

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

Modelos não lineares para dados de contagem longitudinais

Ana Maria Souza de Araujo

Tese apresentada para obtenção do título de Doutor em
Agronomia. Área de concentração: Estatística e Experi-
mentação Agronômica

Piracicaba

2007

Ana Maria Souza de Araujo

Estatística

Modelos não lineares para dados de contagem longitudinais

Orientadora:

Prof^ª Dr^ª **CLARICE G. B. DEMÉTRIO**

Co-orientador:

Prof. Dr. **DALTON F. DE ANDRADE**

Tese apresentada para obtenção do título de Doutor em
Agronomia. Área de concentração: Estatística e Experi-
mentação Agronômica

Piracicaba

2007

**Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP**

Araujo, Ana Maria Souza de

Modelos não lineares para dados de contagem longitudinais / Ana Maria Souza de Araujo. -
- Piracicaba, 2007.
197 p. : il.

Tese (doutorado) - - Escola Superior de Agricultura Luiz de Queiroz, 2007.
Bibliografia.

1. Análise de dados longitudinais 2. Dados de contagem 3. Distribuição de Poisson
4. Método da máxima verossimilhança 5. Modelos não lineares I. Título

CDD 519.5

“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”

Dedicatória

A **DEUS, Pai, Filho e Espírito Santo**,
fonte de todas as graças e bênçãos derramadas durante o curso de
doutorado e ao longo da minha vida.

À **Maria Santíssima**, minha Mãe do Céu,
constante intercessora junto a Jesus.

Aos meus pais,
Mário Pereira de Araujo (*in memoriam*) e
Maria Estela Sousa de Araujo,
de quem recebi e recebo amor incondicional, doação,
dedicação e apoio constante em todas as fases da minha
vida.

AGRADECIMENTOS

Aos meus orientadores, Prof^{za} Clarice Garcia Borges Demétrio e Prof. Dalton Francisco de Andrade, a orientação, o apoio e a amizade.

Aos professores e funcionários do Departamento de Estatística e Matemática Aplicada da Universidade Federal do Ceará, a amizade, o grande incentivo e a confiança em mim depositada.

Aos professores e funcionários do Departamento de Ciências Exatas da ESALQ/USP, a atenção e a amizade.

Ao amigo Padre Carlos Antônio Alcântara de Lima, as orações, a constante presença e a assistência espiritual dada a mim e à minha família.

A Genevile Carife Bergamo, a amizade e o companheirismo ao longo do curso e a convivência alegre durante os tempos de república.

Aos demais amigos e colegas de turma, David José Miquelutti, Denise Nunes Viola, Elisabeth Strapasson, José Carlos Fogo e Pedro Ferreira Filho, a partilha de conhecimentos, a solidariedade nos momentos de angústia e todos os momentos de alegria e descontração.

A todos os amigos dos cursos de mestrado e doutorado do Departamento de Ciências Exatas da ESALQ/USP, que se tornaram a minha família Piracicabana.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenadoria para o Aperfeiçoamento de Pessoal de Nível Superior (CAPES) o apoio financeiro em forma de bolsa.

A todos que, de forma direta ou indireta, contribuíram para a realização deste trabalho.

SUMÁRIO

RESUMO	7
ABSTRACT	8
LISTA DE FIGURAS	9
LISTA DE TABELAS	12
1 INTRODUÇÃO	14
2 DESENVOLVIMENTO	16
2.1 Revisão de Literatura	16
2.1.1 Modelos clássicos para dados longitudinais	16
2.1.2 Modelos generalizados para dados longitudinais	19
2.1.3 Distribuição Poisson Multivariada	22
2.2 Metodologia	25
2.2.1 Material	25
2.2.2 Métodos	27
2.2.2.1 Extensão de Karlis e Meligkotsidou (2005) para acomodar o modelo não linear	27
2.2.2.2 Extensão para delineamento em blocos casualizados com efeitos fixos de bloco	30
2.2.2.3 Extensão para delineamento em blocos casualizados com efeitos aleatórios de bloco	30
2.2.2.4 Modelagem da superdispersão via distribuição Poisson multivariada mista	33
2.2.2.5 Simulação	36
2.2.2.6 Análise dos dados de número de folhas em bromélias	39
2.3 Resultados	40
2.3.1 Simulação	40
2.3.1.1 Delineamento completamente casualizado	40
2.3.1.2 Delineamento em blocos casualizados com efeitos fixos de bloco	41
2.3.1.3 Delineamento em blocos casualizados com efeitos aleatórios de bloco	42
2.3.1.4 Delineamento completamente casualizado com superdispersão	44
2.3.1.5 Delineamento com efeitos fixos de blocos na presença de superdispersão	45
2.3.2 Análise dos dados de número de folhas em bromélias	46
2.3.2.1 Modelo com efeitos fixos de blocos	46
2.3.2.2 Modelo com efeitos aleatórios de blocos	48

	6
3 CONSIDERAÇÕES FINAIS	51
REFERÊNCIAS	53
APÊNDICES	58
ANEXOS	171

RESUMO

Modelos não lineares para dados de contagem longitudinais

Experimentos em que medidas são realizadas repetidamente na mesma unidade experimental são comuns na área agrônômica. As técnicas estatísticas utilizadas para análise de dados desses experimentos são chamadas de análises de medidas repetidas, tendo como caso particular o estudo de dados longitudinais, em que uma mesma variável resposta é observada em várias ocasiões no tempo. Além disso, o comportamento longitudinal pode seguir um padrão não linear, o que ocorre com frequência em estudos de crescimento. Também são comuns experimentos em que a variável resposta refere-se a contagem. Este trabalho abordou a modelagem de dados de contagem, obtidos a partir de experimentos com medidas repetidas ao longo do tempo, em que o comportamento longitudinal da variável resposta é não linear. A distribuição Poisson multivariada, com covariâncias iguais entre as medidas, foi utilizada de forma a considerar a dependência entre os componentes do vetor de observações de medidas repetidas em cada unidade experimental. O modelo proposto por Karlis e Meligkotsidou (2005) foi estendido para dados longitudinais provenientes de experimentos inteiramente casualizados. Modelos para experimentos em blocos casualizados, supondo-se efeitos fixos ou aleatórios para blocos, foram também propostos. A ocorrência de superdispersão foi considerada e modelada através da distribuição Poisson multivariada mista. A estimação dos parâmetros foi realizada através do método de máxima verossimilhança, via algoritmo EM. A metodologia proposta foi aplicada a dados simulados para cada uma das situações estudadas e a um conjunto de dados de um experimento em blocos casualizados em que foram observados o número de folhas de bromélias em seis instantes no tempo. O método mostrou-se eficiente na estimação dos parâmetros para o modelo considerando o delineamento completamente casualizado, inclusive na ocorrência de superdispersão, e delineamento em blocos casualizados com efeito fixo, sem superdispersão e efeito aleatório para blocos. No entanto, a estimação para o modelo que considera efeito fixo para blocos, na presença de superdispersão e para o parâmetro de variância do efeito aleatório para blocos precisa ser aprimorada.

Palavras-chave: Dados de contagem; Dados longitudinais; Modelo não linear; Distribuição Poisson multivariada; Algoritmo EM

ABSTRACT

Non linear models for count longitudinal data

Experiments in which measurements are taken in the same experimental unit are common in agriculture area. The statistical techniques used to analyse data from those experiments are called repeated measurement analysis, and longitudinal study, in which the response variable is observed along the time, is a particular case. The longitudinal behaviour can be non linear, occurring frequently in growth studies. It is also common to have experiments in which the response variable refers to count data. This work approaches the modelling of count data, obtained from experiments with repeated measurements through time, in which the response variable longitudinal behaviour is non linear. The multivariate Poisson distribution, with equal covariances between measurements, was used to consider the dependence between the components of the repeated measurement observation vector in each experimental unit. The Karlis and Meligkotsidou (2005) proposal was extended to longitudinal data obtained from completely randomized. Models for randomized blocks experiments, assuming fixed or random effects for blocks, were also proposed. The occurrence of overdispersion was considered and modelled through mixed multivariate Poisson distribution. The parameter estimation was done using maximum likelihood method, via EM algorithm. The methodology was applied to simulated data for all the cases studied and to a data set from a randomized block experiment in which the number of Bromeliads leaves were observed through six instants in time. The method was efficient to estimate the parameters for the completely randomized experiment, including the occurrence of overdispersion, and for the randomized blocks experiments assuming fixed effect, with no overdispersion, and random effect for blocks. The estimation for the model that considers fixed effect for block, with overdispersion and for the variance parameters of the random effect for blocks must be improved.

Keywords: Count data; longitudinal data; non linear model; Multivariate Poisson distribution; EM Algorithm

LISTA DE FIGURAS

Figura 1 - Número (média de 4 blocos) de folhas em bromélias por tratamento ao longo do tempo	26
Figura 2 - Curva teórica (média de cinco blocos) utilizada na simulação em cada um dos tratamentos	38
Figura 3 - Funções ajustadas e médias de quatro blocos observadas para os cinco tratamentos em cada tempo - Efeitos fixos de blocos	48
Figura 4 - Funções ajustadas e médias de quatro blocos observadas para os cinco tratamentos em cada tempo - Efeitos aleatórios de blocos	49
Figura 5 - Histograma representativo das estimativas para os parâmetros da curva logística e de covariância do Tratamento 1 - Delineamento completamente aleatorizado	181
Figura 6 - Histograma representativo das estimativas para os parâmetros da curva logística e de covariância do Tratamento 2 - Delineamento completamente aleatorizado	181
Figura 7 - Histograma representativo das estimativas para os parâmetros da curva logística e de covariância do Tratamento 3 - Delineamento completamente aleatorizado	182
Figura 8 - Histograma representativo das estimativas para os parâmetros da curva logística e de covariância do Tratamento 4 - Delineamento completamente aleatorizado	182
Figura 9 - Histograma representativo das estimativas para os parâmetros da curva logística e de covariância do Tratamento 1 - Delineamento em blocos casualizados com efeitos fixos de bloco	183
Figura 10 - Histograma representativo das estimativas para os parâmetros da curva logística e de covariância do Tratamento 2 - Delineamento em blocos casualizados com efeitos fixos de bloco	183
Figura 11 - Histograma representativo das estimativas para os parâmetros da curva logística e de covariância do Tratamento 3 - Delineamento em blocos casualizados com efeitos fixos de bloco	184

Gracias por visitar este Libro Electrónico

Puedes leer la versión completa de este libro electrónico en diferentes formatos:

- HTML(Gratis / Disponible a todos los usuarios)
- PDF / TXT(Disponible a miembros V.I.P. Los miembros con una membresía básica pueden acceder hasta 5 libros electrónicos en formato PDF/TXT durante el mes.)
- Epub y Mobipocket (Exclusivos para miembros V.I.P.)

Para descargar este libro completo, tan solo seleccione el formato deseado, abajo:

