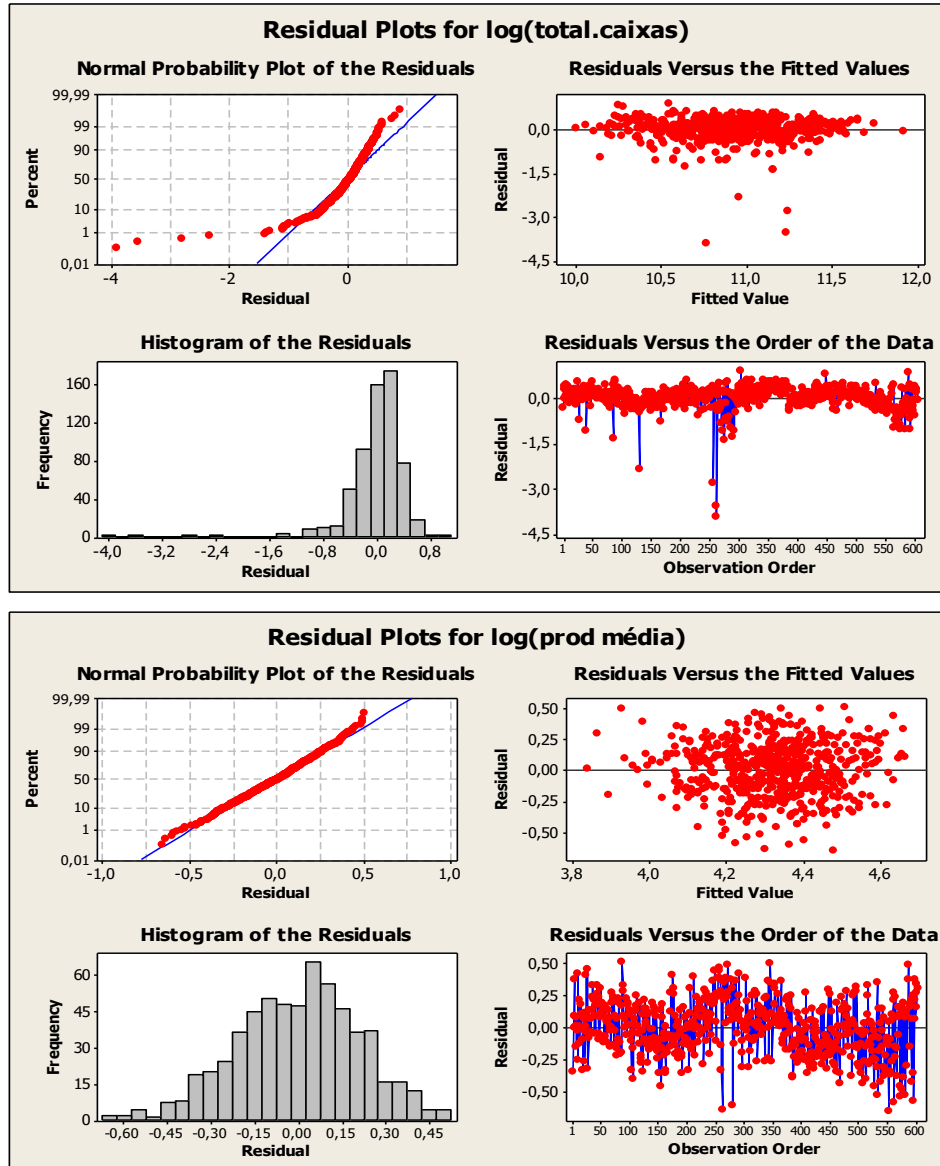
---

Predictor	Coef	SE Coef	T	P
Constant	4,4234	0,1744	25,36	0,000
quant.safras	0,008086	0,002289	3,53	0,000
sexo	0,00312	0,03018	0,10	0,918
idade.lider	-0,002954	0,001325	-2,23	0,026
est.civ.lid	-0,02780	0,02100	-1,32	0,186
escol.lid	-0,05944	0,01923	-3,09	0,002
região	0,07104	0,02288	3,11	0,002
quant.colhedores	-0,003509	0,001295	-2,71	0,007
%homens	0,33051	0,09050	3,65	0,000
Idade.média	-0,000208	0,004240	-0,05	0,961
%casados	0,0458	0,1155	0,40	0,692
média.faltas	-0,032363	0,006817	-4,75	0,000
diária-média-colh	-0,0029555	0,0002852	-10,36	0,000
distância.média	0,0003789	0,0001309	2,90	0,004
%trab experiência	0,05731	0,05002	1,15	0,252

S = 0,212293    R-Sq = 30,3%    R-Sq(adj) = 28,7%

---

**Figura 4.1.** Gráficos dos resíduos



A partir dos gráficos dos resíduos dados na Figura 4.1, verificamos que as suposições necessárias para a validade do modelo estatístico (normalidade dos resíduos, variância constante dos erros) não são verificadas para o modelo assumindo a resposta total de caixas, mas as suposições necessárias para a validade do modelo estatístico (normalidade dos resíduos, variância constante dos erros) são verificadas para o modelo assumindo a resposta produção média diária.

#### 4.4 Análise de regressão logística para a resposta binária produção diária acima do piso (sim = 1 e não = 0) - uso do software MINITAB versão 16

Para análise dos dados relativos à produção diária acima do piso, observa-se que os dados são binários, isto é, (sim = 1 e não = 0). Dessa forma, não podemos usar um modelo de regressão linear múltiplo usual, mas podemos usar uma transformação logística para analisar os dados.

Na presença do vetor de covariáveis  $x_{1i}$  denotando a quantidade de safras da turma;  $x_{2i}$  denotando o sexo do líder da turma;  $x_{3i}$  denotando a idade do líder da turma;  $x_{4i}$  denotando o estado civil do líder da turma;  $x_{5i}$  denotando a escolaridade do líder da turma;  $x_{6i}$  denotando a região onde atua a turma;  $x_{7i}$  denotando a quantidade de colhedores;  $x_{8i}$  denotando a porcentagem de trabalhadores do sexo masculino na turma;  $x_{9i}$  denotando a idade média dos trabalhadores;  $x_{10i}$  denotando a porcentagem de trabalhadores casados na turma;  $x_{11i}$  denotando a média de faltas ao trabalho da turma;  $x_{12i}$  denotando a média diária de colheita;  $x_{13i}$  denotando a distância média percorrida até o local da colheita; e  $x_{14i}$  denotando a porcentagem de trabalhadores % trabalhadores experiência de outras regiões ou estados na turma.

Para relacionar as probabilidades de sucessos  $p_i$  (probabilidade de produção diária acima do piso- respostas binárias) com as covariáveis  $x_{1i}$ ,  $x_{2i}$ ,  $x_{3i}$ ,  $x_{4i}$ ,  $x_{5i}$ ,  $x_{6i}$ ,  $x_{7i}$ ,  $x_{8i}$ ,  $x_{9i}$ ,  $x_{10i}$ ,  $x_{11i}$ ,  $x_{12i}$ ,  $x_{13i}$  e  $x_{14i}$ , consideramos o seguinte modelo de regressão logística,

$$\begin{aligned} \log [p_i/(1 - p_i)] = & \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_3 x_{3i} + \alpha_4 x_{4i} + \alpha_5 x_{5i} + \alpha_6 x_{6i} + \alpha_7 x_{7i} + \alpha_8 x_{8i} \\ & + \alpha_9 x_{9i} + \alpha_{10} x_{10i} + \alpha_{11} x_{11i} + \alpha_{12} x_{12i} + \alpha_{13} x_{13i} + \alpha_{14} x_{14i} \end{aligned} \quad (4.2)$$

Observa-se que, com a transformação logística, temos um modelo linear dado por (4.2). Dos dados, temos 508 observações iguais a 1 (produção diária acima do piso) e 97 observações igual à zero (produção diária abaixo do piso). Estimadores de máxima verossimilhança (EMV) para os coeficientes de regressão do modelo (4.2) são obtidos usando o software MINITAB versão 16 (Resultados na Tabela 4.3).

**Tabela 4.3** – Estimadores (EMV) e valores-p para os coeficientes de regressão logística

---

Predictor	Coef	SE Coef	Z	P	Ratio	Odds 95% CI	
						Lower	Upper
Constant	2,36438	2,37843	0,99	0,320			
quant.safras	0,111084	0,0330684	3,36	0,001	1,12	1,05	1,19
sexo	0,546409	0,398510	1,37	0,170	1,73	0,79	3,77
idade.lider	-0,0475404	0,0194376	-2,45	0,014	0,95	0,92	0,99
est.civ.lid	-0,573142	0,329105	-1,74	0,082	0,56	0,30	1,07
escol.lid	-0,231634	0,302002	-0,77	0,443	0,79	0,44	1,43
região	0,847188	0,317106	2,67	0,008	2,33	1,25	4,34
quant.colhedores	-0,0327219	0,0197565	-1,66	0,098	0,97	0,93	1,01
%homens	2,94304	1,41127	2,09	0,037	18,97	1,19	301,61
Idade.média	-0,0270716	0,0585024	-0,46	0,644	0,97	0,87	1,09
%casados	3,74795	1,75494	2,14	0,033	42,43	1,36	1322,97
média.faltas	-0,103528	0,0979601	-1,06	0,291	0,90	0,74	1,09
diária-média-colh	-0,0260537	0,0038178	-6,82	0,000	0,97	0,97	0,98
distância.média	0,0024311	0,0020388	1,19	0,233	1,00	1,00	1,01
%trab. experiência	1,22570	0,752631	1,63	0,103	3,41	0,78	14,89

A partir dos resultados da Tabela 4.3, concluímos que as covariáveis significativas na produção diária acima do piso (valores-p menores do que 0,05) são: quantidade safras, idade do líder, região, % homens, % casados e diária média colhida. Considerando um nível de significância igual à 0,10, também teríamos a significância de outras duas covariáveis: estado civil do líder e quantidade de colhedores (valores-p menores do que 0,10).

## 5. Análise Bayesiana dos dados

Uma outra alternativa para analisar os dados seria o uso de métodos Bayesianos, assumindo as respostas na escala original sem necessidade de transformação logarítmica para total de



caixas e produção média. Para isso, consideramos modelos de regressão de Poisson sob um enfoque Bayesiano.

Seja  $Y_i$  uma variável aleatória com uma distribuição de Poisson dada por,

$$P(Y_i = y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}, \quad (5.1)$$

onde  $y_i = 0, 1, 2, \dots$  denota o número total de caixas ou produção média da  $i$ -ésima turma de trabalhadores,  $i = 1, 2, \dots, 605$ . Observar que a média e a variância da distribuição de Poisson (5.1) são iguais à  $\lambda_i$ .

Para relacionar o parâmetro  $\lambda_i$  com as covariáveis  $x_{1i}$  denotando a quantidade de safras da turma;  $x_{2i}$  denotando o sexo do líder da turma;  $x_{3i}$  denotando a idade do líder da turma;  $x_{4i}$  denotando o estado civil do líder da turma;  $x_{5i}$  denotando a escolaridade do líder da turma;  $x_{6i}$  denotando a região onde atua a turma;  $x_{7i}$  denotando a quantidade de colhedores;  $x_{8i}$  denotando a porcentagem de trabalhadores do sexo masculino na turma;  $x_{9i}$  denotando a idade média dos trabalhadores;  $x_{10i}$  denotando a porcentagem de trabalhadores casados na turma;  $x_{11i}$  denotando a média de faltas ao trabalho da turma;  $x_{12i}$  denotando a média diária de colheita;  $x_{13i}$  denotando a distância média percorrida até o local da colheita e  $x_{14i}$  denotando a porcentagem de trabalhadores % trabalhadores experiência de outras regiões ou estados na turma, consideramos o seguinte modelo de regressão:

$$\begin{aligned} \log(\lambda_i) = & \beta_0 + \beta_1(x_{1i} - 9,2595) + \beta_2 x_{2i} + \beta_3(x_{3i} - 43,1587) + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \\ & \beta_7(x_{7i} - 47,5223) + \beta_8 x_{8i} + \beta_9(x_{9i} - 35,8777) + \beta_{10} x_{10i} + \beta_{11} x_{11i} + \beta_{12}(x_{12i} - 170,891) + \\ & \beta_{13} x_{13i} + \beta_{14} x_{14i} \end{aligned} \quad (5.2)$$

Observe-se que algumas covariáveis foram centralizadas em suas médias para maior estabilidade do procedimento de simulação usado para gerar amostras da distribuição *a posteriori* de interesse.

A formulação (5.2) garante que  $\lambda_i$  seja positivo, para  $i = 1, 2, \dots, n$ .

Assumindo o modelo definido acima, a função de verossimilhança para o vetor  $\theta$  de parâmetros associados a cada modelo é dada por,

$$L(\theta) = \prod_{i=1}^{605} f(\text{dados}/\theta) \quad (5.3)$$

onde  $\theta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14})$  e  $f(\text{dados}/\theta)$  é a função de probabilidade de Poisson (5.1) para os dados.

Para a análise Bayesiana, assumimos as seguintes distribuições a priori para os parâmetros de regressão:  $\beta_0 \sim N(0,1000)$ ,  $\beta_r \sim N(0,10)$ ,  $r = 1,2,\dots,14$ , onde  $N(a,b^2)$  denota uma distribuição normal com média igual à  $a$  e variância igual à  $b^2$ .

Combinando-se a distribuição *a priori* conjunta para  $\theta$  (um produto de distribuições normais) com a função de verossimilhança  $L(\theta)$ , dada em (5.3), determina-se, a partir da fórmula de Bayes, a distribuição a posteriori para  $\theta$  (BOX E TIAO, 1973).

Os sumários *a posteriori* de interesse foram obtidos usando métodos de Monte Carlo em Cadeias de Markov (GELFAND E SMITH, 1990; OU CHIB E GREENBERG,1995). Uma grande simplificação na geração de amostras da distribuição *a posteriori* para  $\theta$  é obtida usando o software *OpenBugs* (SPIEGELHALTER ET AL, 2003), que só requer a especificação da distribuição para os dados e as distribuições *a priori* para os parâmetros.

### 5.1 Variável resposta: total caixas

Assumindo a regressão de Poisson definidas por (5.1) e (5.2), e usando o software *OpenBugs* com uma amostra simulada de aquecimento (“*burn-in-sample*”) de tamanho 5.000, descartada para eliminar o efeito dos valores iniciais usados no algoritmo Gibbs Sampling, e simulando outras 50.000 amostras escolhidas de 50 em 50 para se ter amostras aproximadamente não-correlacionadas, foi obtida uma amostra final de tamanho 1.000 de valores gerados para  $\beta_0$  e  $\beta_r$   $r = 1,2,\dots,14$ . Os sumários obtidos (média *a posteriori*, desvio-padrão *a posteriori* e intervalos de credibilidade com probabilidade igual à 0,95) são dados na Tabela 5.1. A convergência do algoritmo foi monitorada usando métodos gráficos (PAULINO ET AL, 2003, OU GAMERMAN,1997) e obtidas diretamente do software *OpenBugs*.

Dos resultados da Tabela 5.1, observa-se que todas as covariáveis têm efeitos significativos na produção diária de caixas de frutas, pois o valor zero não está incluído nos intervalos de credibilidade 95% dos parâmetros de regressão.

**Tabela 5.1** – Sumários a posteriori-regressão de Poisson (total caixas)

	Media	DP	Int.Credib.95%	
$\beta_0$	10.84	0.001649	10.83	10.85
$\beta_1$	0.005654	4,27E-02	0.00557	0.005735
$\beta_{10}$	0.6237	0.002101	0.6196	0.6281
$\beta_{11}$	-0.07275	0,134	-0.07301	-0.07246
$\beta_{12}$	-0,0464	0,0000024	-0,000051	-0,0417
$\beta_{13}$	-0.07443	0,967	-0.07634	-0.07256
$\beta_{14}$	-0.005321	0,00587	-0.005333	-0.005309
$\beta_2$	0.02675	0,565	0.0256	0.02782
$\beta_3$	0,593	0,246	0,000546	0,000643
$\beta_4$	-0.03691	0,376	-0.03767	-0.03619
$\beta_5$	-0.03225	0,369	-0.03301	-0.03149
$\beta_6$	0.006684	0,438	0.005829	0.007542
$\beta_7$	0.02472	0,0258	0.02467	0.02476
$\beta_8$	0.5781	0.001768	0.5746	0.5816
$\beta_9$	-0.02054	7.89E-5	-0.0207	-0.02038

## 5.2 Variável resposta: produção média

Também usando o software *OpenBugs* com uma amostra simulada de aquecimento (“*burn-in-sample*”) de tamanho 5.000, descartada para eliminar o efeito dos valores iniciais usados no algoritmo Gibbs Sampling, e simulando outras 50.000 amostras escolhidas de 50 em 50 para se ter amostras aproximadamente não-correlacionadas, foi obtida uma amostra final de tamanho 1.000 de valores gerados para  $\beta_0$  e  $\beta_r$   $r=1,2,\dots,14$ . Os sumários obtidos (média *a posteriori*, desvio-padrão *a posteriori* e intervalos de credibilidade com probabilidade igual à 0,95) são dados na Tabela 5.2.

Dos resultados da Tabela 5.2, observa-se que as covariáveis quantidade safras, idade do líder, estado civil do líder, escolaridade do líder, região, quantidade de colhedores, porcentagem de homens, média faltas, media diária colheita, têm efeitos significativos na produção diária de caixas de frutas, pois o valor zero não está incluído nos intervalos de credibilidade dos parâmetros de regressão. Ou seja, praticamente as mesmas conclusões usando um modelo de

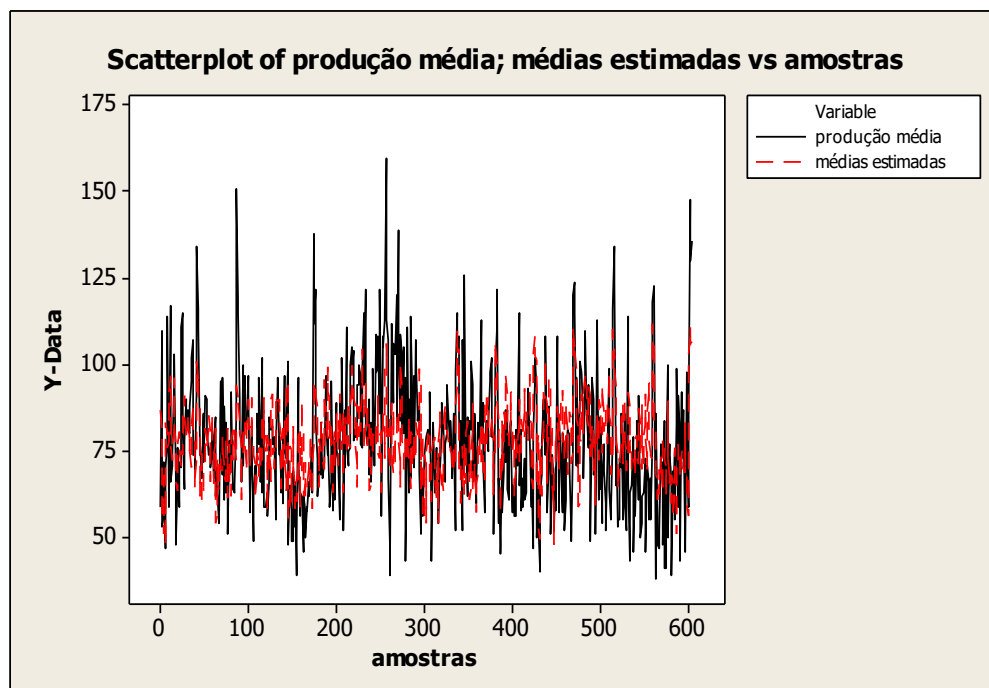
regressão linear tradicional (erros normais) para os dados transformados para a escala logarítmica (ver Tabela 4.2).

Para observar o ajuste dos modelos de regressão de Poisson aos dados, temos na Figura 5.1 os gráficos dos valores observados e médias ajustadas versus observações. Observa-se bom ajuste.

**Tabela 5.2** – Sumários a posteriori-regressão de Poisson (produção média)

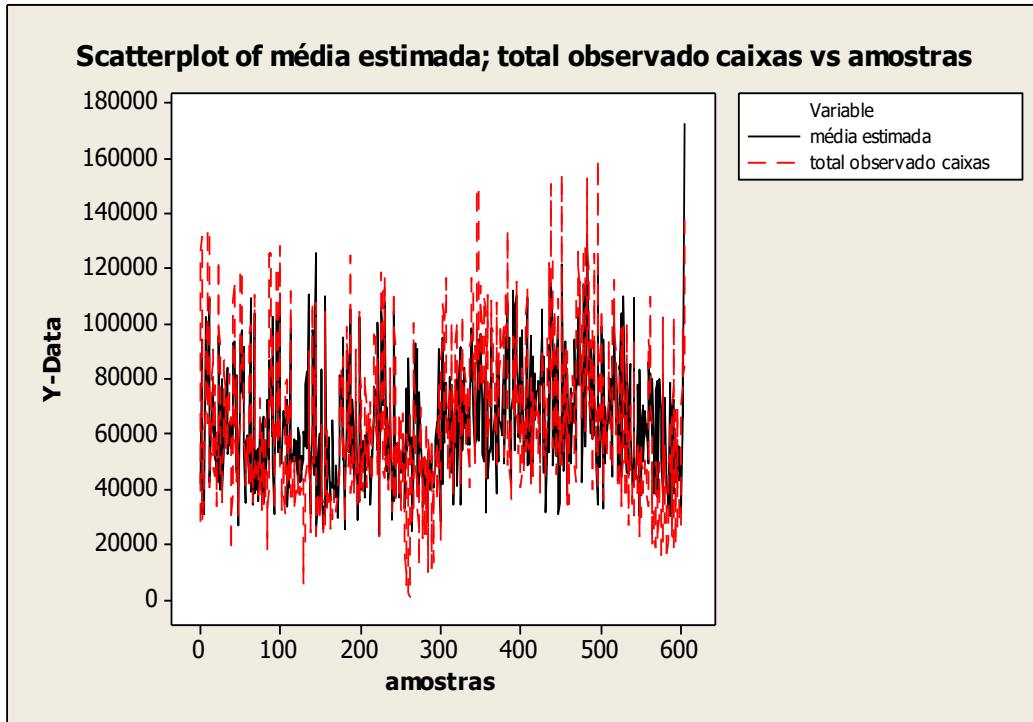
	Media	DP	Int.Credib.95%	
$\beta_0$	4.305	0.04485	4.22	4.397
$\beta_1$	0.007196	0.001194	0.004911	0.009576
$\beta_{10}$	-0.001435	0.06162	-0.1183	0.1186
$\beta_{11}$	-0.03581	0.003637	-0.04296	-0.02861
$\beta_{12}$	0,4	0,0688	0,257	0,000536
$\beta_{13}$	0.04933	0.02647	-0,192	0.1017
$\beta_{14}$	-0.002987	0,164	-0.003297	-0.002655
$\beta_2$	-0.00207	0.01621	-0.03247	0.02997
$\beta_3$	-0.002559	0,729	-0.00403	-0.001038
$\beta_4$	-0.02667	0.01103	-0.04738	-0.003429
$\beta_5$	-0.06551	0.009771	-0.08548	-0.04746
$\beta_6$	0.07091	0.01257	0.04643	0.09558
$\beta_7$	-0.003753	0,706	-0.00512	-0.00237
$\beta_8$	0.3487	0.04718	0.2564	0.4396
$\beta_9$	-0,49	0.002284	-0.004997	0.003996

**Figura 5.1.** Gráficos dos valores observados e médias ajustadas versus observações



**Figura 5.2.** Gráficos dos valores observados de caixas e médias

ajustadas versus observações



### 5.3 Análise Bayesiana da regressão Logística para produção acima do piso

Assumir um modelo de regressão logística com algumas covariáveis centralizadas nas suas médias para a variável binária produção acima do piso dado por,

$$\log[p_i/(1-p_i)] = \alpha_0 + \alpha_1*(\text{quant.safras}[i]-9.25950) + \alpha_2* \text{sexo}[i] + \alpha_3*(\text{idade.lider}[i] - 43.1587) + \alpha_4*\text{est.civ.lider}[i] + \alpha_5*\text{escolar.lider}[i] + \alpha_6* \text{região}[i] + \alpha_7*(\text{quant.cat}[i]- 47.5223) + \alpha_8*\text{perc.homens} [i] + \alpha_9* (\text{idade.média} [i]- 35.8777) + \alpha_{10}*\text{perc.casados}[i] + \alpha_{11}* \text{média.faltas} [i] + \alpha_{12}* (\text{distância.média} [i]-170.891) + \alpha_{13}* \text{perc.\%trabalhadores} \text{ experiência} [i] + \alpha_{14}*\text{diária.média.colh}[i]$$

(5.4)

Considerando o modelo de regressão logística (5.4) para a produção diária acima do piso e as distribuições a priori para os parâmetros de regressão dadas por  $\alpha_0 \sim N(0,1000)$ ,  $\alpha_r \sim N(0,10)$ ,  $r=1,2,\dots,14$ , temos, na Tabela 5.3, os sumários o posteriori obtidos usando o software

*OpenBugs* com um burn-in sample de 5000 amostras e 1000 amostras finais escolhidas de 50 em 50.

Dos resultados da Tabela 5.3, observa-se que as covariáveis quantidade de safras, idade do lider, região, porcentagem de casados e distância média percorrida até o local da colheita têm efeitos significativos na produção diária acima da média, pois o valor zero não está incluído nos intervalos de credibilidade dos parâmetros de regressão correspondentes.

**Tabela 5.3** – Sumários a posteriori- regressão logística(produção diária acima da média)

	Media	DP	Int.Credib.95%	
$\alpha_0$	-0.1608	1.195	-2.536	2.13
$\alpha_1$	0.1132	0.03296	0.04903	0.1763
$\alpha_{10}$	2.984	1.564	0.1587	6.155
$\alpha_{11}$	-0.1272	0.09986	-0.3263	0.06843
$\alpha_{12}$	0.002787	0.002139	-0.0014	0.006907
$\alpha_{13}$	1.27	0.7128	-0.1101	2.77
$\alpha_{14}$	-0.02666	0.003909	-0.03436	-0.01884
$\alpha_2$	0.5191	0.4162	-0.3182	1.319
$\alpha_3$	-0.04895	0.01884	-0.08758	-0.01289
$\alpha_4$	-0.5924	0.3188	-1.212	0.03379
$\alpha_5$	-0.2613	0.3112	-0.8537	0.3634
$\alpha_6$	0.8783	0.3161	0.2839	1.51
$\alpha_7$	-0.03422	0.01974	-0.07306	0.004055
$\alpha_8$	2.598	1.334	-0.09156	5.161
$\alpha_9$	-0.0197	0.05621	-0.1285	0.08807

## 6. Reanálise dos dados considerando as covariáveis contínuas padronizadas

Nesta seção, vamos considerar uma reanálise dos dados considerando as covariáveis contínuas padronizadas, isto é, dadas pela transformação,

$$Z = (X - \text{média})/DP \quad (6.1)$$

Com essa transformação, as interpretações das associações entre cada covariável com a resposta fica simplificada.

### 6.1 Variável resposta: log(total de caixas)

Assumindo o modelo de regressão (4.1), temos o seguinte modelo ajustado obtido usando o software MINITAB, versão 16:

$$\begin{aligned} \log(\text{total.caixas}) = & 11,0 + 0,0372 \text{ quant.safras} + 0,0491 \text{ sexo} + 0,0040 \text{ idade.lider} - \\ & 0,0725 \text{ est.civ.lid} - 0,0129 \text{ escol.lid} - 0,0163 \text{ região} + 0,192 \text{ quant.colhedores} + 0,0668 \\ & \%homens - 0,0814 \text{ Idade.média} + 0,0717 \%casados - 0,107 \text{ média.faltas} - 0,178 \\ & \text{diária.média.colh} + 0,0056 \text{ distância.média} - 0,0221 \%trab.exp \end{aligned} \quad (6.2)$$

Na Tabela 6.1, temos os sumários das inferências obtidas para esse modelo.

**Tabela 6.1** – Estimadores e valores-p para os coeficientes de regressão considerando a resposta  $\log(\text{total caixas})$

Predictor	Coef	SE Coef	T	P
Constant	10,9583	0,0709	154,62	0,000
quant.safras	0,03722	0,02157	1,73	0,085
sexo	0,04908	0,05920	0,83	0,407
idade.lider	0,00396	0,02474	0,16	0,873
est.civ.lid	-0,07246	0,04119	-1,76	0,079
escol.lid	-0,01287	0,03772	-0,34	0,733
região	-0,01629	0,04487	-0,36	0,717
quant.colhedores	0,19164	0,02089	9,17	0,000
%homens	0,06681	0,02318	2,88	0,004
Idade.média	-0,08143	0,03284	-2,48	0,013
%casados	0,07174	0,02271	3,16	0,002
média.faltas	-0,10724	0,02041	-5,25	0,000
diária.média.colh	-0,17777	0,01780	-9,99	0,000
distância.média	0,00565	0,02024	0,28	0,780
%trab.exp	-0,02214	0,03197	-0,69	0,489

S = 0,416401    R-Sq = 39,7%    R-Sq(adj) = 38,3%

A partir dos resultados da Tabela 6.1, verificamos novamente que as covariáveis significativas no total de caixas diárias (valores-p menores do que 0,05) são as seguintes: quantidade colhedores, % homens, idade média, % casados, média faltas e diária média colhida. Considerando um nível de significância igual à 0,10, também teríamos a significância de outras duas covariáveis: quantidade de safras e estado civil do líder (valores-p menores do que 0,10).

Uma vantagem do modelo em termos das covariáveis contínuas padronizadas pode ser dada nas interpretações da magnitude dos efeitos (grau de associação) entre cada covariável e a resposta de interesse.

## 6.2 Variável resposta: $\log(\text{produção média})$

Assumindo o modelo de regressão (4.1), temos o seguinte modelo ajustado por mínimos quadrados e obtido usando o software MINITAB:

$$\log(\text{prod.média}) = 4,34 + 0,0388 \text{ quant.safras} + 0,0031 \text{ sexo} - 0,0281 \text{ idade.lider} - 0,0278 \text{ est.civ.lid} - 0,0594 \text{ escol.lid} + 0,0710 \text{ região} - 0,0289 \text{ quant.colhedores} + 0,0432 \text{ \%homens} -$$



$0,0008 \text{ Idade.média} + 0,0046 \% \text{ casados} - 0,0494 \text{ média.faltas} - 0,0941 \text{ diária.média.colh} + 0,0299 \text{ distância.média} + 0,0187 \% \text{ trab.exp}$  (6.3)

Na Tabela 6.2, temos os sumários das inferências obtidas para esse modelo. A partir desses resultados, também verificamos que as covariáveis significativas na produção média (valores-p menores do que 0,05) são as seguintes: quantidade safras, idade líder, escolaridade do líder, região, quantidade colhedores, % homens, média faltas, diária média colhida e distância média percorrida.

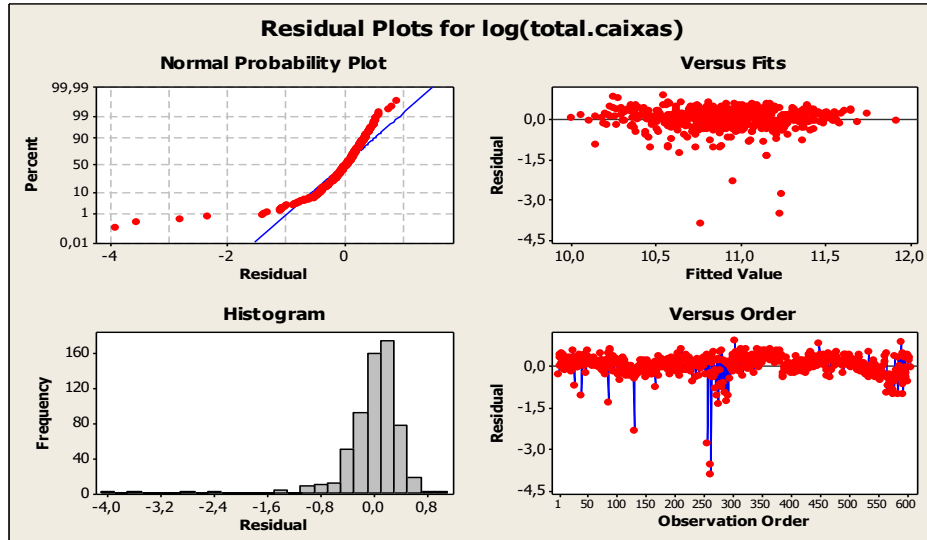
Para verificação da validade dos modelos temos, na Figura 6.1, os gráficos dos resíduos para os dois casos.

**Tabela 6.2** – Estimadores e valores-p para os coeficientes de regressão considerando a resposta  $\log(\text{produção média})$

Predictor	Coef	SE Coef	T	P
Constant	4,33533	0,03613	119,98	0,000
quant.safras	0,03884	0,01100	3,53	0,000
sexo	0,00312	0,03018	0,10	0,918
idade.lider	-0,02811	0,01261	-2,23	0,026
est.civ.lid	-0,02780	0,02100	-1,32	0,186
escol.lid	-0,05944	0,01923	-3,09	0,002
região	0,07104	0,02288	3,11	0,002
quant.colhedores	-0,02886	0,01065	-2,71	0,007
%homens	0,04316	0,01182	3,65	0,000
Idade.média	-0,00082	0,01674	-0,05	0,961
%casados	0,00459	0,01158	0,40	0,692
média.faltas	-0,04940	0,01041	-4,75	0,000
diária.média.colh	-0,094052	0,009075	-10,36	0,000
distância.média	0,02987	0,01032	2,90	0,004
%trab.exp	0,01867	0,01630	1,15	0,252

S = 0,212293    R-Sq = 30,3%    R-Sq(adj) = 28,7%

**Figura 6.1.** Gráficos dos resíduos



A partir dos gráficos dos resíduos dados na Figura 6.1, verificamos que as suposições necessárias para a validade do modelo estatístico (normalidade dos resíduos, variância constante dos erros) não são verificadas para o modelo assumindo a resposta total de caixas, mas as suposições necessárias para a validade do modelo estatístico (normalidade dos resíduos, variância constante dos erros) são verificadas para o modelo assumindo a resposta produção média diária.

### 6.3 Análise de regressão logística para a resposta binária produção diária acima do piso (sim = 1 e não = 0) - uso do software MINITAB versão 16

Vamos assumir o modelo de regressão logística (4.2) com as covariáveis contínuas padronizadas (ver (6.1)).

Estimadores de  $m$

áxima verossimilhança (EMV) para os coeficientes de regressão do modelo (2.2) são obtidos usando o software MINITAB (Resultados na Tabela 6.3).

**Tabela 6.3** – Estimadores (EMV) e valores-p para os coeficientes de regressão logística

Predictor	Coef	SE Coef	Z	P	Odds	95% CI	
					Ratio	Lower	Upper
Constant	1,61725	0,512199	3,16	0,002			
quant.safras	0,533549	0,158831	3,36	0,001	1,70	1,25	2,33
sexo	0,546409	0,398510	1,37	0,170	1,73	0,79	3,77
idade.lider	-0,452477	0,185002	-2,45	0,014	0,64	0,44	0,91
est.civ.lid	-0,573142	0,329105	-1,74	0,082	0,56	0,30	1,07
escol.lid	-0,231634	0,302002	-0,77	0,443	0,79	0,44	1,43
região	0,847188	0,317106	2,67	0,008	2,33	1,25	4,34
quant.colhedores	-0,269125	0,162489	-1,66	0,098	0,76	0,56	1,05
%homens	0,384303	0,184285	2,09	0,037	1,47	1,02	2,11
Idade.média	-0,106893	0,230999	-0,46	0,644	0,90	0,57	1,41
%casados	0,375643	0,175891	2,14	0,033	1,46	1,03	2,06
média.faltas	-0,158023	0,149525	-1,06	0,291	0,85	0,64	1,14
diária.média.colh	-0,829091	0,121491	-6,82	0,000	0,44	0,34	0,55
distância.média	0,191678	0,160745	1,19	0,233	1,21	0,88	1,66
%trab.exp	0,399357	0,245222	1,63	0,103	1,49	0,92	2,41

A partir dos resultados da Tabela 6.3, temos as mesmas conclusões obtidas anteriormente com covariáveis não padronizadas. As covariáveis significativas na produção diária acima do piso (valores-p menores do que 0,05) são: quantidade safras, idade do líder, região, % homens, % casados e diária média colhida. Considerando um nível de significância igual à 0,10, também teríamos a significância de outras duas covariáveis: estado civil do líder e quantidade de colhedores (valores-p menores do que 0,10).

## 6.4 Análise Bayesiana assumindo um modelo de regressão de Poisson

### 6.4.1 Variável resposta: total caixas

Assumindo a regressão de Poisson definidas por (5.1) e (5.2) com todas as covariáveis contínuas padronizadas (ver (6.1)), e usando o software *OpenBugs* com uma amostra simulada de aquecimento (“*burn-in-sample*”) de tamanho 5.000, descartada para eliminar o efeito dos valores iniciais usados no algoritmo Gibbs Sampling, e simulando outras 70.000 amostras escolhidas de 50 em 50 para se ter amostras aproximadamente não-correlacionadas, foi obtida

uma amostra final de tamanho 1.000 de valores gerados para  $\beta_0$  e  $\beta_r$   $r = 1, 2, \dots, 14$ . Os sumários obtidos (média *a posteriori*, desvio-padrão *a posteriori* e intervalos de credibilidade com probabilidade igual à 0,95) são dados na Tabela 6.4.

Dos resultados da Tabela 6.4, observa-se que todas as covariáveis têm efeitos significativos na produção diária de caixas de frutas, pois o valor zero não está incluído nos intervalos de credibilidade dos parâmetros de regressão.

**Tabela 6.4** – Sumários a posteriori; regressão de Poisson (total caixas)

	Média	DP	Int.Credib.95%	
$\beta_0$	11.02	6,42E-01	11.01	$\beta_0$
$\beta_1$	0.02678	2,04E-01	0.02637	$\beta_1$
$\beta_{10}$	0.0627	2,28E-01	0.06223	$\beta_{10}$
$\beta_{11}$	-0.111	1,93E-01	-0.1113	$\beta_{11}$
$\beta_{12}$	-0.003789	1,89E-01	-0.004173	$\beta_{12}$
$\beta_{13}$	-0.02445	3,13E-01	-0.02507	$\beta_{13}$
$\beta_{14}$	-0.1693	1,88E-01	-0.1697	$\beta_{14}$
$\beta_2$	0.02675	5,43E-01	0.02569	$\beta_2$
$\beta_3$	0.006534	2,39E-01	0.006062	$\beta_3$
$\beta_4$	-0.03752	3,87E-01	-0.03825	$\beta_4$
$\beta_5$	-0.03208	3,54E-01	-0.03277	$\beta_5$
$\beta_6$	0.006533	4,54E-01	0.005658	$\beta_6$
$\beta_7$	0.2034	2,04E-01	0.203	$\beta_7$
$\beta_8$	0.07545	2,20E-01	0.075	$\beta_8$
$\beta_9$	-0.08162	3,24E-01	-0.08227	$\beta_9$

#### 6.4.2 Variável resposta: produção média

Também usando o software *OpenBugs* com uma amostra simulada de aquecimento (“*burn-in-sample*”) de tamanho 5.000, descartada para eliminar o efeito dos valores iniciais usados no algoritmo Gibbs Sampling, e simulando outras 50.000 amostras escolhidas de 50 em 50 para se ter amostras aproximadamente não-correlacionadas, foi obtida uma amostra final de tamanho 1.000 de valores gerados para  $\beta_0$  e  $\beta_r$   $r=1, 2, \dots, 14$ . Os sumários obtidos (média *a posteriori*, desvio-padrão *a posteriori* e intervalos de credibilidade com probabilidade igual à 0,95) são dados na Tabela 6.5.

**Tabela 6.5** – Sumários a posteriori-regressão de Poisson (produção média)

	Media	DP	Int.Credib.95%	
$\beta_0$	4.364	0.01871	4.326	4.403
$\beta_1$	0.03467	0.005777	0.02352	0.04552
$\beta_{10}$	0.0554	0.006048	-0.01078	0.01236
$\beta_{11}$	-0.05414	0.005511	-0.0649	-0.0431
$\beta_{12}$	0.03097	0.005293	0.02004	0.04139
$\beta_{13}$	0.01534	0.008868	-0.002016	0.03388
$\beta_{14}$	-0.09496	0.004719	-0.1042	-0.0857
$\beta_2$	-0.001644	0.01542	-0.03197	0.02871
$\beta_3$	-0.02338	0.006658	-0.036	-0.0099
$\beta_4$	-0.02758	0.01134	-0.04987	-0.0042
$\beta_5$	-0.06515	0.01047	-0.08575	-0.0442
$\beta_6$	0.07082	0.01224	0.04551	0.09347
$\beta_7$	-0.03112	0.005212	-0.04107	-0.0210
$\beta_8$	0.0457	0.006402	0.03314	0.05809
$\beta_9$	-0.003433	0.009101	-0.02045	0.01419

Dos resultados da Tabela 6.5, novamente observa-se que as covariáveis quantidade safras, idade do líder, estado civil do líder, escolaridade do líder, região, quantidade de colhedores, porcentagem de homens, média faltas, média diária colheita, têm efeitos significativos na produção média diária de caixas de frutas, pois o valor zero não está incluído nos intervalos de credibilidade dos parâmetros de regressão. Ou seja, praticamente as mesmas conclusões alcançadas usando um modelo de regressão linear tradicional (erros normais) para os dados transformados para a escala logarítmica (ver Tabela 4.2).

## **7. Algumas conclusões e perspectivas futuras**

O presente estudo teve como objetivo identificar, sob o enfoque de engenharia de produção, os principais fatores que contribuem para o bom desempenho das equipes de colheita na citricultura. Para isso, foi considerado como um estudo de caso, a coleta de um conjunto de indicadores referentes a um expressivo número de equipes de colheita de uma empresa do setor de citros do interior do estado de São Paulo. Foi verificada a relação existente entre essas variáveis e os indicadores gerais de desempenho, sendo importante em trabalhos futuros estudar outras variáveis no intuito de melhorar o percentual de explicação dos indicadores gerais de desempenho pelos modelos estatísticos.

Para a análise dos dados, consideramos modelos de regressão linear múltipla para os dados transformados na escala logarítmica, e modelos de regressão de Poisson, sob um enfoque Bayesiano.

Sob o enfoque Bayesiano, conclui-se que a quantidade de safras trabalhadas por equipe, o líder (idade e sexo), a quantidade de colhedores, a porcentagem de trabalhadores do sexo masculino (dentre outras covariáveis) têm impacto significativo sobre o indicador de desempenho (volume colhido).

Com base no modelo estabelecido é possível auxiliar na seleção de equipes de colheita que podem apresentar melhor produtividade e menor custo. Vale salientar que outras covariáveis que não fizeram parte deste estudo podem ser testadas no futuro para construção de um modelo com maior poder de explicação.

Esses resultados podem ser de grande interesse para o setor cítrico.

## Referências

- ALBERT, J.H.; CHIB, S. (1993) Bayesian analysis of binary and polychotomus response data. **Journal American Statistical Association**, Washington, v.88, n.422, p.669-679.
- BAROSSO-FILHO, M.; ACHCAR, J.A.; SOUZA, R.M.(2010). Modelos de volatilidade estocástica em séries financeiras: uma aplicação para o IBOVESPA.**Economia Aplicada**, Ribeirão Preto,14(1): 25-40.
- BENBASAT, I., GOLDSTEIN, D.K., AND MEAD, M., (1987), “**The case research strategy in studies of information systems**”, MIS Quarterly, 11, 3, 369 – 386.
- BERNARDO, J.M.; SMITH, A.F.M. **Bayesian theory**. New York: Wiley, 1994.
- BERTRAND, J. W. M.; FRANSOO. J. C. (2002). Operations management research methodologies using quantitative modeling. **Journal of Operations & Production Management**, v. 22, n. 2, p. 241-261.
- BOX, G. E. P. AND COX, D. R. (1964) An analysis of transformations (with discussion). **Journal of the Royal Statistical Society B**, 26, 211–252.
- BOX, G.; TIAO,G. (1973). **Bayesian inference in statistical analysis**; New York: Addison-Wesley.
- CARVALHO,C.; VENCATO,A.Z.; KIST,B.B.; SANTOS,C.; SILVEIRA,D.; REETZ,E.R.; BELING, R.R.; CORREA,S (2010) **Brazilian fruit yearbook**, Editora Gazeta Santa Cruz: Santa Cruz do Sul,RS,Brazil,128 p.
- CAVALCANTE, C.A. V.; ALMEIDA, A.T (2011) Modelo multicritério de apoio a decisão para o planejamento de manutenção preventiva utilizando PROMETHEE II em situações de incerteza. **Pesquisa Operacional**, Rio de Janeiro,25(2): 279-296.

CEPEA (CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA-ESALQ/USP): harvesting costs 2003/2012

CHIB, S.; GREENBERG, E. (1995). **Understanding the Metropolis-Hastings algorithm.**

The American Statistician, 49, 327-335.

CORTEZ;L.A.B; BRAUNBECK,O.A; CASTRO, L.R.; ABRAHÃO,R.F. CARDOSO., J.L.

(2002)**Revista Frutas e Legumes**, Sistemas de Colheita para Frutas e Hortaliças: oportunidades para sistemas semi-mecanizados, Revista Frutas e Legumes, n.12, p. 26-29, 2002.

DAVIES, F. S.; ALBRIGO, L. G. **Citrus**. Wallingford:CAB International, 254 p, 1994.

DRAPER,N.R.; SMITH, H. **Applied regression analysis, Wiley series in probability and mathematical statistics, 1981.**

DROGETT, E. L.; MOSLEH, A. Análise bayesiana da confiabilidade de produtos em desenvolvimento.**Gestão da Produção**, São Carlos,13(1): 57-69, 2011.

FERREIRA, R.J. P.; ALMEIDA FILHO, A.T. ; SOUZA F. M.C. A decision model for portfolio selection. **Pesquisa Operacional**, Rio de Janeiro,29(2): 403-417, 2009.

FREITAS, M.A. ET AL . Reliability assessment using degradation models: Bayesian and classical approaches. **Pesquisa Operacional**, Rio de Janeiro,30(1): 195-219, 2010.

GAMERMAN, D. MARKOV CHAIN MONTE CARLO: stochastic simulation for Bayesian inference. London: Chapman and Hall, 1997.

GELFAND, A. E.; SMITH, AFM . Sampling-based approaches to calculating marginal distributions, **Journal of the American Statistical Association**, 85, 410, 398- 409, 1990

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2008.

HAIR, JOSEPH F. JR.; ANDERSON, ROLPH E.; TATHAM, RONALD L.; BLACK, WILLIAM C. **Análise Multivariada de Dados**. 5 ed. Porto Alegre: Bookman, 2005.



- HENDERSON,R., SHIMAKURA, S. A serially correlated gamma frailty model for longitudinal count data. *Biometrika* 90, 355–366, 2003.
- HILL, R. C.; GRIFFITHS, W. E.; JUDGE, G. G. **Econometria**. 2. ed. São Paulo: Saraiva, 2003.
- IBGE - Instituto Brasileiro de Geografia e Estatística. **Levantamento sistemático da produção agrícola**. Rio de Janeiro: IBGE, v.25, p.1-88, 2012.
- IBRAHIM, J. G.; CHEN, M.-H.; SINHA, D. **Bayesian survival analysis**. New York: Springer Verlag, 2001.
- KALATZIS,A. E. G.; AZZONI, C. R.; ACHCAR, J. A. Uma abordagem bayesiana para decisões de investimentos. **Pesquisa Operacional**, Rio de Janeiro,26(3): 585-604, 2006.
- LAKATOS, E. M.; MARCONI, M. A. Fundamentos de metodologia científica. 6. ed. São Paulo: Atlas, 2008.
- LEWIS, M. W. Iterative Triangulation; a Theory Development Process using Existing Case Studies. **Journal of Operations Management**, v. 16, p. 455-469, 1998.
- MARCONI, M. DE A.; LAKATOS, E. M. **Fundamentos** de metodologia científica. 7. ed. São Paulo: Atlas, 2010.
- MIGUEL, P. A. C. Estudo de caso na engenharia de produção: estruturação e recomendações para sua condução. **Produção**, v. 17, n. 1, p. 216-229, 2007.
- MONTGOMERY,D.C.; RUNGER, G.C. **Applied statistics and probability for engineers**, fifty edition, New York: Wiley, 2011.
- MOTTA, J. Decisões de preço em clima de incerteza: uma contribuição da análise Bayesiana. **Revista de Administração de Empresas**,São Paulo,37(2):31-46, 1997.
- MOURA, M.C. ET AL. Avaliação Bayesiana da eficácia da manutenção via processo de renovação generalizado, **Pesquisa Operacional**, Rio de Janeiro,27(3): 569-589, 2007.

MOALA, F.A.; RAMOS, P.L.; ACHCAR, J.A. Bayesian Inference for Two-Parameter Gamma Distribution Assuming Different Noninformative Priors, *Revista Colombiana de Estadística*, Colombia, v.36, n.2, p.319-336, 2013.

PAULINO,C.D; TURKMAN,M.; MURTEIRA,B. *Estatística Bayesiana*, Lisboa: Fundação Calouste Gulbenkian, 2003.

PETTO NETO, A. & POMPEU JUNIOR, J. Colheita, beneficiamento e transporte. In: RODRIGUEZ, O.; VIÉGAS, F.C.P.; POMPEU JUNIOR, J. & AMARO, A.A. (Ed).

**Citricultura brasileira**. Campinas: Fundação Cargill,v.2, p.892-897, 1991.

POZZAN, M.; TRIBONI, H. R. Colheita e qualidade do fruto. In: MATTOS JUNIOR, D., NEGRI, J.D.; PIO, R. M.; POMPEU JUNIOR , J. (ED.). **Citros**. Campinas: Instituto Agronômico e Fundag, 2005. p. 801.

QUININO, R.C.; BUENO NETO, P. R. Avaliação bayesiana de inspetores no controle estatístico de atributos. **Gestão da Produção**, São Carlos, 4(3): 296-304, 1997.

RAMIREZ PONGO, R.M.; BUENO NETO, P.R. Uma metodologia bayesiana para estudos de confiabilidade na fase de projeto: aplicação em um produto eletrônico. **Gestão da Produção**, São Carlos ,4(3): 305-320, 1997.

SANDERS, N. R. Managing the forecasting function. **Industrial Mangement & Data Systems**, v.95, n.4, p. 12, 1995.

SEBER,G.A.F.; LEE,A.J. (2003). *Linear regression analysis*, second edition, Wiley series in probability and mathematical statistics.

SPIEGELHALTER, D. J.; THOMAS, A.; BEST, N. G.; LUND, D. *Winbugs user manual*. Cambridge: MRC Biostatistics Unit., 2001.

SPIEGELHALTER, D. J.; THOMAS, A.; BEST, N.G.; LUNN, D. **WinBugs: user manual, version 1.4**, MRC Biostatistics Unit, Cambridge, U.K, 2003.

SCHWITZKY, M. Acuracidade dos métodos de previsão e a sua relação com o dimensionamento dos estoques de produtos acabados. 2001. Dissertação (Mestrado em Engenharia de Produção) -UFSC, Florianópolis, 2001.

TACHIBANA, A.; RIGOLIN, A. DE T.; Análise da produtividade das operações de colheita e carregamento mecanizado de laranja. **Laranja**, Cordeirópolis, v. 23, n.1, p. 57-75, 2002.

## Apêndices

### Apêndice 1 – Descritivo das variáveis a serem estudadas

Código	Variável	Descrição
V1	Safra	Ano em que ocorreu a colheita
V2	Mês	Mês da Colheita
V3	Equipe	Código da Equipe
V4	Líder da Equipe	Nome do Líder da Equipe
V5	Primeira safra do Líder	Ano em que o Líder começou a trabalhar na função
V6	Qtd. Safras	Número de safras que o Líder já trabalhou na função
V7	Sexo do Líder	Masculino ou Feminino
V8	Idade do Líder	Idade do Líder em anos
V9	Estado Civil do Líder	Solteiro, Casado, Divorciado, etc.
V10	Grau de Instrução do Líder	Ecolaridade do Líder
V11	Cidade da Equipe	Nome da cidade que a Equipe é oriunda (cidade do Líder)
V12	Região	Região de atuação da Equipe
V13	Admissão	Data que o Líder foi contratado
V14	Demissão	Data que o Líder foi desligado
V15	Qtd. Colhedores	Número de colhedores presentes na Equipe
V16	Qtd. Homem	Número de homens presentes na Equipe
V17	Qtd. Mulher+C20	Número de mulheres presentes na Equipe
V18	% de Homem	Percentual de colhedores homens
V19	% de Mulher	Percentual de colhedores mulheres
V20	Idade Média	Idade dos colhedores em anos
V21	Idade Média Homem	Idade dos colhedores homens em anos
V22	Idade Média Mulher	Idade das colhedoras mulheres em anos
V23	Qtd. de Casados	Número de colhedores casados
V24	Qtd. de Solteiros	Número de colhedores solteiros
V25	Qtd. de Outros	Número de colhedores não casados e não solteiros
V26	% de Casados	Percentual de colhedores casados
V27	% de Solteiros	Percentual de colhedores solteiros
V28	% de Outros	Percentual de colhedores não casados e não solteiros
V29	% de Casados Homem	Percentual de colhedores homens casados
V30	% de Solteiros Homem	Percentual de colhedores homens solteiros
V31	% de Outros Homem	Percentual de colhedores homens não casados e não solteiros
V32	% de Casados Mulher	Percentual de colhedoras mulheres casadas
V33	% de Solteiros Mulher	Percentual de colhedoras mulheres solteiras

## Continuação

<b>Código</b>	<b>Variável</b>	<b>Descrição</b>
V34	% de Outros Mulher	Percentual de colhedoras mulheres não casados e não solteiros
V35	Média de faltas por colhedor	Número de faltas dos colhedores
V36	Média de faltas - Homem	Número de faltas de colhedores homens
V37	Média de faltas - Mulher	Número de faltas de colhedoras mulher
V38	% Faltas	Percentual de falta dos colherores
V39	% Faltas Homem	Percentual de falta dos colherores homens
V40	% Faltas Mulher	Percentual de falta das colheroras mulheres
V41	Turn Over (%)	Percentual de admissões e demissões de colhedores
V42	Total Caixas Colhidas	Soma das caixas colhidas pelas Equipes
V43	Total Caixas Colhidas Homem	Soma das caixas colhidas por homens das Equipes
V44	Total Caixas Colhidas Mulher	Soma das caixas colhidas por mulheres das Equipes
V45	% Caixas Colhidas Homem	Percentual de caixas colhidas por homens
V46	% Caixas Colhidas Mulher	Percentual de caixas colhidas por mulheres
V47	Total Caixas Colhidas - Tarifa 1	Soma das caixas colhidas na complexidade de colheita 1
V48	Total Caixas Colhidas - Tarifa 2	Soma das caixas colhidas na complexidade de colheita 2
V49	Total Caixas Colhidas - Tarifa 3	Soma das caixas colhidas na complexidade de colheita 3
V50	Total Caixas Colhidas - Tarifa 4	Soma das caixas colhidas na complexidade de colheita 4
V51	Total Caixas Colhidas - Tarifa 5	Soma das caixas colhidas na complexidade de colheita 5
V52	Total Caixas Colhidas - Tarifa 6	Soma das caixas colhidas na complexidade de colheita 6
V53	Total Caixas Colhidas - Tarifa 7	Soma das caixas colhidas na complexidade de colheita 7
V54	Produtividade Média	Média de caixas colhidas por um colhedor
V55	1º Quartil de produtividade	1º quartil de produtividade das Equipes
V56	Mediana de produtividade	Mediana de produtividade das Equipes
V57	3º Quartil de produtividade	3º quartil de produtividade das Equipes
V58	Produtividade Média - Homem	Média de caixas colhidas por colhedores homens
V59	Produtividade Média - Mulher	Média de caixas colhidas por colhedoras mulheres
V60	Produtividade Média - Tarifa 1	Média de caixas colhidas por um colhedor na complexidade 1
V61	Produtividade Média - Tarifa 2	Média de caixas colhidas por um colhedor na complexidade 2
V62	Produtividade Média - Tarifa 3	Média de caixas colhidas por um colhedor na complexidade 3
V63	Produtividade Média - Tarifa 4	Média de caixas colhidas por um colhedor na complexidade 4
V64	Produtividade Média - Tarifa 5	Média de caixas colhidas por um colhedor na complexidade 5
V65	Produtividade Média - Tarifa 6	Média de caixas colhidas por um colhedor na complexidade 6
V66	Produtividade Média - Tarifa 7	Média de caixas colhidas por um colhedor na complexidade 7
V67	Produtividade Acima do piso (S/N)	Produtividade média do colhedor acima do piso (Sim ou Não)

## Continuação

<b>Código</b>	<b>Variável</b>	<b>Descrição</b>
V68	Acelerador Médio por Colhedor (R\$)	Valor médio pago na linha de acelerador para os colhedores (R\$)
V69	% de colhedores que receberam Acelerador	Percentual de colhedores que receberam acelerador na safra
V70	Complemento Médio por Colhedor (R\$)	Valor médio pago na linha de complemento de piso para os colhedores (R\$)
V71	% de colhedores que receberam Complemento	Percentual de colhedores que receberam complemento de piso na safra
V72	Diária Média por Colhedor (R\$)	Valor médio pago na linha de diárias para os colhedores (R\$)
V73	Chuva Média por Colhedor (R\$)	Valor médio pago na linha de chuva para os colhedores (R\$)
V74	Frete Médio (R\$)	Valor médio pago na linha de Frete para as Equipes de colheita (R\$)
V75	Distância Média percorrida - Ida e Volta (Km)	Distância Média percorrida por uma Equipe de colheita (Km)
V76	Grau de instrução - Analfabeto (%)	Percentual de colhedores analfabetos
V77	Grau de instrução - % Primário Incompleto (%)	Percentual de colhedores com primario incompleto
V78	Grau de instrução - Primário Completo (%)	Percentual de colhedores com primario completo
V79	Grau de instrução - Ginásial Incompleto (%)	Percentual de colhedores com ginásial incompleto
V80	Grau de instrução - Ginásial Completo (%)	Percentual de colhedores com ginásial completo
V81	Grau de instrução - Colegial Incompleto (%)	Percentual de colhedores com colegial incompleto
V82	Grau de instrução - Colegial Completo (%)	Percentual de colhedores com colegial completo
V83	Grau de instrução - Superior Incompleto (%)	Percentual de colhedores com superior incompleto
V84	Grau de instrução - Superior Completo (%)	Percentual de colhedores com superior completo
V85	Produtividade Média - Analfabeto (cxs)	Número de colhedores com produtividade média - analfabeto
V86	Produtividade Média - Primário Incompleto (cxs)	Número de colhedores com produtividade média - primário incompleto
V87	Produtividade Média - Primário Completo (cxs)	Número de colhedores com produtividade média - primário completo
V88	Produtividade Média - Ginásial Incompleto (cxs)	Número de colhedores com produtividade média - ginásial incompleto
V89	Produtividade Média - Ginásial Completo (cxs)	Número de colhedores com produtividade média - ginásial completo
V90	Produtividade Média - Colegial Incompleto (cxs)	Número de colhedores com produtividade média - colegial incompleto
V91	Produtividade Média - Colegial Completo (cxs)	Número de colhedores com produtividade média - colegial completo
V92	Produtividade Média - Superior Incompleto (cxs)	Número de colhedores com produtividade média - superior incompleto
V93	Produtividade Média - Superior Completo (cxs)	Número de colhedores com produtividade média - superior completo
VV94	% Trabalhadores com experiência	% de Trabalhadores com experiência
VV95	% Trabalhadores sem experiência	% Trabalhadores sem experiência

## Apêndice 2: Uma breve revisão da metodologia bayesiana

Os métodos bayesianos têm sido considerados alternativas muito eficazes e poderosas na análise de dados. Esse método possui uma filosofia muito diferente do método frequentista. No método clássico, os parâmetros do modelo são considerados constantes desconhecidas, e no método bayesiano todos os parâmetros são considerados quantidades aleatórias. Uma característica muito importante é o fato desta análise permitir a incorporação da informação de um especialista junto à informação dos dados.

A inferência bayesiana tem como fundamentação a Fórmula de Bayes, a qual combina os dados com a informação *a priori* e, então, se obtém *a posteriori* (*priori* já complementada pela informação dos dados), onde é realizado todo o processo inferencial (BOX; TIAO, 1973; PAULINO; TURKMAN; MURTEIRA, 2003).

### Fórmulas de Bayes

Considere uma partição do espaço amostral  $\Omega$ , onde os eventos  $A_1, A_2, \dots, A_k$  formam uma sequência de eventos mutuamente exclusivos e exaustivos, isto é,  $\cup_{j=1}^k A_j = \Omega$  e  $A_i \cap A_j = \emptyset$  (conjunto vazio) para  $i \neq j$  tal que  $P(\cup_{j=1}^k A_j) = \sum_{j=1}^k P(A_j) = 1$ .

Sendo assim para qualquer outro evento  $B$  ( $B \subset \Omega$ ), temos,

$$P(A_i | B) = \frac{P(A_i | B) P(A_i)}{\sum_{j=1}^k P(A_j | B) P(A_j)} \quad (1)$$

Para todo  $i$  variando de 1 a  $k$ .

Assumindo agora um vetor de dados  $y = (y_1, \dots, y_n)'$  e  $\theta$  (quantidades desconhecidas) os parâmetros de uma distribuição de probabilidade associada com a variável aleatória  $Y_i$  com valores observados em  $y_i, i = 1, \dots, n$ .

Considere  $y = (y_1, \dots, y_n)$  uma amostra aleatória onde os dados são independentes e identicamente distribuídos. Suponha que  $y$  é um vetor de observações de uma distribuição conjunta dada por  $f(y | \theta)$  e seja  $\pi(\theta)$  uma distribuição a priori para  $\theta$ . Então, assumindo os valores discretos  $\theta_1, \dots, \theta_k$ , temos de (1) a distribuição a posteriori para  $\theta_i$  dado  $y$  é dado por

$$\pi(\theta_i | y) = \frac{f(y | \theta_i) \pi(\theta_i)}{\sum_{j=1}^k f(y | \theta_j) \pi(\theta_j)} \quad (2)$$

Onde o parâmetro  $\theta$  também é considerado como uma quantidade aleatória, no enfoque bayesiano.

Para  $\theta$  assumindo valores contínuos num dado intervalo, podemos escrever (2) por

$$\pi(\theta | y) = \frac{f(y | \theta) \pi(\theta)}{\int f(y | \theta) \pi(\theta) d\theta} \quad (3)$$

Em que a integral no denominador de (3) é definida no intervalo de variação de  $\theta$ .

### **Distribuição *a Priori***

Na análise bayesiana, a distribuição *a priori* é utilizada a fim de representar o que já é conhecido sobre parâmetros desconhecidos, antes de se avaliar os dados. Deve-se ter muita cautela ao definir uma distribuição *a priori*, afinal se esta informação não for bem definida pode-se chegar a interpretações errôneas. Uma distribuição *a priori* para um parâmetro pode se dar de várias formas, sendo possível ocorrerem a partir de procedimentos subjetivos ou objetivos.

A distribuição *a priori* conjugada é uma *priori* informativa, onde a distribuição *a priori* e a *posteriori* pertencem à mesma classe de distribuições. A passagem de *priori* para a *posteriori* envolve apenas uma simples mudança nos parâmetros, sem a necessidade de cálculos adicionais (PAULINO; TURKMAN; MURTEIRA, 2003).

A distribuição *a priori* também pode ser a incorporação do conhecimento de um pesquisador, ou seja, o pesquisador se baseia na sua prática e no seu “*feeling*” para definir a *priori*. Em muitas situações práticas esse conhecimento do especialista não existe ou, se existe, não é fidedigno. Nesse caso, caracteriza-se uma “*ignorância a priori*” (PAULINO; TURKMAN; MURTEIRA, 2003).

Nestas situações são utilizadas *as prioris* não informativas, ou seja, *prioris* de referência “neutras” (BOX; TIAO, 1973). A utilização deste tipo de distribuição *a priori* permite a comparação com os resultados obtidos pela inferência clássica, haja visto que através de uma *priori* não informativa, o modelo é baseado apenas na informação dos dados amostrais. Existem vários métodos para se definir a *priori* não informativa, como por exemplo: Método de Bayes-Laplace, Método de Jeffreys, entre outros (BOX; TIAO, 1973; PAULINO; TURKMAN; MURTEIRA, 2003).



## Métodos de Simulação para amostras da distribuição *a posteriori*

Para a obtenção de sumários *a posteriori* de interesse, geralmente é necessário resolver integrais bayesianas que não apresentam solução analítica. Na prática, observa-se que os modelos utilizados nem sempre são simples para se obter os resumos *a posteriori*. Mesmo que se tenha uma *priori* e uma verossimilhança simples, a junção delas pode produzir uma distribuição *a posteriori* muito complicada (PAULINO; TURKMAN; MURTEIRA, 2003).

Os métodos com base em amostragem, como, por exemplo, o método de Monte Carlo com cadeias de Markov (MCMC), passaram a ser utilizados com o avanço das técnicas computacionais. Esse método consiste na simulação de uma variável aleatória através de uma cadeia de Markov, no qual a sua distribuição assintoticamente se aproxima da distribuição *a posteriori* (BERNARDO; SMITH, 1994).

A cadeia de Markov é um processo estocástico no qual o próximo estado da cadeia depende somente do estado atual e dos dados. No entanto, existe uma relação com o estado inicial, que é descartado após um período de aquecimento, o chamado “*Burn-in*”.

As formas mais usuais dos métodos MCMC são os amostradores de Gibbs e o algoritmo de Metropolis-Hastings. Estas duas formas simulam amostras da distribuição *a posteriori* conjunta a partir das distribuições condicionais (GELFAND; SMITH, 1990; CHIB; GREENBERG, 1995).

O amostrador de Gibbs nos permite gerar amostras da distribuição *a posteriori* conjunta desde que as distribuições condicionais completas possuam formas fechadas ou conhecidas. Por outro lado, o algoritmo de Metropolis-Hastings permite gerar amostras da distribuição *a posteriori* conjunta com distribuições condicionais completas possuindo ou não uma forma conhecida ou fechada.

### Amostrador de Gibbs

Suponha  $\pi(\theta | y)$  uma distribuição *a posteriori* conjunta, sendo  $\theta = (\theta_1, \dots, \theta_k)$ , no qual desejamos obter inferências. Para isso, simulam-se quantidades aleatórias de distribuições condicionais completas  $\pi(\theta_i | y, \theta_{(i)})$ .

Considere os valores iniciais (arbitrários) para  $\theta: \theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_k$ . Desta forma, segue o seguinte algoritmo:

- Gerar  $\theta_1^{(1)}$  de  $\pi(\theta_1 | y, \theta_2^{(0)}, \dots, \theta_k^{(0)})$ ;

- Gerar  $\theta_2^{(1)}$  de  $\pi(\theta_2 | y, \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)})$ ;

(...)

- Gerar  $\theta_k^{(1)}$  de  $\pi(\theta_2 | y, \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_k^{(1)})$ .

Substitua os valores iniciais por  $\theta^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_k^{(1)})$ , para uma nova realização.

Os valores  $\theta_1^{(z)}, \theta_2^{(z)}, \dots, \theta_k^{(z)}$ , para  $z$  suficientemente grande, convergem para um valor da quantidade aleatória com distribuição  $\pi(\theta | y)$  (BERNARDO; SMITH, 1994, p. 353; CASELA e GEORGE, 1992).

### Algoritmo de Metropolis Hastings

Suponha uma amostra de densidade não regular  $\pi(\theta_i | \theta_{(i)})$ , em que  $\theta_{(i)} = \theta_{i-1}, \theta_{i+1}, \dots, \theta_k$ . Seja  $q(\theta, \beta)$  o núcleo de transição da distribuição  $p(\theta)$  que representa  $\pi(\theta_i | \theta_{(i)})$  e que transforma  $\theta$  em  $\beta$ .

Desta forma o algoritmo dado por:

- Inicie com  $\theta^{(0)}$  e indicador de estado  $j = 0$ ;

- Gerar um ponto  $\beta$  do núcleo de transição  $q(\theta^{(j)}, \beta)$ ;

- Atualizar  $\theta^{(j)}$  por  $\theta^{(j+1)} = \beta$ , com probabilidade,  $p = \min\left\{\frac{p(\beta)q[\theta^{(j)}, \beta]}{p[\theta^{(j)}]q[\beta, \theta^{(j)}]}\right\}$ , ficar

com  $\theta^{(j)}$  com probabilidade  $1 - p$ :

- Repetir os dois últimos passos até conseguir uma distribuição estacionária.

Importante observar: se um valor candidato é rejeitado, então o valor atual é considerado na próxima etapa; no terceiro passo o valor de  $p$  não depende da constante normalizadora; o algoritmo de Metropolis Hasting é especificado pela densidade candidata para geração  $q(x, y)$  (BERNARDO; SMITH, 1994; CHIB; GREENBERG, 1995).

Na prática podemos usar alguns programas computacionais na simulação de amostras da distribuição *a posteriori* de interesse. Um software muito popular e livre é o software *OpenBugs* (versão nova do software *WinBugs*). Com o uso desse software o trabalho de simulação fica muito simplificado.

### Apêndice 3: Artigos produzidos

(1) SIMPEP-2013-Bauru,SP

#### Evento em que o artigo foi submetido

**Evento:** 2013 - XX SIMPEP

**Tema:** Engenharia de produção & objetivos de desenvolvimento do milênio

#### Dados da Apresentação

**Apresentação:** Sessão 4 - 06/11/2013 (Quarta-feira), das 08:00 às 10:00 - Sala Mercúrio

**Apresentadores:**

- Gustavo José Caçador
- Marcelo Edmundo Alves Martins

#### Dados do Artigo

**Inscrição:** 936

**Submetido em:** 04/08/2013 - 14:41:39

**Status:** **Aprovado**

**Título:** Desempenho de equipes de colheita na área de fruticultura: um estudo de caso

**Title:** The team performance of manual fruit harvesting: a study of case

**Resumo:** O presente estudo tem como objetivo identificar sob o enfoque de engenharia de produção, os principais fatores que contribuem para o bom desempenho das equipes de colheita na citricultura. para isso foi considerado como um estudo de caso, a coleta de

um conjunto de indicadores referentes a um expressivo número de equipes de colheita de uma empresa do setor de citros do interior do estado de São Paulo. pretende-se verificar a relação existente entre estas variáveis e os indicadores gerais de desempenho. para a análise dos dados, consideramos modelos de regressão linear múltipla para os dados transformados e modelos de regressão de Poisson, sob um enfoque bayesiano. sob o enfoque bayesiano, conclui-se que a quantidade de safras trabalhadas por equipe, o líder (idade e sexo), a quantidade de colhedores, a porcentagem de trabalhadores do sexo masculino (dentre outras covariáveis) têm impacto significativo sobre o indicador de desempenho (volume colhido).

**Abstract:**

This study aims to identify under a production engineering approach, the main factors that affect the performance of teams in fruit harvest. in this way, it was considered in a case study, some indicators which could be related to the performance of different teams working in the fruit harvest of a industry of the sector in the state of São Paulo. in this study, we want to examine the relationship between these variables and some general performance indicators. for the data analysis, we considered multiple linear regression models where the response of interest was transformed to other scale to satisfy standard statistical properties and Poisson regression models. under a bayesian approach considered for the Poisson regression models, it was concluded that the fruit harvest volume was affected by some factors as the team leader (age and sex), the amount of lanyards and the percentage of male workers.

**Palavras-**

indicadores de desempenho; modelos de regressão linear;

**Chaves:**

modelo de regressão de Poisson, análise bayesiana, métodos de Monte Carlo em cadeias de Markov.

**Keyword:** performance indicators; linear regression models, Poisson regression model, bayesian analysis, Markov chains Monte Carlo methods.

**Área:** 3 - GESTÃO ECONÔMICA

**Sub-área:** 3.5 - Gestão de Desempenho de Sistemas de Produção e Operações

#### Dados dos Autores

<b>Autor 1:</b>	Marcelo Edmundo Alves Martins	marcelo@makegestao.com.br
<b>Autor 2:</b>	Kelly Roberta Pacheco Martins	kelly@makegestao.com.br
<b>Autor 3:</b>	Gustavo José Caçador	gustavo.cacador@solvingefeso.com
<b>Autor 4:</b>	Jorge Alberto Achcar	achcar@fmrp.usp.br
<b>Autor 5:</b>	Claudio Luís Piratelli	clpiratelli@uniara.com.br

---

(2) Artigo aceito para publicação: o artigo intitulado **UTILIZAÇÃO DE MÉTODOS BAYESIANOS NA AVALIAÇÃO DO DESEMPENHO DE EQUIPES DE COLHEITA: UMA APLICAÇÃO NO SETOR DE CITRICULTURA DO ESTADO DE SÃO PAULO**, de autoria de Marcelo Edmundo Alves Martins, Claudio Luis Piratelli e Jorge Alberto Achcar, foi aceito para publicação na **Revista Eletrônica Fafit/Facic (ISSN: 2176-9443)**.

**Fatores que afetam o desempenho de equipes de colheita no setor de citricultura do estado de São Paulo: um estudo de caso**

**“Factors affecting the performance of teams in the citrus harvest sector of the state of São Paulo: a case study”**

Marcelo Edmundo Alves Martins

Centro Universitário de Araraquara – Uniara – Araraquara - Brasil

[marcelo@makegestao.com.br](mailto:marcelo@makegestao.com.br)

Jorge Alberto Achcar

Universidade de São Paulo – FMRP-USP – Ribeirão Preto – Brasil

[achcar@fmrp.usp.br](mailto:achcar@fmrp.usp.br)

Cláudio Luis Piratelli

Centro Universitário de Araraquara – Uniara – Araraquara - Brasil

[clpiratelli@uniara.com.br](mailto:clpiratelli@uniara.com.br)

**Resumo**

Nesse estudo avaliamos via modelos de regressão os principais fatores que contribuem para o bom desempenho das equipes de colheita na citricultura. Como um estudo de caso, foi considerado um conjunto de dados relacionados à coleta de várias equipes de colheita de uma empresa do setor de citros do interior do estado de São Paulo. Usando modelos de regressão múltipla, pretende-se verificar a relação existente o desempenho da equipe com os indicadores gerais de desempenho. Para a análise dos dados, também consideramos modelos de regressão logística para a resposta binária relacionada à turmas com produção acima do pico e turmas com produção abaixo do pico estipulado pela indústria.

**Palavras-chave:** Indicadores de desempenho; modelos de regressão linear; modelo de regressão logística, estimadores de mínimos quadrados, estimadores de máxima verossimilhança.

---

(3) Artigo submetido para publicação na Revista Brasileira de Biometria

**STATISTICAL MODELING FOR PERFORMANCE OF TEAMS IN THE  
CITRUS HARVEST: CLASSICAL VS BAYESIAN APPROACH**

Marcelo Edmundo Alves MARTINS<sup>1</sup>

Jorge Alberto ACHCAR<sup>1,2</sup>

Claudio Luis PIRATELLI<sup>1</sup>

- **ABSTRACT:** This study aims to identify the main factors that contribute to the performance of different teams of workers in the citrus harvest from a production engineering viewpoint. Statistical modeling was adopted as a quantitative approach in order to analyze a dataset from a citrus company in the state of São Paulo, Brazil. Specially, we intend to study the relationship between these variables and the general performance indicator “number of boxes”. The manager indicated several variables related to a large number of teams of workers in the citrus harvest. For the data analysis, we consider a multiple linear regression model assuming transformed responses and Poisson regression models, under a Bayesian approach. The Bayesian approach had the best adherence to the data and shows us that the fruit harvest volume was affected by factors such as the team leader, the number of pickers, the percentage of male workers, among other variables.
- **KEYWORDS:** citrus; performance of teams, harvest, linear regression models, Poisson regression model, Bayesian analysis, Markov Chain Monte Carlo (MCMC) methods.